De Novo Design of Nanopore-like Architectures and Applications of Nanopores to Investigate the SARS-CoV-2 helicase

Sinduja Karl Marx

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

Doctor of Philosophy

University of Washington

2021

Reading Committee:
David Baker, Chair
Jens Gundlach, Chair
Philip Bradley
William Zagotta

Program Authorized to Offer Degree:
Molecular Engineering and Sciences

©Copyright 2021
Sinduja Karl Marx

University of Washington

Abstract

De novo design of nanopore-like architectures and applications of nanopores to investigate the SARS CoV-2 helicase

Sinduja Karl Marx

Chair of the Supervisory Committee:
David Baker
Department of Biochemistry

Jens Gundlach
Department of Physics

Protein nanopores form nanometer-sized holes in a lipid bilayer and are widely used as analytical tools to study the flow of ions and other target molecules through them. Recent advancements in nanopore technologies have enabled use of nanopores for nucleic acid sequencing, peptide or protein analysis, and small-molecule identification and filtration. Custom designed nanopores are a promising area of research that may lead to increased accuracy and specificity of nanopore assays, or enable assays that are not possible or otherwise unknown with the existing library of protein nanopores. Designing protein structures *de novo* requires two steps: designing protein backbones that will result in the desired function, and identifying amino acid sequences that will fold into the desired backbones at their lowest energy states. In chapter 1, I introduce the geometrical and energetic considerations in building protein backbones, with a central channel and a defined sensing region, for the membrane. Using the Rosetta framework, I present *de novo* design strategies of oligomeric aqueous α-helical channels around an ion-channel filter and transmembrane β-barrels. I show the results of biochemical validation of hexameric aqueous α-helical channels, and the ion-conductance properties of the first *de novo* designed 8-stranded transmembrane β-barrel. This work is a step towards reducing our depen-

dence on naturally occurring ion-channels that are challenging to modify, and towards creating entirely *de novo* nanopores with an un-naturally optimal fit for their targets. In chapter 2, I present the application of a naturally occuring nanopore to investigate the helicase found in Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Non-structural protein 13 (nsp 13) helicase is involved in the replication of the SARS-CoV-2 genome and a potential drug target for antivirals. Single-molecule Picometer Resolution Nanopore Tweezers (SPRNT) is a new technique that uses MspA, a biological nanopore, to monitor the movement of single enzyme molecules on DNA at high spatiotemporal resolution. Using SPRNT, I present the first single-molecule assay to monitor the kinetics of single-nucleotide steps of the SARS-CoV-2 helicase, nsp13. This assay provides a single-molecule platform for understanding how antiviral compounds affect nsp13 function and glean insights for their development.

Table of Contents

List of Figures

ACKNOWLEDGEMENTS

First and foremost I thank my supervisors, Professor David Baker and Professor Jens Gundlach for the incredible opportunity to pursue my PhD research in the fields that they have pioneered. David has taught me to consider several trajectories of scientific inquiry in parallel when solving a problem. From Jens, I've learned to leave no hypothesis unturned when proving a result and to strive for precision. For my technical education, I owe thanks to the amazing scientists in their labs. Chunfu Xu, Gustav Oberdorfer, Scott Boyken, Peilong Lu, and TJ Brunette introduced me to *de novo* design of α-helical channels, and I am thankful for their guidance. Anastassia Vorobieva is an inspirational scientist, and I owe her thanks for spearheading the work on *de novo* β-barrels and contributing to the transmembrane β-barrel work in this thesis. I am thankful to Gyu Rie, Ivan Anishchenko and Samuel Berhanu Lemma for their contribution to the work on larger transmembrane β-barrels. Derrick Hicks and Stacey Gerben are impressive *de novo* designers, colleagues, friends and I owe them thanks for their insightful comments on my work. Jon Craig and Andrew Laszlo are extraordinary scientific mentors and I owe them thanks for getting me stoked about single-molecule enzymology and for their support in shaping and critiquing my results. My special thanks to Andrew for launching the single-molecule efforts in SARS-CoV-2 enzymes and his guidance on the nsp13 work in this thesis. I thank Keith J. Mickolajczyk at the Rockefeller University for the SARS-CoV-2 nsp13 protein. Last but not least, I owe sincere thanks to Chris Kim, Jesse Huang and Sarah Abell for collecting the single-molecule data on nsp13 through the Covid-19 pandemic. This work was supported through the National Institutes of Health, National Genome Research Institute (NHGRI) COVID-19 Grant, number R01HG005115.

I owe thanks to Paul Neubert and Christine Luscombe for their guidance throughout my graduate school career. I also thank the 2015 MoleS Cohort and Baker Lab cohort who have made this journey so enjoyable and memorable. I am thankful to my family, especially my parents, Karl and Indu for both fostering and challenging my interests. My in-laws Kaino and Marvin Hopper have been a source of unwavering support and strength as well.

DEDICATION

For Matthew K. Hopper

Chapter 1

De novo design of nanopore-like architectures

1.1 Introduction

Nanopore sensing

A nanopore is a nanometer-sized hole in an insulating material. Ions and charged molecules can be driven through the pore by electrophoresis. These molecules alter the flow of the ionic current through the pore depending on their shape, charge and interactions with the pore. The amplitude and duration of these current blockades can be used to identify them.

In DNA sequencing by a nanopore, single-stranded DNA molecule (1-nm wide) is threaded through the constriction (0.5 nm by 1.2 nm wide) of Mycobacterium smegmatis porin A (MspA) [1]. The nucleotides fit through the constriction of the pore in such a way as to tell the difference between sets of 3-4 nucleotides [2–4]. MspA is exceptionally well suited for this measurement as the constriction closely matches the size of DNA. Similarly, polypeptides occupy different regions in the v-shaped Fragaceatoxin C (FraC) nanopore, 6-nm wide, creating recognizable current blockades. On the other hand, unfolded polypeptides translocate through the constriction of FraC, 1.5 nm wide, and block the pore in a manner that corresponds to their difference in molecular weight [5].

Engineered nanopores

While MspA's fortuitous geometry allows sensitive recognition of DNA bases, the wild-type (WT) pore has been engineered to favor interaction with negatively charged DNA. But to achieve single base resolution, better control of the chemical and physical properties of the constriction zone of MspA are needed. MspA is a symmetric homo-octamer β-barrel and an assymetric constriction site can be engineered in a single chain version. However, single chain constructs misfolded and only a dimer subunit could be engineered [6]. Similarly, charge environment at the constriction of FraC was tuned to allow the translocation of oppositely charged polypeptides. So far, the geometry and resilience to mutation of these pores has allowed them to be used as a nanopore. In some cases, drastic structural changes such as removing domains or varing the number of subunits has been possible [7,8]. Still, it is not always easy to anticipate how changes to a protein's amino acid sequence might influence its stability, structure and function. The complex networks of amino acid interactions have evolved through eons of natural selection. Thus, enhancing existing nanopore technology, or accessing new functions is limited by available channel conformations.

De novo design

De novo protein design provides a new way to investigate the determinants of protein folding and stability on a global scale [9]. Modifying proteins from nature usually has the problem of having to discern which amino acids support function, lead to fragility and have resistance to modification or change. Though strenuous to create, *de novo* designed proteins can eliminate this problem and provide a minimal framework on which to build function. Ultimately, a tractable design process for nanopores is one in which the roles of each individual amino acid are more easily determined and controlled [10,11]. The *de novo* design of membrane proteins lags behind that of water-soluble globular proteins. This is due to our gaps in knowledge for the determinants of protein folding in a membrane environment and the experimental difficulty associated with studying membrane proteins in lipids or detergents. So far, *de novo* design of membrane nanopores has been limited to the design of short α-helical, self-assembling oligomers [12,13]. This is owing to the designability of coiled coils which are analgous to their membrane spanning counterparts. Mahendran et al. (2017) demonstrated the spontaneous formation of an octameric transmembrane ion channel from a 35 amino-acid peptide sequence modified from the membrane domain of Wza pore. Notably, they observed that the peptide channel was dynamic with a 'precursor', low and high conductance state in single channel experiments. Joh et al. (2014) designed a *de novo* four-helix bundle held together by hydrophobic interactions and a central channel lined by negatively charged residues and histidines. They demonstrated that the helices self-assemble in membrane and dynamically conduct first row transition metal ions (Zn^{2+}, Co^{2+}) but not Ca^{2+}.

De novo design strategies for nanopores

Membrane channels inhabit three chemically different environments simultaneously - hydrophobic lipid bilayer, lipid-water interface, and polar water. The outer surface of the protein is surrounded by hydrophobic carbon chains of phospholipids, steroids and other membrane proteins. In the membrane core, there is no opportunity for hydrogen bonding between non-polar lipid tails and the protein backbone; thus, intramolecular hydrogen bonds contribute greatly to stable conformation of membrane proteins. Next, at the lipid-water interface there is a diverse environment with opportunities for dipole-dipole interactions and hydrogen bonding. Finally, membrane channels could contain a passage large enough for water or ions through them and also large domains that extend into the water surrounding the membrane.

The two principal structural motifs found in membrane ion channels are α-helical bundles and β-barrels, as shown in Figure 1.1. α-helical bundles are oligomers of transmembrane helices that are tightly associated via van der waals and inter-helix hydrogen bonding between pairs of polar residues. β-barrels are single closed β-sheets consisting of an even number of anti-parallel β-strands. In this conformation, neighboring strands are connected by intramolecular hydrogen bonds between the hydrogen bond donors and acceptors in the polypeptide backbone.

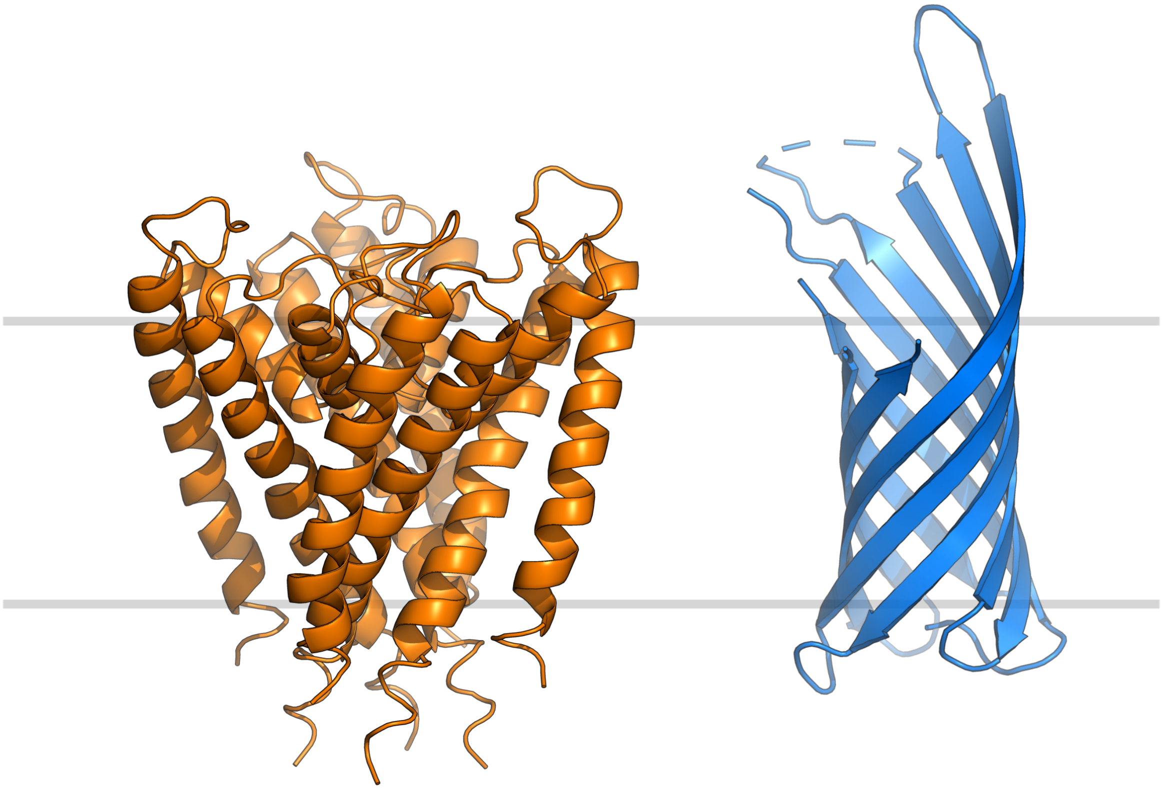

Figure 1.1: Two types of membrane architecture: An α-helical bundle (Orange, PDB ID: 1BL8) and a β-barrel (Blue, PDB: 1QJP). Membrane boundary is shaded in gray.

At the minimum, a nanopore spontaneously assembles in membrane and forms a rigid channel preferrably with a constriction. While the α-helical bundles can form a v-shaped geometry, β-barrel proteins are limited to a cylindrical lumen [14]. In exceptional cases, like MspA (PDB ID: 1UUN), oligomeric β-sheet proteins with multiple domains assemble into a funnel geometry [15]. Spontaneous insertion of membrane proteins in lipid bilayers in very dilute solutions have been reported in the case of many monomeric and oliogmeric outer membrane β-barrel proteins [1,2,16–18] and a few membrane α-helical peptide barrels [5,12,13]. Typically, both require the assistance of protein folding machinery or liposome-mediated insertion into planar bilayers [19–22]. The specificity and stability of helix associations in membrane is attributed to Van der Waals interactions, intrahelical backbone–backbone and interhelical backbone–sidechain hydrogen bonds [23–26]. On the other hand, the stability of β-barrels in membrane is driven by the extensive intrastrand hydrogen bond network [27]. β-barrel subunit specificity can be forced with hydrophobic interactions in the soluble domains [28] or weakly-stable regions in the membrane domain [29].

In this research, we sought to apply the above principles to *de novo* design nanopores. Recently, Rosetta software suite has been used to *de novo* design coiled-coil helical bundles, symmetric α-helical oligomers, and multi-pass transmembrane helices [30–33]. Using a combination of these established computational approaches for α-helical motifs, we proposed to develop *de novo* design methods for α-helical channels. Also, the first *de novo* design of water-soluble barrels was recently demonstrated by Dou et al. (2018). Building on the principles of soluble barrels, we aimed to develop *de novo* design methods for membrane β-barrels.

1.2 Results

1.2.1 PROCEDURAL GENERATION OF SYMMETRIC α-HELICAL CHANNELS

The first step in *de novo* design is the generation of protein backbones in the desired conformation. The second step is the design of low-energy amino acid sequences for these desired backbones. The first challenge of designing α-helical proteins for the membrane is the similarity of the hydrophobic core of the protein to the surface. This high hydrophobic content of membrane proteins makes them insoluble in aqueous solutions and is the second challenge to efficiently purifying and studying the designed proteins. To overcome these challenges, we aimed to first design and experimentally validate desirable channel backbones in aqueous solution and then optimize their sequences for membrane. Further, to simplify the sequence space for large proteins we create oligomers of subunits with rotational or cyclic symmetry (C_n). An anti-parallel 3-helix bundle with N- and C- termini facing opposite ends was chosen as the subunit for the cyclic assemblies. The hydrophobic core of lipid bilayers are often 2-5 nm thick. As a result, to span the membrane, the length of the helices were set in the range of 35 amino acids. Next, we set coordinate constraints within the channel to define the constriction for a particular function, followed by enumerating all possible protein backbones around it. In Rosetta, new backbones are stochastically assembled with short peptide fragments from the Protein Data Bank (PDB) and are guided by the defined blueprint of the structure. The computational design strategy for a C6-symmetric α-helical channel is described in Figure 1.2.

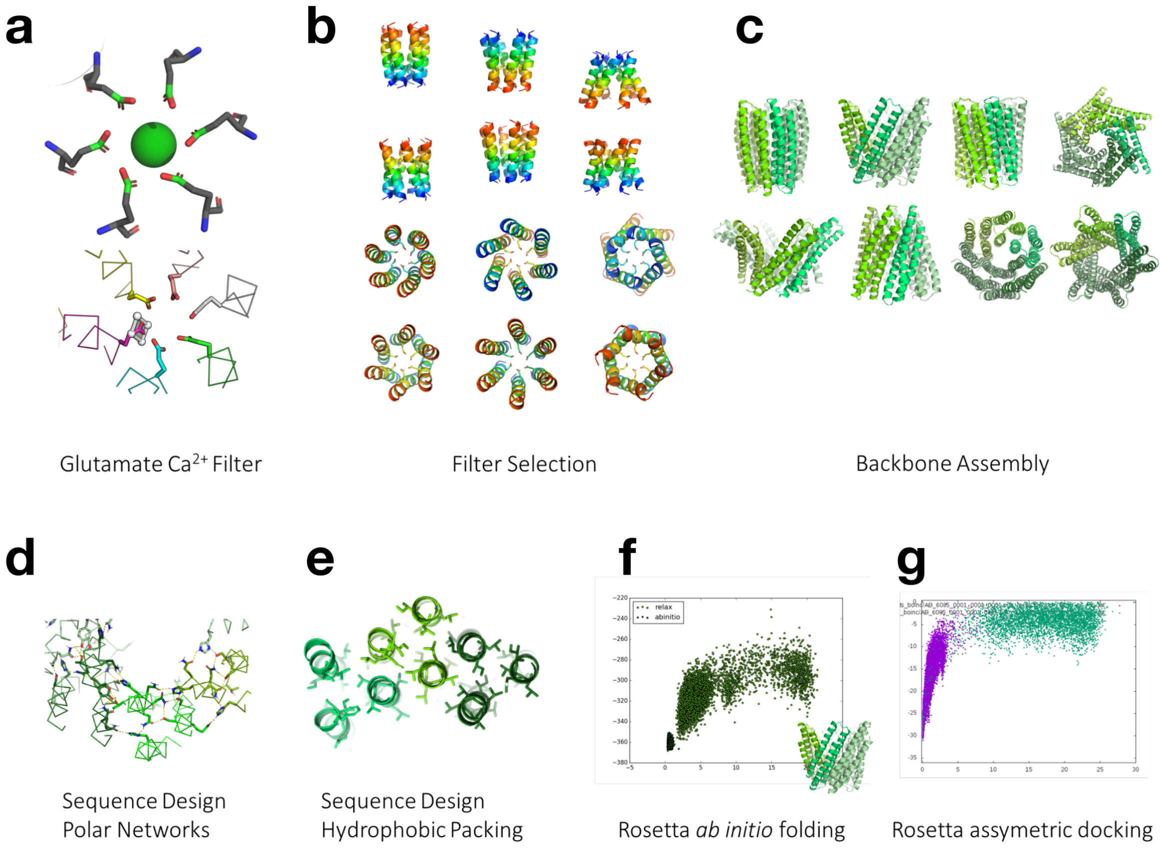

Figure 1.2: Computational design method for C6-symmetric α-helical channel. (a) A hexameric pseudo-symmetric glutamate filter is chosen from the native calcium release-activated channel (CRAC, PDB ID: 4HKS). The atoms CD, OE1 and OE2 on the glutamate residues that co-ordinate Ca^{2+} ion (green sphere) are fixed in orientation (shown as a plane). Possible glutamate residues in this conformation are generated. (b) All possible constrictions of the channel are sampled by housing the glutamate residue anywhere along an ideal 10 amino acid helix in a C6 symmetric fashion. (c) Each helix is a seed for an antiparallel 3-helix bundle built from fragments from the PDB according to a blueprint with varying helix and loop lengths. This phase is used to exhaustively search for compact backbones and diverse geometries in Rosetta's centroid mode. Side and top-down view of example backbones are shown. (d) Backbones are searched for potential hydrogen-bond networks between pairs of residues both within and across chains using Rosetta HBNet protocol. (e) Rosetta Full-atom sequence design is performed on fixed backbones and designed proteins are filtered for satisfaction of buried polar groups and hydrophobic packing. (f) & (g) Designed proteins are further filtered using Rosetta ab initio structure prediction and Rosetta assymetric docking protocol. Sequences that fold back into designed 3-helix bundle and form unique C6 symmetric assemblies are selected for experimental screening.

