SINGLE-MOLECULE ANALYSIS OF DNA ABASIC SITES AND THE HUMAN TELOMERIC G-QUADRUPLEX THROUGH THE ALPHA-HEMOLYSIN ION CHANNEL

by

Na An

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

The University of Utah

December 2012

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Na An**

has been approved by the following supervisory committee members:

Cynthia J. Burrows	, Chair	**08/17/2012** Date Approved
Henry S. White	, Member	**08/17/2012** Date Approved
Marc D. Porter	, Member	**08/27/2012** Date Approved
Kenneth J. Woycechowsky	, Member	**08/17/2012** Date Approved
Saveez Saffarian	, Member	**08/17/2012** Date Approved

and by **Henry S. White**, Chair of the Department of **Chemistry**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Nanopore technology has been at the center of attention during the past decade as one of the most promising methods for next-generation DNA sequencing. It proceeds by electrically drawing an individual single-stranded DNA (ssDNA) strand through a nanoscale pore, leading to detectable changes in the electrical current. Chief advantages of this technology are the minimal requirements of expensive reagents and data storage and handling equipment. The bacterial protein α-hemolysin is the most commonly used biological nanopore for this platform, due to its excellent stability, reproducibility and precise tuning properties by site-directed mutagenesis. One of the most difficult challenges of this technology arises from the low current amplitude resolution between the DNA nucleotides.

The first work presented here approaches these problems by DNA site-specific chemistry to attach detectable labels to one of the most commonly occurring lesions in cells, DNA abasic (AP) sites. Several amines were used to attach to AP site, which showed detectable current changes when the ssDNA was immobilized inside of the nanopore. However, only the 18-crown-6 (18c6) moiety produced distinct current signatures during translocation, when bound to Na^+. The bulky adduct also slowed down the DNA motion to more easily recordable levels, achieving the detection of individual

AP sites at a single-molecule level. 18c6 can form different shapes of complexes, dictated by the surrounding ions, which was used to precisely manipulate its electrical behaviors.

Secondly, the nanocavity of this protein was used to provide insights into secondary structures of the human telomeric G-quadruplexes at a single-molecule level. The folding of the repeat sequence at the end of the chromosome was shown to have significance to genome protection, and depending on the surrounding ions, it could form various quadruplexes. The interactions of these structures and the α-HL were correlated to different current patterns when the DNA was encapsulated inside of the channel, providing better understanding into the polymorphism of the human telomere sequence.

Lastly, combining the above two findings, the 18c6 label was used to detect the oxidative damage of the G-quadruplexes, and the effect on the stability of these structures were also evaluated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

5-mC: 5-methylcytosine

8-BrG: 8-bromoguanine

15c5: 2-aminomethyl-15-crown-5

18c6: 2-aminomethyl-18-crown-6

AP site: DNA abasic site

α-HL: α-hemolysin

BER: base-excision repair

Btn: biotin

CD: circular dichroism

CPD: cyclobutane-pyrimidine dimer

DPhPC: 1,2-diphytanoyl-*sn*-glycero-3-phophocholine

DSB: double-strand break

GlcN: D-(+)-glucosamine

GNM: glass nanopore membrane

GPRP: Gly-Pro-Arg-Pro amide

MspA: *Mycobacterium smegmatis* porin A

NER: nucleotide-excision repair

ODN: oligodeoxynucleotide

OG: 8-oxo-7,8-dihydroguanine

RH: Arg-His

ROS: reactive oxygen species

ssDNA: single-stranded DNA

STM: streptomycin

Strep: streptavidin

TBA: thrombin-binding aptamer

t_D: duration

THF: tetrahydrofuran

T_m: melting temperature

Trn: taurine

UDG: uracil-DNA glycosylase

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my research advisor, Dr. Cynthia J. Burrows, who has been instrumental in guiding and inspiring me through these five years. Her dedication and enthusiasm in science, her critical and creative way of thinking and most importantly, her support and understanding not only have made me a better scientist, but also a better person.

No one could ask for a better group to work with than the one I have had for my graduate school career. The members of Burrows laboratory have been very helpful in providing perceptive, discussing technical problems, as well as being great friends. Especially, I am very grateful for all the support and expertise from Dr. Aaron M. Fleming, who made the last two chapters possible. Also, I would like to thank Dr. Henry S. White for being such a great collaborator and mentor through my research. His example guided me to approach scientific questions in an analytical fashion. I am thankful for all the helpful discussion and suggestions from Dr. Anna Schibel, Qian Jin, Dr. Jin Liu, and Dr. John Watkins from the White group.

I would like to attribute my achievement to my parents for encouraging and loving me unconditionally. And I also want to thank my friend Yiwen Yao for her support and respect.

CHAPTER 1

INTRODUCTION

DNA damage within cellular environments

The integrity of genetic information is constantly at risk due to irreversible, indiscriminate modifications by highly reactive chemicals of exogenous and endogenous origin (Figure 1.1). As one of the major targets for UV-induced damage, cellular DNA, in various organisms ranging from bacteria to humans, is subject to photoreactions, forming mutagenic and cytotoxic lesions such as cyclobutane-pyrimidine dimers (CPDs), 6-4 photoproducts and their Dewar valence isomers.[1, 2] Ionizing radiation also induces a variety of DNA damages, in which DNA double-strand breaks (DSBs) are the most biologically important, due to their high lethality if unrepaired or mis-repaired.[3] DNA single-strand breaks (SSBs), abasic sites (AP), and base lesions are other types of radiation-induced damages and they occur more frequently than DSBs; however, they can be more efficiently repaired with high fidelity and in a timely manner.[4] Additionally, major classes of frontline chemotherapeutic drugs consist of assorted alkylating agents, including triazenes, nitrogen mustards, aziridines, chloroethylating nitrosoureas etc. These highly reactive chemicals transfer alkyl groups onto DNA by reacting with the ring nitrogen and extracyclic oxygen atoms of the base to form a number of covalent adducts.[5]

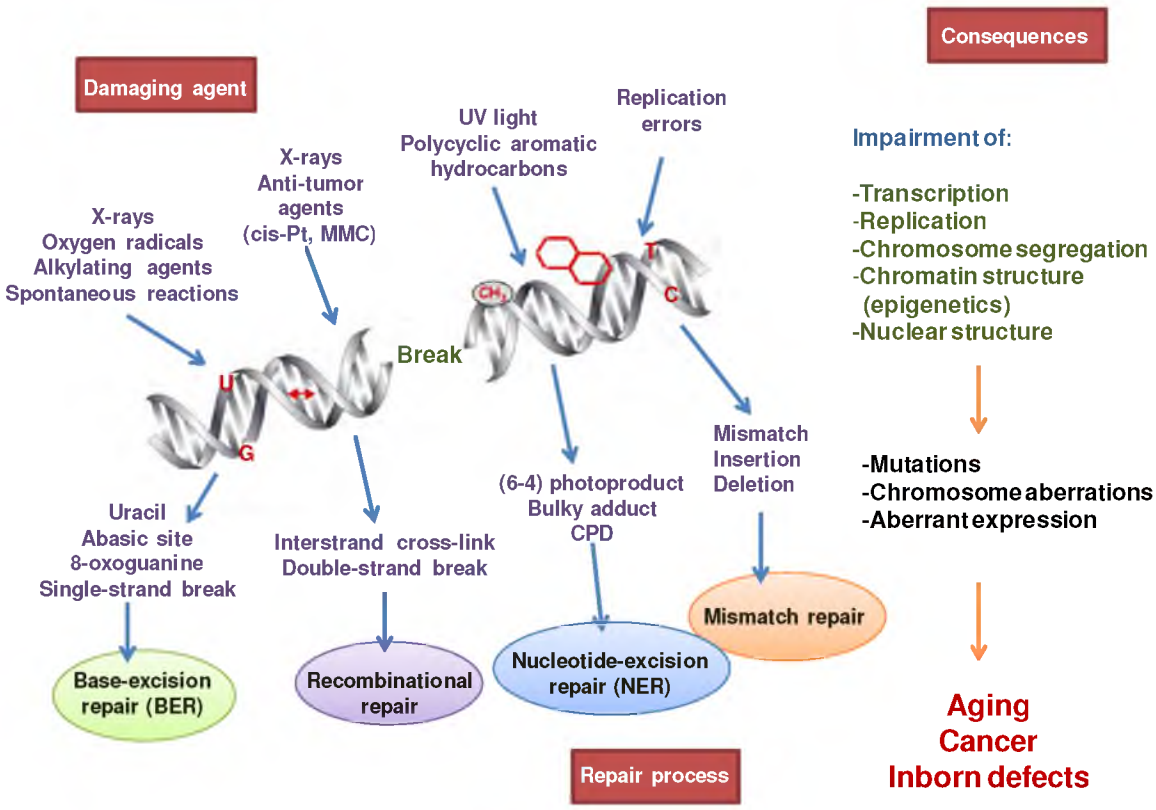


Figure 1.1. DNA damage, repair pathways and consequences.

Besides unavoidable alkylating agents also found in the air, water and food, such as biological byproducts and pollutants (e.g., tobacco smoke, fuel combustion products),[6, 7] ambient particulate matter with an aerodynamic diameter $\leq$ 2.5 μm is associated with elevated urinary concentrations of the biomarker 8-oxo-7,8-dihydroguanine (OG) in humans, indicating increased oxdative stress.[8]

In addition to the above-mentioned exogenous factors, there are sources of DNA damage that arise continuously within living cells. Metabolic processes of the organism generate reactive oxygen species (ROS), which include the free radicals superoxide radical anion ($O_2^{\bullet-}$) and hydroxyl radical (•OH), as well as nonradical oxygen species such as singlet oxygen ($^1O_2$), peroxynitrite ($ONOO^-$) and hydrogen peroxide (H_2O_2).[9, 10] These oxidants are formed as a result of an inflammatory response, mitochondrial respiration,[11] cytochrome P450 metabolism,[12] and other biological reactions. ROS can attack DNA bases and the deoxyribose moieties, leading to oxidized bases, abasic sites, cross-links, bulky adducts, SSBs and DSBs (Figure 1.2).[13, 14] Simultaneously, besides nucleic acids, other cellular components also suffer from ROS-induced damage. Their various subsequent reaction products lipid peroxidation products, *O*-estrogen metabolites, reactive carbonyl species, along with endogenous alkylating agents that modify DNA through the formation of mutagenic adducts.[15]

The chemical lability of DNA itself also contributes to the instability of the genome. Spontaneous hydrolysis of the glycosylic bond results in the formation of DNA abasic (AP) sites, which, if not repaired prior to replication, would stall the polymerase bypass, leading to cellular dysfunction.[16] More details of AP sites will be discussed in Chapter 2. Deamination of cytosine (C) to uracil (U) or 5-methylcytosine (5-mC) to

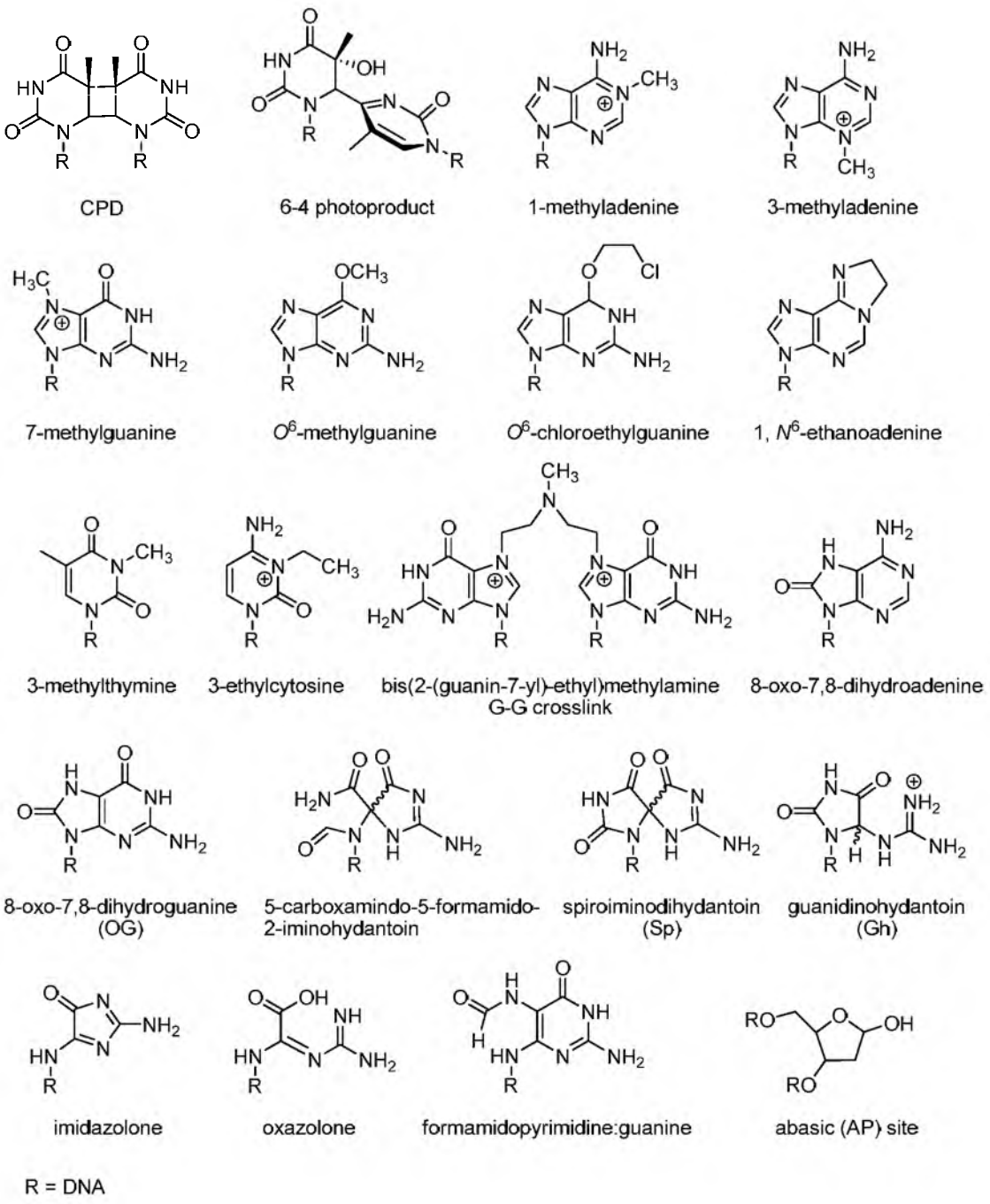


Figure 1.2. Structures of common DNA base lesions.

thymine (T) occurs at 1/500 the rate of depurination; however, the conversion of 5-mC to T generates a C → T transition mutation at methylated CpG sites, if not properly repaired, which is a frequent occurrence in oncogenes.[15, 17] Furthermore, the enormous size of the human genome significantly increases its probability of reaction. Additionally, mutations can arise from errors in DNA synthesis, replication and repair processes due to misincorporation by DNA polymerases. Depending on the identity and function of the enzyme, the error rates of mammalian DNA polymerases range from one in five thousand to ten million nucleotides.[18] It has been estimated that the total number of DNA damage events, including those due to spontaneous damage, may be 100,000 per cell per day, and they can be greatly enhanced under certain conditions, such as when cells are exposed to sunlight or oxidative stress.[15, 19]

Over time, DNA accumulates alterations due to damage and errors made by the DNA machinery, which inactivate tumor-suppressor genes and activate proto-oncogenes.[20] Direct effects of disturbed DNA metabolism are the triggering of cell-cycle arrest and cell death. This is mainly caused by damage-induced transcriptional and replicational stress, in which the lesions block the bypass of DNA and RNA polymerases. More polymerases that prefer modified templates have been found than those for the intact ones; however they temporarily solve the problem at the expenses of higher error rates.[21, 22] Another known pathway to cope with the stalling injury is to reinitiate DNA replication downstream, followed by recombination to fill in the gap.[23] Unfortunately, the endpoint of both pathways is the persistent existence of the lesions, which will keep on posing challenges. Damages that are not efficiently recognized by mammalian repair systems are particularly dangerous as a result, such as UV-induced CPDs, which cause

various forms of skin cancer.[24] Additionally, cells with specialized DNA recombination machinery are particularly sensitive to DSBs, such as B- and T-cells when they are rearranging the immunoglobin or T-cell-receptor genes. Cancers (e.g., leukaemia and lymphomas), preferentially induced by ionizing irradiation, feature the frequent involvement of these genetic loci in oncogenic translocations.[25] Since intact chromosomes are a prerequisite for correct chromosome segregation during cell division, these lesions impair mitosis, inducing chromosomal aberrations and translocations. When too significant, the damage forces cells to choose the ultimate rescue mode, apoptosis, at the expense of a whole cell.

The long-term effects of DNA damage rise from irreversible mutations. The mutational spectra of the *RAS* genes indicate that alterations at codon 12 of *KRAS* are the most common mutations among the three *RAS* genes and the types of mutations are tumor specific. For instance, *RAS* often undergoes transition mutations in colorectal cancers (GGT → GAT), while exhibiting predominantly transversion mutations in lung cancers (GGT → GTT). As for the tumor suppressor gene *TP53*, methylated CpG sites constitute more than one third of all the cancer mutations and they are mostly transition mutations due to deamination of 5-mC, making them the most important hotspots for this gene in human internal cancers.[26] In some cases, these alterations can be linked to exogenous carcinogens, such as sunlight-induced skin cancers,[24] tobacco-associated lung cancers,[27] and aristolochic acid-related urothelial tumors.[28] Unfortunately, a potential carcinogen-induced origin of a great number of other human genetic mutations has not been discovered.

In addition to the quantity and identity of DNA lesions, the distributions also determine their repairability and lethality, especially in the case of damage induced by IR and certain chemotherapies.[29, 30] With similar types of DNA modifications, IR and bleomycin sulfate require fewer lesions to produce a lethal event, compared to endogenous ROS. This is explained by the fact that radiation and bleomycin introduce clustered DNA damage in the genome in a nonrandom fashion, which is defined as ≥ 2 lesions within 1-2 helical turns. The spatial distribution of these multiply damaged sites can impair the ability of the repair system in cells. Two lesions at opposite strands situated ≥ 3 bp apart generally create two SSB from repair, resulting in a DSB, while two opposing damages situated less than 3 bp apart generate only one SSB-repair intermediate, which inhibits the binding and activity of the repair enzymes for the opposite strand. Similar observations were found for tandem lesions on the same strand, and the delay of repair at one site due to the SSB intermediate provided extra time for replication to occur, enhancing the mutagenicity of the individual lesions.[4, 31]

As the blueprint of all cellular proteins and RNA, the nuclear DNA is irreplaceable; consequently, its maintenance is crucial for survival during the entire lifetime of an organism.[32] Thus, cells have invested heavily in elaborate DNA damage repair systems, all of which, collectively, cope with the inflicted insults. Nucleotide-excision repair (NER) handles the wide class of helix-distorting lesions that hinder base pairing, and obstruct transcription and replication. Complementarily, base-excision repair (BER) targets small chemical modifications of the native bases. Moreover, homologous recombination and end joining are the two pathways that tackle the more problematic DSBs. Finally, adding to the cocktail are some suicide repair proteins that directly revert

certain injuries, such as O^6-methylguanine methyltransferase. As sophisticated as this network is, when it becomes overwhelmed and the damages are beyond repair, it leads to aging and more catastrophic events, such as cancer and inborn disease. For example, people with a compromised NER pathway have an increased predisposition to xeroderma pigmentosum;[19] individuals with an impaired DSB detection and repair system are predisposed to ataxia telangiectasia and Nijmegen breakage syndrome.[33] Therefore, a quantitative evaluation of cellular DNA damage would provide insights into the correlation between DNA damage and cancer or other diseases, and will also be essential to the development of rational strategies directed against the initiation of malignant neoplasm and to the design of anticancer or anti-aging drugs.

Current DNA damage detection methods

The single-cell gel electrophoresis technique, also known as the comet assay, is one of the most reliable methods for DNA damage detection and has been used for more than 20 years in environmental toxicology and nutritional studies.[34] Typically, detergent and a high concentration of salt are utilized to lyse cells in agarose on a frosted slide, allowing histone-free DNA to unwind in the buffer before applying an electrical field. Intact DNA remains tightly supercoiled, while strand breaks move much faster through the agarose gel, forming the tail. The comet-like structure can be visualized using an intercalating dye and quantified by designed software. The ratio of the amount of DNA in the tail to the amount of DNA in the head indicates the level of damage. An upgrade of this assay, which introduces strand breaks at the sites of certain types of damage using lesion-specific repair enzymes (e.g., Fpg, hOGG, Endonuclease III),[35] helps increase both

sensitivity and most importantly, the specificity of the technique for lesion type. Combined with primer extension methodology, the comet assay can also provide surrounding sequence information along with the quantity of the damage, and it requires a subpicomole amount of DNA sample.

Despite being such a well-established method, the inability of the comet assay to detect multiple lesions on the same strand and the fact that most DNA modifications are not good substrates for polymerases become fundamental problems for the comet assay. Additionally, different experimental parameters all contribute to the results, such as solution molarity, pH, washing times, unwinding times and electrophoresis times. Moreover, poor understanding of the underlying mechanism of the comet structure limits this method to showing only a vague correlation between comet shape and the amount of DNA damage.[36]

Another common approach is the use of combined chromatography and mass spectrometry techniques (LC-MS/MS). DNA isolated from tissue samples is digested to produce single nucleosides by nuclease and phosphatase treatment, and the nucleosides are then separated by liquid chromatography (LC), followed by mass spectrometric detection.[37-39] LC is great at discriminating different damages; however, the enormous background of the native bases along with the induction of artificial alternations during sample preparation processes generate results that are always higher than that from the comet assay for the same sample.[40] Furthermore, no sequence information can be provided, and the digestion of bulky adducts is normally challenging.

Fluorescence-based detection platforms were also developed to monitor changes in DNA organization, fragmentation, chromosomal aberrations and apoptosis, such as

halo assay, fluorescence in situ hybridization (FISH), annexin V labeling.[41, 42] Immunological assays with radiochemical, fluorescent or enzyme-conjugated secondary antibodies present promising qualities for the quantification of thymine glycol, cis-platin-induced adducts, CPDs, and their Dewar isomers with high efficiency; however, they all suffer from the cross-reactivity of the antibodies with normal DNA bases.[43,44] Moreover, Barton et al. have estimated the level of DNA damage by examining the change in the efficiency of DNA-mediated charge transport caused by base modifications and by the formation of bulky adducts.[45, 46] However, none of these techniques possess all of the following attributes: the ability to detect single molecules, the ability to provide accurate sequence information, the ability to recognize multiple lesions on the same strand, high sensitivity and specificity, and low cost.

Single-molecule analysis of DNA with the α-hemolysin ion channel

The α-hemolysin (α-HL) toxin is secreted by *Staphylococcus aureus* and has been identified as a suitable ion channel due to its ability to self-assemble into a heptameric complex across a lipid bilayer, forming an aqueous pore with a diameter just large enough to translocate single-stranded DNA (ssDNA).[47-51] This mushroom-shaped protein consists of a larger vestibule with an interior dimension of ~3.6 nm that in turn leads to its β-barrel, penetrating the lipid bilayer. The external size of the pore is 10 × 10 nm, while the inner constriction of the narrower β-barrel is ~1.4 nm, which nicely accommodates ssDNA.[52] Upon the application of an electrical field, the background electrolytes are driven through the α-HL ion channel, creating an open channel current (I_o). In the presence of negatively charged ssDNA, a given molecule would occasionally

diffuse into a small volume near the mouth of the pore, then be actively captured by the voltage bias, and drawn from the *cis* to the *trans* side of the protein, or vice versa (Figure 1.3). As a result, a current blockage as large as 85% of I_o is generated; potentially, its duration and blocking levels could provide the information of the identities of the sequential bases.

Emerging as a rapid and inexpensive single-molecule DNA sequencing platform, nanopore ion channel technology has been under intensive investigation. Chief advantages of the α-HL protein nanopore include its excellent stability, reproducibility, and precise tuning properties via site-directed mutagenesis, in addition to the minimal requirements of expensive reagents and data storage and handling equipment. Furthermore, it eliminates the need for enzyme-dependent DNA amplification, and consequently, the costs of nanopore analysis are projected to be far lower than traditional methods. The high speed (1-20 μs/nucleotide) at which ssDNA is driven through the ion channel above the threading voltage promises rapid and long reads; however, it is this very feature of free translocation that prohibits resolution of single bases that exhibit characteristic signatures of less than a few picoamperes current difference using available electronics.[50, 53] Another challenge is the difficulty of achieving single-nucleotide discrimination, given that there are ~10-15 nucleotides, instead of one occupying the sensing region of the protein channel, which contribute to the current signature simultaneously.[54, 55] Most efforts to approach these obstacles are devoted to engineering the protein,[56, 57] embedding adapters,[58] synthesizing peptide-conjugated nucleotides,[59] controlling DNA movement with molecular motors,[60] and exploring experimental conditions.[61, 62]

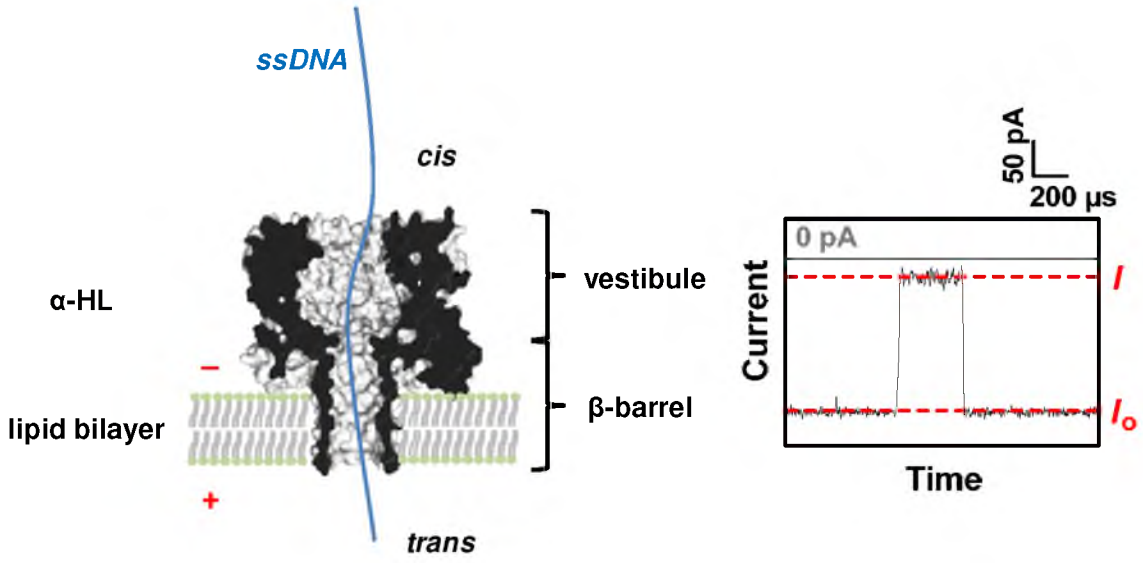


Figure 1.3. Single-molecule analysis of DNA using α-hemolysin (α-HL) nanopore.

As a single-molecule technology, the α-HL nanopore has also been explored to examine epigenetic markers,[63] DNA damage products,[64, 65] secondary structures,[66] enzymatic activity[67] and various chemical reactions.[68] Other protein channels have also been evaluated, such as *Mycobacterium smegmatis* porin A (MspA), which has a smaller constriction and sensing zone in comparison to α-HL.[69] At the same time, solid-state nanopores are being manufactured to provide better physical properties and various dimensions that may eventually be suitable for DNA sequencing.[70]

References

1. Sinha, R. P.; Hader, D.-P. UV-induced DNA damage and repair: a review. *Photochem. Photobiol. Sci.* **2002,** 1, 225-236.

2. Batista, L. F. Z.; Kaina, B.; Meneghini, R.; Menck, C. F. M. How DNA lesions are turned into powerful killing structures: insights from UV-induced apoptosis. *Mutat. Res.* **2009,** 681, 197-208.

3. Jeggo, P. A.; Geuting, V.; Lobrich, M. The role of homologous recombination in radiation-induced double-strand break repair. *Radiotherapy and Oncology* **2011,** 101, 7-12.

4. Shikazono, N.; Noguchi, M.; Fujii, K.; Urushibara, A.; Yokoya, A. The yield, processing, and biological consequences of clustered DNA damage induced by ionizing radiation. *J. Radiat. Res.* **2009,** 50, 27-36.

5. Fu, D.; Calvo, J. A.; Samson, L. D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* **2012,** 12, 104-120.

6. Oikawa, S. Mechanism of oxidative DNA damage induced by environmental carcinogens and antioxidants. *Genes and Environment* **2008,** 30, 1-9.

7. Moller, P.; Flokmann, J. K.; Forchhammer, L.; Brauner, E. V.; Danielsen, P. H.; Risom, L.; Loft, S. Air pollution, oxidative damage to DNA, and carcinogenesis. *Cancer Lett.* **2008,** 266, 84-97.

8. Wei, Y.; Han, I.-K.; Shao, M.; Hu, M.; Zhang, J.; Tang, X. $PM_{2.5}$ constituents and oxidative DNA damage in humans. *Environ. Sci. Technol.* **2009,** 43, 4757-4762.

9. Pluskota-Karwatka, D. Modifications of nucleosides by endogenous mutagens-DNA adducts arising from cellular processes. *Bioorg. Chem.* **2008,** 36, 198-213.

10. Jackson, A. L.; Loeb, L. A. The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutat. Res.* **2001,** 477, 7-21.

11. Yang, J.-L.; Weissman, L.; Bohr, V. A.; Mattson, M. P. Mitochondrial DNA damage and repair in neurodegenerative disorders. *DNA Repair* **2008,** 7, 1110-1120.

12. Heine, T.; Glatt, H.; Epe, B. Human cytochrome P450 reductase can act as a source of endogenous oxidative DNA damage and genetic instability. *Free Radical Biol. Med.* **2006,** 40, 801-807.

