Development of Single Molecule Electronic SNP Assays using Polymer Tagged Nucleotides and Nanopore Detection

Youngjin Cho

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Development of Single Molecule Electronic SNP Assays
using Polymer Tagged Nucleotides and Nanopore Detection

Youngjin Cho

As knowledge of the human genome has accelerated, various diseases and individuals' responses to drugs have been pinpointed to specific DNA variations in one's genome. Among many different types of variants, the most common and simplest is the single nucleotide polymorphism (SNP) in which a single base substitution occurs. Although there have been considerable improvements in technologies that can reveal a single base difference in a DNA strand, simple and affordable methods that have high detection sensitivity and require small sample volume are expected to facilitate widespread adoption of routine SNP analysis in clinical settings.

One such method that meets these requirements is to use nanopore as a single molecule detector, an emerging analytic system that detects changes in current related to molecules occupying a nanometer aperture. This dissertation thus chronicles our endeavors in developing a nanopore-based SNP assay using polymer tagged dideoxynucleotides (ddNTPs). The fundamental principles of this method rely on single base extension (SBE) of a primer by DNA polymerase using polymer tagged ddNTP analogs for allele discrimination and simple electronic readout of an alpha hemolysin (αHL) nanopore current for allele detection at the single molecule level. Using four uniquely tagged ddNTPs, a characteristic current level that is specific to each base is produced, thus identifying the SNP alleles present and the genotype at the site.

To demonstrate the feasibility of this approach, four polymer attached ddNTPs, each with a different tag that generates a characteristic current blockade level in the αHL nanopore, were

designed and synthesized. To search for a DNA polymerase that can accept these tagged ddNTP analogs as substrates, several candidate DNA polymerases were surveyed and their relative efficiencies for incorporation of the analogs were compared (Chapter 2).

To generate a steady and stable blockade event for accurate SNP analysis, two different means of positioning a tag molecule in the αHL nanopore after the SBE reaction have been explored: covalent conjugation of DNA primer to the pore and immobilization of biotinylated primer within the pore by streptavidin. To find a suitable position for primer attachment on the pore, three αHL mutants, each with a different single conjugation point, were constructed. Using these mutants, different DNA-pore conjugates were produced and purified via various chromatography systems (Chapter 3).

In the nanopore system, charged molecules such as DNA are electrophoretically driven through the pore under an applied voltage, thereby modulating the ionic current through the nanopore. This current reveals useful information about the structure and dynamic motion of the molecule at the single molecule level. Before performing SNP analysis, we first studied single molecule behaviors of oligonucleotides of different lengths and structures in the αHL pore and their ensuing current signatures in the system (Chapter 4).

Finally, harnessed with tools and insights from the nanopore single molecule studies, actual SNP assays were performed in our nanopore system using the polymer tagged ddNTPs and SBE. Chapter 5 discusses the integrated approach where SBE is achieved on a primer-conjugated αHL nanopore and Chapter 6 presents the results using a biotin-streptavidin complex for immobilization of tag molecules in the pore. Overall, this thesis validates adaptation of the nanopore detection system for SNP analysis using the polymer tagged ddNTPs.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLE

LIST OF SYMBOLS

ddNTPs – dideoxynucleotides

GWAS – genome wide association studies

SNPs – single nucleotide polymorphisms

SBE – single base extension

FP – fluorescence polarization

FRET – fluorescence resonance energy transfer

MALDI-TOF MS – matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

m/z – mass-to-charge ratio

PPi – pyrophosphate

KCl – potassium chloride

pA – pico-Ampere

Si_3N_4 – silicon nitride

αHL – alpha hemolysin

dsDNA – double-stranded DNA

CMOS – complementary metal-oxide semiconductor

dNTPs – deoxynucleotides

Mn^{2+} – manganese ion

Mg^{2+} – magnesium ion

NHS – N-hydroxysuccinimide ester

TEAC – triethylammonium acetate

TBTA – tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine

dSp – deoxyribose phosphate

Imp – inosine methyl phosphonate

Tmp – thymidine methyl phosphonate

sSMCC – sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate

AA – amino acid

WT – wide type

Cys – mutant containing a cysteine residue

8x-Histag – 8x histidine tag

SEC – size exclusion chromatography

DPhPC – 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine

βOG – octyl β-D-glucopyranoside

IMAC – immobilized metal affinity chromatography

IEC – ion exchange chromatography

ssDNA – single stranded DNA

MD – molecular dynamics

VKORC1 – vitamin K epoxide reductase complex subunit 1 gene

ACKNOWLEDGEMENTS

This is the culmination of my spiritual journey in the Ju Group of the Pharmacology Department at Columbia University. Although there is a single name on this thesis, this work would not have been achievable without the care and guidance that I have received from so many. First and foremost, I am deeply indebted to Dr. Jingyue Ju for his support and direction throughout the years. He took a chance on me, accepting me as a student in his group and introduced me into the magic world of organic synthesis and nanopore stochastic sensing. I am grateful for all the opportunities that I have been granted to try, fail and grow and also the safe environment that he provided for me to do so. His energy and unyielding passion for scientific research have been a great inspiration for me and it is my hope to embody commensurate enthusiasm in my future endeavors.

I have benefited significantly from discussions and assistance from all of the members of the Ju group. My special thanks go to Dr. James Russo, who provided constant care and interest in my project throughout the entire process. In particular, the emotional support and encouragement that he rendered along the way has been crucial, especially during the challenging times. Dr. Sergey Kalachikov taught me various experimental techniques and I was privileged to learn from a scientist of such high caliber. Dr. Shiv Kumar introduced me to and greatly helped me with synthesizing molecules which are critical components of my project. Dr. Carl Fuller provided great advice and suggestions and shared a wealth of his scientific knowledge. There is no doubt that the support and encouragement that I received from the rest of the Ju lab were the key elements that led me to finish this work. Without their kindness and willingness to assist, I would not be able to reach this far.

My thesis committee has been greatly instrumental in keeping me on track. Dr. Robert Kass has helped me keep my perspective on the big picture and move the project forward. Dr. Jonathan Javitch has been caring with his keen and to the point advice which often made me leave the room with a new angle on the problem. Thus, I am grateful for being challenged by them, and for their time and effort to provide me clear guidance. I would also like to thank Dr. Timothy Bestor and Dr. Qing Fan for agreeing to serve on my dissertation committee.

I am grateful to the Pharmacology department as well. Knowing that I have had the support of the department helped me endure difficult moments and it truly made a difference.

I must also thank Dr. Wonhee Jang for encouraging me to go study abroad and suggesting doctoral study in the first place. I am grateful to Dr. Keiji Hashimoto and Dr. Benedikt Westphalen for their constant care and advice about research in general.

There is no word to express the level of love and unwavering support my family has shown me, without which I would not be able to achieve this goal. Thus, I deeply thank my dad, mom, brother and sister-in-law for standing behind me always. I was fortunate to have a family here in New York. I am grateful to my aunt and cousins for including me in their family events and celebrating the successes along the way with me. I would also like to express love to my little sister, Jadoo Cho, whose charm and innocence always make me feel easy and relaxed.

Lastly, I would like to acknowledge my two heroes, my grandparents Youngsul Cho and Haesoon Kim. Whatever I am and whatever I have to contribute to this world is a product of these two people. By accomplishing this work, I hope to make them proud and wish to repay their unending love and belief in me.

For my grandfather, Youngsul Cho

who made me dare to dream big and stretch my potential

CHAPTER 1: Introduction

1.1 Implication of SNPs in precision medicine and pharmacogenomics

Over the past few decades, advances in molecular science and genomic technology have enabled us to achieve remarkable progress in the service of medicine such as the Human Genome Project, the International HapMap Project and genome wide association studies (GWAS).[1,2] On account of these endeavors, there has been an increased awareness of the scale and significance of numerous subtle genetic variations found in the human genome. Among these, the most common genetic markers are single nucleotide polymorphisms (SNPs) which account for remarkable genetic diversity and are a vital resource for studying complex genetic traits.[3,4]

By definition, SNPs are single-nucleotide substitutions of one base for another that are stable enough to occur in more than one percent of the general population.[3,9,16] In general, SNPs occur throughout the genome about once every 100 to 300 nucleotide base pairs and this can be interpreted as roughly 10 million SNPs within the 3 billion nucleotide human genome.[9,16] The majority of SNPs do not occur within genes and usually have no influence on protein function.[3,6,16] Nevertheless, many do correlate with a particular drug response or disease susceptibility in an individual and thus they are still considered as important markers for comparative or evolutionary genomic studies.[6,7] On the other hand, SNPs within the coding region or the regulatory sequences of a gene directly affect protein structure or function by changing the amino acid sequence of the product or level of gene expression.[8,16] In this case, SNPs can elucidate direct associations between sequence difference and phenotypic variation.

Recently, a new approach to disease treatment and prevention called precision medicine has emerged and is leading toward a paradigm shift in medical practice. In essence, precision medicine aims to help better understand and predict the onset and progression of disease, reactions to treatment as well as health outcomes by taking into account genetic, lifestyle and environmental variations of each individual.[2] One core component of this promising avenue is pharmacogenomics which seeks to make use of genetic information to maximize effectiveness of drugs and minimize the likelihood of their adverse events.[12-15] It is well reported that individuals respond in different ways to the same drug because of SNPs in genes encoding molecules such as drug-metabolizing enzymes, drug targets and drug transporters.[11,18,19] However, at present, most medical treatments still involve empirical selection of medications with widely varied efficacy and common side effects and this has led to mounting cases of drug-induced side effects and even adverse drug reactions in patients each year, representing a major social and financial burden.[20,22] In response of the importance of this matter, pharmacogenomics essentially hopes to improve overall healthcare by predicting an individual's genetic response to a specific medication as a means of delivering a better choice of drug with optimal safety and efficacy according to the patient's genetic profile.[5]

Despite such potential clinical benefits of pharmacogenomics, there have been many barriers for it to be widely adopted into clinical practice. This includes the cost of analysis and the turnaround time of the genetic testing.[21,25,26] Although advances in technologies have led to a continuous decline in these two factors, most conventional assays are still not economical for screening of the general population and rapid enough to be useful as a standard of care in the clinic.[21,22] These have been impediments to the routine clinical utilization of pharmacogenomic testing.[25,27] In light of this, a novel method that aims at faster and more reduced costs than the

current technologies is expected to help remove such obstacles and ease the integration of SNP genotyping into clinical practice, providing a more effective, precise and less empirical approach to healthcare.

1.2 Review of traditional SNP genotyping methods

In general, SNP analysis involves two steps: the generation of allele-specific products for SNPs of interest (allele discrimination) and their detection for genotype identification (allele detection).[41]

First, an allele discrimination step is performed to introduce specificity. There are four common methods to achieve this: hybridization, ligation, enzymatic cleavage and single base extension of a primer (SBE).[16,35] For instance, the hybridization strategy uses differences in binding energy of the probe-target duplex to discriminate between perfectly matched and mismatched strands.[31] Since the hybridization method does not involve any enzymatic reactions, it has been applied on high-throughput platforms such as microarrays.[16] However, it requires rather stringent conditions where the annealing occurs only between probe and target DNA that have perfectly complementary sequences to each other and not to the probe with a single base mismatch.[17,41] The enzymatic cleavage method is based on the activity of a particular enzyme to cleave DNA having specific cognate sequences and structures. Although probes are not required for this method, it has low throughput and can only be applied for SNP sites that are located in enzyme recognition sequences.[16,17] Assays involving ligation utilize the specificity of the ligase enzyme and require three probes in total: two allele specific probes for annealing to the template at the SNP site and one common probe that binds immediately adjacent to the allele-specific probe.[39] If the allele specific probe anneals to the SNP site with a perfect match, the enzyme ligates it to the common probe to produce a single long oligonucleotide. Unfortunately, the assay

itself is not very sensitive and thus it has not been widely adopted.[17] Lastly, the assays using the SBE strategy provide great flexibility and robustness, producing reliable results. They are based on allele-specific incorporation of nucleotides in a primer extension reaction with a DNA template using a single primer for detecting both alleles and it exploits the specificity of DNA polymerase to achieve allele distinction.[30] In contrast to the above four methods for allele discrimination, the SBE strategy is not sensitive to the overall DNA sequence and thus allows for analyzing various SNPs under similar reaction conditions, minimizing the effort for assay design and optimization.[16,17,41]

As mentioned earlier, these various allele discrimination steps are followed by allele detection for the specific base identification. There are three conventional detection methods: fluorescence, mass spectrometry and chemiluminescence.[16,41] For fluorescence based detection, fluorescence polarization (FP) and fluorescence resonance energy transfer (FRET) are the two widely used strategies. In FP measurements, the fluorophores are excited with linearly polarized light and fluorescence measurements are made parallel and perpendicular to the plane of the polarized excitation light.[5] The relative fluorescence intensity of these two measurements changes as the molecule rotates while in its excited state lifetime.[126] In general, the degree to which a molecule rotates is indicative of its size. If a fluorescently labeled probe binds to a target molecule, the rate of rotation decreases due to its higher molecular size and this preserves the fluorescence polarization. Therefore, the polarization value is higher in the bound state compared to the unbound state.[16,43] On the other hand, FRET occurs by a dipole-dipole interaction between two dyes in close proximity in which a donor fluorophore in the excited state transfers its energy to an acceptor fluorophore in the ground state.[42] Consequently, the fluorescent signature varies

depending on the distance between the two dyes and thus it can be employed where reaction results bring together or separate the neighboring donor/acceptor dyes.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is an ionization and separation technique for the mass analysis of various biomolecules. It uses a matrix of small organic molecules which absorb laser energy to generate ions from large molecules with minimum fragmentation.[16] Resulting ions are then accelerated at a fixed voltage and separate from each other based on their different mass-to-charge ratios (m/z).[32] During MALDI-TOF analysis, the m/z ratio of an ion is determined by the time required for it to travel the distance of the flight tube.[38] Unlike the above fluorescent detection methods, MS does not require any labels. However, its current drawback is the loss of signal intensity and mass resolution with increasing size of the analyte.[65]

Lastly, pyrosequencing based on chemiluminescence detection employs a cascade of enzyme reactions for SNP detection: DNA polymerase, ATP sulfurylase, luciferase and apyrase.[40] In this method, nucleotides are added one at a time to a primer extension reaction mixture and the result is analyzed each time to detect which base is actually incorporated at each step. When the nucleotide is incorporated into the primer, pyrophosphate (PPi) is released as a result of the catalytic reaction by DNA polymerase. In a series of enzymatic reactions, ATP sulfurylase uses the PPi to produce ATP, which participates in luciferase-mediated conversion of luciferin to oxyluciferin.[16] The oxyluciferin then emits visible light to an extent relative to the amount of ATP. At the end of each cycle, the unincorporated nucleotide is degraded by apyrase and the cycle can be repeated if desired with another nucleotide.[44]

Over the last decades, there have been considerable improvements towards high detection sensitivity and throughput of genotyping assays. However, these traditional detection methods

still require sizable and rather expensive equipment and thus they are not quite ideal for a point of care diagnostic platform. Novel methods working especially towards low concentration detection in small sample volume, simple and fast instrumentation, as well as high sensitivity such as single molecule readout are thus in great demand.

1.3 Nanopore detection

1.3.1 Principle of stochastic sensing of analytes in a nanopore

Nanopores are nanoscale apertures in electrically insulating materials that are used to characterize biochemical and physical properties of biomolecules at the single molecule level. In general, a single nanopore is placed in an electrochemical reservoir separated into *cis*- and *trans*-compartments with an electrode on each side, typically Ag/AgCl, and filled with a conductive electrolyte solution such as potassium chloride (KCl). Under an applied voltage, two reversible electrochemical reactions occur at the electrodes. On the *trans* side of the membrane, an oxidative electrochemical reaction $Ag(s) + Cl^- \rightarrow AgCl(s) + e^-$ takes place at the anode (+). As a result of Cl^- deposition from the solution to the electrode, an electron migrates through the wire to the amplifier. This creates a charge imbalance at the electrode, leading to cation migration towards the *cis* side of the membrane.[46] The reverse reaction $AgCl(s) + e^- \rightarrow Ag(s) + Cl^-$ occurs at the cathode (-) in which an electron is deployed from the circuit and the resulting Cl^- ion migrates towards the *trans* side of the membrane (Figure 1.1A).[46] Such ion transport across the pore and completion of an electrochemical circuit generates a stable baseline ionic current, typically at the pico-Ampere (pA) scale. The current is detected using a patch clamp system with highly sensitive electronics. In this technique, capture and subsequent translocation of charged biomolecules such as DNA are characterized by monitoring modulations in current with time (Figure 1.1B). For instance, in the presence of target analytes, the passage of the electrolyte ions

through the pore is impeded as the electrophoretic force pulls an analyte through the pore, leading to an observable drop in ionic current. As the molecule completely transits through the pore, more ions are able to move across the pore again, resulting in a rebound in the current. Different molecules generate distinct event signatures, typically represented by three parameters: the mean current amplitude of current blockades triggered by the analyte when it transits through the pore (δI), the duration of its residence in the pore [dwell time (t_d)], and the time delay between two consecutive events (δt). In general, the δI can vary due to several factors such as size, structure and net charge of the molecule in the pore.[48] The t_d is usually affected by the type and strength of the interactions between the analytes and the nanopore such as aromatic interaction or electrostatic force, and δt changes with the concentration of the molecule in solution.[46] Therefore, investigating the events in this system not only reveals the dynamic motion of the analytes in the pore but also provides valuable information about their structure and concentration.

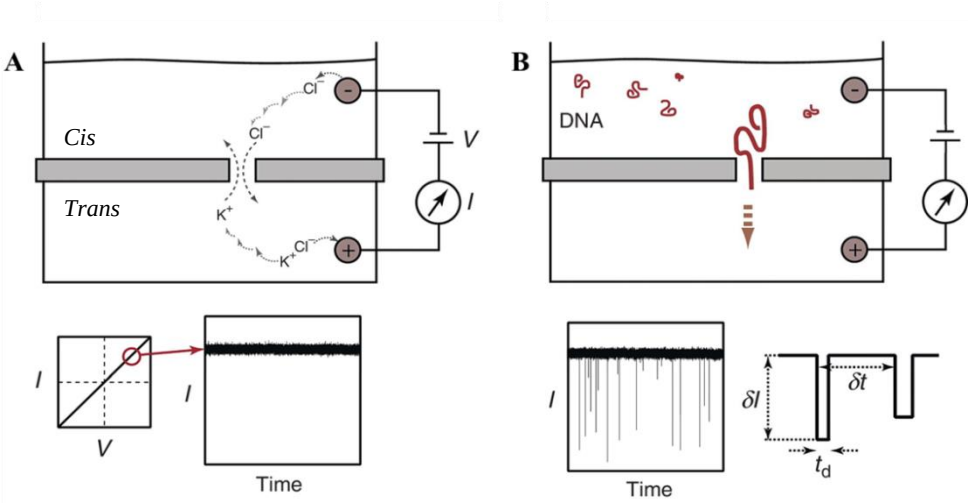


Figure 1.1 Schematic of ionic current detection of biomolecules in a nanopore. **A.** An insulating material with a nanoscale aperture divides two chambers containing an electrolyte solution such as KCl. When a voltage bias is applied across the partition, ions are transported across the pore resulting in electric current. In general, the relationship between current and voltage of a pore is linear and thus holding the potential at a constant voltage generates a steady-state DC current signal. **B.** When a charged biopolymer such as DNA is added to the grounded (*cis*) chamber, it diffuses towards the positively charged (*trans*) side. The passage of each individual molecule causes a transient drop in the ionic current through the pore. This event can be characterized by parameters such as the average event amplitude (δI), the dwell time (t_d) and the time between two consecutive events (δt).[46]

1.3.2 Types of nanopores

The nanopore may be either a naturally occurring protein channel embedded in a lipid bilayer or a synthetic nanoscale cavity fabricated in a solid state membrane. Naturally occurring biological protein pores possess advantageous features such as their atomic precision and reproducibility; each pore is atomically identical to the previous one.[45] Also, they are made in biological cells and therefore it is relatively easy to modify amino acids at defined positions or fuse other functional groups to them using genetically engineered DNA templates.[49] On the other hand, solid state nanopores are usually formed by etching pores in a thin sheet of materials such as silicon nitride (Si_3N_4) or graphene and thus their pore geometry is variable.[50] While biological nanopores are rather sensitive to temperature, pH and solution composition and also have a rather limited life span due to the fragility of lipid bilayers, solid state pores present superior mechanical, chemical and thermal robustness and they can be reused repeatedly.[51] Although the recent advances in microfabrication techniques are beginning to rival biological pores, protein pores still offer much greater flexibility for chemical and structural modification by site-directed mutagenesis.[49,50] Thus, they have been serving as the main workhorses in many applications of nanopore technology and single molecule studies.

1.3.3 Features of the alpha hemolysin pore

There are varied protein pores, each with their own distinct shape and dimensions as well as benefits and limitations. Among them, alpha hemolysin (αHL) is one of the best studied protein pores for use as a nanopore detector. αHL is a heptameric cytotoxin that is secreted by the bacterium *Staphylococcus aureus* as a water-soluble monomer.[47,52,53] Upon binding to the membrane, seven protomers self-assemble to form a 232.4 kDa mushroom-shaped heptameric transmembrane channel of defined structure consisting of three domains: the cap, rim and stem

(Figure 1.2).[52] The overall dimensions of the pore are 10 nm x 10 nm with the cap of the protein concealing a large vestibule with a 2.6 nm wide opening and 4.6 nm at its widest.[47,52] The overall stem region is 5.2 nm in length and 2 nm wide with a 1.4 nm wide limiting aperture at the junction of the vestibule and transmembrane domain.[47,52,53] This channel remains constitutively open, meaning it does not present any voltage or ligand induced gating and is also stable over a wide range of temperature and pH.[50] All these features provide αHL with significant advantages over other protein pores and thus make it ideally suited for serving as a stochastic sensing element.

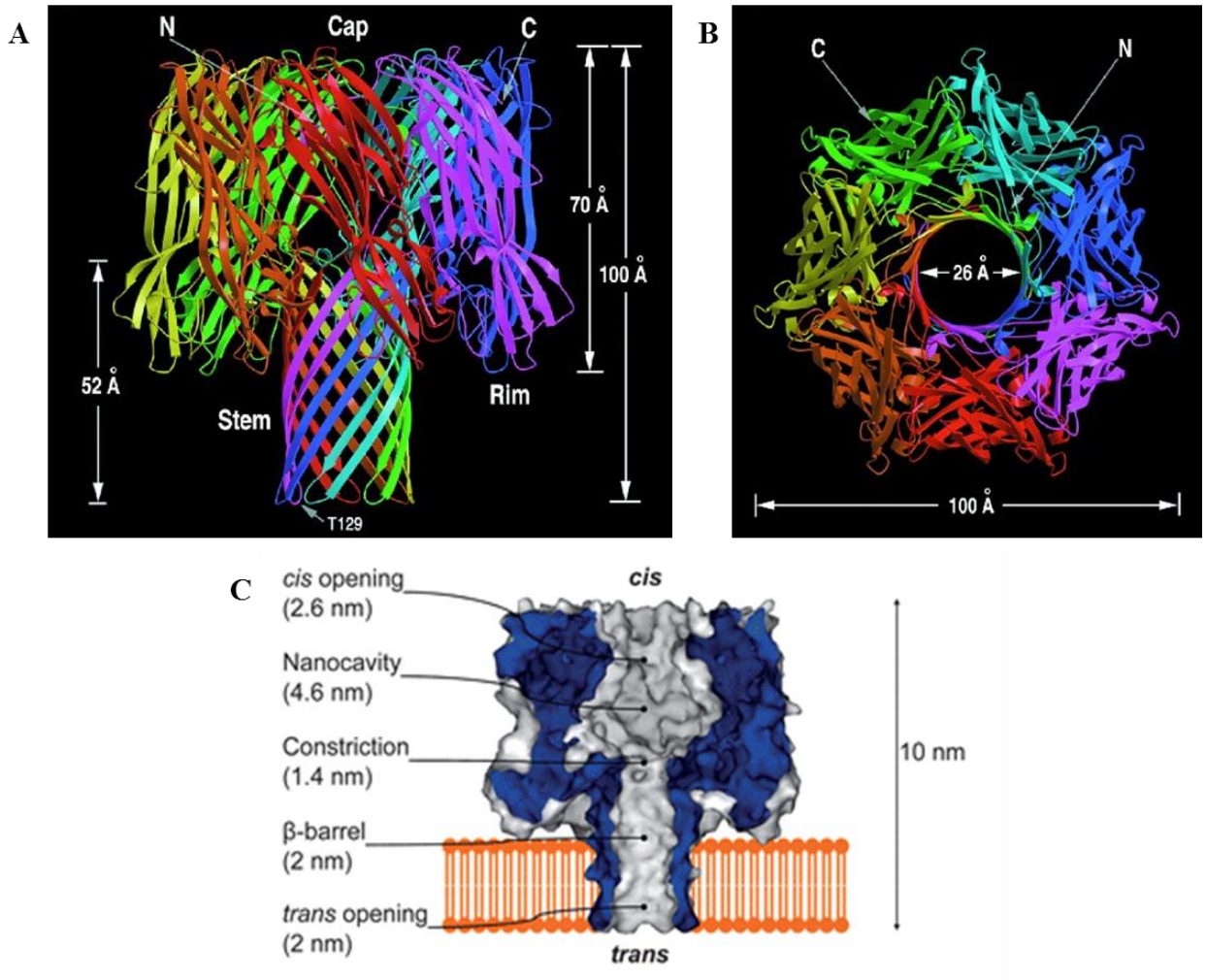


Figure 1.2 Structural details of the mushroom-shaped αHL pore. **A.** Side view of crystal structure of the oligomerized heptamer at 1.9 Å resolution.[52] **B.** Top view **C.** Detailed dimensions of the pore.[47]

1.3.4 αHL as a single molecule sensor for genomic applications

The first biomolecule sensing with αHL dates back to 1996 when Kasianowicz et al. demonstrated the electrophoretic translocation of polynucleotides through this channel.[54] As described earlier, the inner diameter of the channel is 1.4 nm and the size of a single stranded DNA (ssDNA) molecule is approximately 1.2 nm which means ssDNA can pass through the pore but double-stranded DNA (dsDNA) cannot. Because of this close correspondence between them, a single DNA molecule nearly occupies the limiting aperture as it traverses the pore and this generates a current signal that is exquisitely related to the physical and chemical properties of the molecule. Since only a single ssDNA can fit in the pore at a time, this provides a means to transduce single molecule events into observable current changes making αHL a very promising single molecule analytic tool for biomolecules, especially ssDNA and RNA (Figure 1.3).[45-48]

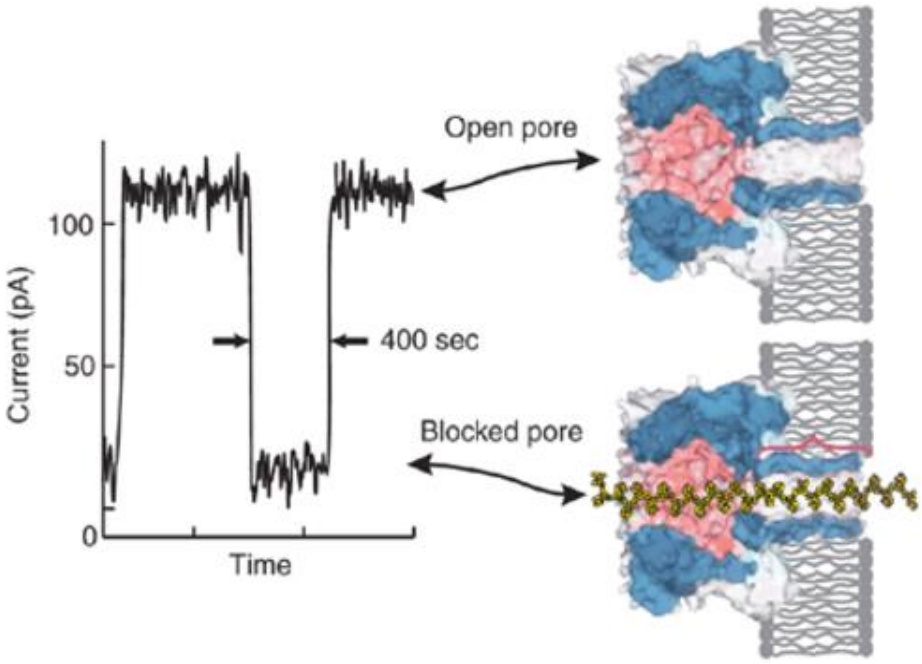


Figure 1.3 A typical current blockade signal of a strand of DNA passing through an αHL pore.[59]

To date, many interesting applications have been developed based on the use of αHL as a biosensor. Among them, the most attention has been paid to developing a single molecule medical diagnostic tool using this nanopore. Compared to traditional diagnostic platforms, a nanopore-based one is revolutionary in that it delivers sensitive single molecule resolution, requiring a very low analyte concentration and small sample volume, and also provides rapid analysis at relatively low cost which is especially important in point of care diagnostics. Of particular interest is single molecule nanopore-based DNA sequencing. Because of the ability of the αHL to translocate ssDNA in strict single file fashion, it was suggested that this channel could be an ultimate DNA sequencer if individual bases produce characteristic modulations in the current signal during translocation.[54] Unfortunately, ssDNA crosses the pore too quickly to achieve single-base resolution.[59] Thus, over the last decades, much research has gone into controlling the translocation of DNA and several variations of the approach including the use of adaptors, molecular motors, exonuclease, and polymer tagged nucleotides have demonstrated their capability for nanopore DNA sequencing.[55-61]

Besides DNA sequencing, there are other potential applications where nanopore technology could excel in diagnostics.[48] Among these, an important diagnostic application is genetic analysis of SNPs. As mentioned earlier, SNPs are the most common variants found in the human genome and implicated in various diseases and responses to treatment.[62] However, there are only a few studies which show discrimination of single nucleotide variation in a nanopore.[63-65] Although SNP analysis can be achieved by direct sequencing of DNA, reading every nucleotide in order is time-consuming and costly for SNP analysis unless it is for discovery purposes. Besides, sequencing often generates errors at a rate of one base per 100 and this is about the rate at which SNPs are found to occur in the genome. Thus, it is not quite ideal for SNP

analysis even though the accuracy can be improved by increasing sequencing depth. In particular, the relatively new nanopore-based DNA sequencing approaches are still in their infancy and thus the accuracy of current methods remains to be improved. Therefore, a more targeted nanopore assay such as focusing on clinically relevant variations would be of more benefit in terms of improving overall efficiency of care and decreasing the associated cost for screening patients.

1.3.5 Literature review for nanopore-based SNP anlaysis

To date, three studies have demonstrated SNP discrimination using nanopores; two of them are based on synthetic nanopores and the last one employs the αHL channel as a detection system.[63-65]

First, in proof-of-principle experiments, Zhao et al. have demonstrated detection of SNPs using a solid-state nanopore.[63] The method is based on the binding of a restriction enzyme to DNA and they used the pore as a local force knob to discriminate the binding energies between a restriction enzyme and its cognate sequence compared to a SNP sequence (Figure 1.4). Specifically, they measured the permeability of *EcoR*I bound dsDNA through a 2.6 nm pore in a Si_3N_4 membrane as a function of applied voltage. The pore is large enough to pass dsDNA but too small for the restriction enzyme. When the enzyme-DNA complex is drawn to the pore under the electric field, the potential in the pore pulls the DNA to the *trans* side of the membrane, generating a shear force between the enzyme and the binding sites in DNA. In this study, they observed that a threshold voltage required for breaking the bond between the enzyme and DNA depends on the sequence of the binding site and thus the congnate DNA sequence could be differentiated from SNP sites based on the differences in the required voltage to reverse the binding. As in Figure 1.4B, one nucleotide difference in the binding sequence caused the threshold potential to fall, therefore allowing SNPs to be discovered. The voltage is plotted

against the number of copies of 105 bp dsDNA that have threaded through the pore as a result of protein dissociation. The DNA was measured by quantitative PCR.

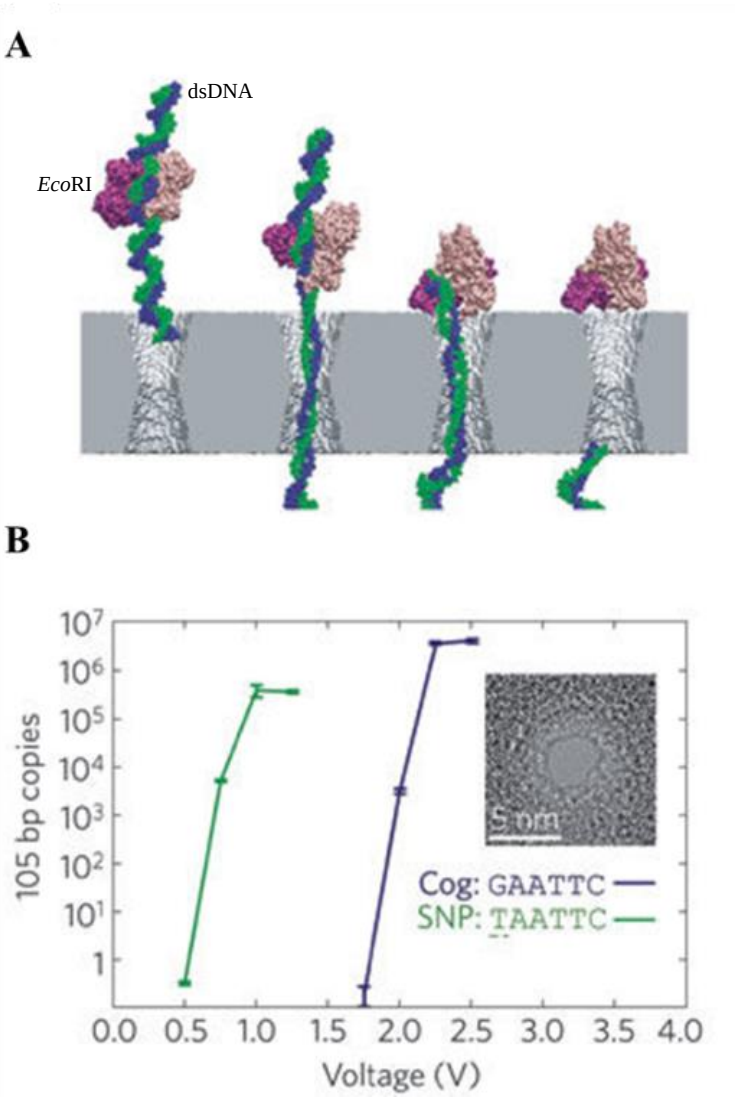


Figure 1.4 SNP detection based on the difference in binding energy between protein and DNA using a synthetic nanopore. **A.** An *Eco*RI bound dsDNA complex was electrophoretically drawn into a nanopore and then dissociated. **B.** The shearing takes place when the potential across the pore is above a threshold, thereby dissociating the enzyme and enabling the DNA to be translocated through the pore. Cog indicates cognate enzyme binding sequence.[63]

Another SNP method using a synthetic nanopore is based on probe strand displacement by the SNP allele (Figure 1.5).[64] In detail, a 7.2 kbp DNA carrier having a streptavidin (black)-biotin (yellow) complex at its 3' end is employed; this generates a unique current blockade signature when it is applied to a glass nanopore due to the large biotin-streptavidin complex. The carrier also includes a complementary sequence for the polymorphic allele (blue) and the biotin probe is placed in such a way that binding of the polymorphic but not the wild type allele to the complementary sequences displaces the biotin probe in the carrier. Without the biotin label, streptavidin no longer binds to the carrier and thus the characteristic current signal of the biotin-streptavidin complex disappears, indicating the presence of the SNP allele in the solution.

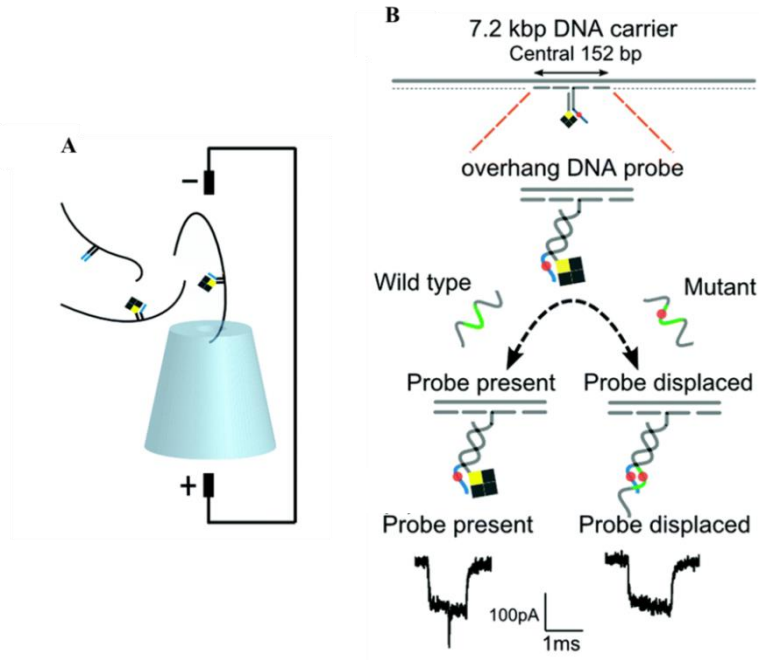


Figure 1.5 SNP detection with DNA carriers and glass nanopore. **A.**Schematic of DNA carriers passing through a nanopore under an applied potential. **B.** Schematic of the designed DNA structure for SNP detection. A pair of overhang DNA probes is placed in the middle of the carrier. The central 152bp portion is zoomed in to display the displacement event caused by the target DNA strand (green). Blue indicates a complementary sequence to the target DNA strand and the SNP site on the mutant strand is marked as a pink spot. The biotin (yellow)-streptavidin (black) complex is shown as a square. The bottom shows the two examples of current signatures when the biotin probe is present or displaced.[64]

Recently, another paper has demonstrated a method to detect a point mutation in DNA molecules using the αHL nanopore (Figure 1.6).[65] The technique uses a "nanolock-nanopore" sensor in which a sequence-specific DNA interstrand structure with metal ions called a nanolock is constructed. Specifically, it is based on the stabilization effect of metal ions such as Hg^{2+} on the T-T basepair in the DNA duplex structure. The metal bound DNA displays a distinct current signature for dehybridization in the αHL nanopore due to the effect of metal binding on the T-T basepair compared to regular A-T bound DNA. The difference in the two current signatures allows the SNP site to be detected.

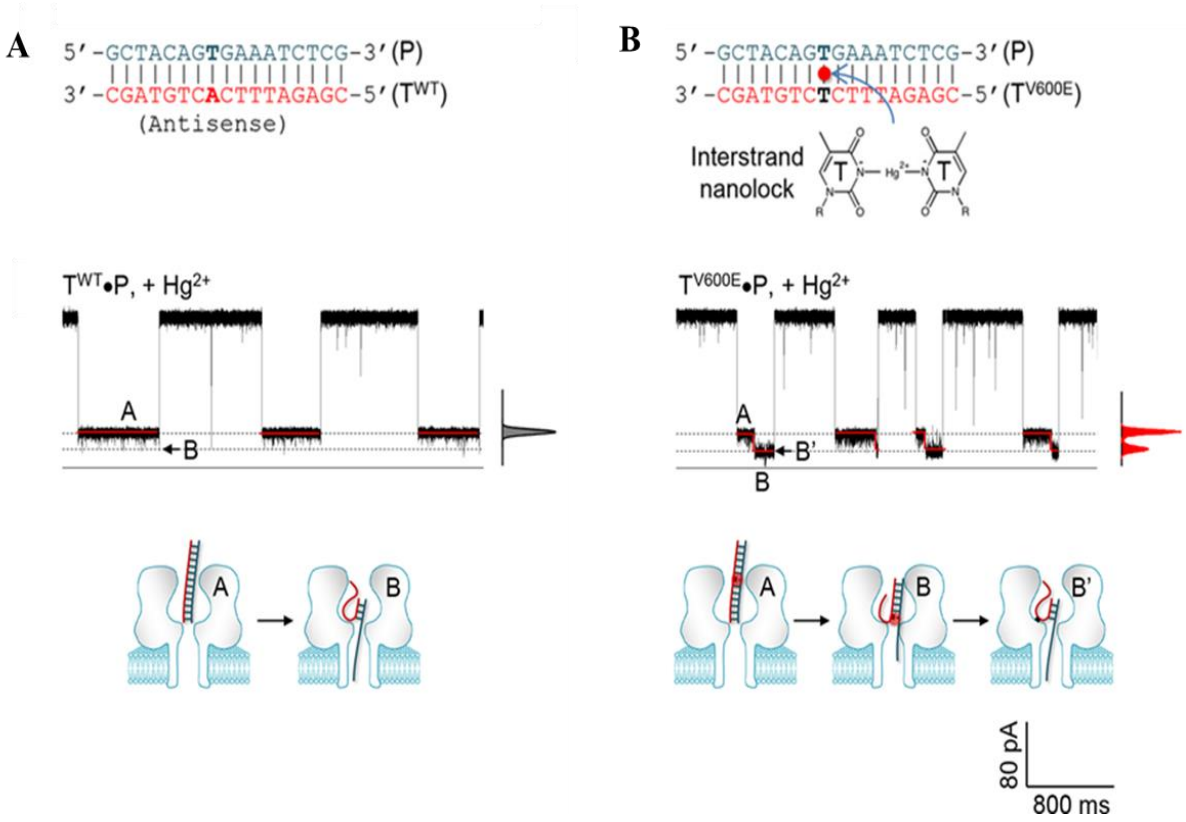


Figure 1.6 'Nanolock-nanopore' sensors for SNP detection. **A.** Sequences of probe (P) and WT strand. **B.** Sequences of probe (P) and target mutant strand forming an interstrand nanolock on the T-T mismatch with Hg^{2+}. The bottom shows the comparison of current signatures between the WT duplex and the mutant duplex carrying a nanolock. The nanolock causes a distinct pattern of signal, allowing the SNP site to be detected. Cartoons show the proposed stepwise process of unzipping of WT *vs* nanolock duplexes in the αHL pore.[65]

The above three assays have proven their concepts for nanopore-based SNP applications. Although they are all very original and intriguing, each of them can be applied under limited conditions. For instance, even though the first approach can be extended to other protein binding sites such as transcription factors or nucleases, it is still limited to these specific sequences and thus cannot be employed for other SNP sites. On the other hand, the second method is analogous to the traditional hybridization based allele discrimination and as mentioned earlier, it is often challenging to find a condition where a single-base mismatch is sufficient to avoid hybridization of the nonmatching probe.[17,41] The last method requires four different metal ions or other equivalent compounds that specifically bind to each interstrand basepair in the sequence and in this paper, the authors only demonstrated binding of Hg^{2+} to the T-T basepair. More importantly, all these assays provide only a simple binary absence or presence readout. In other words, they do not reveal base identity at the SNP site nor the genotype of alleles, whether they are homozygous or heterozygous. Therefore, another nanopore-based SNP assay that can provide full information about the variation in the sequence and genotype of alleles would be of more assistance for diagnostics.