1.2.2 EXPERIMENTAL CHARACTERIZATION AND DESIGN ITERATION OF HEXAMERIC α-HELICAL CHANNELS

Designed backbones with good interface shape complementarity, satisfaction of buried polar groups and hydrophobic packing are chosen for experimental screening. A small-scale experimental screen was conducted by testing 6 designs sampling a few channel geometries and oligomeric interfaces. Synthetic genes that code for the designed proteins were transformed into *E. coli.* 3/6 designs expressed with very low yield and were prone to aggregation at high concentrations. We iterated on one of the designs, hereafter referred to as CR5, by optimizing sequence and by modifying the polar networks at oligomer interfaces (Figure 1.3 and Figure 1.4). CR5 variants 3, 5 and 7 expressed solubly and in 10-fold higher yield than the original (Figure 1.5 a. and Figure 1.6 a.). Notably, these designs did not carry histidines in the channel lumen (Figure 1.4 b.). Native mass spectrometry analysis revealed that designs CR5_var_3, and CR5_var_7 were predominantly hexamers and design CR5_var_5 was a quadromer (Figure 1.6 b.). From a qualitative design analysis, CR_var_3 and CR_var_7 retain more residues from the original designed hydrogen bond network between the chains (where a network is more than 2 residues involved in hydrogen bonding), while CR_5 has no network (Figure 1.4 a.). These data suggest that the polar networks at the subunit interfaces can alter subunit specificity. Initial crystallization screens showed a positive crystal hit for two of the design variants, but poor diffraction. This suggests that while the sequence designs for the CR5 channel are close, further sequence optimization was necessary.

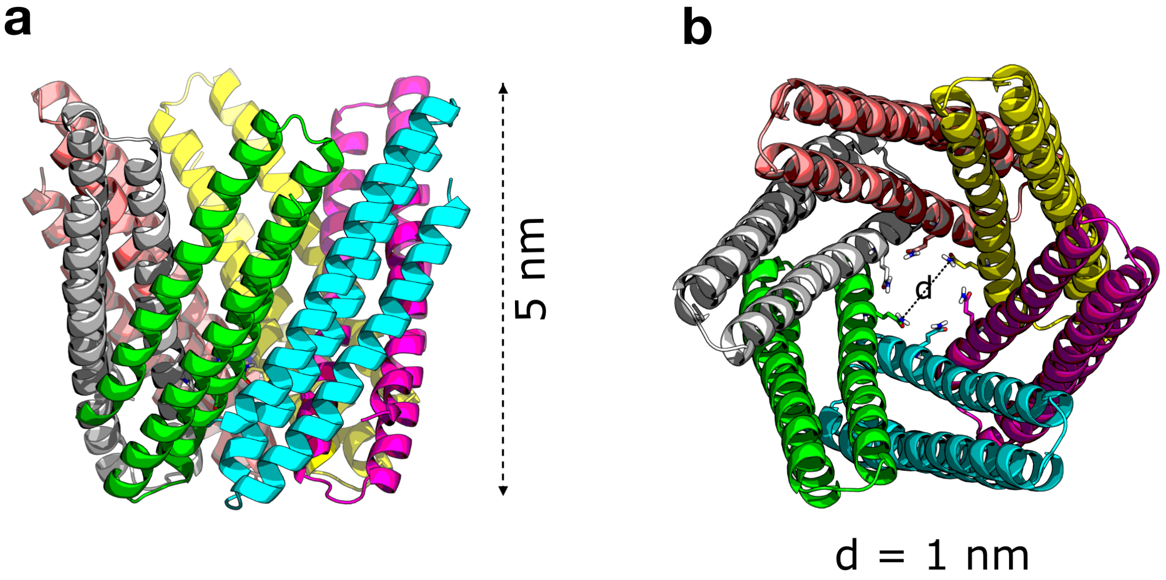

Figure 1.3: Computational model of *de novo* designed hexameric α-helical channel "CR5". (a) & (b) CR5 side and top-down view respectively.

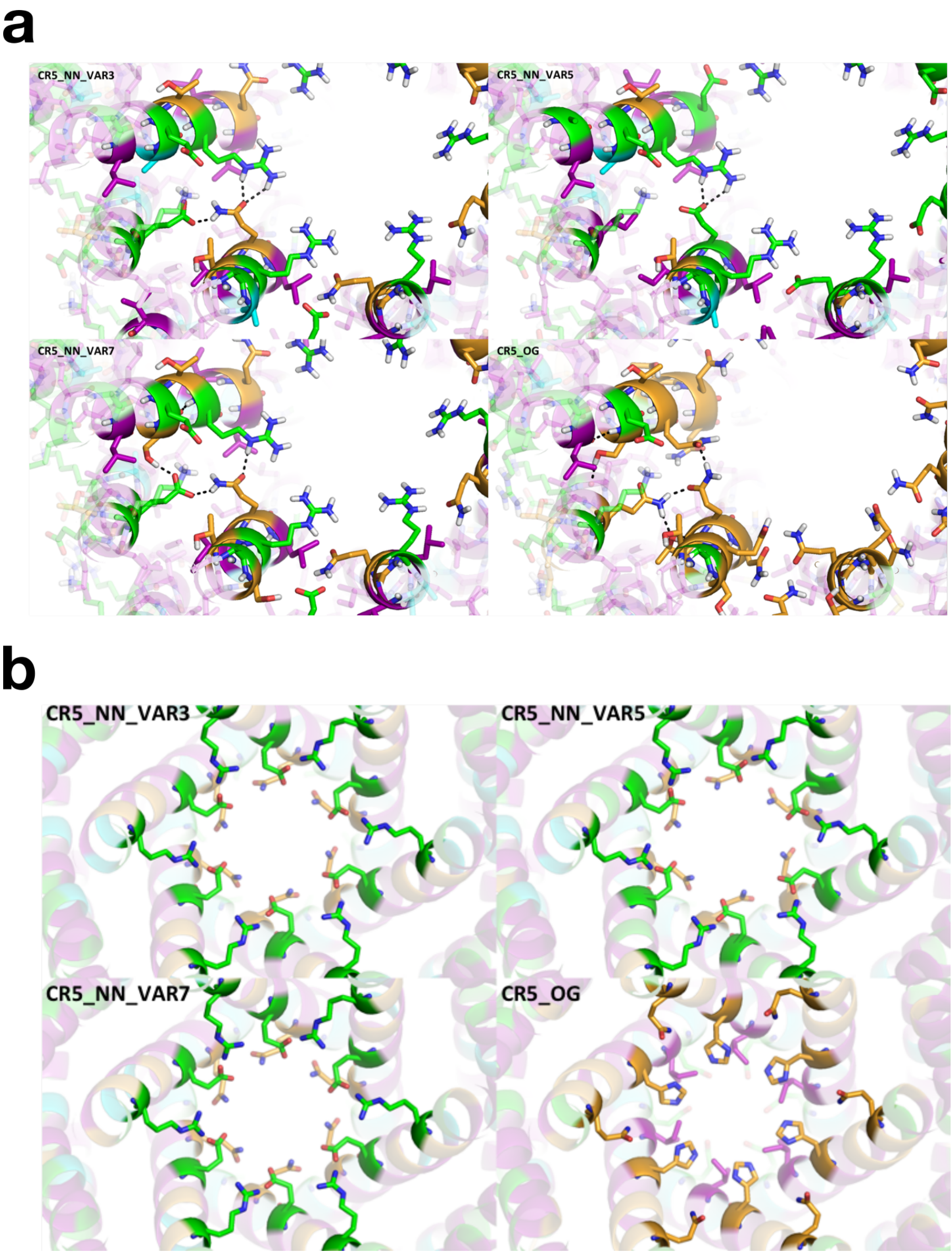

Figure 1.4: Variants of hexameric α-helical channel “CR5”.
CR5 variant no. 3, 5, 7, and original CR5 are shown clockwise. (a) Differences in polar networks between CR5 variants. (b) Histidines in original design were replaced.

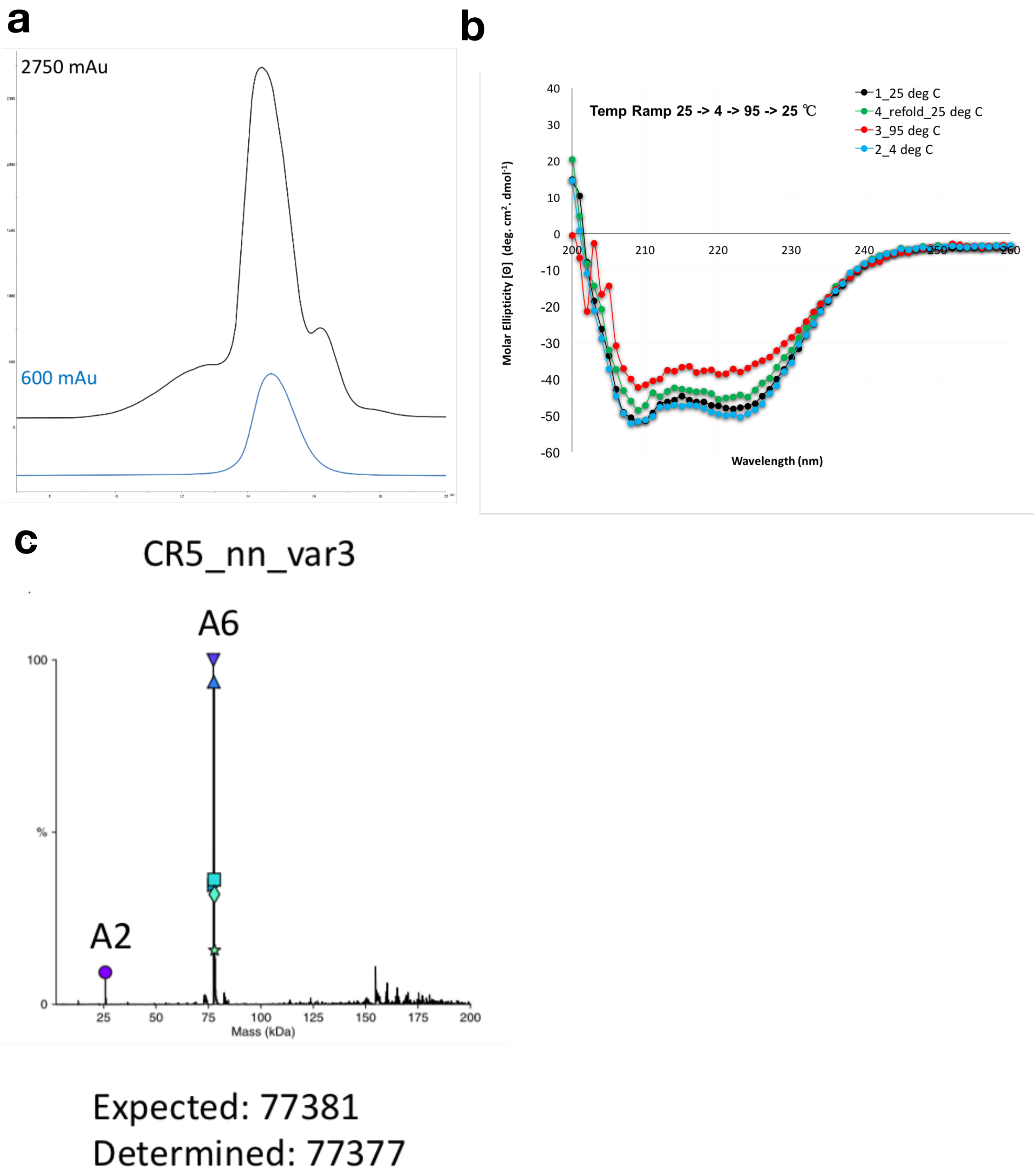

Figure 1.5: Biophysical characterization of a hexameric α-helical channel.
(a) Size exlusion profile of original design ("CR5", blue) and improved design ("CR5_nn_var3", black) shows the latter has greater solubility and yield. Further characterization of CR5_nn_var3 to determine its secondary structure and oligomer number are shown in (b) and (c). (b) Circular dichroism spectra shows that the protein has an α-helix secondary structure in solution and has a high thermal stability of upto 95○C. (c) Native mass spectrometry analysis shows that the protein assembles predominantly as a hexamer. Expected and determined molecular weight in Daltons are listed.

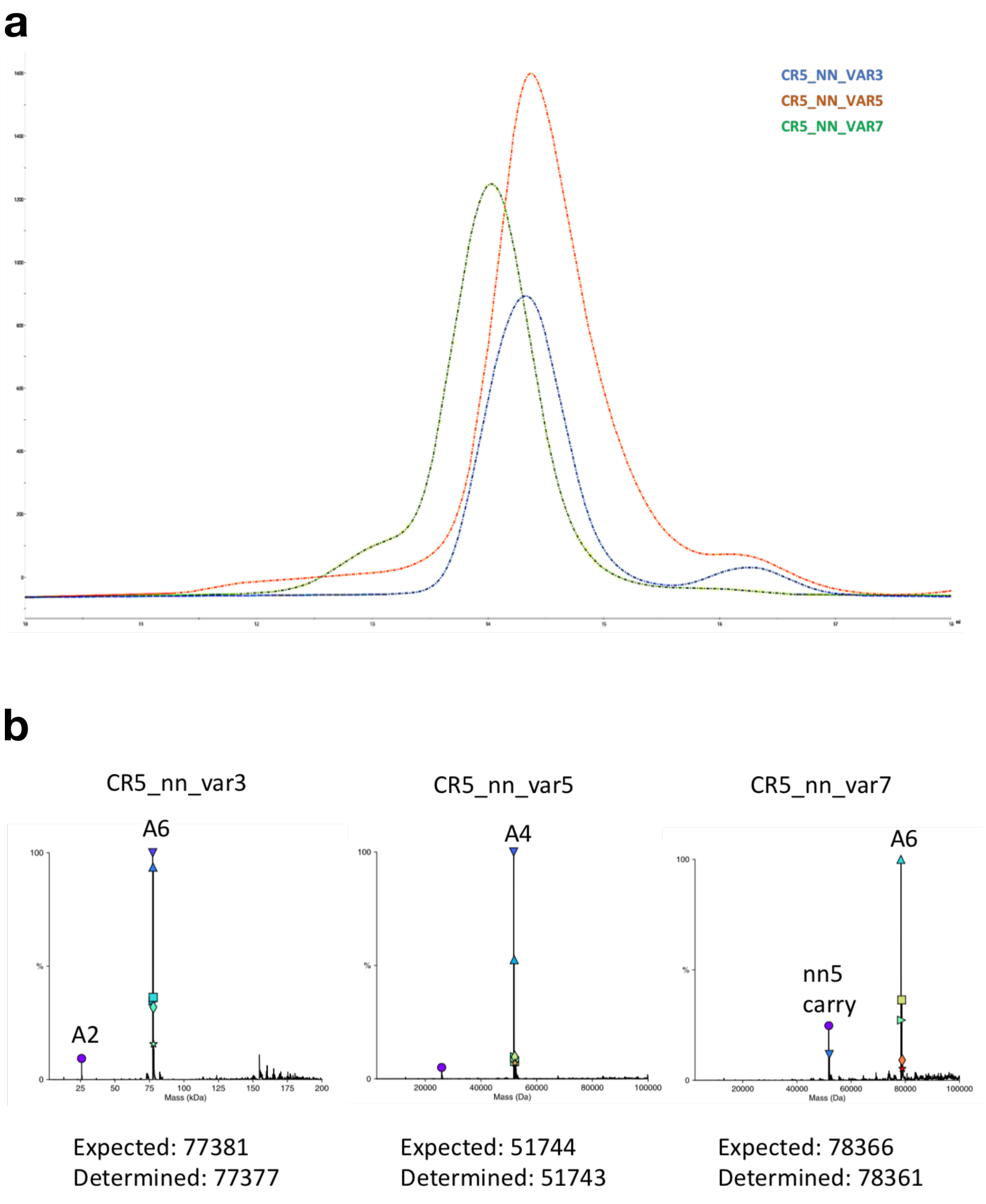

Figure 1.6: Biophysical characterization of variants of hexameric α-helical channel "CR5". (a) Size exclusion profile of three CR5 variants. (b) Native mass spectrometry analysis of oligomer number of each of the CR5 variants. CR5 variant 3 and 7 are predominantly hexamers, and variant 5 is quadromer.