13. Cadet, J.; Douki, T.; Ravanat, J.-L. Oxidatively generated damage to the guanine moiety of DNA: mechanistic aspcets and formation in cells. *Acc. Chem. Res.* **2008,** 41, 1075-1083.

14. Wang, Y. Bulky DNA lesions induced by reactive oxygen species. *Chem. Res. Toxicol.* **2008,** 21, 276-281.

15. Bont, R. D.; Larebeke, N. V. Endogenous DNA damage in humans: a review or quantitative data. *Mutagenesis* **2004,** 19, 169-185.

16. Boiteux, S.; Guillet, M. Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair* **2004,** 3, 1-12.

17. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993,** 362, 709-715.

18. Aziz, K.; Nowsheen, S.; Pantelias, G.; Iliakis, G.; Gorgoulis, V. G.; Georgakilas, A. G. Targeting DNA damage and repair: embracing the pharmacological era for successful cancer therapy. *Pharmacology & Therapeutics* **2012,** 133, 334-350.

19. Garinis, G. A.; Horst, G. T. J. V. d.; Vijg, J.; Hoeijmakers, J. H. J. DNA damage and ageing: new-age ideas for an age-old problem. *Nat. Cell Biol.* **2008,** 10, 1241-1247.

20. Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature* **2001,** 411, 366-374.

21. Goodman, M. F.; Tippin, B. Sloppier copier DNA polymerases involved in genome repair. *Curr. Opin. Genet. Dev.* **2000,** 10, 162-168.

22. Kunkel, T. A.; Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **2000,** 69, 497-529.

23. Lawrence, C. The Rad6 DNA repair pathway in *Saccharomyces cerevisiae*: what does it do and how does it do it? *BioEssays* **1994,** 16, 253-258.

24. Pfeifer, G. P.; Besaratinia, A. UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. *Photochem. Photobiol. Sci.* **2012,** 11, 90-97.

25. Helmink, B. A.; Sleckman, B. P. The response to and repair of RAG-mediated DNA double-strand breaks. *Annu. Rev. Immunol.* **2012,** 30, 175-202.

26. Pfeifer, G. P.; Besaratinia, A. Mutational spectra of human cancer. *Hum. Genet.* **2009,** 125, 493-506.

27. Loeb, L. A.; Harris, C. C. Advances in chemical carcinogenesis: a historical review and prospective. *Cancer Res.* **2008,** 68, 6863-6872.

28. Wang, Y.; Meng, F.; Arlt, V. M.; Mei, N.; Chen, T.; Parson, B. L. Aristolochic acid-induced carcinogenesis examined by ACB-PCR quantification of H-*Ras* and K-*Ras* mutant faction. *Mutagenesis* **2011,** 26, 619-628.

29. Radford, I. R.; Lobachevsky, P. N. Clustered DNA lesion sites as a source of mutations during human colorectal tumourigenesis. *Mutat. Res.* **2008,** 646, 60-68.

30. Leloup, C.; Garty, G.; Assaf, G.; Cristovao, A.; Breskin, A.; Chechik, R.; Shchemelinin, S.; Paz-Elizur, T.; Livneh, Z.; Schulte, R. W.; Bashkirov, V.; Milligan, J. R.; Grosswendt, B. Evaluation of lesion clustering in irradiated plasmid DNA. *Int. J. Radiat. Biol.* **2005,** 81, 41-54.

31. Sage, E.; Harrison, L. Clustered DNA lesion repair in eukaryotes: relevance to mutagenesis and cell survival. *Mutat. Res.* **2011,** 711, 123-133.

32. Harper, J. W.; Elledge, S. J. The DNA damage response: ten years after. *Mol. Cell* **2007,** 28, 739-745.

33. Maynard, S.; Schurman, S. H.; Harboe, C.; Souza-Pinto, N. C. D.; Bohr, V. A. Base excision repair of oxidative DNA damage and association with cancer and aging. *Carcinogenesis* **2009,** 30, 2-10.

34. Azqueta, A.; Shaposhnikov, S.; Collins, A. R. DNA oxidation: investigation its key role in environmental mutagenesis with the comet assay. *Mutat. Res.* **2009,** 674, 101-108.

35. Mangal, D.; Vudathala, D.; Park, J.-H.; Lee, S. H.; Penning, T. M.; Blair, I. A. Analysis of 7,8-dihydro-8-oxo-2'-deoxyguanosine in cellular DNA during oxidative stress. *Chem. Res. Toxicol.* **2008,** 22, 788-797.

36. McArt, D. G.; Mckerr, G.; Howard, C. V.; Saetzler, K.; Wasson, G. R. Modelling the comet assay. *Biochem. Soc. Trans.* **2009,** 37, 914-917.

37. Yu, H.; Venkatarangan, L.; Wishnok, J. S.; Tannenbaum, S. R. Quantitation of four guanine oxidation products form reaction of DNA with varying doses of peroxynitrite. *Chem. Res. Toxicol.* **2005,** 18, 1849-1857.

38. Watanabe, K.; Liberman, R. G.; Skipper, P. L.; Tannenbaum, S. R.; Guengerich, F. P. Analysis of DNA adducts formed in vivo in rats and mice form 1,2-dibromoethane, 1,2-dichloroethane, dibromomethane, and dichloromethane using HPLC/accelerator mass spectrometry and relevance to risk estimates. *Chem. Res. Toxicol.* **2007,** 20, 1594-1600.

39. Mangerich, A.; Knutson, C. G.; Parry, N. M.; Muthupalani, S.; Ye, W.; Prestwich, E.; Cui, L.; McFaline, J. L.; Mobley, M.; Ge, Z.; Taghizadeh, K.; Wishnok, J. S.; Wogan, G. N.; Fox, J. G.; Tannenbaum, S. R.; Dedon, P. C. Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer. *Proc. Nat.l Acad. Sci. U.S.A.* **2012,** 109, Early Edition.

40. Collins, A. R.; Cadet, J.; Moller, L.; Poulsen, H. E.; Vina, J. Are we sure we know how to measure 8-oxo-7,8-dihydroguanine in DNA from human cells? *Arch. Biochem. Biophys.* **2004,** 423, 57-65.

41. Kumari, S.; Pastogi, R. P.; Singh, K. L.; Singh, S. P.; Sinha, R. P. DNA damage: detection strategies. *EXCLI Journal* **2008,** 7, 44-62.

42. Clark, T. A.; Spittle, K. E.; Turner, S. W.; Korlach, J. Direct detection and sequencing and damaged DNA bases. *Genome Integrity* **2011,** 2, 10.

43. Rossner, P. J.; Sram, R. J. Immunochemical detection of oxidatively damaged DNA. *Free Radical Res.* **2012,** 46, 492-522.

44. Liedert, B.; Pluim, D.; Schellens, J.; Thomale, J. Adduct-specific monoclonal antibodies for the measurement of cisplatin-induced DNA lesions in individual cell nuclei. *Nucleic Acids Res.* **2006,** 34, e47.

45. Boal, A. K.; Barton, J. K. Electrochemical detection of lesions in DNA. *Bioconjugate Chem.* **2005,** 16, 312-321.

46. Buzzeo, M. C.; Barton, J. K. Redmond red as a redox probe for the DNA-mediated detection of abasic sites. *Bioconjugate Chem.* **2008,** 19, 2110-2112.

47. Kasianowicz, J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of individual polynucleotide molecules usinga membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** 93, 13770-13773.

48. Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** 97, 1079-1084.

49. Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D. W.; Haussler, D.; Akeson, M. Rapid discrimination among individual DNA molecules at single nucleotide resolution using a nanopore instrument. *Nat. Biotechnol.* **2001,** 19, 248-250.

50. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Ventra, M. D.; Garaj, S.; Hibbs, A.; Huang, X.; Jovanovich, S. B.; Drstic, P. S.; Lindsay, S.; Ling, X. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **2008,** 26, 1146-1153.

51. Deamer, D. W.; Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* **2002,** 35, 817-825.

52. Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of stephylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **1996,** 274, 1859-1866.

53. Howorka, S.; Siwy, Z. Nanopore analytics: sensing of single molecules. *Chem. Soc. Rev.* **2009,** 38, 2360-2384.

54. Gyarfas, B.; Olasagasti, F.; Benner, S.; Garalde, D.; Lieberman, K. R.; Akeson, M. Mapping the position of DNA polymerase-bound DNA templates in a nanopore at 5A resolution. *ACS Nano* **2009,** 3, 1457-1466.

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

56. Maglia, G.; Restrepo, M. R.; Mikhailova, E.; Bayley, H. Enhanced translocation of sinlge DNA molecules through α-hemolysin nanopores by manipulation of internal charge. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** 105, 19720-19725.

57. Rincon-Restrepo, M.; Mikhailova, E.; Bayley, H.; Maglia, G. Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes. *Nano Lett.* **2011,** 11, 746-750.

58. Banerjee, A.; Mikhailova, E.; Cheley, S.; Gu, L.; Montoya, M.; Nagaoka, Y.; Gouaux, E.; Bayley, H. Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. *Proc Natl Acad Sci USA* **2010,** 107, 8165-8170.

59. Mitchell, N.; Howorka, S. Chemical tags facilitate the sensing of individual DNA strands with nanopores. *Angew. Chem., Int. Ed.* **2008,** 47, 5565-5568.

60. Cherf, G.; Lieberman, K.; Rashid, H.; Lam, C.; Karplus, K.; Akeson, M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precesion. *Nat Biotechnol.* **2012,** 30, 343-348.

61. Kawano, R.; Schibel, A.; Cauley, C.; White, H. Controlling the translocation of single-stranded DNA through α-hemolysin ion channels using viscosity. *Langmuir* **2009,** 25, 1233-1237.

62. Japrung, D.; Henricus, M.; Li, Q.; Maglia, G.; Bayley, H. Urea facilitates the translocation of single-stranded DNA and RNA through the α-hemolysin nanopore. *Biophys. J.* **2010,** 98, 1856-1863.

63. Wallace, E.; Stoddart, D.; Heron, A.; Mikhailova, E.; Maglia, G.; Donohoe, T.; Bayley, H. Identification of epigenetic DNA modifications with a protein nanopore. *Chem Commun* **2010,** 46, 8195-8197.

64. Schibel, A.; An, N.; Jin, Q.; Fleming, A.; Burrows, C.; White, H. Nanopore detection of 8-oxo-7,8-dihydro-2'-deoxyguanosine in immobilized single-stranded DNA via adduct formation to the DNA damage site. *J. Am. Chem. Soc.* **2010,** 132, 17992-17995.

65. Schibel, A.; Fleming, A.; Jin, Q.; An, N.; Liu, J.; Blakemore, C.; White, H.; Burrows, C. Sequence-specific single-molecule analysis of 8-oxo-7,8-dihydroguanine lesions in DNA based on unzipping kinetics of complementary probes in ion channel recordings. *J. Am. Chem. Soc.* **2011,** 133, 14778-14784.

66. Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M. Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nat. Biotechnol.* **2001,** 19, 248-252.

67. Olasagasti, F.; Lieberman, K.; Benner, S.; Cherf, G.; Dahl, J.; Deamer, D.; Akeson, M. Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat. Nanotechnol.* **2010,** 5, 798-806.

68. Lu, S.; Li, W.; Rotem, D.; Mikhailova, E.; Bayley, H. A primary hydrogen-deuterium isotope effect observed at the single-molecule level. *Nat. Chem.* **2010,** 2, 921-928.

69. Manrao, E.; Derrington, I.; Laszlo, A.; Langford, K.; Hopper, M.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **2012,** 30, 349-353.

70. Pennisi, E. Genome sequencing. Going solid-state. *Science* **2012,** 336, 536.

CHAPTER 2

SINGLE-MOLECULE DETECTION OF DNA ABASIC SITES THROUGH THE ALPHA-HEMOLYSIN ION CHANNEL

Introduction

As one of the most frequent lesions in the genome, DNA abasic (AP) sites are derived from either spontaneous hydrolysis of the glycosidic bonds, often as a result of alkylation or oxidation of purines, or the enzymatic removal of modified bases by glycosylase enzymes in the base excision repair (BER) pathway (Figure 2.1).[1-3] Failure to repair AP sites poses a severe challenge to polymerases and also leads to strand breaks, DNA cross-links and transcriptional mutations, resulting in cellular dysfunction.[1, 2, 4] Common detection methods for AP sites include using aldehyde-reactive probes,[5, 6] followed by an ELISA-like assay,[7] fluorescence microscopy,[8] or atomic force microscopy;[9] various metalloinsertors,[10] mass spectrometry[11] and isotope labels[12] can also help visualize this lesion. There is urgent need for methods that provide surrounding sequence information and have the ability to detect multiple damage sites per strand, a phenomenon of significant biological importance.[13] Furthermore, given their abundance in human cells (~18,000 per cell per day) and ability to stall polymerases, AP lesions

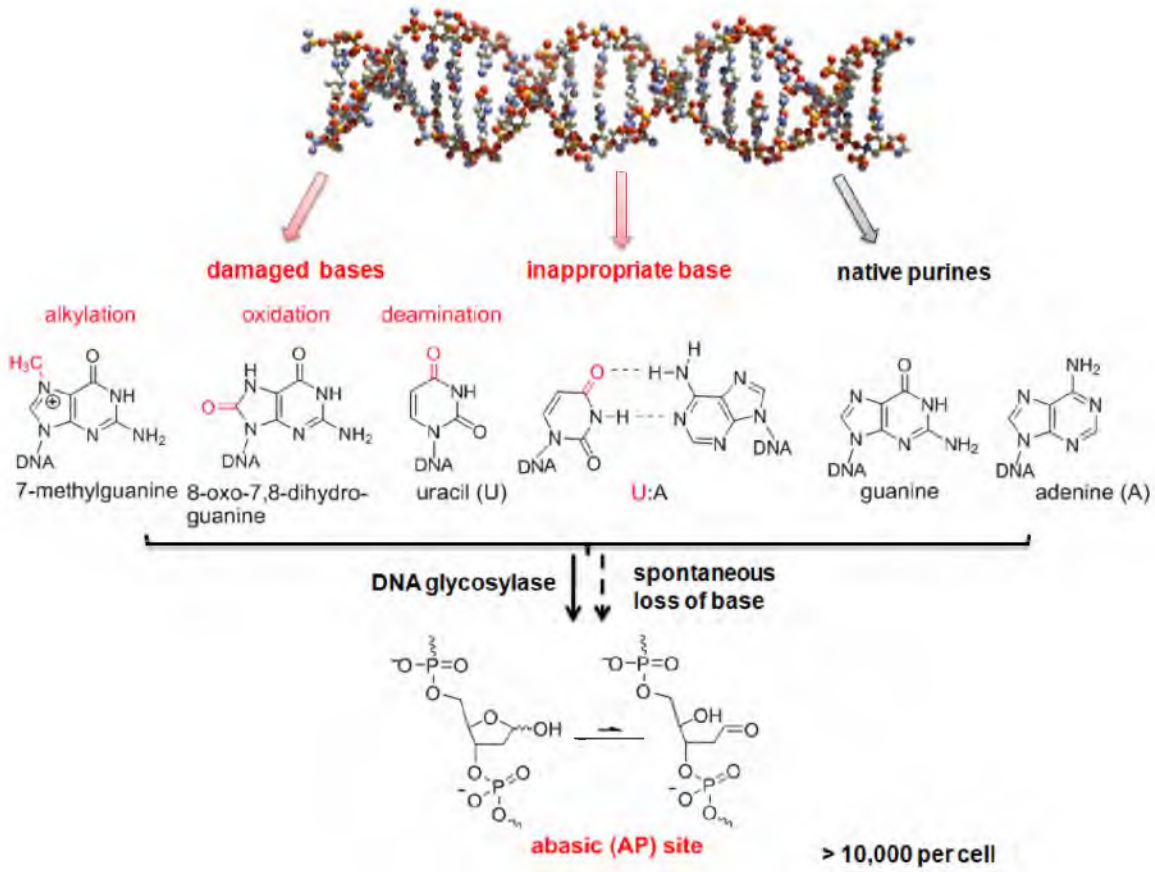


Figure 2.1. Sources of DNA abasic (AP) sites in cells.

become a technical hurdle to DNA sequencing efforts that require PCR amplification and sequencing-by-synthesis processes.

Emerging as a rapid and inexpensive single-molecule DNA analysis platform, nanopore ion channel technology has been under intensive investigation with the biological pore α-hemolysin (α-HL), which proceeds by electrically drawing an individual single-stranded DNA (ssDNA) strand through this self-assembled bacterial protein ion channel embedded in a lipid bilayer.[14, 15] Due to the limited size of the constriction of α-HL (~1.4 nm),[16] the translocation of ssDNA through the channel significantly reduces the electrolyte ion flow, resulting in a deep current blockage, potentially revealing the identity of the nucleotide sequence.[17-19] Recently, this method has also been explored to examine epigenetic markers,[20] DNA oxidative damage products,[21, 22] secondary structures,[23] enzymatic activity[24-26] and chemical reactions.[27] Among the studies that approach the drawbacks of this technology, which were discussed in Chapter 1, the stable AP analog tetrahydrofuran (THF) has been used as a marker to inspect DNA polymerase activities and to map the recognition sites of the α-HL ion channel;[24-26] however, the native AP lesion has never previously been studied in α-HL or related ion channels.

Due to the prevalent role of AP sites in normal and disease-related processes, the ability to obtain sequence information surrounding this lesion at a single-molecule level would be of great importance in understanding disease mechanisms and in medical diagnostics, for which no convenient method currently exists. The α-HL nanopore presents a potential solution to this challenge. In previous work, we showed that adducts to 8-oxo-7,8-dihydro-2'-deoxyguanosine (OG) produced different current levels than

native bases while being immobilized inside of the α-HL ion channel. However, all of these OG adducts failed to generate resolvable current signatures during free translocations.[21] In this work, we demonstrate that the selective attachment of an 18-crown-6 (18c6) moiety to an AP site produces very unique current amplitude signatures that allow the identification of individual sites along a single homopolymeric or heteropolymeric DNA strand as a result of the specific interactions between the crown ether and electrolyte cations. This is the first demonstration of current modulation during ssDNA translocation via adduct formation to native strands. Furthermore, the AP-18c6 adduct translocation behavior can be precisely manipulated by simply altering the electrolyte cation. Additionally, the speed of DNA translocation was dramatically slowed as the AP-18c6 adduct passes through the protein nanopore, resulting in millisecond-long current signatures that are readily observed.[28]

Experimental Section

Materials

KCl, NaCl, LiCl, EDTA, 2-aminomethyl-18-crown-6, $NaBH_3CN$, wild-type α-HL monomer, phospholipid 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC), streptavidin and uracil-DNA glycosylase (UDG) were purchased from commercial suppliers and used without further purification.

Preparation and characterization of AP-18c6 adduct

The oligodeoxynucleotides (ODN) were synthesized from commercially available phosphoramidites (Glen Research, Sterling, VA) by the DNA-Peptide Core Facility at the

University of Utah. Each ODN was cleaved from the synthetic column and deprotected according to the manufacturer's protocols, followed by purification using a semi-preparation ion-exchange HPLC column with a linear gradient of 25% to 100% B over 35 min while monitoring absorbance at 260 nm (A = 20 mM Tris, 1 M NaCl pH 7 in 10% CH_3CN/90% ddH_2O, B = 10% CH_3CN/90% ddH_2O). Uridine-containing ODNs (10 μM, 1 nmole) and 1 unit UDG were thermally equilibrated in UDG buffer (pH 8.0) at 37 °C for 30 min, followed by dialysis against ddH_2O for 12 h. The resulting AP-containing ODNs (10 μM, 1 nmole) were dried and resuspended in MOPS buffer (pH 6.5), followed by the addition of 2-aminomethyl-18-crown-6 (20 mM, 2 μmoles) and $NaBH_3CN$ (100 mM, 10 μmoles); then the reactions were kept at 37 °C for 24 h. Unreacted AP-containing ODNs were cleaved by 0.1 M NaOH at 90 °C for 10 min. After dialysis against ddH_2O for 12 h, all products were purified by analytical ion-exchange HPLC running a linear gradient of 25% to 100% B over 35 min while monitoring absorbance at 260 nm. Analysis of the crude reaction products indicated yields of ~95% for the mono adduct and ~80% for the bis adduct. The identities of the ODNs containing AP-18c6 were determined by negative ion electron spray (ESI^-) mass spectrometry on a Micromass Quattro II mass spectrometer equipped with Zspray API source in the mass spectrometry laboratory at the Department of Chemistry, University of Utah (calculated: 12259.4, experimental: 12259.2).

Ion channel recordings

Ultrapure water (>18 MΩ•cm) from a Barnstead E-pure water purifier and used to make buffered electrolyte solutions (1.0 mM EDTA, 25 mM Tris, pH=7.9 and either

LiCl, NaCl or KCl, at concentrations listed in the text and figure captions) for the single ion channel current recordings. The electrolyte solutions were filtered with a sterile 0.22 mm Millipore vacuum filter immediately prior to ion channel recordings. The protein α-HL was diluted to a 0.5 mg/mL solution in ultra pure water and the DPhPC was dissolved in decane to a concentration of 10 mg/mL, both of which were stored at -20 ^{o}C. The glass nanopore membrane (GNM) (radius = 600 nm) was fabricated as previously established,[29] and was silanized in 2% (v:v) 3-cyanopropyldimethylchlorosilane in CH_3CN for 6 h. Ag/AgCl electrodes were prepared by soaking Ag wires (diameter = 0.25 mm) in bleach. DPhPC bilayers spanning across the orifice of the GNM were prepared as previously described.[30] A pipette holder with a pressure gauge and a 10-mL gas-tight syringe were used to attach the GNM to the DC system. Two Ag/AgCl electrodes were positioned inside and outside of the GNM to apply a voltage. The DPhPC solution (1 μL, 10 mg/mL) was painted on the GNM surface using a plastic pipette tip. After addition of monomer α-HL (0.2 μL, 1 mg/mL), pressure was applied to form a suspended bilayer, followed by reconstitution of a single α-HL ion channel in the bilayer.

A custom built high-impedance, low-noise amplifier and data acquisition system, designed and constructed by Electronic Biosciences (EBS), San Diego, CA, was used for the current-time (*i-t*) recordings. In the immobilization studies, Strep-Btn DNA (40 pmol, 200 nM) was added in the cell and more than 200 capture/release events were collected under 120 mV bias (*trans* versus *cis*) with a 10 kHz low pass filter, and 50 kHz data acquisition rate. Then, the same amount of Strep-Btn dC_{40} control was added as an internal standard. For the translocation studies, ssDNA (2 nmol, 10 μM) was added and

more than 1000 events were collected for each voltage with a 100 kHz low pass filter, and 500 kHz data acquisition rate.

The biotin linkage that was used in the immobilization studies was the same as previously reported.[31]

Data analysis

Density plots were analyzed with software donated by EBS. Events were extracted using QUB 1.5.0.31 and fitted using Igor Pro 6.1. Individual translocation *i-t* traces were refiltered to 50 kHz for presentation.

Results and Discussion

Modification of AP sites with a crown ether adduct

It is thought that 10-15 nucleotides contribute to the current blockage signal when ssDNA translocates through the α-HL ion channel;[31] thus, a single AP lesion is unlikely to be resolved from the native nucleotides via a modest change in the current amplitude. The goals of chemical modification of DNA bases are both to amplify the signal differences and to stabilize the AP site against strand breaks. AP sites tautomerize between the ring-closed hemi-acetal form and the ring-opened aldehyde form; the latter offers the possibility of functionalization via reductive amination coupling with 2-aminomethyl-18-crown-6 to form the stable adduct AP-18c6 (Figure 2.2). In the present work, uridine-containing oligodeoxynucleotides (ODNs) were treated with uracil-DNA glycosylase (UDG) to generate AP sites. With the possible exception of trace amounts of 5-formyl-pyrimidines, oxidation products of 5-methylC and thymidine,[32] the AP site

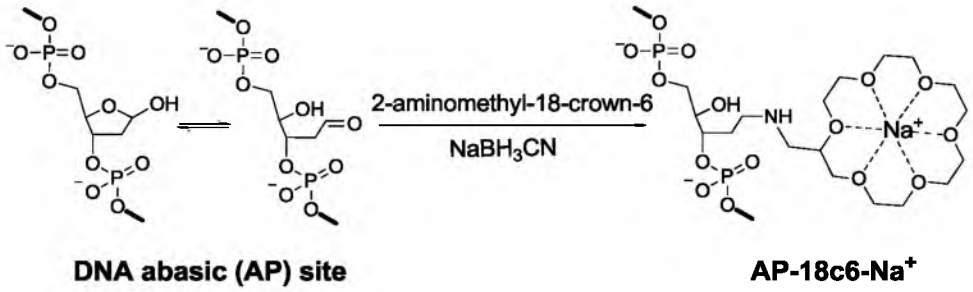


Figure 2.2. Synthetic scheme for AP-18c6 adduct.

presents the only aldehyde group existing in DNA, and reductive amination occurs in very high yield and specifically at the AP site using $NaBH_3CN$ in the presence of a primary amine. We chose 2-aminomethyl-18c6 as the amine of choice due to its water solubility and ability to interconvert between a large, rigid, disc-like structure when bound to alkali metal ions and a flexible, collapsed one when the metal ion dissociates.

Ion channel recording of immobilized AP and AP-18c6

As a preliminary study, the electrical current signatures of the AP site and AP-18c6 were measured in a static experiment in which ssDNA was immobilized inside the protein nanopore using streptavidin-biotin (Strep-Btn) complex formation, suspending ssDNA in the ion channel to accurately measure the current blockage level.[31, 33, 34] A voltage (120 mV *trans* vs. *cis*) applied across a lipid bilayer spanning the orifice of a glass nanopore membrane (GNM) drove the ssDNA to enter the *cis* side of the protein nanopore, reducing the open channel current I_o to a lower level I. Both I and I_o were recorded, and $\%I/I_o$ is quoted as the percentage residual current (Figure 2.3). The Strep-Btn ssDNA complex was retained for ~1 s; the polarities of the electrodes were then reversed for ~150 ms to remove DNA from the channel, and the voltage polarity was switched back again for the next event.

The 5-nm long β barrel of wild-type α-HL possesses multiple recognition sites that simultaneously contribute to the residual current, and one of the most current-sensitive regions resides at position 14 of the DNA strand relative to the biotin linkage.[31] Consequently, AP and AP-18c6 were placed at this position in the immobilization analysis. More than 200 capture/release events were collected for each sample, and the

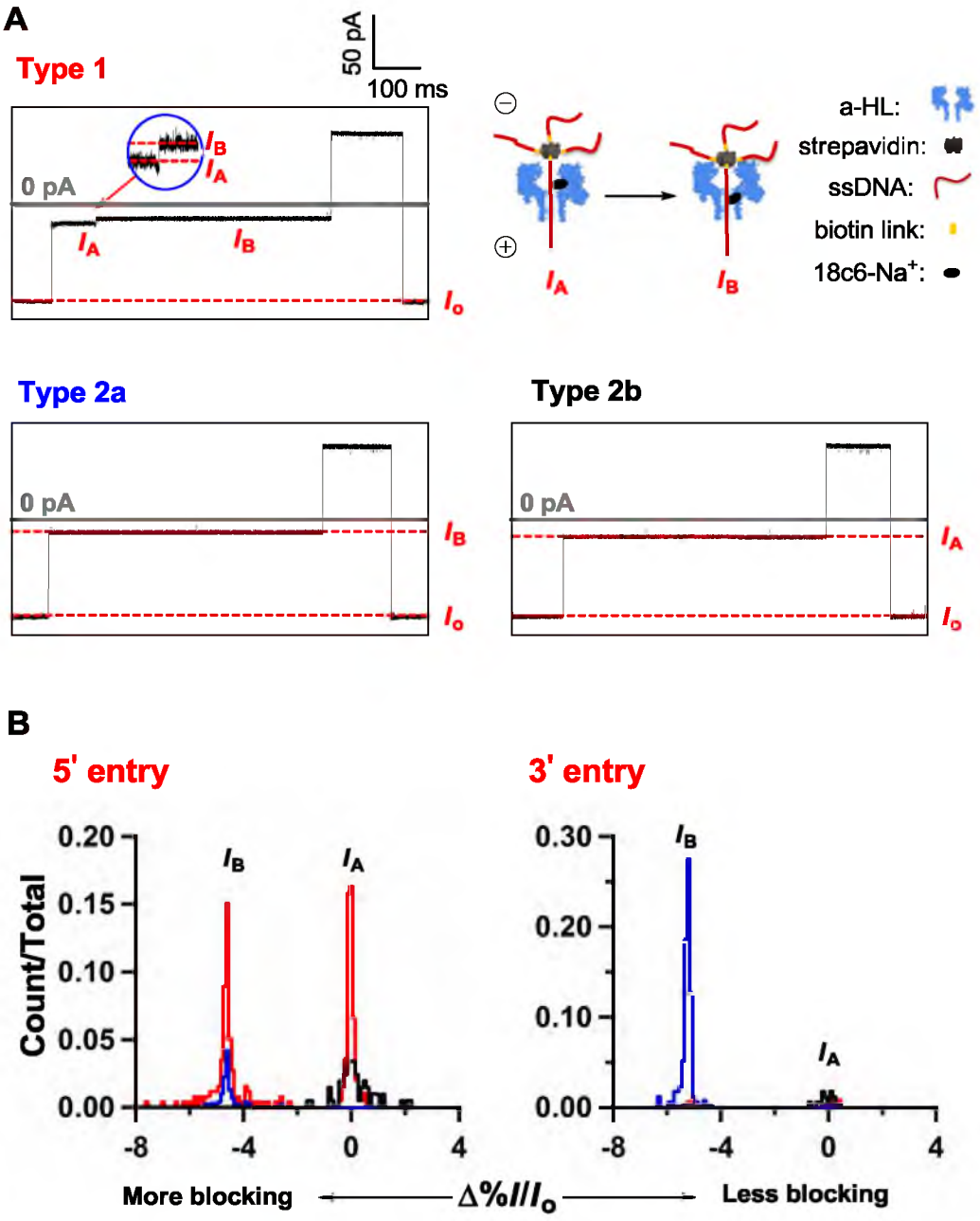


Figure 2.3. Immobilization studies of AP-18c6 with 1 M NaCl as electrolyte. (*A*) Three types of *i-t* traces presented by AP-18c6. (*B*) Histograms of current blockage levels of both 3′ and 5′ entries. Color code: red (type 1), blue (type 2a), and black (type 2b).