1.4 Our approaches for single molecule electronic SNP assay using polymer tagged ddNTPs

As mentioned earlier, routine clinical practice of SNP analysis mandates simple and inexpensive instruments, fast turnaround time and an integrated and robust workflow.[27] In keeping with this demand, this thesis demonstrates a nanopore-based single molecule SNP assay that has the potential to be a favored genotyping method in a clinical setting. Basically, the fundamental structure of our assay is based on SBE of a primer using polymer tagged ddNTPs for allele discrimination and simple ionic current readout of an αHL pore for allele detection. In

an attempt to realize this, we have developed two different assay models. While the first approach is designed to be an integrated method in which allele discrimination is combined with the allele detection step, the second approach involves an initial allele discrimination step, after which the reaction products are applied to the detection system for analysis. More detailed stepwise processes and features of each method are illustrated below.

1.4.1 An integrated SNP assay in which SBE is achieved on a primer-conjugated nanopore.

With the aim of developing an assay that is preferred for a clinical setting, a combined workflow was designed by integrating the allele discrimination step into the nanopore detection system (Figure 1.7). In detail, a primer is covalently conjugated to the αHL nanopore and this serves as a platform for the entire assay. After reconstituting the primer-conjugated pore in the membrane, SBE is performed by applying a circular DNA template, tagged ddNTPs and DNA polymerase to the pore. In this reaction, depending on the sequence of the template, DNA polymerase extends the primer with one of the ddNTPs, each of which has a unique identifying molecule attached to its base. Under an applied voltage, the electric field drives the tag on the extended primer into the pore and this generates a current signal event that is specific to the identifying molecule on each of the nucleotides, thereby revealing the genotype of the template. The results for this assay design are presented in Chapter 5.

1.4.2 An alternative approach utilizing a biotin-streptavidin complex

In the first assay design, primer is covalently conjugated to the pore which is in a stationary phase on the membrane. Although this integrated design is beneficial for increasing throughput and automation of the assay in the future, immobilization of one of the enzyme's

substrates could affect overall efficiency of the enzymatic reaction due to issues of diffusion and the need to bring together the required substrates at a fixed position.

To compensate for this potential limitation of the previous design, we developed an alternative approach where SBE is completed in advance of applying it to the detection system. The key difference in this method compared to the previous one is that it utilizes the strength of the biotin-streptavidin interaction as a means of immobilizing the extended primer near the vestibule of the αHL pore. In detail, SBE is performed using biotinylated primer, tagged ddNTPs and DNA polymerase. This is followed by streptavidin treatment, after which the reaction products are applied to the nanopore system. Under an applied potential, the resulting biotin-streptavidin complex is stalled on the pore due to the large diameter of streptavidin and this essentially serves the same purpose as conjugating primer to the pore in the previous approach. The annealed primer-template is double stranded and only the single stranded tag on the extended primer is designed to be long enough to thread into the channel of the αHL pore. Hence, the electric field drives the tag on the extended primer to the pore under the applied potential and this generates a distinct current blockade signal depending on the identifying molecule on each nucleotide, revealing in turn, the genotype of the template (Figure 1.8). The data validating this assay are shown in Chapter 6.

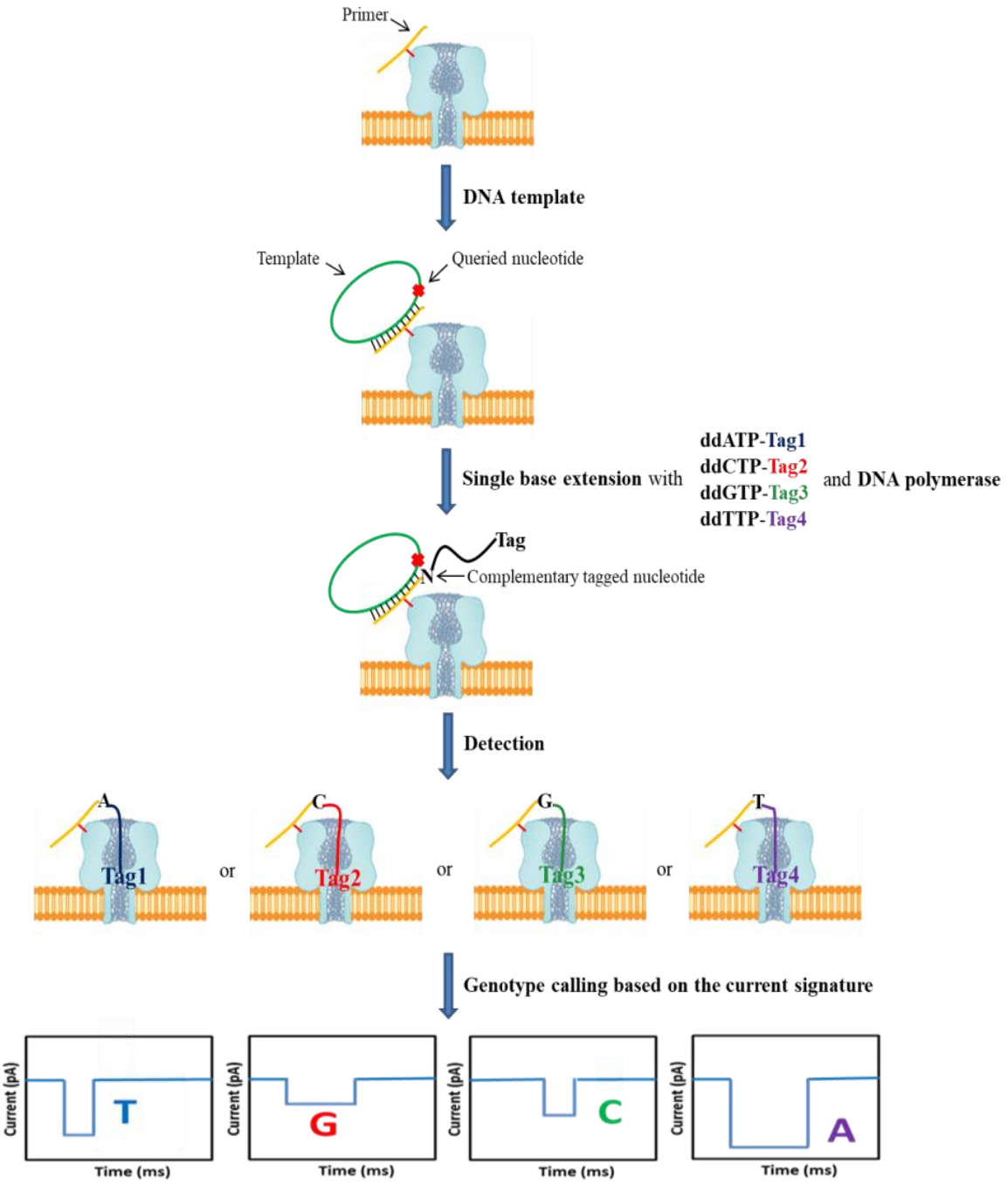


Figure 1.7 Scheme of single molecule electronic SNP assay using four polymer tagged ddNTPs in a primer-conjugated nanopore array.

An αHL nanopore that is conjugated with primer is prepared. SBE is performed by adding ddNTPs with tags on the base position, DNA polymerase and circular template. The primer is extended by the complementary nucleotide to the queried site in the template by DNA polymerase. Under an applied potential, the tag on the extended primer is drawn into the pore and this generates a unique event in the measured current signal, thereby revealing the genotype of the template.

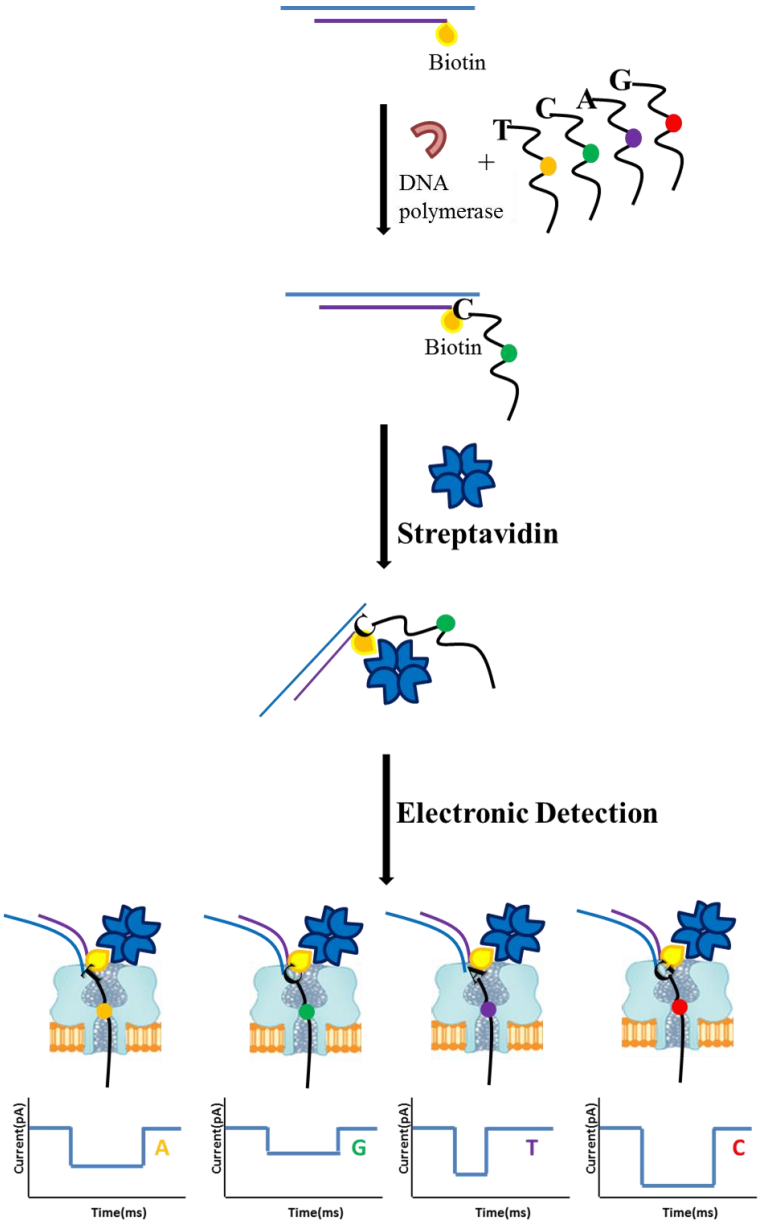


Figure 1.8 An alternative single molecule electronic SNP assay scheme employing a biotin-streptavidin complex in a nanopore array.

SBE is performed with a biotinylated primer, template, DNA polymerase and tagged ddNTPs. The reaction products are subsequently bound to a streptavidin molecule and analyzed in the nanopore system. The biotin-streptavidin complex positions the tag on the extended primer near the pore, allowing the tag to be pulled into the pore under the applied potential. The ensuing current signals are monitored to decipher the base identity in the template.

1.5 The Genia nanopore array

The instrumentation used for nanopore experiments in this thesis was developed by Genia Technologies, Inc. The key part of their system is a complementary metal-oxide semiconductor (CMOS) integrated circuit containing hundreds of individually addressable Ag/AgCl electrodes isolated by independent lipid bilayers (Figure 1.9A).[56] The overall design of the chip is that it consists of one shared fluidic chamber (the *cis* compartment) and membrane-isolated 5 μm diameter wells with a silver electrode deposited at the base of each well (the *trans* compartment). The surface of the chip is covered with SiO_2 and over-coated with C_{18} silane to create a hydrophobic support for building lipid bilayers across the top (Figure 1.9B). All chip functions, voltage application and current measurements are controlled on a single reader circuit board. The fluidic chip can house hundreds of pores in parallel and insertion of a single nanopore per membrane is achieved using individual feedback circuits which monitor currents generated from single pores. First, microgram amounts of phospholipid dissolved in alkanes are applied to the chip. Membrane formation is achieved using an automated lipid spreading protocol utilizing a syringe pump to apply iterative buffer and air bubble flow to mechanically thin the membrane until a single lipid bilayer is shaped on a C_{18} silanized surface. The electrical capacitance, which is directly associated with the structural integrity of the membrane, is measured with each electrode by applying voltage across the membrane during this step. An empirically set capacitance threshold value of 5 $fF/\mu m^2$ is employed to classify the properly formed single lipid bilayer and the thinning process ends when a majority of wells exhibit this value. Once the membrane is formed over the electrode, αHL is pumped into the system and the instrument begins the automated pore insertion step which consists of a linearly increasing voltage gradient ranging from 50 mV to 600 mV in 50 mV increments. The increasing potential stresses the lipid

bilayers, potentiating self-insertion of αHL pores into the bilayer. During this step, the reader board screens the current across each membrane and a sudden rise in the current is considered a pore insertion event based on the increase in conductance across the membrane. When this occurs, the voltage on the specific membrane is turned off to avoid additional pore insertion events. In this way, the likelihood of multiple pore insertions in a single well is minimized. At this point, analytes can be flowed into the *cis* chamber to carry out the desired assay.

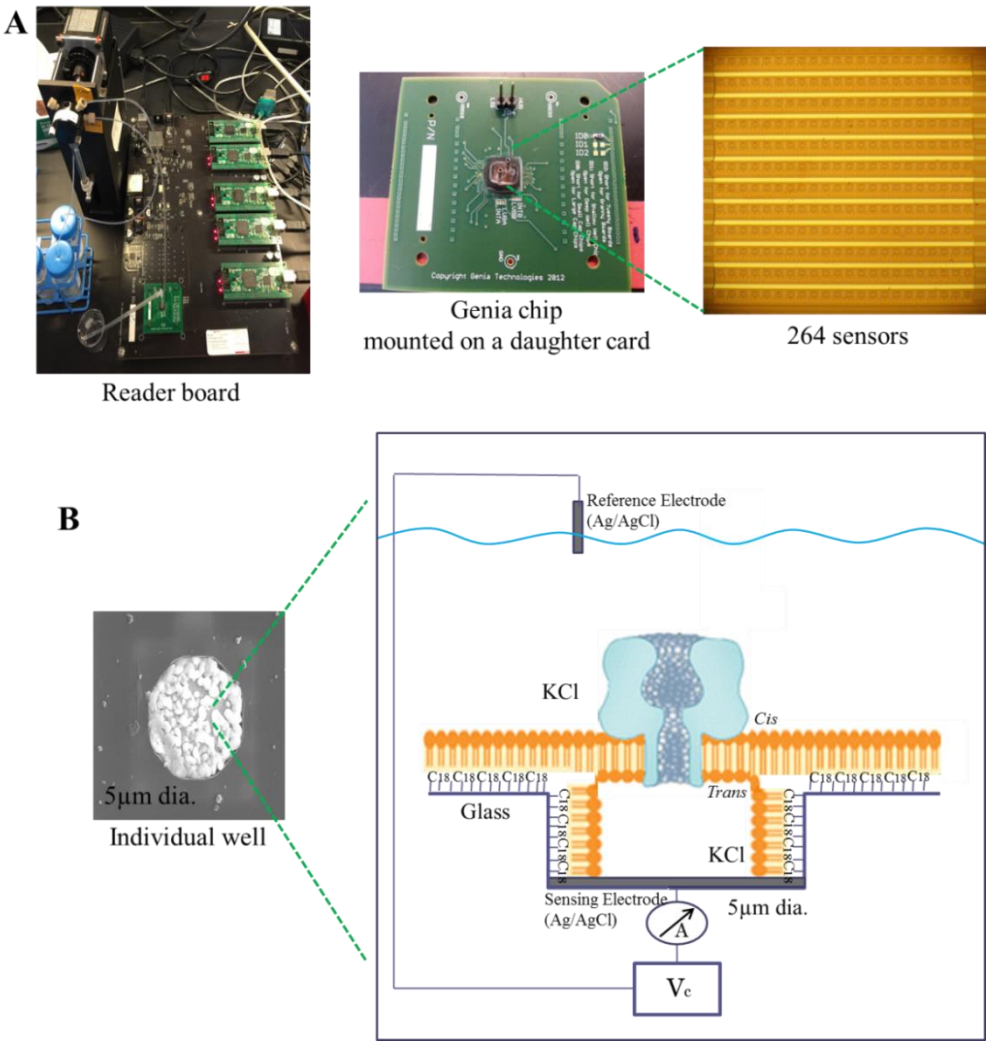


Figure 1.9 Genia nanopore array. **A.** Its reader board and integrated circuit with 264 sensors. **B.** Each sensor contains a 5µm diameter silver electrode at the bottom of a well. A lipid bilayer covers the opening of the well and a single αHL nanopore completes the electrical circuit between the sensing and reference electrodes.

CHAPTER 2: Design and Synthesis of Polymer-Tagged ddNTP Analogs for DNA Polymerase Extension

2.1 Introduction

In our single molecule electronic SNP assay, the capture of a unique identifying molecule that is tagged to each nucleotide generates a distinct current impedance signal in the nanopore system. To realize this concept, four nucleotide analogs possessing distinct tags were designed and synthesized. Previous studies from our laboratory have demonstrated the synthesis of deoxynucleotides (dNTPs) with various polyethylene glycol (PEG) or oligonucleotide polymers with structural variations.[56,68] In those studies, the polymer tags were attached to the terminal phosphate position of the nucleotides and this enabled the tags to be released along with pyrophosphates upon their incorporation into DNA for continuous sequence determination. However, in SNP analysis, only a single base in the original template is of interest and thus further extension is not necessary. Thus, we employed ddNTPs for our SNP assay instead of dNTPs; ddNTPs lack the 3'-hydroxyl group which is crucial for creation of a phosphodiester bond between the incorporating nucleotide and the primer end. Hence, further extension of primer ceases once a ddNTP is included in a growing chain. In addition, since only one base in the template is of concern, the tag molecule does not have to be removed for the next cycle of extension. This means that the tags can be permanently incorporated into the primer by conjugating to places other than the terminal phosphate on the nucleotide. Lastly, our method relies on DNA polymerase to produce allele-specific SBE products using polymer tagged ddNTPs. Therefore, it is imperative to find a suitable position for attaching the tag molecules in

the ddNTPs where they do not disrupt the active sites of DNA polymerases and conserve the DNA structure during nucleotide incorporation.

In view of all these points, an ideal tag conjugation site on ddNTPs was investigated. Previously, a study has shown that certain modified DNA polymerases can accept nucleotide analogs having large groups at the 5-position of pyrimidines (C and U) and the 7-position of purines (A and G).[69] In support of this, the X-ray crystallographic structure of the ternary complex formed by a rat DNA polymerase, a DNA template-primer, and ddCTP has revealed that these specific positions in each base ring are relatively available for sizable tag attachment (Figure 2.1).[70] Indeed, bulky fluorescent dyes have been placed at these positions and they have been used extensively for several genomic applications such as DNA labeling and sequencing.[71-74] Based on these studies, we decided to conjugate a tag to the 5-position of C/U and the 7-position of A/G.

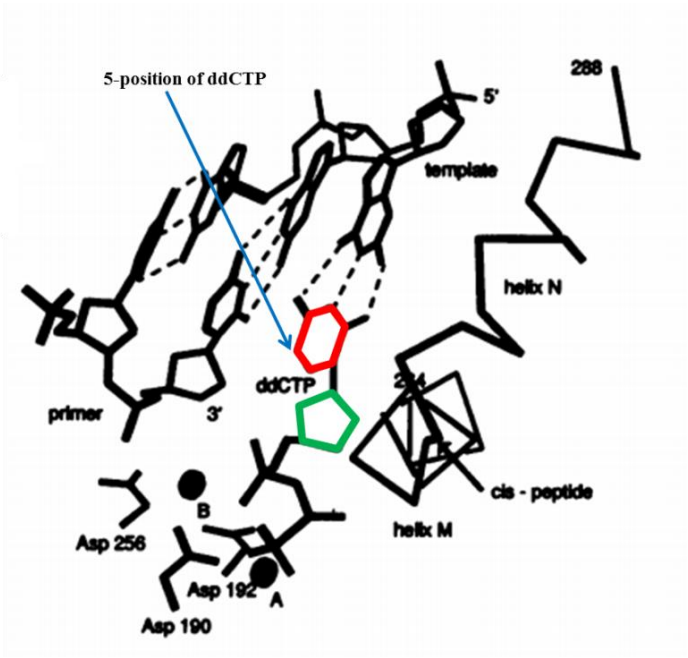


Figure 2.1 X-ray crystallographic structure of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. Green indicates the sugar ring and red marks the pyrimidine base ring in ddCTP. The 5-position of ddCTP (arrow) has a relatively large space for modification with a bulky group.[69]

After synthesizing these tagged ddNTPs, their applicability as substrates for DNA polymerase was tested using a self-primed looped DNA template. In quest of a proper enzyme that can tolerate our newly synthetized tagged ddNTP analogs, five DNA polymerases that have been shown to accept ddNTPs were investigated. These include two Therminator DNA polymerases, mutant versions of the native DNA polymerase from *Thermococcus* species 9°N-7, which have been engineered to possess an improved ability to catalyze incorporation of modified substrates such as ddNTPs, ribonucleotides or other nucleotide analogs with variations at the 3´ ribose position. Therminator II contains D141A / E143A / Y409V / A485L mutations and Therminator III has D141A / E143A / L408S / Y409A / P410V mutations.[75,76] These distinct combinations of mutations enable them to tolerate various nucleotide analogs and incorporate them into a growing primer. The Klenow fragment (exo-) is an N-terminal truncation of DNA polymerase I which retains polymerization but has had its 5' → 3' and 3' → 5' exonuclease activities removed. It has been widely used to sequence DNA using the method of Sanger which employs ddNTPs.[77] Sequenase 2.0 is a genetically engineered version of bacteriophage T7 DNA polymerase which has been shown to incorporate ddNTPs at the same rate as dNTPs. Finally, Thermo Sequenase is a Taq DNA polymerase variant with a genetic change that abolishes discrimination against ddNTPs.[78-82] Moreover, previous studies have shown that the use of the catalytic manganese ion (Mn^{2+}) instead of magnesium ion (Mg^{2+}) can increase the efficiency of incorporation of ddNTPs by some of the enzymes listed above and thus additional extension reactions with Mn^{2+} were also examined.[83-85] Using the enzyme that exhibited the best incorporation efficiency from the above five candidates, a kinetic fluorogenic assay was performed to estimate the efficiency of our newly tagged ddNTP analogs as substrates for DNA polymerase compared to unmodified ddNTPs.

2.2 Methods

2.2.1 Synthesis of polymer tagged ddNTPs (*cf.* Figure 2.4)

Oligonucleotide polymers were conjugated to the 5-position of pyrimidines (C and U) and 7-position of purines (A and G). To synthesize these molecules, first, 5-proparglyamino-ddNTPs for C/U and 7-proparglyamino-ddNTPs for A/G (**1-4**) were extended with amino caproic acid-N hydroxysuccinimide (NHS) ester to install an amino extended chain on the ddNTPs (**5-8**) as described by Duthie et al.[88] The products were purified by reverse phase HPLC. In order to attach polymer tags using a click chemistry approach, based on a cycloaddition reaction between an alkyne and an azide, an azido group was introduced by reacting the extended ddNTPs with azidobutyric acid-NHS ester. The reaction mixture was stirred overnight at room temperature and purified by HPLC using 0.1 M triethylammonium acetate (TEAC) buffer (pH 7.5) and an acetonitrile gradient to provide azido extended ddNTPs (**9-12**). Each 5'-hexynyl-oligonucleotide tag (custom synthesized by TriLink, 500 nmol in 200 µl H_2O) was reacted with the corresponding ddNTP-N_3 nucleotide (750 nmol) followed by the addition of copper bromide (50 µl, 0.1 M solution in 3:1 DMSO/*t*-BuOH) and tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) (100 µl, 0.1 M solution in 3:1 DMSO/*t*-BuOH). The reaction mixture was stirred at 40 °C for 16 hr and purified by HPLC using 0.1 M TEAC buffer (pH 7.5) and an acetonitrile gradient.

2.2.2 Screening several DNA polymerases for incorporation of the polymer tagged ddNTPs

In search of a polymerase that can recognize and incorporate our polymer tagged ddNTP analogues, five DNA polymerases (Therminator II, Therminator III, Klenow, Sequenase 2.0 and Thermo Sequenase) that are known to have the ability to incorporate modified ddNTPs were screened for SBE with self-primed looped templates and different ratios of our polymer tagged ddNTP relative to the templates. Sequences of these templates are as follows: 5′-CGCGGCGCGGTTCCGCGCCGCGAGCT-3′ for T and 5′-CGCGGCGCGGTTCCGCGCCGCGGCTA-3′ for C. The reactions were performed at 37˚C for 1 hour in a 20μL volume containing 1 μM template, 4 units of Thermo Sequenase (Affymetrix) and 20 or 30 μM tagged ddNTPs. Reactions with 5 units of Klenow(exo-) (New England Biolabs), 13 units of Sequenase 2.0 (Affymetrix), or 4 units of Therminator II or III (NEB) having the same concentration of the template and the tagged ddNTPs were also investigated. The extension products were analyzed by 8M urea 15% polyacrylamide gel electrophoresis.

2.2.3 Kinetic measurements of nucleotide activity using a fluorogenic assay

Sequence of the primer is shown in Figure 2.7. Extension reactions were prepared with the primer, Thermo Sequenase, dCTP and various dilutions of ddCTP or tagged ddCTP. First, a series of SBE reactions were carried out with final ddCTP concentrations ranging from 0 to 10 μM at a concentration of dCTP of 10 μM and 0.3 μM of labeled primer. Under these conditions, the primer is extended by the incorporation of dCTP until it encounters ddCTP incorporation. Next, the same experiments were performed in the presence of the polymer tagged ddCTP at

concentrations ranging from 0 to 5 μM with 5 μM dCTP and 0.3 μM labeled primer. Fluorescence intensity was measured on a SpectraMax3 Microplate Reader (Molecular Devices, LLC).

2.3 Results and Discussion

2.3.1 Design and synthesis of linear oligonucleotide tagged ddNTPs

In our assay, allele discrimination is achieved by capturing the tag on the extended primer in a nanopore system. Therefore, in order to generate characteristic current blockade signatures, several distinct oligonucleotide polymers having various structural modifications were attached to the 5 position of C/T and 7 position of A/G. In general, each oligonucleotide tag starts with a hexynyl moiety at the 5' position for conjugation to azido-ddNTPs followed by unique sequential units of molecules as identifiers to generate different current blockade signatures. This includes relatively neutral phosphodiester building blocks, different numbers of repeated abasic nucleotides or simple spacers that have a smaller diameter. These differences in charge and size create unique current impedance signals when the tags thread through an αHL nanopore in response to an applied voltage. These modifications are envisioned to be localized to the narrowest constriction zone in the αHL pore where they will directly affect ionic current and thus generate the largest changes in current amplitude. Figure 2.2 shows details of the 4 tagged ddNTPs that we selected based on their unique current blockade levels from the nanopore experiments. Three nucleotides contain different numbers of abasic nucleotides, deoxyribose phosphate units (dSp) (3, 8 or 17), and one nucleotide has a chain of relatively neutral molecules, 6 inosine methyl phosphonate (Imp) linkages within the tag as identifiers. An optional Cy3 dye was included in the tag design as a gel electrophoresis tracer to verify that the primer was extended with the polymer tagged ddNTP.

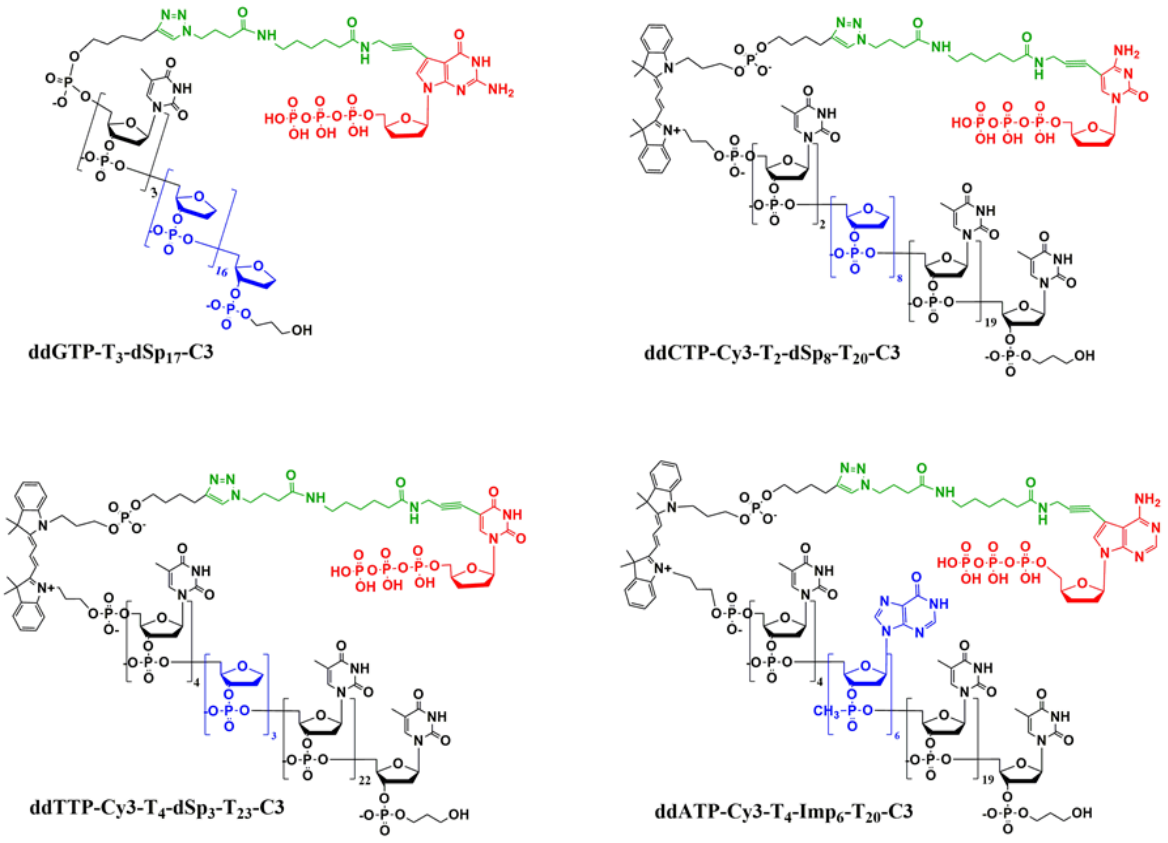


Figure 2.2 Structures of polymer tagged ddNTPs. A nucleotide (red) is attached to a linker (green) and an oligonucleotide polymer tag having an optional Cy3 dye and a dT chain as a backbone (black) with interspersed abasic nucleotides (dSp) or inosine methyl phosphonates (Imp) as identifiers (blue).

To synthesize these tagged ddNTP molecules, we employed the copper catalyzed azide-alkyne 1,3-dipolar cycloaddition chemistry system that generates a 1,2,3-triazole (Figure 2.3). In detail, 5-aminopropargyl-ddNTPs for C/T and 7-aminopropargyl-ddNTPs for A/G (**1-4**) were reacted with aminocaproic acid-NHS ester in order to produce an extended chain for the ddNTPs (**5-8**). They were subsequently reacted with azidobutyric acid-NHS ester to generate azido terminated ddNTPs (**9-12**) that were eventually coupled with an alkyne moiety at the 5' position of each tag through the azide-alkyne click chemistry system (**13-16**) (Figure 2.4). Detailed reaction conditions are provided in the Methods section. The synthesized molecules were

purified by HPLC and the final products were characterized by MALDI-TOF MS analysis. MALDI-TOF mass data for the four tagged nucleotides are shown in Figure 2.5.

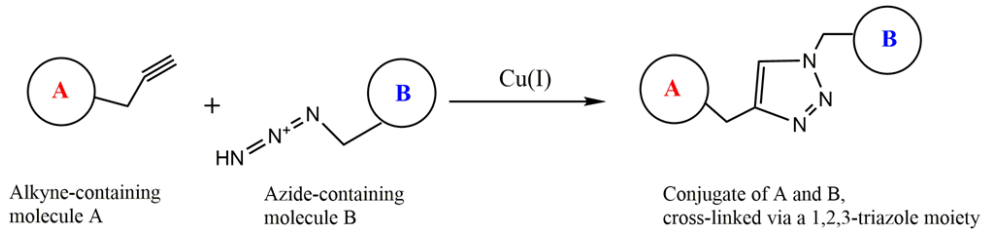


Figure 2.3 Cu (I)-catalyzed 1,3-dipolar Huisgen cycloaddition of alkyne and azide.

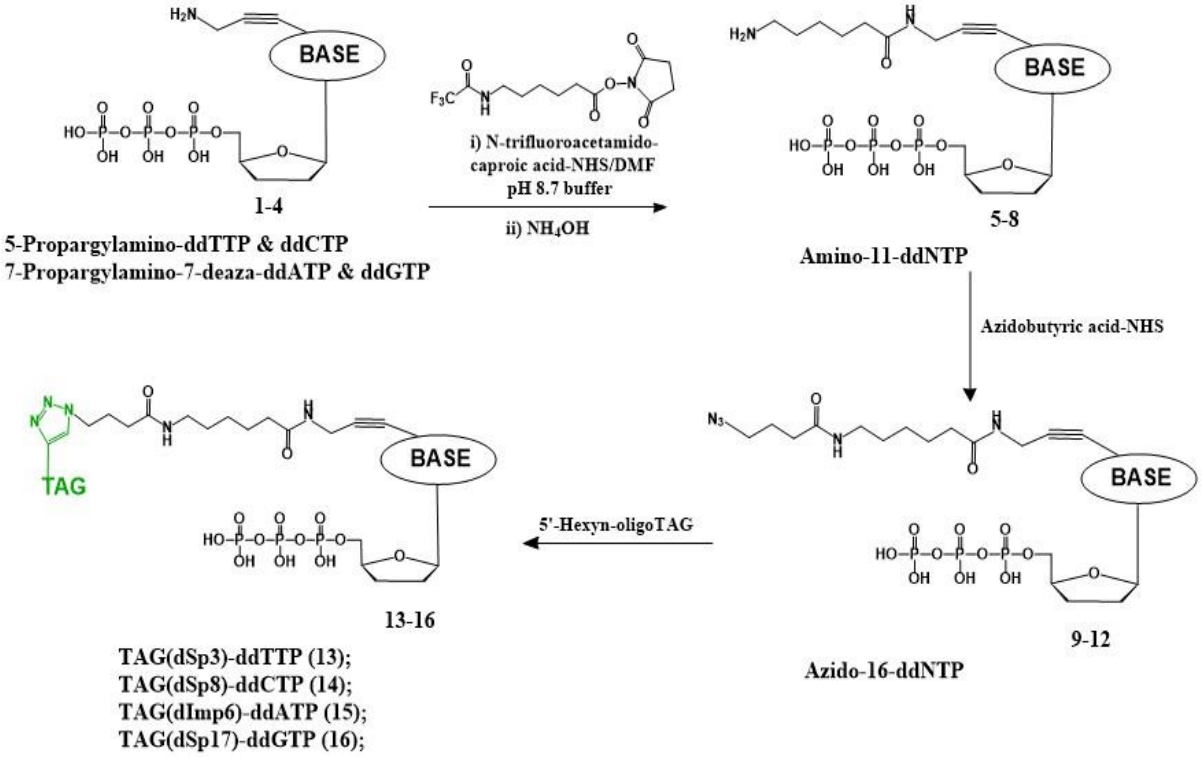


Figure 2.4 Synthesis scheme for the four polymer tagged ddNTPs.

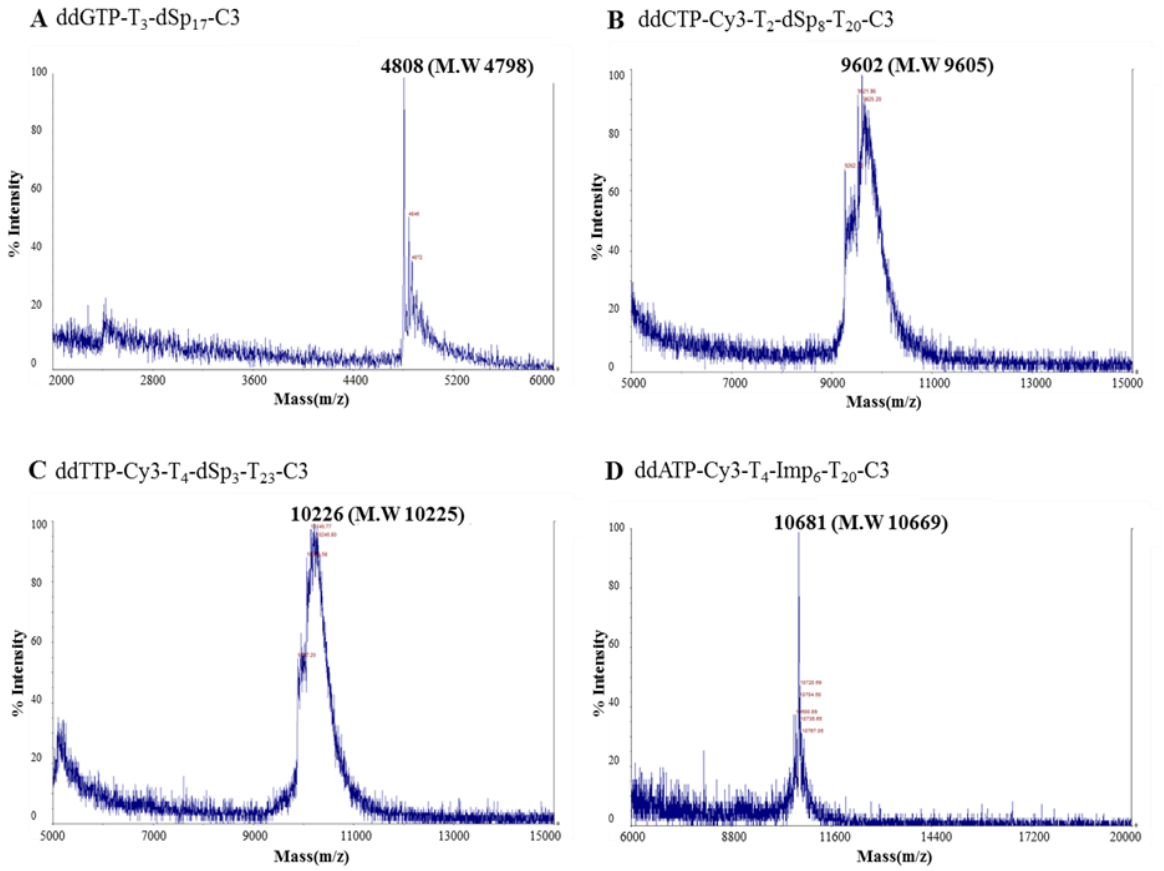


Figure 2.5 MALDI-TOF MS spectra of the four tagged ddNTPs. The expected molecular weights of these nucleotides are (A) 4798 Da for ddGTP-T_3-dSp_{17}-C3, (B) 9605 Da for ddCTP-Cy3-T_2-dSp_8-T_{20}-C3**,** (C) 10225 Da for ddTTP-Cy3-T_4-dSp_3-T_{23}-C3, (D) 10669 Da for ddATP-Cy3-T_4-Imp_6-T_{20}-C3. The measured values are shown in the spectra.