1.3 Discussion

In sum, we were able to use a design-test-hypothesis approach to iteratively develop a *de novo* design protocol for aqueous hexameric α-helical channels. We find that a fragment assembly based approach gives access to funnel shaped alpha-helical pore geometries. The computational protocol was straightforward to extend to design channels with C4 and C5 symmetries, accessing a variety of pore geometries and sizes (data not shown). While assemblies larger than C6 are possible using this approach, it was significantly more challenging to design oligomeric interfaces against alternative higher or lower order assemblies. The experimental 'testing' phase would benefit from screening more sequences from widely different backbone clusters. Among the different backbones generated, the oligomeric interface ranged from large uninterrupted buried surface areas between subunits to only a few hydrophobic interactions. The oligomeric interface of CR5 had high shape complementarity with 2 patches of hydrophobic packing interspersed with plausible polar interactions. As a result, the hypothesis testing was limited to optimizing polar networks at the oligomer interfaces. This design protocol is also ripe for systematically deriving the principles of oligomer complex formation as a function of the buried surface area and molecular weight (size) of the protein. Finally, structurally validated CR5 like channels can serve as scaffolds for their membrane counterparts.

1.4 Results

1.4.1 *De novo* DESIGN OF 8-STRANDED TRANSMEMBRANE β-BARREL (TMB)

The principles behind the *de novo* design of a transmembrane β-barrel was published on February 19, 2021 in Science [34]. I worked with Anastassia Vorobieva, and Stacey Gerben on the Rosetta *de novo* design of the 8-stranded transmembrane β-barrels. For the detailed computational and experimental protocol, I refer the readers to the article's Supplementary Materials and Methods. In brief, the simplest transmembrane β-barrel (TMB) architecture was designed and used as a model to understand the determinants of TMB folding. For the 23 proteins that passed biochemical screens, suggesting that they may fold into a TMB, we probed them for function in a single-molecule study (described in next section). Among them, TMB 2_17, shown in Figure 1.7, was the most successful. Notably, TMB 2_17 also crystallized and the structure (2.05 resolution) was found to be identical to that of the designed model with a 1.1 root mean square deviation (RMSD) between the protein backbone coordinates. This work was the first demonstration of a *de novo* designed transmembrane β-barrel, and validation of a *de novo* design strategy that accesses high sequence and shape diversity.

Figure 1.7: TMB 2_17 Crystal Structure.
Crystal structure of TMB 2_17 (pink, 2.05 resolution) superimposed on designed (gray). Water molecules represented as crosses. The RMSD between the actual and design structure is 1.1 .

1.4.2 Conductance properties of *de novo* 8-stranded transmembrane β-barrel

Bacterial porins are known to spontaneously self-assemble in planar lipid bilayers, a property that dictates the ease of their use in nanopore technologies. To test the ability of the designed TMBs for spontaneous insertion and ion-conduction, we screened them in a single-molecule planar bilayer assay (Methods detailed in Appendix A.1). First, the detergent type (DDM, neutral maltoside; OG, nonionic; DPC, zwitterionic) influenced the frequency of channel interaction with the bilayer. Of the 23 designs, 2 designs in DPC detergent interacted with the bilayer at high frequency. Of the 2, 1 had the most consistent results. The electrical recordings recordings suggest a single-channel conductance of 24pA at +100 mV in 1M KCL (Figure 1.8 a. & b.). In Figure 1.8 a., we also observe that the conductance fluctuates between two states (5 pA and 24 pA). In 1.8 b., we observe step-wise increases and decreases in conductance of 20 pA which are likely to be additional channels in the bilayer opening, and closing. The all-points current amplitude histogram of possible single-channels, from n=3 independent TMB 217 refolding preparations, are compared with each other in Figure 1.8 c. We find that histogram peak of the single-channel conductance varies with each refolded batch; this suggests that the TMB conformation may be sensitive to potential changes in refolding conditions.

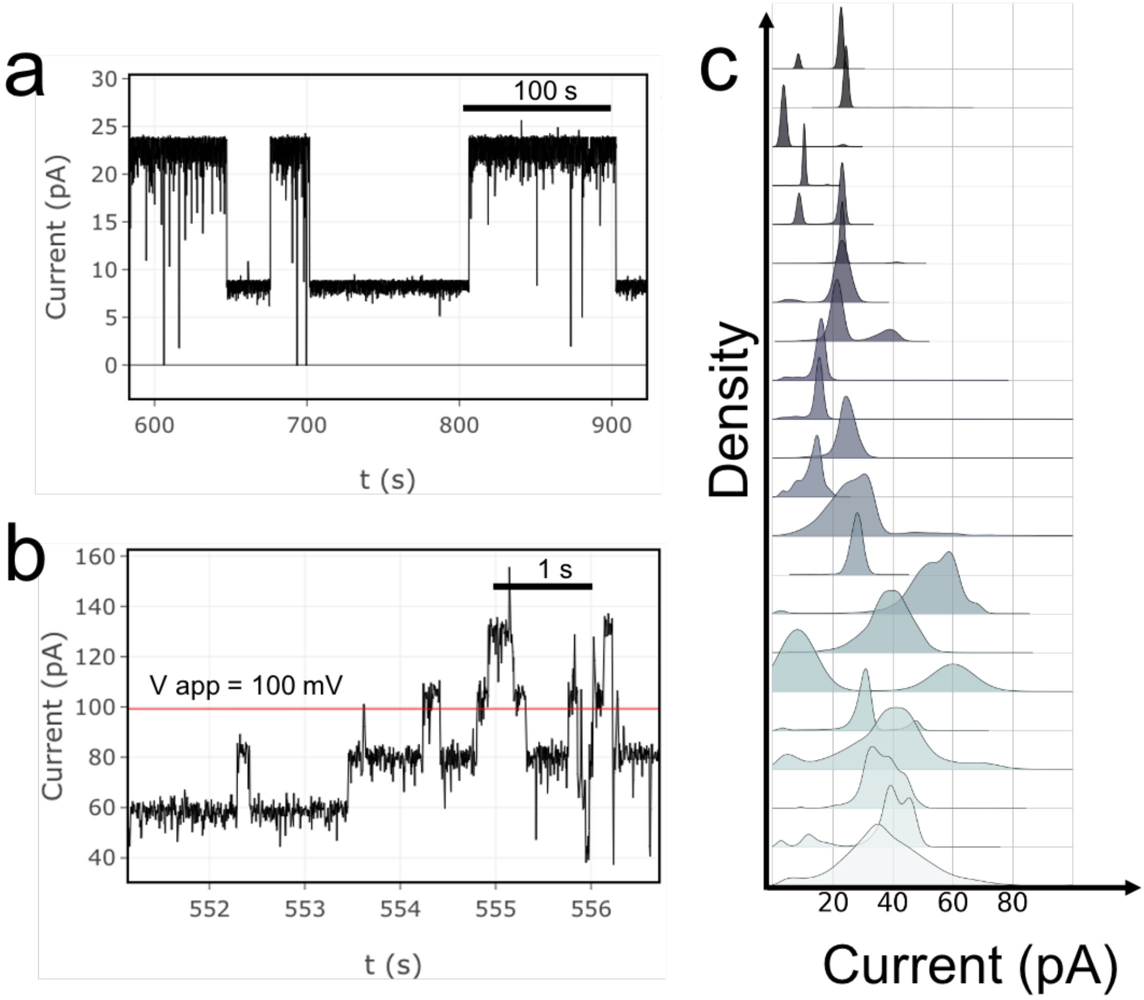

Figure 1.8: Single-channel Electrical Recordings of the TMB 2_17
When TMB 2_17 protein was reconstituted into a planar lipid bilayer, we observed discrete jump in conductance at an applied membrane potential of +100 mV in 1 M KCl, 10 mm Hepes, pH 8.0. (a) Representative electrical trace of potentially single channel of TMB 2_17 at 24 pA conductance. (b) Step-wise insertions of potentially multiple TMB 2_17. (c) All-points current amplitude histogram of individual single-channel electrical traces like in (a). β-barrel

1.5 Discussion

TMB are prepared from inclusion bodies using a rapid-dilution refolding protocol [34]. One key difference in subsequent refolding preparations was the increase in concentration of prepared proteins in detergent; changes in detergent-protein ratio can alter TMB conformation or oligomerization, which would result in non-uniform conductances. A limitation of the reconstitution process for porins in planar bilayers is the potential of proteins to aggregate or misfold before insertion [20]. The observation of two conductances (low and high) occurred in most cases together, which indicate that they are two states of one channel. Similar multi-state conductances have been reported for other porins, and in the case of native 8-stranded OmpA [16,35,36]. Larger TMBs share similar structure and sequence properties with smaller TMBs, suggesting that the design process can be extended to create larger barrels with wider channels.

Chapter 2

Investigating nsp13 using SPRNT

2.1 Introduction

Non-structural protein nsp13 is a SARS CoV-2 drug target

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) relies on its helicase, nonstructural protein 13 (nsp13) for viral replication [37–40]. Nsp13 from SARS-CoV-2 differs from that of SARS-CoV-1 (responsible for SARS 20 years earlier) by only a single amino acid (V570I). In addition, both share motifs with helicases from the larger family of nidoviruses [41]. This high degree of conservation suggests these helicases could have a shared unwinding mechanism, and thus make them an attractive drug target for broad spectrum antivirals.

Helicases are motor enzymes that couple the energy from ATP hydrolysis to unwind double-stranded nucleic acids. Previously, bulk activity assays using SARS-CoV-1 nsp13 show it unwinds both double-stranded DNA (ds DNA) and RNA (ds RNA) equally, hydrolyzing all dNTPs and NTPs in a 5'-3' direction [42]. The reported velocity for the helicase was on the order of only a few hundred bases per second ($bp - s^{-1}$), but enhanced by cooperativity and interactions with SARS-CoV's RNA-dependent RNA polymerase (RdRP), nonstructural protein 12 [42–45]. More recently, Mickolajczyk et al. (2020) biochemically characterized nsp13 from SARS-CoV-2 and estimated velocities around 5 - 10 $bp - s^{-1}$ depending on stability of nucleic acid substrate. They also use optical tweezers, a single-molecule technique, on a RNA hairpin duplex to show >50-fold nsp13 velocity enhancement from applying 10–20 pN destabilizing force on the substrate.

Viral helicases are druggable targets

Vaccines activate the human body's immune response to produce antibodies that target the outer shell of viruses and block them from entering cells. Another form of viral control are small-molecule inhibitors that target the virus' enzymes and self-replication machinery. Examples of viral inhibitors that target replication enzymes include the treatment to human immunodeficiency virus (HIV-1), hepatitis C virus (HCV) and Herpes simplex virus (HSV). Specifically, the FDA approved inhibitor Pritelivir for HSV disrupts the HSV helicase's ability to unwind its genome, thereby stopping viral replication [46]. While helicases are druggable targets, inhibitor development can be challenging without a structure and the ability to assay the mode of action of the inhibitor. Inhibitors targeting Adenosine triphosphate (ATP) hydrolysis or metal-binding domains need to be highly specific to viral helicases without affecting the host cellular enzymes with similar functions.

Viral inhibitor development is aided by mechanism of action studies

For a monomeric helicase, the sequence of reactions that translate 1 nucleotide triphosphate (NTP) binding and hydrolysis to 1 base translocation is described by the inchworm model [47]: nucleic acid binding in pocket A, NTP binding in NTP binding cleft, closure of NTP binding cleft, this conformational change causes the displacement of base from pocket A to pocket B, NTP hydrolysis, nucleotide diphosphate (NDP) and phosphate release, re-opening of NTP binding cleft and pocket A, entry of new base into pocket A, translocation by one base. In theory, an inhibitor can target any of these steps in addition to specific co-factor interactions to prevent helicase unwinding. So, understanding the course and kinetics of viral helicase unwinding can help rationally engineer different types of inhibitors. Conventionally, half-maximal inhibitory concentration (IC_{50}) values are used to evaluate how well a drug binds to and reduces enzyme activity. But, IC_{50} values are primarily a measure of binding affinity and lack information about the mechanism of inhibition. Finally, measuring the time for which a drug resides on its target has been proposed as a more robust metric for early drug selection and optimization [48].

SPRNT has 40 picometer and sub-millisecond resolution to model helicase kinetics

Single-molecule picometer resolution nanopore tweezers (SPRNT) is a new technique for monitoring the motion of individual enzymes along a nucleic acid template. In SPRNT, a single *Mycobacterium smegmatis porin A* (MspA) nanopore is established on a bilayer. Enzyme DNA complexes are drawn into the pore by an electric field until the enzyme rests on the rim of the pore. Single stranded DNA is released into the pore as the enzyme walks along the nucleic acid template. The nucleotides within the sensing region of MspA produce a unique ionic current and the duration of these currents is controlled by the enzyme. Thus, the dwell-times between transitions in current corresponds to the successive motor-enzyme states. In SPRNT, we use a known DNA sequence to measure dwell-times of ion current levels to make kinetics measurements of the controlling motor enzyme, which is used to infer mechanism of enzyme motion [49–52]. Additionally, the measured nucleic acid sequence in the pore reveals the exact sequence in the motor enzyme, thus SPRNT is also ideal for investigating how sequence affects enzyme kinetics.

The detailed biochemical pathway [52] revealing how a Super Family 2 helicase hel308 couples ATP hydolysis to ssDNA translocation is exemplary of the possibilities with SPRNT. In sum, SPRNT can resolve 40 picometer (pm) enzyme motions along a DNA track at enzyme state durations of 10 ms. Prior to SPRNT, the highest resolution was acquired with optical tweezers, which can resolve enzyme motions of ~ 300 picometer (pm) at ~ 1 second timescales.

Proposal: Develop a SPRNT assay to monitor nsp13 kinetics and establish a platform for inhibitor studies

With SPRNT's extraordinary spatialtempo resolution, we seek to explain the mechanism of SARS CoV-2 helicase nsp13 and provide a foundation for inhibitor mechanism of action studies. In this work, we establish an assay to observe individual nsp13 molecules in real-time stepping along a nucleic acid track in single steps. From these measurements, we determine the kinetic step size, ATP-coupling stoichiometry and rate of unwinding of nsp13. Next, we vary external parameters such as substrate type and [ATP] to change the rates of the chemical processes to elucidate the underlying mechanism of nsp13 unwinding kinetics. While not published with this work, we also vary applied force, temperature and sequence to further understand nsp13 mechanochemical regulation. These results enable us to understand the behavior of the broader family of nidovirus helicases and superfamily 1B helicases for the first time using a single-molecule approach. Finally, this assay forms the basis for measuring inhibitor-nsp13 binding mode and duration which are key indicators to a drug's efficacy *in vivo* [48].

2.2 Results

2.2.1 Monitoring nsp13 helicase unwinding duplex DNA base-by-base

To measure the nucleotide stepping rate we constructed nucleic acid sequences to load nsp13 and monitor the enzyme's progression on single- and double-stranded tracks in a 5' - 3' direction. First, from the loading site geometries tested, a partial DNA duplex with a 5' poly-T tail yielded full length reads at a high event frequency (Figure 2.1). This sequence and geometry is used throughout this study to understand nsp13 duplex unwinding kinetics. The same sequence without it's complementary blocking oligo performed similarly and is used to study nsp13 single-stranded translocation kinetics.

Second, we found empirically that nsp13 events were only possible in the reverse orientation of the pore, where the enzyme is in contact with residues closer to pore's constriction rather than those on the wider rim. The SPRNT schematic is illustrated in Figure 2.1. This is in contrast with all previous nanopore sequencing and SPRNT enzyme studies on MSPA.

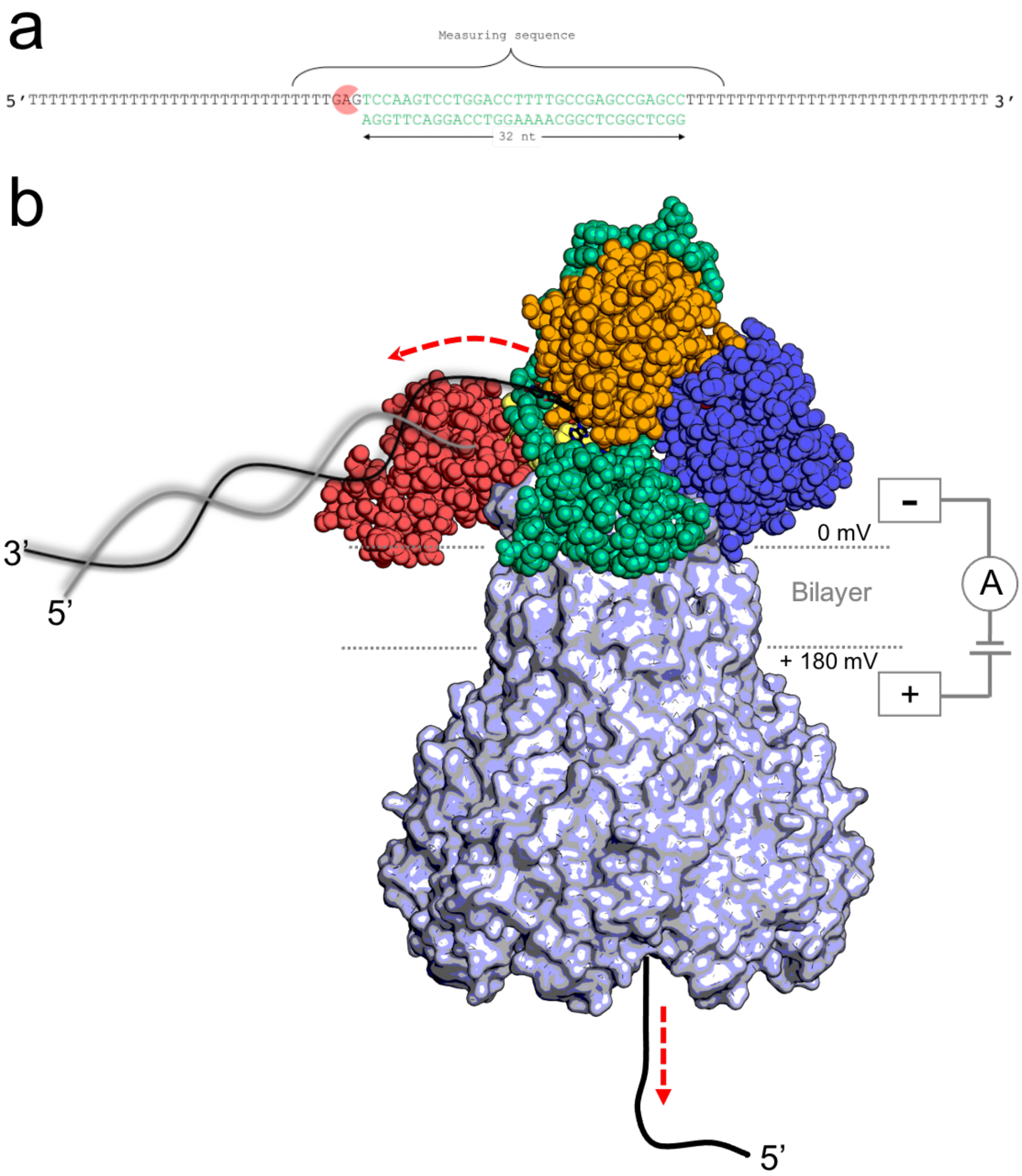

Figure 2.1: Schematic of nsp13 unwinding experiments
(a) Partial double-stranded DNA used for single-molecule nsp13 unwinding experiment. Nsp13 (red) loads on the 5' end of the duplex junction and begins unwinding in a 5' - 3' direction. (b) A single nsp13 molecule bound to partially duplex DNA is drawn into the pore by the electric field. The 5' tail threads into the pore until the enzyme comes into contact with pore. The rate of release of DNA into the pore is now controlled by the nsp13 motion on the DNA. The voltage applies a force to the DNA in the constriction, resulting in an assisting force on the enzyme when the enzyme travels in a 5' to 3' direction.