5′- or 3′-biotinylated poly-C_{40} (5′- or 3′-Btn-C_{40}), whose $\%I/I_o$ values were set to 0, were used as the internal standards for the corresponding substituted strands.

A single AP lesion was generated by treating 3′-Btn $C_{39}U_{\omega 14}$ with UDG, where ω14 represented the position of the substituted base, and current measurements for $\Delta\%I/I_o$ were performed immediately afterwards. The AP site without an adduct showed an average current level in 1 M KCl that was 1.1% less blocking than the control 3′-Btn-C_{40} (Figure 2.4*A*). Previous efforts suggested that the $\Delta\%I/I_o$ of DNA native bases fall into the range of 0-1.2% $\Delta\%I/I_o$ relative to C,[21, 31] and thus unfortunately, the AP site yielded a current blockage almost identical to G under these circumstances. This suggests that an AP site would produce a false readout as G in sequencing efforts with wild-type α-HL, and given the multiple sources and abundance of AP sites in cells, this would be problematic.

The behaviors of ssDNA strands are rather different with respect to directionality when entering the protein channel, in terms of both event frequency and current blockage [35, 36]. Thus, the current signatures of both 5′-Btn-C_{39}AP-18c6$_{14}$ (3′ entry) and 3′-Btn-C_{39}AP-18c6$_{\omega 14}$ (5′ entry) were recorded with 5′- and 3′-Btn-C_{40} as the references, respectively, in 1 M NaCl. Unlike the immobilization events of unmodified ssDNA and other adducts we have previously studied, AP-18c6 presented two different types of individual *i-t* traces (Figure 2.3*A*). Type 2 events exhibited a constant current level of magnitude I_A or I_B that indicated the capture of ssDNA. More interestingly, a transition between two current blockage levels ($I_A \rightarrow I_B$) was observed in type 1 events. Given that I_A had the same residual current as the control poly-dC_{40} ($\Delta\%I_A/I_o = 0$), this result suggests that the initial current level I_A in type 1 events corresponds to the 18c6-Na^+

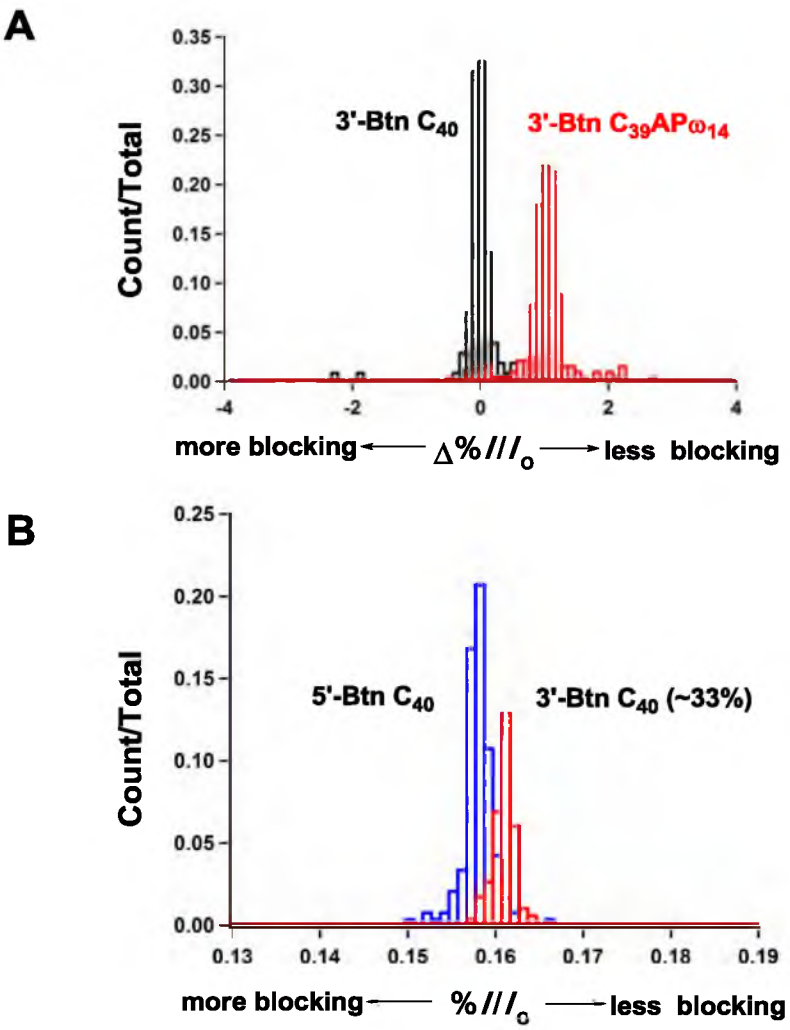


Figure 2.4. Immobilization studies of AP sites and directionality effects (*A*) Current histogram of 3′-Btn $C_{39}AP_{\omega 14}$ with 3′-Btn C_{40} as control in 1 M KCl. (*B*) Current histogram of 5′-Btn C_{40} and 3′-Btn C_{40} (200 nM of each) in 3 M NaCl.

complex hesitating above the constriction of α-HL, while the sensing β barrel was recording signals of the poly-dC part of the strand. The 18c6-Na^+ complex can convert between several conformations of similar shape, one of which (IC_1) features a tight coordination associated with a significantly shorter O-Na^+ distance than the quasiplanar symmetric D_{3d} conformer that 18c6-K^+ almost exclusively adopts. In the Na^+ complex, 18c6 is curved to provide a balance between the O-Na^+ attraction and O-O repulsion, resulting in a more compact shape. The other conformer is D_{3d} with a lower stability constant (K_s) due to the longer distance between oxygen atoms and Na^+ compared to 18c6-K^+.[37] We therefore hypothesize that after entering the vestibule, a period of time is required for 18c6-Na^+ to optimize its conformation, allowing it to enter into the narrow β barrel and position itself at the recognition site, causing a much deeper current blockage level (I_B, $\Delta\%I_A/I_o \sim$ -4.6%) that lasts for the remainder of the event before voltage reversal (type 1). As demonstrated below, the $I_A \rightarrow I_B$ transition provides a signature for detecting 18c6 adducts during translocation of DNA. Type 2a events correspond to entry of 18c6-Na^+ into the vestibule of α-HL with an orientation that allowed it to pass into the constriction without hesitation, resulting immediately in current level I_B. Contrarily, type 2b events correspond to the adducted complex located at a position such that interactions between the protein, 18c6 and Na^+ prevented it from entering into the β barrel before the reversal of polarity; the poly-dC portion of the DNA strand was solely contributing to the signal resulting in current level I_A.

Similarly to unmodified ssDNA, the directionality of AP-18c6 entering into the protein channel also affects its current signature (Figure 2.3*B*) with 3′ entry displaying a smaller residual I_B ($\Delta\%I_A/I_o \sim$ -5.3%). Many more 5′ entry events (58%) presented the

desirable two-step current feature (type 1) compared to those of 3′ entry (< 2%). In nearly 90% of 3′ entry events, the 18c6-Na^+ complex appears to pass immediately through the constriction into the β barrel (type 2a), while only 12% of 5′ entry events were able to do so. Also, there were many AP-18c6 events (30%) that could not enter the β barrel from the 5′ end (type 2b); yet, only ~8% of the events had the same problem from the other terminus. These directionality effects can be explained by the tendency of ssDNA bases to tilt towards the 5′ end in a strongly confined environment, producing the "Christmas tree" geometry.[36] As a result, DNA strands entering from the 5′ direction experience more friction from the protein, along with steric hindrance from the bulky 18c6-Na^+, causing more type 2b events to occur than during 3′ entry. The direction of DNA entry may position the crown ether differently relative to the protein constriction, and slight differences in the position/orientation of 18c6 would likely change its interaction with the protein, thus affecting the current signature during translocation.

Immobilized DNA measurements were also performed in 1 M KCl and LiCl to interrogate the interactions between 18c6 and cations as an approach to altering the electrical signature of AP-18c6, and also to provide insight into the correlation between these interactions and the electrical signals. Because 5′ entry produced a more sensitive and desirable signature as described above, the current blockages of 3′-Btn C_{39}AP-$18c6_{\omega 14}$ were recorded in 1 M KCl and LiCl, respectively, and not surprisingly, the two 18c6-M^+ experiments showed very distinct $\Delta\%I/I_o$ histograms compared to 18c6-Na^+ (Figure 2.5). While 18c6-Li^+ presented a single $\Delta\%I/I_o$ peak, 2.1 % more blocking than the control, 18c6-K^+ had one broader distribution of $\Delta\%I/I_o$, centered at –4.1 %, and one peak lining up with the 3'-Btn-dC_{40} control. These observations can be explained as

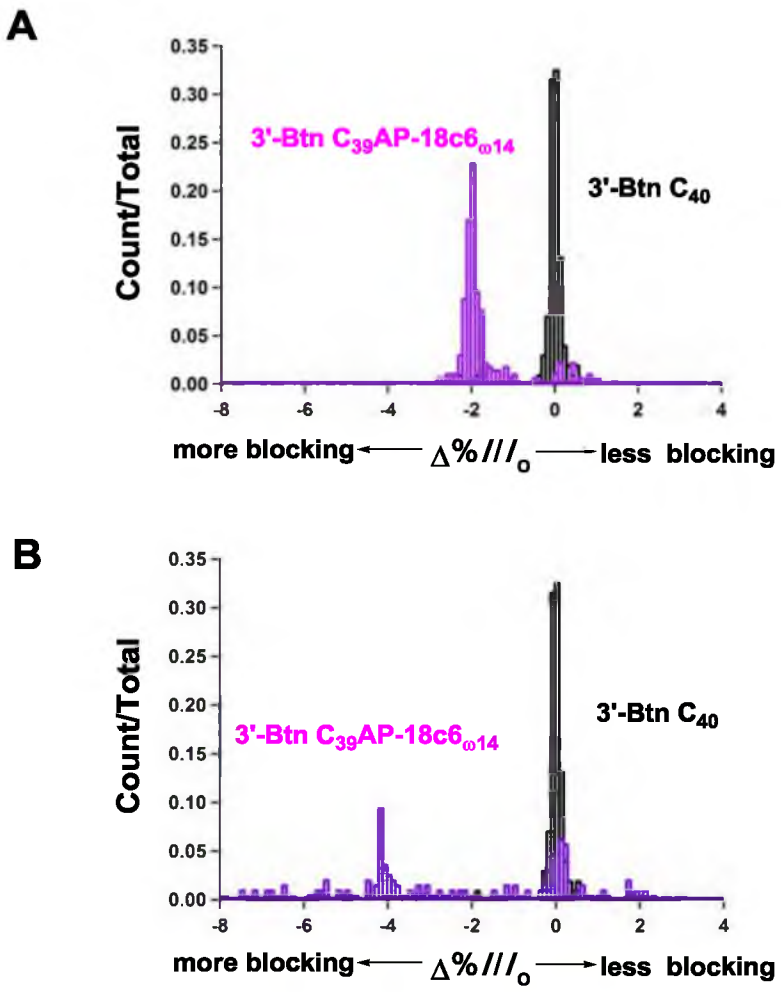


Figure 2.5. Current histograms of 3′-Btn C_{39}AP-18c6$_{\omega 14}$ with 3′-Btn C_{40} as control in 1 M (*A*) LiCl (*B*) KCl.

follows. Due to its small size and high desolvation energy, Li^+ does not form a very stable complex with 18c6.[38] Thus, the slightly deeper current blockage of the AP-18c6 adduct observed in LiCl electrolyte compared to unmodified DNA is likely due to the added steric bulk of the uncomplexed crown ether in an elliptical conformation. In the case of K^+, the 18c6-K^+ complex exists as a single quasiplanar symmetric D_{3d} conformer, resulting from the ideal size matching between the cation and the polyether cavity. Additionally, this conformer provides ideal solvation, and consequently, stabilizes the system and maximizes its size (~ 5.3 Å in radius).[37] Given the high K^+ concentration and the high energy cost for 18c6-K^+ to change conformation (>13 kJ mol^{-1}), we anticipated that the passage of the 18c6-K^+ through the constriction of the α-HL would be difficult. In Figure 2.5*B*, only the poly-dC tail is responsible for the small current blockage peak at $\Delta\%I/I_o = 0$, equivalent to an unmodified 3′-Btn-C_{40} strand with the 18c6-K^+ complex sitting above the constriction. The broader peak of the histogram, which contains the majority of the events, suggests that when the metal complex is forced into the constriction zone, it is not in a reproducible orientation that gives a single deep current blockage. Together, these results demonstrate that the interactions between 18c6 and electrolyte cations can be correlated to their electrical signatures, and that the unique two-level current feature of 18c6-Na^+ can be used to detect AP sites at a single-molecule level.

Translocation studies of AP-18c6 in a homopolymeric strand

Immobilization studies allowed us to closely examine the potential of detecting AP lesions using the characteristic current signature obtained by the interaction between

18c6 and various alkali metal cations. In the series, Li^+ is too weakly complexed by 18c6 to affect the current signature, K^+ is too strongly bound by 18c6 leading to a rigid complex that is too large to reproducibly enter the constriction of the ion channel, and Na^+ is "just right," demonstrating both a step change in current blockage level and a readily measurable delay time for proper entry into the β barrel. Next, we examined the behavior of ssDNA while translocating through the α-HL ion channel. To be consistent with the immobilization studies, individual AP-18c6 were also embedded in a poly-dC context, and two DNA constructs were studied: a mono adduct 5′-d($C_{43}XC_{43}$) and a bis adduct 5'-d($C_{25}XC_{35}XC_{25}$), where X = AP-18c6. Measurements were performed in a buffer containing 3 M NaCl as the electrolyte over a range of applied voltages (80, 100, 120, 160 mV *trans* vs. *cis*).

The translocation of unmodified homopolymeric strand dC_{87} was investigated first as a control, focusing on current level (*I*), duration time (t_D) and directionality effects. Individual *i-t* traces from either 3' or 5' entry displayed a single constant current during translocation with t_D ~0.25 ms, consistent with a translocation velocity of ~4 μs/nucleotide (Figure 2.6*A*). Two populations of events were observed in the *I* vs. t_D density plots, (Figure 2.6*B*); in accordance with the immobilization studies of 5′-Btn-C_{40} and 3′-Btn-C_{40} under the same conditions (%I/I_o = 15.6 and 16.2, respectively), 3′ entry of this homopolymer produced a deeper current blockage (Figure 2.4*B*). Thus, these clusters were assigned to represent 3′ and 5′ entry of the strand into the protein vestibule, respectively. Consistent with diffusional broadening of translocation times, the t_D histogram of each population was well described by a Gaussian curve with the peak value t_{max}, which decreased with increased applied voltages for both directions (Figure 2.6*C*).

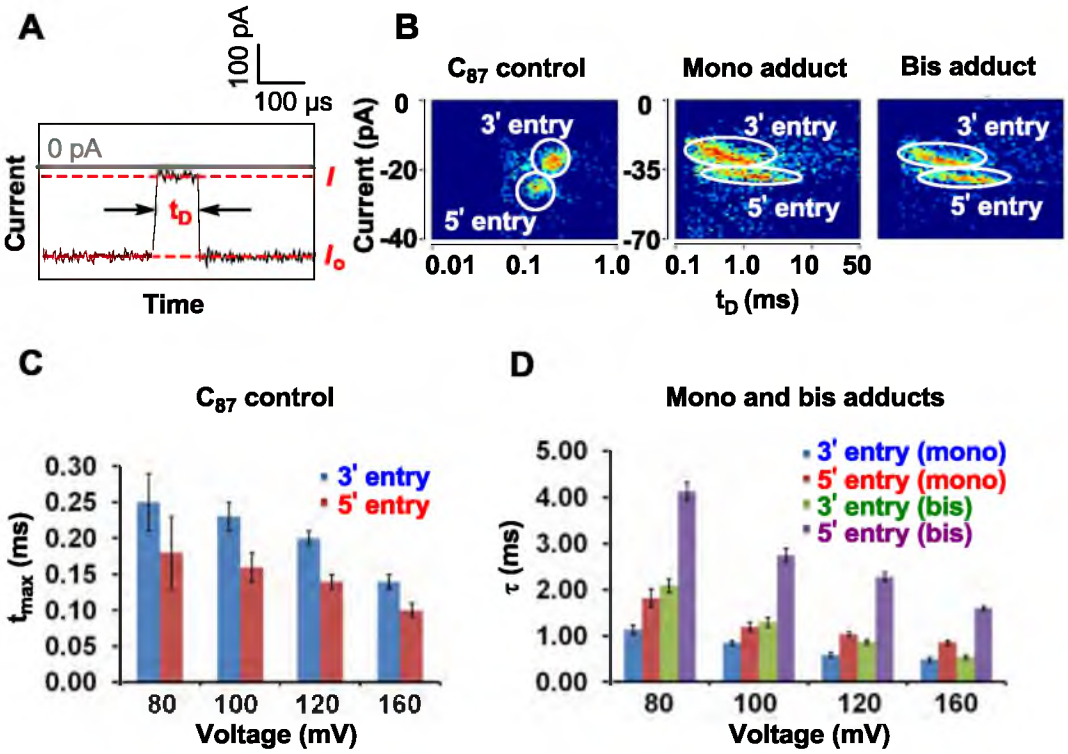


Figure 2.6. Translocation studies of AP-18c6 in homopolymeric strands with 3 M NaCl as the electrolyte. (*A*) Typical *i-t* trace of the control strand and definition of translocation duration (t_D). (*B*) Sample density plots of control and adducted strands (120 mV *trans* vs. *cis*). (*C*) Plot of t_{max} versus voltage for the control strand. (*D*) Plot of τ versus voltage for the mono and bis adducts.

Entry from the 3' terminus was favored under lower driving force, resulting in higher event frequencies, while the slightly faster 5′ population grew with increasing voltage until such discrimination faded around 160 mV (*trans* vs. *cis*).

Similarly, both mono and bis adduct strands also showed distinct directionality effects with 3′ entry events occurring more frequently under low voltages and generating deeper current blockages (Figure 2.6*B*). However, the population clusters of both directions presented more oval-shaped distributions that were better approximated by a single exponential decay with a time constant $\tau_{5'}$ or $\tau_{3'}$ (5′ or 3′ entry, respectively) of several milliseconds, indicating that the slow conformational motion of AP-18c6 required for it to pass through the constriction region of α-HL determines the overall translocation time. In contrast to the values for the C_{87} control, $\tau_{5'}$ was considerably longer than $\tau_{3'}$ for both mono and bis adducted strands (Figure 2.6*D*).

Individual *i-t* traces of both mono and bis-adduct strands exhibited rather different features depending on the entry direction (Figure 2.7). For 3′ entry, the presence of the AP-18c6 adduct was not evident in the shape of the *i-t* trace (Figure 2.7*A*). However, *i-t* traces with unique pulse-shaped signatures were observed upon translocation of AP-18c6 from the 5′ terminus (Figure 2.7*B*). Taking the *i-t* trace of the mono adduct at 120 mV as an example, an initial current blockage level of I_{A1} was observed with characteristic exponential time distribution (~2.5 ms, Table 2.1). Given that the value of $\%I_{A1}/I_o$ was consistent with the percentage residual current of the immobilized 3′-Btn-C_{40} ($\%I/I_o$ = 16.2%), this segment of the *i-t* trace is assigned to be the temporary pausing of poly-dC in the narrow constriction zone of the ion channel, while 18c6-Na^+ remains in the larger

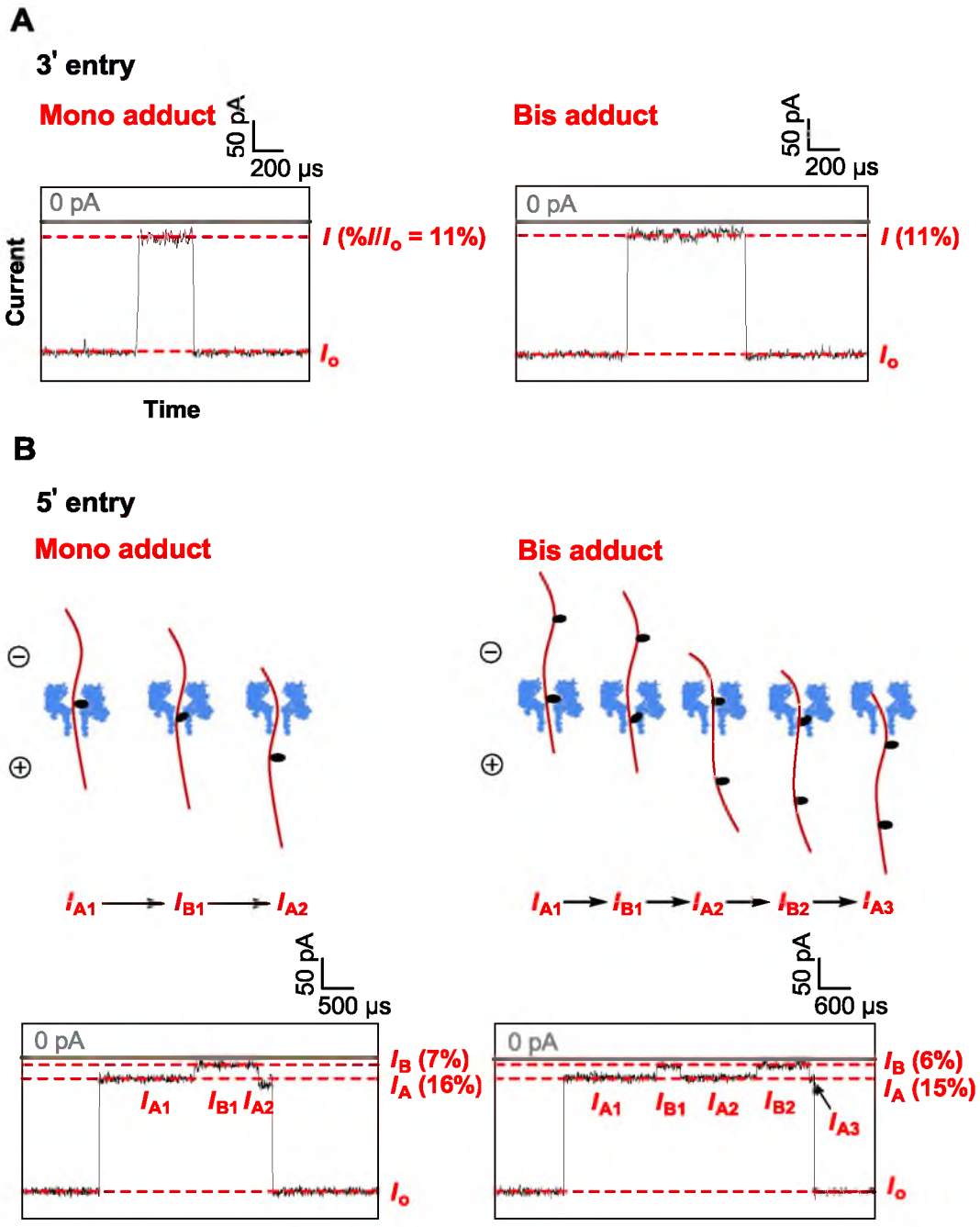


Figure 2.7. Individual i-t traces of AP-18c6 in homopolymeric strands. (*A*) Sample *i-t* traces of 3′ entry for mono and bis adducts (120 mV *trans* vs. *cis*). (*B*) Sample *i-t* traces of 5′ entry for mono and bis adducts (120 mV *trans* vs. *cis*).

vestibule, waiting to undergo ion dissociation and/or conformational change that would permit entry into the narrower, sensing part of the channel. The mid-section of the event (I_{B1}) showed a nearly 9% deeper current blockage, and exponential decay constant of ~ 0.59 ms, which we interpret to be associated with the bulky crown ether moving through the narrow β barrel. Notably, the percentage residual current changes ($\%I_A/I_o$-$\%I_B/I_o$) in the presence of 18c6 adduct during translocation fell into the same range compared to the immobilization experiment (~4.6%), which statically positioned the adduct at a single location. Finally, when the AP-18c6 exited the constriction zone, the current level returned to that of poly-dC (I_{A2}) for a short period of time while the C_{43} tail completed translocation, resulting in an overall $I_{A1}\rightarrow I_{B1}\rightarrow I_{A2}$ pulse shape. The short I_{A2} segment was well described by a Gaussian distribution of times, with t_{max} ~ 0.06 ms, consistent with a free translocation velocity of ~1.5 μs/nucleotide. For the bis-adduct, the 35 cytosines between two AP-18c6 adducts were sufficient to position the first adduct out of the *trans* side of α-HL, while the second trailing adduct undergoes conformational changes and passes through the constriction, causing a second current modulation ($I_{A2}\rightarrow I_{B2}\rightarrow I_{A3}$). Overall the bis adduct displays two pulse-shaped modulations ($I_{A1}\rightarrow I_{B2}\rightarrow I_{A2}\rightarrow I_{B2}\rightarrow I_{A3}$), corresponding to the sequential translocation of the two AP-18c6 adducts (Figure 2.7*B*).

Consistent with the immobilization studies, not every 5′ entry event of AP-18c6 adduct has to hesitate for a long time to pass the constriction if 18c6-Na^+ enters into the vestibule with the proper orientation and conformation. Accordingly, 20% of the mono-adduct events did not have a resolvable pulse-like modulation in the current, nor did 5% of the bis-adduct events. Additionally, ~26 % of the bis-adduct events showed only one

current modulation, similar to the mono-adduct, but with slightly longer duration, indicating one of the adducts had the proper orientation and conformation to pass unimpeded through the constriction. Translocation studies were performed at different voltages (80, 100, 120, 160 mV *trans* vs. *cis*) in order to verify that the adducted strands did successfully reach the trans side of α-HL instead of being held in the vestibule and diffusing back to the *cis* side. We observed that the duration of both directions was inversely correlated with the driving force (Figure 2.6*D*), suggesting full translocations of the 18c6-Na^+ adducted DNA.

In the above analyses, the duration of the current segments I_{A1} (for the mono adduct) and I_{A1} and I_{A2} (for the bis adducts) are associated with 18c6-Na^+ conformational changes, including the likely dissociation of the complex, prior to 18c6 moving through the constriction of α-HL. At 120 mV, these times are 2.50 ms (I_{A1}) for the mono adduct, and 2.75 (I_{A1}) and 1.78 ms (I_{A2}) for the bis adduct (Table 2.1), much longer than the value expected (2.9×10^{-8} s) based on the Na^+ dissociation rate constant in bulk solution.[39] This five order of magnitude difference indicates that the constricted geometry of the ion channel dramatically decreases the dissociation of the 18c6-Na^+ complex.

Translocation studies of AP-18c6 in a heteropolymeric strand

Both immobilization and translocation studies showed promising results for using the interactions between 18c6 and electrolyte cations to detect individual AP sites in a homopolymeric oligo-dC strand. Next, we further explored the sequence context effect on the adduct behavior by embedding one AP-18c6 adduct in a heteropolymeric DNA

Table 2.1. Kinetic analysis of current segments from 5′ entry of various adducts[1].

DNA	Voltage (mV) *trans* vs. *cis*	I_{A1} (τ/ms)	I_{B1} (τ/ms)	I_{A2} (ms)	I_{B2} (τ/ms)	I_{A3} (t_{max}/ms)
Mono adduct	120	2.50	0.59	t_{max}= 0.06	NA	NA
Mono adduct	160	0.92	0.23	t_{max}= 0.04	NA	NA
Bis adduct	120	2.75	0.32	τ = 1.78	0.82	0.05
Bis adduct	160	1.04	0.12	τ = 0.83	0.48	0.02
Heterosequence adduct	120	2.90	0.71	t_{max}= 0.15	NA	NA
Heterosequence adduct	160	1.32	0.35	t_{max}= 0.09	NA	NA

[1] Kinetics values have an error of ~10%.

strand that was designed to be free of secondary structure for simplicity (Figure 2.8*A*). Similarly to the homopolymeric context, both adducted and unmodified strands showed a distinct directionality effect with respect to current blockages, and translocation times that decreased at higher voltages (Figure 2.8*B* and *D*). The more dynamic and flexible secondary structures of the heterosequence, compared to the stick-like structure proposed for a poly-dC strand,[40] results in an intermediate current level I_M before the deep blockage in the individual *i-t* traces. As previously reported, the intermediate level I_M corresponds to the presence of DNA in the large vestibule, exploring orientations to thread into the β barrel (Figure 2.8*C*).[23] Most significantly, the 5′ entry events still exhibited the pulse-shaped current modulation ($I_{A1} \rightarrow I_B \rightarrow I_{A2}$) due to the presence of the AP-18c6 adduct. These results are important as a demonstration that the current amplitude signature of AP-18c6 is due to its unique motif and interactions with the electrolyte cations; thus, it can potentially serve as a universal tool for AP site identification, independent of sequence context.