2.3.2. Surveying polymerases for incorporating base labeled polymer tagged ddNTP analogs

After synthesizing these nucleotides, five DNA polymerases (Therminator II, Therminator III, Klenow(exo-), Sequenase 2.0 and Thermo Sequenase) that have shown the capability to accept modified ddNTPs were tested for incorporating our base labeled polymer tagged ddNTP analogs.[75-82] The extension reactions were performed using self-primed looped templates and varying enzymes or catalytic ions. Tagged ddCTP or tagged ddTTP were

employed for the screening as examples. The sequences of the templates and detailed reaction conditions are provided in the Methods section. The extension results were compared and analyzed by urea-PAGE analysis. To ensure that the extended primers contain the tagged ddNTPs, the molecules on the gel were first illuminated under UV light without any DNA staining. As mentioned earlier, an optional Cy3 dye was included in the tag design and in those cases the primer extended by tagged ddNTP also retains the dye. The gels on the left in Figure 2.6 show the two species that contain a Cy3 dye: the residual tagged ddNTP and the extended products. The presence of an extra band containing Cy3 dye on the gel confirms that the primer was extended by the expected molecule. The gels on the right show the results after DNA staining with SYBR® Gold to display all DNA containing molecules. Figure 2.6A shows reaction results with Therminator II (T2) and Therminator III (T3) DNA polymerases. Figure 2.6B displays the results with Klenow (exo-), Sequenase 2.0 and Thermo Sequenase enzymes. As shown in the figure, Thermo Sequenase revealed the best incorporation efficiency with our polymer tagged ddNTPs among the tested DNA polymerases, indicating it does not discriminate against them. While Klenow, Therminator II and III polymerases demonstrated modest incorporation results, Sequenase 2.0 did not show any extension of primer by our tagged nucleotides. Also, previous studies have shown that certain enzymes preferentially utilize Mn^{2+} instead of Mg^{2+} ions.[82,83] Thus, these enzymes were also tested in reactions containing Mn^{2+}. A summary table of the incorporation results by all DNA polymerases tested with a specific catalytic ion is provided in Table 2.1. The detailed reaction conditions are provided in the Methods section.

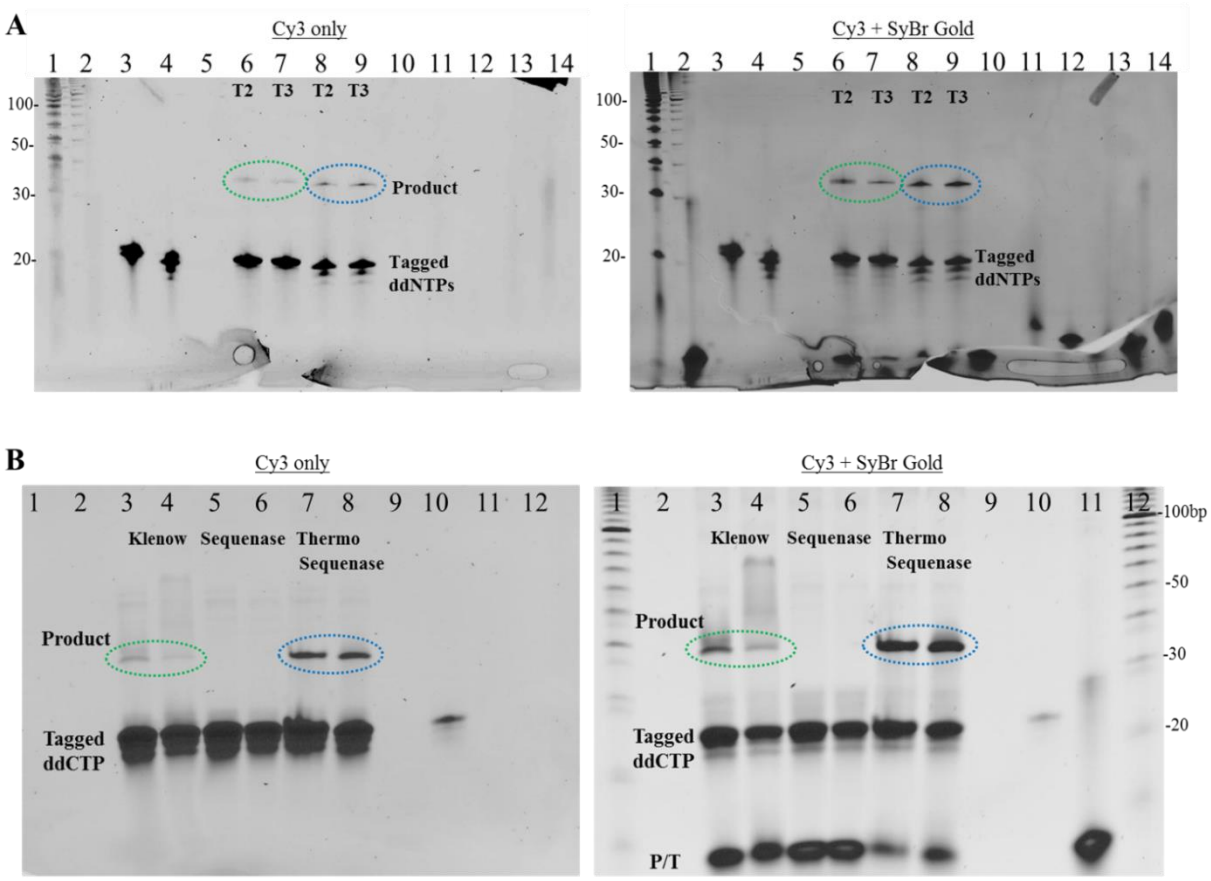


Figure 2.6 Screening five DNA polymerases for incorporation of tagged ddNTPs. Reactions were analyzed on a denaturing polyacrylamide gel. The gels were first illuminated under UV light without any DNA staining to monitor molecules having a Cy3 dye (left) followed by DNA staining (right). **A.** Therminator II (T2) and III (T3) were tested for SBE with self-primed looped template and tagged ddTTP or tagged ddCTP. Lanes 6 and 7 show the incorporation results of tagged ddTTP with either T2 or T3. Lanes 8 and 9 are the reaction results with tagged ddCTP. Lane 2, 3 and 4 are controls: the template, tagged ddTTP and tagged ddCTP respectively. **B.** SBE reaction results with Klenow, Sequenase 2.0 and Thermo Sequenase with different ratios of the template relative to the tagged ddCTP (1:20 and 1:10). Lanes 3 and 4 show the results with the Klenow enzyme. Lanes 5 and 6 are with Sequenase 2.0 polymerase and lanes 7 and 8 are with Thermo Sequenase polymerase. Lane 9 is a blank lane. Lanes 10 and 11 are controls with tagged ddCTP and the template respectively.

DNA Polymerase	Efficiency
Therminator II (Mn^{2+})	+
Therminator III (Mg^{2+})	+
Klenow (Mg^{2+})	+
Thermo Sequenase (Mn^{2+})	+++
Thermo Sequenase (Mg^{2+})	++++
Sequenase 2.0 (Mg^{2+})	-
Sequenase 2.0 (Mn^{2+})	-

Table 2.1 A summary table showing the relative incorporation efficiency of tagged ddNTPs by the tested enzymes.

2.3.3 Kinetic measurements of activity of tagged ddNTP analogs as substrates for Thermo Sequenase

A simple kinetic enzymatic assay was performed to test how efficiently our polymer tagged ddNTP analogs are consumed as substrates for Thermo Sequenase compared to unmodified ddNTPs. The assay is based on fluorescence polarization of a self-primed template labeled at its 3' end with a single fluorophore. A previous study has shown that the fluorescence intensity of the primer changes as DNA polymerases extend it.[86] The sequences of the template stretches of the molecules are composed of 7 identical nucleotides (Figure 2.7) and thus, in this design, primer can be extended by between one and seven bases, depending on the ratio of dNTPs and ddNTPs added in the mixture. With this template, control extension reactions were performed with an increasing concentration of ddCTP and a constant concentration of dCTP. Based on the fluorescent intensity emitted in each experiment, a relative IC_{50} of ddCTP was estimated by measuring the concentration of ddCTP required to bring the intensity down to the

half-way point between the maximum and bottom plateaus. The same experiment was repeated with our tagged ddCTP and its relative IC_{50} was compared to that of unmodified ddCTP.

Figure 2.7 Structure of oligonucleotides used in this experiment. The fluorescein label (FL) is attached to the 3′-most T residue. The sequence of the single-stranded part of the oligonucleotide consists of seven identical bases (G's shown here) and thus allows the enzymatic extension of the 3′ end by between one and seven bases.

Figure 2.8 shows a plot of fluorescence intensity versus time. First, a series of SBE reactions were carried out with final ddCTP concentrations ranging from 0 to 10 µM with 10 µM dCTP and 0.3 µM labeled primer. Under these conditions, the primer is extended by the incorporation of dCTP until it encounters ddCTP incorporation. As shown in Figure 2.8A, the concentration of unmodified ddCTP that produces the half maximum fluorescent intensity is approximately 0.5 µM. Next, the same experiments were repeated in the presence of our polymer tagged ddCTP at concentrations ranging from 0 to 5 µM with 5 µM dCTP and 0.3 µM labeled primer. The result shows that the apparent IC_{50}, a concentration of nucleotide required to bring the intensity down to the half-way point, of polymer tagged ddCTP is approximately 0.25 µM (Figure 2.8B). Thus, we concluded that our tagged ddCTP analog has comparable kinetic properties to the unmodified ddCTP at least with the DNA polymerase tested.

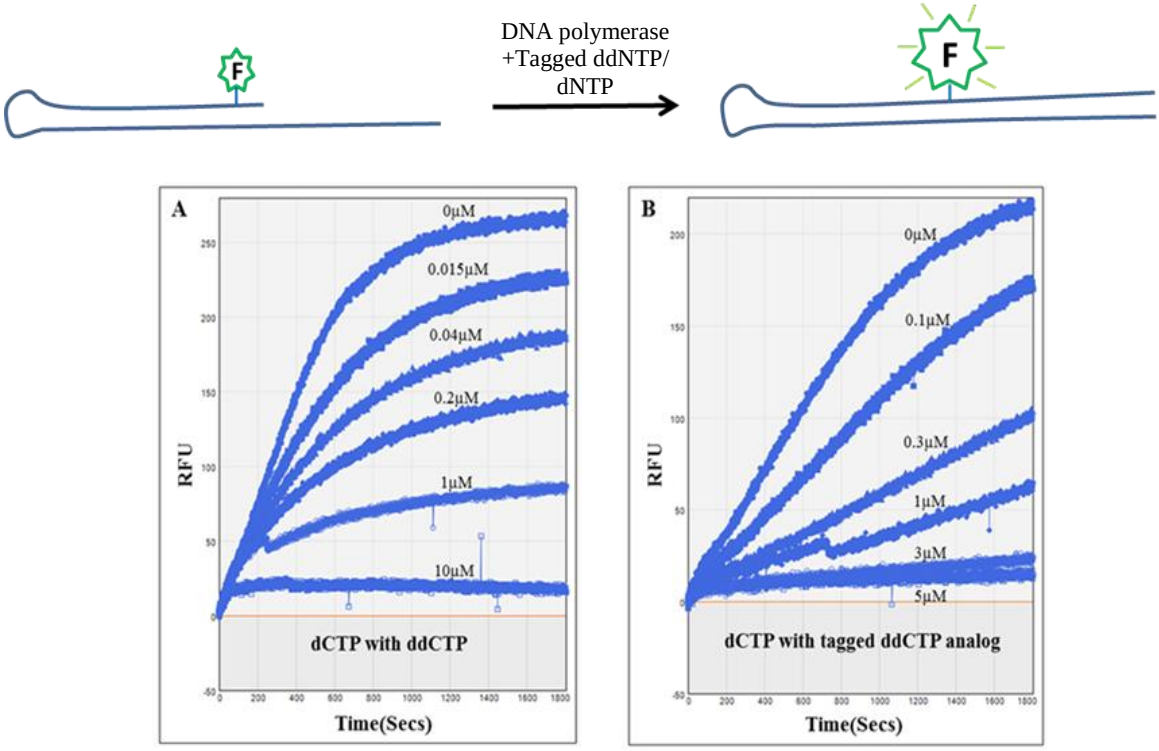


Figure 2.8 Fluorescence intensity change *vs* time using self-primed looped primers having a fluoresceinated dT near their 3' ends. **A.** Varying concentrations of ddCTP (0 to 10 μM) were added to solutions containing 0.3 μM single fluorescein labeled self-primed template and 10 μM dCTP. The concentration of unmodified ddCTP that produces the half maximum fluorescent intensity is approximately 0.5 μM. **B.** The same experiments performed with the tagged ddCTP (0 to 5 μM) in the presence of the same template and 5 μM dCTP. The result shows that the apparent IC_{50} of polymer tagged ddCTP is approximately 0.25 μM.

2.4 Conclusion

In most currently available genomic assays, DNA polymerase is required to catalyze necessary biochemical reactions for determining the identity of bases in DNA sequences and the fundamental difference among these assays is the type of nucleotide substrate incorporated.[21] In other words, the capability of DNA polymerase to process certain nucleotide analogs lies at the heart of the assays.

In this chapter, novel ddNTP analogs labeled with oligonucleotide polymer tags were synthesized and various DNA polymerases were investigated for their ability to accept our new

compounds. Our results indicate that Thermo Sequenase with Mg^{2+} exhibits the best incorporation efficiency of our polymer modified ddNTPs and the kinetic experiment shows that the rate of the enzymatic reaction was barely affected despite the presence of long tags on the nucleotides. As mentioned earlier, Thermo Sequenase is an engineered version of Taq DNA polymerase. In this enzyme, the phenylalanine at amino acid (AA) position 667 in the original Taq polymerase is replaced with tyrosine and studies have shown that this single residue substitution eliminates discrimination against ddNTPs.[79] Since this discovery, many studies have demonstrated the incorporation of ddNTPs labeled with fluorescent molecules by this enzyme.[80,81] In line with these previous results, our data indicate that even a rather extensive modification such as a 20-30 unit long polymer attached to the 5 position of pyrimidine and 7 position of purine bases does not interrupt the catalytic activity of Thermo Sequenase and thus they can be efficiently incorporated into the growing chain by this enzyme.

CHAPTER 3: Building DNA Conjugated Nanopores: Strategies for Mutagenesis, Conjugation and Purification

3.1 Introduction

In one of our approaches for the single molecule electronic SNP assay, a ssDNA molecule is conjugated to the αHL heptameric pore and this complex serves as a platform for the entire assay (Chapter 5). Thus, this chapter will discuss several experimental strategies for constructing DNA-pore conjugates.

The attachment of DNA to the pore was achieved via a sulfosuccinimidyl-4-(N-maleimidomethyl) cyclohexane-1-carboxylate (sSMCC) hetero bifunctional crosslinker that contains an amine-reactive NHS ester and a thiol-reactive maleimide group at opposite ends. The native form of αHL does not possess cysteine residues. In order to conjugate DNA to the pore, we thus engineered an αHL construct where a single amino acid (AA) has been substituted by cysteine. To find a suitable position for DNA attachment on the pore, three attachment sites at various distances from the *cis* opening of the pore were explored using three mutant DNA constructs having a cysteine residue (Cys) at a different AA site.

As mentioned earlier, αHL is expressed as a water-soluble monomer (33.2 kD) that naturally assembles into a heptameric pore-forming channel (232.4 kD) upon inserting into the membrane. Thus, to build DNA-pore conjugates, αHL monomers were first expressed in *E. coli* and purified. To attach a single DNA molecule to the pore, the formation of pore was induced by incubating a mixture of WT and Cys αHL monomers in the presence of lipids. Both monomers were recombinant proteins with a specific tagging system for purification. Specifically, the WT

construct contains a neutral 2x-Streptag® and the Cys construct has an 8x histidine tag (8x-Histag) that becomes protonated below its pI. Taking advantage of the charge difference between the two tagging systems, a 6:1 stoichiometry of the pore having 6 units of WT monomers and one unit of Cys mutant was separated from the rest of the pores by ion exchange chromatography (IEC). The step-wise process for the purification of 6(WT):1(Cys) stoichiometry pore and the possible permutations of hetero-heptameric pore are shown in Figure 3.1.

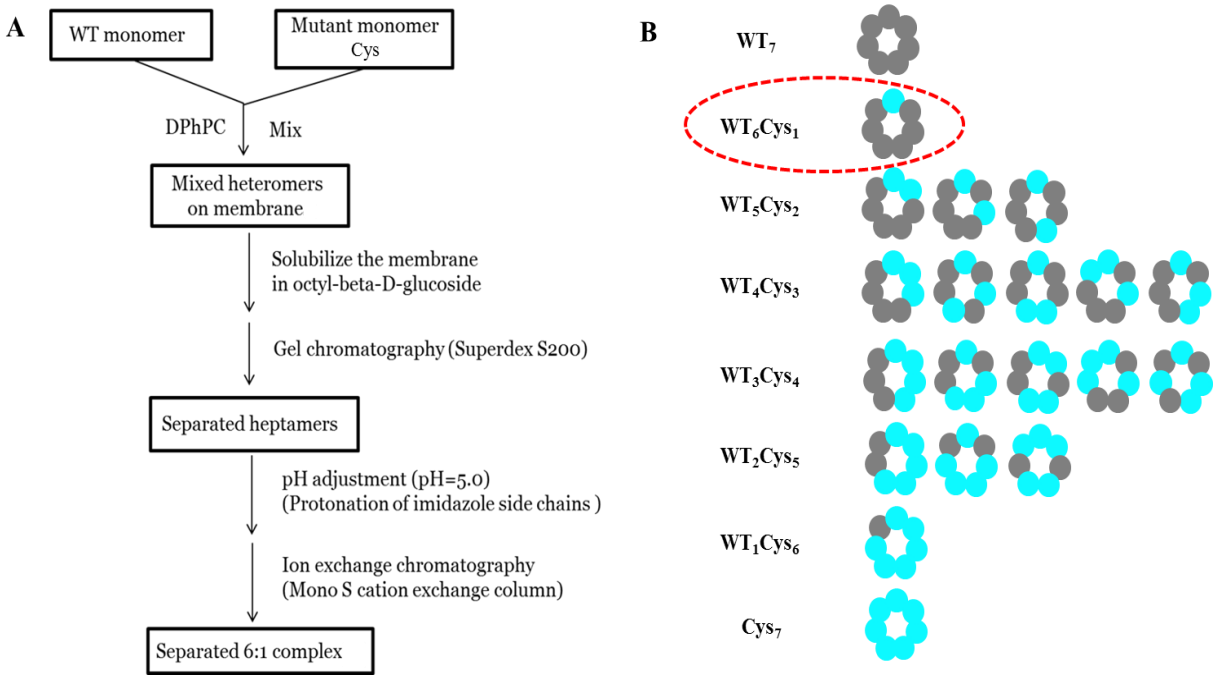


Figure 3.1 Building a 6:1 pore assembly. **A.** A stepwise process of isolating a 6:1 pore assembly. **B.** Possible permutations of eight stoichiometric ratios of WT and Cys with the desired arrangement highlighted.

The sSMCC linker attached DNA molecule (1) was reacted with the 6:1 pore and the resulting single DNA molecule attached DNA-pore conjugate (2) was finally purified using size exclusion chromatography (SEC). The overall process of conjugation is illustrated in Figure 3.2.

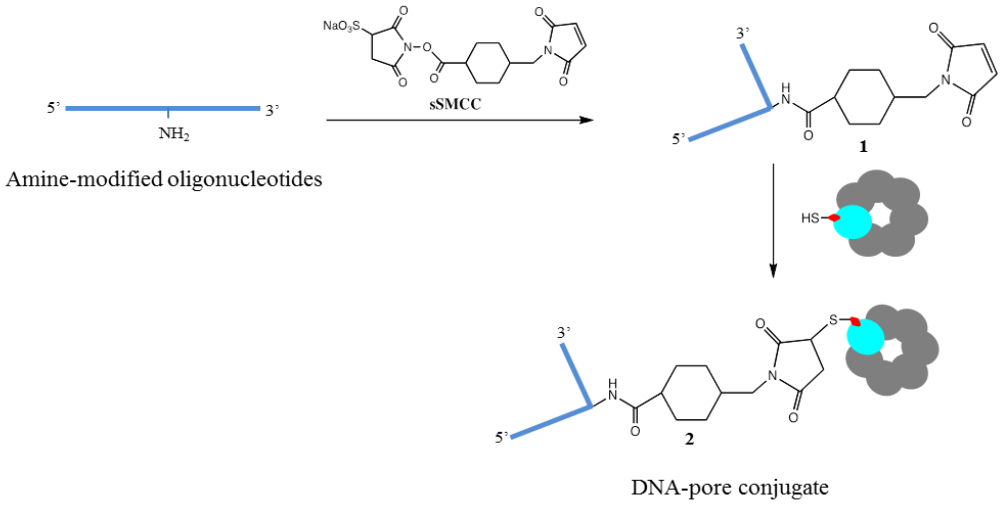


Figure 3.2 Synthesis of DNA-conjugated αHL.

3.2 Methods

3.2.1 PCR based one-step mutagenesis

A single pair of partially overlapping primers was designed to anneal to the template sequences flanking the target regions (underlined) where the sequences are to be substituted. The primer sequences that were used for each mutagenesis experiment are as follows. **8xHistag**: **F**- 5'-GAAGAAATGACAAATAAGGGTCATCATCACCATCACCATCATCATTGAGTTTAAAC TATATAGAAT-3', **R**- 5'-TTAAACTCAATGATGATGGTGATGGTGATGATGACCCTTAT TTGTCATTTCTTCTTTTTCCCA-3', **S239C**: **F**- 5'-GTGATTACCATGGATCGCAAAGCGT GCAAACAGCAGACCAACATTGATGTGATTTATGAACGC-3', **R**- 5'-CACATCAATGTT GGTCTGCTGTTTGCACGCTTTGCGATCCATGGTAATCACGGTCGCAAAATCCGG-3', **N93C**: **F**- 5'-GTACAGTTGCAACTACCTGATTGCGAAGTAGCTCAAATATCTGATTACT ATCCAAGAAAT-3', **R**- 5'-GTAATCAGATATTTGAGCTACTTCGCAATCAGGTAGTTG CAACTGTACCTTAAAGGCTGAAGG-3'. Using these primers and the original plasmid as template, a PCR reaction was performed with high-fidelity Phusion® DNA polymerase (NEB).

A 20 µl PCR reaction contained 1 µl of template (~100 ng), 1 µl of each primer (~10 pM each), 1 µl of 10 mM dNTP mixture, 1X buffer and 1 µl Phusion® DNA polymerase (2 U). The reaction started with a pre-denaturation step for 2 mins at 94°C followed by 34 cycles of 1 min at 94 °C and 1 min at 68 °C with a final extension step for 1 min at 72 °C. After the PCR reaction, the mixture was treated with 1 µl of *Dpn*I (20 U) (NEB) at 37 °C for 1 hour to remove the original template plasmids that contain methylated sites.

3.2.2 Transformation

Following the *Dpn*I digestion, 1 µl of the mixture was used to transform DH10B™ electrocompetent cells (Thermo Fisher). After an hour recovery in 600 µl SOC medium, the transformed cells were spread on LB plates containing 100 µg/ml carbenicillin and incubated at 37 °C overnight. Eight colonies were selected and grown in 100 µl of LB with 100 µg/ml carbenicillin overnight. To confirm the presence of the desired plasmid, a DNA fragment containing the αHL gene was amplified in a PCR reaction containing 1 µl of the culture, 1 µM forward and reverse primers, 0.2 mM dNTP mix, 0.1 unit of 2X Advantage DNA polymerase mix and 2 µl 10X Advantage buffer (Clontech). The PCR products were analyzed on a 1% agarose gel. After confirming the size of products, plasmid DNA was purified using a QIAprep Spin Miniprep Kit (Qiagen) and sequenced.

3.2.3 αHL protein expression

WT and Cys mutant proteins were expressed in One Shot® BL21 (DE3) Star cells (Thermo Fisher). This *E. coil* strain is deficient in Lon protease (cytoplasm) and OmpT protease (outer membrane). Therefore, it is suitable for expressing recombinant proteins. The αHL gene was originally cloned using the pEXP5 vector system that contains the T7-lac promoter for inducible expression. To induce the protein expression, MagicMedia™ medium (Thermo Fisher) with 100

μg/ml carbenicillin was inoculated with the seed culture at 30 °C for 18 hours. Cells were harvested and lysed by sonication in buffer: 100 mM Tris-HCl, 1M NaCl, pH 7.4 for WT and 1M PBS, 40 mM imidazole, 1M NaCl, 50 mM TCEP, pH 7.4 for the Cys mutants. Cell debris was removed by ultracentrifugation at 4 °C for 30 min at 20,000 rpm and the supernatants were collected.

3.2.4 αHL monomer purification

The proteins were purified using a Streptactin® affinity column for WT and a nickel-NTA column for Cys mutants on an ÄKTA chromatography system (GE). After loading with cleared lysates, columns were first washed with the equilibration buffers to eliminate nonspecific binding to the column and then proteins were eluted in elution buffer: 100 mM Tris-HCl, 1 M NaCl, 2.5 mM desthiobiotin, pH 7.5 for WT and 1M PBS, 500 mM imidazole, 1 M NaCl, 50 mM TCEP, pH 7.4 for Cys mutants. The purified fractions were analyzed by SDS-PAGE gel analysis.

3.2.5 Formation and purification of αHL heptameric pores

The two purified αHL monomers were mixed at a ratio of 6(WT):1(Cys) and incubated at 40°C for 1 hour in the presence of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC) at a concentration of 5mg/ml. Lipid vesicles were subsequently solubilized by adding 300 mM *n*-octyl-β-D-glucoside (βOG) to the reaction to a final concentration of 5% (vol/vol). Fully formed heptameric pores were separated from the remaining monomers and lipid vesicles by Superdex 200 SEC in 1 M PBS, 1 M NaCl, pH 7.5 and 30 mM βOG. The collected protein fractions were characterized by SDS-PAGE analysis.

3.2.6 Separation of 6:1 stoichiometry pore assembly

The pH of the purified heptameric pore solution was adjusted to pH 5.0 to protonate the 8x-Histag on the Cys mutant and then loaded on a Mono S column (GE) in 20 mM MES buffer, pH 5.0, 0.1 % Tween 20 and eluted with a linear gradient of 0 M to 2 M NaCl. The chromatographic fractions were collected and the presence of two monomers and their relative ratio were confirmed by SDS-PAGE.

3.2.7 Conjugation of DNA to pore

1 mM oligonucleotide with a single amine modification was incubated with 2 mg of sSMCC crosslinker (Thermo Scientific) at room temperature for 1 hour and the linker attached DNA was first purified from the unreacted sSMCC by Sephadex G25 SEC in 1 X PBS, pH 7.4. The purified 6(WT):1(Cys) pore was added at a 60X molar deficit relative to the concentration of the sSMCC-activated DNA and incubated for 4 hours. The final DNA-pore conjugate was isolated from the residual DNA by Superdex 200 SEC. The attachment of DNA to the Cys αHL mutant and the completion of the reaction were observed by a mobility shift of the Cys monomer in the denatured samples on an SDS-polyacrylamide gel.

3.2.8 Hemolytic capability of the purified proteins on blood agar plate

3.2 µg of each protein was spotted on a 5% sheep blood agar plate (Thermo Fisher) and incubated at 37 °C overnight.

3.2.9 Leakage kinetic assay based on vesicles

DPhPC vesicles were generated by the extrusion method using membrane with pore size 100 nm according to literature procedures.[89] Kinetics of αHL-induced calcein dye leakage from lipid vesicles was measured by fluorescence spectroscopy using a Fluorolog 3P fluorometer (HORIBA Jobin Yvon). Vesicle solutions (700 µL) were excited at 488 nm and fluorescence

over time was recorded at 515 nm. After collecting baseline fluorescence for 1 min, 50 µL pore solution was injected into the vesicle solution.

3.3 Results and Discussion

3.3.1 Introducing an 8xHistag and a single cysteine residue in αHL construct

As described above, the mutant αHL monomer with a single cysteine residue was designed to contain an 8x-Histag. Thus, we first introduced this tag to the C-terminus of the αHL plasmid construct using a one-step PCR-based mutagenesis strategy. Specifically, this method employs a pair of partially overlapping primers that is designed to anneal to the template sequences flanking the target regions. Using these primers and the original plasmid as template, PCR was performed and the reaction was subsequently treated with *Dpn*I restriction enzyme which is specific for methylated and hemimethylated DNA. Since only the parental plasmid previously grown in *E. coli* strains was Dam methylated, this step removes the unmodified original template, leaving the newly amplified PCR products containing an 8x-Histag intact. The PCR products become viable circular plasmids via an *in vivo* recombination mechanism in the transformed *E. coli* cells. A schematic of the procedures is illustrated in Figure 3.3A.

Based on this method, a pair of primers containing new sequences of histidine residues at the C-terminus were designed (Figure 3.3B). After PCR, the original plasmid without the modification was digested by *Dpn*I. The remaining *Dpn*I resistant PCR products were confirmed by agarose gel electrophoresis to contain the expected plasmid size (Figure 3.4A). After confirming their size, they were used to transform DH10B electrocompetent cells. After plating on LB agar, 8 colonies were analyzed and a fragment of the plasmid (1.2kb) was amplified from each of them to confirm they possessed the desired αHL modification (Figure 3.4B). After

verifying the sequences of the 8x-Histag at the C-terminus, it was employed for expression and purification of αHL monomers.

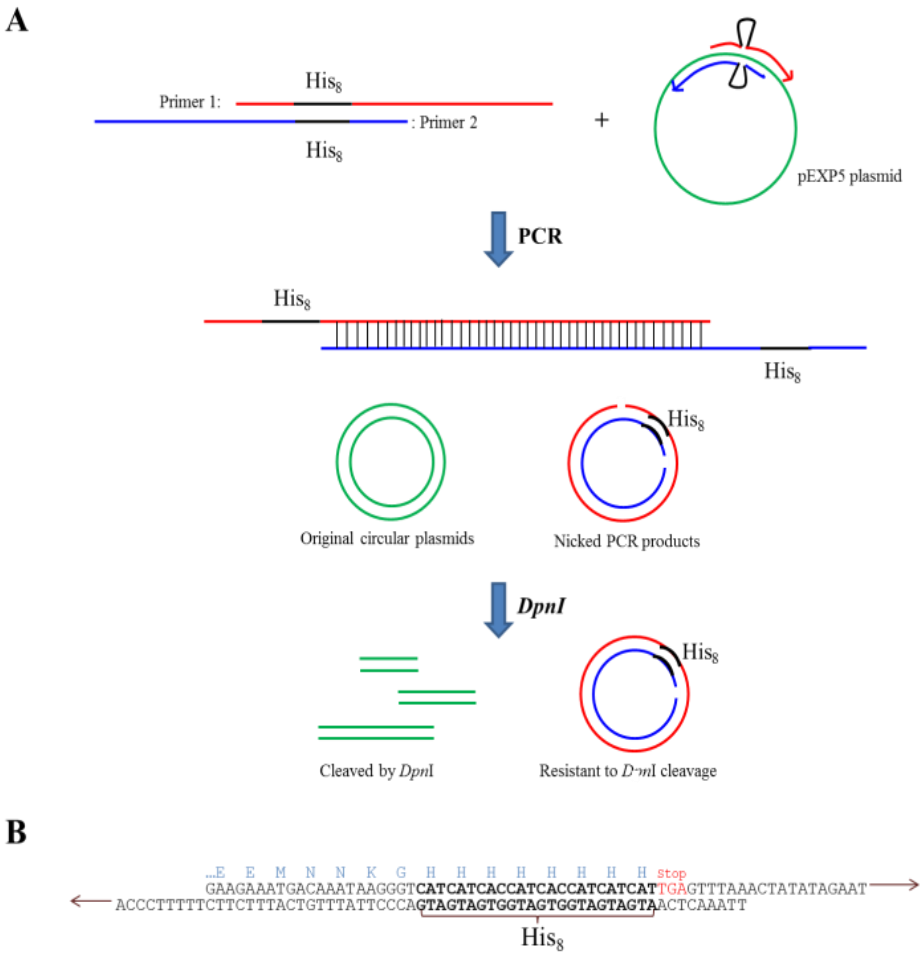


Figure 3.3 A. Scheme of one-step PCR-based site-directed mutagenesis. **B.** Sequences of forward and reverse primers.

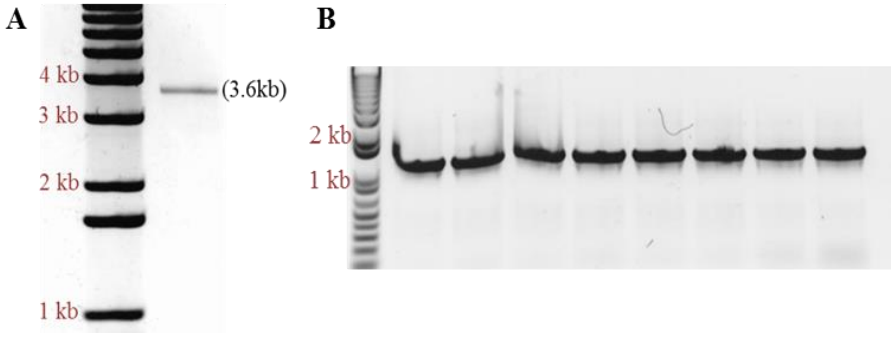


Figure 3.4 Analysis of PCR products after mutagenesis. **A.** PCR products after *Dpn*I digestion separated on a 1% agarose gel. **B.** Analysis of αHL amplicons from 8 selected colonies in a 1% agarose gel.

To attach DNA at different distances from the entrance of the pore, we made several versions of αHL mutants in which a single cysteine residue was introduced in different positions of the pore. In total, three different AA positions on the cap of the αHL pore were selected: C46, C239 and C93 (Figure 3.5). Position 46 is in the middle of the cap and position 239 is at the edge of the top surface and thus is further from the pore entrance. Position 93 is on the side of the pore instead of the top surface and therefore is the furthest from the pore mouth. The C46 mutant was previously available in our laboratory and the other two αHL mutants were created using the previously made αHL plasmid with 8xHistag via the same site-directed mutagenesis strategy.

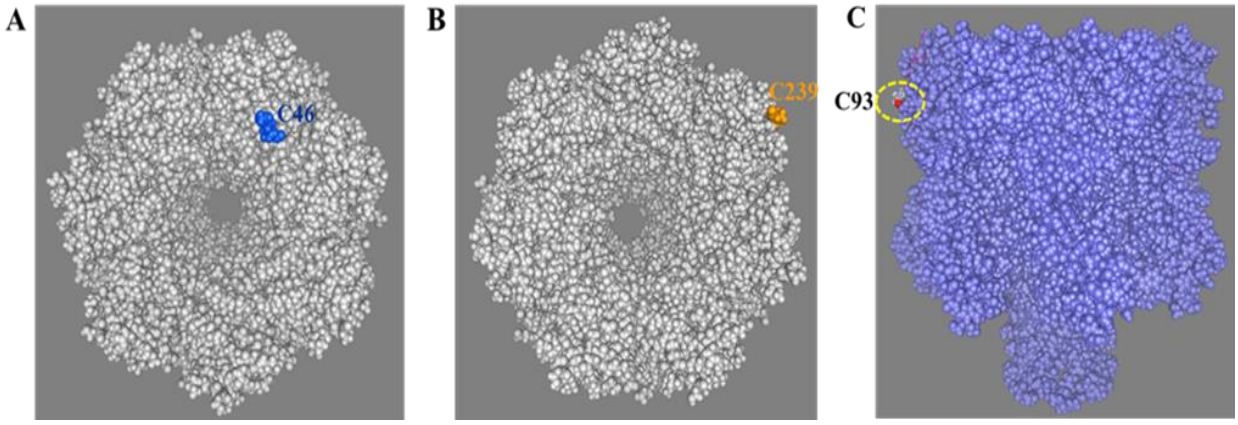


Figure 3.5 Positions of cysteine residues in three different αHL mutants.
A. C46, top view of αHL pore **B.** C239, top view **C.** C93, side view.

3.3.2 Purifications of monomers, 6(WT):1(Cys) heptameric pore and DNA-pore conjugate

The two αHL monomers, WT and Cys, were first expressed in *E. coli* strain BL21 (DE3) and purified using affinity chromatography. As described above, the WT αHL construct contains a 2x-Streptag and therefore has a strong affinity for Strep-Tactin® which is an engineered form of streptavidin. The cleared cell lysate containing WT αHL monomer was applied to a Strep-

Tactin column (GE). After a brief washing step which removes non-specifically bound proteins from the column, elution was performed with desthiobiotin, which competes for the binding site of the Strep-Tactin®. Figure 3.6A shows a chromatogram of the WT monomer purification and the elution peak collected. The Cys αHL mutants contain an 8xHistag and thus they were purified using an immobilized metal affinity chromatography (IMAC) column charged with Ni^{2+} which interacts with the imidazole rings of the histidine tag introduced in the Cys mutants. The bound proteins were then eluted with imidazole which competes with the histidine tag for binding to the nickel ion. A chromatogram of the C239 αHL monomer purification is shown as an example (Figure 3.6B). The other two Cys mutants (C46 and C93) were purified in the same way. Figure 3.6C shows the result of SDS-PAGE analysis of each peak in the two chromatograms above. The WT αHL having the 2xStreptag (~35kD) is slightly larger than the one with an 8xHistag (~34kD) and the gel shows the difference in size between the two proteins. The last two lanes on the gel represent denatured forms of the molecules. To verify that the purified proteins possess hemolytic activity, all four recombinant αHL monomers were tested on a blood agar plate. They all produced hemolytic plagues in blood agar, confirming that they are indeed functional hemolysin proteins (Figure 3.6D).

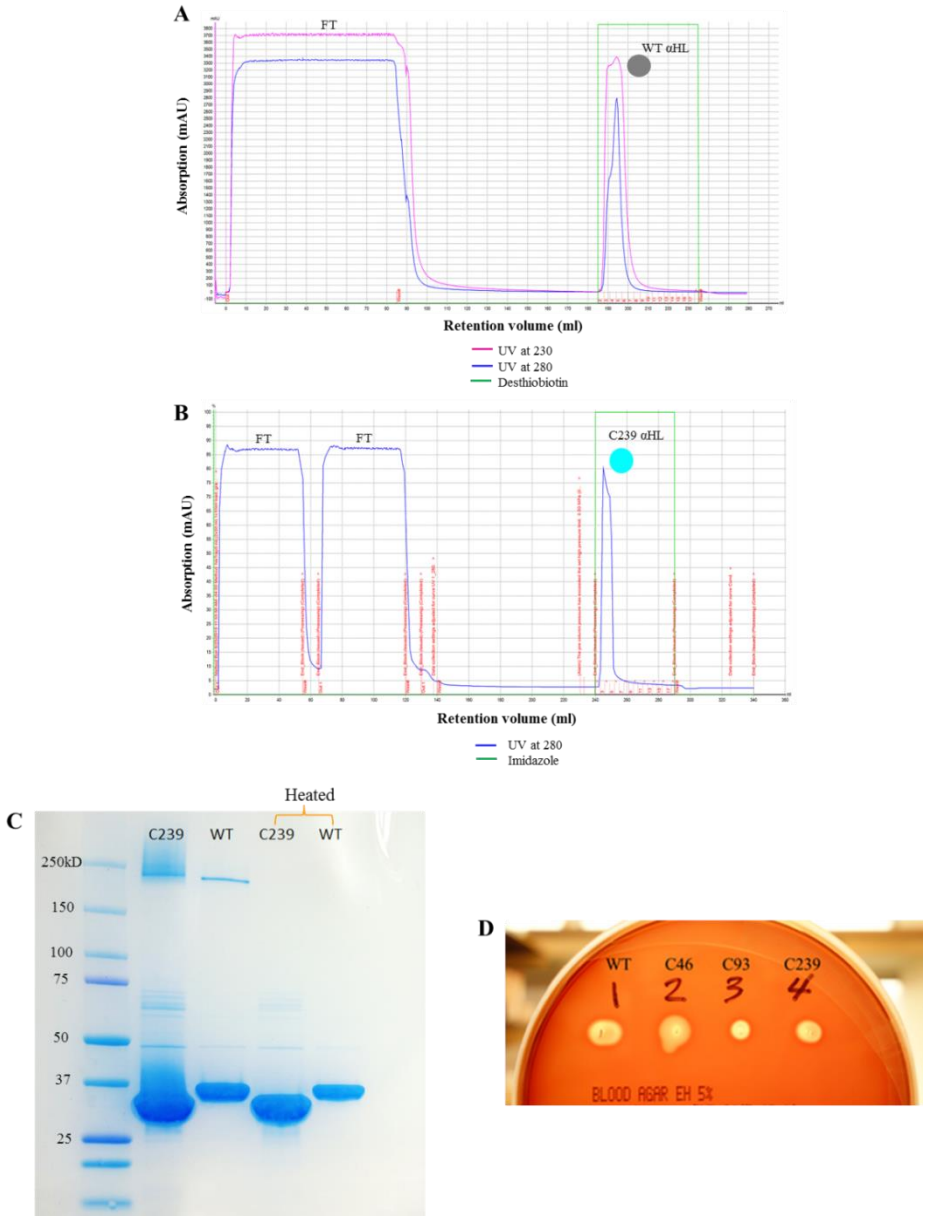


Figure 3.6 Purification of αHL monomers. **A.** Chromatogram of the purification result for WT monomer on a Streptactin column. **B.** Chromatogram of purification of C239 monomer on a nickel column. FT = flow through. **C.** SDS-PAGE analysis of purified monomers. **D.** Confirmation of hemolytic activities of the purified proteins on a 5% sheep blood agar plate.