With this new experimental setup, we recorded several hundred traces of nsp13 unwinding and translocation on ds DNA at saturating ATP conditions at an applied voltage of 180 mV and saturating ATP. An example nsp13 DNA unwinding trace is shown in Figure 2.2 a. Each level in the ion-current trace yields two pieces of information - DNA sequence in the pore and the duration for which enzyme resides on one nucleotide before proceeding to the next (Figure 2.2 b). For a known DNA sequence, measuring current amplitude gives us the position on the nucleic acid track and measuring dwell-time at each position gives us the single-molecule turn-over rate (k_{cat}) of nsp13. The conversion of ion current to DNA position and dwell-time is detailed in Appendix B.3.

By comparing 54 nsp13 unwinding events, we produced a consensus set of current levels for nsp13 controlled DNA sequence through the pore, shown in Figure 2.2 c. Aligning the consensus levels to a predicted set of levels for the known DNA sequence, we deduce the orientation of the read as 5' - 3' and that the levels are spaced 1 nt apart (Figure B.1). These results show that we observe nsp13 unwinding in single-nucleotide steps.

For each of the levels observed, we computed the average median dwell-time ($\tau_{1/2}$), shown in Figure 2.2 d. With the exception of the degenerate positions, the dwell times correpond to single-nucleotide residence times. The average median dwell-time per position is 1.2 ms $\pm$ 0.4 ms at saturating ATP with an applied voltage of 180 mV. This translates to an average enzyme rate of ~ 800 nucleotides per second with assisting force of ~35 pN.

From Figure 2.2 d, $< \tau_{1/2} >$ varies between 1.9 ms to 0.4 ms depending on DNA position. This indicates that nucleotide identity within nsp13 nucleic acid binding site influences unwinding kinetics.

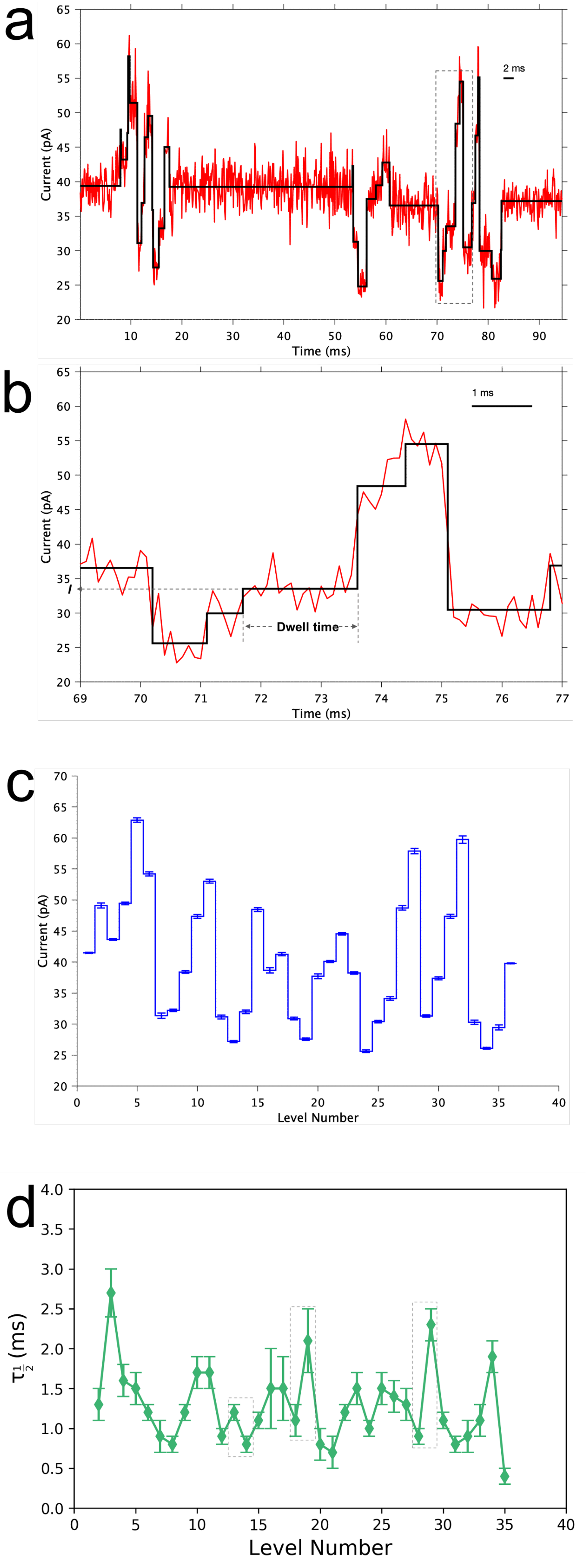

Figure 2.2: SPRNT analysis of nsp13 helicase

Figure 2.2: SPRNT analysis of nsp13 helicase
(a) Enzyme motion is observed as a time series of ion current levels. Black lines mark the average current of observed levels. (b) For each level, ion current amplitude (I) provides DNA sequence in the MspA nanopore and dwell time provides kinetics of the motor enzyme. (c) Consensus of current level partterns for 54 reads or unwinding events with nsp13. (d) Median duration of corresponding current levels in (c) at saturating ATP conditions. Degenerate current levels are boxed in gray.

2.2.2 Comparing nsp13 helicase unwinding and translocation on DNA

Next, we sought to examine the difference in kinetics between nsp13 unwinding and translocation. In Figure 2.3, we compared the median duration ($\tau_{1/2}$) of each current level for single-strand (ss) translocation events to those of double-stranded (ds) unwinding events by dividing $\tau_{1/2}$ (ds) by $\tau_{1/2}$ (ss); We found that single-stranded translocation events were on average 2-fold faster than double-stranded unwinding events. Again, we observe a fluctuating $\tau_{1/2}$ that indicates sequence dependent unwinding and translocation. The magnitude of ion current was the same for both event types.

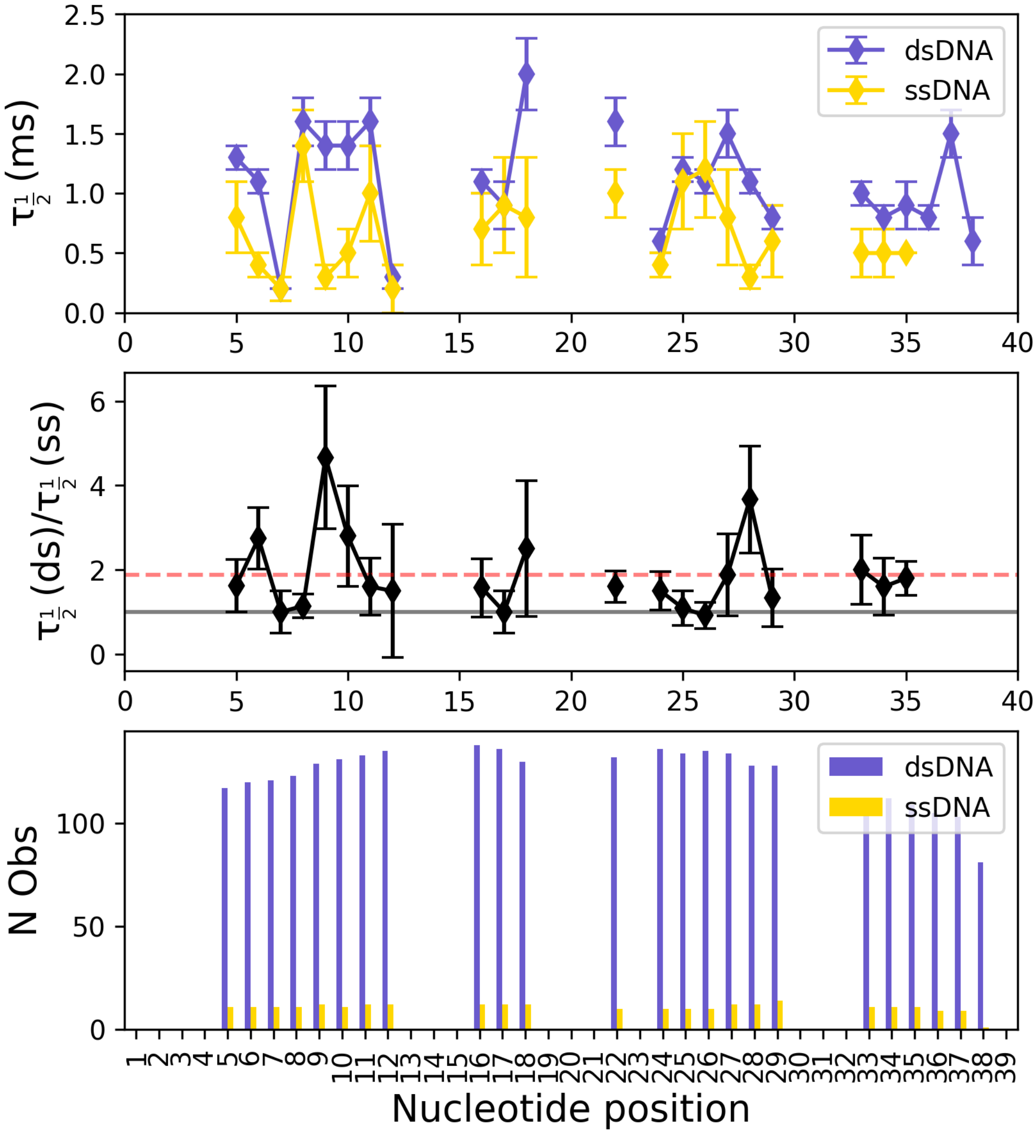

Figure 2.3: Comparing nsp13 helicase double-stranded DNA (dsDNA) unwinding to single-stranded DNA translocation (ssDNA)
(Top) Median duration of current levels for ds DNA unwinding (purple) and ss DNA translocation (yellow). (Middle) The ratio of the median durations in (top) to remove sequence dependence that also influences the level durations. (Bottom) The number of times that a given level whose durations are shown in (top) was observed. Only few ssDNA translocation experiments were performed.

2.2.3 TWO TIME-SCALES OF ENZYME KINETICS AT CONSTANT CONDITIONS

Across independent nsp13 experiments on MspA nanopore, we qualitatively observed roughly two time-scales of unwinding events. Figure a. shows a representative unwinding event from each of the two classes observed at constant conditions of 1 mM ATP, 180 mV and duplex dsDNA. To understand the average behavior of the enzymes within an experiment, we analyze the distribution of mean duration per position of each nsp13 molecule. The distributions can be roughly classified as fast-decaying exponential distribution or a combination of uniform and slow-decaying exponential distributions, as shown in Figure B.2. Combining enzyme events of the same distribution type yields a 7-fold separation in nsp13 rate under saturating ATP and identical conditions (Figure 2.4 b).

The catalytic turnover rate of helicases are ATP dependent. Thus, the single-nucleotide dwell time is ATP dependent and we examine ATP concentration as the source of this variation in enzyme rates. Figure 2.4 b shows an increase in mean dwell time per nucleotide with decrease in [ATP] in fast enzyme category and no change in the slow enzyme category. There is still a 5-fold difference in rates of the two enzyme categories. We further investigate the ATP-dependence of the two enzyme event types in the next section.

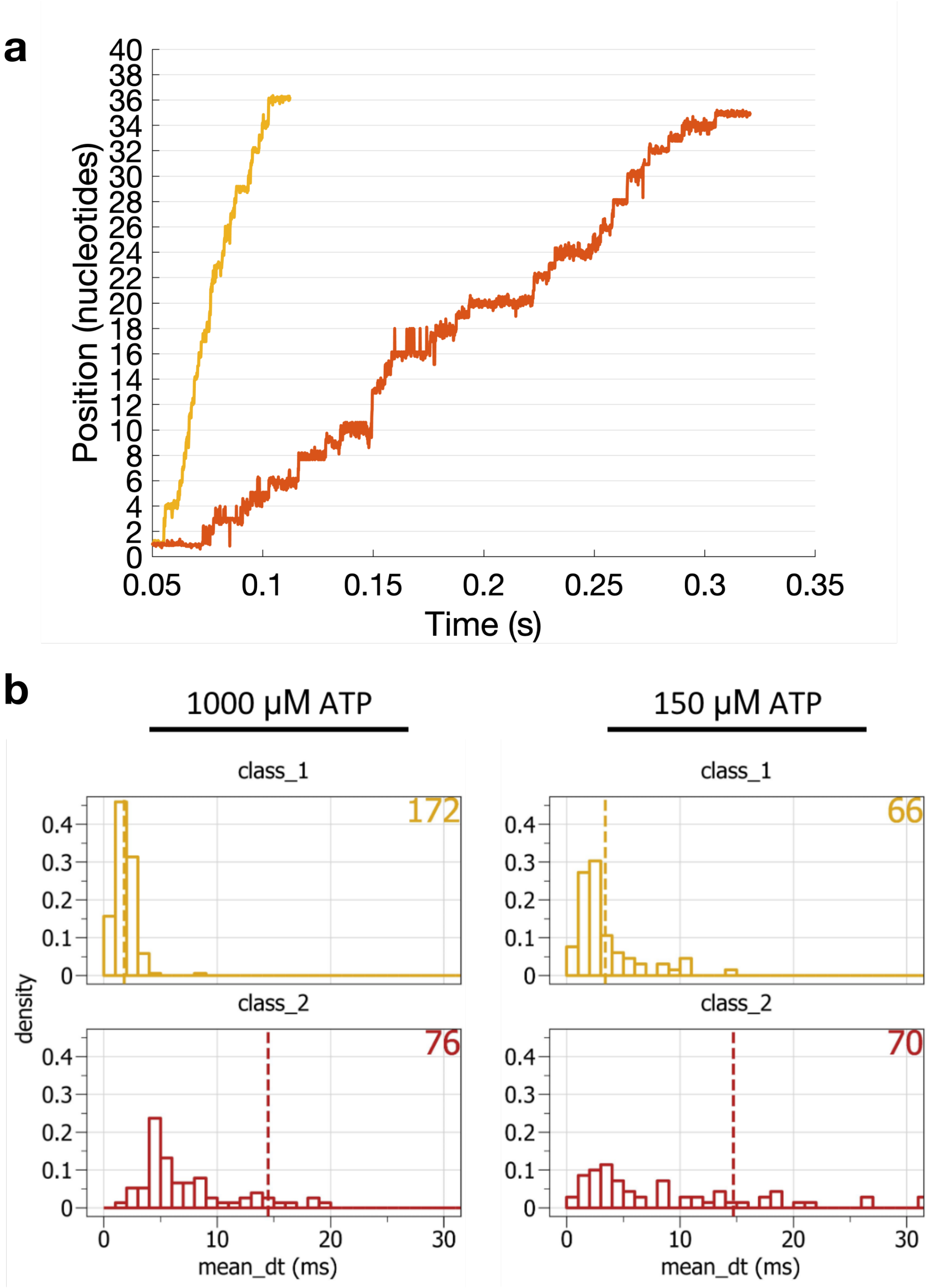

Figure 2.4: Fast and slow nsp13 unwinding events
(a) Position versus time trace for two different nsp13 unwinding events: fast class 1 event (gold) and slow class 2 event (red) (b) Distribution of average dwell time per enzyme for fast enzymes versus slow enzymes at two different ATP conditions: (Left) Saturating ATP condition (Right) Low ATP condition. Counts in top right corner are number of enzymes.

2.2.4 ANALYZING ATP DEPENDENCE OF NSP13 UNWINDING AND TRANSLOCATION KINETICS

We varied [ATP] to analyze its effect on nsp13 unwinding and translocation. Because of the variation in dwell time with DNA sequence, we analyze each DNA position individually to avoid averaging over sequence-dependent effects. We also consider the fast enzyme events separately from slow enzyme events to avoid averaging over potentially different enzyme pathways.