Manipulation of AP-18c6 behaviors by altering the electrolyte

The distinct behaviors of AP-18c6 with different electrolytes in the immobilization studies provided us with insight into the correlation between 18c6-M^+ interaction and its electrical signal, and these specific properties also presented the possibility to tune the current signature of the adduct by altering the electrolyte contents. In addition to NaCl, translocation studies of the mono adduct (5′ $C_{43}XC_{43}$, X = AP-18c6) were also performed in KCl and LiCl, and consistent with the immobilization results, 18c6-K^+ was not able to reproducibly enter the constriction zone due to its stable D_{3d}

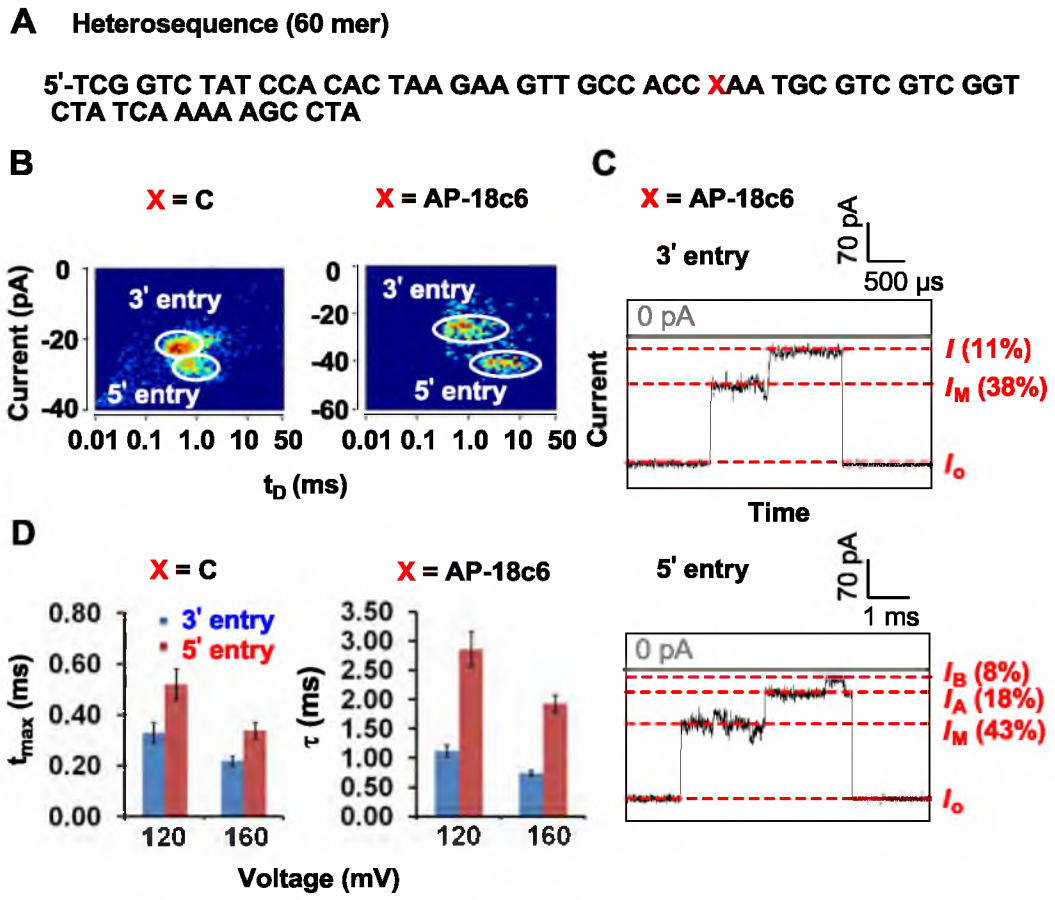


Figure 2.8. Translocation studies of AP-18c6 in heteropolymeric sequence with 3 M NaCl as electrolyte. (*A*) Heteropolymeric sequence (60 mer). (*B*) Sample density plots of control (X = C) and adducted strand (X = AP-18c6) (120 mV *trans* vs. *cis*). (*C*) Individual *i-t* traces of the adducted strand. Top: 3′ entry event. Bottom: 5′ entry event. (*D*) Left: plot of t_{max} vs. voltage for the control strand. Right: plot of τ vs. voltage for adducted strand.

conformation and the high [K^+]. At higher voltages, the 18c6-K^+ could be held in the vestibule and released only upon the reversal of the polarities, causing 'immobilization-like' *i-t* traces (Figure 2.9*A*). On the other hand, 18c6-Li^+ showed no barrier in translocating through the channel, leaving unidentifiable current signatures (Figure 2.9*B*).

Next, the same experiments were performed in LiCl electrolytes doped with varying amounts of KCl. Given the selectively strong interaction between 18c6 and K^+ and the much weaker binding affinity of 18c6 towards Li^+, the latter could be neglected; thus, the fractional percentage of K^+-bound 18c6 could be calculated based on the equilibrium stability constant (K_s = 114.8 M^{-1}) under these conditions (Figure 2.10).[39] With a low driving force (80 mV *trans* vs. *cis*), three distinct populations of events (A, B and C) appeared with different durations (Figure 2.10), and the ratio of event counts changed with the electrolyte content. While populations A and B showed normal translocation *i-t* traces, C presented "immobilization-like" ones. Based on the results obtained in the pure LiCl electrolyte, population A is assigned to the K^+-free AP-18c6. Populations B and C were much longer, and their combined percentage event counts, for the fractional percentage of K^+ ranging from 0.1 to 20%, were in excellent agreement with the percentage of K^+-bound 18c6 calculated based on K_s (Figure 2.10), suggesting they were both due to 18c6-K^+ complexes. Because the ratios of population C to B were always roughly 3:1 under these conditions, we suggest that they correspond to the directionality effect with B being 3′ and C being 5′ entry of the DNA strand into the vestibule. The agreement between nanopore-measured and K_s–based calculated values of the percentage of K^+-bound 18c6, coupled with the above-noted orders of magnitude slower dissociation time of the Na^+-complex in the nanopore relative to bulk solution,

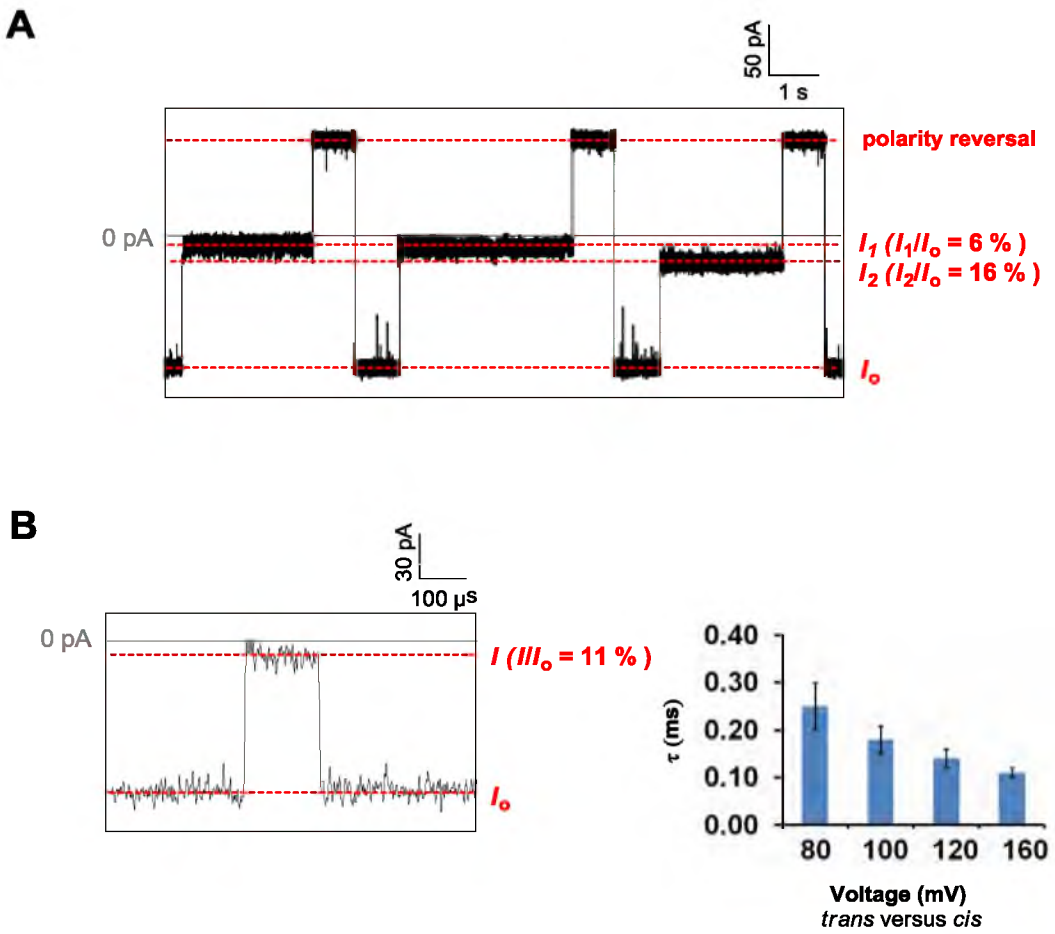


Figure 2.9. Translocation studies of 18c6-K^+ and 18c6-Li^+. (*A*) Sample *i-t* trace of mono adduct translocation in 1 M KCl under 120 mV (*trans* vs. *cis*). (*B*) Free translocation of the mono adduct in 1 M LiCl. Left: sample *i-t* trace under 120 mV (*trans* vs. *cis*); Right: plot of time constant τ (ms) versus applied voltages.

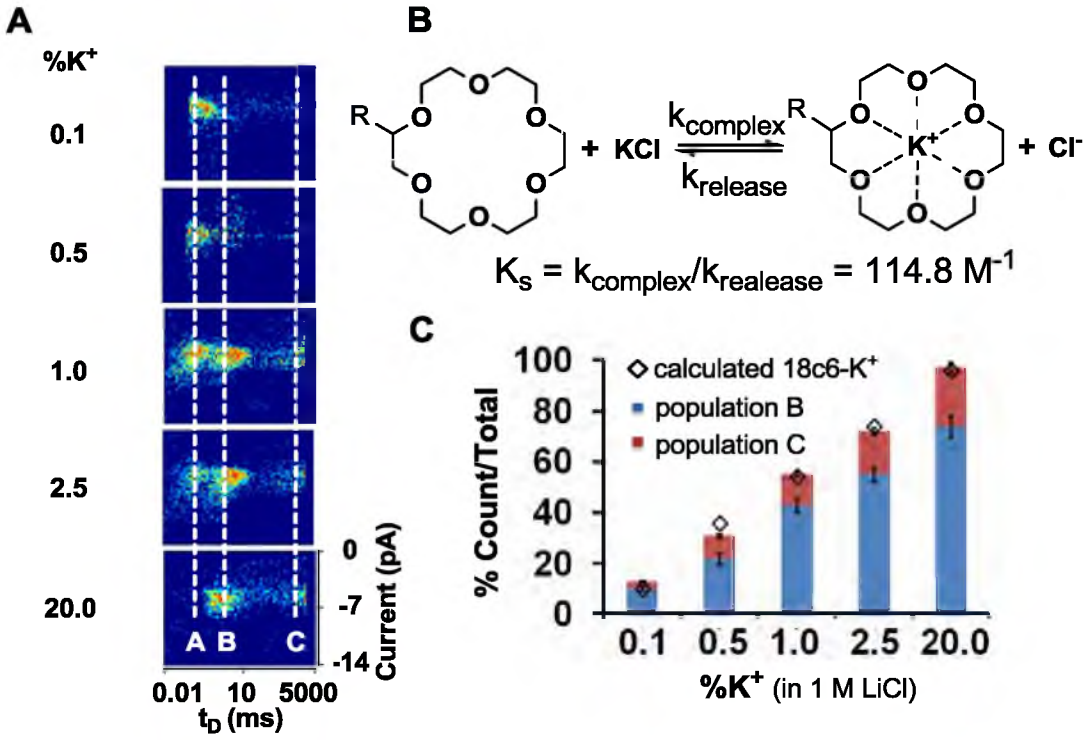


Figure 2.10. Electrolyte-dependent translocation studies of the mono adduct strand. (*A*) Density plots of various $\%K^+$ (in 1 M LiCl) conditions under 80 mV (*trans* vs. *cis*). (*B*) Complexation and dissociation of 18c6-Na^+ (R = DNA) with the equilibrium constant K_s [39]. (*C*) Plot of combined percentage event counts of populations B and C vs. $\%K^+$ (in 1 M LiCl), compared to calculated values.

indicate that the measured values of the percentages of K^+-bound 18c6 are determined by equilibrium concentrations of bound and unbound 18c6 external to the α-HL ion channel.

In summary, these results demonstrate a crucial role for 18c6-M^+ interactions in producing electrical signatures of 18c6-adducted DNA while translocating through a protein nanopore, and also provide a promising tool to manipulate the behavior of the adduct precisely on a single-molecule level. By attaching 18c6 to an AP site, the electrical signature of ssDNA translocation could be dramatically modified. Specifically, pulse-like current modulations during translocation of DNA caused by the presence of individual AP-18c6 adducts demonstrate the possible application of this method as a detection tool for multiple AP sites on single molecules of DNA. Employing the same chemical strategy combined with the assistance of lesion-specific glycosylases, one can envision that a variety of DNA base damages and mismatches can be specifically converted to AP sites and then functionalized to the AP-18c6 adduct, permitting their sequence-specific detection in single molecules.

Translocation studies of AP-18c6 adduct in a 120 mer heteropolymeric strand

A major challenge using α-hemolysin (α-HL) nanopore arises from the fast speed at which ssDNA is translocated, resulting in the requirement of fast monitoring of small currents (pA). Thus, site-specific chemical reactions are being explored to slow down the ssDNA movement, as well as to generate unique current signatures. Another difficulty of analyzing ssDNA with this method is the secondary structure mediated by base-paring and base-stacking that are prevalent in natural samples. As a result, long reads of ssDNA

with heterosequences commonly adopt multiple conformations that are too stable to be unraveled or threaded into the protein channel.

In order to test the feasibility to measure biological samples, a 120 mer ssDNA with heterosequence was synthesized and driven through the α-HL ion channel (Figure 2.11*A*). First, the ssDNA was designed not to form any stable hairpins or possess any self complementary regions to simplify the situation and the successful translocation rate was evaluated. Under standard conditions, it presented two different event types I and II, both of which initiated with long intermediate current blockages, while type I ended with a deep blockage, indicating successful translocation (Figure 2.11*B*). Due to the highly dynamic and flexible structure of the long heterosequence, only 9% of the captured DNA was able to thread its terminus into the β barrel and travel through the channel to the *trans* side of the protein before diffusion drives it back to the *cis* side as shown in type II. In addition, the successful translocation rate was not affected significantly with respect to the voltages.

Different conditions were explored to facilitate the translocation, including additions of the denaturing agent urea, a short complementary strand, and making chemical modifications to DNA. The percentage translocation rates are shown in Figure 2.11*C*. The most dramatic effect came from the addition of a short complementary strand, which could be explained by the fact that forming an orderly duplex structure at the center of the DNA strand helps to expose its termini to be threaded into the channel. Surprisingly, 2 M urea did not seem to change the translocation rate. And according to published results, higher concentrations of urea (>4 M) might be needed to fully denature the secondary structure; however, the overall event frequency would considerably

A **120 mer heterosequence**

5'- AGT TGC CAC CTA ATG CGT CGT CGG TCT AXC AAA AAG CCT ACA CAG AAA AAT CA*G TTG TCG TCG GTC T*AT CCA CAC TAA GAA GTT GCC ACC XAA TGC GTC GTC GGT CTA TCA AAA AGC CTA

X = C or AP-18c6 adduct

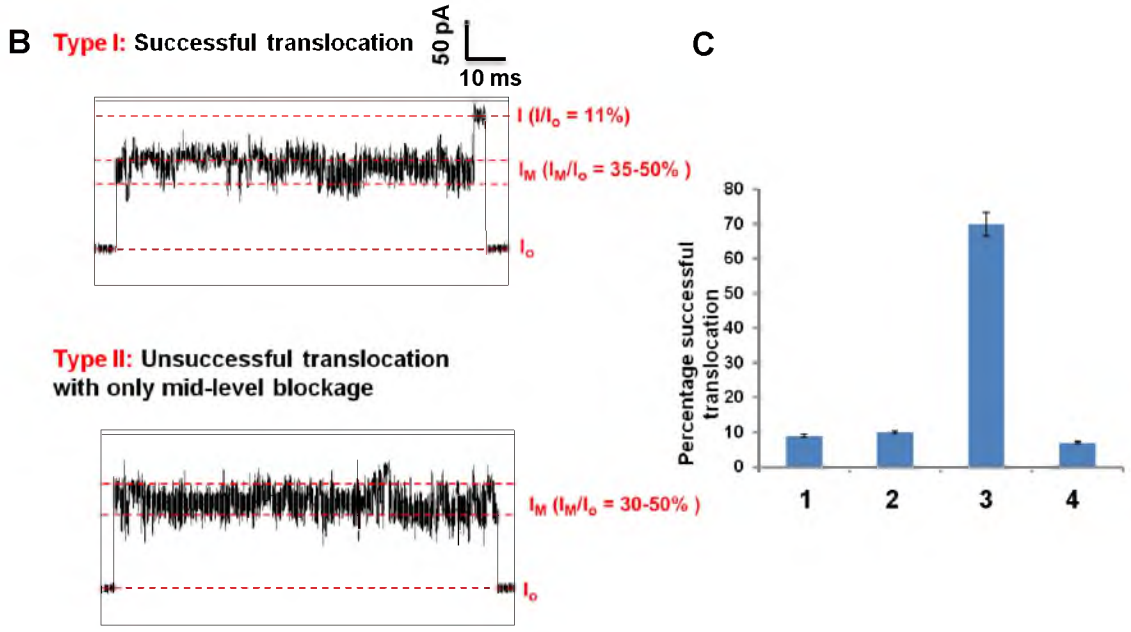


Figure 2.11. Translocation studies of 120 mer heterosequence ssDNA. (*A*) Sequence of the120 mer heterosequence ssDNA. (*B*) Two different event types. (*C*) Percentage successful translocation at 120 mV under different conditions: **1.** standard buffer (3 M NaCl 25 mM Tris pH 7.90, 1 mM EDTA). **2.** standard buffer with 2 M urea. **3.** standard buffer with 5 folds of a 14 mer complementary to the italic and underlined portion of 120 mer ssDNA in (*A*). **4.** bis 18-crown-6 adduct in standard buffer.

decrease.[41] At last, making two point modifications to DNA was not distorting enough to unravel the whole DNA strand.

Due to the threading difficulty caused by the stable secondary structures occurring in the heterosequence, the successful translocation events were studied under different conditions. The best results were obtained by annealing a 14mer complementary strand at the center of the 120mer, as discussed above. Therefore, translocation of the bis-adduct strand (X = AP-18c6) was measured after the annealing of said complementary strand to obtain a substantial number of events. About 65 % of events showed promising current modulations for adduct recognition, while the rest of them did not present resolvable signatures (Figure 2.12). The percentage of desirable events was much higher than with the poly-dC background, which may be due to the long stretch of polynucleotides between the adducts, and the delay imposed for unzipping of the short complement.

Conclusion

Site-specific attachment of 18c6 to AP sites presented unique current amplitude signatures that enabled the detection of this lesion at a single-molecule level in both homopolymeric and heteropolymeric sequences. Due to the various binding affinities of 18c6 towards different cations, the behaviors of the adduct could be precisely tuned by alternating the electrolyte contents. Additionally, a longer heteropolymer was studied to better understand the effect of secondary structures on the threading of ssDNA into the protein channel under several conditions. As a result, the successful translocation rate was greatly increased, providing the recognition of multiple lesions along the same strand.

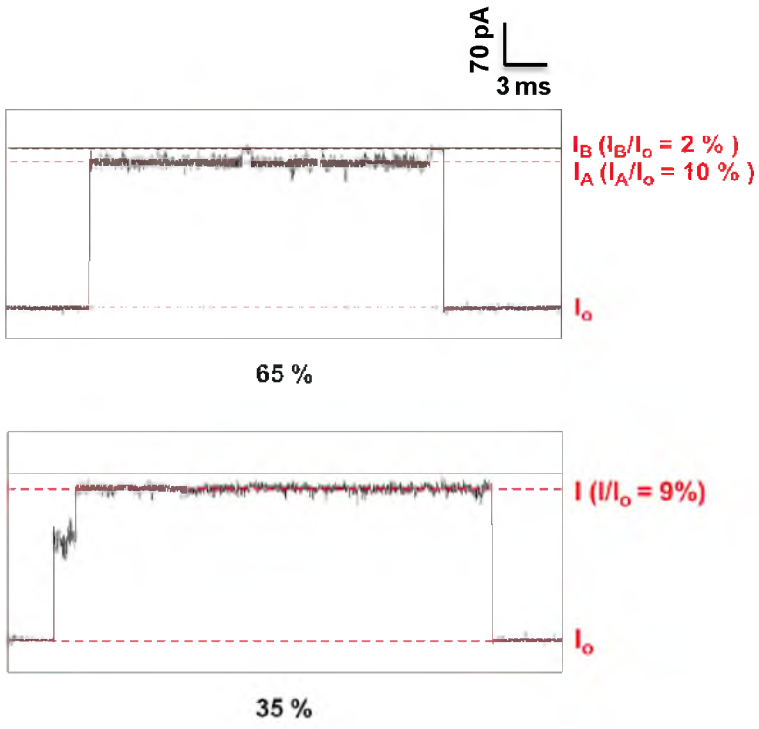


Figure 2.12. Sample *i-t* traces of 120mer bis-adduct translocation in 3 M NaCl (120 mV). About 65% of the traces show two events as in the top trace, while 35% look like the lower trace.

References

1. Boiteux, S.; Guillet, M. Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair* **2004,** 3, 1-12.

2. Wilson III, D. M.; Barsky, D. The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutat. Res.* **2001,** 485, 283-307.

3. Clauson, C. L.; Oestreich, K. J.; Austin, J. W.; Doetsch, P. W. Abasic sites and strand breaks in DNA cause transcriptional mutagenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** 107, 3657-3662.

4. Gates, K. S. An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol.* **2009,** 22, 1747-1760.

5. Kubo, K.; Ide, H.; Wallace, S. S.; Kow, Y. W. A novel, sensitive, and specific assay for abasic sites, the most commonly produced DNA lesion. *Biochemistry* **1992,** 31, 3703-3708.

6. Komatsu, Y.; Ohtsuka, E.; Mikami, A.; Takebayashi, T.; Kojima, N. Construction of highly reactive probes for abasic site detection by introduction of an aromatic and a guanidine residue into an aminooxy group. *J. Am. Chem. Soc.* **2009,** 131, 13208-13209.

7. Atamna, H.; Cheung, I.; Ames, B. N. A method for detecting abasic sites in living cells: Age-dependent changes in base excision repair. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** 97, 686-691.

8. Greco, N. J.; Tor, Y. Simple fluorescent pyrimidine analogues detect the presence of DNA abasic sites. *J. Am. Chem. Soc.* **2005,** 127, 10784-10785.

9. Sun, H. B.; Qian, L.; Yokota, H. Detection of abasic sites on individual DNA molecules using atomic force microscopy. *Anal. Chem.* **2001,** 73, 2229-2232.

10. Zeglis, B. M.; Boland, J. A.; Barton, J. K. Targeting abasic sites and single base bulges in DNA with metalloinsertors. *J. Am. Chem. Soc.* **2008,** 130, 7530-7531.

11. Bowler, F. R.; Diaz-Mochon, J. J.; Swift, M. D.; Bradley, M. DNA analysis by dynamic chemistry. *Angew. Chem. Int. Ed.* **2010,** 49, 1809-1812.

12. Wang, Y.; Liu, L.; Wu, C.; Bulgar, A.; Somoza, E.; Zhu, W.; Gerson, S. L. Direct detection and quantification of abasic sites for in vivo studies of DNA damage and repair. *Nucl. Med. Biol.* **2009,** 36, 975-983.

13. Sutherland, B. M.; Hada, M. Spectrum of complex DNA damages depends on the incident radiation. *Radiat. Res.* **2006,** 165, 223-230.

14. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Ventra, M. D.; Garaj, S.; Hibbs, A.; Huang, X.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **2008,** 26, 1146-1153.

15. Deamer, D. W.; Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* **2002,** 35, 817-825.

16. Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of straphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **1996,** 274, 1859-1866.

17. Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** 93, 13770-13773.

18. Clarke, J.; Wu, H.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **2009,** 4, 265-270.

19. Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** 97, 1079-1084.

20. Wallace, E. V. B.; Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Donohoe, T. J.; Bayley, H. Identification of epigenetic DNA modifications with a protein nanopore. *Chem. Commun.* **2010,** 46, 8195-8197.

21. Schibel, A. E. P.; An, N.; Jin, Q.; Fleming, A. M.; Burrows, C. J.; White, H. S. Nanopore detection of 8-oxo-7,8-dihydro-2'-deoxyguanosine in immobilized single-stranded DNA via adduct formation to the DNA damage site. *J. Am. Chem. Soc.* **2010,** 132, 17992-17995.

22. Schibel, A. E. P.; Fleming, A. M.; Jin, Q.; An, N.; Liu, J.; Blakemore, C. P.; White, H. S.; Burrows, C. J. Sequence-specific single-molecule analysis of 8-oxo-7,8-dihydroguanine lesions in DNA based on unzipping kinetics of complementary probes in ion channel recordings. *J. Am. Chem. Soc.* **2011,** 133, 14778-14784.

23. Vercoutere, W. A.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M. Rapid discrimination among individual DNA hairpin molecules at

single-nucleotide resolution using an ion channel. *Nat. Biotechnol.* **2001,** 19, 248-252.

24. Olasagasti, F.; Lieberman, K. R.; Benner, S.; Cherf, G. M.; Dahl, J. M.; Deamer, D. W.; Akeson, M. Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat. Nanotechnol.* **2010,** 5, 798-806.

25. Gyarfas, B.; Olasagasti, F.; Benner, S.; Garalde, D.; Lieberman, K. R.; Akeson, M. Mapping the position of DNA polymerase-bound DNA templates in a nanopore at 5 A resolution. *ACS Nano* **2009,** 6, 1457-1466.

26. Lieberman, K. R.; Cherf, G. M.; Doody, M. J.; Olasagasti, F.; Kolodji, Y.; Akeson, M. Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. *J. Am. Chem. Soc.* **2010,** 132, 17961-17972.

27. Lu, S.; Li, W.; Rotem, D.; Mikhailova, E.; Bayley, H. A primary hydrogen-deuterium isotope effect observed at the single-molecule level. *Nat. Chem.* **2010,** 2, 921-928.

28. An, N.; Fleming, A. M.; White, H. S.; Burrows, C. J. Crown ether-electrolyte interactions permit nanopore detection of individual DNA abasic sites in single molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** 109, 11504-11509.

29. Zhang, B.; Galusha, J.; Shiozawa, P. G.; Wang, G.; Bergren, A. J.; Jones, R. M.; White, R. J.; Ervin, E. N.; Cauley, C. C.; White, H. S. Bench-top method for fabricating glass-sealed nanodisk electrodes, glass nanpore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **2007,** 79, 4778-4787.

30. White, R. J.; Ervin, E. N.; Yang, T.; Chen, X.; Daniel, S.; Cremer, P. S.; White, H. S. Single ion-channel recordings using glass nanopore membranes. *J. Am. Chem. Soc.* **2007,** 129, 11766-11775.

31. Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** 106, 7702-7707.

32. Frelon, S.; Douki, T.; Ravanat, J.; Pouget, J.; Tornabene, C.; Cadet, J. High-performance liquid chromatography-tandem mass spectrometry measurement of radiation-induced base damage to isolated and cellular DNA. *Chem. Res. Toxicol.* **2000,** 13, 1002-1010.

33. Purnell, R. F.; Mehta, K. K.; Schmidt, J. J. Nucleotide identification and orientation discrimination of DNA homopolymers immobilized in a protein nanopore. *Nano Lett.* **2008,** 8, 3029-3034.

34. Henrickson, S. E.; Misakian, M.; Robertson, B.; Kasianowicz, J. J. Driven DNA transport into an asymmetric nanometer-scale pore. *Phys. Rev. Lett.* **2000,** 85, 3057-3060.

35. Purnell, R.; Mehta, K.; Schmidt, J. Nucleotide identification and orientation discrimination of DNA homopolymers immobilized in a protein nanopore. . *Nano Lett* **2008,** 8, 3029-3034.

36. Muzard, J.; Martinho, M.; Mathe, J.; Bockelmann, U.; Viasnoff, V. DNA translocation and unzipping through a nanopore: some geometrical effects. *Biophys. J.* **2010,** 98, 2170-2178.

37. Martinez-Haya, B.; Hurtado, P.; Hortal, A. R.; Hamad, S.; Steill, J. D.; Oomens, J. Emergence of symmetry and chirality in crown ether complexes with alkali metal cations. *J. Phys. Chem. A* **2010,** 114, 7048-7054.

38. Smetana, A. J.; Popov, A. I. Lithium-7 nuclear magnetic resonance and calorimetric study of lithium crown complexes in various solvents. *J. Solution Chem.* **1980,** 9, 183-196.

39. Izatt, R. M.; Bradshaw, J. S.; Nielsen, S. A.; Lamb, J. D.; Christensen, J. J. Thermodynamic and kinetic data for cation-macrocycle interaction. *Chem. Rev.* **1985,** 85, 271-339.

40. Broido, M. S.; Kearns, D. R. ^{1}H NMR evidence for a left-handed helical structure of poly(ribocytidylic acid) in neutral solution. *J. Am. Chem. Soc.* **1982,** 104, 5207-5216.

41. Japrung, D.; Henricus, M.; Li, Q.; Maglia, G.; Bayley, H. Urea facilitates the translocation of single-stranded DNA and RNA through the α-hemolysin nanopore. *Biophys. J.* **2010,** 98, 1856-1863.

CHAPTER 3

MODULATION OF THE CURRENT SIGNATURE OF DNA ADDUCTS IN THE ALPHA-HEMOLYSIN ION CHANNEL

Introduction

The α-hemolysin (α-HL) protein ion channel has been under intensive investigation as a DNA analysis platform.[1, 2] Most of the studies are focused on rapid, low-cost sequencing technique development via monitoring the different current blockage levels and the durations of normal DNA bases as an electric field drives single-stranded DNA (ssDNA) through the channel.[3-8] One of the biggest challenges of this technology arises from the low current amplitude resolution between the nucleobases while ssDNA translocates through the protein channel. Along with the high speed of the ssDNA electrophoretic movement, the available electronics cannot provide a high enough signal-to-noise ratio to achieve single-nucleotide identification.