Additionally, to compare the kinetic properties of the purified monomers, a vesicle leakage experiment was performed with the WT and one of the Cys αHL mutants (C93). In this experiment, DPhPC vesicles containing calcein dye (Figure 3.7A) were prepared. The

encapsulated calcein dye exhibits almost no fluorescence at high concentration because of self-quenching. However, when an αHL monomer is added to the calcein vesicles, it forms pores in the vesicles causing the dye to escape into the solution and thus the fluorescence from calcein increases upon dilution (Figure 3.7B). Therefore, αHL pore-forming activity can be measured by the increase in fluorescence intensity after addition of the αHL monomer. Figure 3.7C shows changes in fluorescence intensity over time. Two concentrations of αHL monomers were tested (25 nM and 50 nM) and we observed that the initial slope for the dye leakage induced by the C93 mutant was greater than that of the WT αHL, indicating that the C93 αHL mutant has relatively faster kinetics of pore formation than the WT. The upper black trace represents a control experiment with Triton X-100, which solubilizes the vesicle membrane and thus causes all calcein dye to be released into the solution. This trace corresponds to the fluorescence intensity of complete leakage and the same fluorescence intensity was absorbed after addition of Triton X-100 to the αHL containing solutions at the end of experiments (not shown). Overall, the data show that the mutant αHL has a slightly faster kinetics of pore formation than WT αHL and both should be suitable for preparation of heterologous pores.

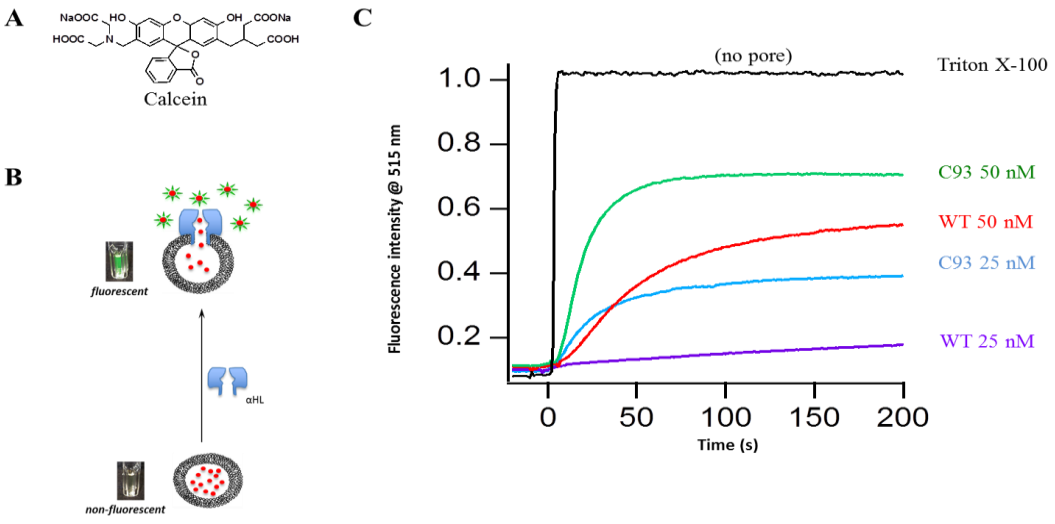


Figure 3.7 A kinetic experiment using calcein dye leakage from vesicles. **A.** Structure of calcein dye. **B.** Scheme of fluorescence emission after the pore insertion. **C.** A graph of fluorescent intensity over time comparing WT and the C93 mutant.

As described earlier, αHL monomers spontaneously form a heptameric transmembrane channel upon contact with a lipid membrane. To induce the formation of heterologous heptameric pores, a mixture of the purified WT and Cys αHL monomers was incubated in the presence of DPhPC bilayers at 40°C for 1h. After solubilizing the vesicles with octyl-beta-glucopyranoside (βOG), αHL oligomers were separated from the residual monomers and the lipid by SEC. Figure 3.8 shows the chromatogram of the reaction mixture and SDS-PAGE analysis of the three peaks. In order to confirm that the high-molecular weight fraction represents the αHL oligomers, the collected fractions were heat denatured (95 °C, 5 min) and compared with the equivalent native forms. The first six lanes display intact samples and the last six lanes show the denatured ones (Figure 3.8B). As can be seen in the gel, the protein that came out as the first peak (P1) possess the mobility corresponding to the expected heptameric pore size (~240kD) and dissociates to monomers upon heat denaturation (~34kD). Based on the chromatogram and results of this electrophoretic analysis, we thus concluded that P1 contained the heptameric oligomers of hemolysin.

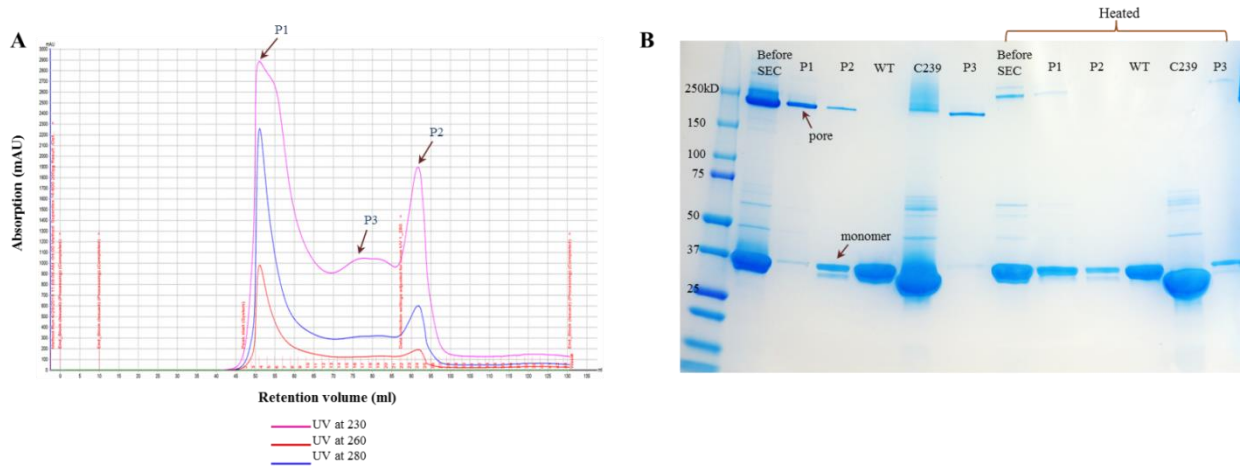


Figure 3.8 Purification of heptameric pores. **A.** A chromatogram of the SEC purification result. **B.** Fraction analysis by SDS-PAGE after Coomassie blue staining.

At this step, a mixture of hetero-heptameric pores of various stoichiometries is present in the sample. To isolate 6(WT):1(Cys) pore assemblies, we used mono S (GE) IEC. Before applying the sample to the column, the pH of the solution was adjusted to pH 5.0 to protonate the histidine tag residues in the Cys αHL monomer of the assembled pores. Thus, the positively charged Cys αHL pores bind to the negatively charged groups on the cation exchange column. After a brief washing step to remove the unbound materials from the column, elution was begun using a salt gradient to separate molecules based on their overall charge. Figure 3.9A shows the resulting chromatogram. Five different peaks were collected and analyzed by SDS-PAGE (Figure 3.9B). The first six lanes on the left show the native forms of the proteins and the last six lanes represent their denatured counterparts. While the protein migration pattern of all five collected peaks indicate that they contain approximately the same size heptameric pores, the denatured forms demonstrate that there are two different monomers in these fractions except for those that came out in the first peak (P1), which is comprised of only WT αHL monomers and thus represents a 7:0 pore. The other peaks show progressively different ratios of WT and Cys αHL monomers. Based on the combined gel and chromatography results, we concluded that the second peak (P2) contains heptameric pores with 6:1 stoichiometry.

As described earlier, DNA oligonucleotide was conjugated to this purified 6:1 pore using the sSMCC crosslinker and the final product was separated from the residual DNA by SEC. Based on the chromatogram (Figure 3.10A), the first peak was analyzed in the gel along with controls of different monomers and pores. The presence of DNA conjugated Cys monomer was confirmed by a mobility shift of the Cys monomer upward due to its increase in molecular weight after conjugation of the DNA.

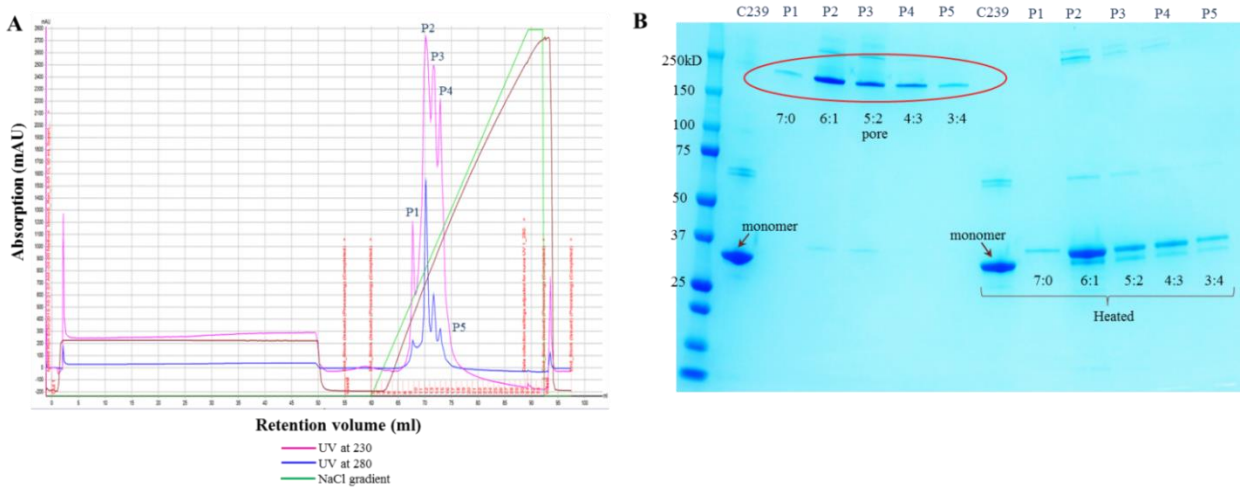


Figure 3.9 Purification of pores of various stoichiometries by ion exchange chromatography. **A.** A chromatogram of the purification result on a Mono S column **B.** SDS-PAGE analysis of five collected fractions along with the C239 monomer control.

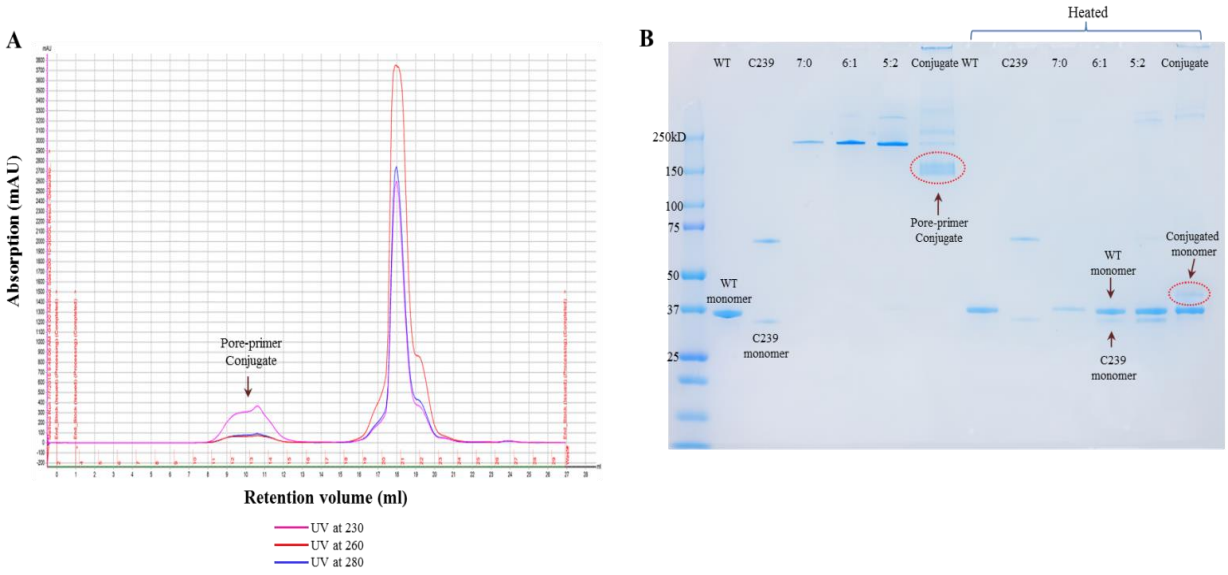


Figure 3.10 Purification of 'DNA-pore' conjugates. **A.** A chromatogram of the SEC purification result. **B.** SDS-PAGE gel analysis of each peak fraction along with monomer and previous pores with different stoichiometries as controls.

3.4 Conclusion

The overall process of constructing DNA-pore conjugates is described in this chapter. First, several Cys-containing aHL mutants were created using one-step site directed mutagenesis and the corresponding proteins were purified by affinity chromatography according to their respective tagging system; WT: Streptag and Cys: Histag. By mixing WT and Cys αHL monomers, αHL oligomers with various stoichiometries were generated and purified by SEC. Because of the presence of the histidine tag in the Cys-containing αHL, pores of different monomer stoichiometry could be separated based on their charge. After isolating the 6:1 pore, DNA was conjugated to it via an sSMCC crosslinker, which was confirmed by a change in mobility of the Cys monomer after conjugation. In short, these data validate our methodology for building nanopores with single DNA molecules.

CHAPTER 4: Single Molecule Measurements of Different Conformation DNAs Conjugated to Nanopore

4.1 Introduction

In our first approach, an integrated single molecule electronic SNP assay, ssDNA is covalently tethered to the pore to serve as a platform for the entire assay (Chapter 5). Therefore, to gain insights about single molecule behavior of DNA coupled to the pore, several DNA-pore conjugates having DNAs of different lengths and structures were prepared and their current signatures were analyzed in the nanopore system.

As described earlier, αHL is a pore-forming membrane channel that is well suited for single molecule DNA experiments due to its geometry and stability.[90-96] The mushroom shaped heptameric transmembrane pore contains a large 4.6 nm cavity that is located entirely within the extramembranous domain. In the transmembrane domain, the pore lumen constricts to form a 14-stranded β-barrel that is 2 nm wide and 5.2 nm long with a 1.4 nm limiting constriction between the vestibule and the stem.[52] Because its diameter is large enough to accommodate ssDNA but not dsDNA, each polynucleotide must uncoil when it is forced into the narrow lumen of the pore and traverse through it as an extended single stranded chain, scanning the full contour length of the molecule.[101-103] During translocation, the ssDNA partially occludes ion flow through the pore increasing its electrical resistance, and this allows for straightforward sensing of the DNA molecule, which in turn provides information about its structure and dynamic motions.[100-104]

In 1996, Kasianowicz et al. first demonstrated electrophoretic transport of individual ssDNA and ssRNA molecules through αHL pores.[54] Since then, polynucleotide translocation

through this pore has drawn much interest, producing numerous theoretical and experimental studies about how extended polymers enter and pass through a constricted pore under an applied potential.[54,98,99,104,105] The dynamics of polymer transport through a narrow pore is central to many biological processes including RNA export across nuclear pores, phage infection and protein transport through membrane channels, and is also the fundamental principle behind various new methods for nucleic acid analysis using nanopores.[95,96,101,102,106-108]

Under physiological conditions, biopolymers such as ssDNA form into a random coil geometry. This geometry maximizes the conformational entropy of the DNA by allowing it to access a large range of different conformations.[101,104] As previously discussed, ssDNA can only thread the limiting constriction of the αHL pore as an extended chain. There is a significant entropic penalty associated with this process, which constrains and reduces the degrees of freedom of the ssDNA.[101,104] This effectively presents an energetic barrier which makes it very unlikely that ssDNA will be able to diffuse through the pore. This has been experimentally demonstrated in a study that measured the capture of DNA as a function of voltage. In 2003, Meller found that the capture rate in αHL pores is exponentially dependent on the applied potential, indicating the presence of an energetic barrier to threading the DNA into the pore.[102] The barrier is overcome by increasing voltage, the driving force, thus exponentially increasing the DNA capture rate. Several experiments have supported the notion that the location of the energy barrier for polymer transport probably corresponds to the narrowest geometric path inside the pore.[98,100] This means that the polymer must cross this threshold before it is committed to transport through the pore against thermal agitation and repulsive forces; otherwise, it is likely to escape on the same side of the membrane from which it entered.[46,103,104] Molecular Dynamics (MD) simulation has also shown that the steepest gradient of the electrostatic potential arises

around the smallest cross-sectional area inside the αHL pore.[99,105] Experimental studies have supported the idea that this position in the channel is the place that is most sensitive to blockade and thus where the recognition for different translocating molecules is at its strongest.[98,103]

Another study has shown that pore geometry also influences the capture of biopolymers. In 2000, Henrickson et al. studied translocation frequency of ssDNA into an αHL as a function of applied potential and compared the rate when the DNA was added to the *cis* or *trans* side of the membrane.[101] Consistent with the above study, for both *cis* and *trans* experiments, DNA capture into the pore had an exponential relationship with the applied potential, suggesting a voltage-reduced energy barrier for capture. Additionally, the rate of capture was considerably different for *cis* and *trans* experiments; *cis* entry was noticeably more favorable than *trans* entry at any given potential.[101] These results can be interpreted to mean that the wider opening of the αHL vestibule provides a relatively lower energy barrier for DNA entry since ssDNA can enter the *cis* side in many configurations while it can only insert itself to the *trans* entrance of the pore in a stretched conformation. In other words, DNA entry into the *cis* side is energetically favored. Together with the first study, these data suggest that size and geometry of the pore are critical factors in polymer dynamics.

In light of the above published results, the dynamics of threading DNA through the pore were characterized at varying applied potentials, DNA lengths and structure and their impacts on the ionic current reduction within the pore were compared. For these experiments, DNAs were attached to the same position on the pore using C46 heptamers and thus DNA in each conjugate spanned the same distance to the pore mouth. Since DNA is covalently conjugated to the pore, in our experiments, the molecule was captured and immobilized inside the pore by electrical force but they were not completely translocated into the *trans* compartment. As mentioned above,

previous experimental and MD simulation studies have indicated that the major component of the electric potential drop occurs in the transmembrane beta barrel, rather than over the entire length of the lumen of the αHL.[98,100,103] Considering this, for the analysis of current signatures, the pore in this chapter is defined as the stem region that penetrates the membrane, and excludes the extramembranous larger vestibule and pore mouth. The final conjugates were purified using the same methods as described in Chapter 3.

4.2 Methods

These experiments were performed in the Genia nanopore array and a detailed description of the instrument and the common automated procedures such as membrane formation and pore insertion steps are provided in Chapter 1.

4.2.1 DNA capture experiment with DNA-pore conjugates

DNA capture experiments were performed in a buffer containing 300 mM KCl, 20 mM Hepes, pH 7.5, unless otherwise stated. The conjugate was diluted in the buffer to a final concentration of 8 nM. After applying a 10 µl aliquot to the *cis* compartment, the automated pore insertion process was started. Several trapezoidal voltage pulses were generated to measure the current signature of the inserted pore.

4.2.2 Binding of complementary DNA to the oligonucleotide on the 'DNA-pore' conjugate

After recording the current signature of the DNA-pore conjugate, 3 µM "dumbbell" shaped complementary DNA was applied to the system. The sequence of the complementary DNA is as follow: 5'-CGCGGCGCGTAAGCGCCGCGACAACAGTACTAAACCCCCATAAATAGGAGCGCCGGCCGTAAGGCCGGCG-3'.

4.2.3 Exonuclease treatment of the conjugate

This experiment was performed in a buffer containing 150 mM KCl, 20 mM Hepes, and pH 7.5. 10 units of 3' → 5' Phi29 exonuclease was diluted in the buffer and added to the system after the pore insertion step.

4.2.4 Data analysis

Data were sampled at 1-kHz bandwidth in an asynchronous configuration at each cell using circuit-based analog-to-digital conversion and noise filtering (Genia Technologies, Inc.) and the I *vs* V plot was built using MATLAB.

4.3 Results and Discussion

In these experiments, current passing through the nanopore was measured at varying voltages. Specifically, the voltage was linearly ramped for 2 seconds, held at a fixed voltage for 6 seconds, and then ramped back to the ground level for 2 seconds, forming a trapezoidal waveform with rising and falling periods. Such slow ramping of voltage allows us to investigate instances of capture and release of the molecule into and out of the pore. In this type of waveform, the molecule escapes from the pore at the end of each cycle when the polarity of the potential difference is reversed. This allows the conjugated DNA to be analyzed repeatedly from the same pore by generating repeated waveforms.

4.3.1 Current signature in the absence of DNA attachment

Before testing the DNA-pore conjugates, the conductance of the C46 heptameric pore without bound DNA was measured as a control. First, 300 mM KCl solution was added to the system. After checking the current measurement, DPhPC lipid was applied to form the bilayers followed by addition of the pore. A trapezoidal waveform that ramps up from 0 to 160 mV in 2 seconds and returns to the ground level after 6 seconds of plateau was generated and repeated 10

times. Figure 4.1 shows a plot of the applied voltage (red) and induced current (blue) over time. As seen in this figure, the current is directly proportional to the applied voltage with minor fluctuations. This demonstrates that the purified pore conducts appropriate current and also establishes the baseline current signature of the pore without any DNA captures.

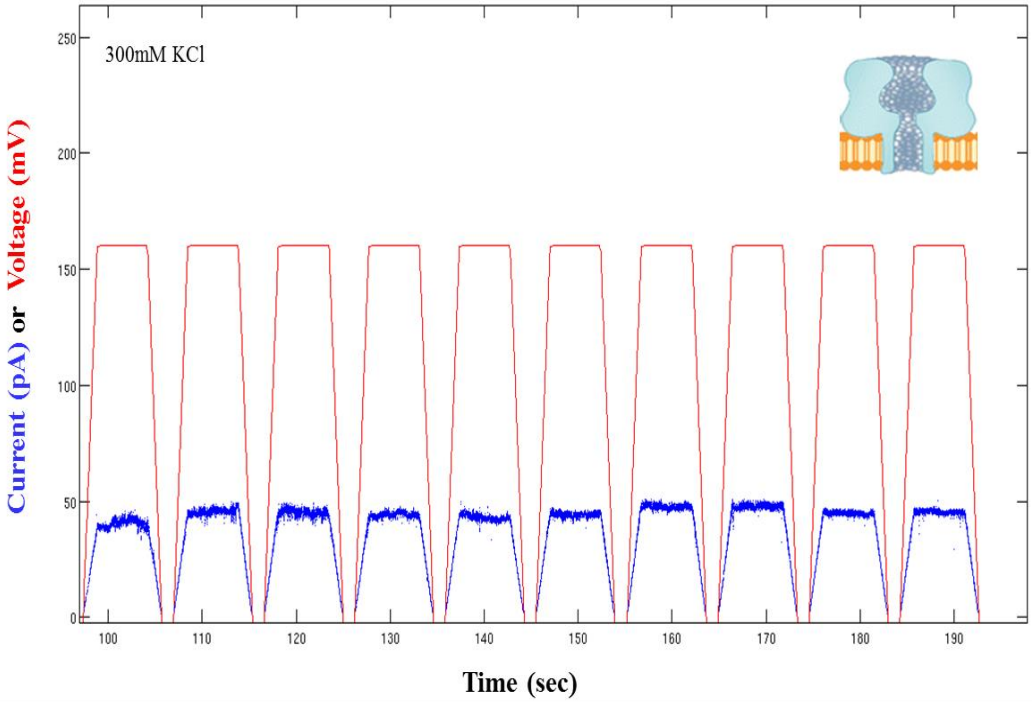


Figure 4.1 A graph of voltage and current traces against time from the C46 pore without bound DNA.

4.3.2 Current signature of ssDNA tethered to the pore

To characterize the intramolecular capture of DNA by αHL pores, several DNA-pore conjugates were made with DNAs of different lengths and structures, and tested in the nanopore system. As mentioned earlier, we employed the sSMCC crosslinker to attach an amine-modified DNA to Cys 46 in αHL. Therefore, the position of an amine modification (NH_2) in the DNA sequence along with the position of the cysteine residue on the pore determines the final configuration of the DNA-pore conjugate. First, a "T" shaped conjugate was made by placing an

amino dT in the middle of a 25-mer oligonucleotide sequence, 17 bases from the 5' end and 7 bases from the 3' end. Figure 4.2A shows the NH_2 position in the sequence of the oligonucleotide along with a graph of the voltage and current traces obtained with this conjugate. Figure 4.2B is an expanded image of a single trapezoidal trace. Unlike the DNA-free pore, there is a sudden decrease in current with increasing voltage. We interpret this to mean that the end of the DNA is threaded into the pore. Such a DNA capture event causes a decrease in current because the DNA present inside the pore blocks ion flow through the pore. During the 160 mV plateau period, the DNA stays within the pore, maintaining the current until the voltage ramps back to the ground level. As the potential difference decreases, there is an abrupt increase in current as the DNA escapes from the pore restoring the current to its open pore value. Interestingly, this conjugate exhibits a symmetrical behavior, where capture and release occur at about the same voltage. Compared to the previous current trace of the DNA-free pore, these signals by this DNA-pore conjugate represent the dynamic motions of a single molecule of DNA being captured into and released from the pore in response to changes in the electric field. Along with the results from the gel shift assay in Chapter 3, the differences in current traces demonstrate the presence of DNA conjugated to the pore.

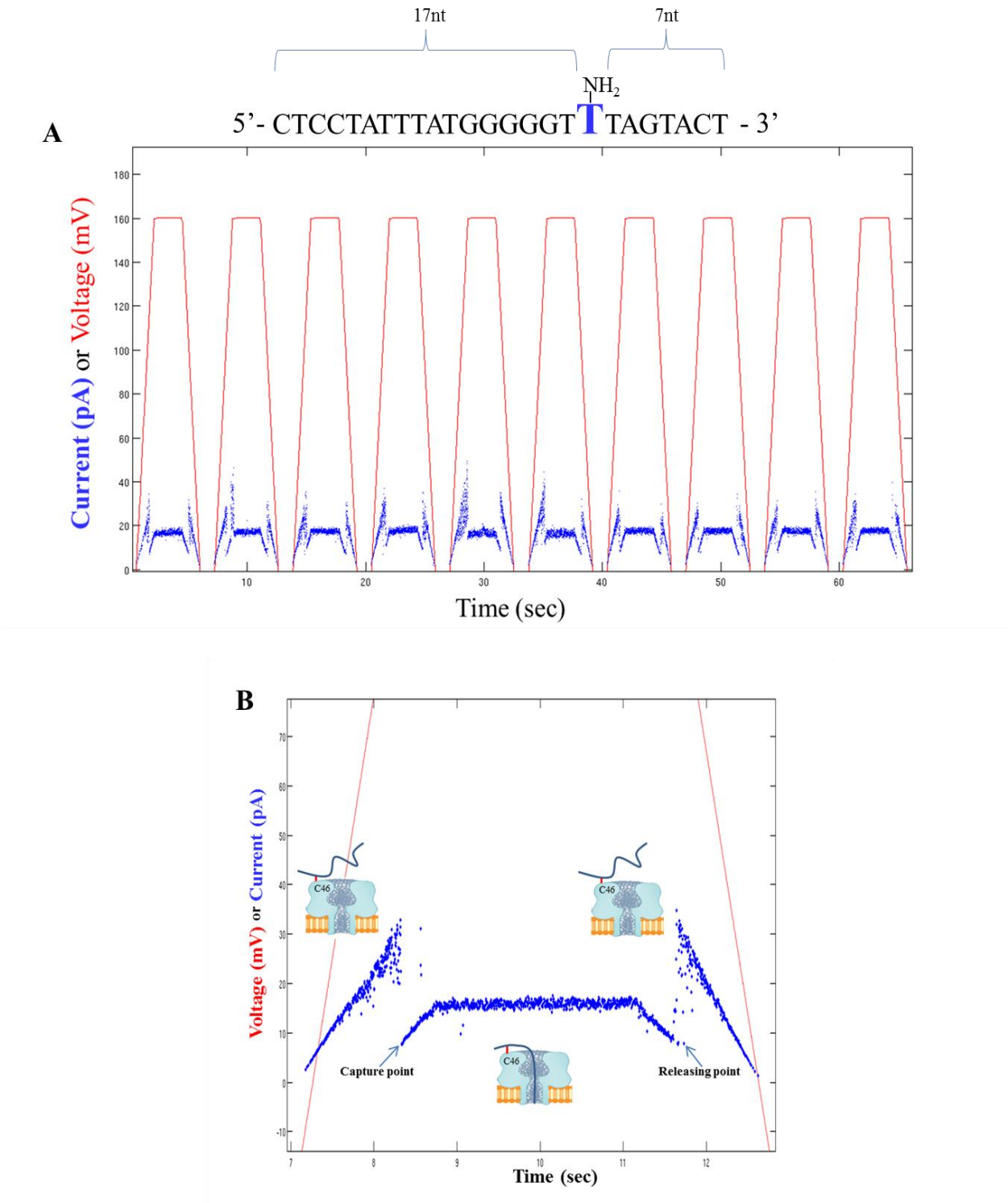


Figure 4.2 **A.** A graph of voltage and current traces against time with the DNA-pore conjugate. **B.** A zoomed-in picture of a single current trace.

4.3.3 Threshold voltages for threading the DNA into the pore

Using the recorded data points shown in Figure 4.2, we built the I *vs* V plot of this conjugate shown in Figure 4.3. Three conductance states are observed: one with a low resistance at low voltage (1), another with a high resistance at higher voltage (3) and an intermediate one with intermittently increased resistance (2). The first state corresponds to the conductivity of the open pore with no DNA threaded. The second state is an intermediate one when the DNA mostly stays outside the pore, but occasionally and briefly threads in. The third state represents when the DNA more continuously threads and stays inside the pore. According to this plot, the transition from the first state to the second state appears to occur at approximately 80 mV and transition from the second state to the third occurs at about 110 mV. We interpret this to mean that the DNA starts approaching the pore at about 80 mV and it shifts to a stable threaded position when the potential reaches approximately 110 mV. Between these two voltages, DNA appears to be in between the two states, briefly interrupting the pore current. The fact that current is not as low in state 2 may indicate that the DNA can only partially traverse the pore before it comes back out, essentially probing the vestibule or entrance to the channel, but is not fully threaded.

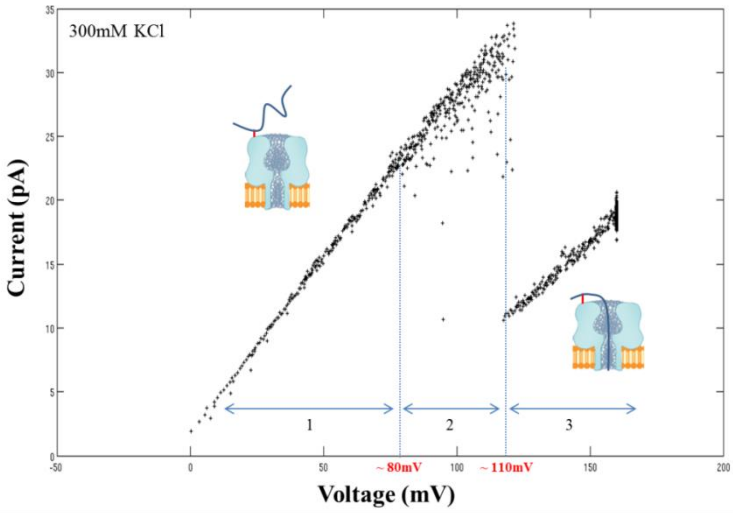


Figure 4.3 A plot of I *vs* V from the DNA-pore conjugate. This shows three conductance states: a higher conductance when the DNA molecule remains outside the pore (1), a lower one corresponding to a state when the DNA is stably trapped inside (3), and an intermediate state (2) when the DNA does not fully thread.

4.3.4 Voltage dependent current signatures

Interestingly, similar data were also obtained from current traces, which were generated by applying a range of voltages to the conjugate as shown in Figure 4.4. In this experiment, trapezoidal waveforms with peak voltages of 80 – 200 mV with an increment of 20 mV after every 5 cycles were applied. In the current trace at 80 mV, we observe brief interruptions in current, suggesting that the DNA is briefly captured in the pore, but it escapes. At 100 mV, the DNA also briefly threads, this time long enough to display fully threaded current levels during the plateau period. At 120 mV and above, the DNA threads more continuously, producing steady blockade events under the applied potential. This result is in agreement with the I *vs* V plot of Figure 4.3 where we see one state below 80 mV and another above 110 mV and an intermediate situation between these voltages. In terms of thermodynamics, this indicates that approximately 80 mV corresponds to the minimum potential required to overcome the energetic barrier for DNA to enter the pore and thus the DNA starts threading inside. However, it appears to require a potential greater than 110 mV for it to be stably confined within the pore.

4.3.5 Absence of current signature in exonuclease treated conjugate

As a control experiment, the conjugate was also exposed to the exonuclease activity of phi29 DNA polymerase which digests the DNA of the conjugate starting from the 3’ end. As shown in Figure 4.5, the signature blockade pattern disappeared in the current trace upon exonuclease treatment. Instead, the trace appears to resemble that of the heptameric pore without attached DNA (Figure 4.1). This result verifies that the capture events in the previous experiments are attributable to the presence of the oligonucleotide attached to the αHL pore.

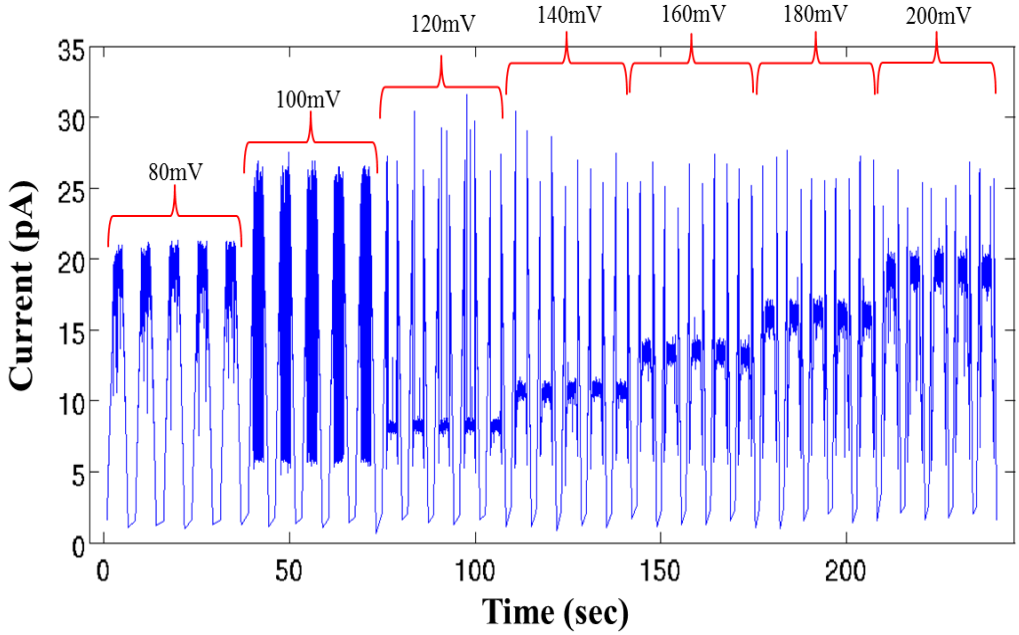


Figure 4.4 Current traces against time at various voltages showing dynamics of DNA capture events into the pore.

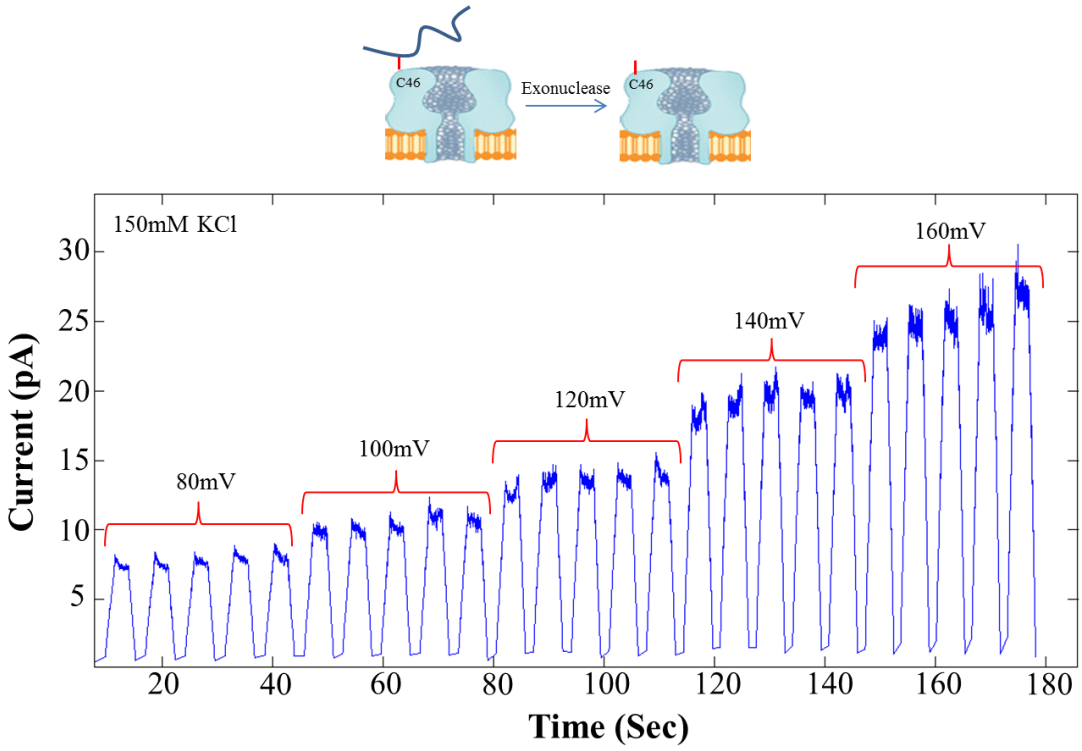


Figure 4.5 Current signatures of the exonuclease treated conjugate at various voltages.

4.3.6 Current changes resulting from duplex formation of the conjugated DNA

In a separate experiment, a "dumbbell"-shaped complementary DNA strand was added to the conjugate to test if hybridization would abolish DNA capture. Figure 4.6 displays the current traces before and after the addition of the complementary strand to the oligonucleotide in the conjugate. While the same capture behavior as in the previous experiments was observed before applying the complementary DNA strand, its characteristic signature disappeared after the addition of the strand. This change in the current trace upon the addition of the complementary DNA could be interpreted as indicative of its binding to the ssDNA on the pore, with the resulting duplex no longer being captured as the ssDNA end is no longer available for threading into the pore. This shows that the secondary structural changes of the conjugated DNA also can be detected by monitoring current signals.