For fast unwinding enzyme events, we compare the median duration of each current level at 150 μM ATP to those at 1 mM ATP by dividing $\tau_{1/2}$(150 μM) by $\tau_{1/2}$(1 mM) as shown in Figure 2.5. We find that the dwell time increases with decreased ATP and on average is 2-fold greater at 150 μM ATP. Because the dwell time changes along the DNA position in a sequence dependent manner, we analyzed the dwell time distribution per position across the high and low ATP conditions (2.6). The dwell time distributions shift towards the right with decreasing [ATP] across many DNA positions. A two- fold increase in dwell time is interpreted as reduction in half of maximal velocity of the enzyme at 150 μM ATP, which is close to the previously reported K_m value of ~100 μM ATP [53]. Thus, it is likely that the fast unwinding enzyme events are [ATP] dependent and that each nucleotide step observed is the result of nsp13 driven ATP catalysis. In all traces, the ion current magnitude did not change with ATP concentration.

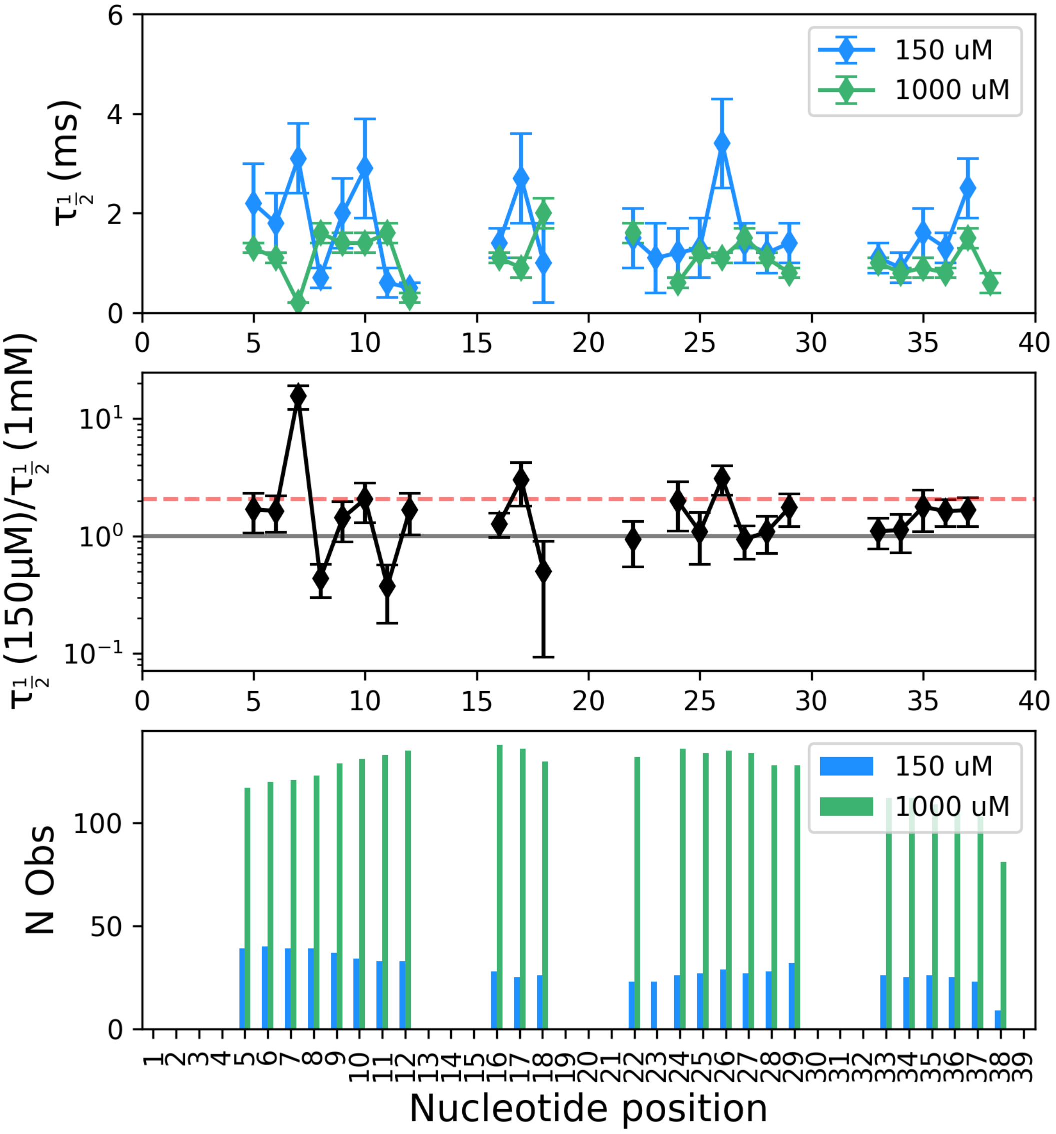

Figure 2.5: Analyzing ATP dependence of nsp13 fast unwinding events
(Top) Median duration of current levels for two different ATP conditions: 150 μM (blue) and 1000 μM (green). (Middle) The ratio of the median durations in (top) to remove sequence dependence that also influences the level durations. (Bottom) The number of times that a given level whose durations are shown in (top) was observed.

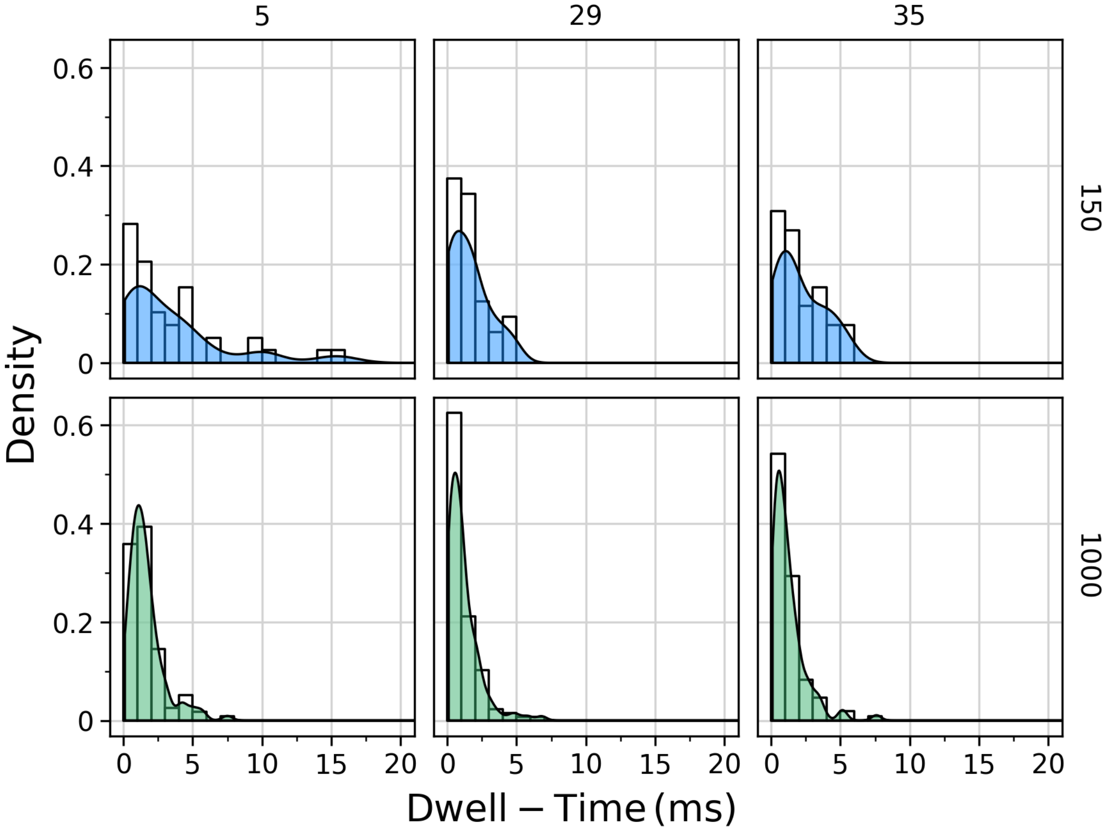

Figure 2.6: Analyzing ATP and sequence dependence of nsp13 fast unwinding events The distribution of dwell times for three positions (left to right: Position 5, 29, and 35) across two different ATP conditions (150 μM in blue and 1000 μM in green). The kernel density estimate is plotted on top in black.

Similarly, Figure 2.7 compares the ATP dependence of fast translocation enzyme events. The dwell time increases with decreased [ATP] and on average is 1.6-fold greater at 150 μM ATP.

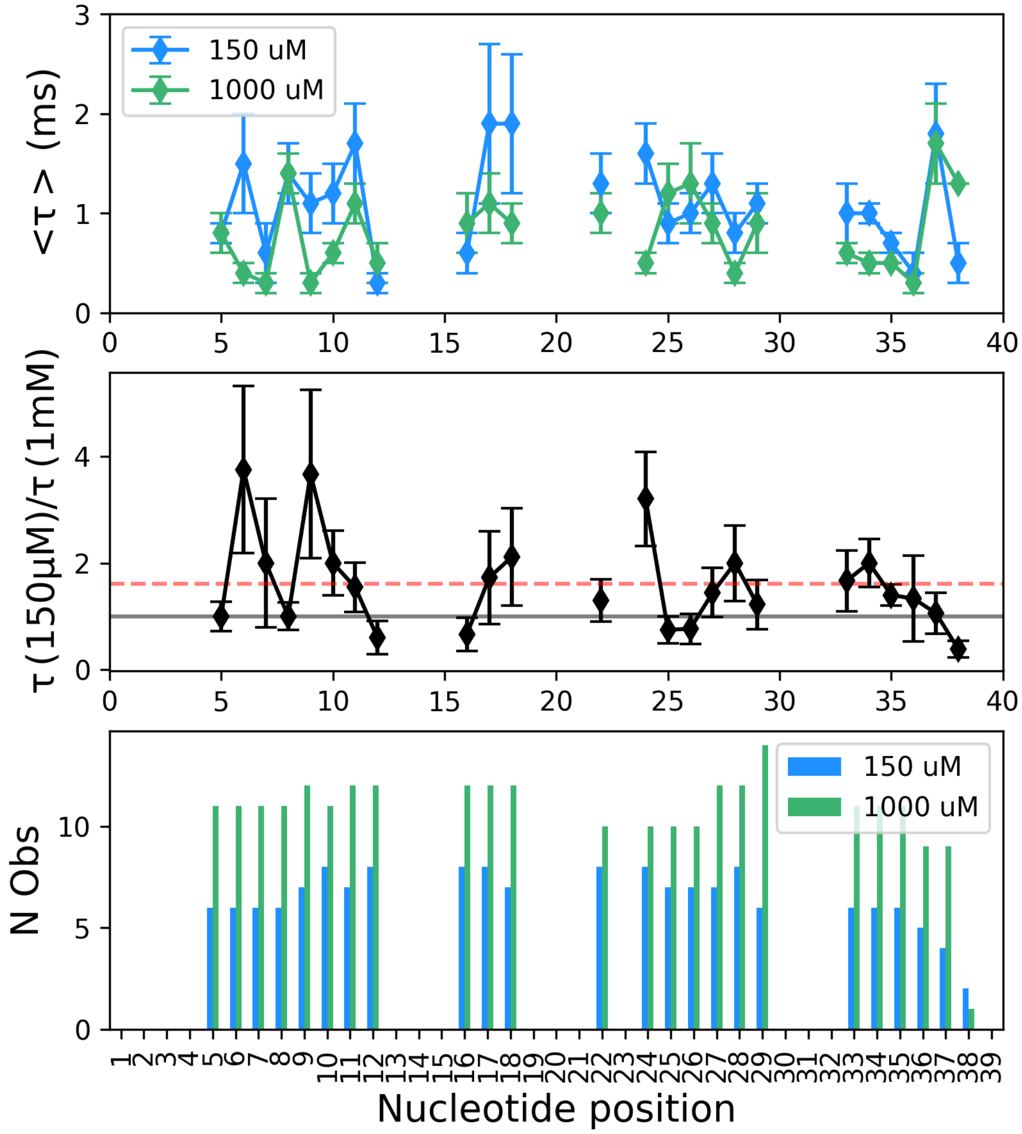

Figure 2.7: Analyzing ATP dependence of nsp13 fast ssDNA translocation events (Top) Median duration of current levels for two different ATP conditions: 150 μM (blue) and 1000 μM (green). (Middle) The ratio of the median durations in (top) to remove sequence dependence that also influences the level durations. (Bottom) The number of times that a given level whose durations are shown in (top) was observed.

An identical analysis of the slow unwinding enzyme events shows that the median duration of each level is independent of [ATP], as seen in Figure 2.8. Notably, the processivity of the helicase shows a sharp decline with decrease in ATP. Processivity is the probability that the helicase will step forward along the nucleic acid without falling off during the unwinding event. These results suggest that while the unwinding rate in slow events is not limited by ATP at 150 μM, the nsp13 off-rate is dependent on [ATP]. While we have qualitatively observed slow translocation enzyme events (not shown), we do not yet have enough statistics on it to report the [ATP] dependence.

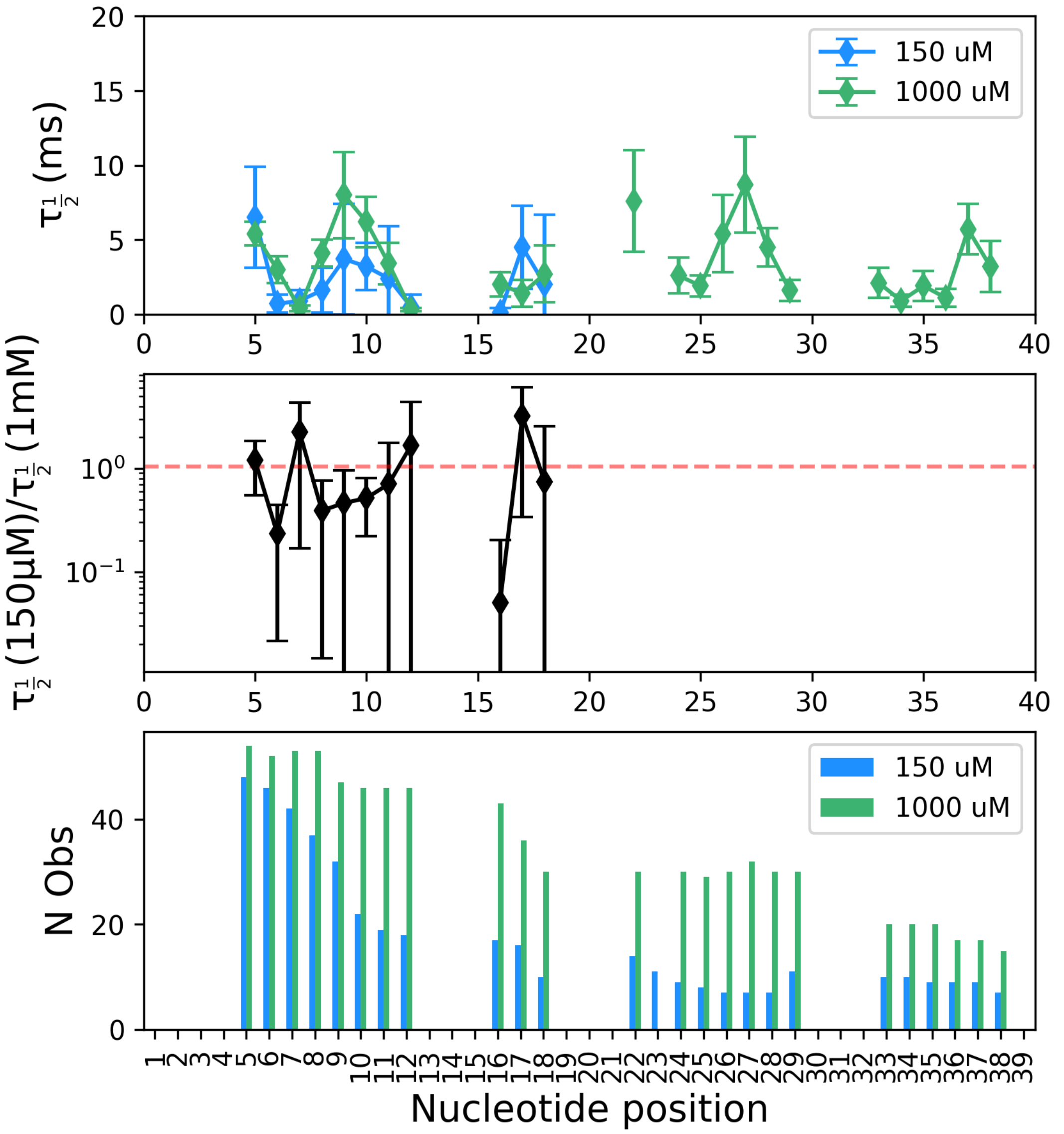

Figure 2.8: Analyzing ATP dependence of nsp13 slow unwinding events
(Top) Median duration of current levels for two different ATP conditions: 150 μM (blue) and 1000 μM (green). (Middle) The ratio of the median durations in (top) to remove sequence dependence that also influences the level durations. (Bottom) The number of times that a given level whose durations are shown in (top) was observed.

2.2.5 INVESTIGATING THE EFFECT OF NSP13 ATP-SITE MUTATION ON UNWINDING KINETICS

K288A is a knock out mutation of ATP binding site lysine that is inactive in bulk helicase activity assays (data not shown). Figure 2.9 shows a representative unwinding event and its underlying distribution from catalytically inactive K288A nsp13 compared with the fast and slow unwinding events from WT nsp13. Comparing the distributions of mean dwelltime of each nsp13 molecule of K288A to that of WT show that K288A mean dwell times are more uniformly distributed. The mean dwell time of K288A is 7-fold greater than that of WT's fast unwinding events and approximately the same as that of WT's slow unwinding events.