In order to overcome these technical hurdles, several approaches have been developed. First, strategies were used to immobilize ssDNA inside of the α-HL ion channel, which allows close inspection of the various current levels produced by different bases. One of these approaches is to attach a 3'-terminal biotinylated DNA to streptavidin, forming a complex so large that it is not able to enter into the protein

vestibule. As a result, the ssDNA is kept inside of the ion channel, providing enough time for the electronics to measure its conductance. [9-12] By changing the positions of the substituted nucleotide, three recognition zones along the protein β-barrel were discovered, and this means that at least 10-15 nucleotides are contributing to the current blockage signals with position 14, relative to the 3' biotin, possessing the best discrimination.[10] Further, DNA polymerases have also been also utilized to ratchet the DNA molecule one nucleotide at a time to control the translocation speed, consequently increasing the accuracy.[7, 8] Simultaneously, site-directed mutagenesis was applied to engineer the protein ion channel, producing narrower recognitions and higher capture rates.[9, 13] Experimental conditions were explored to assist the unraveling of DNA secondary structures, and to slow down the translocation speed.[14-16] Additionally, small molecules have been synthesized to bind the α-HL ion channel as adapters for the ssDNA translocation.[17, 18]

Another approach to increase the current-amplitude resolution is through chemical modification of DNA. As discussed in Chapter 2, the DNA abasic (AP) site 18-crown-6 adduct had exhibited unique current signatures that are rather different form the native-base current level registry.[19] This provided the detection of individual AP sites along the same strand at a single-molecule level. There are also other adducts that produced current signatures different from those of the normal bases in the immobilization experiments; unfortunately, they were unable to generate distinct signatures during the free translocation through the wild-type α-HL. However, by studying these adducts, we can gain more insights into the correlation between the size, shape, rigidity and hydrophobicity of the molecules attached to DNA and the electrical

signals so yielded. Additionally, these observations can be of great potential to produce distinguishable signatures for the mutant α-HL ion channels.

Experimental Section

Materials

KCl, EDTA, Tris-HCl, taurine (Trn), Arg-His (RH), D-(+)-glucosamine (GlcN), Gly-Pro-Arg-Pro amide (GPRP), streptomycin (STM), $NaBH_3CN$, wild-type α-HL, phospholipid 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC), 2-aminomethyl-15-crown-5 (15c5), 2-aminomethyl-18-crown-6 (18c6), streptavidin and uracil-DNA glycosylase (UDG) were purchased from commercial suppliers and used without further purification.

Preparation and characterization of DNA AP adducts

The oligodeoxynucleotides (ODN) were synthesized from commercially available phosphoramidites (Glen Research, Sterling, VA) by the DNA-Peptide Core Facility at the University of Utah. After synthesis, each ODN was cleaved from the synthetic column and deprotected according to the manufacturer's protocols, followed by purification using a semipreparative ion-exchange HPLC column with a linear gradient of 25% to 100% B over 35 min while monitoring absorbance at 260 nm (A = 20 mM Tris, 1 M NaCl pH 7 in 10% CH_3CN/90% ddH_2O, B = 10% CH_3CN/90% ddH_2O, flow rate = 3 mL/min). Uridine-containing oligomers (10 μM, 1 nmole) and 1 unit UDG were thermally equilibrated in UDG buffer (pH 8.0) at 37 ^{o}C for 30 min, followed by dialysis against ddH_2O for 12 h. The resulting AP-containing ODNs (10 μM, 1 nmole) were dried down

and resuspended in MOPS buffer (pH 6.5), followed by the addition of the appropriate amine (20 mM, 2 μmoles) and $NaBH_3CN$ (100 mM, 10 μmoles); then the reactions were kept at 37 °C for 24 h. Unreacted AP-containing ODNs were cleaved with 0.1 M NaOH. After dialysis against ddH_2O for 12 h, all products were purified by analytical ion-exchange HPLC running a linear gradient of 25% to 100% B over 35 min while monitoring absorbance at 260 nm (A = 20 mM Tris, 1 M NaCl pH 7 in 10% CH_3CN/90% ddH_2O, B = 10% CH_3CN/90% ddH_2O, flow rate = 1 mL/min) Analysis of the crude-reaction products indicated yields of approximately 85-90%. The identities of the 3'-biotinylated ODNs were determined by negative ion electron spray (ESI^-) mass spectrometry on a Micromass Quattro II mass spectrometer equipped with Zspray API source in the mass spectrometry laboratory at the Department of Chemistry, University of Utah (Table 3.1).

Ion channel recordings

Ultrapure water (>18 MΩ•cm) was prepared by a Barnstead E-pure water purifier and used to make buffered electrolyte solution (1.0 M KCl, 1.0 mM EDTA, 25 mM Tris, pH=7.9) that was used for the single ion channel current recordings. The electrolyte was filtered with a sterile 0.22 mm Millipore vacuum filter before the measurement. The protein α-HL was diluted to a 0.5 mg/mL solution in ultra pure water and the lipid DPhPC was dissolved in decane to a concentration of 10 mg/mL, both of which were stored in a -20 °C freezer. The glass nanopore membrane (GNM) (radius = 600 nm) was fabricated as previously established,[20] and was silanized in 2% (v:v) 3-cyanopropyldimethylchlorosilane in CH_3CN for 6 h. Ag/AgCl electrodes were prepared

Table 3.1. Mass spectrometry characterization of the oligonucleotides. The ODNs presented here were biotinylated at the 3' terminus, and position 14, relative to the biotin tag, was substituted with various amines (X) in a poly-dC background.

$C_{39}X_{\omega 14}$ (X =)	Calculated Mass	Experimental Mass
C	12075.2	12075.3
U	12076.2	12076.0
Trn	12091.2	12091.9
GlcN	12145.3	12144.8
RH	12276.4	12276.8
GPRP	12390.4	12390.1
STM	12547.5	12546.9
15c5	12215.4	12215.2
18c6	12259.4	12259.2

by soaking silver wires (diameter = 0.25 mm) in bleach.[21] In the immobilization studies, the 3'-biotinyliated ODNs (160 pmol) were mixed with streptavidin (40 pmol) and kept at 23 °C for 20 min before the measurements, while in the translocation studies, the 87-mer ODNs were used directly after purification and dialysis.

A custom built high-impedance, low noise amplifier and data acquisition system, donated by Electronic Bio Sciences, San Diego, CA, was used for the current-time (*i-t*) recordings. The GNM was rinsed with CH_3CN, ethanol and ultrapure water, and then filled with the electrolyte described above. A pipette holder with a pressure gauge and a 10 mL gas-tight syringe was used to locate the GNM to the DC system. Two Ag/AgCl electrodes were positioned inside and outside of the GNM to apply an electrical potential.

The lipid DPhPC solution (1 μL) was painted on the GNM surface using a plastic pipette tip (flat gel-loading tips, 1-200 μL) to form a suspended bilayer, which was confirmed by the resistance of approximately 100 GΩ, a dramatic decrease from that of an open GNM orifice (10 MΩ). After the addition of α-HL (0.2 μL), a pressure was applied to assist the insertion of the ion channel protein, which had a resistance of around 1 GΩ under these conditions.

In the immobilization studies, Strep-Btn DNA (40 pmol, 200 nM) was added in the cell and more than 200 capture/release events were collected under -120 mV bias with a 10 kHz low pass filter, and 50 kHz data acquisition rate. Then the same amount of Strep-Btn C_{40} was added as an internal standard, and ~200 events were collected for each strand under the same conditions. As for the translocation studies, 87-mer DNA (2 nmol, 10 μM) was added and more than 2000 events were collected under different voltages

(80, 100, 120, 160 mV *trans* vs. *cis*) with a 100 kHz low pass filter, and 500 kHz data acquisition rate.

Immobilization data analysis

The Strep-Btn DNA was retained in the ion channel for 1 s and a histogram of the percentage residual current $\%I/I_o$ was plotted with a bin width of 0.1%, setting Strep-Btn C_{40} to 0%.

Translocation data analysis

As for the poly-dC_{87}, the histogram of events longer than 0.01 ms were fit into a Gaussian model with a peak location t_p. As for the poly-$dC_{43}GPRPdC_{43}$, the events were fit into an exponential decay model. A bin width of 0.01 ms was applied for all the analysis.

Results and Discussion

Formation of DNA AP adducts

Given the fact that at least 10-15 nucleotides are contributing to the current blockage signals,[9, 10] one AP could not efficiently change the signal to be resolved from the normal bases; therefore, we amplified the differences by generating AP adducts. DNA AP tautomerize between a ring-closed and ring-open aldehyde form, whose existence confers to the AP lability toward basic conditions and also offers the possibility to be further functionalized via coupling with amines to form Schiff bases, followed by reduction with $NaBH_3CN$ to produce stable amine adducts (AP-NR) (Figure 3.1).

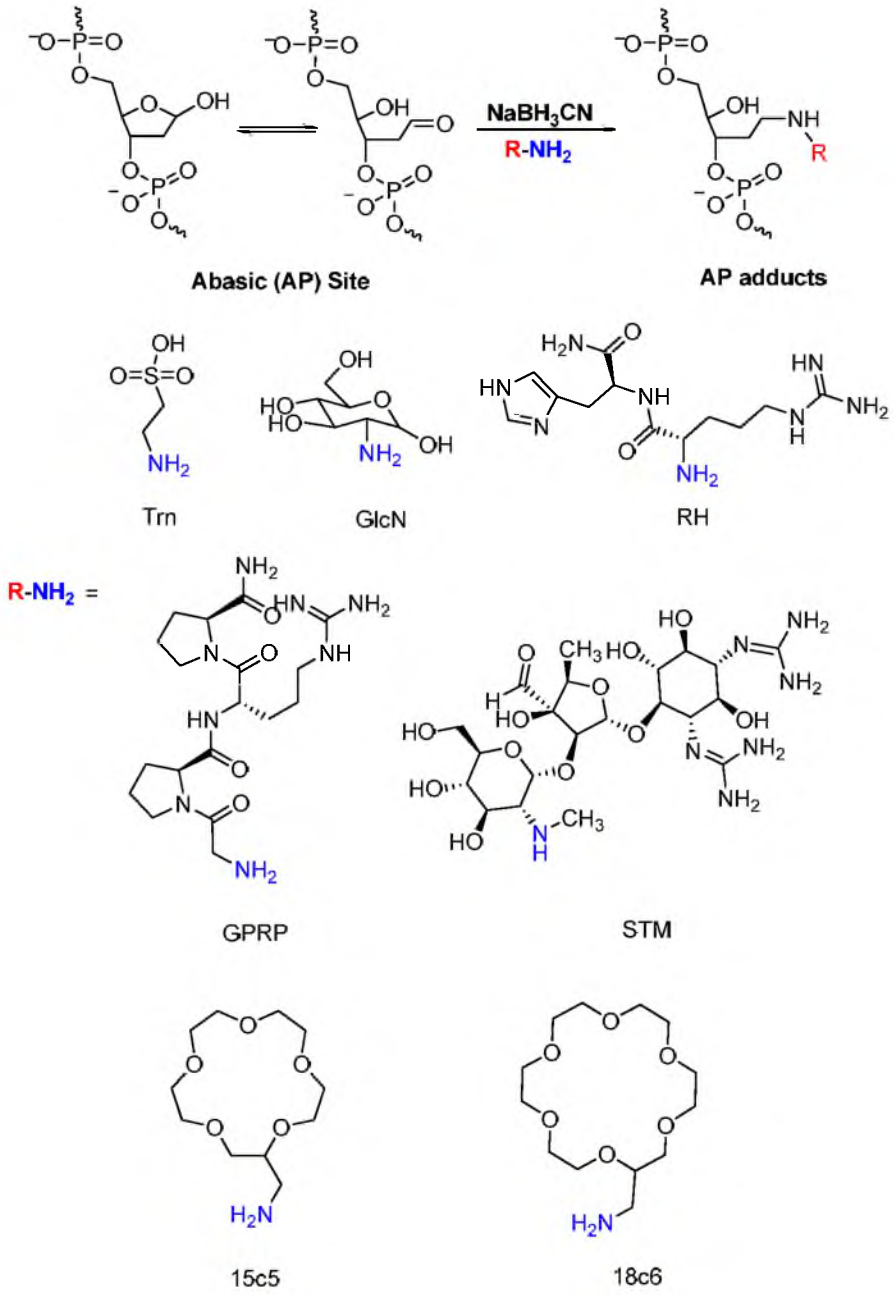


Figure 3.1. DNA AP adducts synthesis scheme. Taurine (Trn), glucosamine (GlcN), Arg-His (RH), Gly-Pro-Arg-Pro amide (GPRP) and streptomycin (STM), 2-aminomethyl-15-crown-5 (15c5) and 2-aminomethyl-18-crown-6 (18c6) were attached to DNA AP to produce adducts (AP-NR) via reductive amination.

Immobilized adducts in a homopolymeric background

Reproducible insertion of a single α-HL ion channel was constituted into a spanning lipid bilayer, supported by a glass nanopore membrane (GNM). The orifice radius of the GNM was fabricated to be as small as 500 ~ 1000 nm to reduce systematic noise and to increase current, temporal resolution;[20] the surface of the glass was modified with 3-cyanopropyldimethylchlorosilane, allowing the formation of a suspended lipid bilayer structure and preventing protein adsorption on glass.[21] We used the streptavidin-biotin (Strep-Btn) complex strategy to immobilize ssDNA in the ion channel to accurately measure its electrical signature. A voltage (-120 mV) applied across the GNM drove the complex to enter the *cis* side of the protein nanopore, reducing the open channel current I_o to a lower level I, both of which were recorded, and $\%I/I_o$ was quoted as the percentage residual current. The Strep-Btn ssDNA complex was retained for 1 s; the polarities of the electrodes were reversed to remove it and then were switched back for the next event (Figure 3.2).

The 5 nm long β barrel of α-HL possesses multiple recognition sites that simultaneously contribute to the residual current, and one of the most sensitive spots resides in position 14 (ω14), relative to the biotin linkage at the 3'-DNA terminus.[10] Therefore, the substitutes were placed at this position in both a homopolymer sequence and heterosequences in the immobilization analysis. More than 200 capture/release events were collected for every DNA strand in each sample and the 3'-biotinylated poly-dC_{40} (C_{40}) was used as an internal standard.

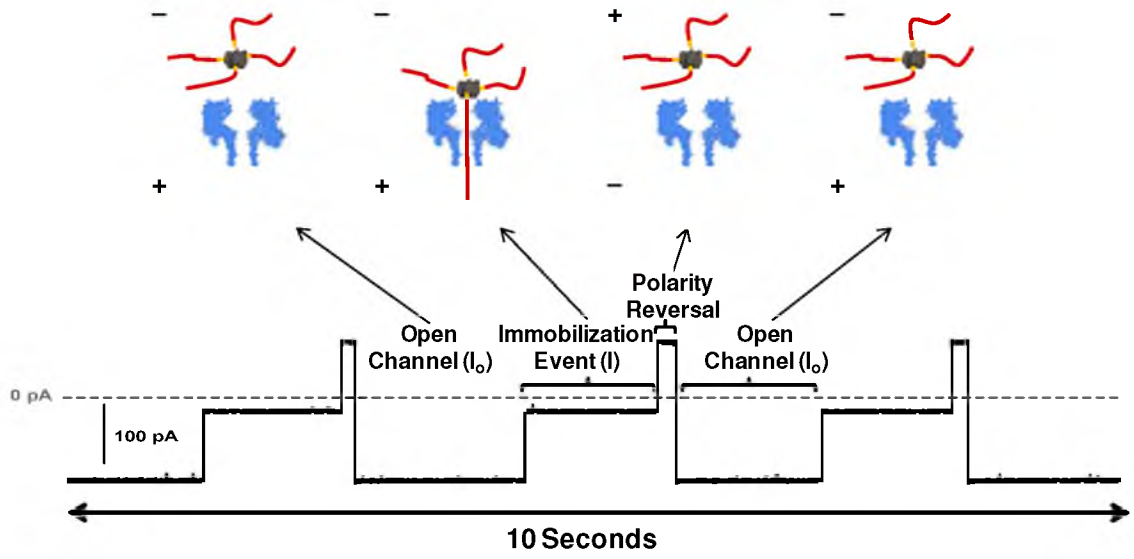


Figure 3.2. Illustration of the immobilization events. Four ssDNA (red) with the biotin linkage (orange) were bound to the streptavidin (black) to form a complex. When driven by the electric force, the complex entered the α-HL ion channel (blue) and remained captured, causing a reduction of the open channel current I_o to the blockage level I. The polarities of electrodes were reversed to control the motion of the complex.

As discussed in Chapter 2, natural DNA AP sites generated a current level that fell into the range of the native bases, and it would therefore produce a false readout in sequencing efforts.[19] In order to shift the $\%I/I_o$ of AP sites to be distinguishable from the native base range, a series of AP adducts were synthesized at the ω14 position with different sizes and shapes via reductive amination, and their electrical signatures within α-HL ion channel were modulated and recorded.

The amines that were attached to AP included taurine (Trn), Arg-His (RH), glucosamine (GlcN), Gly-Pro-Arg-Pro amide (GPRP) and streptomycin (STM). Figure 3.3 shows the $\%I/I_o$ histograms of the AP adducts relative to C_{40}, and except for STM, all adducts exhibited a single distributions of residual current levels. $C_{39}Trn_{\omega 14}$ had a broader but similar $\%I/I_o$ compared to C_{40}, not surprisingly as the molecular weight of this adduct was fairly close to C_{40}, and the disturbance of poly-dC homopolymer secondary structure caused different interactions of the ssDNA with the interior of the protein channel, resulting in a broader $\%I/I_o$ range, a phenomenon that is observed for all the adducts. Additionally, the more bulky adducts $C_{39}RH_{\omega 14}$ and $C_{39}GlcN_{\omega 14}$ blocked the open channel current about 2% and 3% more than an AP site, respectively. Even though RH is heavier than GlcN, $C_{39}GlcN_{\omega 14}$ had a bigger influence on the current levels than $C_{39}RH_{\omega 14}$ did, and this could be explained by the different geometries these two amines displayed. RH was more linear than GlcN, which suggested that there were more factors than merely size affecting the electrical signatures of the DNA AP adducts. The tetrapeptide GPRP produced even deeper blockage, and the $\Delta\%I/I_o$ = 8% compared to AP presented an absolute current difference of ΔI ~ 10 pA in 1 M KCl under -120 mV. The only AP adduct that had more than one $\%I/I_o$ distribution was $C_{39}STM_{\omega 14}$; besides a 9% more

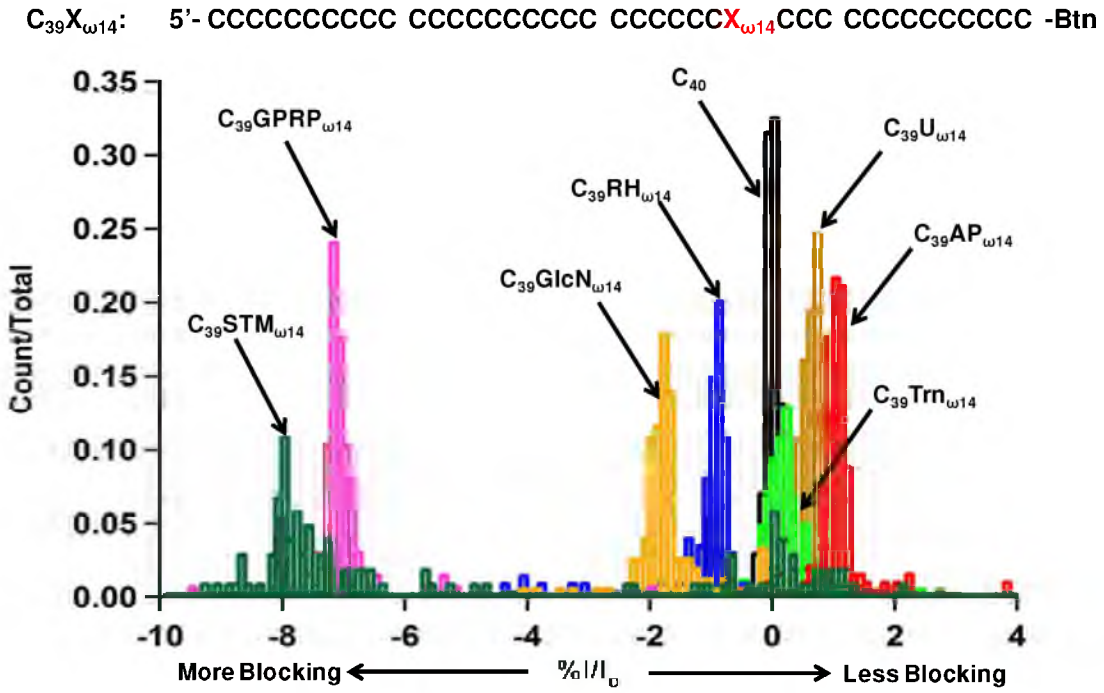


Figure 3.3. *%I/I*$_o$ histograms of U, AP and AP adducts at ω14 position relative to C in a poly-dC$_{40}$ sequence. Every histogram has more than 200 events with a bin increment of 0.1%, and *%I/I*$_o$ for dC$_{40}$ was set to 0.

blocking $\%I/I_o$ relative to AP, there was also another that lined up with C_{40}. Given the enormous size of STM, this is indicating that $C_{39}STM_{\omega 14}$ could not immediately enter the β barrel of α-HL, and thus it hesitated at the constriction, causing the sensing of the poly-dC signal.

After being attached to a 3'-biotinylated poly-dC_{40} at the ω14 position and trapped inside the α-HL ion channel via binding to streptavidin in 1 M KCl under -120 mV, the electrical signatures of crown ethers were also recorded. The $\%I/I_o$ histograms were plotted, setting C_{40} as a reference that indicated different behaviors of the adducts (Figure 3.4). While $C_{39}15c5_{\omega 14}$ produced a single more blocking current level than C_{40} ($\%I/I_o$ = 4.7 %), $C_{39}18c6_{\omega 14}$ showed an extra peak, lining up with C_{40}. This could be explained by the fact that the 18c6-K^+ complex has a single quasiplanar symmetric D_{3d} conformer, resulting from the ideal size matching between the cation and the polyether cavity. Additionally, this conformer provides a planar complex that is proper for hydration, and consequently, stabilizes the system and maximizes its size (~ 3.8 Å in radius).[22] When AP-$18c6_{\omega 14}$ was trying to pass the constriction in order to enter the β barrel, the complex exceeded the dimension of the protein, causing the sensing of the poly-dC tail that gave the extra C_{40} peak. Given the low release rate (3.7×10^6 s^{-1}) and the high concentration of K^+,[23] it is reasonable to assume that its dissociation from 18c6 was unfavorable. In the case of 15c5-K^+, the ion is too big to fit into the cavity and the crown ether was forced to distort its backbone to achieve the greatest interaction with the cation lying above it. This provides the ability of the 15c5-K^+ system to have smaller and more flexible conformation, and high concentration of the K^+ ions could, to some extent, screen the complexation. Thus, 15c5-K^+ complex could change its conformation to pass

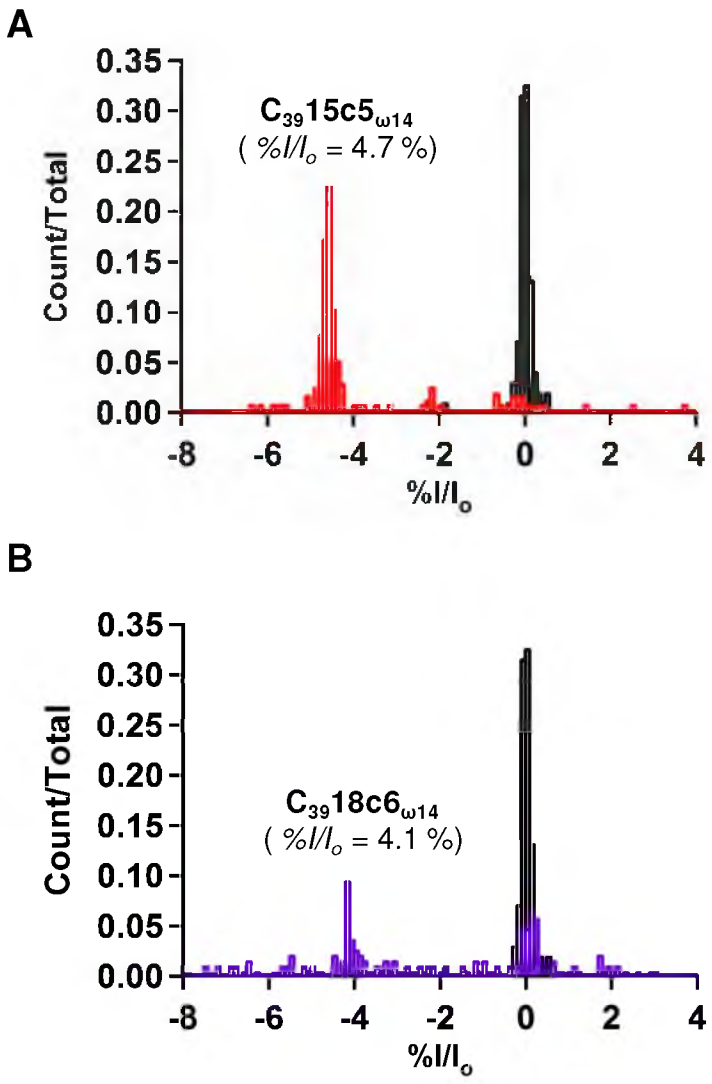


Figure 3.4. *%I/I$_o$* histograms of the crown ether adducts at ω14 position in a poly-dC$_{40}$ sequence relative to C$_{40}$ in 1 M KCl (A: C$_{39}$15c5$_{ω14}$; B: C$_{39}$18c6$_{ω14}$). Every histogram had more than 200 immobilization events with 0.1% bin increment. The *%I/I$_o$* of the reference C$_{40}$ was set to 0.

the constriction and fully dip into the sensing zone, giving only one blocking peak. The observation that the other peak of C_{39}18c6-K^+ blocked less than C_{39}15c5-K^+ was another indicator that 18c6-K^+ was not at the same sensitive site as 15c5-K^+. The 18c6 data presented here were taken from Chapter 2 and the published article for comparison purposes.[19]

Immobilized adducts in a heteropolymeric background

In order to demonstrate the sequence context effect on the $\%I/I_o$, U, AP and the GPRP adduct were incorporated at the ω14 position in a 40-mer oligonucleotide with the K-*ras* gene sequence embedded in the poly-dC background. This particular site that U, AP and GPRP adduct were positioned at is highly prone to G→A transition mutation, predominantly in colon, breast, lung and brain cancers.[24] As one major source of the G→A transition mutation, deamination of C to U could be specifically recognized and removed by UDG, generating AP that was further functionalized to form a GPRP adduct. Figure 3.5 presents the $\%I/I_o$ histograms for K-*ras*$C_{\omega 14}$, K-*ras*$U_{\omega 14}$, K-*ras*$AP_{\omega 14}$ and K-*ras*$GPRP_{\omega 14}$ compared to C_{40}, and they all showed single residual current distributions. K-*ras*$C_{\omega 14}$ and K-*ras*$U_{\omega 14}$ both had similar but sharper $\%I/I_o$ histograms compared to C_{40}, while K-*ras*$AP_{\omega 14}$ blocked 0.5% less than K-*ras*$C_{\omega 14}$ ($\Delta\%I/I_o$ = 1% for $C_{39}AP_{\omega 14}$ and C_{40}), indicating that sequence context has a slight effect on the electrical signatures of AP and native bases. Recently, it has been reported that the sequence context could change the blocking orders of DNA damaged bases and using an engineered α-HL nanopore could also alternate the percentage residual current of native bases,[10, 25] suggesting that it is essential for DNA AP site detection to shift its $\%I/I_o$ beyond the normal range and to

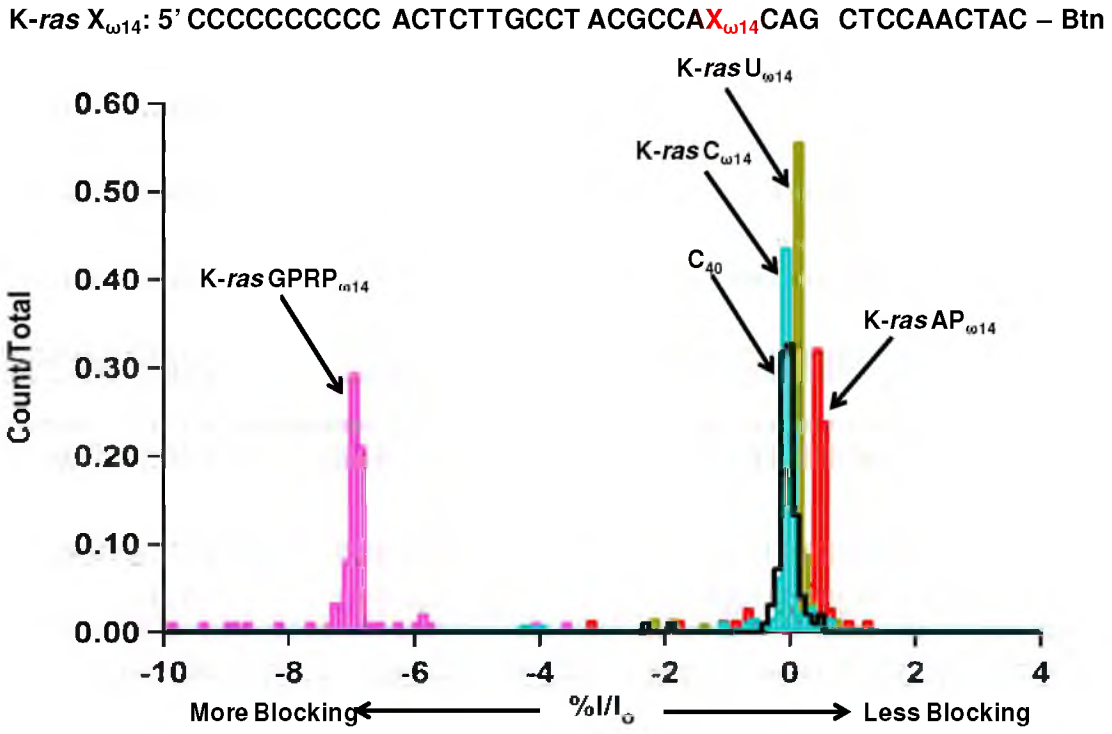


Figure 3.5. *%I/I*$_o$ histograms of U, AP sites and GPRP adduct at ω14 position relative to C in the K-*ras* sequence. Every histogram has more than 200 events with a bin increment of 0.1%, and *%I/I*$_o$ of C_{40} was set to 0.

keep the relative blockage order independent of the sequence background. In the K-*ras* gene sequence, GPRP adduct reduced the open channel current 8% more than K-*ras*$AP_{\omega 14}$ (Figure 3.5), which is significant enough to be distinguished from native bases despite the sequence context and α-HL mutants.