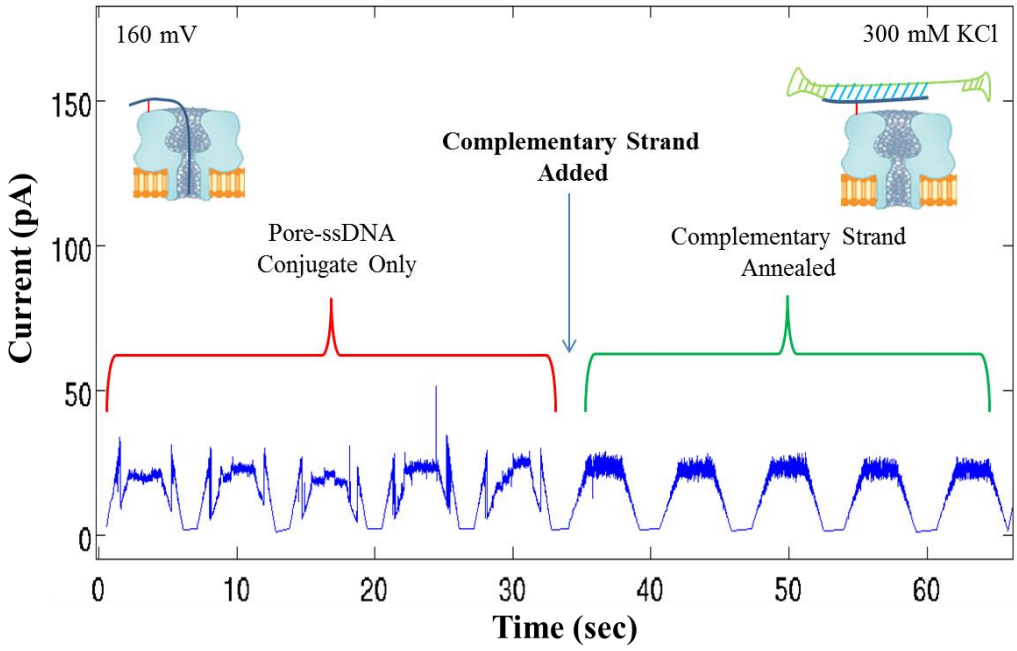


Figure 4.6 Current trace against time before and after the addition of a complementary DNA strand to the DNA in the conjugate. The characteristic current signature disappeared after binding the complementary DNA.

4.3.7 Current signature of conjugate having shorter oligonucleotide

As described above, the ssDNA molecule on the previous conjugate had an amine modification at the 18th position of a 25 base long ssDNA, resulting in branches of 17 nucleotides and 7 nucleotides (Figure 4.2). Given the highly stable and uniform blockade level generated by this conjugate, we speculated that the longer branch is preferentially captured and threads, generating a steady blockade signature. To verify that this was the case and also to further probe the length dependence of capture, another 25-mer oligonucleotide containing shorter branches was designed and conjugated to the pore. The new DNA consists of branches of 13 and 11 bases from the conjugation point as shown in Figure 4.7. Unlike the DNA on the previous conjugate, the DNA of this new conjugate could only thread briefly and partially even at a relatively high voltage such as 160 mV. In fact, the signature of this conjugate was rather noisy with transitions between the two conductance states and the percentage of time spent at the higher current was greater than at the lower one, indicating that the DNA was in continuous motion, captured briefly in the pore but mostly staying outside.

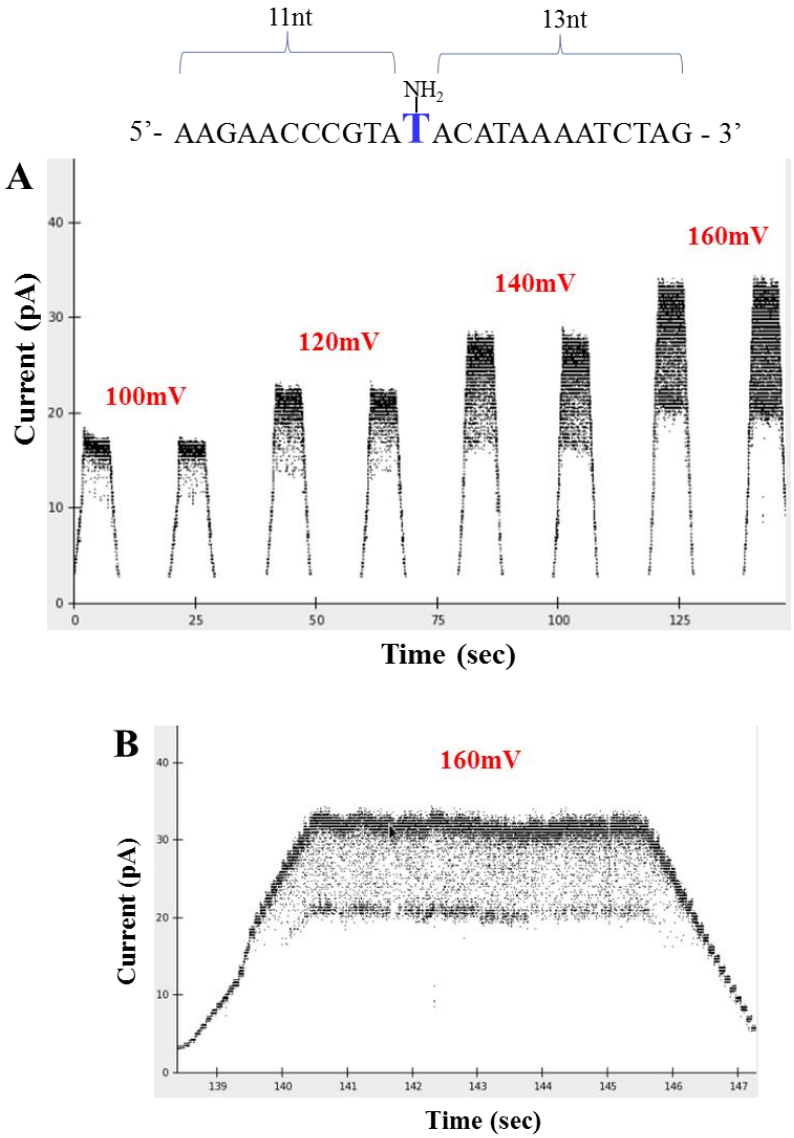


Figure 4.7 A. Current trace against time from the conjugate with the DNA sequence shown in the figure. **B.** A zoomed in picture of a single trapezoidal voltage pulse at 160 mV.

As discussed earlier, these conjugates were made with the pore having a cysteine residue at position 46 that is in the middle of the top surface of the pore (Figure 3.5) estimated to be 17 Å from the pore opening.[52] The distance from the *cis* entrance to the constriction site is 48 Å.[52] Therefore, the total distance from the point at which DNA is conjugated to the constriction zone is 65 Å. Given the 5.6 Å length per base of ssDNA,[110] the current signatures of the conjugates

are consistent with the lengths of the attached DNA and the signals produced by them. Specifically, 65 Å is translated to approximately 12 nucleotides and thus the 13 nucleotide long DNA branch is just about able to pass the constriction zone but it might not be long enough to be stably threaded into the pore, thus generating such a fluctuating current signature. On the other hand, the previous conjugate produced stable and long-lived events. We interpret this to mean that the length of 17 nucleotides (95.2 Å) is sufficient to almost extend through the full length of the pore and thus form a more stable threaded structure. Note that the reduction in current level for captures with the new conjugate was smaller (~35%) than that of the previous one (~67%). This suggests that the molecule reached a different threaded structure, presumably extending a shorter distance into the barrel. This is consistent with the notion that the blockade level increases with the fractional volume of the pore occupied by the polymer.[98,100] Overall, these data are in good agreement with structural considerations and roughly map the distance from the attachment to the barrel of the pore.[52,98-100] A summary schematic of behaviors of the DNA molecules on the two conjugates under the electric field is presented in Figure 4.8.

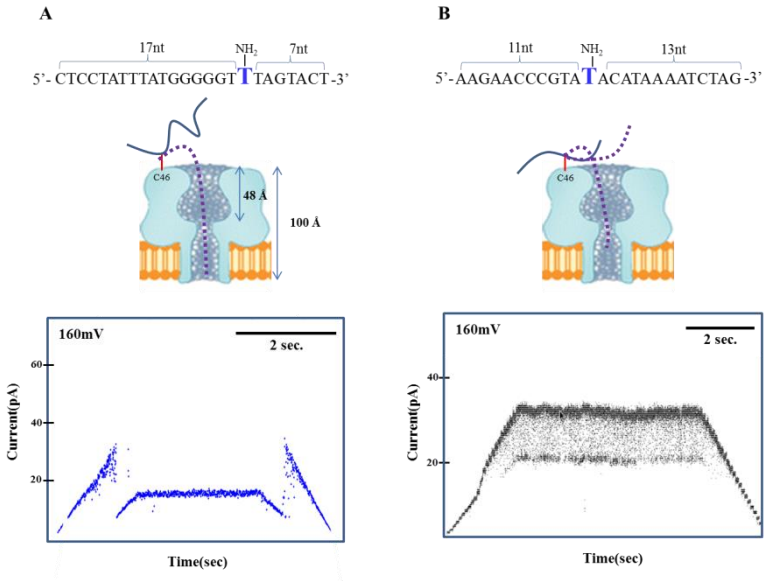


Figure 4.8 A schematic of possible behaviors of the two DNA molecules at equivalent voltage.
A. The ssDNA having 17 nucleotides is long enough to almost extend through the full length of the pore, generating a steady blockade event. **B.** The ssDNA with 13 nucleotides forms a less stable and weaker blockade level when threaded. These signatures are consistent with the dimension of the channel.

4.3.8 Current signature of hairpin DNA attached to the pore

We also investigated how a DNA molecule containing a secondary structure with an intramolecular base-paired hairpin would behave when conjugated to the αHL pore. Hairpins are sequences of ssDNA that fold back on themselves to form base pairs. As explained earlier, the narrow path of the αHL pore allows passage of ssDNA but not dsDNA. Therefore, the higher-order structure of a molecule such as a hairpin or duplex region must denature as it is driven through the channel by electrical force. Figure 4.9 shows the structure and sequence of the tested DNA molecule. It consists of a 16 base-pair double stranded stem portion that is connected with a small loop, and has a 13 base ssDNA tail with an amino modification at its 5' end serving as the conjugation site. Thus, the 3' end of the hairpin DNA is the only end that can thread into the pore. However, it is in a double-stranded region of the hairpin and thus, if we observe DNA threading, we will know that the DNA has been denatured.

To investigate the behavior of the molecule having a hairpin structure under an electrical potential, varying voltages of 60 – 200 mV were applied and changes in current were recorded (Figure 4.9). In this conjugate, the prolonged blockade events start to appear at substantially higher voltages than needed for conjugates with linear DNA, indicating that a higher potential difference is required to unzip and pull the 3' end of the DNA into the pore. The length of the unzipped DNA is sufficient to fully traverse the pore. Thus, we interpret this to mean that the DNA is denatured by electrical force and its leading end is completely driven through to the *trans* side. After the 6 seconds of pulse duration, the voltage starts to ramp back to the ground level and the current begins decreasing proportionally. This is followed by a small but rapid "jump" in current and this tail signature is indicative of the molecule finally escaping the pore as the applied potential approaches 0 mV. Note that the threaded DNA is released at the very end of

the falling period. This delayed release event suggests that the threaded DNA tends to stay in the pore even when the external force is not sufficient to hold it there.

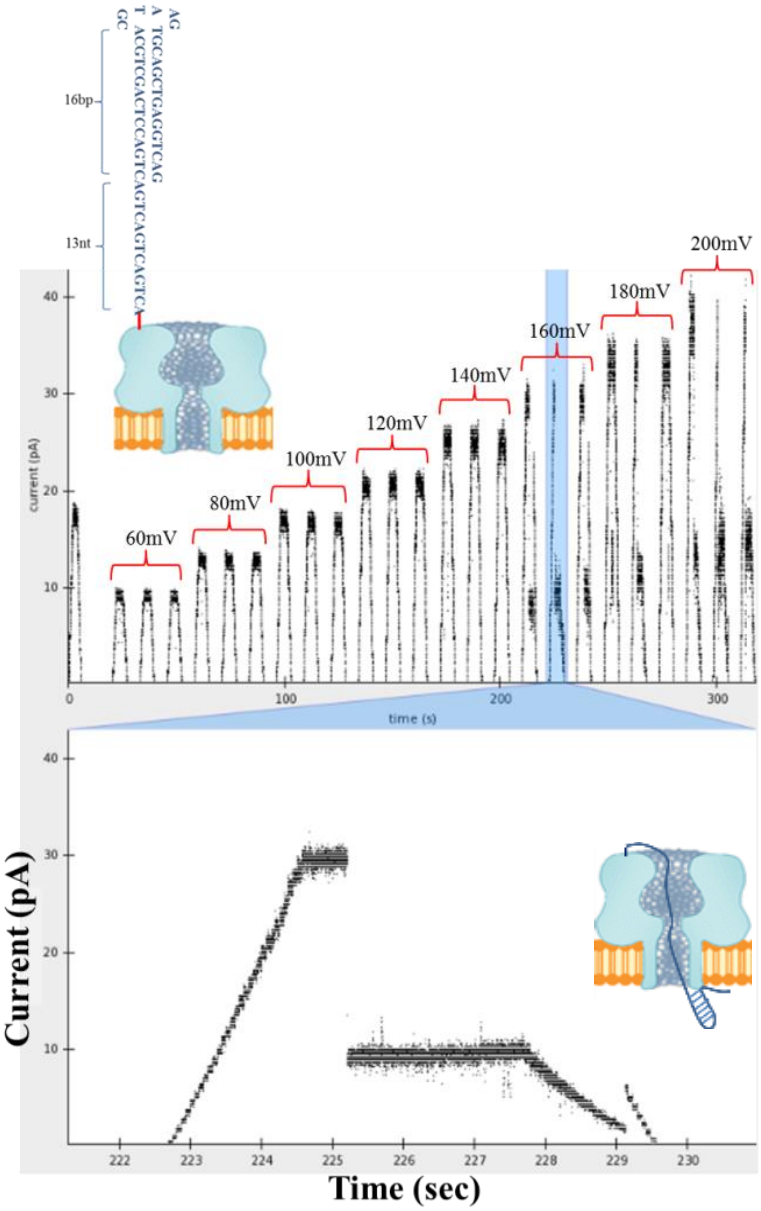


Figure 4.9 A graph of current trace against time from the conjugate with hairpin DNA.

An interesting characteristic of the current signature of this conjugate is that it showed asymmetrical behavior, unlike the previous ssDNA attached conjugate that produced a mirror image of blockade event with similar threshold potentials for both capture and release. This

asymmetry in its signature suggests that the potential required to capture the molecule to the pore is different from the level needed to release it from the pore. As discussed earlier, a previous study showed that the narrower diameter pore entrance on the *trans* side presents a higher entropic barrier and thus the DNA capture rate from the *trans* compartment is significantly lower than from the *cis* side.[101] Based on this, we interpret the asymmetrical trend in this conjugate to imply that a substantial portion of the unzipped DNA is completely translocated to the *trans* side where the entropic barrier is greater and thus the DNA end is less easily re-captured back into the pore from the *trans* side, thus taking a longer time to be released. Furthermore, due to the presence of complementary sequences, it is also possible that the end of the molecule might partially fold back on itself again after completely traversing the pore to the *trans* side, forming a secondary structure with sufficient stabilizing energy to basically lock itself inside the pore (Figure 4.9). This, along with the increased entropic barrier due to the narrow entry, might contribute to its extended stay inside the pore.

4.3.9 Removing the hairpin structure

As a means to confirm the structure of this molecule, another control experiment was performed with this conjugate. A *Pvu*Π restriction site is included in the double-stranded portion of the molecule (Figure 4.10 (Top)). Treatment with *Pvu*Π cuts off the hairpin loop, resulting in a short intermolecular base-paired duplex. To test if this structural change can be detected in the current traces, several stepped waveforms were generated at various voltages. Currents were recorded before and after the direct addition of the enzyme to the same conjugate. As shown in Figure 4.10, the asymmetrical current blockade signature was produced before the addition of the restriction enzyme. However, the shape of the current trace is pronouncedly different after the addition of the enzyme, consistent with the expected structural change in the molecule.

Furthermore, a stable blockade event is detected even at a relatively low voltage such as 80 mV. We interpret this to mean that the shorter duplex DNA region without the loop is more easily unzipped and the remaining 21mer ssDNA threads into the pore at a lower voltage. Overall, these data indicate that structural differences of the molecule could be detected in the nanopore system by monitoring changes in current signals.

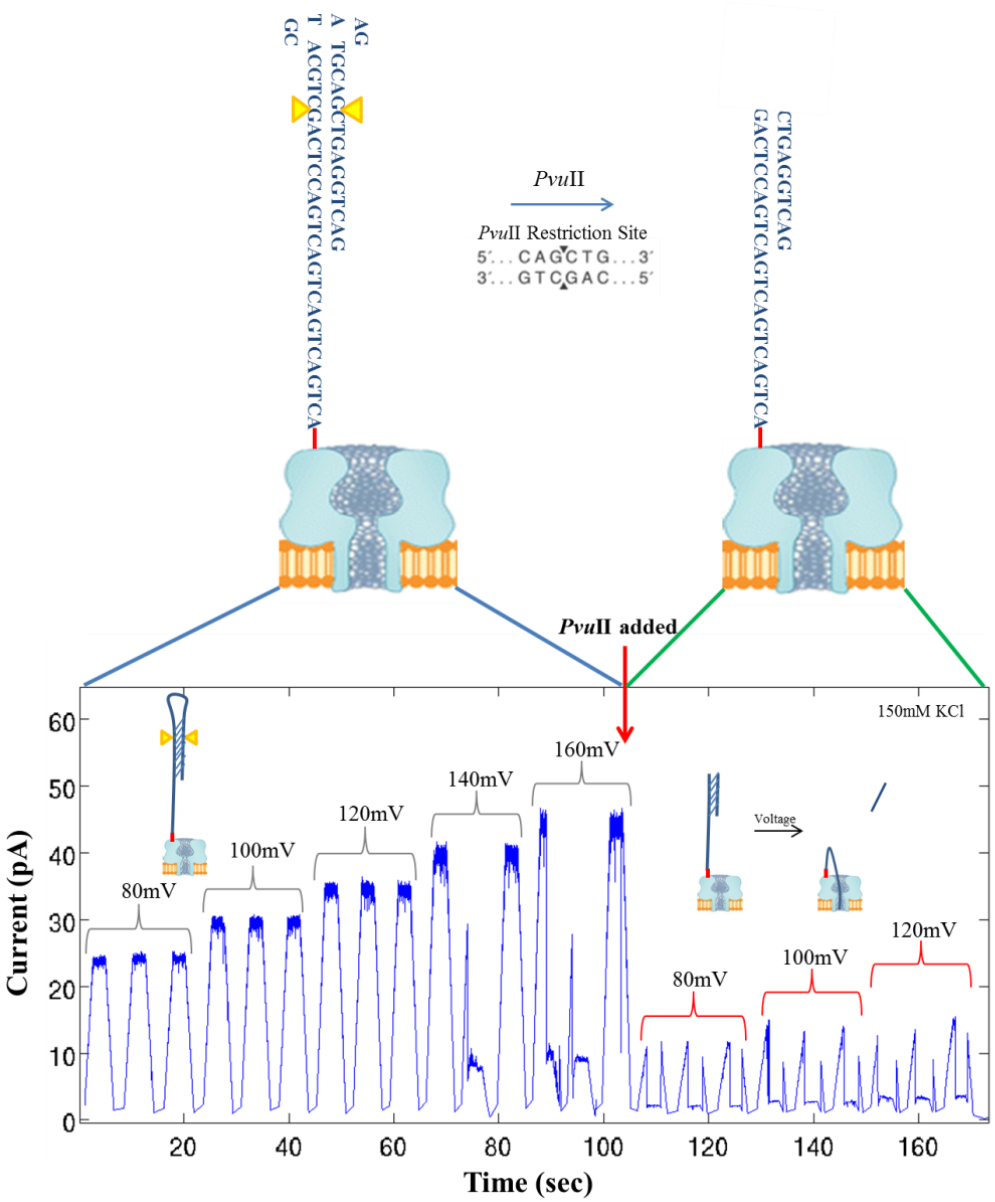


Figure 4.10 Current trace against time before and after *Pvu*II treatment of the conjugate with hairpin DNA.

4.4 Conclusion

Analysis of biomolecules at the single molecule level can reveal valuable information that is often disguised by ensemble averaging such as their detailed dynamic movements or temporary intermediate states. In this chapter, current signatures elicited by changes in length and structure of ssDNA attached on the pore were studied in the nanopore system. Consistent with previous studies, our data support the notion that the current signals collected at various voltages reveal useful information about the molecule such as its structural details and dynamic behaviors.[98,100,102,103] At the same time, the results here also demonstrate that tethering DNA to the channel could serve as a molecular ruler to probe the geometry and electric potential within the pore. [98,100]

CHAPTER 5: Single Molecule Electronic SNP Assay in a Primer-conjugated Nanopore

5.1 Introduction

Armed with the insights from the DNA capture experiments in the previous chapter, a novel single molecule electronic SNP assay that employs SBE of a primer using polymer-tagged ddNTPs for allele discrimination and an electrical readout of the αHL nanopore for allele detection is established. As described earlier, a ssDNA primer conjugated αHL pore is first prepared and serves as a platform for the entire assay (Figure 5.1). In this assay, the primer is attached to the pore like the DNA-pore conjugates in Chapter 4. Therefore, the primer can also thread into the pore and produce background events if it is attached close to the pore opening. To avoid this, other conjugation sites in the pore, AA positions 239 and 93, which are located further from the pore entrance are explored (Figure 3.5). We observed that conjugates with 18-mer DNA attached to the C239 site had a tendency to produce weak blockade events at higher voltages (180 mV or 200 mV). Thus, in order to have a full window for measurements, we chose C93 as the conjugation point in the 6(WT):1(C93) pores for this assay. After verifying conductance of the conjugate, *in-situ* SBE reactions are performed by applying circular template, DNA polymerase and four polymer tagged ddNTPs that contain unique tag structures for identification. The polymerase extends the primer with one of the ddNTPs depending on the sequence of the template and the long polymer tag on the extended primer is then pulled into the pore by electrical force, generating a distinct current blockade level specific to each of the nucleotides. By monitoring current signatures, the genotype of the template is deciphered.

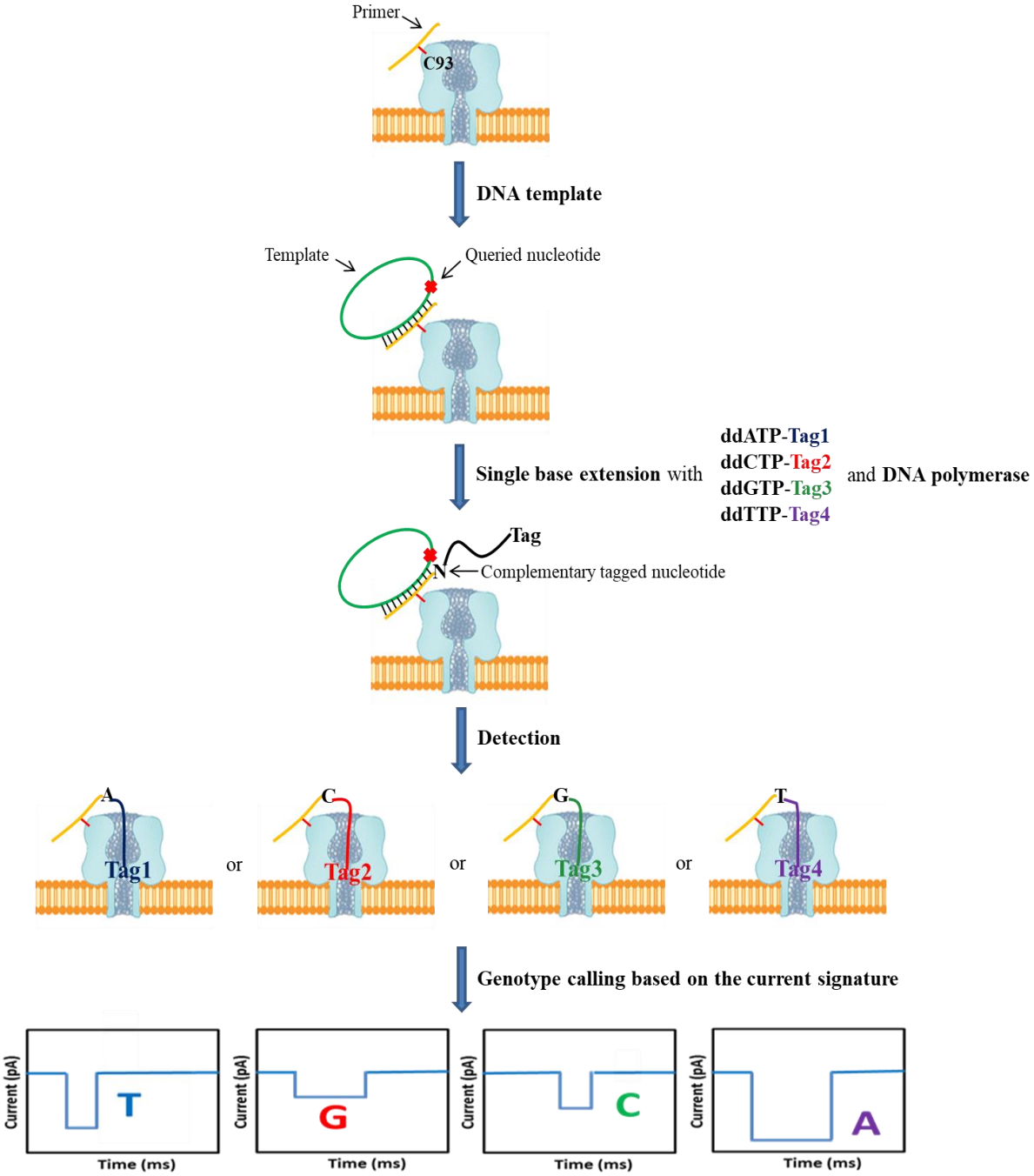


Figure 5.1 A scheme of a single molecule electronic SNP assay using four tagged ddNTPs in a primer-conjugated nanopore array.

Nanopore that is conjugated with primer is prepared. SBE is performed by adding ddNTPs with tags on the base position, DNA polymerase and circular template. The primer is extended by the complementary nucleotide to the queried site in the template by DNA polymerase. Under an applied potential, the tag on the extended primer is drawn into the pore and this generates a characteristic event in the measured current signal, thereby revealing the genotype of the template.

In this assay design, the position of the tag molecule in the pore after extension is dependent on the initial primer length that is conjugated to the pore. In other words, a change in its length might move the identifying moiety in tags away from the constriction zone where the largest current change occurs in the pore, compromising the accuracy of tag recognition. Although potential variations in primer length are likely to be minor and thus the position certainly can be adjusted by remapping the molecules, the assay is not very flexible if the tag molecule has to be calibrated based on the primer length.

To improve this aspect of the assay, another version of tags containing an internal positioning structure was also explored. As mentioned above, the entry of dsDNA is limited to the vestibule of the αHL pore and there is barely enough room for ssDNA to pass through the narrowest constriction point of the channel. Taking advantage of this structural barrier, tagged ddNTPs with a hairpin structure in the middle of the tag sequence were designed (Figure 5.2). The structure begins with a nucleotide with a short linear left overhang (A) which will be covalently attached to the primer after SBE followed by a hairpin structure in the middle (B), and it ends with another linear overhang with an identifying moiety (identifier) positioned right after the hairpin structure (C) for producing a unique current signature. Since this tag design contains a double stranded secondary structure, the hairpin on the extended primer is held within the vestibule under the applied potential locating the identifying moiety at a fixed position inside the channel until it unzips. In this way, the tag position within the pore is solely defined by the hairpin structure and no longer depends on the primer length. Several studies have measured the time required for thermally activated opening of hairpin DNAs and they typically last several hundreds of milliseconds depending on the strength of their base pairing energy; this is sufficient time to record the current signatures of the identifying moiety in the tags.[114,115]

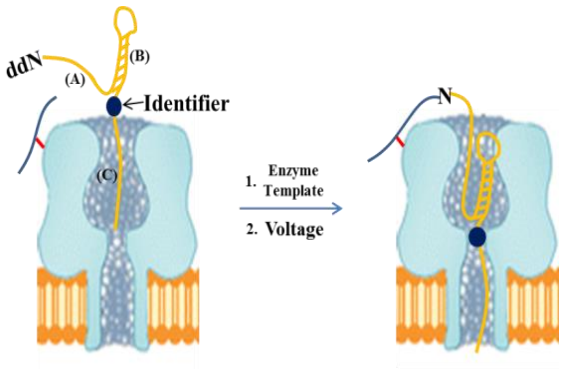


Figure 5.2 A schematic of a likely position of a hairpin tag in the pore under the applied voltage.

To employ these secondary structure-containing tags, the assay was slightly modified to include an additional washing step before acquiring data for tag recognition on the extended primer (Figure 5.3). The primer-conjugated nanopore is first prepared as in the previous approach. After performing *in situ* SBE by applying hairpin tagged ddNTPs, template and DNA polymerase onto the conjugate, several rounds of a washing step is performed to ensure removal of residual hairpin tagged ddNTPs in the solution, which might otherwise generate false positive signals. After thorough washing, only the hairpin tag on the nucleotide that is permanently incorporated into the primer remains in the system and produces the appropriate current signature when the voltage is applied for the identification of the tag.

To test the feasibility of this version of the assay using hairpin tagged nucleotides, four hairpin tagged ddNTPs were synthesized following the same synthetic strategy described in Chapter 2. After confirming that they can be also recognized by the DNA polymerase, we first investigated the conditions for completely washing away the residual tag molecules in the solution and then examined their characteristic blockade levels.

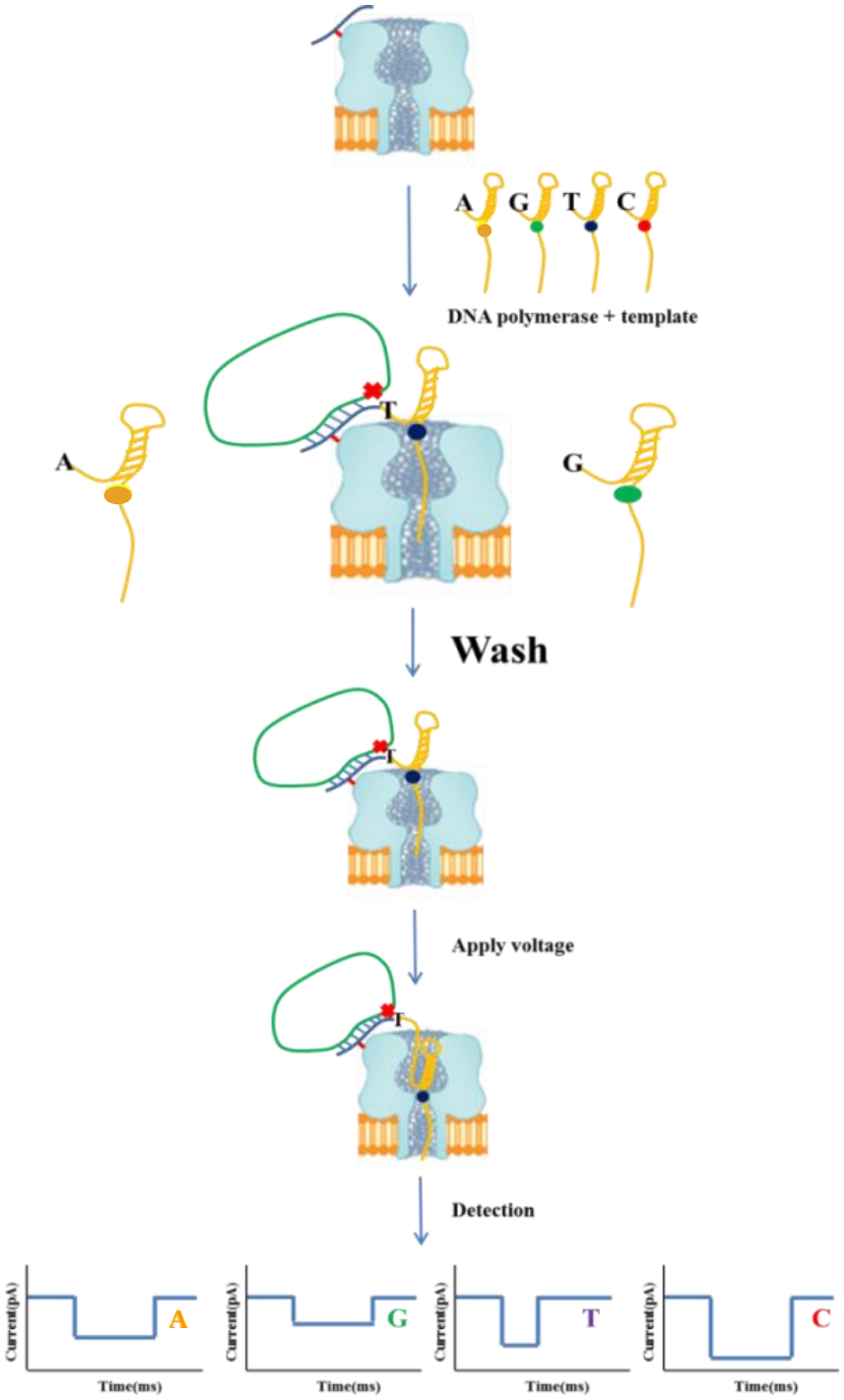


Figure 5.3 A schematic of the single molecule electronic SNP assay using hairpin tagged ddNTPs. An extra washing step is included before applying voltage for tag recognition to remove the residual tagged nucleotides in the system.

5.2 Methods

5.2.1 Template circularization

5’ phosphorylated synthetic 70mer ssDNA templates having a different base at the queried site were designed. Sequences of template are as follows: Temp G 5’-Phos/ATGACTCCAGGCCCTCTGAG**G**GATAGCAGGAAGCAGAACCCACCAGACCAGGCCCCTGA-3’, Temp T 5’-Phos/ATGACTCCAGGCCCTCTGAG**T**GATAGCAGGAAGCAGAACCCACCAGACCAGGCCCCTGA-3’, Temp C 5’-Phos/ATGACTCCAGGCCCTCTGAG**C**GATAGCAGGAAGCAGAACCCACCAGACCAGGCCCCTGA-3’, Temp A 5’-Phos/ATGACTCCAGGCCCTCTGAG**A**GATAGCAGGAAGCAGAACCCACCAGACCAGGCCCCTGA-3’. The ligation reaction was performed in a 20 µl volume with 500 pmol DNA, 1000 units of CircLigase (Epicentre), 2 µl 10X reaction buffer, 50 µM ATP and 2.5 mM $MnCl_2$, and incubated at 37 °C overnight. The reaction products were subsequently treated with 300 units of exonuclease I (Epicentre) and incubated for 4 hours. The final products were purified using a QIAquick PCR purification kit (Qiagen) and the circular nature of the products was confirmed on an 8M urea 15% polyacrylamide gel.

5.2.2 *In situ* SBE on the conjugate in the nanopore system

After the automated membrane formation and pore insertion processes, SBE reactions were performed by applying a final concentration of 3 µM circular template, 5 µM tagged nucleotides, 0.04 unit of Thermo Sequenase, 6.5 mM $MgCl_2$ in a 150 mM KCl, 20 mM Hepes, pH 7.4 buffer.

5.2.3 Testing hairpin tagged nucleotides in the nanopore system

The experiment was performed in a buffer containing 300 mM KCl, 20 mM Hepes, pH 7.4. Each nucleotide was diluted in the buffer to a final concentration of 3 µM. Washing was achieved by flowing buffer into the system.

5.2.4 Primer extension by hairpin tagged ddNTPs and *Hha*I digestion of the hairpin structure

The extension reactions were performed in a 20 µl volume with 4 units of Thermo Sequenase, 1X reaction buffer, 20 pmol of specific self-primed looped template and 200 pmol of matching hairpin tagged ddNTPs. The sequences of the four templates are as follows: Temp C 5'-GAT**C**GCGCGCGCCTTGGCGCGGCGC-3', Temp G 5'-GAT**G**GCGCCGCGCCTTGGCGC GGCGC-3', Temp T 5'-GAT**T**GCGCCGCGCCTTGGCGCGGCGC-3', Temp A 5'-GAT**A**GC GCCGCGCCTTGGCGCGGCGC-3'. They were incubated at 65 °C for 1 hour and half of the reaction products were treated with 20 units of *Hha*I restriction enzyme for 30 mins at 37 °C. The extension and digestion products were confirmed on an 8M urea 15% polyacrylamide gel.

5.3 Results and Discussion

5.3.1 Single molecule electronic SNP assay using a primer-conjugated nanopore and linear polymer tagged ddNTPs

To avoid potential background events due to the ssDNA primer conjugated on the pore, AA position 93 on the side of the pore was explored as a conjugation site. This mutant was generated using the same PCR-based site-directed mutagenesis method and 18-mer ssDNA primer molecule was attached to 6(WT):1(C93) pore using the same conjugation strategy as described in chapter 3. Figure 5.4 shows the current trace for this conjugate at various voltages. Unlike the previous conjugate that generated strong blockade signals, the DNA conjugated at this position (C93) of the pore no longer threaded into the pore and thus a clean and steady open current trace without any interruptions was produced. The presence of the primer conjugated αHL monomer was confirmed after pore denaturation by SDS-PAGE analysis (Chapter 3).

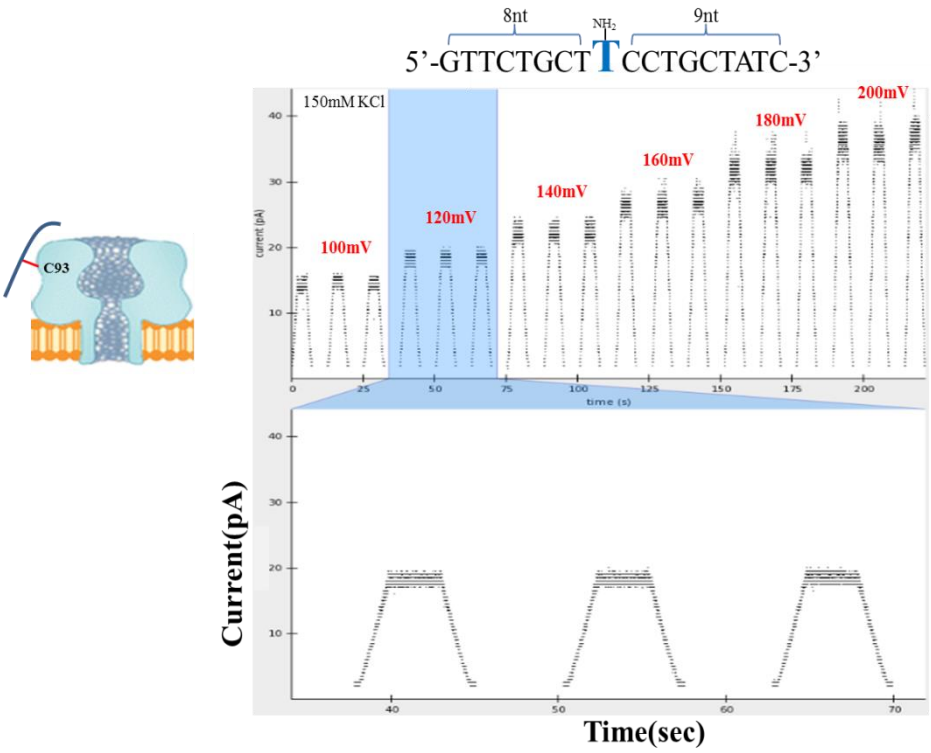


Figure 5.4 Current trace against time from the conjugate with DNA on the side of the pore.

To prevent the template from being pulled into the pore and producing a background signal, a 40mer ssDNA template was circularized using CircLigase™ (Epicentre), which catalyzes intramolecular ligation of the ends of a ssDNA template containing a 5'-phosphate and a 3'-hydroxyl group. The conversion of ssDNA into a circular DNA form can be confirmed by either mobility shift following Urea-PAGE analysis or digesting the reaction with exonuclease I which specifically digests linear ssDNA having a free 5' or 3' end. Figure 5.5A shows the results of circularization reactions with the four different templates along with control linear ssDNAs before the reaction (Linear). The conformational change of the molecule was verified by both the change in its mobility on the gel (Pre-Exo) and also by its resistance to exonuclease I (Post-Exo). The circular nature of the molecule was further verified by applying it to the nanopore system with the pore-primer conjugate (Figure 5.5B). The clear trace at various

voltages in this experiment shows that neither the primer nor template produces any background current blockade events.