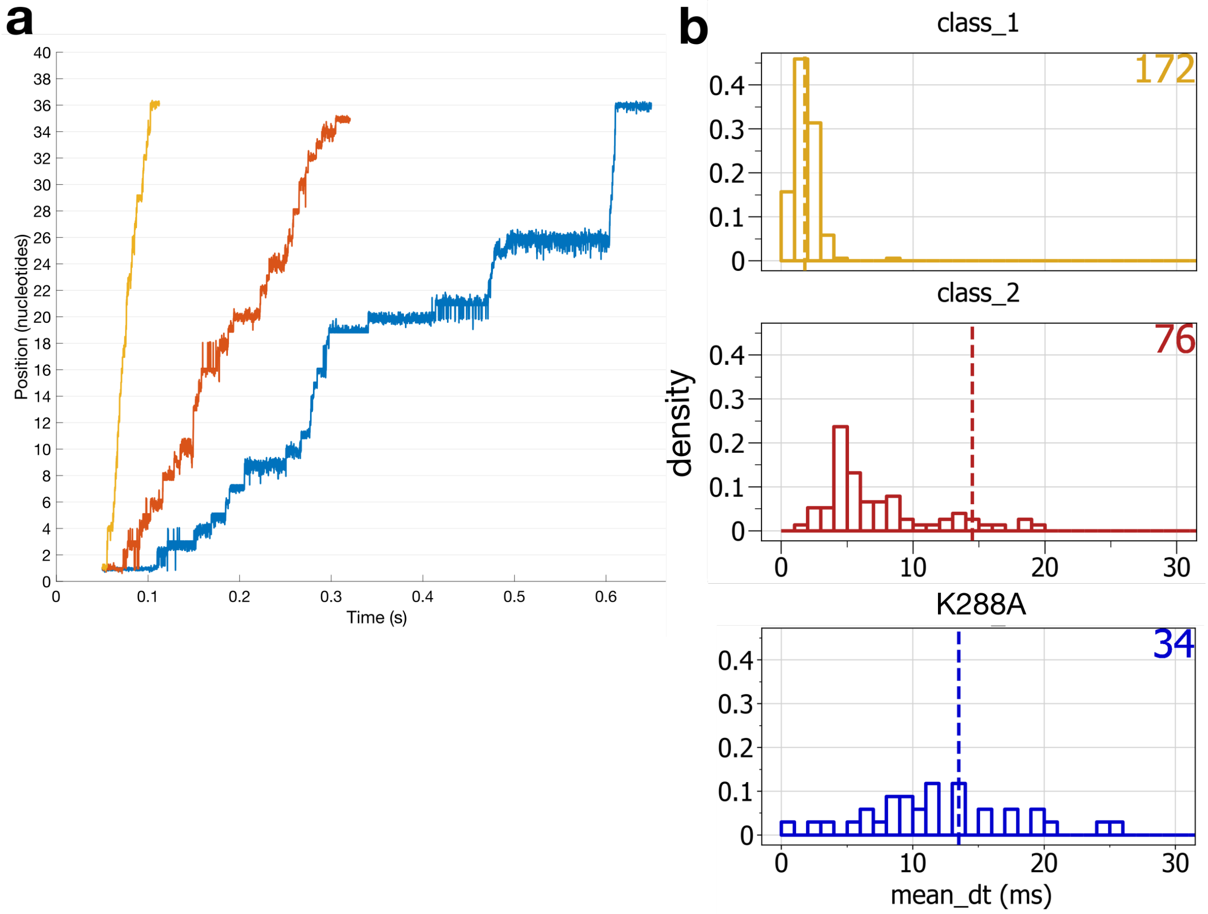

Figure 2.9: Unwinding events of active-site mutant K288A nsp13
(a) Comparision of position versus time trace for three different nsp13 unwinding events: fast class 1 event (gold), slow class 2 event (red), and K288A nsp13 unwinding event (blue). (b) Distribution of average dwell time per enzyme for each enzyme class from (a). Counts in top right corner are number of enzymes.

2.3 Discussion

Single nucleotide steps

In this work, we show that SARS-CoV-2 nsp13 unwinds duplex DNA in single-nucleotide steps. We also measure changes in nsp13 kinetic parameters such as duration of each step, backstep probability and processivity by varying experimental parameters such [ATP], and DNA form (single- vs double-stranded). Based on the Betterton-Jülicher model [54], viral helicases are predicted to travel in single nucleotide steps, each forward nucleotide step the result of a single ATP hydrolysis cycle. Bulk assays estimate that viral helicases such as NS3 and T7 gp4 travel in single nucleotide steps [55,56]. Moreover, our collaborators estimate that nsp13 unwinds RNA hairpin duplexes in single-nucleotide steps using a single-molecule optical tweezer measurements [53]. Our measurements are the first single-molecule single-step observations of nsp13 moving along a nucleic acid track, and the first direct confirmation that a viral helicase such as nsp13, belonging to the superfamily class of 1B helicases, unwinds duplex nucleic acids in single-nucleotide steps.

Nsp13 velocity

The mean nsp13 unwinding velocity we observed cluster at 150 $bp - s^{-1}$ and 600 $bp - s^{-1}$ under saturating ATP conditions and a constant force of ~ 35 pN in the direction of unzipping (+35 pN). Most of our experiments were dominated by one of the two populations of enzyme events, either fast or slow. Additionally, we are the first to measure nsp13 ssDNA translocation and find a translocation velocity of 1000 $b - s^{-1}$ under saturating ATP conditions and +35 pN force. (While we have qualitatively observed slower ssDNA translocation events on the order of few hundred base pairs per second, we have not yet measured it rigorously to report an exact value).

Catalysis and ATP-dependence

While the fast enzyme events show a 2-fold reduction in speed at 150 μM ATP, the slow enzyme events show no change. Mickolajczyk et al. (2020) estimate the K_m of ATP to be 100 μM from their optical tweezer measurements at 18 pN force. Thus, it is likely that our fast enyme measurements are rate-limited by [ATP] at ~150 μM and that the measured dwell times for individual nucleotide steps provide kinetic data of nsp13 ATP hydrolysis cycle. However, since we have only measured a two-fold change so far, futher ATP concentration titration in <100 μM range is needed to confirm the ATP K_m, and subsequently nsp13 k_{cat} at our applied forces.

Further, K288A nsp13, catalytically inactive in bulk assays, is driven by +35 pN force in the presence of saturating ATP at a rate of ~75 $bp - s^{-1}$; this is comparable to that of the slowest enzyme events of WT nsp13 and 8-fold slower than the fast ‘ATP-catalyzed’ enzyme events of WT nsp13. In the absence of ATP and +35 pN force, no unwinding events were observed with K288A nsp13. This suggests that either ATP is required for nsp13 to bind to nucleic acid or that chemical processes, other than ATP hydrolysis are rate-limiting. These data also offer one possible explanation for the occurence of slow unwinding events which may be the result of enzyme inactivation.

Using single molecule optical tweezer measurements, Mickolajczyk et al. (2020) report nsp13 unwinding velocity of upto 280 $bp - s^{-1}$ at 18 pN on a RNA hairpin and observe an exponentially rising velocity with applied force in the range of 4 - 20 pN. While this velocity was measured with different nucleic acid substrate and under different applied force (and geometry of force application), our fast enzyme results of 600 $bp - s^{-1}$ track with the much greater applied force of ~35 pN directly on the enzyme.

While *in vivo* velcity of nsp13 and helicases have not yet been measured, the *E. Coli* replication complex is estimated to travel at 1000 $bp - s^{-1}$ or 1 ms per base-pair. This suggests that *in vivo* viral helicases may work as fast or faster. Notably, SPRNT is the only available single molecule technique that can resolve single nucleotide steps (~ 6) that occur at durations of 1ms [49].

Nsp13 regulation

In addition to force and stability of nucleic acid substrate, the velocity of nsp13 is reported to be influenced by sequence, cooperativity, stability of nucleic acid substrate, presence of RdRp and cysteine residues [38,39,42,43,53]. More detailed investigation into the effect of applied force, substrate type (double vs single-stranded, RNA vs DNA) and [ATP] on the velocity and processivity of K288A nsp13 and WT nsp13 can reveal the exact chemical pathways underlying unwinding and translocation.

Other biologically relevant clues

Notably, our nsp13 events were only possible in the reverse orientation of the MspA nanopore. Only in this orientation, can the enzyme-DNA complex be in contact with the lipid bilayer. A possible explanation for this result could stem from the fact that the replication-transcription complex of SARS CoV (consisting of nsp13, RdRp, and other non structural proteins) are known to be membrane bound during genome replication [57].

Interestingly, we find that nsp13 DNA unwinding rate is sequence dependent. *In vitro* helicase unwinding have been shown to speed up in AT-rich regions and slow down in GC-rich regions due to the energetic differences between DNA base pairs [58]. Our findings also demonstrate sequence dependence of nsp13 translocating on ssDNA. Though the exact mechanism is unknown, helicases such as UvrD [59] and Hel308 [51] show sequence dependent kinetics even while translocating on ss DNA. Two ways this could occur is differences in interaction energies between helicase amino acid and nitrogenous base and/or steric effect of bases in nucleic acid binding site affecting the binding affinity of helicase walker domains to the nucleic acid.

Chapter 3

Conclusions

Nanopore technologies harness the atomic precision of transmembrane proteins to create a highly confined channel within which to study molecules at the angstrom-scale in real-time. In my research, I explored de novo design principles to procedurally generate diverse channel geometries with Rosetta and studied the ability of the designed proteins to form pores within membranes. Our ability to create customized pores is a step towards enhancing the accuracy of nanopore DNA sequencing and single-molecule identification.

While a precise nanopore allows greater discrimination between DNA bases, yet another protein is responsible for the controlled release of DNA into the pore. Nanopore DNA sequencing relies on enzymes that bind and walk on DNA to slow the rate of DNA translocation from the speed of ions to that of the enzymes themselves. This in turn gives access to the kinetics of the enzyme at the microsecond timescale. In my research, I also show our ability to monitor the SARS CoV-2 helicase walking in single-nucleotide steps. This provides a platform to study how viral inhibitors target the helicase and glean insights for their development.

Appendix A

Supplementary information for chapter 1

A.1 Single channel analysis of *de novo* TMBs

De novo TMB is inserted into planar bilayers by detergent dilution, an approach widely used to reconstitute bacterial porins into planar lipid bilayers. Refolded proteins in detergent (n-dodecyl β–D-maltoside (DDM) or n-Dodecylphosphocholine (DPC) or Tetraethyleneglycolmonooctylether (C8E4)) at above the critical micelle concentration (CMC) is diluted into the cis chamber of the bilayer apparatus.

Experimental setup

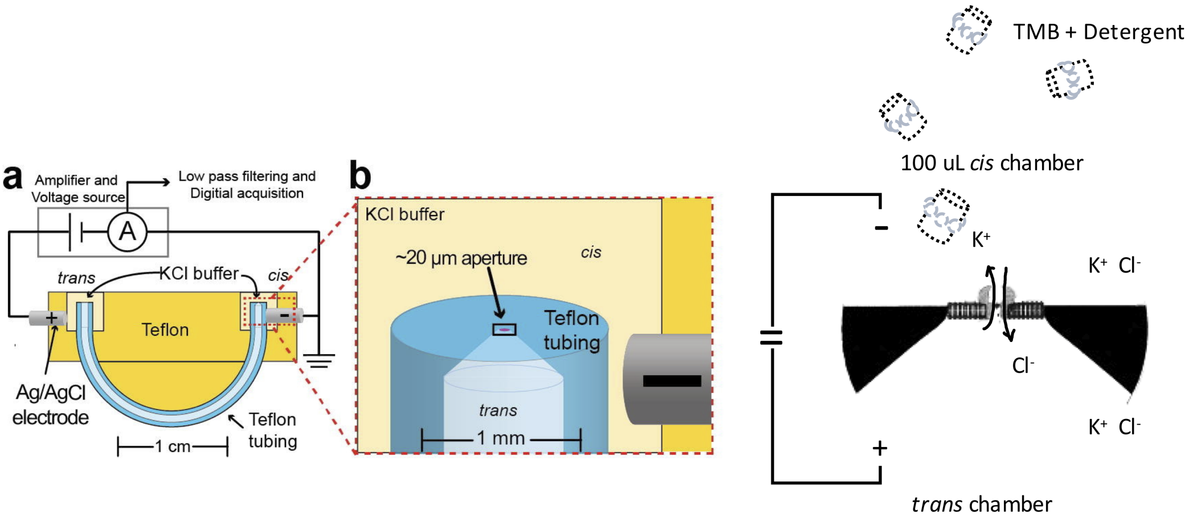

Figure A.1: Single channel experiment setup. Electrical circuit that connects two chambers separated by a planar bilayer (B) Zoomed aperture across which planar bilayer is formed (C) Direction of flow of ions through a pore. Adapted from [60].

- As shown in figure A.1, two chambers (cis- and trans+) are separated by an artificial lipid bilayer. Each chamber is filled with electrolyte solution consisting of 1M KCL, 10 mM Hepes pH 8.0. Lipid paint consisting of a mixture of DOPhPC lipid and hexadecane is painted across a 20 μm aperture. A planar bilayer is established by placing a bubble on the aperture and creating an air-water interface.
- A potential of + 100 mV is applied across the bilayer by two Ag/AgCl electrodes dipped in the cis and trans chamber. A positive current is one in which cations flow from the trans to the cis side. To ensure bilayers do not rupture during an experiment, the bilayer is verified to withstand > 500 mV and has a noise under 1 pA in the current measurement. A control experiment is recorded with no reagents.
- Proteins are then added into the cis chamber. The protein:detergent ratio varies with critical micelle concentration of the detergent and the activity of the protein. Typically, *de novo* TMBs have been diluted to 50 ug/mL in either 2 CMC or 1 CMC of detergent in TBS pH 7.5. They are further diluted 25 to 50 fold in the cis chamber. These ratios result in single channels on average.
- When a discrete jump is observed, the cis chamber is flushed to remove any further proteins from inserting. The potential offset was adjusted to zero such that at 0 mV applied voltage, the measured current is 0 pA.
- In order to capture the first jump in sometimes a series of conductance increases, the data was continuously recorded for this exploratory phase in research.
- Time to insertion varied greatly:
    - When time to insertion was very long (> 200 s): more proteins were added, new bilayers would be created frequently and/or the cis chamber would be reset and new proteins would be added.
    - When time to insertions was even longer (> 1000 s), new apertures would be tried.

Current recording

- Single channel conductance was measured in two ways: (a) currents recorded at a constant holding potential (b) currents at a scanning potential between - 100 mV to + 100 mV to derive the I-V curve.
- I-V curves were determined for the primary conducting states of single TMB channels by recording current values immediately after the transient capacitative currents arising from the voltage steps.
- Voltage induced gating experiments were recorded by switching the voltage on and off to > 250 mV. This yeilds a representive series of votlage-dependent opening and closing events of one channel.

- The current was amplified with an Axopatch 200B integrating patch clamp amplifier (Axon Instruments). Signals were filtered with a Bessel filter at 10 kHz and then acquired by a computer at 50 kHz (sampling at 20 μs) after digitization with a PCI-6251 DAQ (National Instruments).
- The recording conditions for all the experiments are the same unless specified: Symmetric cis and trans 1M KCl, 10 mM Hepes pH 8.0, room temperature, voltage = + 100 mV.

Data Analysis

Data from continuously recorded experiments needed to be simply visualized (a) to quickly develop an understanding of channel behavior (b) to record patterns or reproducibility across several independent experiments and experimental setups (c) to slice low-noise sections of experiment that correspond to single channels, multiple channels, a different holding potential and voltage ramps. Consequently, a custom visualization application for exploratory analysis was written in Dash, a Python framework for web applications.

- Raw data is first gaussian filtered at 250 Hz and loaded into the visualization application which runs on any browser. Every recorded data file is available in a data table and individual experiments can be viewed chronologically.
- Events corresponding to single channels, their associated voltage sweep, multiple channels, their associated voltage sweep are sliced and tagged. Given that there can be sub-conducting states, the highest conducting state for an isolated channel is considered the fully open state of the single channel.
- At this stage, the analysis of these curated events can be automated by setting current thresholds to detect ion-current state changes and finding the distribution in amplitude of currents.
- Raw data traces for single and multiple channels along with their all-points histogram at +100 mV are presented in the results. This is a first order assesment of channel reproducibility and gating behavior.
- Analysis of I-V curves and voltage induced closures become identifying features of types of channels.
- With more than one channel and more than one conducting state per channel, the number of possible ion-current states are discerned by applying signal convolution.

A.2 *De novo* design of 10-, 12-, 16-stranded repeat TMB

Single chain β-barrels of a tunable diameter are desirable for nanopore applications. The regular and repeating architecture of anti-parallel beta barrels is thought to have evolved from repeated duplication of a single ancestral β- hairpin [61]. Simply duplicating the sequence of an 8-stranded β-barrel, as was done in the case of the naturally occuring OmpX, may form a larger barrel as observed by its increased conductance [62].

TMB 2_17 and TMB 2_3, structurally validated 8-stranded *de novo* TMBs, were chosen as templates to build larger barrels. The repeat unit of the larger barrel can be a single or double β-hairpin from the smaller barrel. A β-hairpin is two adjacent antiparallel β-strands linked by a turn of a few residues and connected by a hydrogen bond ladder between them. An 8-stranded barrel has 4 unique β-hairpins, that is referred to as H1 H2 H3 H4 going from N-to-C termini of the protein, and 2 double hairpins H1H2 and H3H4. Figure A.2 a. shows the design models for larger 10- and 12- stranded barrels created by duplicating hairpins in 8-stranded barrels. While this design strategy preserves the backbone hydrogen bonding between the strands, the structural motifs and polar networks, formed by side-chain residues, are not necessarily in place. A subset of these barrels expressed in *E.coli* and were refolded in (DPC) detergent (methods as described in [34]). Figure A.2 b. shows that these larger transmembrane β-barrels purify as a single species in detergent. Further biochemical screens are necessary to understand the stability and structure of these designs; however, their intial success suggests that the repeat strategy may be a feasible way to obtain starting backbones to further optimize sequence upon.