Employing the same strategy combined with the assistance of glycosylases, a variety of DNA base damages and point mutations can be specifically converted to AP and then functionalized to adducts whose electrical signatures were closely investigated above. Besides the uridine/UDG example above, there are 8-oxo-7,8-dihydro-2'-deoxyguanosine (OG):A/MutY,[26] T:G/TDG, 5-methylcytocsine/AID,TDG,[27] 1,N^6-etheno-2'-deoxyadenosine (εdA)/AlkA, and so on.[28]

Translocation of the adducts through the α-HL ion channel

To demonstrate that DNA AP adducts could readily pass through α-HL constriction to the *trans* side, a poly-dC_{87} homopolymer sequence was used to incorporate two of the most blocking adducts: GPRP and STM. The adducts were positioned in the middle of the ssDNA (poly$dC_{43}XdC_{43}$) and polydC_{87} was measured as a standard. Four different voltages (80, 100, 120, 160 mV *trans* vs. *cis*) were used to drive ssDNA through the ion channel and more than 2000 events were collected to plot the duration histograms under each condition. The STM adduct exhibited immobilization-like events, similar to the 18c6 adduct[19] and reversal of the polarities were required to release the ssDNA, suggesting that this adduct could not translocate through the protein nanopore due to its size.This helped confirm our assumption that the prominent $\%I/I_o$ distribution in the same location as C_{40} in the immobilization measurement was

indicating the limitation of the bulky adducts (Figure 3.6). While the translocation of polydC_{87} displayed nearly Gaussian distributions under four voltages, polydC_{43}GPRPdC_{43} was more reliably evaluated by an exponential decay model with the temporal constant τ (Figure 3.7). The analysis model change here was indicative of the extra energy barrier required to drive the bulky adduct through the ion channel. Both polydC_{87} and polydC_{43}GPRPdC_{43} showed a decrease in duration time with increasing voltage, which was consistent with results from other laboratories,[6, 29] and also indicated the complete translocation of the adduct GPRP.

Conclusion

The electrical signature investigation of DNA AP site adducts in both homopolymer sequences and heterosequences demonstrated the power to shift the residual current of AP sites beyond the range of normal bases, an important development for both lesion detection and sequencing efforts. Combined with various glycosylases, the method described here can be used to specifically detect a number of DNA base damages and point mutations. Translocation studies proved the complete migration of the GPRP adduct from *cis* side of the α-HL to the *trans* side, while the STM adduct did not, and this provided, to a certain extent, the size limitation of bulky adducts.

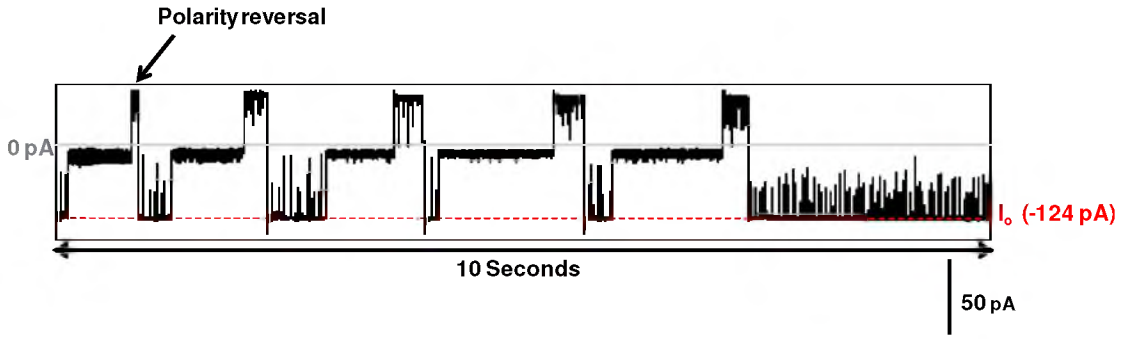


Figure 3.6. Example of *i-t* trace for poly-$dC_{43}STMdC_{43}$ at 120 mV (*trans* vs. *cis*).

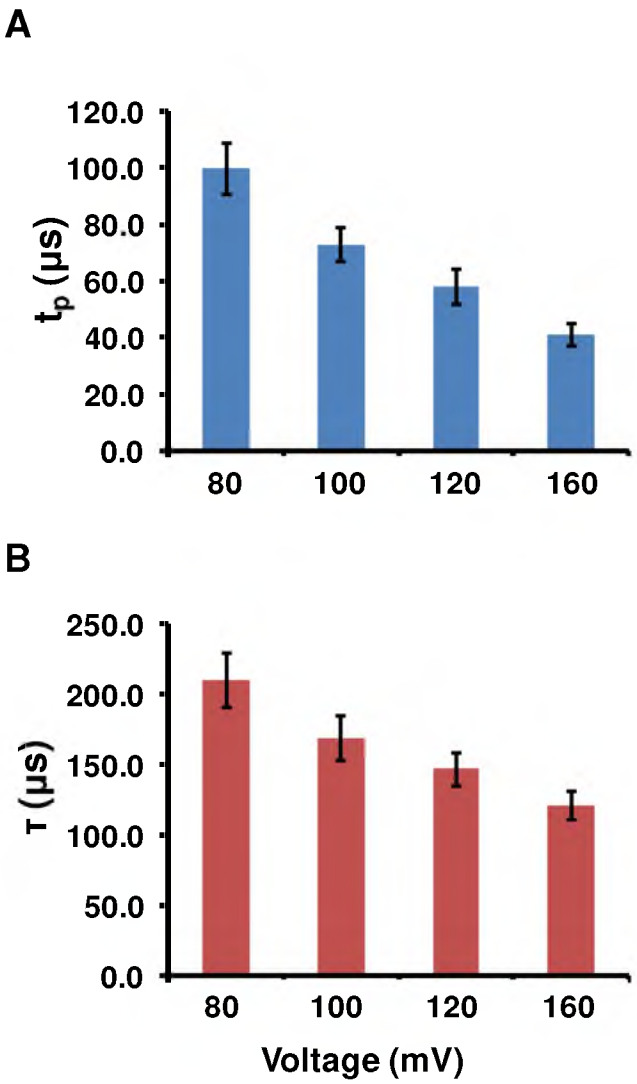


Figure 3.7. Translocation durations (A) Gaussian distribution peak value t_p for poly dC87 and (B) exponential decay constant τ for poly$dC_{43}GPRPdC_{43}$ as a function of applied voltages.

References

1. Branton D; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Ventra, M. D.; Garaj, S.; Hibbs, A.; Huang, X.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **2008,** 26, 1146-1153.

2. Deamer, D. W.; Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* **2002,** 35, 817-825.

3. Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** 93, 13770-13773.

4. Akeson, M.; Branton, D.; Kasianowicz, J. J.; Brandin, E.; Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* **1999,** 77, 3227-3233.

5. Clarke, J.; Wu, H.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **2009,** 4, 265-270.

6. Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** 97, 1079-1084.

7. Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Karplus, K.; Akeson, M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precesion. *Nat. Biotechnol.* **2012,** 30, 343-348.

8. Manrao, E. A.; Derrington, I. M.; Laszlo, A. H.; Langford, K. W.; Hopper, M. K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **2012,** 30, 349-353.

9. Stoddart, D.; Maglia, G.; Mikhailova, E.; Heron, A. J.; Bayley, H. Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angew. Chem. Int. Ed.* **2010,** 49, 556-559.

10. Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** 106, 7702-7707.

11. Purnell, R. F.; Schmidt, J. J. Discrimination of single base substitutions in a DNA strand immobilized in a biological nanopore. *ACS Nano* **2009,** 3, 2533-2538.

12. Purnell, R. F.; Mehta, K. K.; Schmidt, J. J. Nucleotide identification and orientation discrimination of DNA homopolymers immobilized in a protein nanopore. *Nano Lett.* **2008,** 8, 3029-3034.

13. Maglia, G.; Restrepo, M. R.; Mikhailova, E.; Bayley, H. Enhanced translocation of single DNA molecules through α-hemolysin nanopores by manipulation of internal charge. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** 105, 19720-19725.

14. Japrung, D.; Henricus, M.; Li, Q.; Maglia, G.; Bayley, H. Urea facilitates the translocation of single-stranded DNA and RNA through the α-hemolysin nanopore. *Biophys. J.* **2010,** 98, 1856-1863.

15. Franceschini, L.; Mikhailova, E.; Bayley, H.; Maglia, G. Nucleobase recognitino at alkaline pH and apparent pK_a of single DNA bases immobilized within a biological nanopore. *Chem. Commun.* **2012,** 48, 1520-1522.

16. Kawano, R.; Schibel, A. E. P.; Cauley, C.; White, H. S. Controlling the translocation of single-stranded DNA through α-hemolysin ion channels using viscosity. *Langmuir* **2009,** 25, 1233-1237.

17. Banerjee, A.; Mikhailova, E.; Cheley, S.; Gu, L.; Montoya, M.; Nagaoka, Y.; Gouaux, E.; Bayley, H. Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. . *Proc. Natl. Acad. Sci. U.S.A.* **2010,** 107, 8165-8170.

18. Rincon-Restrepo, M.; Mikhailova, E.; Bayley, H.; Maglia, G. Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes. *Nano. Lett.* **2011,** 11, 746-750.

19. An, N.; Fleming, A. M.; White, H. S.; Burrows, C. J. Crown ether-electrolyte interactions permit nanopore detection of individual DNA abasic sites in single molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** 109, 11504-11509.

20. Zhang, B.; Galusha, J.; Shiozawa, P. G.; Wang, G.; Bergren, A. J.; Jones, R. M.; White, R. J.; Ervin, E. N.; Cauley, C. C.; White, H. S. Bench-top method for fabricating glass-sealed nanodisk electrodes, glass nanpore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **2007,** 79, 4778-4787.

21. White, R. J.; Ervin, E. N.; Yang, T.; Chen, X.; Daniel, S.; Cremer, P. S.; White, H. S. Single ion-channel recordings using glass nanopore membranes. *J. Am. Chem. Soc.* **2007,** 129, 11766-11775.

22. Martinez-Haya, B.; Hurtado, P.; Hortal, A. R.; Hamad, S.; Steill, J. D.; Oomens, J. Emergence of symmetry and chirality in crown ether complexes with alkali metal cations. *J. Phys. Chem. A* **2010,** 114, 7048-7054.

23. Gokel, G. W.; Leevy, W. M.; Weber, M. E. Crown ethers: sensors for ions and molecular scaffolds for materials and biological models. *Chem. Rev.* **2004,** 104, 2723-2750.

24. Pfeifer, G. P.; Besaratinia, A. Mutational spectra of human cancer. *Human Genetics* **2009,** 125, 493-506.

25. Schibel, A. E. P.; An, N.; Jin, Q.; Fleming, A. M.; Burrows, C. J.; White, H. S. Nanopore detection of 8-oxo-7,8-dihydro-2'-deoxyguanosine in immobilized single-stranded DNA via adduct formation to the DNA damage site. *J. Am. Chem. Soc.* **2010,** 132, 17992-17995.

26. Kundu, S.; Brinkmeyer, M. K.; Livingston, A. L.; David, S. S. Adenine removal activity and bacterial complementation with the human MutY homologue (MUTYH) and Y165C, G382D, P391L and Q324R variants associated with colorectal cancer. *DNA Repair* **2009,** 8, 1400-1410.

27. Rai, K.; Huggins, I. J.; James, S. R.; Karpf, A. R.; Jones, D. A.; Cairns, B. R. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and Gadd45. *Cell* **2008,** 135, 1201-1212.

28. David, S. S.; Williams, S. D. Chemistry of glycosylases and endonucleases involved in base-excision repair. *Chem. Rev.* **1998,** 98, 1221-1261.

29. Rincon-Restrepo, M.; Mikhailova, E.; Bayley, H.; Maglia, G. Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes. *Nano Lett.* **2011,** 11, 746-750.

CHAPTER 4

STRUCTURE-DEPENDENT INTERACTIONS OF HUMAN TELOMERE SEQUENCE WITH THE ALPHA-HEMOLYSIN NANOPORE

Introduction

Nucleic acids possess the ability to fold to a myriad of secondary structures, which are dependent on the primary sequence as well as the physical conditions in which the structures are prepared and characterized. Duplexes and hairpins are traditional structural motifs, while triplexes, i-motifs, and G-quadruplexes provide additional structures that fold in a highly sequence- and experimentally-depended fashion.[1] As for G-quadruplexes, the folding topology is very sensitive to the cation of the salt chosen to study.[2] The G-quadruplex that folds from the human telomere sequence, which is comprised of the hexanucleotide sequence 5'-TTAGGG-3' repeated thousands of times, provides a fascinating case in which the cation and the physical context play a critical role in defining the overall structural topology.

The following studies highlight the cation and context-dependent conditions in which the human telomere sequence topology is affected. The sequence 5'-AGGG(TTAGGG)$_3$-3' studied by NMR in NaCl solutions was observed to fold to a structure referred to as the basket fold (Figure 4.1).[3] Key features for this structure

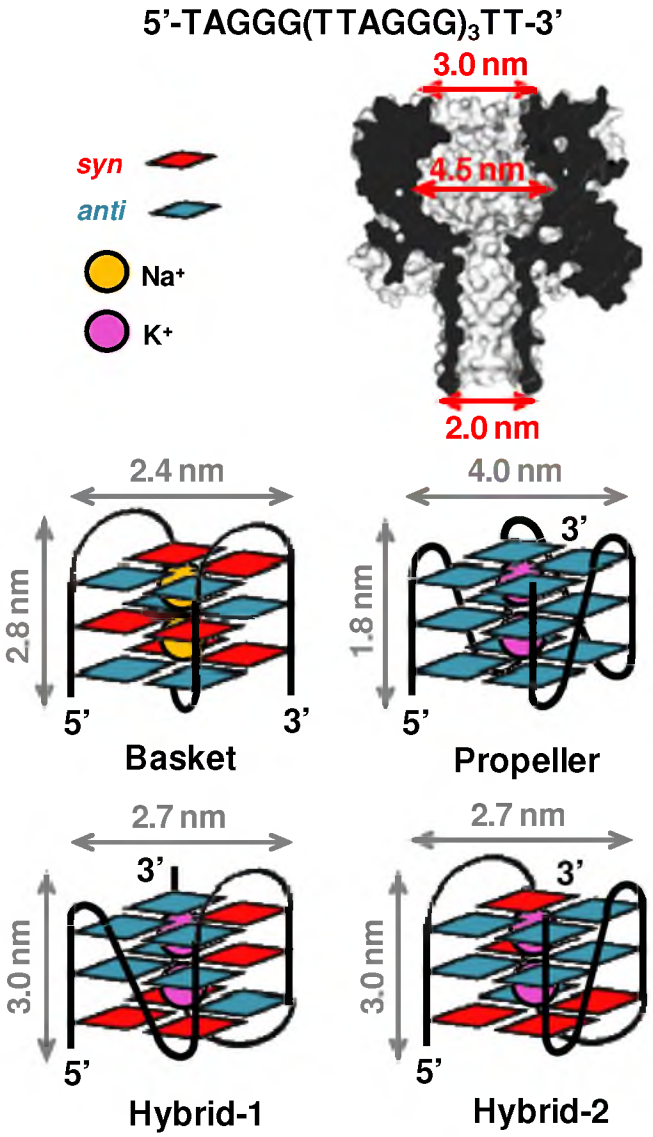


Figure 4.1. Structures and dimensions of the α-HL, basket, propeller, hybrid-1 and hybrid-2 folds.

include an antiparallel strand arrangement with alternating *syn* and *anti* orientation of the G nucleotides and three tetrads, which give two edgewise loops and one diagonal loop. NMR-based studies with the sequence 5'-TAGGG(TTAGGG)$_3$-3' in KCl solution show that it predominantly folds to the hybrid-1 structure (Figure 4.1).[4-6] Characteristics of this structure are an antiparallel strand orientation with a 3+1 core of *syn* and *anti* G nucleotides giving three tetrads. The hybrid-1 fold loop topology consists of a double-chain-reversal loop closest to the 5'-end and two edgewise loops. Counter to the previous observation, the addition of two nucleotides on the 3'-end, which yields the natural human telomere sequence 5'-TAGGG(TTAGGG)$_3$TT-3', was studied by NMR in KCl solution to give predominantly the hybrid-2 structure (Figure 4.1).[7] The key difference between the hybrid folds is with regard to the loop topology in which the hybrid-2 fold has the double-chain-reversal loop closest to the 3'-end, and the other two loops are edgewise. Further studies conducted with the hybrid folds have found them to be highly polymorphic.[5, 8] In contrast to the hybrid folds, the sequence 5'-AGGG(TTAGGG)$_3$-3' studied by X-ray crystallography was observed to give the propeller fold (Figure 4.1), which consists of a parallel strand arrangement with all G nucleotides adopting the *anti* conformation giving three tetrads, and all three loops are double-chain-reversals.[9] These studies highlight the intricacies in which analysis context and cation identity drive the folding topology for the human telomere sequence.

Until recently, the propeller fold was thought to be the solid-state structure of the human telomere sequence, while the hybrid folds were the solution structures in KCl salts. However, the propeller fold has been observed in KCl solutions bearing high concentrations of organic solvents. The leading hypothesis concerning this observation is

that decreased water activity favors the propeller fold because G-quadruplexes are stabilized by dehydration.[10] Therefore, under dehydrating conditions the fold with the largest solvent accessible area will be observed, which is the propeller fold.[11] More recently, the propeller fold has also been observed under conditions of high viscosity.[12] With these observations in hand, we can finely tune the analysis conditions to give the desired structures for study.

The α-hemolysin (α-HL) nanopore has been used to investigate DNA secondary structures for the past decade, including the hairpins,[13-16] duplexes,[17-20] and quadruplexes.[21] By monitoring the duration and current signatures while electrically drawing the DNA molecules through the nanopore ion channel, kinetics of the unzipping processes and DNA-protein interactions can be evaluated at a single-molecule level. Recently, Shim et al. reported studies of the thrombin-binding aptamer (TBA), which is a guanine-rich 15-base single-stranded DNA (ssDNA) folding into a two-tetrad and one-cation quadruplex, by encapsulating it in the α-HL nanocavity.[22] The unraveling of these structures was found to be spontaneous inside of the larger vestibule, which did not present any voltage dependence. Additionally, the cation-selective regulation of folding and unfolding of this aptamer was examined by analyzing signature current blockages.[23] Furthermore, Rotem et al. utilized an α-HL equipped with the TBA to interrogate the interactions between thrombin and its aptamer as a protein detection platform that also provided the thermodynamic and kinetic properties of this complex.[24]

In the present work, we described the structure-dependent interactions of the human telomeric G-quadruplexes with the α-HL nanopore ion channel, mediated by the identity of the electrolyte's cation and increased viscosity. For these studies, we have

chosen the natural human telomere sequence 5'-TAGGG(TTAGGG)$_3$TT-3', which has a two nucleotide overhang on both the 5'- and 3'-ends. Compared to TBA, telomeric G-quadruplexes contain three tetrads, two cations and different loop topologies, dictated by the surrounding metal ions. We have chosen to capitalize on the ability to utilize salt identity to form known structures of the human telomere sequence in which their interactions with α-HL nanopore were studied.

Experimental Section

Materials

KCl, NaCl, LiCl, LiOH, EDTA, Tris, phospholipid 1,2-diphytanoyl-sn-gly-cero-3-phosphocholine (DPhPC), 3-cyanopropyldimethylchlorosilane, wild-type α-HL monomer, ethanol, acetonitrile, and decane were purchased from commercial suppliers and used without further purification.

Oligodeoxynucleotide preparation and characterization

The oligodeoxynucleotides (ODN) were synthesized by the DNA-Peptide Core Facilty at the University of Utah with the commercially available phosphoramidites (Glen Research, Sterling, VA). After being cleaved from the synthetic column and deprotected according to the manufacturer's protocols, ODNs were purified using a semi-preparative ion-exchange HPLC column with a linear gradient of 20% to 100% B over 30 min while monitoring absorbance at 260 nm (A = 10% CH_3CN/90% ddH_20, B = 25 mM LiOAc (pH 8) 1 M LiCl, flow rate = 3 mL/min). Next, the excess salts were removed by dialysis. The ODN concentrations were determined by UV-vis in ddH_2O at 75 ^oC utilizing the

following extinction coefficients: ε (5'-TAGGG(TTAGGG)$_3$TT) = 0.2805 L*μmole^{-1}*cm^{-1}. Annealing occurred by heating the ODN samples at a 10.0 μM concentration in reaction buffer to 90 °C followed by slowly cooling them to room temperature (~3 h).Next, the G-quadruplex samples were placed at 4 °C for 48 h prior to their study. T_m analysis was conducted by monitoring the absorbance at 260 nm for the propeller sample, and 295 nm for the hybrid and basket samples, at a 3.0 μM concentration while heating the samples from 15.0 to 85.0 °C at a ramp rate of 0.3 °C/min, and a measurement was taken every minute.

Ion channel recordings

A custom-built, high-impedance, low-noise amplifier and data acquisition system was used for the ion channel recordings, and it was designed and constructed by Electronic Biosciences (EBS), San Diego, CA. Ultrapure water (> 18 MΩ•cm) was made by a Barnstead E-pure water purifier and used to make buffered electrolyte solutions (1.0 mM EDTA, 25.0 mM Tris, pH = 7.9, and either LiCl, NaCl or KCl, at concentrations listed in the text and figure legends), followed by filtration with a sterile 0.22 mm Millipore vacuum filter before the measurements were made. The final DNA concentrations were 5.0 μM, except for the propeller studies for which it was 20.0 μM.

A glass nanopore membrane (GNM) was fashioned to be ~600 nm in radius, according to the protocol previously published.[25] The surface of GNM was chemically modified with 2% (v/v) 3-cyanopropyldimethylchlorosilane in acetonitrile,[26] and it was rinsed with acetonitrile, ethanol and ultrapure water prior to use and filled with buffered electrolyte. An Ag/AgCl electrode was positioned inside the GNM and attached to a

pressure gauge and a 10 mL gastight syringe. The other Ag/AgCl electrode was submerged in the buffered electrolyte in the experimental cell. After the addition of DPhPC, a resistance increase (~10 MΩ to ~100 GΩ) was observed as the indication of a suspended bilayer formation across the GNM orifice. A pressure was applied to the back of the GNM for the protein channel to reconstruct, and a single insertion of the α-HL decreased the resistance to ~1 GΩ. A 100 kHz low-pass filter was used to collect and analyze the data with 500 kHz sampling rate. The sample *i-t* traces presented in Figure 3-5 were refiltered to 20 kHz for presentation.

Data analysis

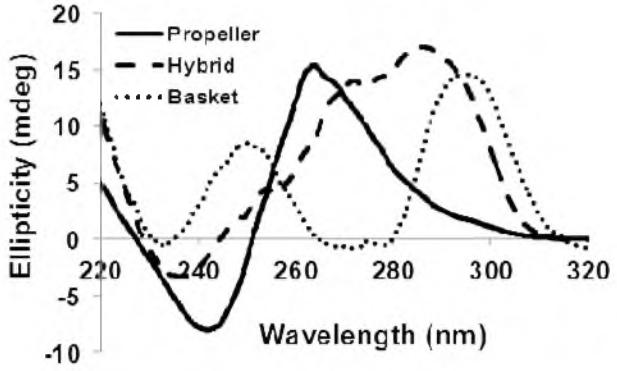
Only events with duration longer than 10 μs (filter bandwidth) were extracted and analyzed using QUB 1.5.0.31 and fitted using Igor Pro 6.1.

Results and Discussion

Characterization of the G-quadruplex structures

To determine if the high-ionic strength of the electrolytes utilized in these studies influenced the structures, CD spectroscopy and thermal melting measurements were conducted. When the sequence, 5'-TAGGG(TTAGGG)$_3$TT-3', was studied in 1 M NaCl solutions, CD characteristics (λ_{max} = 296 and 250 nm, λ_{min} = 270 nm) were observed that are consistent with literature values for the basket fold (Figure 4.2*A*).[27] In a solution bearing 50 mM KCl and 1 M LiCl, CD characteristics (λ_{max} = 286 nm with a shoulder at 270 nm, and λ_{min} = 236 nm) were observed that are consistent with literature reports for the hybrid folds (Figure 4.2*A*).[27] Strikingly, when the LiCl concentration was increased

A



B

[NaCl] (mM)	[KCl] (mM)	[LiCl] (mM)	T_m (°C)
1000	0	0	72.3 ± 0.5
0	50	950	60.2 ± 0.2
0	1	999	40.3 ± 0.5
0	20	5000	60.9 ± 0.8

Figure 4.2. CD spectra and thermal melting values for the G-quadruplexes analyzed. CD spectra (*A*) for the propeller (solid, 20 mM KCl and 5 M LiCl), hybrid (dashed, 50 mM KCl and 1 M LiCl), and basket folds (dotted, 1 M NaCl). All measurements were conducted at 10 μM DNA with the specified salt conditions at 20 °C. Thermal melting data (*B*) for each G-quadruplex analyzed in which the measurements were recorded at 3 μM DNA with the specified salt conditions.

to 5 M with 20 mM KCl, the CD spectra dramatically changed (λ_{max} = 264 nm, and λ_{min} = 242 nm), which is consistent with those reported for the propeller fold (Figure 4.2*A*).[11, 27] We conclude that the 5 M LiCl solution shifts the folding topology toward the propeller fold, and away from the hybrid fold, due to the decreased water activity of the electrolyte solution (5 M LiCl, a_w ~ 0.75),[28] as previously described by Miller et al. with solutions containing 40% CH_3CN (a_w ~ 0.81).[10, 28] The thermal melting values (T_m) for each G-quadruplex studied are reported in Figure 4.2*B*.

Interaction between hybrid folds and the α-HL

Being electrophoretically driven to the proximity of the *cis* side of the α-HL ion channel, the hybrid folds were able to enter into the large vestibule, producing long current blockages (Figure 4.3). There were mainly two types of current-time (*i-t*) traces observed, which correlated to the different interactions between the DNA molecule and the protein. Type 1 events featured alternations between intermediate current levels (I_M) and a deep blockage current level (I), which constituted ~71% of all the recorded events. Interestingly, I_M presented two distinct amplitude populations with percentage residual currents $\%I_M/I_o$ (where I_o represents the open channel current) of ~45% and ~36%, respectively, and a frequency ratio of I_{M1a}: I_{M1b} = 7:18 at 120 mV (*trans* vs. *cis*) (Figure 4.4*A*). Since the G-quadruplex studies in this work have two nucleotides overhanging at both termini, which are protruding from the top and bottom G-tetrads in the hybrid folds (Figure 4.1), the first possibility is that the two I_M populations correspond to either the interactions of 3' and 5' termini with the protein constriction. The second explanation is the coexistence of two hybrid folds, 1 and 2, that provide different I_M values.. In order to

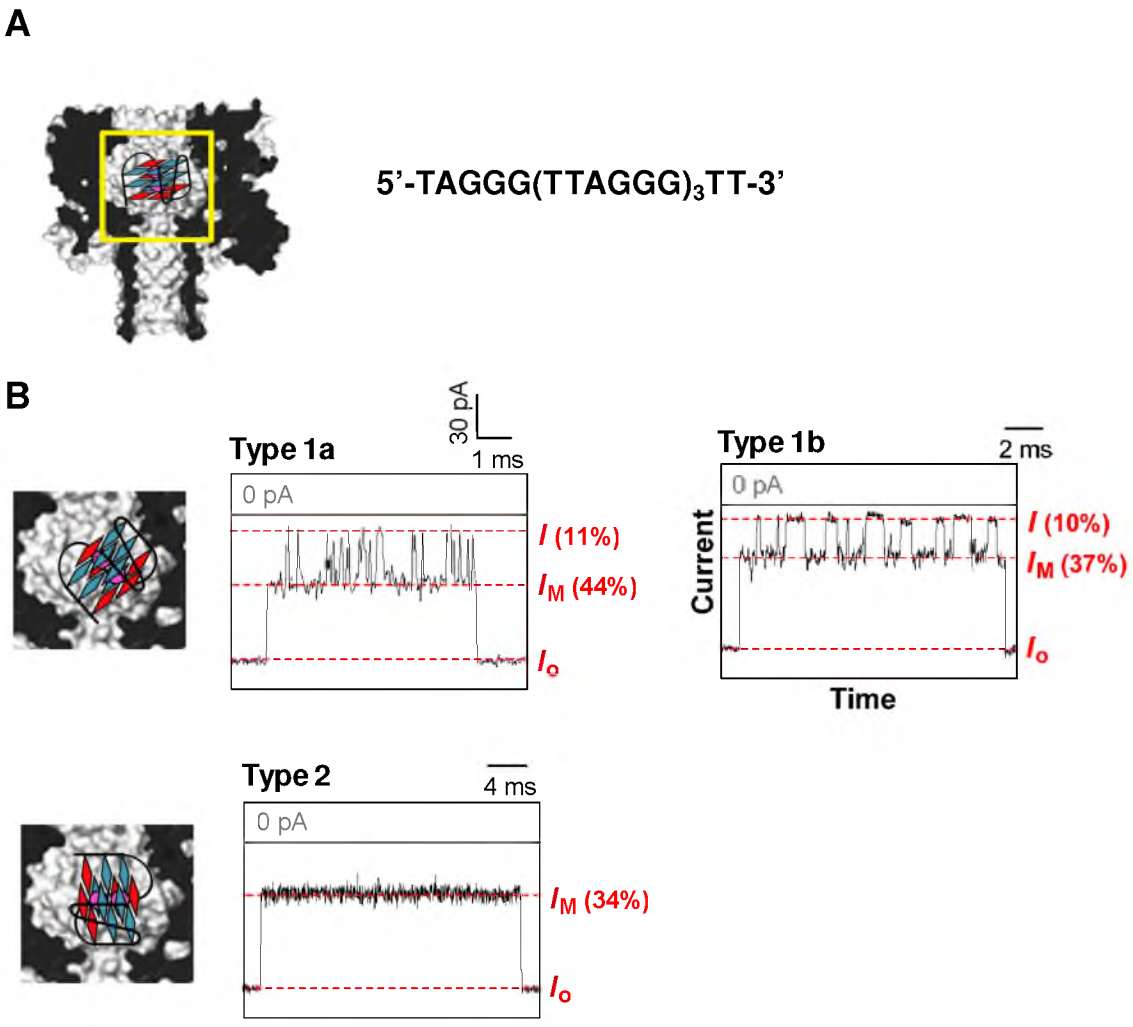


Figure 4.3. Interactions between the hybrid folds and the protein nanopore. (*A*) Illustration of the hybrid folds of the human telomere sequence within the vestibule of α-HL. (*B*) Examples of two main types of current-time traces observed. Buffer condition: 50.0 mM KCl, 950 mM LiCl, 1.0 mM EDTA, 25.0 mM Tris, pH=7.9.