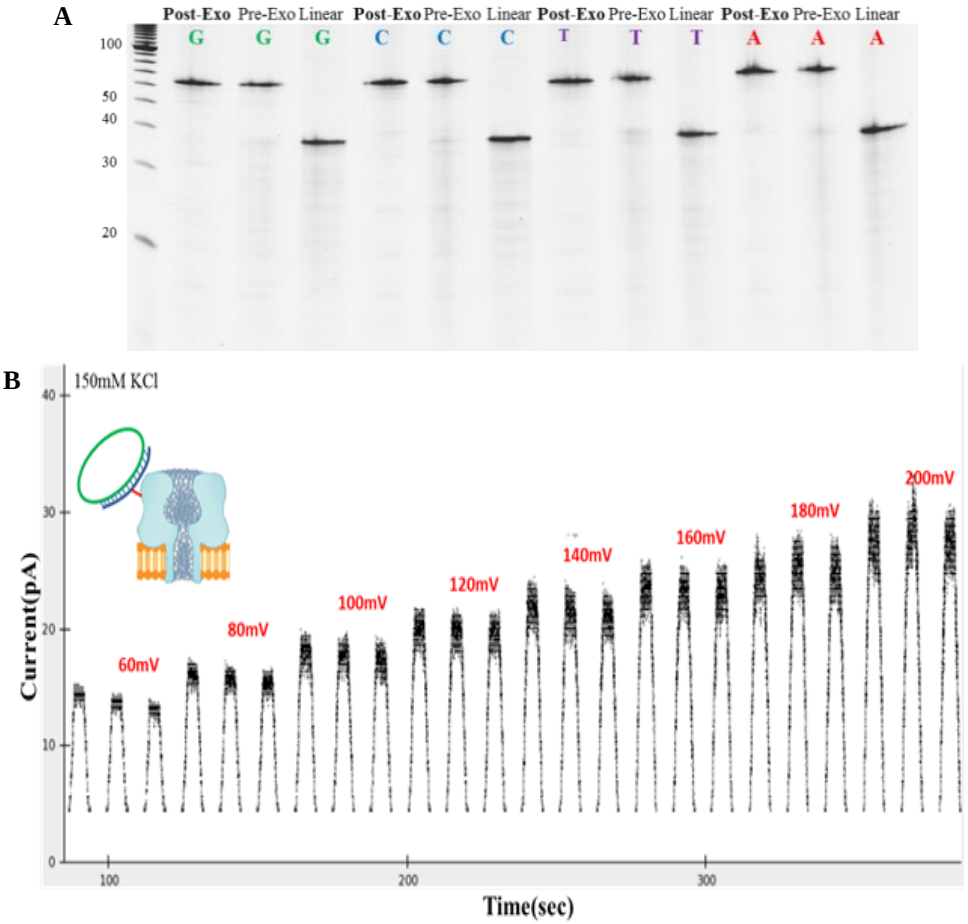


Figure 5.5 A. Urea-PAGE gel analysis of circularization reactions. **B.** A graph of current trace against time after the addition of circular template to the conjugate.

After checking the background, *in-situ* SBE was performed by adding specific template, DNA polymerase and tagged ddNTPs to the membrane-bound pore-primer conjugate. Considering that a complementary nucleotide would be permanently incorporated into the primer that is covalently bound to the pore, we expected to obtain the blockade signal in every trapezoidal pulse if it is indeed from the tag on the incorporated nucleotide. Other signals that occurred intermittently were disregarded. Figure 5.6 shows the current traces and a summary of blockade levels generated in four different reactions with one specific template and matching

nucleotide. The reaction with template C and G nucleotide tagged with 17 adjacent dSps, the moiety with the smallest width, allowed more current to flow, thus resulting in the highest blockade level (I/I_o = 0.68). Reactions with G and A templates where complementary nucleotides, C and T contain less dSp units (8 or 3) elicited slightly lower levels (I/I_o = 0.58 and I/I_o = 0.47, respectively). On the other hand, the A nucleotide with a neutral moiety (6 Imp) blocked most of the current flow and thus its level was the lowest among the four tagged ddNTPs (I/I_o = 0.36). A summary of the four different levels is shown in Figure 5.6E.

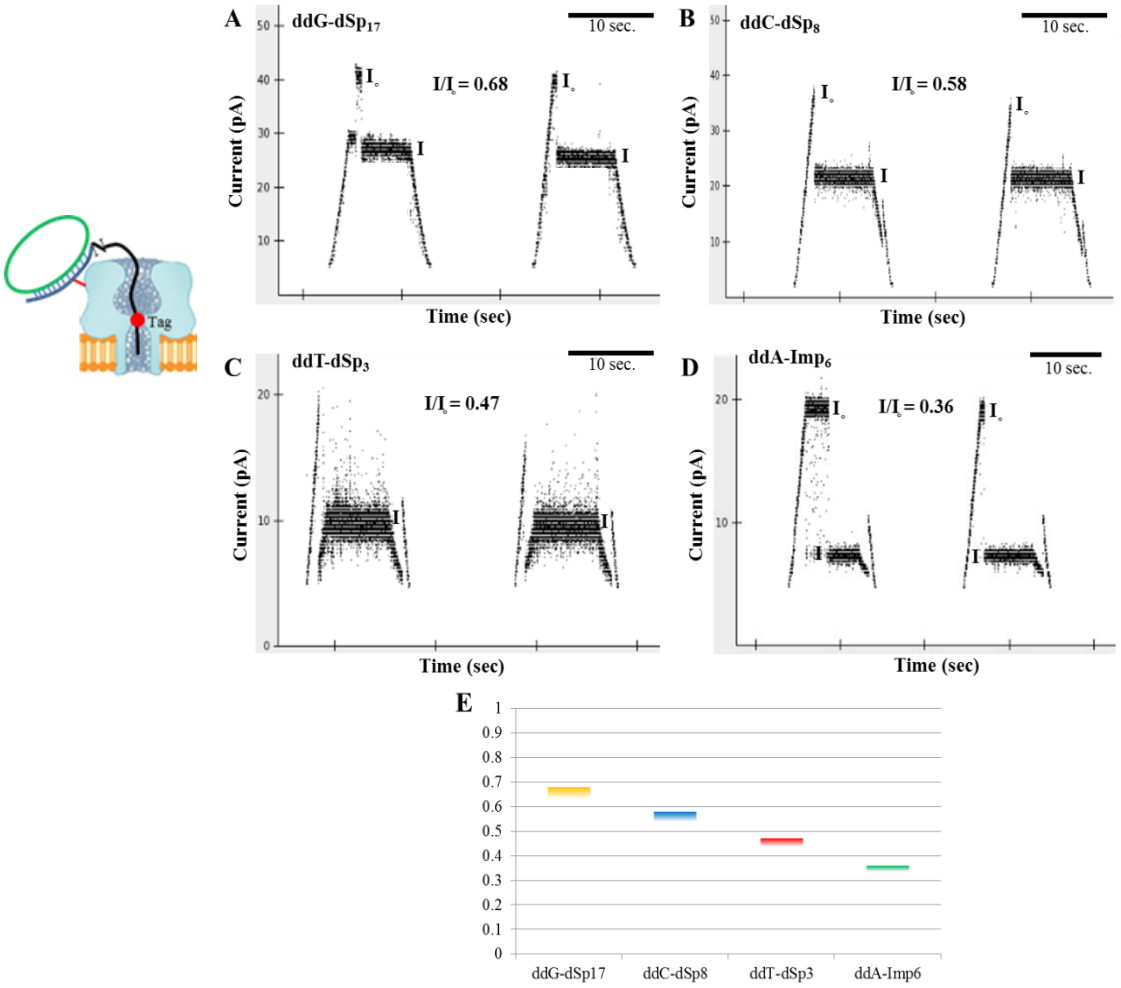


Figure 5.6 Current signatures of four tagged ddNTPs in reactions with single nucleotide and matched template. **A.** Detected current signature of the *in situ* SBE reaction with template C, Thermo Sequenase and tagged ddGTP on the conjugate. **B.** Current signature of the reaction with Template G, Thermo Sequenase and tagged ddCTP on the conjugate. **C.** Current signature of the reaction with template A, Thermo Sequenase and tagged ddTTP on the conjugate. **D.** Current signature of the reaction with Template T, Thermo Sequenase and tagged ddATP on the conjugate. **E.** A summary of current levels of the four reactions.

In this approach, the SBE reaction is achieved by applying DNA polymerase, template and nucleotides to the conjugate in which primer, one of the enzyme's substrates, is immobilized on the membrane-bound pore. Thus, for the reaction to occur, all three components need to diffuse towards the immobilized primer to form the ternary complex; first complementary template should approach and bind to the primer and then enzyme should grip the matching nucleotide and bring it to the active site. Unlike the reaction in solution where all the reactants are free to diffuse and thus collide with each other, the rate of the reaction in this approach was unfortunately too slow to gather enough samples of blockade events.

To improve the reaction efficiency, we have conducted a few straightforward initial tests, such as increasing the amount of template, nucleotide or enzyme in the reaction within the constraints of the system, extending the incubation time up to 2 hours and increasing temperature of the system up to 45 °C to accelerate the diffusion rate. Unfortunately, we could not succeed in obtaining a sufficient number of events for statistical analysis by any of these means.

Nonetheless, we are convinced that this starightforward assay has great potential for genotyping and additional room for improvements. A previous study has shown that the rate of enzymatic reaction on the immobilized substrate is determined by both the intrinsic reaction rate of the enzyme and the diffusion of the enzyme between substrate sites.[119] Thus, in the future, improvement of the enzyme's intrinsic kinetic properties by mutagenesis might lead to an increase in reaction rate. Or, modification of the enzyme to bear more negative charges could also help drive the enzyme to the primer site by increasing electrophoretic strength under the applied voltage. On the other hand, previous studies have attached a cholesterol anchor to a DNA molecule and have shown that this can serve to concentrate them at the lipid bilayer, thereby increasing their capture rate in the nanopore.[116,117] Based on this, the lipophilic concentration can

also be increased by conjugating hydrophobic moieties to the template and enzyme to concentrate them in the vicinity of the membrane-bound primer-conjugated pore.

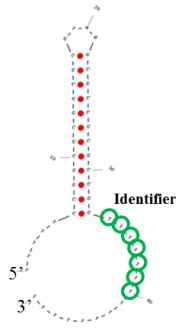
5.3.2 Design and synthesis of tagged nucleotides with a secondary structure

In an attempt to render more flexibility to our approach using primer-conjugated nanopore, a variation of tags having a secondary structure to aid in proper positioning of the identifier in the tag was also explored in our system. As discussed earlier, our previous tags are built on a thymidine (dT) backbone with specific identifiers in the middle (Figure 2.2). Although minor, we noticed occasional interactions between the dT backbone of the tags and DNA molecules possessing a sequence of deoxyadenosine (dA) in one of our control SBE reactions and found that these can produce false-positive background events in the nanopore system. In order to prevent such interaction between the tag backbone and other DNA molecules in the reaction mixture, a backbone of repeated deoxythymidine (dT) molecules but interspersed with a propyl spacer was employed in the new set of hairpin tagged nucleotides (Figure 5.7). As with previous linear tagged ddNTPs, each hairpin tag molecule begins with a 5'-hexynyl moiety for conjugation to ddNTPs via a click chemistry process followed by a short 5-mer backbone ssDNA tail. The rest of the tag is composed of another ssDNA portion consisting of a unique identifier with a double stranded hairpin structure in the middle. In general, the stability of the hairpin structure is mainly influenced by the strength of the stem portion which is determined by its length and GC content and by the size of the loop.[111] Given these factors, a strong hairpin loop structure was designed with 12 G-C complementary sequences in the stem connected by a stable loop sequence that has been shown to display high thermodynamic stability.[112,113] For this version of tagged ddNTPs, we tested four new identifying components to explore different current signatures. As a standard, an oligonucleotide molecule with 6 deoxythymidines (T6)

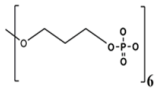
following by the hairpin was designed and expected to reduce the current to approximately 60 % of the open current when it is held in the pore based on a previous study from our laboratory.[56] In order to create a higher current level, another molecule with 6 propyl phosphates, which have much smaller diameters than T6 was tested. For the remaining two tags, 2 units of relatively neutral, thymidine methyl phosphonate (Tmp) were selected to generate a slightly lower level than T6 and 6 consecutive units of Tmp right after the hairpin were designed to produce the lowest current level. These identifiers in the tags have been previously demonstrated to produce distinct signal in our laboratory. To stably capture the molecule in the pore, the tail of the tag was elongated with an additional backbone ssDNA. These molecules were attached to ddNTPs using the same synthesis strategies described in Chapter 2. After their purification by RP-HPLC, the molecular weights of each compound were confirmed by MALDI-TOF MS analysis (Figure 5.8).

Using this new version of hairpin tagged ddNTPs with a secondary structure, we first assayed for their recognition and incorporation by Thermo Sequenase in SBE reactions. Due to the high molecular weights of the hairpin tagged ddNTPs, there is a significant change in molecular weight between the original primer and the extended product. In addition, in the previous version of tagged ddNTPs, an optional Cy3 tracer was added to confirm that the primer was extended by the expected molecules, the tagged ddNTPs. Following similar logic, in this design, two *Hha*I restriction digestion sites (5’-GCG'C-3’) were included in the stem sequences to confirm that the final products include hairpin tagged ddNTPs. As indicated in Figure 5.9, the tagged nucleotides are digested into 5 pieces by the *Hha*I enzyme and thus the high molecular weights of new products should disappear after treatment with the restriction enzyme if they contain the hairpin tagged ddNTPs. Figure 5.10 shows the comparison of molecules in the extension reactions before and after *Hha*I treatment on a denaturing polyacrylamide gel.

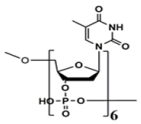
Although it was not possible to clearly separate small fragments of digested DNA molecules, this clearly shows the presence of new high molecular weight products after the extension reaction and their disappearance upon *Hha*I enzyme treatment, demonstrating that our hairpin tagged ddNTPs can also be incorporated into the primer by Thermo Sequenase.



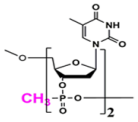
A 5' -(Hexynyl)TT(Propyl)TTTCCGGCGCGGCGCGTAAGCGCCGCGCCGG
(Propyl)(Propyl)(Propyl)(Propyl)(Propyl)(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)-3'



B 5' -(Hexynyl)TT(Propyl)TTTCCGGCGCGGCGCGTAAGCGCCGCGCCGG
TTTTTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)-3'



C 5'- (Hexynyl)TT(Propyl)TTTCCGGCGCGGCGCGTAAGCGCCGCGCCGG
T(mp)T(mp)TTTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)-3'



D 5'- (Hexynyl)TT(Propyl)TTTCCGGCGCGGCGCGTAAGCGCCGCGCCGG
T(mp)T(mp)T(mp)T(mp)T(mp)T(mp)(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)-3'

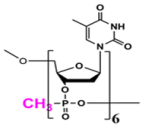


Figure 5.7 Sequences of four hairpin tags and structure of an identifying moiety in each tag. **A.** ddA-hp-propyl$_6$ **B.** ddC-hp-dT$_6$ **C.** ddG-hp-Tmp$_2$ **D.** ddT-hp-Tmp$_6$

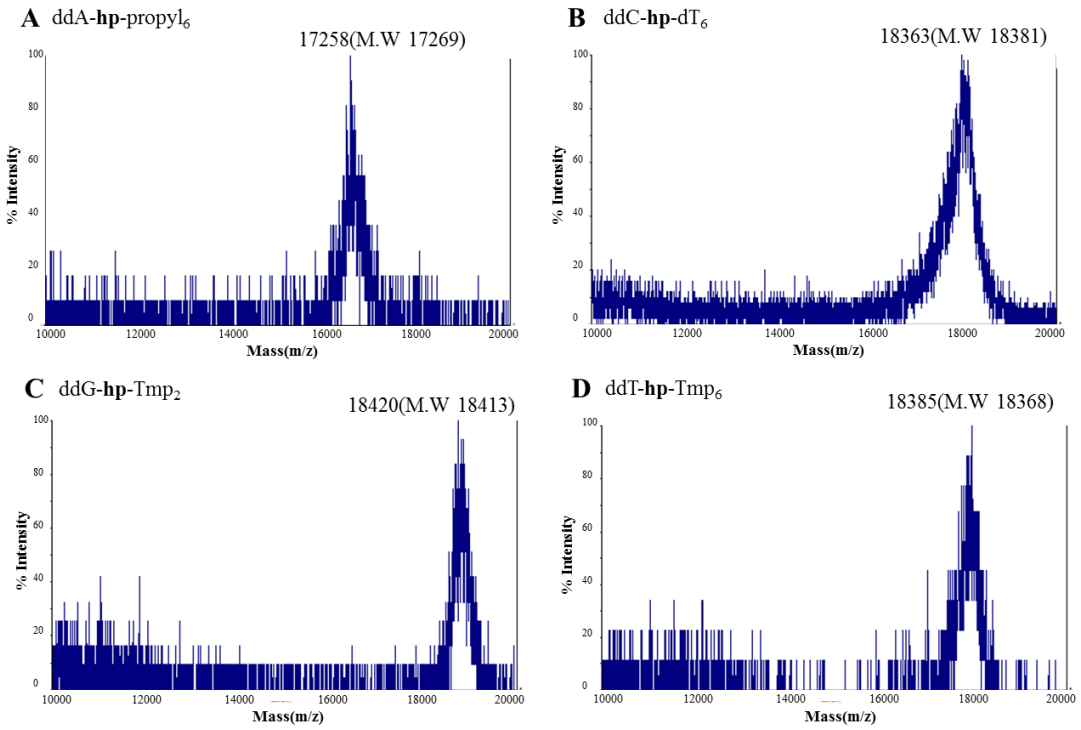
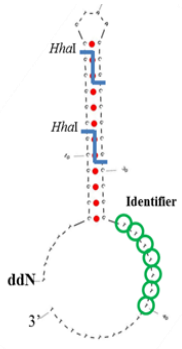


Figure 5.8 MALDI-TOF MS spectra of four hairpin tagged ddNTPs. The expected molecular weights of these nucleotides are **A.** 17269 Da for ddATP-hp-propyl$_6$, **B.** 18381 Da for ddCTP-hp-T$_2$-dT$_6$, **C.** 18413 Da for ddGTP-hp-Tmp$_2$, **D.** 18368 Da for ddTTP-hp-Tmp$_6$. The measured values are shown in the spectra.



A ddA-TT(Propyl)TTTCCG**GCGC**G**GCGC**GTAA**GCGC**C**GCGC**CGG
(Propyl)(Propyl)(Propyl)(Propyl)(Propyl)(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)-3'

B ddC-TT(Propyl)TTTCCG**GCGC**G**GCGC**GTAA**GCGC**C**GCGC**CGG
TTTTTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)-3'

C ddG-TT(Propyl)TTTCCG**GCGC**G**GCGC**GTAA**GCGC**C**GCGC**CGG
T(mp)T(mp)TTTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)-3'

D ddT-TT(Propyl)TTTCCG**GCGC**G**GCGC**GTAA**GCGC**C**GCGC**CGG
T(mp)T(mp)T(mp)T(mp)T(mp)T(mp)(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)TTT(Propyl)-3'

Figure 5.9 The two *Hha*I digestion sites (5'-GCG'C-3') are highlighted and underlined in the stem sequence.

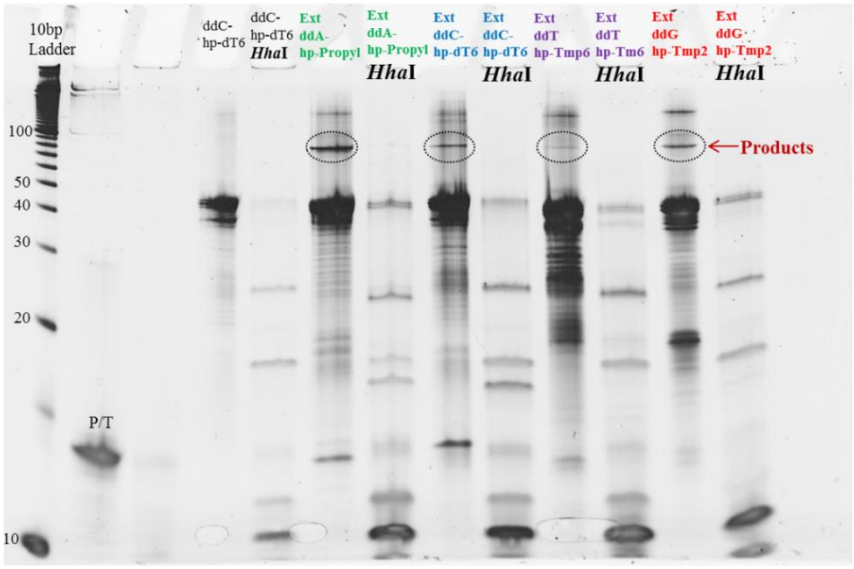


Figure 5.10 Results of extension reactions using self-primed looped template and *Hha*I digestion of stem loop in the hairpin tag. The circled bands indicate extension products. The extended primer and the absence of these molecules after *Hha*I treatment confirms that the primer was extended by hairpin tagged ddNTPs.

5.2.3 Testing a washing step and characterization of current level by the new identifiers

As discussed earlier, an additional washing step that removes any residual nucleotide in the solution is mandatory to employ this type of tag having a secondary structure. Thus, we first tested if the applied hairpin nucleotides could be clearly washed away by flowing excess buffer through the system. To ensure complete removal of the residual nucleotides, the washing was repeated three times. Figure 5.11 shows the current signals at three different points of the experiment: before and after the addition of ddTTP-hp-Tmp_6 and after the washing steps. Noticeably, the signals generated after adding the nucleotide were different from the ones with the empty pore, and those after washing revealed clean traces without any apparent blockade events, indicating the hairpin tagged nucleotide in the system was removed by applying extra buffer to the system.

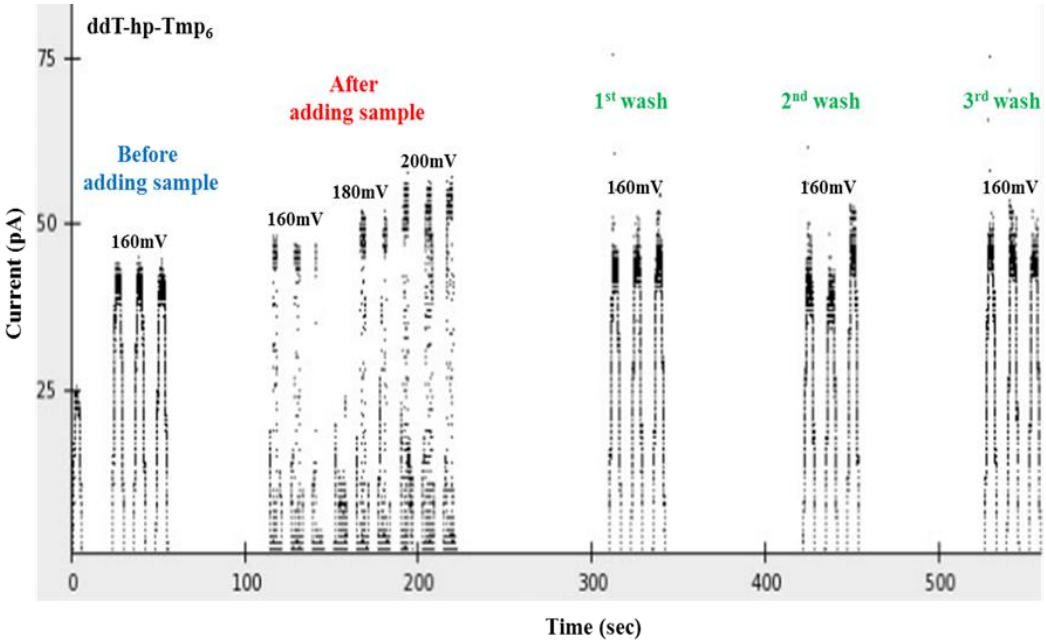


Figure 5.11 Comparison of current traces from the pore before applying the hairpin tagged nucleotide, after the addition and after 3 repetitive washing steps.

After establishing conditions for washing, we then sought to establish reference current blockade levels of the new hairpin tags by applying each one of them to the system. Figure 5.12 shows the current traces of the four newly synthesized hairpin tagged ddNTPs. Unfortunately, these tags generated current signals with high fluctuations. Although there are a few pulses which show stable blockade-like traces, most signals were not easily interpretable due to highly changing levels. Considering the strong stem sequence (Tm = 98.2 °C) of the hairpin structure, we reasoned that this is unlikely to be due to instability or unzipping of the molecule. Nonetheless, one previous study has shown that unlike hairpin structures with a single tail, those with two tails such as ours can in fact be more easily unzipped and translocated through a pore by electric force.[118] The authors argued that hairpins with with two tails are not likely to enter the vestibule of the pore due to their larger structure. This is in contrast to other studies that interpreted their data under the assumption that such hairpins lodge themselves on the bottom of the vestibule of the pore in the same manner as those with a single tail.[125] Rather, these

investigators proposed that the hairpin and one of the overhanging tails stays outside the *cis* entrance of the pore with only the other tail being threaded into the pore (Figure 5.13). Such a position of the hairpin allows it to be more easily opened up under the applied voltage since the force generated is parallel to the base pair plane in such an arrangement compared to when the hairpin is in the vestibule and the unzipping force is perpendicular to the plane.[118] Nevertheless, we used a stronger stem sequence than theirs and our data evidently do not show any uniform signs of unzipping or translocation of the molecule. A few signals, though minor, showed prolonged blockade levels implying that the tags were statically held within the vestibule. Therefore, these puzzling characteristic behaviors of our tags can not be explained clearly with their model either. It is possible that this kind of DNA structure may not produce a uniform level because its motion around the pore entrance is not actually steady under the applied voltage. This might be the reason for the fluctuating current signals in our experiments. Further testing at varying concentrations and voltages and comparisons of molecules having different length or structure hairpins are required to fully clarify the mechanism of these behaviors.

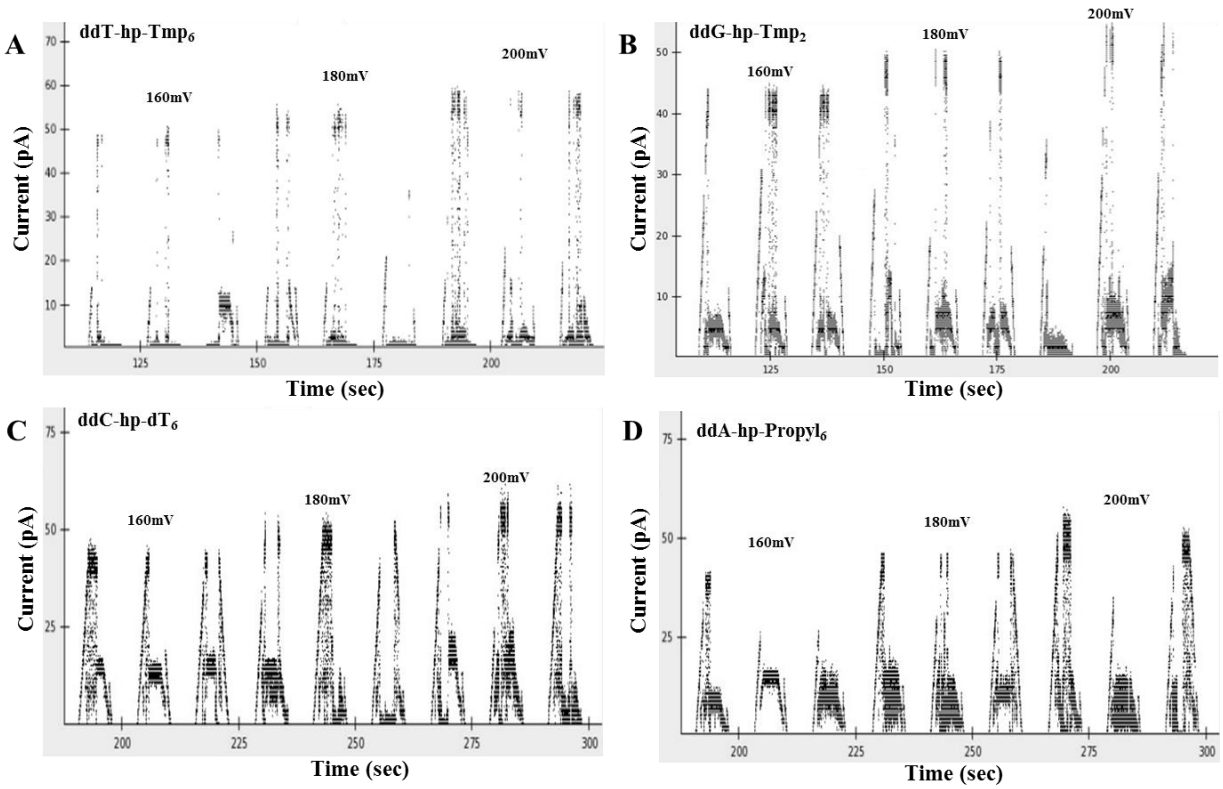


Figure 5.12 Detected current signatures of four hairpin tagged ddNTPs at three different voltages.

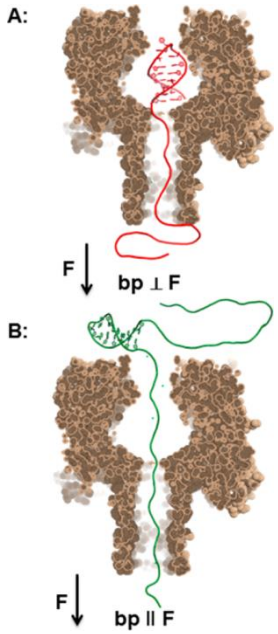


Figure 5.13 The direction of the unzipping force and the base pairs in the two different hairpin structures. **A.** The direction of unzipping force is perpendicular to the direction of the base pair in a hairpin with a single tail. **B.** The direction of unzipping force is parallel to the direction of the base pair, in a hairpin with two tails.[118]

In terms of our assay, generation of four clearly distinguishable levels is a prerequisite for employing these molecules as tags. Thus, the levels must be identifiable in the first place. Also, the hairpin is conjugated to the nucleotide during the synthesis step and eventually it gets incorporated into the primer. Thus, for our purpose, the hairpin structure must have two tails. If future experiments also support the theory that a structure with two tails is more easily disengaged, a hairpin with crosslinked base pairs might stablize the current levels by preventing its unwinding by the electrical force. Such crosslinked tags with unique identifiers at known intervals in the tail can also help further validate the theory of the authors of the above-mentioned studies concerning the hairpin's position under an applied electric potential.[118,125]

5.4 Conclusion

In summary, a single molecule SNP assay was performed on an integrated platform in which allele discrimination is combined with the allele detection step using a primer-conjugated nanopore. Combining an enzymatic allele discrimination process with allele detection in a single system has the advantage that it can ease future automation and scale up of the assay. With our four selected linear polymer tagged ddNTPs, *in situ* SBE was attempted by applying template, enzyme and tagged ddNTPs to the membrane-bound primer-conjugated pore in our detection system. Although we managed to obtain current signatures from each reaction with specific template and matching nucleotide, the reaction rate was not very efficient due to the constraint on primer concentration in this design. Therefore, it was not practical to collect a sufficient number of blockade events from this assay design. In the future, strategies to increase the electrophoretic concentration of the enzyme or lipophilic concentration of the enzyme and template might aid in driving them to the membrane-embedded primer-pore conjugate against

the diffusion rate. This will help verify the current signatures of the tagged ddNTPs and allow for further genotyping experiments on this platform.

In an attempt to remove the dependency of tag position on primer length and consequently to confer more flexibility to the assay, a slight variation of the assay using nucleotides containing a hairpin structure was also examined. To employ this version of tagged ddNTPs, an additional washing step was included to eliminate potential false positive signals by the unincorporated hairpin molecules. To stably seize the molecule in the pore, a structure with a highly stable hairpin and two ssDNA tails was designed and tested. In contrast to our expectation, we found that the motion of this specific structure is not stable in an electric field and thus the current signatures of nucleotides containing tags with hairpin structures could not be clearly defined due to their high fluctuations. Additional tests with variations in structure and under different conditions are expected to provide further insights into the behavior of these molecules under the applied electrical force.

CHAPTER 6: Single Molecule Electronic SNP Assay using Biotin-Streptavidin Interaction

6.1 Introduction

The integrated single molecule SNP assay with immobilized primer on the pore described in Chapter 5 revealed slow reaction rate due to the limited concentration of the primer. Thus, it was not feasible to accumulate enough events for statistical analysis to compare different samples. With the aim of improving the efficiency of the reaction, we developed a modified assay in which the enzymatic allele discrimination process was conducted in advance of the detection step. In this approach, immobilization of tag in the pore was achieved using a biotin-streptavidin complex instead of covalently conjugating primer to the pore. Specifically, SBE is completed in solution with a biotinylated primer, matching template, DNA polymerase and tagged ddNTPs before applying the reaction to the detection system. In this way, the efficiency of the reaction is not limited by the immobilized substrate. After the completion of SBE, the reaction is then treated with streptavidin and applied to the system. Under the applied voltage, molecules are stochastically driven to the pore and since the streptavidin is too large to penetrate the pore, the tag molecule on the streptavidin bound biotinylated primer is immobilized, generating a unique current signature specific to the incorporated nucleotide (Figure 6.2).

To validate this new approach, first, four SBE reactions with biotinylated primer, specific template, DNA polymerase and matching nucleotide were performed and the reference current level of each tag was established in this configuration. Using the verified tag-specific levels for each of the nucleotides, additional experiments were performed to check the specificity of the

enzymatic reaction and also to test the assay's ability to differentiate between homozygous and heterozygous alleles.

After verifying all these properties of the SNP assay with synthetic templates, it was validated again with actual human DNA samples. As a model system, two polymorphic sites in the vitamin K epoxide reductase complex subunit 1 gene (VKORC1), which is implicated in the high inter-individual variability in response to the commonly prescribed anticoagulant drug warfarin, were selected: –1639 C >T (rs9923231) and 1173 G >A (rs9934438).[121,122] The -1639 C >T allele, which is in the promoter region of VKORC1, alters the binding site for transcription factors.[121] Experiments have shown that carriers of the T allele at this position have decreased expression of VKORC1 mRNA, presumably producing lower amounts of functional VKORC1 proteins.[124] Hence, carriers of the T allele require lower doses of warfarin than those with the C allele and they are at the greatest risk for warfarin-related adverse events.[120-123] On the other hand, the 1173G >A site is in the first intron of the VKORC1 gene and is the first SNP found to be associated with the low-dose warfarin phenotype.[121] Although the exact function of this SNP has not yet been determined, studies have shown that in some populations, the common G allele is replaced by the A allele at this site, altering sensitivity to warfarin.[121,123] Thus, carriers of the A allele are considered warfarin sensitive and require lower doses of the drug than those with the G allele.[120-123]

Using these two sites as examples, the surrounding portions of VKORC1 gene was amplified from DNA samples having different genotypes and the double stranded amplicons were then converted to ssDNAs. With these as templates, SBE reactions were performed with biotinylated primer, DNA polymerase and tagged ddNTPs followed by the streptavidin treatment and analysis in the nanopore system.

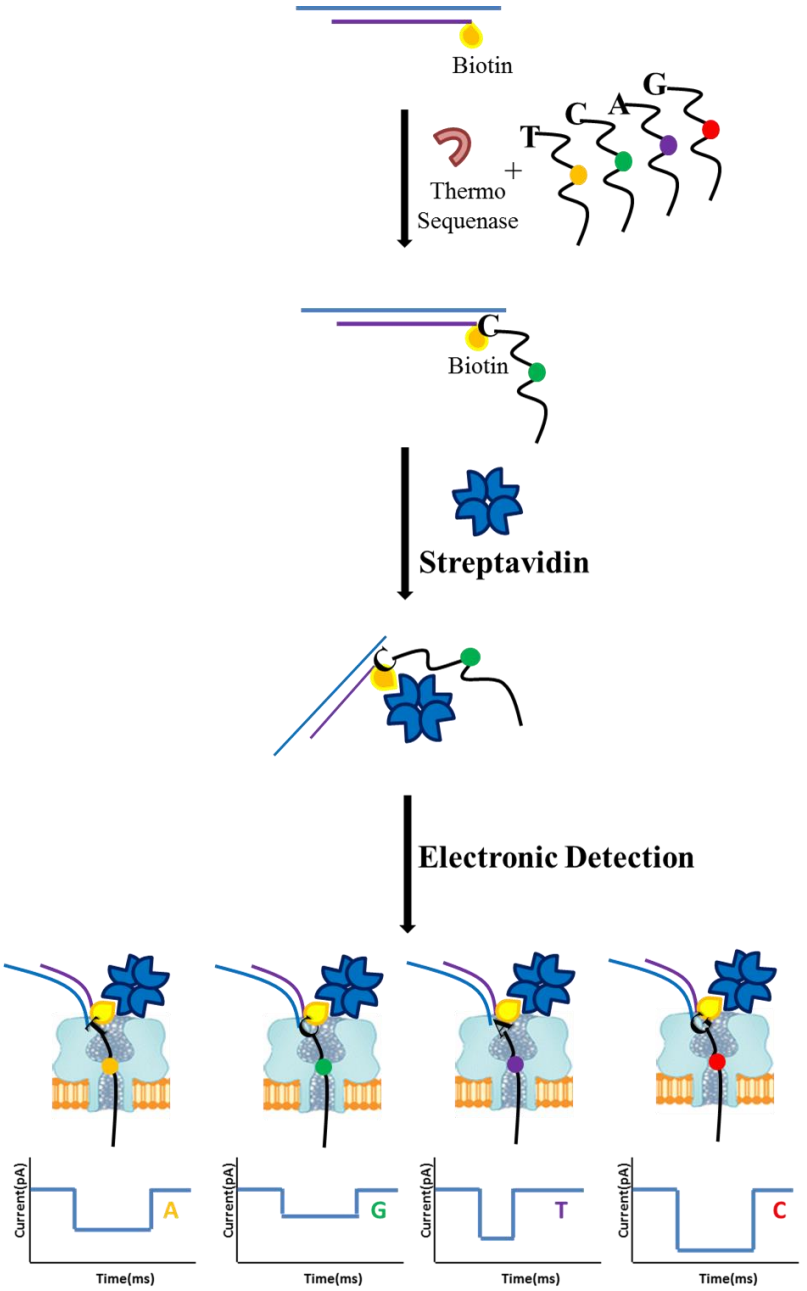


Figure 6.1 Scheme of the single molecule electronic SNP assay. SBE reactions were performed with primers having biotinylated nucleotide at the 3' end, templates, Thermo Sequenase and tagged ddNTPs. After completion of the reaction, they were subsequently treated with streptavidin and then applied to the nanopore detection system. The ensuing current modulations were measured to identify the incorporated nucleotides on the primer.

6.2 Methods

6.2.1 Generation of ssDNA from PCR products by λ exonuclease

PCR reactions were performed using a phosphorylated primer for the unwanted strand, a regular 5' hydroxyl primer for the desired strand, Advantage® 2 polymerase mix (Clonetech) and human DNA samples. Sequences of the primers are as follows: F: 5'-GGTGATTTCCAAGAAGCCA CC-3' R: 5'-/Phos/GATAGGGTCAGTGACATGGAATCCTG-3' (1173 G>A), F: 5'-GGTTTC ACCATG TTGGCCAG-3' R: 5'-/Phos/GGAAGTCAAGCAAGAGAAGACCTG-3' (-1639 C>T). After purifying amplicons with the QIAquick PCR Purification Kit (Qiagen), they were digested with 10 units λ exonuclease (NEB) at 37 °C for an hour. The reactions were purified using ssDNA/RNA Clean & Concentrator™ (Zymo Research) and employed as templates in the SBE reactions.

6.2.2 SBE reactions and nanopore detection

The SBE reactions were performed with 10 pmol biotinylated primer, 15 pmol complementary template, 1 unit Thermo Sequenase, 1X reaction buffer and 100 pmole each tagged ddNTP in a 20 µl total reaction volume. They were incubated at 65 °C for an hour after 15 seconds of an initial denaturation step at 95 °C and then analyzed on a 15% Criterion™ TBE-Urea polyacrylamide gel. After confirming the presence of the products, a quarter of the reaction was applied to the nanopore system after incubation with 1 nmol of streptavidin (Sigma). The sequences of biotinylated primer and synthetic template are as follows: 5'-GTTCTGCTTCCTG C/Biotin-dT/-3' (Figure 6.2) and 5'-GGANAGCAGGAAGCAGAACCC-3' (N = A, T, C and G). The sequences of biotinylated primers for VKORC1 genes are as follows: 5'-AAACAACCATT GGC/Biotin-dC/-3' for -1639 C>T and 5'-CAGGAGATCAT CGA/Biotin-dC/-3' for 1173 G>A.

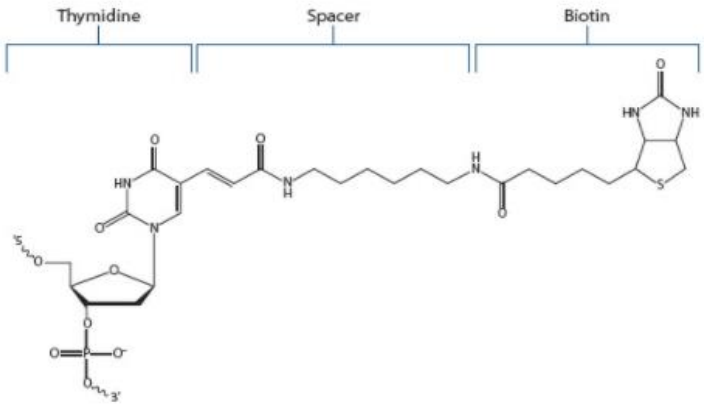


Figure 6.2 Structure of biotin-modified thymidine residue as an example.[124]

6.2.3 Data analysis

After compiling blockade events from each experiment, histograms and box-and-whisker plots were generated in Excel, and Graphpad Prism software was used to fit them to a normal distribution. Statistical significance was determined with a one way ANOVA for comparisons among groups with significance level $p < 0.05$ and individual comparisons were determined using a Bonferroni post-test correction.