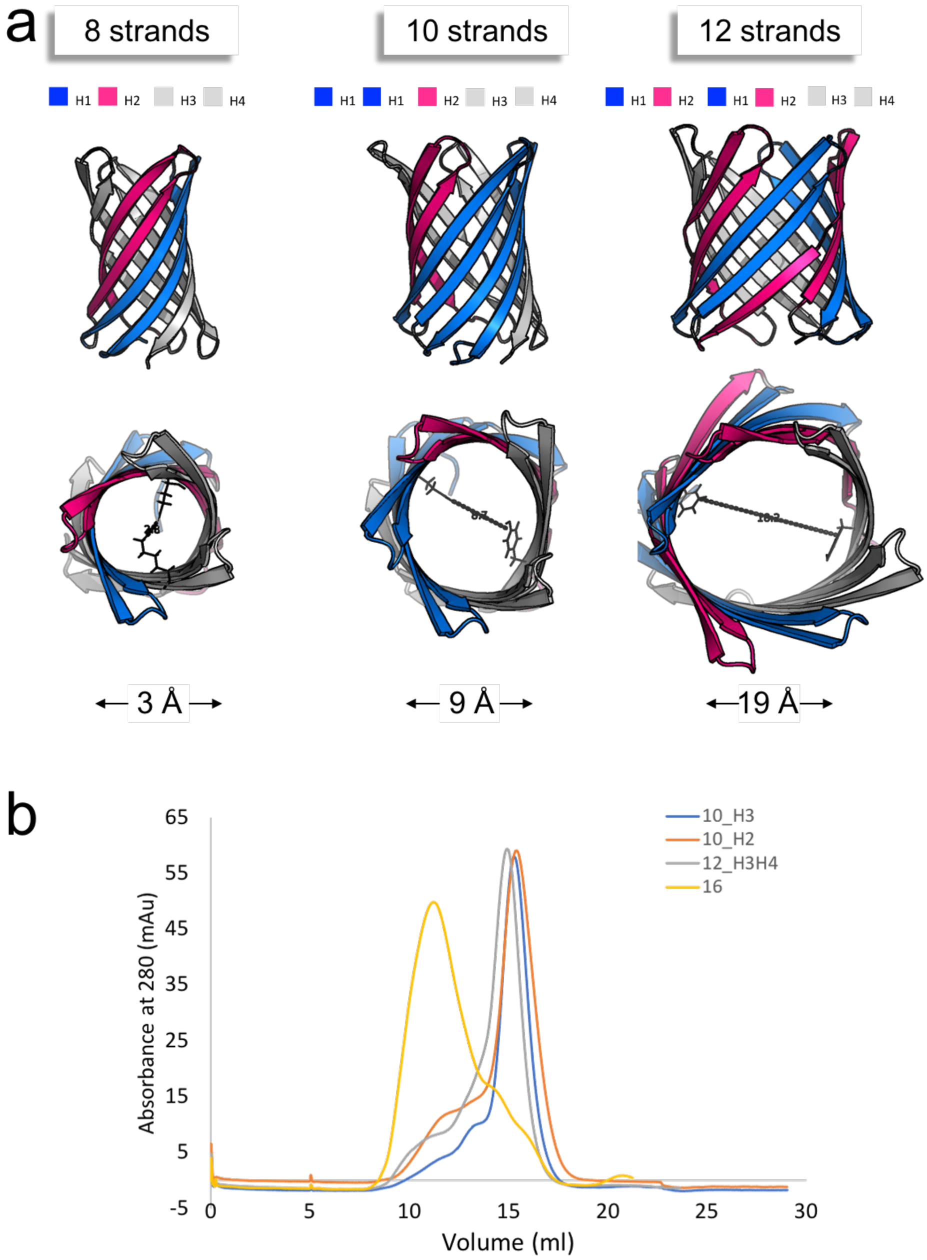

Figure A.2: Repeat β-hairpin design strategy for larger β-barrels
(a) Rosetta design models of 10- and 12-stranded β-barrel. In the 8-stranded β-barrel, each β-hairpin is represented as a square labelled H1, H2, H3, H4. Duplication of a single hairpin (H1, blue) results in 10-stranded barrel. The sequence of the hairpins is listed as H1 H2 H2 H3 H4. Duplication of 2 hairpins (H1, H2 blue, pink) results in 12-stranded barrel. Duplication of all hairpins results in 16-stranded barrel.
(b) Size-exclusion chromatography (SEC) of transmembrane β-barrels refolded in (DPC) detergent. Four β-barrels built from 8-stranded TMB 2_17 are shown. The hairpin sequence with duplication for each of the designs is as follows - 10_H3: N- H1 H2 H3 H3 H4 -C; 10_H2: N- H1 H2 H2 H3 H4 -C; 12_H3H4: N- H1 H2 H3 H4 H3 H4 -C; 16: Full duplication. While the 2 10-strand barrels appear to be folded in DPC detergent, the 16-strand barrel purifies in the void volume and thus, misfolded.

Appendix B

Supplementary Information for chapter 2

B.1 Materials and Methods

Proteins: The M2-NNN-MspA protein [63], a mutant form of WT MspA (PDB ID: 1UUN), in detergent was ordered from GenScript. SARS-CoV-2 nsp13 WT (PDB ID: 6ZSL) and K288A mutant was obtained from collaborators. Nsp13 expression and activity assay are described in [53].

DNA constructs: DNA sequences were designed to produce easily identifiable and high contrast electronic readouts. DNA sequence prediction for a 5' feeding DNA into a backwards orientation of MspA was approximated from previous base-to-current map as described in [64].

DNA sequences were synthesized at Stanford University Protein and Nucleic Acid Facility. The DNA sequences are listed Table B.2.

Single strand translocation experiments were perfomed with Sequence A and duplex unwinding experiments with Sequence A hybridized to complement A. The latter were annealed in a 1:1.2 molar ratio respectively by incubating at 95°C for 3 min followed by slow-cooling to 4°C.

Operating buffers: The nsp13 reaction buffer in the cis chamber was 20 mM KCL, 20 mM Hepes pH 7.5, 1 mM TCEP, 1mM ATP. The ATP concentration was adjusted in cis chamber for varying ATP conditions. The other side of the membrane, trans chamber, consisted of 500mM KCl, 20mM HEPES, pH 7.5. The reaction buffer was informed by bulk assays in literature showing tapering activity of nsp13 beyond 50 mM KCL [REF].

The trans buffer was maintained at higher salt to maximize ion-current signal.

Nanopore experiments: For single-molecule experiments, a single MspA nanopore was established in a bilayer as described in [2,60,65]. Briefly, a lipid bilayer is formed with a mixture of 1,2-di-O-phytanyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids, SKU: 999984C) and n-hexadecane across a 20 μm diameter aperture that separates two 65 μL chambers containing our operating buffers. A potential of 180 mV is applied using an Axopatch 200B or Axopatch 1B amplifiers (Axon Instruments) and a National Instruments PCI-6251 DAQ.

Nsp13 was preloaded onto DNA construct in the absence of ATP by incubating at RT for 30 min, at a final concentration of 0.4 uM DNA, 4uM nsp13, 124 mM KCL, 20 mM Hepes pH 7.5, 5 mM MgCl2, 1 mM TCEP.

Translocation and unwinding experiments were initiated by adding nsp13:DNA complexes to the cis chamber at a final concentration of 100 nM and 10 nM respectively.

All experiments presented in chapter 2 were done at RT (23 ± 1 °C).

Data acquisition and analysis: Data were acquired with a sampling rate of 50 kHz and filtered at 10 kHz. Enzyme events were detected using a thresholding algorithm described in [2]. Good events were selected manually and consisted of enough ion-current levels or context to accurately map to sequence position. Levels within events are found by a level-finding alogrithm and the durations of these levels are calculated to yield dwell-time per DNA position. The level-finding algorithm is described in more detail in Appendix B.3.

B.2 DNA Sequences

Sequence 5'-3'
Worked:
>Sequence A
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGTCCAAGTCCTGGACCTTT
TGCCGAGCCGAGCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
>Complement A
AGGTTCAGGACCTGGAAAACGGCTCGGCTCGG
>Sequence B
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGTCCAAGTCCTGGACCTTT
TTTTTTTTTTTTTTTTTTTTTTTTTTT
>Complement B

AGGTTCAGGACCTGG
>Sequence C
TAACCGGAGCATCTTACTTTTTTTTTTTTTTTTTTTTZ

Failed:
>Sequence C extended
TAACCGGAGCATCTTACGAGCAATGAGGCTAGAGGCTTAACCGGAGCATCTT
ACTTTTTTTTTTTTTTTTTTTTZ
>template_90mer_5pfeed
PTTTTTTTTTTTTTTTAGTGATAGTGCXGTCGTTGGTGTACAAGTAATTGTC
AACACTCCCATTTGTATGTTCGGCTAACTTGTACCGATACA
>complement_90mer_5pfeed
TGTATCGGTACAAGTTAGCCGAACATACAAATGGGAGTGTTGACAATTACTT
TTTTTTTTTTTTTTTTTZ
>hairpin_5pfeed
PTTTTTTTTTTTTTTTAGTGATAGTGCXGTCGTTGGTGTACAAGTAATTGTC
AACACTCCCATTTGTATGTTCGGCTAACTTGTACCGATACATTTTTGTATCG
GTACAAGTTAGCCGAACATACAAATGGGAGTGTTGACAATTACTTGTATTTT
TTTTTTTTTTTTZ
>template_140mer_3pfeed
TTTTTTTTTTTTTTTTTTTCCCTTTCTTTCCTGATACACTTTTCTTCTTTTTT
TTAGACATTTCCTTTTCCTGCGGAAGCTGTCCGGAAGCTTTACATAGTTCCTT
TCGTTGAGTAGCCTTAXCTGTCCGCCACATGTCAP
>complement_140mer_3pf
AAAAGGAAATGTCTAAAAAAAAGAAGAAAAGTGTATCAGGAAAGAAAGG

RNA:
>RNA_template_120mer
GGUUUUUUUUUUAACUAGUCGUAGUGCAGUCGCUACAUUACAGUGUAUUGUUU
AGCGGACGUUCACUAUCGCAUUCUUUAAUACUUGGUAUCGUCGGUUGAGUCGC
GAGUCGCCACCGUC
>RNA_complement
AGTGAACGTCCGCTAAACAATACTTTTTTTTTTTTTTTTTTTTTTTTZ

Table: DNA sequences used in chapter 2.
List of DNA template and complement strands used in this study. Sequence A and complement A was the DNA used in the results of chapter 2. Legend: X = abasic, Z=cholestrol, P = phosphate.

B.3 Extracting dwell-time per DNA position from raw ion current

Nsp13 walking along a nucleic acid strand is observed as a series of discrete ion current steps. Raw ion current is converted to DNA position and corresponding dwell-time as follows:

- Currents in MspA pore are determined by four nucleotides at a given time. Each four nucleotide combination was measured and a ‘quadromer’ ion-current to sequence map was created as described in [64]. For a given sequence, a prediction is made using the quadromer map.
- By comparing measured ion current levels to predicted ion current levels, a consensus of median ion-current per position is created for observed levels. A set of high quality events containing all expected levels are used to create this reference consensus. The comparision also allows to identify current levels separated by indistinguishable current differences. These degenerate current regions are shown in Figure B.1.
- Each current data point is then aligned to reference consensus similar to alignment algorithm described in [64,66]. Briefly, the Needleman-Wunch alignment algorithm is modified to allow for back-steps, and skips in a series of steps caused by the stochastic motion of enzymes. All alignments are manually checked and filtered for misaligned events.
- Given the sampling frequency is 10 kHz, the shortest measurable step corresponds to a single data point and a time interval of 100 μs. The duration per position is calculated as 100 μs * number of data points aligned to a position.

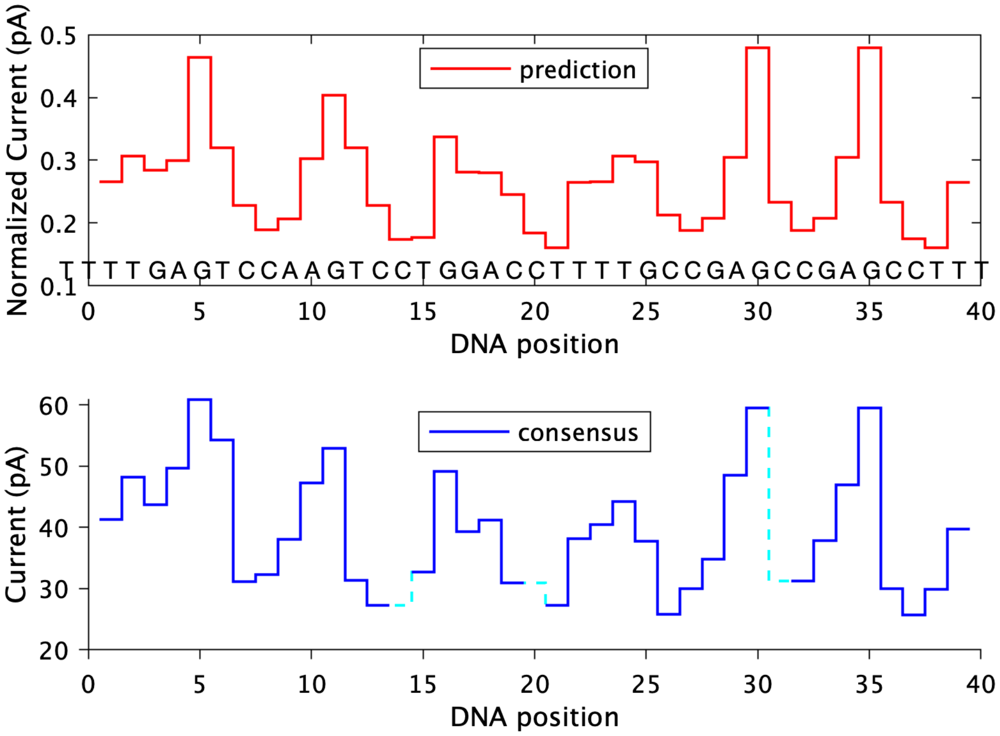

Figure B.1: Prediction versus consensus for a 5' - 3' read with single nucleotide steps (Top) Predicted ion current level patterns for sequence used in chapter 2. (Bottom) Consensus ion current level from 100+ measurements of nsp13 unwinding or translocating and across several ATP conditions. Degenerate ion current levels are shown in light blue dashed lines.

B.4 Supplementary figures for chapter 2

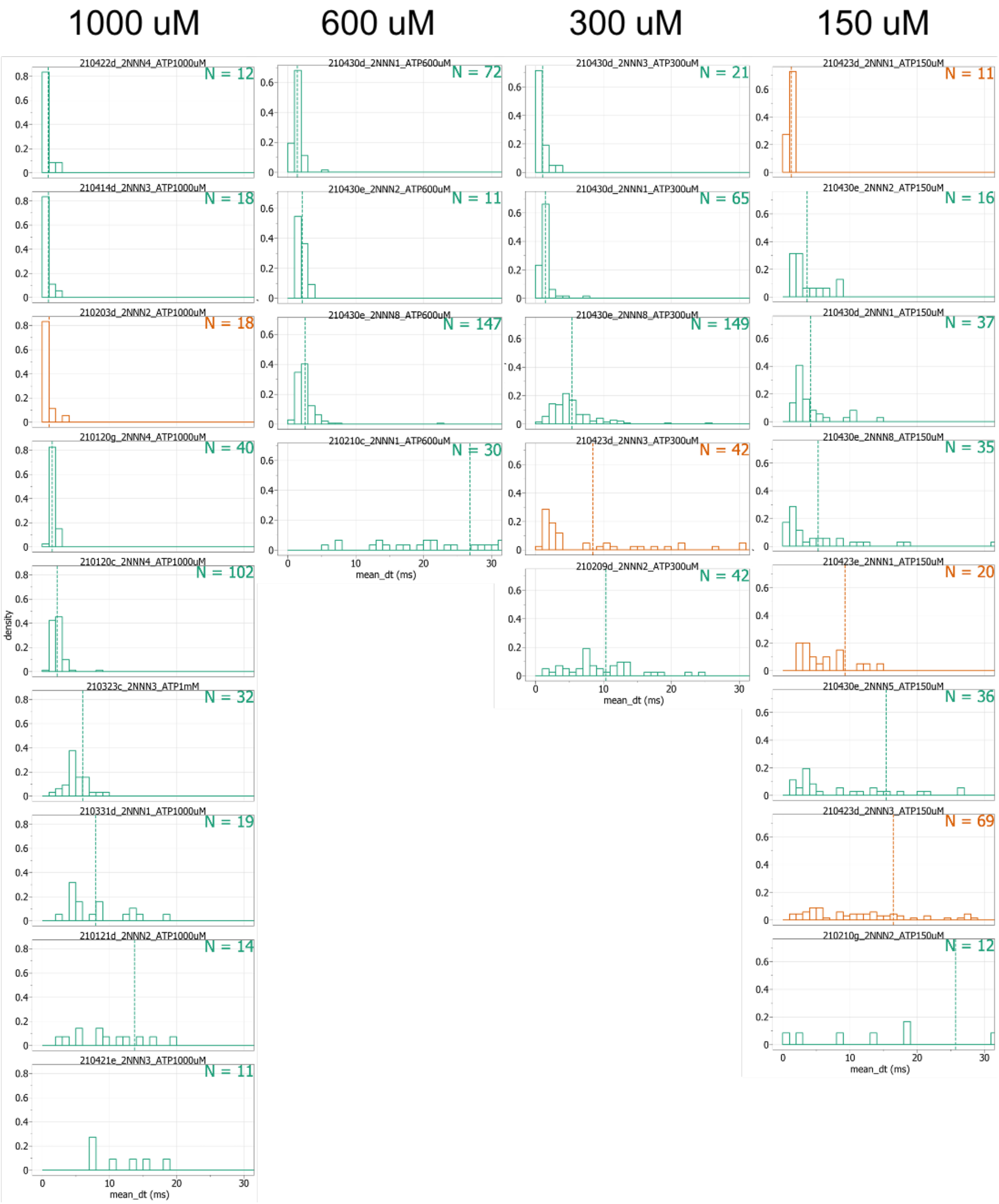

Figure B.2: Distribution of average dwell time per enzyme for varying [ATP]. From left to right: 1000 uM, 600 uM, 300 uM and 150 uM. Each row/panel is an independent nanopore experiment. Dashed line is through the mean for the distribution.