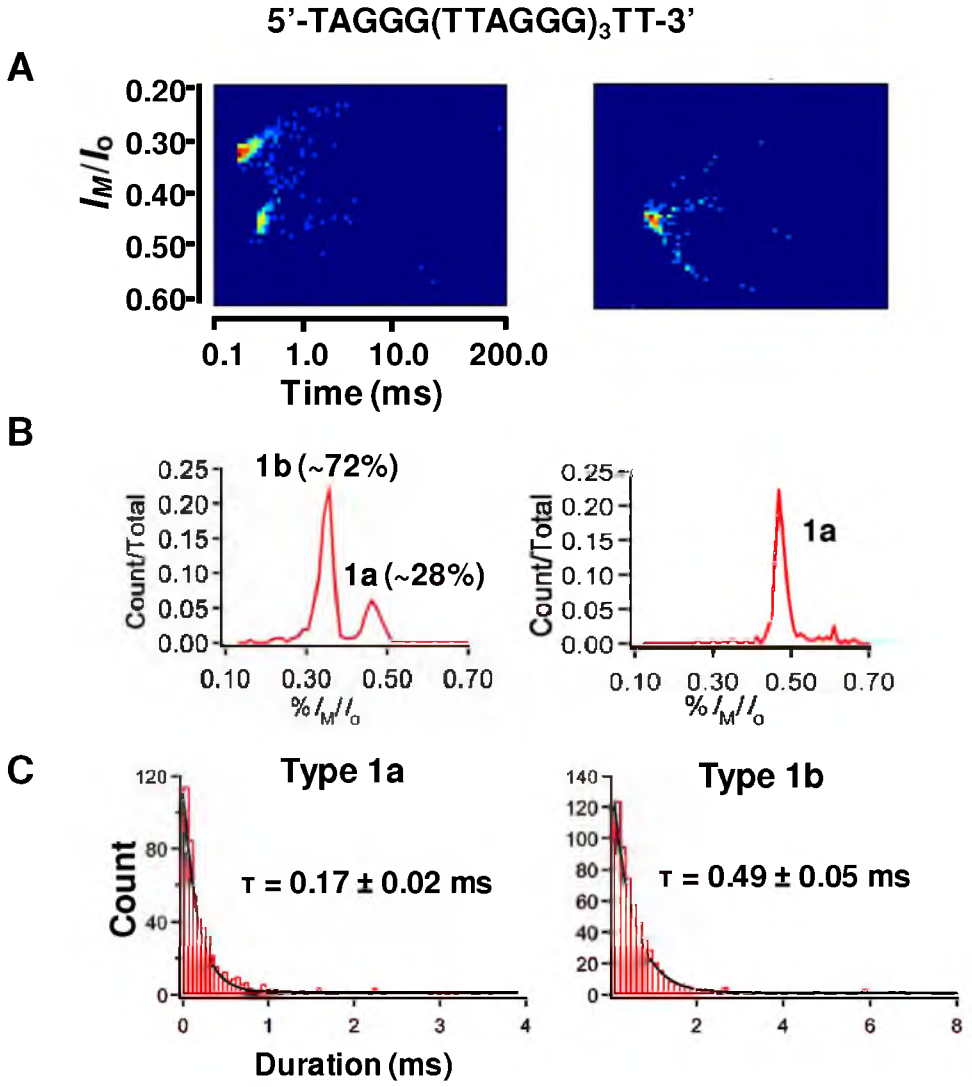
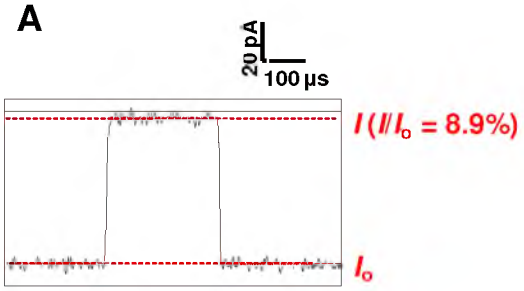


Figure 4.4. Comparison of type 1 events between natural human telomere sequence and the modified sequence (5'-TAGGGTTAGXGTTAXGGTTAGGGTT-3', where X=8-BrG). (*A*) Density plots of the I_M current levels for the natural sequence (left) and the 8-BrG-modified sequence (right). (*B*) Percentage residual current histograms for the I_M current levels for the natural sequence (left) and the 8-BrG-modified sequence (right). (*C*) Duration histograms for the deep blockage current levels *I* of type 1a (left) and 1b (right). Buffer condition: 50.0 mM KCl, 950 mM LiCl, 1.0 mM EDTA, 25.0 mM Tris, pH=7.9.

better understand these behaviors, two 8-bromoguanine bases (8-BrG) were incorporated into the positions 10 and 15, forcing the hybrid-2 fold.[5] As a result, the more blocking I_{M1b} population disappeared, leading to the conclusion that these two different current patterns are correlated to the two different conformations that the human telomere sequence adopts in the K^+ solution, and the assignments of current type 1b to the hybrid-1 fold and type 1a to the hybrid-2 fold (Figure 4.4*A* and *B*). Individual *i-t* traces did not alternate between I_{M1a} and I_{M1b} within one event, indicating the G-quadruplex cannot change its conformation for the time being encapsulated inside of the protein vestibule. Interestingly, the duration distributions of the deep blockage levels *I* were also different between these two sub-types with I_{1b} presenting a larger decay constant τ than I_{1a} in a single-exponential decay model (Figure 4.4*C*). Being able to lock the G-quadruplex conformation inside of the nanocavity while providing resolvable current amplitudes, the α-HL ion channel appears to be a valuable tool to study these structures at a single-molecule level. Not surprisingly, loop side entering was much less frequent (~8%) and exhibited a one-state current blockage ($\%I_M/I_o$ = ~34%), which was also observed with a 'dumbbell' hairpin present inside of the protein channel (Figure 4.3*B*).[29] Unfortunately, there was one event type that did not present voltage dependence, and we could not assign it to a specific interaction (Figure 4.5). This minor population type may best be explained by the fact that hybrid folded G-quadruplexes are notoriously polymorphic, and other uncharacterized structures likely exist in solution.[30]

The addition of the A_{25} tail facilitated the threading of the DNA into the β-barrel, however, it did not seem to assist the unraveling of the secondary structure (Figure 4.6). The resulting deep-current blockages lasted for >20.0 s until reversal of the polarity drove

A



B

Voltage(mV) *trans* vs. *cis*	Percentage event frequency	Decay constant τ (ms)
100	20 %	0.26±0.04
120	21 %	0.20±0.02
160	18 %	0.25±0.03

Figure 4.5. Unknown event type of the human telomere sequence hybrid folds. (A) Sample *i-t* trace. (B) Analysis of its percentage event frequencies and durations (decay constant τ) at 100, 120 and 160 mV (*trans* vs. *cis*). Buffer condition: 50.0 mM KCl, 490 mM LiCl, 1.0 mM EDTA, 25.0 mM Tris, pH=7.9.

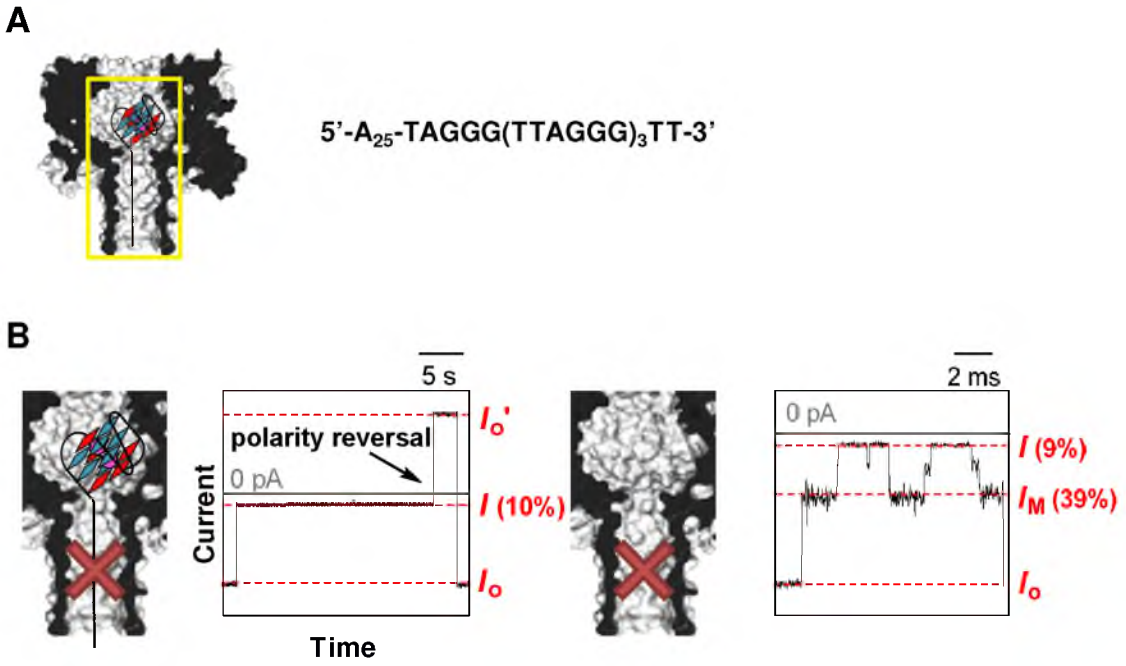


Figure 4.6. Interactions between the hybrid folds with an A_{25} tail and the protein nanopore. (*A*) Illustration of the hybrid folds and the α-HL and the human telomere sequence with an A_{25} tail. (*B*) Examples of two main types of current-time traces observed. Buffer conditions: 50.0 mM KCl, 950 mM LiCl, 1.0 mM EDTA, 25.0 mM Tris, pH=7.9.

the DNA back to the *cis* side of the α-HL, indicating that the hybrid fold cannot unravel under the electrical force, thus the protein was sensing the homopolymeric tail of the strand (Figure 4.6*B*). Increasing the driving force (up to 200 mV), while the DNA was trapped inside of the protein channel, was also attempted to unfold and translocate the G-quadruplex, yet this high force was not successful (Figure 4.7). The long deep blockages still remained until the voltage was decreased to ~75 mV, which was below the threshold for electrophoretically holding the DNA molecule, and release of the DNA was observed. Another method to approach this challenge was to lower the K^+ concentration to 1.0 mM (in 1 M LiCl), under these conditions half-maximal binding to K^+ should occur.[31] In this study, ~43% of the events had the same signatures as the hybrid folds described above, while the rest presented the same translocation *i-t* trace and had a similar duration time that was obtained for the control experiment in 1 M LiCl (Figure 4.8 and 4.9). This observation highlights the fact that once K^+ is bound by the G-quadruplex sequence, it cannot translocate through α-HL.

Interaction between the basket fold and the α-HL

Similarly to other folds, the sodium-bound basket fold could easily pass the *cis* opening of α-HL, entering the large vestibule and yielded two major event types as previously observed with the hybrid fold (Figure 4.10). Compared to the hybrid folds, the basket fold has both termini protruding from the same G-tetrad face; as a result, when the G-quadruplex enters into the ion channel with the terminal overhang facing the protein constriction, a shoulder-spike ($I_M \rightarrow I$) event signature was produced (type 1). Prior to the

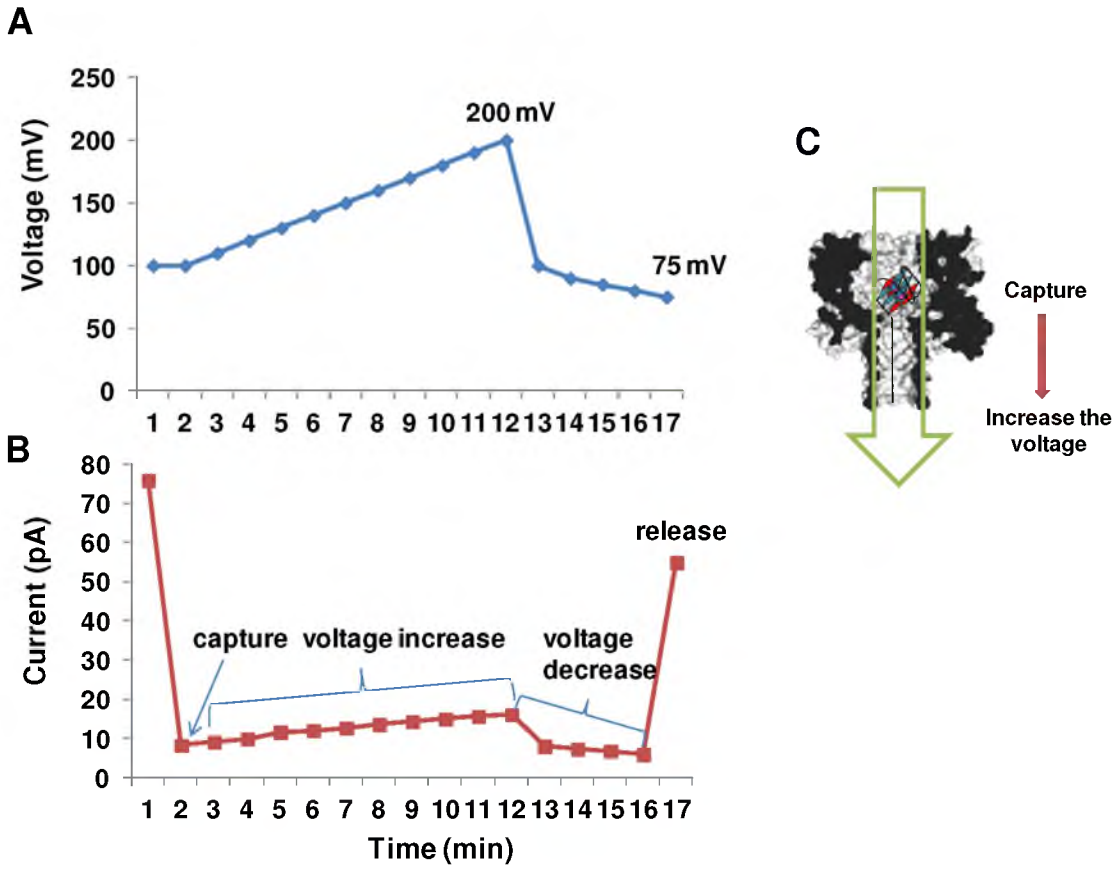
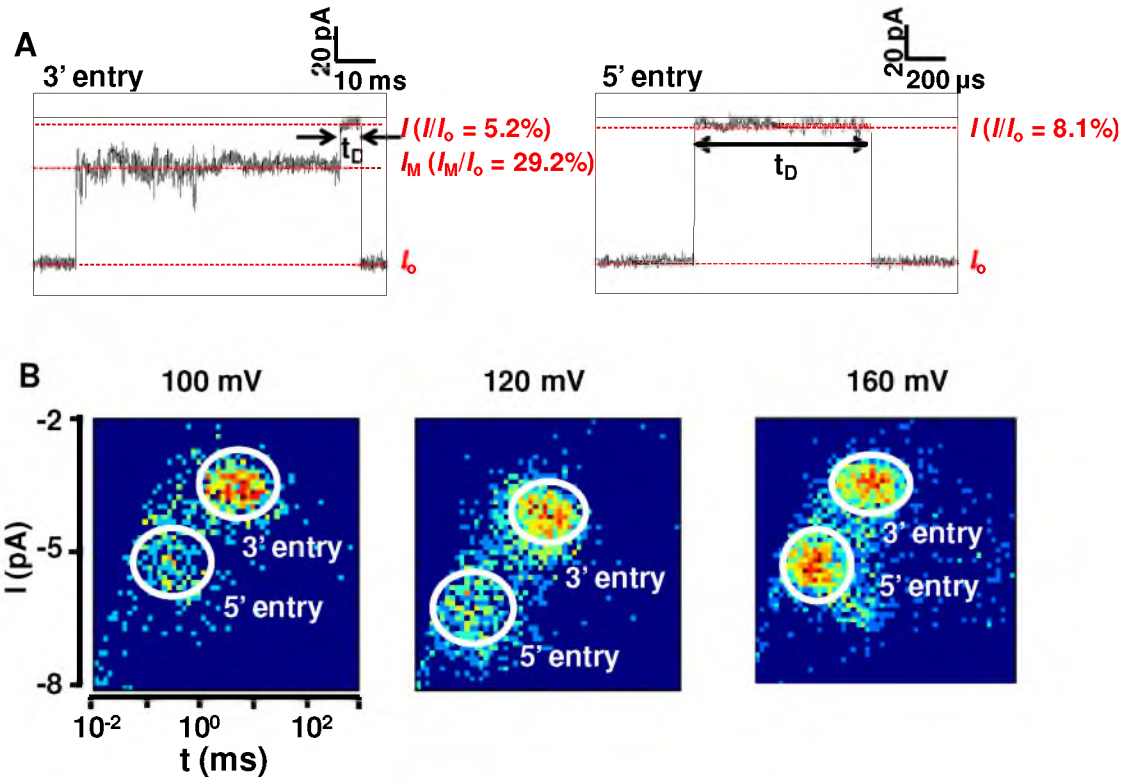


Figure 4.7. Voltage-dependent studies of the human telomere sequence hybrid folds with an A_{25} tail. (A) Plot of applied voltages (*trans* vs. *cis*) as a function of time. (B) Plot of the DC current measured as a function of time. (C) Cartoon illustration of the experiment. Buffer condition: 50.0 mM KCl, 950 mM LiCl, 1.0 mM EDTA, 25.0 mM Tris, pH=7.9.



C

Voltage (mV) *trans* vs. *cis*	Event frequency (population 5'/3')	%$I_{5'}/I_o$	%$I_{3'}/I_o$	T_1 (ms)	T_2 (ms)
100	0.35	8.1 ± 0.7%	5.1 ± 0.4%	1.07 ± 0.12	17.92 ± 1.20
120	0.42	8.2 ± 0.7%	5.0 ± 0.4%	0.55 ± 0.07	12.03 ± 0.91
160	1.39	9.0 ± 0.4%	5.0 ± 0.3%	0.35 ± 0.05	6.50 ± 0.43

Figure 4.8. Studies of the human telomere sequence with an A_{25} tail within the α-HL in 1 M LiCl. (A) Sample *i-t* traces of 3' and 5' entry events at 120 mV (*trans* vs. *cis*). (B) Density plots at 100, 120 and 160 mV (*trans* vs. *cis*). (C) Analysis of durations (decay constant τ), event frequency ratios and percentage residual currents.

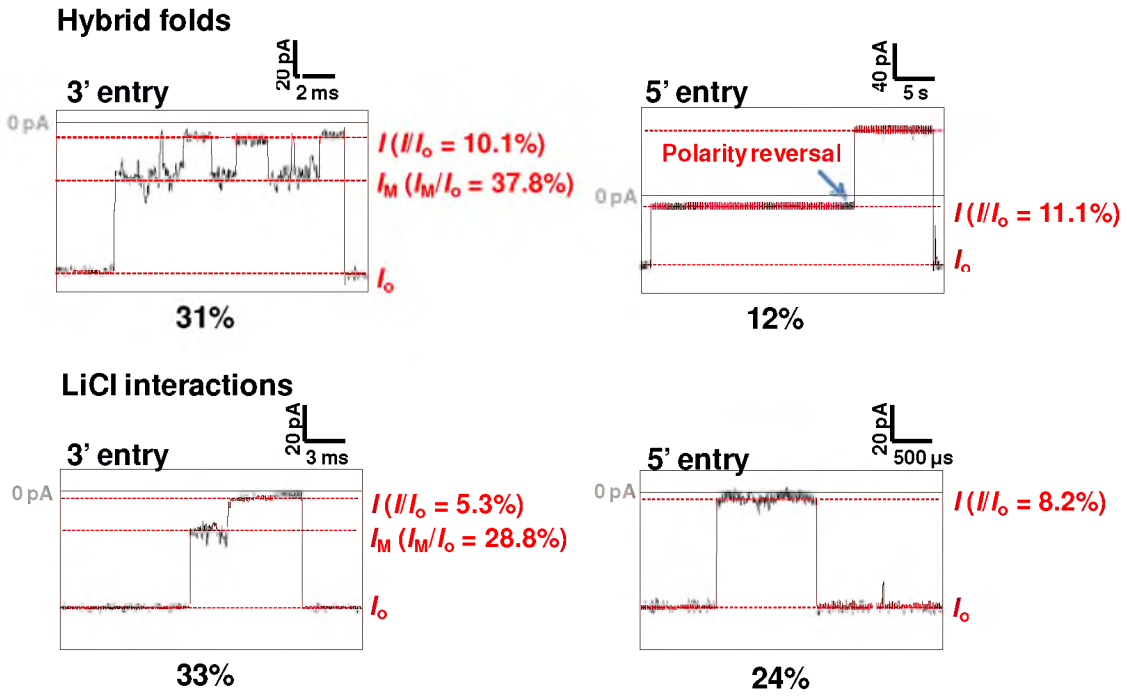


Figure 4.9. Studies of the human telomere sequence with an A_{25} tail with α-HL in low K^+ concentration (1 mM) in 1 M LiCl.

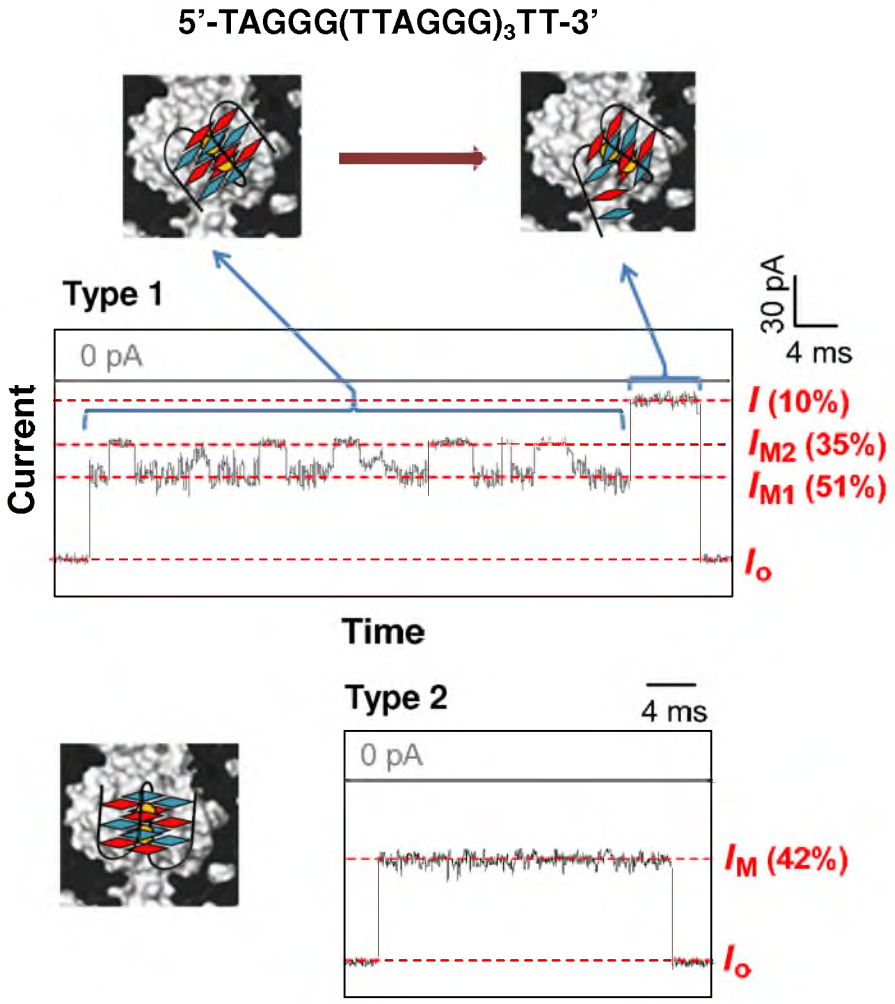


Figure 4.10. Illustrations and sample *i-t* traces for the interactions between the basket fold of the human telomere sequence and the α-HL ion channel at 120 mV (*trans* vs. *cis*) (5.0 μM). Buffer condition: 1.0 M NaCl, 1.0 mM EDTA, 25.0 mM Tris, pH=7.9.

deep-current blockage *I*, transitions between multiple I_M levels were readily observed in the range of $\%I_M/I_o$ = 35-51%, which is also a typically occurrence for hairpins, especially those with terminal base mismatches.[15] Due to its smaller dimensions (Figure 4.1), the basket fold was able to enter the vestibule without much discrimination among its orientations, giving rise to roughly equal amounts of type 1 and 2 events (Figure 4.10).

Human telomeric G-quadruplexes exhibit greater thermostability in the presence of K^+ than Na^+, as measured by melting temperature (Figure 4.2*B*). This is largely dictated by the dehydration energy, ionic radii of the cations, loop topologies and G-quadruplex polymorphism.[8] Consistent with these observations, the basket fold was able to unravel its secondary structure, followed by the translocation to the *trans* side of the α-HL ion channel. This was indicated by the deep-current blockages (*I*) at the end of the events (Figure 4.11), and by the inverse dependence of the durations on the applied voltages.

Interaction between the propeller fold and the α-HL

While both hybrid and basket folds were able to enter into the large vestibule, the diameter of the 'sandwich-like' propeller fold (~4.0 nm) exceeds that of the protein *cis* opening (~3.0 nm) (Figure 4.1), producing only swift and small disturbances to I_o, due to collisions between the vestibule mouth and the propeller G-quadruplex (Figure 4.12). In order to rule out the possibility that the high viscosity of the electrolyte was slowing down the diffusion of the DNA molecules to the protein channel,[32] the same amount of a control sequence (poly-dC_{87}) was added to the same solution, and after being gently mixed, much higher event frequencies were observed with deep current blockages. This

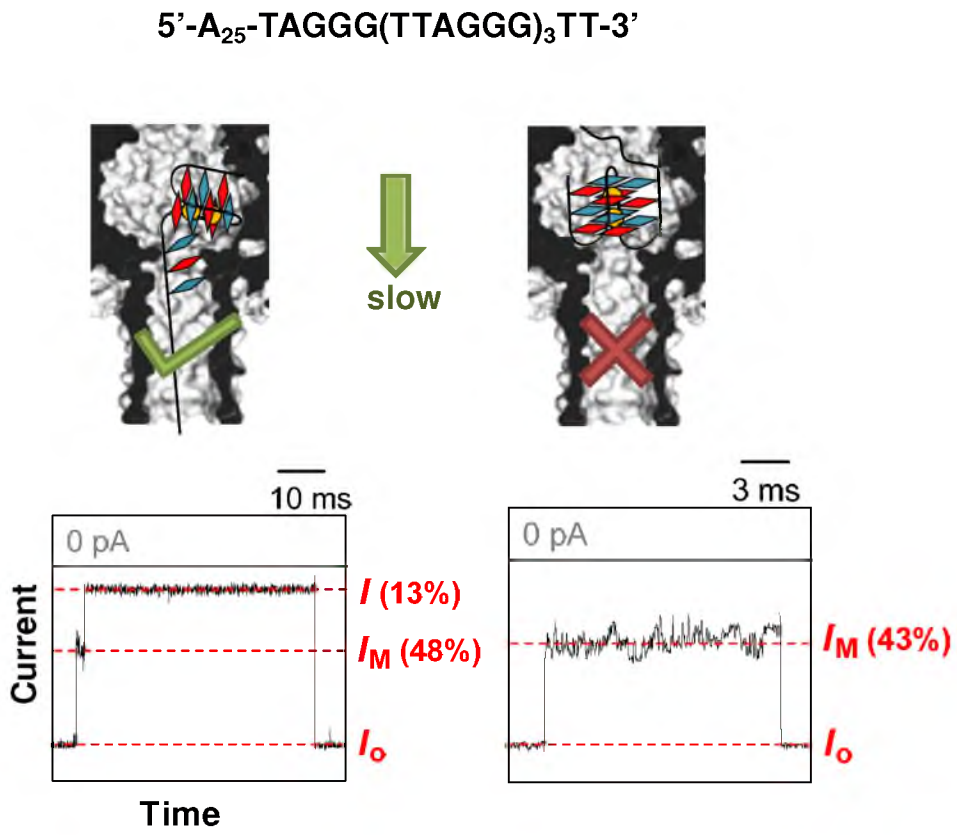


Figure 4.11. Interactions between the basket fold with an A_{25} tail and the protein nanopore at 120 mV (*trans* vs. *cis*). Buffer conditions: 1.0 M NaCl, 1.0 mM EDTA, 25.0 mM Tris, pH=7.9.