6.3 Results and Discussion

6.3.1 Experiments with synthetic DNA templates

First, four SBE reactions with primer, specific template, DNA polymerase and a matching nucleotide were performed to establish a reference level for each tag. Before applying the products to the detection system, the presence of extended products was confirmed by urea-PAGE analysis along with primer, template and nucleotide controls. As shown in Figure 6.3, all four reactions contained an extra band of the expected size confirming that primer was extended

by each of the tagged ddNTPs in these reactions. There was almost no primer left in any of the reactions, indicating their high efficiency. After checking the size of products, each reaction was treated with streptavidin and then applied to the nanopore system for current determination. As in the previous experiments, a trapezoidal pulse was generated by slowly ramping potential to 180 mV for 2 seconds followed by a plateau of 6 seconds and returning to the ground level for 2 seconds at the end. In all four signals, there is an obvious current drop during the 6 seconds of the plateau period, indicating that the tag on the extended primer is stably immobilized in the pore by streptavidin (Figure 6.4). After the plateau period, voltage begins ramping back to baseline and thus the current drops further with a minor increase at the end due to the tag molecule being released from the pore. Specifically, the G nucleotide tagged with 17 adjacent dSp units decreased current flow the least ($\Delta I/I_{\circ}$ 0.74 ± 0.03), producing the highest level. The C nucleotide with 8 dSps produced the next highest current level ($\Delta I/I_{\circ}$ 0.58 ± 0.02). Events elicited by primer extended by the T nucleotide having 3 dSps produced $\Delta I/I_{\circ}$ 0.35 ± 0.03 and those extended by the A nucleotide containing 6 neutral Imps generated $\Delta I/I_{\circ}$ 0.28 ± 0.02 on average. The collected events were quantified via a box-and-whisker plot and a histogram and the results show that the events corresponding to each nucleotide cluster in well-separated regions (Figure 6.4E and 6.4F respectively). We also examined two negative control reactions: one with only biotinylated primer, matched template and DNA polymerase but without any nucleotide (Figure 6.5) and another with all reactants in the absence of enzyme. Neither showed any measurable blockade events after streptavidin treatment most likely due to the relatively short length of the primer and thus it would be difficult to stably arrest the unextended primer by streptavidin in the pore. The absence of detectable events from these control reactions verifies that the ones collected in the previous experiments resulted from the tags on the extended

primers. Overall, these results demonstrate that each tag on a specific nucleotide produces a distinct, well-separated level.

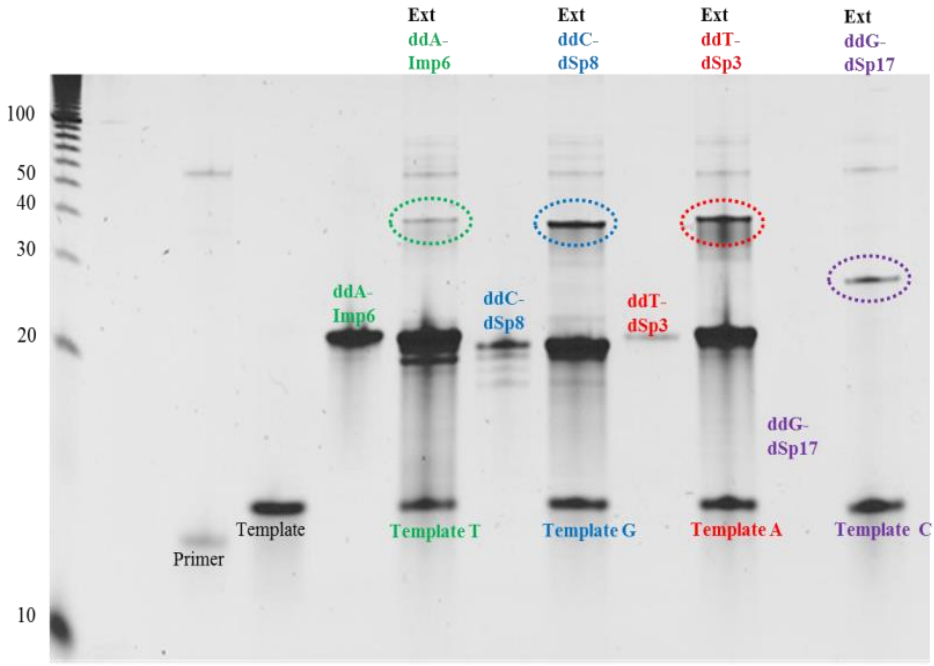


Figure 6.3 SBE results with biotinylated primer, matched template, single tagged ddNTP and Thermo Sequenase on a denaturing polyacrylamide gel. The first two lanes indicate the position of primer and template. The following paired lanes show the size of each nucleotide and reaction results of SBE using the same nucleotides. Due to the absence of bases in dSpacers (dSp), ddG-dSp17 by itself is not visible with DNA staining. The circled bands indicate products of each reaction.

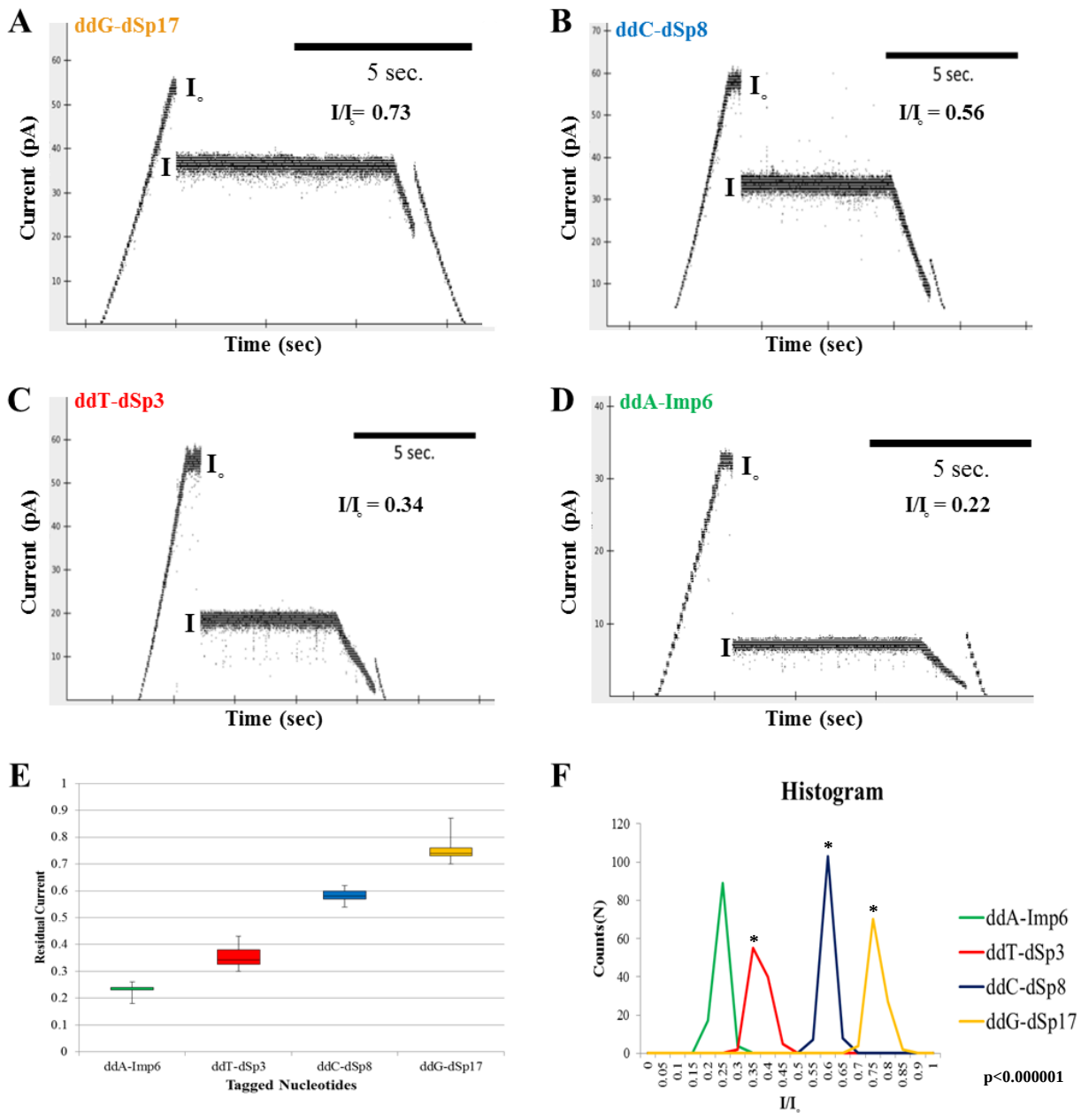


Figure 6.4 Current signatures of four tagged nucleotides in reactions with specific template and matched nucleotide at 180 mV. **A.** Detected current signature of the SBE reaction with biotinylated primer, template C, Thermo Sequenase and tagged ddGTP. **B.** Current signature of the reaction with biotinylated primer, Template G, Thermo Sequenase and tagged ddCTP. **C.** Current signature of the reaction with biotinylated primer, template A, Thermo Sequenase and tagged ddUTP. **D.** Current signature of the reaction with biotinylated primer, Template T, Thermo Sequenase and tagged ddATP. **E.** A box-and-whisker plot of current levels of the four reactions. **F.** A histogram of counted events from these experiments. The differences among them were statistically significant.

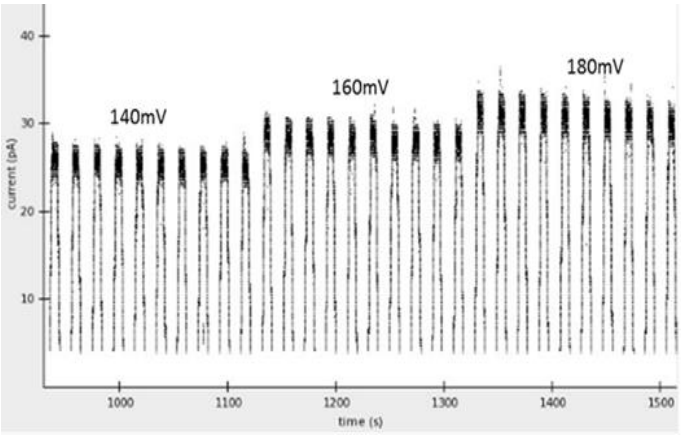


Figure 6.5 Current trace from the negative control reaction with the biotinylated primer, template, DNA polymerase but no nucleotide.

With these levels as guidance, the specificity of the SBE reaction was confirmed by performing reactions with a mixture of tagged ddNTPs, one specific template, biotinylated primer and Thermo Sequenase. Considering that most SNPs are biallelic, we first examined the four SBE reactions with specific template and a mixture of two different tagged ddNTPs. Figure 6.6 shows distinguishable current signatures from each reaction under a constant 160 mV. First, the reaction with template C and a mixture of tagged ddGTP and tagged ddATP produced signals that correspond to the level previously observed by tagged ddGTP ($\Delta I/I_{\circ}$ 0.72 ± 0.02) (Fig. 6.6A) while the reaction with template T and the same mixture of tagged nucleotides generated signatures that are consistent with those detected by tagged ddATP ($\Delta I/I_{\circ}$ 0.28 ± 0.02) (Fig.6.6D). The specificity for template T (Fig. 6.6B) and template A (Fig. 6.6C) were also verified in reactions with a mixture of tagged ddCTP and tagged ddTTP. A box-and-whisker plot and a histogram of the collected events from each experiment are shown in Figure 6.6E and 6.6F respectively. The presence of products in these reactions was also confirmed on a denaturing urea gel (Figure 6.7). We also analyzed the specificity of the reactions containing a mixture of all four tagged ddTNPs and one specific template (Figure 6.8). In these reactions which contain all

four nucleotides, minor non-specific events were observed, notably in the reaction with template C. Nevertheless, the major blockade events in each reaction correlate clearly with the level elicited by the nucleotide complementary to the added template (Figure 6.8). Overall, both results demonstrate that template-specific products were produced under the four different conditions, that Thermo Sequenase exhibits high specificity with our polymer tagged ddNTP analogs, and that these levels are clearly distinguishable in the nanopore system.

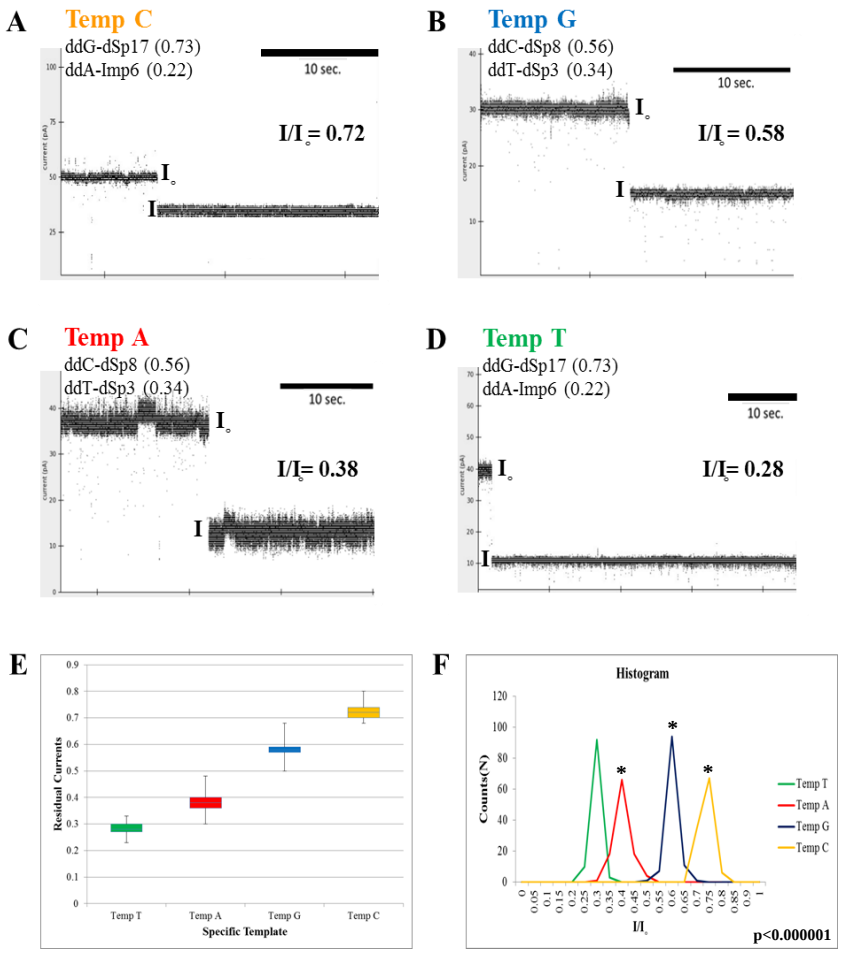


Figure 6.6 Current signatures of four tagged ddNTPs in reactions with a mixture of two nucleotides and single template at 180 mV. **A.** Detected current signature of the reaction with biotinylated primer, template C, Thermo Sequenase and a mixture of tagged ddGTP and tagged ddATP. **B.** Current signature of the reaction with biotinylated primer, template G, Thermo Sequenase and a mixture of tagged ddCTP and tagged ddUTP. **C.** Current signature of the reaction with biotinylated primer, template A, Thermo Sequenase and a mixture of tagged ddUTP and tagged ddGTP. **D.** Current signature of the reaction with biotinylated primer, template T, Thermo Sequenase and a mixture of tagged ddCTP and tagged ddUTP. **E.** A summary of the apparent current levels of the four reactions. **F.** A histogram of counted events from these experiments. The differences among them were statistically significant.

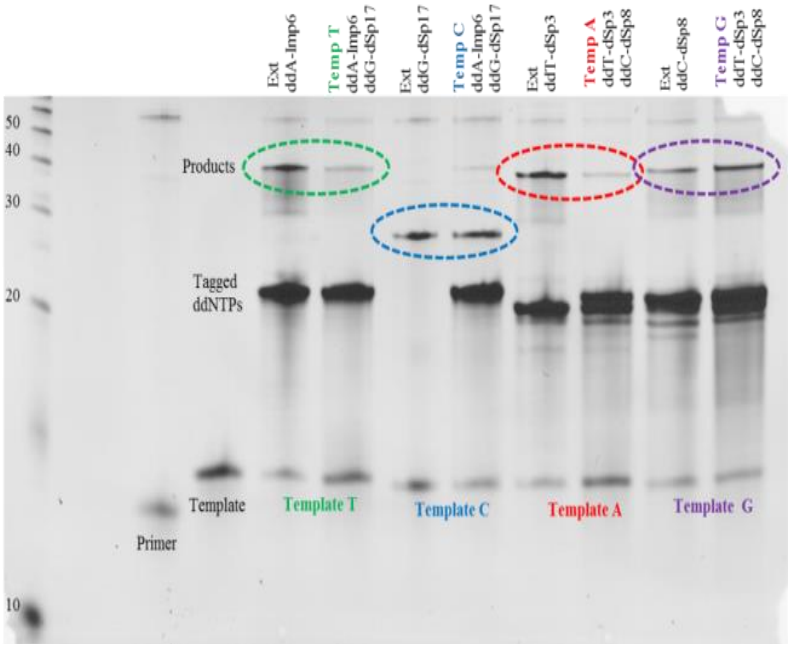


Figure 6.7 SBE reaction results with biotinylated primer, one specific template and a mixture of two tagged ddNTPs on a denaturing polyacrylamide gel. Each reaction is compared with the products from reactions with single matched nucleotides.

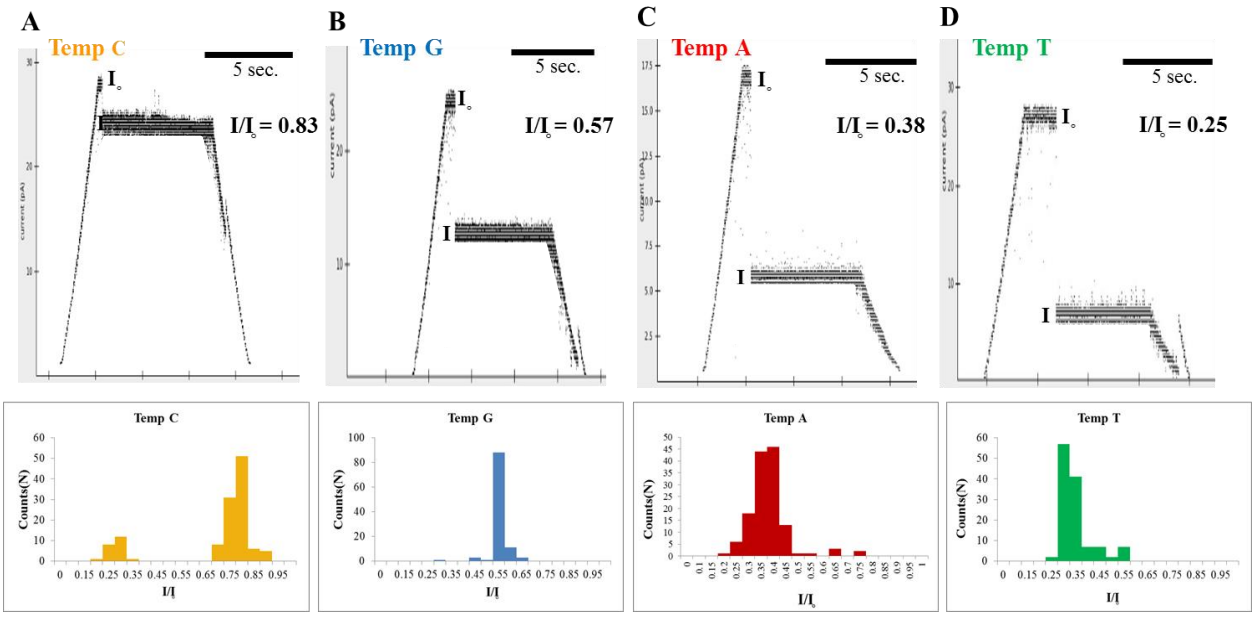


Figure 6.8 Current signatures of four tagged ddNTPs in reactions with a mixture of four nucleotides and single template and corresponding histograms below. **A.** Detected current signature of the reaction with biotinylated primer, template C, Thermo Sequenase and a mixture of four tagged ddNTPs. **B.** Current signature of the reaction with biotinylated primer, template G, Thermo Sequenase and a mixture of four tagged ddNTPs. **C.** Current signature of the reaction with biotinylated primer, template A, Thermo Sequenase and a mixture of four tagged ddNTPs. **D.** Current signature of the reaction with biotinylated primer, template T, Thermo Sequenase and a mixture of four tagged ddNTPs. The differences among them were statistically significant.

Besides the specificity of the assay, another critical aspect of the genotyping is the ability to discern heterozygous from homozygous alleles. To evaluate this feature of our assay, we also performed the reactions with a mixture of two different templates and corresponding tagged nucleotides and analyzed them in the nanopore system. As mentioned earlier, in this assay design, an event is produced by stochastically driving and threading a streptavidin-bound tag molecule into the pore and after detection, the molecule is diffused again into the solution as the polarity of voltage is reversed at the end of each cycle. Thus, by repeatedly generating trapezoidal pulses, multiple independent events are obtained from a single pore. In this setting, it is therefore possible to detect two distinct current signatures generated by two different extended products from a single pore, provided that there were two different templates in the reaction such as would be the case for heterozygous alleles. We tested this by performing SBE with a mixture of two templates. Figures 6.9A and 6.9C show current traces that exhibit two unique levels in a single pore confirming the presence of two products created in these reactions. The histograms of the collected events display two distinct populations of current levels confirming that there were two extension products generated for each allele in these reactions (Figure 6.9B and 6.9D). Although the templates were mixed in a 1:1 ratio, there was a difference in frequencies observed in each case. This might be due to differences in kinetics of threading each tag into the pore due to their different chemical structures and properties. Or the enzyme's binding affinity for each substrate could vary, producing one product slightly more often than the other. Nevertheless, the results clearly show two clusters of blockade levels, demonstrating the presence of two different templates in these reactions.

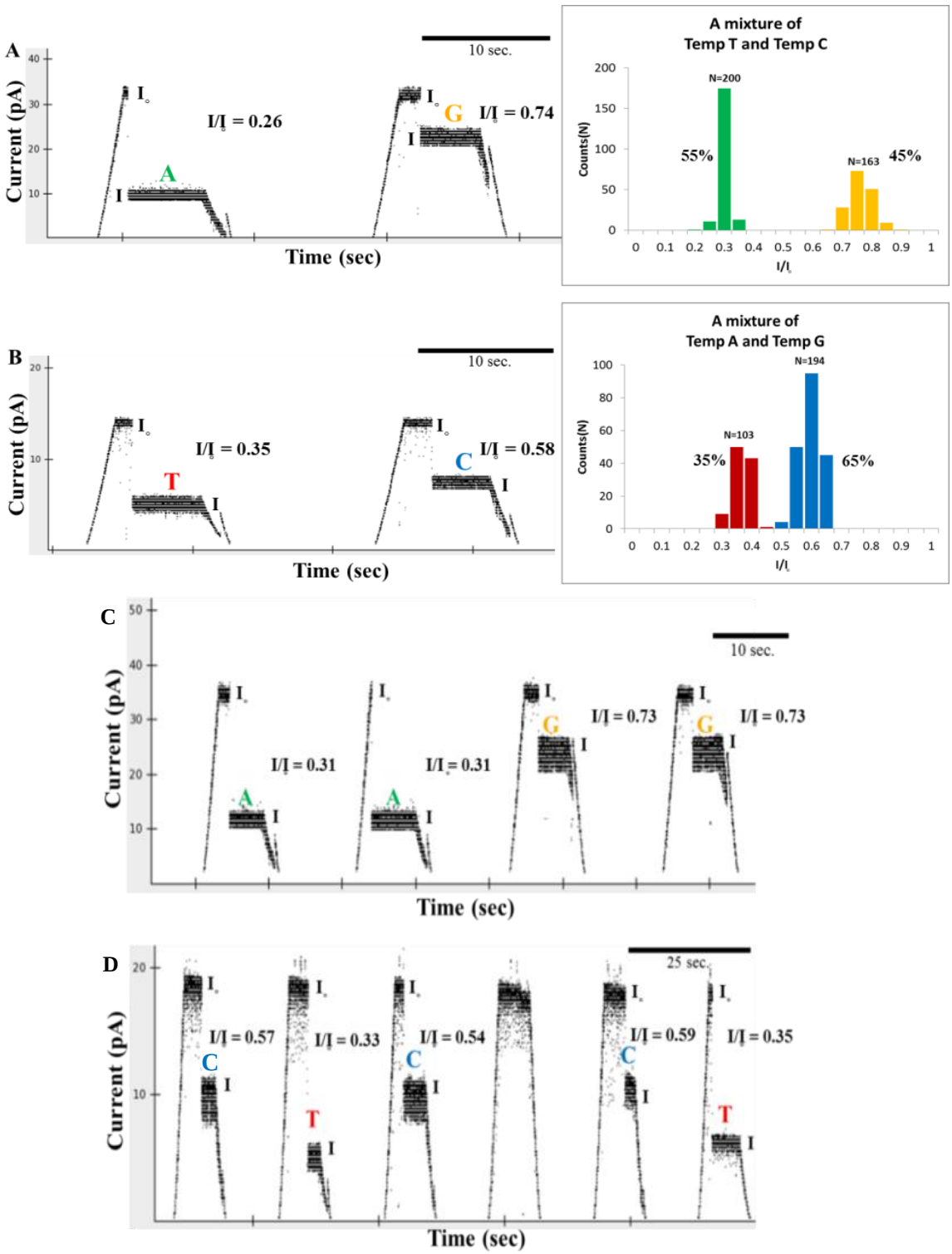


Figure 6.9 Current signatures of two products in reactions with a mixture of two templates from a single pore. **A.** Detected current signature and a histogram of counted events from the reaction with biotinylated primer, a mixture of template T and template C, Thermo Sequenase and a mixture of tagged ddATP and tagged ddGTP. **B.** Current signature and a histogram of counted events from the reaction with biotinylated primer, a mixture of template G and template A, Thermo Sequenase and a mixture of tagged ddCTP and tagged ddTTP. **C&D.** Additional examples of current traces from the two reactions.

6.3.2 Genotyping at two polymorphic sites in VKORC1 using human genomic DNA templates

Lastly, our assay was also validated by detecting different genotypes in actual human DNA samples. As mentioned earlier, warfarin, the most commonly prescribed anticoagulant drug, has a narrow therapeutic window and exhibits wide inter-individual variation in dose requirements.[120] Together with other non-genetic factors, increasing evidence has shown that two variants, 1173 G>A and -1639 C>T, in the VKORC1 gene appear to be the most important genetic determinants for warfarin dosing.[120-123] Thus, as a model system, human DNA samples bearing different genotypes at these two sites: homozygous G, homozygous A and heterozygous G/A at the 1173 G >A site and homozygous C, homozygous T and heterozygous C/T at the -1639 C >T were tested in our assay. After confirming their genotypes by Sanger sequencing of the PCR amplified locus (Figure 6.10), samples were processed to generate ssDNA template by a PCR reaction with a 5' phosphorylated primer for the unwanted strand. After the amplification, the phosphorylated strand was digested by λ exonuclease, which is specific for 5' phosphorylated DNA (Figure 6.11).[97]

Using the resulting ssDNA as template, SBE reactions were performed with tagged ddNTPs and the presence of products was confirmed along with controls by urea-PAGE analysis (Figure 6.12). Each reaction was examined together with two controls: double stranded PCR products and λ exonuclease treated products. Two bands were found in the λ exonuclease digested product lanes; the upper band indicates undigested and thus residual dsDNA and the bottom band contains the digested ssDNA, which migrates slightly faster than the original dsDNA. The last lane in each set is the reaction result from the SBE reactions and the circled bands indicate the extended products. The right panels show the same gel before DNA staining

(Figure 6.12B and D). All our tags except the ddG-dSp17 contain Cy3 dye. Therefore, the gels before staining show the two Cy3 containing-molecules: the residual tagged nucleotides and extended products. These results confirm again that the primer was extended with the expected molecules, the tagged nucleotides. After incubation with streptavidin, each reaction was applied to the detection system and blockade events were examined. Negative control reactions with biotinylated primer, heterozygous templates and a mixture of two tagged ddNTPs but without enzyme were included in the last lane on the gel, demonstrating the absence of products. These control reactions were analyzed in the system as well and any events detected from them were considered background captures.

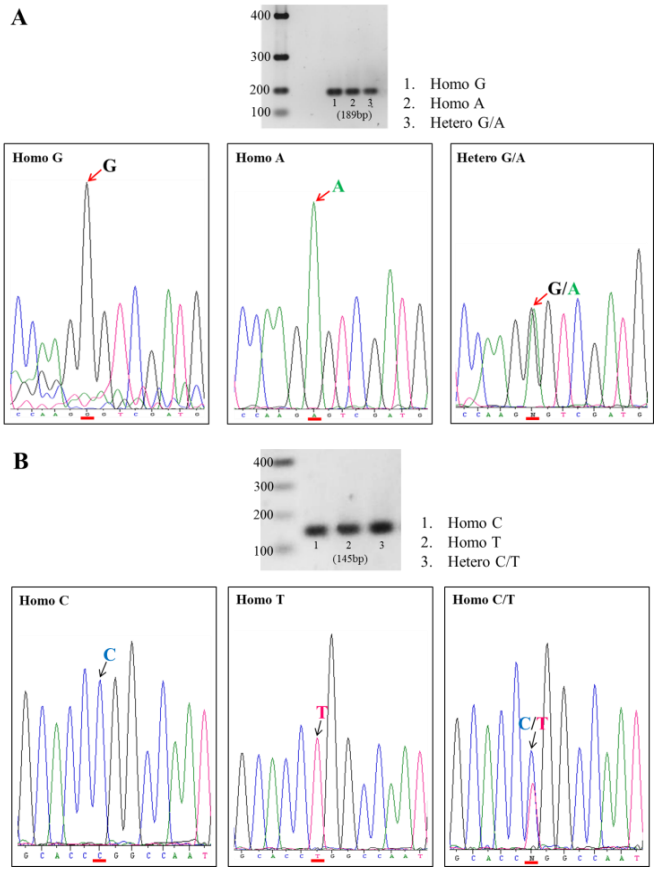


Figure 6.10 PCR amplification and Sanger sequencing results for two sites in VKORC1 gene. **A.** PCR amplification and Sanger sequencing result for VKORC1 gene including 1173 G>A (rs9934438) site. **B.** PCR amplification and Sanger sequencing result of VKORC1 gene including -1693 C>T (rs9923231) site.

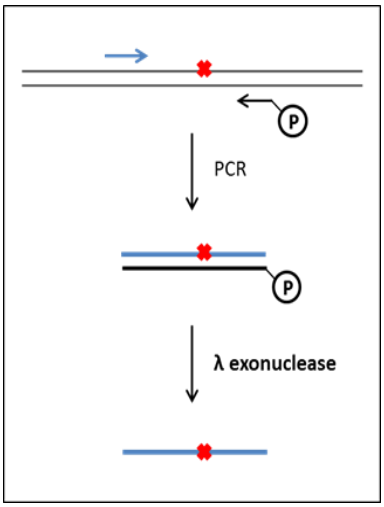


Figure 6.11 A scheme of ssDNA template preparation.

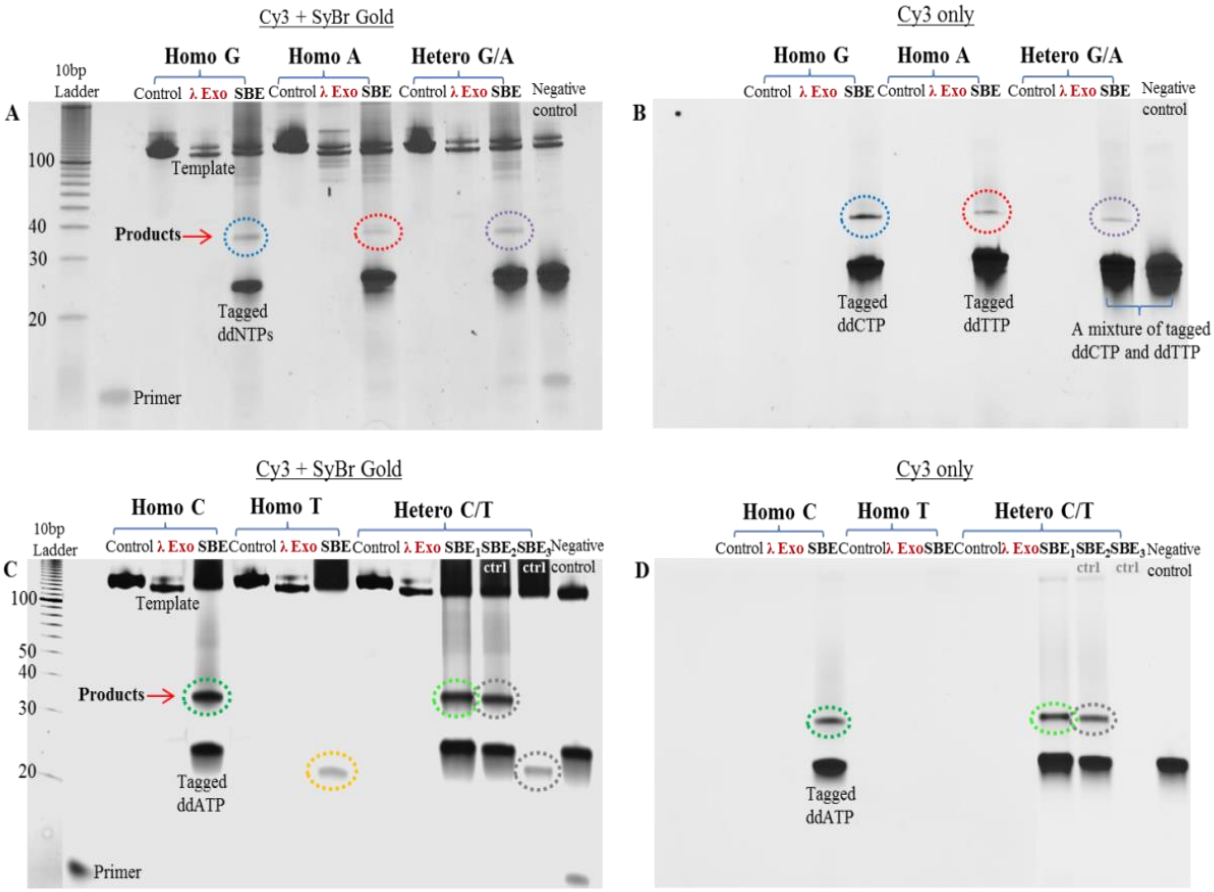


Figure 6.12 SBE reaction results for various VKORC1 genotypes assessed by gel electrophoresis. **A.** SBE reaction results with three –1639 C >T (rs9923231) genotypes with DNA staining. **B.** The same gel without any staining. **C.** SBE reaction results with three 1173G >A (rs9934438) genotypes with DNA staining. **D.** The same gel without any staining.

For the analysis of these experiments, each trapezoidal pulse was treated as an independent event and the four reactions (the three experimental reactions with different genotypes and one negative control) were normalized based on the total number of collected events. Specifically, 4060 measurements including trapezoids without any blockade events were compiled from each experiment for the 1173 site (rs9934438) and a sample size of 5346 was used for the analysis of the -1639 C >T site (rs9923231). As shown in Figure 6.13 and Figure 6.14, the events from homozygous alleles exhibited uniform blockade levels: the homozygous G and homozygote A produced $\Delta I/I_{\circ}$ 0.52 ± 0.09 and $\Delta I/I_{\circ}$ 0.35 ± 0.08 whereas homozygous C and homozygous T generated $\Delta I/I_{\circ}$ 0.63 ± 0.1 and $\Delta I/I_{\circ}$ 0.3 ± 0.07, respectively. On the other hand, mixtures of two signals that are in line with the levels found for homozygous samples were detected from the reactions with heterozygous alleles. These levels also coincide with those obtained for each product in the reactions with synthetic DNA templates. Figure 6.15 shows the histograms of events collected from homozygous and heterozygous alleles. While the four experiments with the homozygotes exhibit unimodal distribution, those with heterozygotes display bimodal distributions with two distinct peaks, demonstrating again the presence of two different extension products. Box-and-whisker plots are also provided in Figure 6.15C and 6.15D.

As mentioned earlier, the events from the negative control reaction without enzyme were also analyzed to estimate the extent of the background in the experimental reactions. Figure 6.16A shows the number of blockade events in the normalized samples; a sample size of 4060 and 5346 measurements were gathered from each reaction for the 1173 G >A and the -1649 C >T sites, respectively. The rate was calculated by dividing the number of capture events by the sample size for each site (Figure 6.16B). The relatively minor rates in the negative controls highlight that the background events are insignificant in these reactions, indicating that a

majority of the events were elicited as a result of the enzymatic reaction. The differences across samples were tested by ANOVA followed by a Bonferroni post t-test correction and all tests were statistically significant. Overall, data in these experiments validate that our approach can be employed for diagnostic screening of SNPs in human genes.

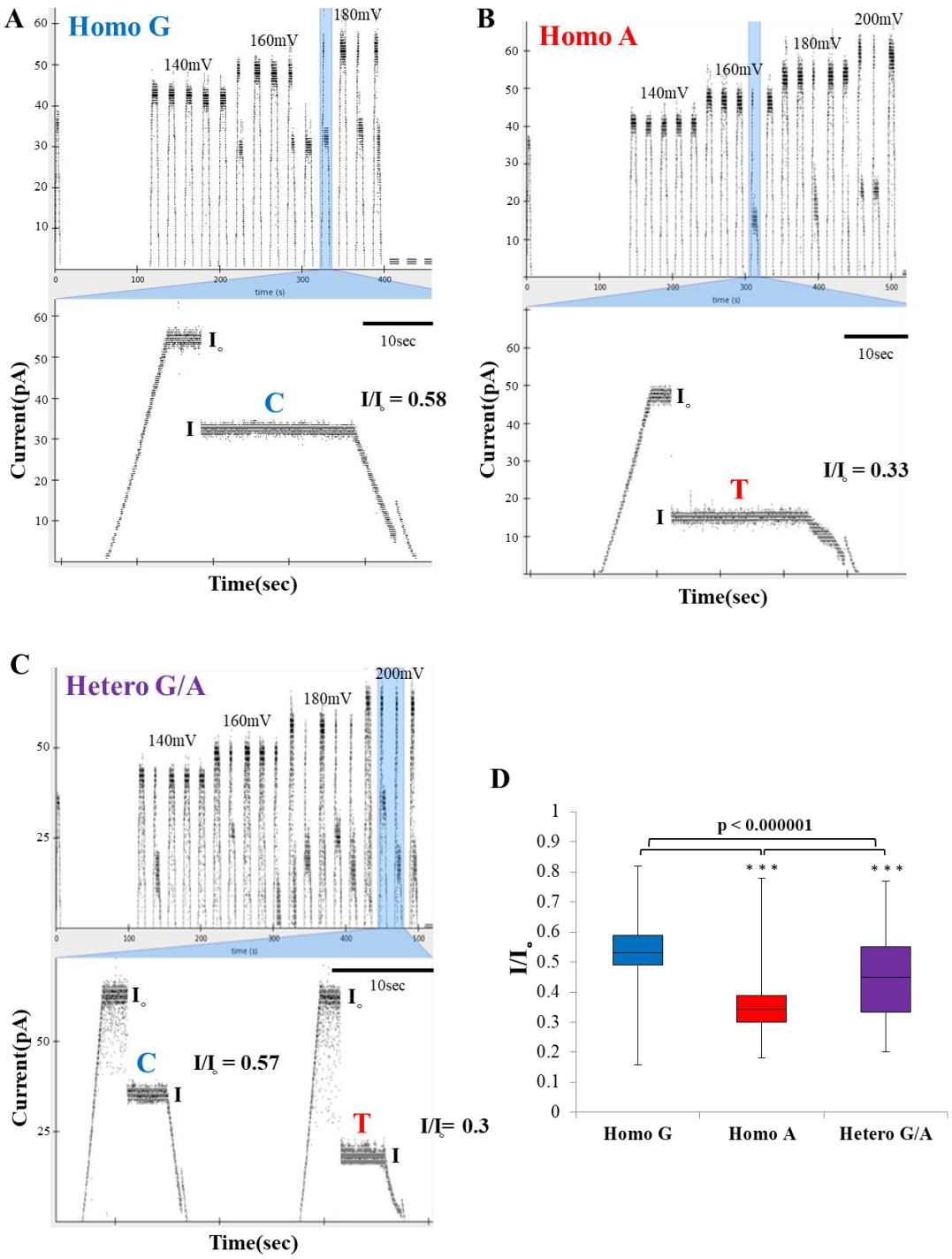


Figure 6.13 Current signatures of products in reactions with different genotypes at VKORC1 1173 G>A. **A.** Detected current signatures of the reaction with biotinylated primer, homozygote G allele, Thermo Sequenase and tagged ddCTP. **B.** Current signatures of the reaction with biotinylated primer, homozygote A allele, Thermo Sequenase and tagged ddTTP. **C.** Current signatures of the reaction with biotinylated primer, heterozygote G/A alleles, Thermo Sequenase and a mixture of tagged ddCTP and tagged ddUTP. **D.** Box-whisker plots of different genotypes. The differences among them were statistically significant.