Bibliography

[1] J.J. Kasianowicz, E. Brandin, D. Branton, D.W. Deamer, Characterization of individual polynucleotide molecules using a membrane channel, Proceedings of the National Academy of Sciences. 93 (1996) 13770–13773.

[2] E.A. Manrao, I.M. Derrington, A.H. Laszlo, K.W. Langford, M.K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, J.H. Gundlach, Reading dna at single-nucleotide resolution with a mutant mspa nanopore and phi29 dna polymerase, Nature Biotechnology. 30 (2012) 349.

[3] A.H. Laszlo, I.M. Derrington, B.C. Ross, H. Brinkerhoff, A. Adey, I.C. Nova, J.M. Craig, K.W. Langford, J.M. Samson, R. Daza, others, Decoding long nanopore sequencing reads of natural dna, Nature Biotechnology. 32 (2014) 829.

[4] M.T. Noakes, H. Brinkerhoff, A.H. Laszlo, I.M. Derrington, K.W. Langford, J.W. Mount, J.L. Bowman, K.S. Baker, K.M. Doering, B.I. Tickman, others, Increasing the accuracy of nanopore dna sequencing using a time-varying cross membrane voltage, Nature Biotechnology. 37 (2019) 651.

[5] G. Huang, K. Willems, M. Soskine, C. Wloka, G. Maglia, Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with frac nanopores, Nature Communications. 8 (2017) 935.

[6] M. Pavlenok, I.M. Derrington, J.H. Gundlach, M. Niederweis, MspA nanopores from subunit dimers, PLoS One. 7 (2012) e38726.

[7] G. Huang, A. Voet, G. Maglia, FraC nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution, Nature Communications. 10 (2019) 835.

[8] M. Ayub, D. Stoddart, H. Bayley, Nucleobase recognition by truncated α-hemolysin pores, ACS Nano. 9 (2015) 7895–7903.

[9] G.J. Rocklin, T.M. Chidyausiku, I. Goreshnik, A. Ford, S. Houliston, A. Lemak, L. Carter, R. Ravichandran, V.K. Mulligan, A. Chevalier, others, Global analysis of protein

folding using massively parallel design, synthesis, and testing, Science. 357 (2017) 168–175.

[10] D. Baker, What has de novo protein design taught us about protein folding and biophysics?, Protein Science. 28 (2019) 678–683.

[11] K.J. Grayson, J.R. Anderson, Designed for life: Biocompatible de novo designed proteins and components, Journal of the Royal Society Interface. 15 (2018) 20180472.

[12] K.R. Mahendran, A. Niitsu, L. Kong, A.R. Thomson, R.B. Sessions, D.N. Woolfson, H. Bayley, A monodisperse transmembrane α-helical peptide barrel, Nature Chemistry. 9 (2017) 411.

[13] N.H. Joh, T. Wang, M.P. Bhate, R. Acharya, Y. Wu, M. Grabe, M. Hong, G. Grigoryan, W.F. DeGrado, De novo design of a transmembrane zn2+-transporting four-helix bundle, Science. 346 (2014) 1520–1524.

[14] K.R. Vinothkumar, R. Henderson, Structures of membrane proteins, Quarterly Reviews of Biophysics. 43 (2010) 65–158.

[15] M. Faller, M. Niederweis, G.E. Schulz, The structure of a mycobacterial outer-membrane channel, Science. 303 (2004) 1189–1192.

[16] A. Arora, D. Rinehart, G. Szabo, L.K. Tamm, Refolded outer membrane protein a of escherichia coliForms ion channels with two conductance states in planar lipid bilayers, Journal of Biological Chemistry. 275 (2000) 1594–1600.

[17] M. Chen, Q.-H. Li, H. Bayley, Orientation of the monomeric porin ompg in planar lipid bilayers, ChemBioChem. 9 (2008) 3029–3036.

[18] P. Goyal, P.V. Krasteva, N. Van Gerven, F. Gubellini, I. Van den Broeck, A. Troupiotis-Tsaïlaki, W. Jonckheere, G. Péhau-Arnaudet, J.S. Pinkner, M.R. Chapman, others, Structural and mechanistic insights into the bacterial amyloid secretion channel csgg, Nature. 516 (2014) 250.

[19] L.-M. Winterstein, K. Kukovetz, O. Rauh, D.L. Turman, C. Braun, A. Moroni, I. Schroeder, G. Thiel, Reconstitution and functional characterization of ion channels from nanodiscs in lipid bilayers, The Journal of General Physiology. 150 (2018) 637–646.

[20] T. Gutsmann, T. Heimburg, U. Keyser, K.R. Mahendran, M. Winterhalter, Protein reconstitution into freestanding planar lipid membranes for electrophysiological characterization, Nature Protocols. 10 (2015) 188.

[21] C.L. Hagan, S. Kim, D. Kahne, Reconstitution of outer membrane protein assembly from purified components, Science. 328 (2010) 890–892.

[22] B. Schiffrin, D.J. Brockwell, S.E. Radford, Outer membrane protein folding from an energy landscape perspective, BMC Biology. 15 (2017) 123.

[23] X. Feng, P. Barth, A topological and conformational stability alphabet for multipass membrane proteins, Nature Chemical Biology. 12 (2016) 167.

[24] M. Eilers, S.C. Shekar, T. Shieh, S.O. Smith, P.J. Fleming, Internal packing of helical membrane proteins, Proceedings of the National Academy of Sciences. 97 (2000) 5796–5801.

[25] P. Curnow, Designing minimalist membrane proteins, Biochemical Society Transactions. 47 (2019) 1233–1245.

[26] M. Mravic, J.L. Thomaston, M. Tucker, P.E. Solomon, L. Liu, W.F. DeGrado, Packing of apolar side chains enables accurate design of highly stable membrane proteins, Science. 363 (2019) 1418–1423.

[27] T. Haltia, E. Freire, Forces and factors that contribute to the structural stability of membrane proteins, Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes. 1241 (1995) 295–322.

[28] A. Henning-Knechtel, J. Knechtel, M. Magzoub, DNA-assisted oligomerization of pore-forming toxin monomers into precisely-controlled protein channels, Nucleic Acids Research. 45 (2017) 12057–12068.

[29] H. Naveed, J. Liang, Weakly stable regions and protein-protein interactions in beta-barrel membrane proteins, Current Pharmaceutical Design. 20 (2014) 1268–1273.

[30] P.-S. Huang, G. Oberdorfer, C. Xu, X.Y. Pei, B.L. Nannenga, J.M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles, Science. 346 (2014) 481–485.

[31] S.E. Boyken, Z. Chen, B. Groves, R.A. Langan, G. Oberdorfer, A. Ford, J.M. Gilmore, C. Xu, F. DiMaio, J.H. Pereira, others, De novo design of protein homo-oligomers with modular hydrogen-bond network–mediated specificity, Science. 352 (2016) 680–687.

[32] J.A. Fallas, G. Ueda, W. Sheffler, V. Nguyen, D.E. McNamara, B. Sankaran, J.H. Pereira, F. Parmeggiani, T. Brunette, D. Cascio, others, Computational design of self-assembling cyclic protein homo-oligomers, Nature Chemistry. 9 (2017) 353.

[33] P. Lu, D. Min, F. DiMaio, K.Y. Wei, M.D. Vahey, S.E. Boyken, Z. Chen, J.A. Fallas, G. Ueda, W. Sheffler, others, Accurate computational design of multipass transmembrane proteins, Science. 359 (2018) 1042–1046.

[34] A.A. Vorobieva, P. White, B. Liang, J.E. Horne, A.K. Bera, C.M. Chow, S. Gerben,

S. Marx, A. Kang, A.Q. Stiving, S.R. Harvey, D.C. Marx, G.N. Khan, K.G. Fleming, V.H. Wysocki, D.J. Brockwell, L.K. Tamm, S.E. Radford, D. Baker, De novo design of transmembrane β barrels, Science. 371 (2021). http://dx.doi.org/10.1126/science.abc8182.

[35] M.M. Mohammad, K.R. Howard, L. Movileanu, Redesign of a plugged β-Barrel membrane protein*, The Journal of Biological Chemistry. 286 (2011) 8000–8013. https://www.sciencedirect.com/science/article/pii/S0021925820539458.

[36] L.K. Tamm, F. Abildgaard, A. Arora, H. Blad, J.H. Bushweller, Structure, dynamics and function of the outer membrane protein a (ompa) and influenza hemagglutinin fusion domain in detergent micelles by solution nmr, FEBS Letters. 555 (2003) 139–143.

[37] S. Fang, B. Chen, F.P.L. Tay, B.S. Ng, D.X. Liu, An arginine-to-proline mutation in a domain with undefined functions within the helicase protein (nsp13) is lethal to the coronavirus infectious bronchitis virus in cultured cells, Virology. 358 (2007) 136–147. http://dx.doi.org/10.1016/j.virol.2006.08.020.

[38] A. Seybert, C.C. Posthuma, L.C. van Dinten, E.J. Snijder, A.E. Gorbalenya, J. Ziebuhr, A complex zinc finger controls the enzymatic activities of nidovirus helicases, Journal of Virology. 79 (2005) 696–704. http://dx.doi.org/10.1128/JVI.79.2.696-704.2005.

[39] J. Chen, B. Malone, E. Llewellyn, M. Grasso, P.M.M. Shelton, P.D.B. Olinares, K. Maruthi, E.T. Eng, H. Vatandaslar, B. Chait, T. Kapoor, S.A. Darst, E.A. Campbell, Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex, bioRxiv : The Preprint Server for Biology. (2020). http://dx.doi.org/10.1101/2020.07.08.194084.

[40] L. Yan, Y. Zhang, J. Ge, L. Zheng, Y. Gao, T. Wang, Z. Jia, H. Wang, Y. Huang, M. Li, Q. Wang, Z. Rao, Z. Lou, Architecture of a SARS-CoV-2 mini replication and transcription complex, Nature Communications. 11 (2020) 5874. http://dx.doi.org/10.1038/s41467-020-19770-1.

[41] K.C. Lehmann, E.J. Snijder, C.C. Posthuma, A.E. Gorbalenya, What we know but do not understand about nidovirus helicases, Virus Research. 202 (2015) 12–32. http://dx.doi.org/10.1016/j.virusres.2014.12.001.

[42] A.O. Adedeji, B. Marchand, A.J.W. Te Velthuis, E.J. Snijder, S. Weiss, R.L. Eoff, K. Singh, S.G. Sarafianos, Mechanism of nucleic acid unwinding by SARS-CoV helicase, PloS One. 7 (2012) e36521. http://dx.doi.org/10.1371/journal.pone.0036521.

[43] N.-R. Lee, H.-M. Kwon, K. Park, S. Oh, Y.-J. Jeong, D.-E. Kim, Cooperative translo-

cation enhances the unwinding of duplex DNA by SARS coronavirus helicase nsP13, Nucleic Acids Research. 38 (2010) 7626–7636. http://dx.doi.org/10.1093/nar/gkq647.

[44] K.-J. Jang, S. Jeong, D.Y. Kang, N. Sp, Y.M. Yang, D.-E. Kim, A high ATP concentration enhances the cooperative translocation of the SARS coronavirus helicase nsP13 in the unwinding of duplex RNA, Scientific Reports. 10 (2020) 4481. http://dx.doi.org/10.1038/s41598-020-61432-1.

[45] Z. Jia, L. Yan, Z. Ren, L. Wu, J. Wang, J. Guo, L. Zheng, Z. Ming, L. Zhang, Z. Lou, Z. Rao, Delicate structural coordination of the severe acute respiratory syndrome coronavirus nsp13 upon ATP hydrolysis, Nucleic Acids Research. 47 (2019) 6538–6550. http://dx.doi.org/10.1093/nar/gkz409.

[46] Y.-C. Jiang, H. Feng, Y.-C. Lin, X.-R. Guo, New strategies against drug resistance to herpes simplex virus, International Journal of Oral Science. 8 (2016) 1–6. http://dx.doi.org/10.1038/ijos.2016.3.

[47] S.S. Velankar, P. Soultanas, M.S. Dillingham, H.S. Subramanya, D.B. Wigley, Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism, Cell. 97 (1999) 75–84. http://dx.doi.org/10.1016/s0092-8674(00)80716-3.

[48] R.A. Copeland, D.L. Pompliano, T.D. Meek, Drug–target residence time and its implications for lead optimization, Nature Reviews. Drug Discovery. 5 (2006) 730–739. https://www.nature.com/articles/nrd2082.

[49] A.H. Laszlo, I.M. Derrington, J.H. Gundlach, MspA nanopore as a single-molecule tool: From sequencing to SPRNT, Methods. 105 (2016) 75–89. http://dx.doi.org/10.1016/j.ymeth.2016.03.026.

[50] J.M. Craig, A.H. Laszlo, H. Brinkerhoff, I.M. Derrington, M.T. Noakes, I.C. Nova, B.I. Tickman, K. Doering, N.F. de Leeuw, J.H. Gundlach, Revealing dynamics of helicase translocation on single-stranded DNA using high-resolution nanopore tweezers, Proceedings of the National Academy of Sciences of the United States of America. 114 (2017) 11932–11937. http://dx.doi.org/10.1073/pnas.1711282114.

[51] J.M. Craig, A.H. Laszlo, I.C. Nova, H. Brinkerhoff, M.T. Noakes, K.S. Baker, J.L. Bowman, H.R. Higinbotham, J.W. Mount, J.H. Gundlach, Determining the effects of DNA sequence on hel308 helicase translocation along single-stranded DNA using nanopore tweezers, Nucleic Acids Research. 47 (2019) 2506–2513. http://dx.doi.org/10.1093/nar/gkz004.

[52] J.M. Craig, A.H. Laszlo, I.C. Nova, J.H. Gundlach, Modelling single-molecule kinetics of helicase translocation using high-resolution nanopore tweezers (SPRNT), Essays in

Biochemistry. 65 (2021) 109–127. http://dx.doi.org/10.1042/EBC20200027.

[53] K.J. Mickolajczyk, P.M.M. Shelton, M. Grasso, X. Cao, S.E. Warrington, A. Aher, S. Liu, T.M. Kapoor, Force-dependent stimulation of RNA unwinding by SARS-CoV-2 nsp13 helicase, (2020) 2020.07.31.231274. https://www.biorxiv.org/content/10.1101/2020.07.31.231274v1.

[54] A. Garai, D. Chowdhury, M.D. Betterton, Two-state model for helicase translocation and unwinding of nucleic acids, Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics. 77 (2008) 061910. http://dx.doi.org/10.1103/PhysRevE.77.061910.

[55] W. Cheng, S.G. Arunajadai, J.R. Moffitt, I. Tinoco, others, Single–base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase, (2011). https://science.sciencemag.org/content/333/6050/1746.abstract?casa_token=NSj8XJkeafAAAAAA:w_ZISDNr77LF6VQKk4ut1BgFtNYY3EdTP8W_co9KuIQGE7lDAznzcjPCZWcRNFdaIx8fzZOMYiC15w.

[56] S. Syed, M. Pandey, S.S. Patel, T. Ha, Single-molecule fluorescence reveals the unwinding stepping mechanism of replicative helicase, Cell Reports. 6 (2014) 1037–1045. http://dx.doi.org/10.1016/j.celrep.2014.02.022.

[57] A.A. Gulyaeva, A.E. Gorbalenya, A nidovirus perspective on SARS-CoV-2, Biochemical and Biophysical Research Communications. 538 (2021) 24–34. http://dx.doi.org/10.1016/j.bbrc.2020.11.015.

[58] R. Galletto, M.J. Jezewska, W. Bujalowski, Unzipping mechanism of the double-stranded DNA unwinding by a hexameric helicase: The effect of the 3' arm and the stability of the dsDNA on the unwinding activity of the escherichia coli DnaB helicase, Journal of Molecular Biology. 343 (2004) 101–114. http://dx.doi.org/10.1016/j.jmb.2004.07.056.

[59] E.J. Tomko, T.M. Lohman, Modulation of escherichia coli UvrD Single-Stranded DNA translocation by DNA base composition, Biophysical Journal. 113 (2017) 1405–1415. http://dx.doi.org/10.1016/j.bpj.2017.08.023.

[60] A.H. Laszlo, I.M. Derrrington, J.H. Gundlach, Subangstrom measurements of enzyme function using a biological nanopore, sprnt, in: Methods in Enzymology, Elsevier, 2017: pp. 387–414.

[61] M. Remmert, A. Biegert, D. Linke, A.N. Lupas, J. Söding, Evolution of outer membrane β-barrels from an ancestral β hairpin, Molecular Biology and Evolution. 27 (2010) 1348–1358.

[62] T. Arnold, M. Poynor, S. Nussberger, A.N. Lupas, D. Linke, Gene duplication of the

eight-stranded β-barrel ompx produces a functional pore: A scenario for the evolution of transmembrane β-barrels, Journal of Molecular Biology. 366 (2007) 1174–1184.

[63] T.Z. Butler, M. Pavlenok, I.M. Derrington, M. Niederweis, J.H. Gundlach, Single-molecule DNA detection with an engineered MspA protein nanopore, Proceedings of the National Academy of Sciences of the United States of America. 105 (2008) 20647–20652. http://dx.doi.org/10.1073/pnas.0807514106.

[64] A.H. Laszlo, I.M. Derrington, B.C. Ross, H. Brinkerhoff, A. Adey, I.C. Nova, J.M. Craig, K.W. Langford, J.M. Samson, R. Daza, K. Doering, J. Shendure, J.H. Gundlach, Decoding long nanopore sequencing reads of natural DNA, Nature Biotechnology. 32 (2014) 829–833. http://dx.doi.org/10.1038/nbt.2950.

[65] I.M. Derrington, T.Z. Butler, M.D. Collins, E. Manrao, M. Pavlenok, M. Niederweis, J.H. Gundlach, Nanopore DNA sequencing with MspA, Proceedings of the National Academy of Sciences of the United States of America. 107 (2010) 16060–16065. http://dx.doi.org/10.1073/pnas.1001831107.

[66] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, Journal of Molecular Biology. 48 (1970) 443–453. http://dx.doi.org/10.1016/0022-2836(70)90057-4.