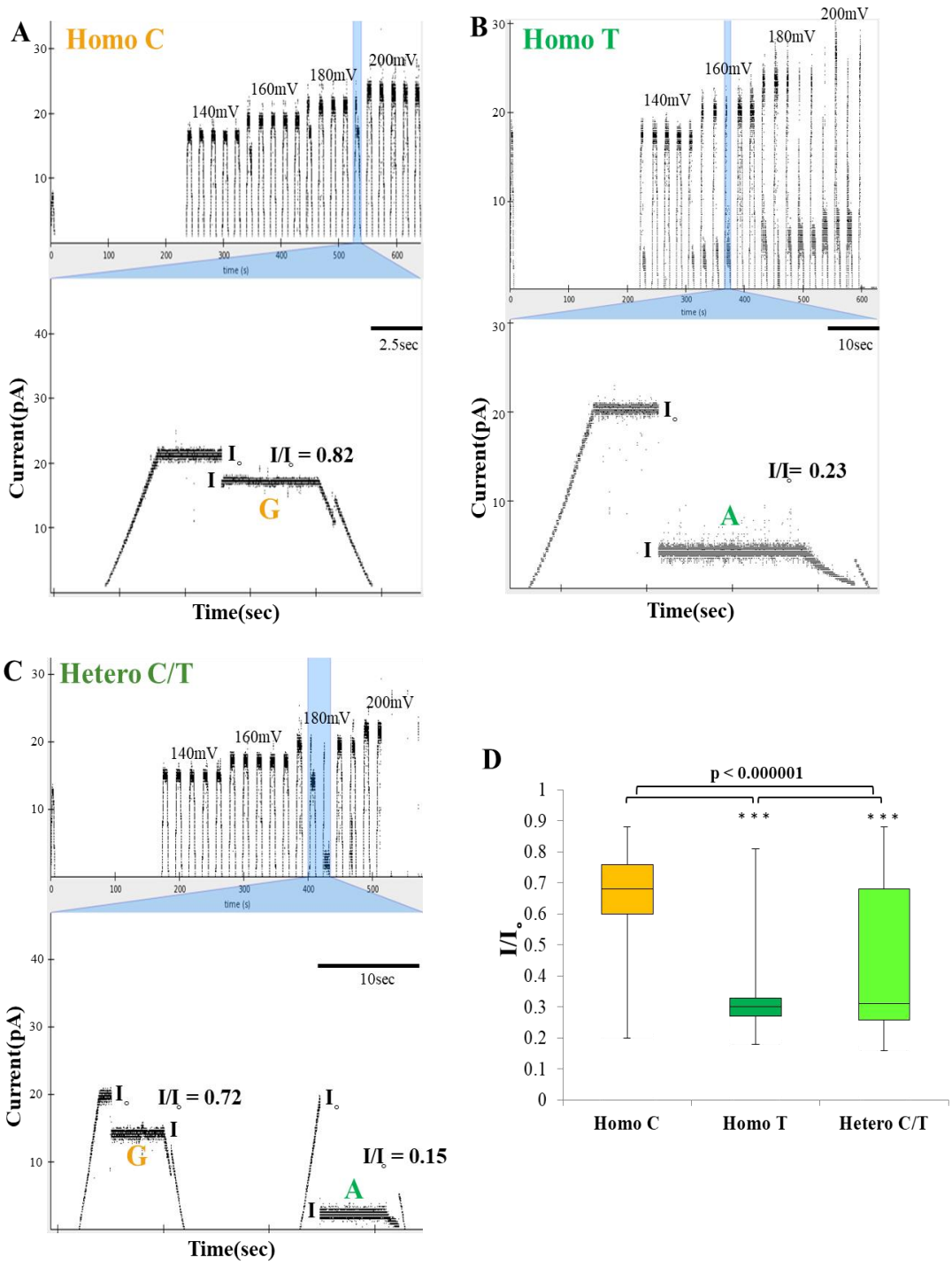


Figure 6.14 Current signatures of products in reactions with different genotypes at VKORC1 -1639 C>T. **A.** Detected current signatures of the reaction with biotinylated primer, homozygote T allele, Thermo Sequenase and tagged ddATP. **B.** Current signatures of the reaction with biotinylated primer, homozygote C allele, Thermo Sequenase and tagged ddGTP. **C.** Current signatures of the reaction with biotinylated primer, heterozygote C/T alleles, Thermo Sequenase and a mixture of tagged ddATP and tagged ddGTP. **D.** Box-whisker plots of different genotypes. The differences among them were statistically significant.

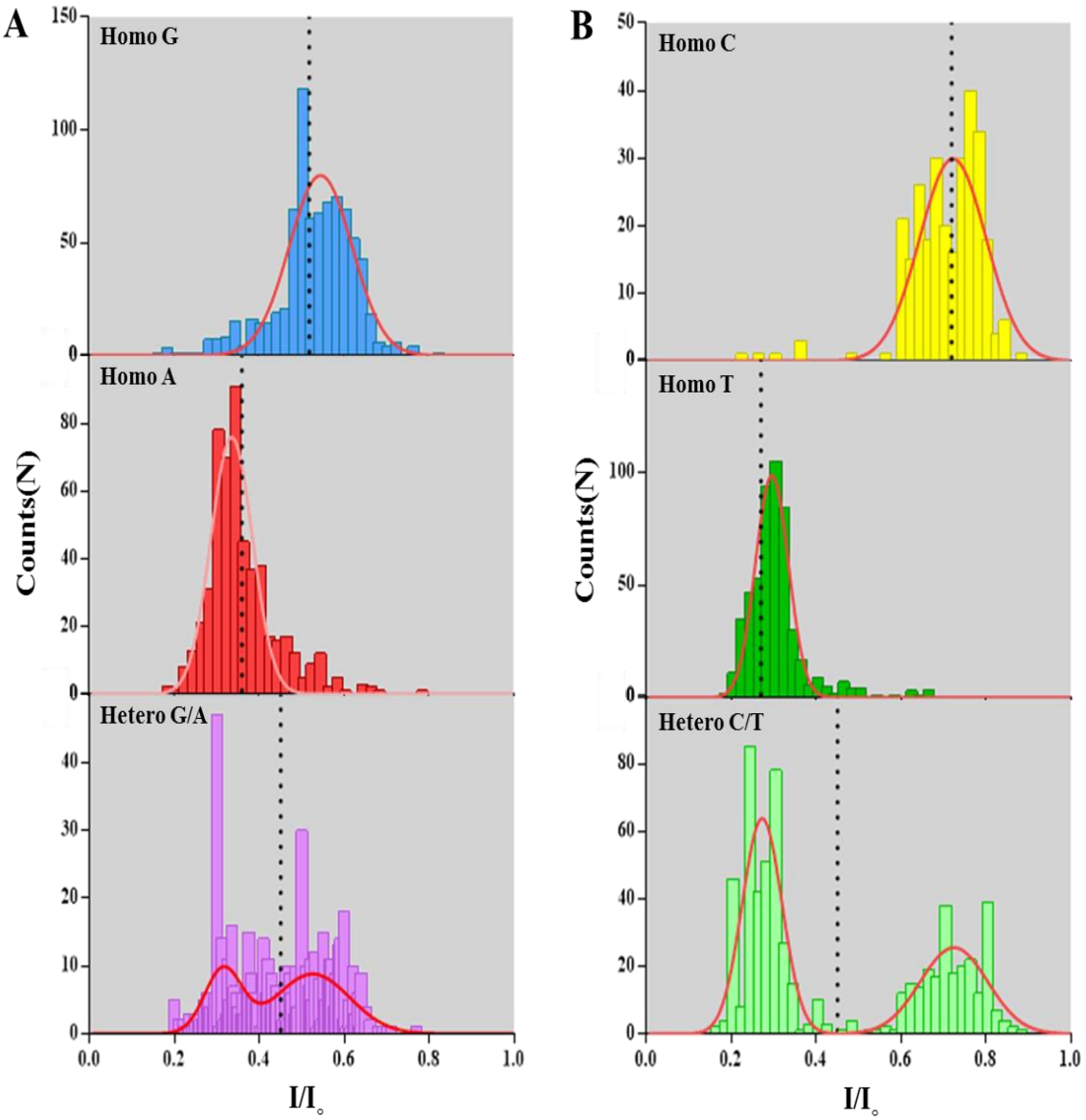


Figure 6.15 Histograms generated from the results exemplified in Figure 6.13 and Figure 6.14. Red line indicates a normal distribution fitted to the data and the black dotted line is the mean of each data set. **A.** Histograms of the collected events from three reactions (three different genotypes) for VKORC1 1173 G>A sites. **B.** Histograms of the collected events from three reactions (three different genotypes) for three different genotypes at VKORC1 -1639 C>T site.

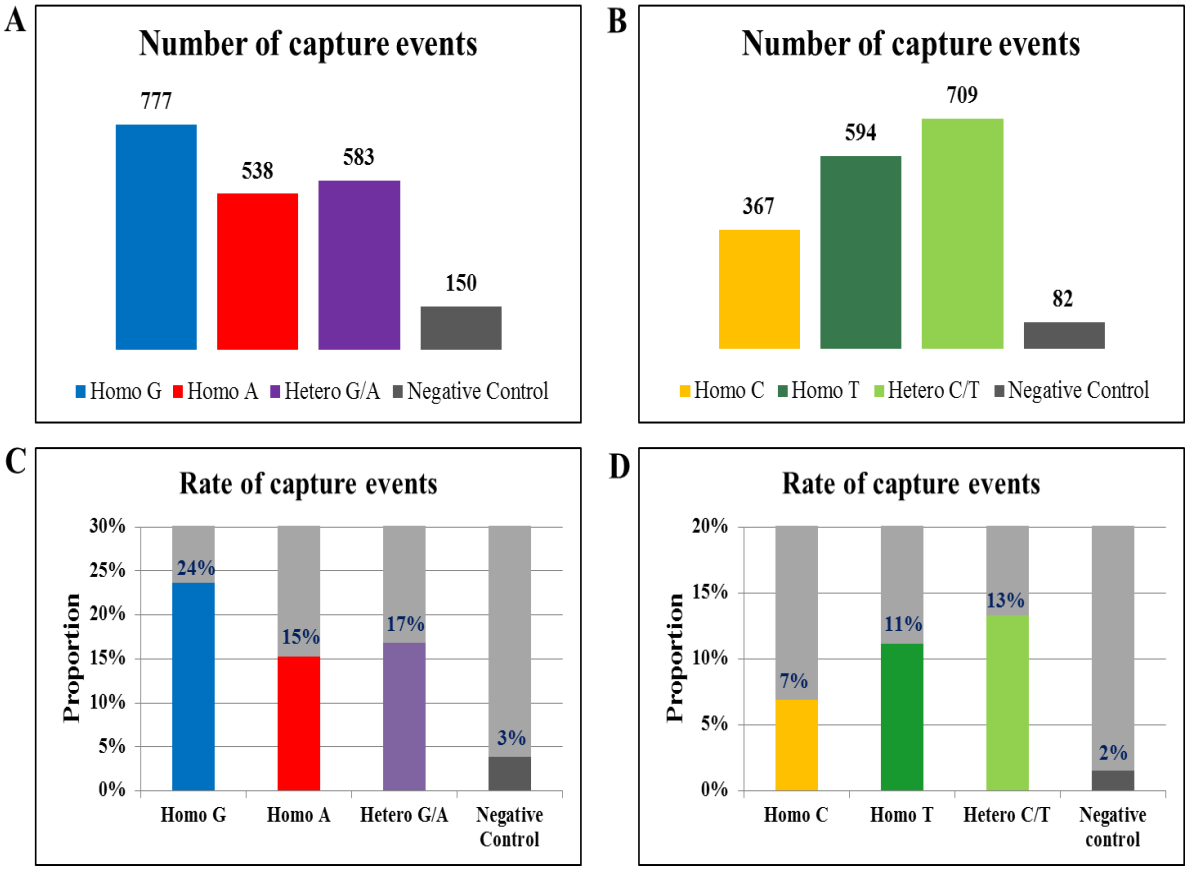


Figure 6.16 The number of blockade events and the percentages of capture events in normalized data sets **A.** The number of capture events analyzing the VKORC1 1173 G>A site. **B.** The number of events examining the VKORC1 -1639 C>T site. **C&D.** The percentages of capture events in each data set.

6.4 Conclusion

To improve the efficiency of the previously described integrated approach (Chapter 5), here we developed another assay that is based on the immobilization of a biotinylated extended primer with a streptavidin molecule. Unlike the former approach, this method first completes the enzymatic reaction in solution followed by incubation with streptavidin and subsequent application to the detection system. Since the reaction rate is not limited by substrate concentrations in this approach, we could achieve very efficient SBE reactions, allowing us to collect a sufficient number of blockade events from each reaction containing different templates for statistical analysis. After establishing the reference level for each tag, we also tested the specificity of our SBE reactions with Thermo Sequenase in the presence of the specific template and a mixture of tagged ddNTPs. Our data demonstrated that template specific products were produced in each reaction and could be clearly differentiated in our nanopore detection system. In addition, the assay's ability to discriminate homozygous and heterozygous alleles was examined; different alleles produced distinct genotype-specific histogram patterns. As a final validation of the assay, we examined two polymorphic sites in the VKORC1 gene using human genomic DNA samples and obtained well-separated genotype-specific levels. Taken together, these data not only prove the concept of our design for base labeled ddNTPs for nanopore detection but also supports its potential to be a nanopore-based companion diagnostic using these nucleotides. By simplifying SNP analysis with such a simple electronic readout that requires small sample volume and low detection concentration, the cost for detecting SNPs could be reduced and this might allow better, faster and cheaper routine screening for SNPs in patients, ultimately realizing an important goal of precision medicine.

Conclusion and Future Directions

In this thesis, we demonstrated a novel single molecule electronic SNP assay using four nucleotide analogs and a nanopore detection system. Specifically, we employed SBE of a primer with ddNTPs containing unique oligonucleotide polymers attached to their bases for allele discrimination, and ionic current readout due to capture of these tags in αHL nanopores for allele detection. To realize this concept, we designed and synthesized four polymer-attached ddNTPs with tags that produce characteristic electrical signatures in the nanopore system and confirmed that they are active substrates for Thermo Sequenase DNA polymerase. Two different assay models were tested to stably immobilize a tag molecule on the extended primer inside the pore; an integrated approach using ssDNA primer-conjugated αHL nanopores and another two-step assay employing a biotin-streptavidin interaction. As discussed earlier, the first approach revealed slow SBE reaction rate due to issues of diffusion and the primer concentration constraint. To overcome this, in the future, strategies to enhance the electrophoretic concentration of the enzyme using mutagenesis or lipophilic concentration of the enzyme and template via a hydrophobic moiety such as cholesterol could help congregate them in the vicinity of the immobilized substrate, the membrane-embedded primer-pore conjugate.

As an alternative approach, we developed an assay using a biotin-streptavidin interaction in which the enzymatic allele descrimation step is performed in advance of the detection step. In this approach, we could achieve very efficient SBE reactions, allowing us to examine the essential properties of the SNP assay using our tagged ddNTPs in the nanopore system. The results demonstrated that our method is an accurate and reliable approach for genotyping analysis that has the potential to be implemented for real-life diagnostic applications. Typically, screening patients or disease detection involves analysis of a panel of biomarkers. As described

in Chapter 1, our nanopore chip is composed of a large number of individually addressable sensors. Therefore, multiple SNP sites could be simultaneously analyzed in a single cycle of the assay by applying different samples over separate areas of the chip.

Furthermore, additional dimensions could be introduced to the output by including another identifying moiety in the tag design. For instance, two different blocks of identifiers can be arranged in the tag and they can be sequentially detected as it traverses the pore during the capture event (Figure A1). In an immobilized setting, the rear identifier in the tag serves as a barcode for the encoded reaction and registers its brief signature first as it crosses the constriction zone. This is followed by a long-lived blockade event that is produced by the former identifier as it is stably confined inside the pore, revealing the identity of the incorporated base as demonstrated throughout this thesis. The combination of two identifying moieties generates a unique pattern of current signatures and this could allow simultaneous detection of multiple SNPs in a single nanopore. Previously, it has been demonstrated that PEG molecules of different lengths can characteristically block the pore current.[93] Thus, a series of barcode probes can be generated to encode different reactions using these molecules. In this study, we constructed 4 distinct identifiers that generate well-separated current signatures in the nanopore system. Using just these four, at least 16 combinatorial libraries of tags consisting of two uniquely ordered molecules can be generated. In addition, previous studies have demonstrated that the resolution of αHL can be improved by mutating AA residues in the barrel.[67,92] The clearer resolution of the new pore might provide better separation of different tag molecules and this could allow additional current levels by other compounds to be discovered. These advances, combined with an array of hundreds of thousands of nanopores should allow parallel analysis of a biomarker panel, eventually realizing a cost-effective multiplexed point-of-care SNP diagnostic.

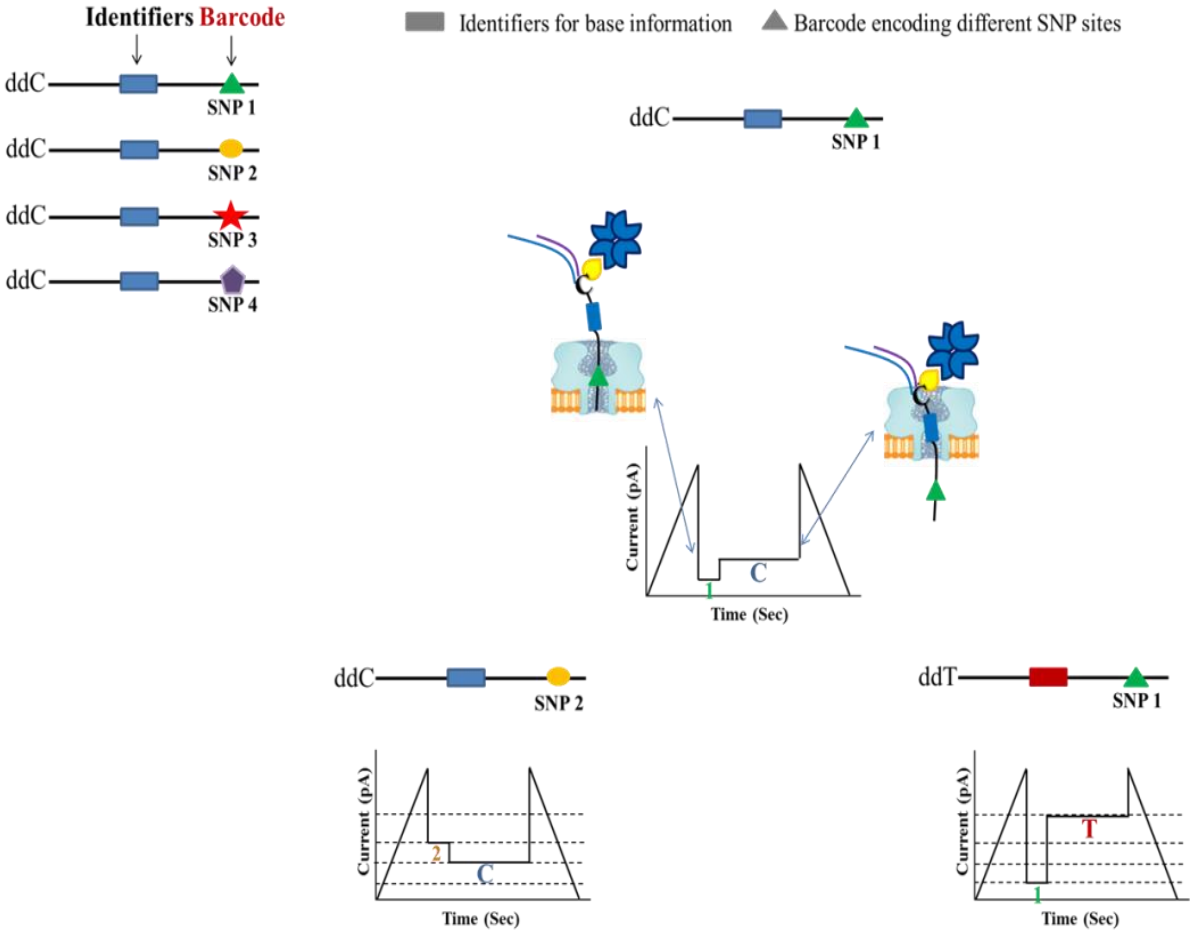


Figure A1 Schematic of current signatures of tags containing a barcode for different SNP sites. The initial transient signature which is generated as the molecule crosses the pore encodes different SNP sites. This, along with a large array of nanopores, should allow multiplexed nanopore SNP genotyping.

BIBLIOGRAPHY

1. Manolio TA, Collins FS, The HapMap and Genome-Wide Association Studies in Diagnosis and Therapy, *Annu Rev Med*., (2009), 60: 443-456
2. Collins FS, Varmus H, A New Initiative on Precision Medicine, *N Engl J Med*., (2015), 372:793-795
3. Hunt R, Sauna ZE, Ambudkar SV, Gottesman MM, Kimchi-Sarfaty C, Silent (Synonymous) SNPs: Should We Care About Them?, *Humana Press*, (2009), 578:23-39
4. Shastry BS, SNPs: Impact on Gene Function and Phenotype, *Humana Press*, (2009), 578:3-22
5. Lakowicz JR, Principles of Fluorescence Spectroscopy, *Kluwer Academic / Plenum Publishers*, *New York*, (1999)
6. Jain KK, Molecular diagnostics as basis of personalized medicine, *Textbook of Personalized Medicine*., (2009), 2:29-58
7. Komar AA, Single nucleotide polymorphisms, *Methods in Molecular Biology*, 2nd edition, (2009),1:3-16
8. Fareed M, Afzal M, Single nucleotide polymorphism in genome-wide association of human population: A tool for broad spectrum service, *The Egyptian Journal of Medical Human Genetics*, (2013), 14:123-134
9. Perkel J, SNP genotyping: six technologies that keyed a revolution, *Nature Methods*, (2008), 5:447-454
10. Whirl-Carrillo1 M, McDonagh1 EM, Hebert1 JM, Gong L, Sangkuhl1 K, Thorn1 CF, Altman RB, Klein TE, Pharmacogenomics Knowledge for Personalized Medicine, *Clin Pharmacol Ther*., (2012), 92:414-417.
11. Carlson B, SNPs- A shortcut to personalized medicine, *Genetic Engineering and Biotechnology News*, (2008), 28:12
12. Lindpaintner K., Pharmacogenetics and the future of medical practice, *Br J Clin Pharmacol*., (2002), 54:221-230.
13. Pirmohamed M., Pharmacogenetics: past, present and future, *Drug Discov Today*., (2011) 16:852-61
14. Ma Q, Lu AYH, Pharmacogenetics, Pharmacogenomics, and Individualized Medicine, *Pharmacological Reviews*, (2011), 63:437-459
15. Pirmohamed M, Pharmacogenetics and pharmacogenomics, *Br J Clin Pharmacol*., (2001), 52: 345–347.
16. Kim S, Misra A, SNP genotyping technologies and biomedical applications, *Annu.Rev.Biomed.Eng.*, (2007), 9:289-320
17. Syvanen AC, Accessing genetic variation: genotyping single nucleotide polymorphisms, *Nature Reviews*, (2001), 2:930-942
18. Barkur S. Shastry, SNPs in disease gene mapping, medicinal drug development and evolution, *J Hum Genet*., (2007), 52:871-880
19. Roden DM, George Jr AL, The genetic basis of variability in drug responses.

Nature Reviews Drug Discovery, (2002), 1:37-44
20. Wei CY, Lee MT, Chen YT, Pharmacogenomics of adverse drug reactions: implementing personalized medicine, *Hum Mol Genet*., (2012), 21:58-65
21. Johnson JA, Pharmacogenetics in clinical practice: how far have we come and where are we going?, *Pharmacogenomics*, (2013), 14:835-843
22. Kaufman AL et al., Evidence for Clinical Implementation of Pharmacogenomics in Cardiac Drugs. *Mayo Clinic Proceedings*., (2015), 90:716-729
23. McKinnon RA, Ward MB, Sorich MJ, A critical analysis of barriers to the clinical implementation of pharmacogenomics, *Ther Clin Risk Manag*., (2007), 3:751-759.
24. Shastry BS, Pharmacogenetics and the concept of individualized medicine, *The Pharmacogenomics Journal*, (2006), 6:16-21
25. Ventola CT, The Role of Pharmacogenomic Biomarkers in Predicting and Improving Drug Response, *P&T*., (2013), 38:624-627
26. Ventola CT, Pharmacogenomics in Clinical Practice, *P&T*., (2011), 36:412-416
27. Relling MV, Altman RB, Goetz MP, Evans WE, Clinical implementation of pharmacogenomics: overcoming genetic exceptionalism, *Lancet Oncol*., (2010), 11:507-509
28. Ragiussis J, Genotyping technologies for all, *Drug Discovery Today: Technologies*, (2006), 3:115-122
29. Podini D, Vallone PM, SNP genotyping using multiplex single base primer extension assays, *Humana Press*, (2009), 23:379-391
30. Syvänen AC, Aalto-Setala K, Kontula K, Söderlund H, A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E, *Genomics*, (1990), 8:684–692
31. Mir KU, Southern EM, Determining the influence of structure on hybridization using oligonucleotide arrays. *Nature Biotechnol.*, (1999), 17:788-792
32. Buetow KH et al. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl Acad. Sci. USA*., (2001), 98:581-584
33. Deshpande A, Valdez Y, Nolan JP, Multiplexed SNP genotyping using primer single-base extension (SBE) and microsphere arrays, *Current Protocols in Cytometry*, (2005) 13:1-11
34. LaFramboise T, Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances, *Nucleic Acids Research.*, (2009), 37:4181-4193
35. Sobrino B, Brion M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies, *Forensic Science International.*, (2005), 154:181-194
36. Twyman RM. Single nucleotide polymorphism (SNP) genotyping techniques-an overview, *Encyclopedia of Diagnostic Genomics and Proteomics*, (2005), 240:1202-1207
37. Zhou G, Gotou M, Kajiyama T, Kambara H, Multiplex SNP typing by bioluminometric assay coupled with terminator incorporation (BATI), *Nucleic Acids Res*., (2005), 33:15
38. Kim S, Edwards JR, Deng L, Chung W, Ju J, Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry, *Nucleic Acids Res.*, (2002), 30:16
39. Landegren U, Kaiser R, Sanders J, Hood L, A ligase-mediated gene detection technique, *Science*, (1988), 1077-1080

40. Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P et al., Single nucleotide polymorphism analysis by pyrosequencing, *Anal. Biochem*., (2000), 280:103-10
41. Kwok PY, Methods for genotyping single nucleotide polymorphisms, *Annual Review of Genomics and Human Genetics*, (2001), 2:235-258
42. Tyagi S, Bratu DP, Kramer FR, Multicolor molecular beacons for allele discrimination, *Nat. Biotechnol*., (1998), 16:49-53
43. Latif S, Bauer-Sardiña I, Ranade K, Livak K, Kwok PY, Fluorescence polarization in homogeneous nucleic acid analysis II: 5′ nuclease assay, *Genome Res.*, (2001), 11:436-40
44. Jose Luis Royo1, Manuel Hidalgo2 & Agustin Ruiz, Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping, *Nature Protocols*, (2007), 2:1734-1739
45. Howorka S, Siwy Z, Nanopore analytics: sensing of single molecules, *Chem. Soc. Rev*., (2009), 38:2360-2384
46. Wanunu M, Nanopores: a journey towards DNA sequencing. *Phys. Life Rev*., (2012), 9:125-158
47. Gu LQ, Shim JW, Single molecule sensing by nanopores and nanopore devices, *Analyst*, (2010), 135:441-451
48. Venkatesan BM, Bashir R, Nanopore sensors for nucleic acid analysis, *Nature Nanotechnology*, (2011), 6:615-624
49. Howorka S, Building membrane nanopore, *Nature Nanotechnology*, (2017), 12:619-630
50. Feng Y, Zhang Y, Ying C, Wang D, Du C, Nanopore-based Fourth-generation DNA Sequencing Technology, *Genomics Proteomics Bioinformatics*, (2015), 13:4-16
51. Dekker C, Solid-state nanopores, *Nat. Nanotech*., (2007), 2:209-215
52. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE, Structure of Staphylococcal α-Hemolysin, a Heptameric Transmembrane Pore, *Science*, (1996) 274:1859-1865
53. Montoya M, Gouaux E, β-Barrel membrane protein folding and structure viewed through the lens of α-hemolysin, *Biochimica et Biophysica Acta (BBA)*, (2003), 1609:19-27
54. Kasianowicz JJ, Brandin E, Branton D, Deamer DW, Characterization of individual polynucleotide molecules using a membrane channel, *Proc. Natl. Acad. Sci. U.S.A.*, (1996), 93:13770-13773
55. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H, Continuous base identification for single-molecule nanopore DNA sequencing, *Nat. Nanotech*., (2009), 4:265-270
56. Fuller WC et al., Real-time single molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array, *Proc. Natl. Acad. Sci. U.S.A.*, (2016), 113:5233-5238
57. Mirsaidov UM, Wang D, Timp W, Timp G, Molecular diagnostics for personal medicine using a nanopore, *WIREs Nanomed Nanobiotechnol*., (2010), 2:367-381
58. Stranges PB et al., Design and chacterization of a nanopore-coupled polymerase for single-molecule DNA sequencing by synthesis on an electrode array, *Proc. Natl. Acad. Sci. U.S.A.*, (2016), 113:6749-6756
59. Branton D et al., The potential and challenges of nanopore sequencing, *Nature Biotechnology*, (2008), 10:1146-1153
60. Astier Y, Braha O, Bayley H, Toward Single Molecule DNA Sequencing: Direct Identification of Ribonucleoside and Deoxyribonucleoside 5‘-Monophosphates by Using

an Engineered Protein Nanopore Equipped with a Molecular Adapter, *J. Am. Chem. Soc*., (2006), 128:1705-1710
61. Manrao EA et al., Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase, *Nature Biotechnology*, (2012), 30:349-353
62. Botstein D, Risch N, Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease, *Nature Genet*., (2003), 33: 228-237
63. Zhao Q et al., Detecting SNPs using a synthetic nanopore, *Nano Lett*., (2007), 7:1680-1685
64. Kong J, Zhu J, Keyser UF, Single molecule based SNP detection using designed DNA carriers and solid-state nanopores, *Chemical Communications*, (2017), 53:436-439
65. Wang Y et al., Nanolock–Nanopore Facilitated Digital Diagnostics of Cancer Driver Mutation in Tumor Tissue, *ACS sensors*, (2017), 2:975-981
66. Jehan T, Lakhanpaul S, Single nucleotide polymorphism (SNP)-Methods and applications in plant geneticsL A review, *Indian Journal of Biotechnology*, (2006), 5:435-459
67. Stoddart D et al., Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore, *Proc. Natl. Acad. Sci. U.S.A., (*2009), 106:7702-7707
68. Kumar S et al., PEG-Labeled Nucleotides and Nanopore Detection for Single Molecule DNA Sequencing by Synthesis, *Scientific Reports*, (2012), 2:684
69. Rosenblum BB et al., New dye-labeled terminators for improved DNA sequencing patterns, *Nucleic Acids Research*, (1997), 25: 4500-4504
70. Pelletier H, Sawaya MR, Kumar A, Wilson SH, Kraut J, Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP, *Science*, (1994), 264,1891-1903
71. Ju J, Kim DH, Bi L, Meng Q, Bai X, Li Z, Li X, Marma MS, Shi S, Wu J, Edwards JR, Romu A, Turro NJ, Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators, *Proc. Natl. Acad. Sci U.S.A.*, (2006), 103:19635-19640
72. Turcatti G, Romieu A, Fedurco M, Tairi AP, A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis, *Nucleic Acids Res*., (2008), 36:25
73. Zhu Z, Chao J, Yu H, Waggoner AS, Directly labeled DNA probes using fluorescent nucleotides with different length linkers., *Nucleic Acids Res*., (1994), 22: 3418-3422
74. Edwards JR, Itagaki Y, Ju J, DNA sequencing using biotinylated dideoxynucleotides and mass spectrometry, *Nucleic Acids Res*., (2001), 29:104.
75. Gardner AF, Jack WE, Determinants of nucleotide sugar recognition in an archaeon DNA polymerase, *Nucleic Acids Res*., (1999), 27:2545-2555
76. Gardner AF, Jack WE, Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and Taq DNA polymerases, *Nucleic Acids Res*., (2002), 30:605-613
77. Sanger F, Nicklen S, Coulson AR, DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, (1977), 74 5463-5467
78. Tabor S, Richardson CC, DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, (1987c). 84 4767-4771

79. Tabor S, Richardson CC, A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides, *Proc. Natl. Acad. Sci. U.S.A.*, (1995), 92:6339-43.
80. Vander Horn PB, Davis MC, Cunniff JJ, Ruan C, McArdle BF, Samols SB, Szasz J, Hu G, Hujer KM, Domke ST, Brummet SR, Moffett RB, Fuller CW. Thermo Sequenase DNA polymerase and *T. acidophilum* pyrophosphotase: New thermostable enzymes for DNA sequencing, *BioTechniques*, (1997), 22:758-765
81. Zhu B, Bacteriophage T7 DNA polymerase-sequenase, *Front Microbiol*., (2014), 5:181
82. Tabor S, Huber HE, Richardson CC, *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7, *J. Biol. Chem.*, (1987a), 262:16212-16223
83. Korch C, Drabkin H, Improved DNA Sequencing Accuracy and Detection of Heterozygous Alleles Using Manganese Citrate and Different Fluorescent Dye Terminators, *Genome Res*., (1999), 9: 588-595
84. Tabor S, Richardson CC, Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis, *J. Biol. Chem.*, (1989a), 264 6447-6458
85. Frank EG, Woodgate R, Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese, *J Biol Chem*., (2007), 282:24689-96
86. Nikiforov TT, Fluorogenic polymerase, endonuclease, and ligase assays based on DNA substrates labeled with a single fluorophore, *Anal Biochem.*, (2011), 412:229-36
87. Chen CY, DNA polymerase drive DNA sequencing-by-synthesis technologies: both past and present, *Frontiers in Microbiology*, (2014), 5:305
88. Duthie RS et al., Novel cyanine dye-labeled dideoxynucleoside triphosphates for DNA sequencing, *Bioconjugate Chemistry*, (2002), 13:699-706
89. Zhu TF, Budin I, Szostak JW, Vesicle extrusion through polycarbonate track-etched membranes using a hand-held mini-extruder, *Methods Enzymol*, (2013), 533:275-282
90. Ashkenasy N, Sánchez-Quesada J, Bayley H, Ghadiri MR, Recognizing a single base in an individual DNA strand: a step toward DNA sequencing in nanopore, *Angew. Chem. Int. Ed.*, (2005), 44:1401-1404
91. Howorka S, Cheley S, Bayley H, Sequence-specific detection of individual DNA strands using engineered nanopores, *Nat. Biotechnol.*, (2001), 19:636-639
92. Stoddart D et al., Nucleobase recognition in ssDNA at the central constriction of the α-hemolysin pore, *Nano Lett*., (2010), 10:3633-3637
93. Robertson JW et al., Single-molecule mass spectrometry in solution using a solitary nanopore., *Proc Natl Acad Sci. U.S.A*, (2007), 104:8207-11
94. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M, Automated forward and reverse ratcheting of DNA in a nanopore at 5Å precision, *Nat. Biotechnol.*, (2012), 30:344-348
95. Venkatesan BM, Bashir R, Nanopore sensors for nucleic acid analysis, *Nature Nanotechnology.*, (2011), 6:615-624
96. Howorka S, Siwy Z, Nanopore analytics: sensing of single molecules, *Chem. Soc. Rev*., (2009), 38:2360-2384
97. Lim BN, Choong YS, Ismail A, Glökler J, Konthur Z, Lim TS, Directed evolution of nucleotide-based libraries using lambda exonuclease, *BioTechniques*, (2012), 53:357-364

98. Howorka S, Bayley H, Probing distance and electrical potential within a protein pore with tethered DNA, *Biophys. J.*, (2002), 83:3202-3210
99. Aksimemtiev A, Schulten K, Imaging αHL with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map, *Biophys. J.*, (2005), 88:3745-3761
100. Henrickson SE et al., Probing single nanometer-scale pores with polymeric molecular rulers, *J Chem Phys.*, (2010), 132:135101
101. Henrickson SE et al., Driven DNA transport into an asymmetric nanometer-scale pore, *Phys Rev Lett.*, (2000), 85:3057-60
102. Meller A, Nivon L, Branton D, Voltage-driven DNA translocations through a nanopore, *Phys Rev Lett.*, (2001), 86:3435
103. Deamer DW, Characterization of nucleic acids by nanopore analysis, *Acc. Chem. Res.*, (2002), 35:817-825
104. Muthukumar M, Plesa C, Dekker C, Single-molecule sensing with nanopores, *Physics Today*, (2015), 68:40
105. Noskov SY, Im W, Roux B, Ion permeation through the α-hemolysin channel: theoretical studies based on Brownian Dynamics and Poisson-Nernst-Plank electrodiffusion theory, *Biophysical Journal*, (2004), 87:2299-2309
106. Simon SM, Blobel G, A protein-conducting channel in the endoplasmic reticulum, *Cell*, (1991), 65:371-380
107. Berrier C, Bonhivers M, Letellier L, Ghazi A, High-conductance channel induced by the interaction of phage lambda with its receptor maltoporin, *FEBS Letters*, (2000), 476:129-133
108. Garcia LR, Molineux IJ, Transcription-independent DNA translocation of bacteriophage T7 DNA into Escherichia coli, *J. Bacteriol*, (1996), 178:6921
109. Meller A, Dynamics of polynucleotide transport through nanometer-scale pores, *J Phys Cond Matt*, (2003), 15:581-607
110. Chen H, Meisburger SP, Pabit SA, Sutton JL, Webb WW, Pollack L, Ionic strength-dependent persistence lengths of sing-stranded RNA and DNA, *Proc. Natl. Acad. Sci. U.S.A.*, (2012), 109:799-804
111. Antao VP, Lai SY, Tinoco I, A thermodynamic study of unusually stable RNA and DNA hairpins, *Nucleic Acids Res.*, (1991), 19:5901-5905
112. Nakano M, Moody EM, Liang J, Bevilacqua PC, Selection for Thermodynamically Stable DNA Tetraloops Using Temperature Gradient Gel Electrophoresis Reveals Four Motifs: d(cGNNAg), d(cGNABg), d(cCNNGg), and d(gCNNGc), *Biochemistry*, (2002), 41:14281-14292
113. Antao VP, Tinoco I, Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops, *Nucleic Acids Res.*, (1991), 20:819-824
114. Deamer DW, Branton D, Characterization of nucleic acids by nanopore analysis, *Acc. Chem. Res.*, (2002), 35:817-825
115. Branton D, Sauer-Budge AF, Nyamwanda JA, Lubensky DK, Unzipping kinetics of double-stranded DNA in a nanopore, *Phys. Rev. Lett.*, (2003), 90:238101-238104
116. Smith AM, Abu-Shumays R, Akeson M, Bernick DL, Capture, unfolding, and detection of individual tRNA molecules using a nanopore device, *Front Bioeng Biotechnol.*, (2015), 3:91

117. Henley RY, Carson S, Wanunu M, Studies of RNA sequence and structure using nanopores, *Prog Mol Biol Transl Sci*, (2016), 139:73-99
118. Ding Y, Fleming AM, White HS, Burrows CJ, Internal *vs* fishhook hairpin DNA: unzipping locations and mechanisms in the α-Hemolysin nanopore, *J. Phys. Chem.*, (2014), 118:12873-12882
119. Gaspers PB, Gast AP, Robertson CR, Enzymes on immobilized substrate surfaces: reaction, *J. Colloid Interface Sci.*, (1995), 172:518-529
120. Rieder MJ et al. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose, *N Engl J Med*., (2005), 352:2285-2293
121. Owen RP, Gong Li, Sagrelya H, Klein TE, Altman B, VKORC1 pharmacogenomics summary, *Pharmacogenetics and Genomics*, (2010), 20:642-644
122. Dan M. Roden, Russell A. Wilke, Heyo K. Kroemer, C. Michael Stein, Pharmacogenomics; The Genetics of Variable Drug Responses, *Circulation*, (2011), 123:1661-1670
123. Johnson JA et al., Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing, *Clinical pharmacology and therapeutics*., (2011), 90:625-9
124. http://www.idtdna.com/pages/decoded/decoded-articles/core-concepts/decoded/2012/09/20/which-biotin-modification-to-use- (accessed 2017 Aug 30)
125. Schink S, Renner S, Alim K, Arnaut V, Simmel FC, Gerland U, Quantitative analysis of the nanopore translocation dynamics of simple structured polynucleotides, *Biophys. J*, (2012), 102:85-95