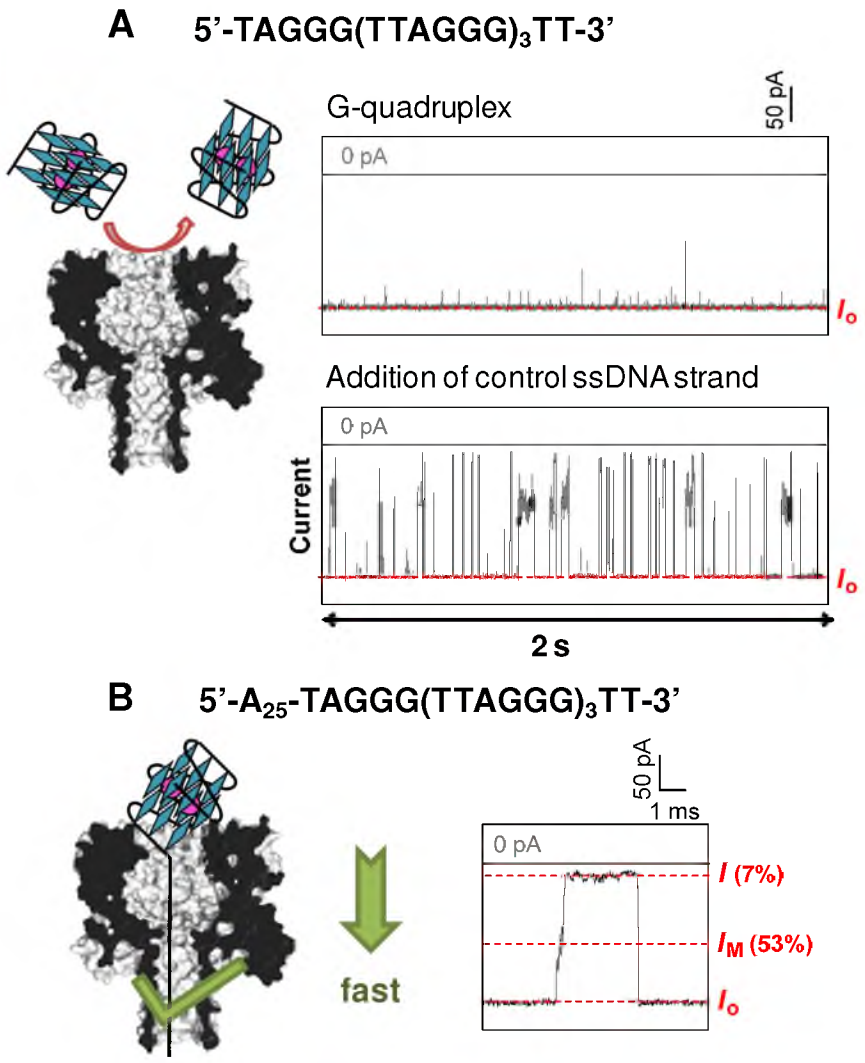


Figure 4.12. Illustrations and sample *i-t* traces for the interactions between the basket fold of the human telomere sequence and the α-HL ion channel at 120 mV (*trans* vs. *cis*) (5.0 μM) (A) and human telomere sequence with an A_{25} tail (5.0 μM) (B). Buffer condition: 1.0 M NaCl, 1.0 mM EDTA, 25.0 mM Tris, pH=7.9.

is an indication that the α-HL could exclude DNA molecules based on their size and shape of their secondary structures. Next, the A_{25} tail was added as a handle for the protein ion channel to unfold the propeller structure, and most interestingly, the G-quadruplex could not only unfold and translocate through the α-HL, but it could also finish these processes extremely fast in comparison to the basket and hybrid folds (Figure 4.13). The quick translocation time likely results from the decrease T_m (Figure 4.2*B*) of the propeller fold under these conditions, and the non-restrictive unraveling environment found outside the vestibule mouth.

Translocation of the human telomeric G-quadruplexes

In order to evaluate the durations of G-quadruplexes unfolding in various buffer conditions, a control sequence was designed with the same overall G content, but was not able to fold up into the quadruplexes. The translocation durations were analyzed in three different electrolytes, and as expected, high viscosity slowed down the DNA movement dramatically.[33] Instead of a Gaussian-like distribution observed for the A- or C-rich sequences,[34] a single-exponential decay model best described the translocation of this G-rich sequence with the decay constant τ (Figure 4.13*A*). When the basket fold was translocating through the ion channel, this stable secondary structure took a long time to unfold, resulting in much longer durations compared to the control sequence (Figure 4.13*B*). However, the propeller fold went through the unraveling and translocation processes only slightly slower than the control strand. Based on the discussion of the results presented above, this could be due to the different locations at which the unfolding occurs. The absence of steric hindrance outside of the α-HL ion channel, compared to the

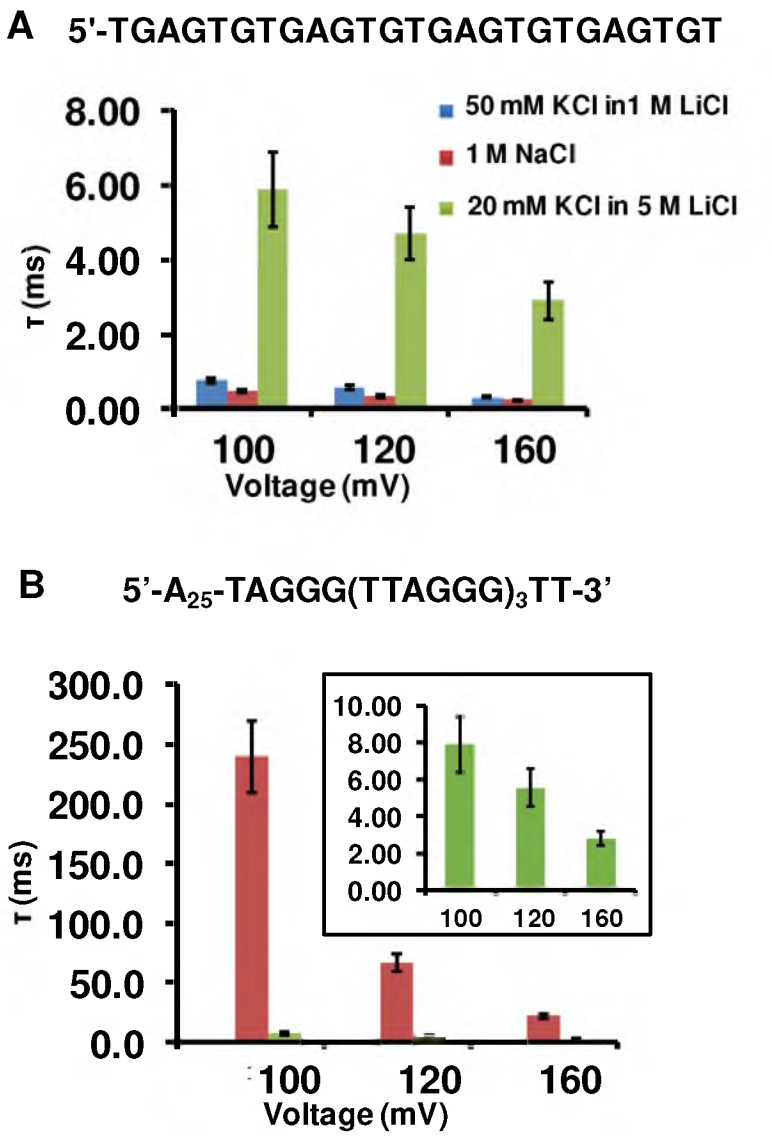


Figure 4.13. Voltage-dependent studies of the event durations for a control sequences (A) and the human telomere sequence (B) in different buffer conditions.

small volume of the protein vestibule (~39.5 nm^3),[35] provided more freedom for the propeller fold to change its secondary structure, while the basket fold had difficulties doing so being trapped inside of the vestibule. Since the hybrid folds were not able to unfold for the time being investigated, they were not plotted.

Discussion

In these studies, the three known structural motifs for the human telomere sequence were studied utilizing the size-limiting properties of the α-HL nanopore. We observed that the hybrid fold, with and without a 5'-tail, does not translocate to the *trans* side of the pore when bound to K^+; the basket fold is capable of translocation through the pore, though, this process is very slow compared to ssDNA; and last, the propeller fold cannot enter into the vestibule because it is wide and flat and unable to enter the vestibule of α–HL. However, the addition of a 5'-tail to the propeller folded sequence facilitates its translocation, which was observed to be fast. What factors determine if a G-quadruplex fold can unravel and traverse to the other side? The T_m values of the structures give an indication of the thermodynamic stabilities of the various folds. Under these high-ionic strength conditions, the hybrid, basket and propeller folds have T_m values of 60.2 ± 0.2, 72.3 ± 0.5 and 60.9 ± 0.8 °C, respectively (Figure 4.2*B*). Immediately, it is apparent that T_m (thermodynamics) is not a predictor of whether a G-quadruplex fold is capable of translocating through the nanopore, because the basket fold, with the highest T_m, gave voltage-dependent data that is characteristic of translocation while the hybrid fold did not show a similar profile. Next, the kinetics of unfolding was compared to our translocation data; previously it was found that in equal concentrations of K^+ and Na^+ (100 mM) the

unfolding rates for the human telomere sequence were 1.3 X 10^{-3} s^{-1} and 4.6 X 10^{-3} s^{-1}, respectively.[36] From these values, the K^+ fold unravels slightly slower than the Na^+ fold; however, these rates are ion-concentration dependent, and in our studies even analysis with 0.001 M K^+, did not provide data characteristic of translocation, while the basket fold in 1 M Na^+ did unfold and translocate. Unfolding kinetic data for the propeller fold were not found in the literature, thus a comparison to the basket and hybrid folds could not be made. From these data, we conclude that translocation of a G-quadruplex through the α-HL nanopore is not very dependent on the thermodynamics or unfolding kinetics of the secondary structure.

The next set of comparisons was made with respect to geometric factors associated with the G-quadruplex secondary structures. Molecular volumes were determined from the ^{3}V webserver[37] in which the values were determined as follows: hybrid folds (hybrid 1 = 10.3 nm^3, pdb 2JSK[5]; hybrid 2 = 11.0 nm^3, pdb 2JSQ[5]), basket fold (12.5 nm^3, pdb 143D[3]), and the propeller fold (11.0 nm^3, pdb 1K8P[9]). Based on volumes alone, α-HL's vestibule (39.5 nm^3) should easily accommodate all three G-quadruplex folds that do not have a 5'-tail. However, it was observed that only the hybrid and basket folds can gain entry into the vestibule, but the propeller fold could not, even though it has the smallest overall volume. We envision the propeller cannot enter the vestibule, because of its "sandwich-like" shape as previously described. Moreover, the basket fold has a larger volume than the hybrid folds, though the basket can unravel within the vestibule while the hybrid cannot. Therefore, the ability to translocate is not wholly dependent on the requirement that there to be enough room to unravel within the vestibule, because the basket fold translocates and is the largest. Lastly, as previously

stated the hybrid fold cannot translocate even under high force (200 mV). Thus, we conclude that the following G-quadruplex parameters do not determine if a fold can translocate: Thermodynamic stability, kinetics of unfolding, or molecular volume. Hence, by process of elimination we hypothesize that molecular shape is the key factor that determines if a G-quadruplex fold can traverse through α-HL or not.

Molecular shape for G-quadruplexes is determined by multiple factors, which include strand polarity, the *syn* vs. *anti* conformation for the G nucleotides in which these properties determine the loop topologies for a given fold. The basket fold is capable of translocation, which means it can unravel such that the single-strand can pass the β-barrel. These events are initiated from either the 3' or 5' side, and start by opening up the first tetrad. For the strand to move through the β-barrel the remaining G-quadruplex has to roll within the vestibule, which is a favorable process because the basket fold is cylindrically shaped (Figure 4.14). The addition of a 5'-tail to the basket fold helps to initiate threading and unraveling for translocation to occur. The propeller fold represents the other limiting case in which this shape is not capable of entry into the mouth of the vestibule. Based on the propeller fold's dimensions (4.0 X 2.1 nm) it cannot pass the circular mouth of the vestibule (d = 3.0 nm; Figure 4.14). This represents a similar phenomenon that is observed with trying to place a square peg into a round hole. Addition of the 5'-tail circumvents the entry problem into the vestibule in which unraveling the propeller fold can now occur outside the mouth of the vestibule mouth. The data present a fast unraveling time for this fold. We conclude this result is from the freedom to rotate outside the vestibule without any steric or molecular interactions that would occur within the interior of the vestibule.

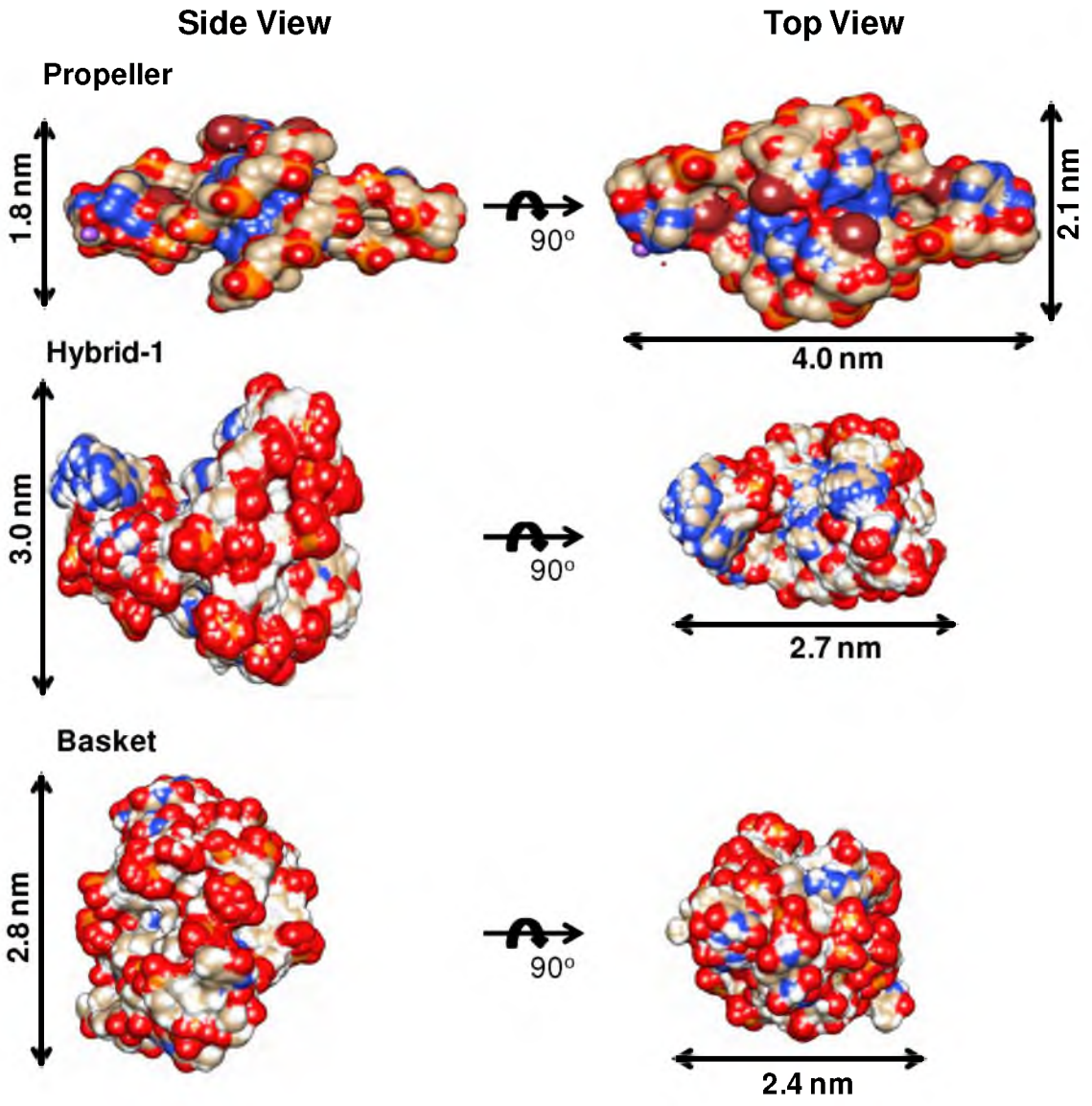


Figure 4.14. Molecular shapes for the G-quadruplexes studied. The folds shown are the propeller (top, pdb 1K8P),[9] hybrid-1 (middle, pdb 2JSK),[5] and basket (bottom, pdb 143D).[3]

Conclusions

Interactions of these G-quadruplexes with the α-HL ion channel produced various current patterns that were characteristic for the dimensions and loop topologies of different cation-selective secondary structures. While the basket fold was able to unravel within the ion channel, the hybrid folds had difficulty unfolding for the time period being investigated (~ 10 min). On the contrary, the propeller fold could not enter into the α-HL ion channel due to its 'sandwich-like' shape and larger dimensions presenting a square peg/round hole problem; however, it could unfold outside of the protein vestibule much faster compared to the basket fold with the assistance of the A_{25} tail. Given the substantial amount of time the G-quadruplexes can be trapped inside of the protein nanocavity, we have the opportunity to investigate these secondary structures at a single-molecule level. Additionally, the α-HL ion channel could interrogate DNA molecules based on their size and shape. Biological nanopores have been under extensive investigation as a popular technology for Next-Generation DNA sequencing,[38-46] and various sequences of ssDNA, especially telomere and promoter regions of certain oncogenes,[47, 48] can fold up into stable G-quadruplexes under physiological and nanopore measurement conditions. Thus, the work presented here could provide some insight into the potential challenges that will occur once nanopore sequencing is commenced on actual eukaryotic samples.

References

1. Choi, J.; Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011,** 40, 5893-5909.

2. Patel, D. J.; Phan, A. T.; Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **2007,** 35, 7429-7455.

3. Wang, Y.; Patel, D. J. Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex. *Structure* **1993,** 1, 263-282.

4. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **2006,** 34, 2723-2735.

5. Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K^+ solution. *Nucleic Acids Res.* **2007,** 35, 6517-6525.

6. Xu, Y.; Noguchi, Y.; Sugiyama, H. The new models of the human telomere d[$AGGG(TTAGGG)_3$] in K^+ solution. *Bioorg. Med. Chem.* **2006,** 14, 5584-5591.

7. Phan, A. T.; Luu, K. N.; Patel, D. J. Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K^+ solution. *Nucleic Acids Res.* **2006,** 34, 5715-5719.

8. Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008,** 36, 5482-5515.

9. Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002,** 417, 876-880.

10. Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3'-Sequence of d($AG_3(TTAG_3)_3$). *J. Am. Chem. Soc.* **2010,** 132, 17105-17107.

11. Heddi, B.; Phan, A. T. Structure of Human Telomeric DNA in Crowded Solution. *J. Am. Chem. Soc.* **2011,** 133, 9824-9833.

12. Lannan, F. M.; Mamajanov, I.; Hud, N. V. Human Telomere Sequence DNA in Water-Free and High-Viscosity Solvents: G-Quadruplex Folding Governed by Kramers Rate Theory. *J. Am. Chem. Soc.* **2012**.

13. Muzard, J.; Martinho, M.; Mathe, J.; Bockelmann, U.; Viasnoff, V. DNA translocation and unzipping through a nanopore: some geometrical effects. *Biophys. J.* **2010,** 98, 2170-2178.

14. Mathe, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Meller, A. Nanopore unzipping of individual DNA hairpin molecules. *Biophys. J.* **2004,** 87, 3205-3212.

15. Vercoutere, W. A.; Winters-Hilt, S.; DeGuzman, V. S.; Deamer, D.; Ridino, S. E.; Rodgers, J. T.; Olsen, H. E.; Marziali, A.; Akeson, M. Discrimination among individual Watson-Crick base pairs at the termini of single DNA hairpin molecules. *Nucleic Acids Res.* **2003,** 31, 1311-1318.

16. DeGuzman, V. S.; Lee, C. C.; Deamer, D. W.; Vercoutere, W. A. Sequence-dependent gating of an ion channel by DNA hairpin molecules. *Nucleic Acids Res.* **2006,** 34, 6425-6437.

17. Jin, Q.; Fleming, A. M.; Burrows, C. J.; White, H. S. Unzipping kinetics of duplex DNA containing oxidized lesions in an α-hemolysin nanopore. *J. Am. Chem. Soc.* **2012,** 134, 11006-11011.

18. Schibel, A. E. P.; Fleming, A. M.; Jin, Q.; An, N.; Liu, J.; Blakemore, C. P.; White, H. S.; Burrows, C. J. Sequence-specific single-molecule analysis of 8-oxo-7,8-dihydroguanine lesions in DNA based on unzipping kinetics of complementary probes in ion channel recordings. *J. Am. Chem. Soc.* **2011,** 133, 14778-14784.

19. Sauer-Budge, A. F.; Nyamwanda, J. A.; Lubensky, D. K.; Branton, D. Unzipping kinetics of double-stranded DNA in a nanopore. *Phys. Rev. Lett.* **2003,** 90, 23801.

20. Howorka, S.; Cheley, S.; Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat. Biotechnol.* **2001,** 19, 636-639.

21. Shim, J. W.; Gu, L.-Q. Single-molecule investigation of G-quadruplex using a nanopore sensor. *Methods* **2012,** ahead of print.

22. Shim, J. W.; Gu, L.-Q. Encapsulating a single G-quadruplex aptamer in a protein nanocavity. *J. Phys. Chem. B* **2008,** 112, 8354-8360.

23. Shim, J. W.; Tan, Q.; Gu, L.-Q. Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity. *Nucleic Acids Res.* **2008,** 37, 972-982.

24. Rotem, D.; Jayasinghe, L.; Salichou, M.; Bayley, H. Protein detection by nanopores equipped with aptamers. *J. Am. Chem. Soc.* **2012,** 134, 2781-2787.

25. Zhang, B.; Galusha, J.; Shiozawa, P. G.; Wang, G.; Bergren, A. J.; Jones, R. M.; White, R. J.; Ervin, E. N.; Cauley, C. C.; White, H. S. Bench-top method for

fabricating glass-sealed nanodisk electrodes, glass nanpore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **2007,** 79, 4778-4787.

26. White, R. J.; Ervin, E. N.; Yang, T.; Chen, X.; Daniel, S.; Cremer, P. S.; White, H. S. Single ion-channel recordings using glass nanopore membranes. *J. Am. Chem. Soc.* **2007,** 129, 11766-11775.

27. Karsisiotis, A. I.; Hessari, N. M. a.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. Topological Characterization of Nucleic Acid G-Quadruplexes by UV Absorption and Circular Dichroism. *Angew. Chem., Int. Ed.* **2011,** 50, 10645-10648.

28. Blandamer, M. J.; Engberts, J. B. F. N.; Gleeson, P. T.; Reis, J. C. R. Activity of water in aqueous systems; A frequently neglected property. *Chem. Soc. Rev.* **2005,** 34, 440-458.

29. Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M. Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nat. Biotechnol.* **2001,** 19, 248-252.

30. Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H. Folding Pathways of Human Telomeric Type-1 and Type-2 G-Quadruplex Structures. *J. Am. Chem. Soc.* **2010,** 132, 14910-14918.

31. Gray, R. D.; Chaires, J. B. Kinetics and mechanism of K^+- and Na^+-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res.* **2008,** 36, 4191-4203.

32. Kawano, R.; Schibel, A. E. P.; Cauley, C.; White, H. S. Controlling the translocation of single-stranded DNA through α-hemolysin ion channels using viscosity. *Langmuir* **2009,** 25, 1233-1237.

33. Kowalczyk, S. W.; Wells, D. B.; Aksimentiev, A.; Dekker, C. Slowing down DNA Translocation through a Nanopore in Lithium Chloride. *Nano Lett.* **2012,** 12, 1038-1044.

34. Meller, A.; Nivon, L.; Brandin, E.; Golovchenoko, J.; Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** 97, 1079-1084.

35. Jung, Y.; Cheley, S.; Braha, O.; Bayley, H. The internal cavity of the staphylococcal α-hemolysin pore accommodates ~175 exogenous amino acid residues. *Biochemistry* **2005,** 44, 8919-8929.

36. Zhao, Y.; Kan, Z.-y.; Zeng, Z.-x.; Hao, Y.-h.; Chen, H.; Tan, Z. Determining the Folding and Unfolding Rate Constants of Nucleic Acids by Biosensor.

Application to Telomere G-Quadruplex. *J. Am. Chem. Soc.* **2004,** 126, 13255-13264.

37. Voss, N. R.; Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010,** 38, W555-W562.

38. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Ventra, M. D.; Garaj, S.; Hibbs, A.; Huang, X.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **2008,** 26, 1146-1153.

39. Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Karplus, K.; Akeson, M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precesion. *Nat. Biotechnol.* **2012,** 30, 343-348.

40. Manrao, E. A.; Derrington, I. M.; Laszlo, A. H.; Langford, K. W.; Hopper, M. K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **2012,** 30, 349-353.

41. Banerjee, A.; Mikhailova, E.; Cheley, S.; Gu, L.; Montoya, M.; Nagaoka, Y.; Gouaux, E.; Bayley, H. Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. . *Proc. Natl. Acad. Sci. U.S.A.* **2010,** 107, 8165-8170.

42. Clarke, J.; Wu, H.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **2009,** 4, 265-270.

43. Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** 106, 7702-7707.

44. Maglia, G.; Restrepo, M. R.; Mikhailova, E.; Bayley, H. Enhanced translocation of single DNA molecules through α-hemolysin nanopores by manipulation of internal charge. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** 105, 19720-19725.

45. Howorka, S.; Siwy, Z. Nanopore analytics: sensing of single molecules. *Chem. Soc. Rev.* **2009,** 38, 2360-2384.

46. Maglia, G.; Heron, A. J.; Stoddart, D.; Japrung, D.; Bayley, H. Analysis of single nucleic acid molecules with protein nanopores. *Methods in Enzymology* **2010,** 475, 591-623.

47. Paramasivam, M.; Cogoi, S.; Xodo, L. E. Primer extension reactions as a tool to uncover folding motifs within complex G-rich sequences: analysis of the human *KRAS* NHE. *Chem. Commun.* **2011,** 47, 4965-4967.

48. Qin, Y.; Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **2008,** 90, 1149-1171.

CHAPTER 5

NANOPORE STUDIES OF OXIDATIVE DAMAGE TO THE HUMAN TELOMERIC G-QUADRUPLEXES

Introduction

The human telomere is composed of 5-15 kbps of a repeat sequence 5'-TTAGGG-3', which is terminated with a 100-200 nt single-stranded overhang.[1] Though essential for many cellular functions, telomeres are prone to progressive shortening due to the end-replication problem, which was proposed to serve as a molecular clock for cellular lifespan.[2, 3] The existence of telomerase and a homologous recombination mechanism were shown to elongate telomeres;[4, 5] however, there are still fundamental problems that need addressing. Natural chromosome ends can be recognized as double-stranded breaks, which initiates the DNA-repair response, but the telomere structure provides protection from inappropriate pathway. Besides various telomere-binding proteins, higher order structures formed by these repeat sequences were suggested to have great contributions to this function, which is also known as the telomeric cap.[6-10]

Human telomeres and telomerase have been targeted in cancer therapy, because cancer cells were shown to be able to maintain their telomere length.[11] Meanwhile, a correlation has been found between telomere shortening and the exposure to oxidative

and inflammation stress.[12, 13] As the most popular higher order telomeric structure proposed for the telomere capping, quadruplexes are composed from a G-rich sequence, which is a dominant site for oxidative damage, given that this base has the lowest redox potential.[14-16] The G-oxidation products can no longer form Hoogsteen base pairs with the other Gs on the same tetrad, leading to disruption of the secondary structure (Figure 5.1). Recently, O'Callaghan et al. have discovered that upon oxidative stress, oxidized-base lesions in telomeric DNA are higher than in other genomic regions. However, this estimate is complicated by the lack of methods that could provide both sequence information and identity of the DNA lesions.

The α-hemolysin (α-HL) nanopore has been shown to present promising potential as a tool to study secondary structures, including hairpins, duplexes, and quadruplexes. As discussed in Chapter 4, we used this technology to understand the interactions between human telomeric G-quadruplexes and this protein ion channel, which provided insights into the stabilities and shapes of these structures. Herein, we combined such understanding and the site-specific chemistry of attaching the detectable label 18-crown-6 to DNA, which was discussed in Chapter 2,[17] to address some questions at a molecular level. And we chose 8-oxo-7,8-dihydroguanine (OG), a keybiomarker for oxidative stress, as a point of interest.

Experimental Section

Materials

KCl, NaCl, LiCl, LiOH, EDTA, Tris, phospholipid 1,2-diphytanoyl-sn-gly-cero-3-phosphocholine (DPhPC), 3-cyanopropyldimethylchlorosilane, 2-aminomethyl-18-

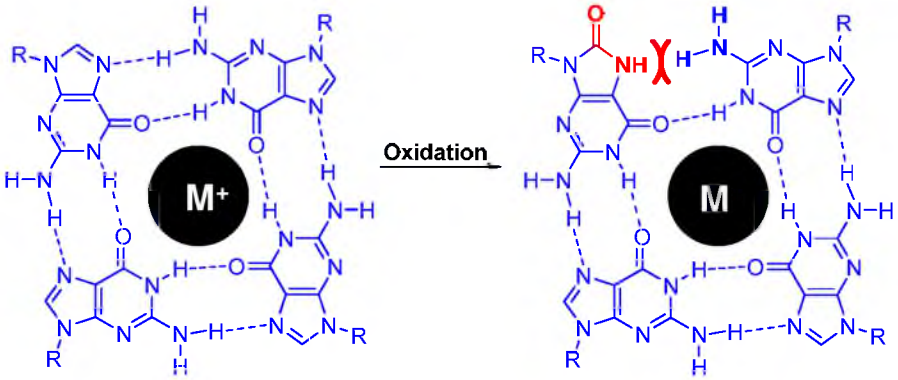


Figure 5.1. Oxidative damage to the G-quadruplex disrupts the hydrogen bonding of that tetrad.

crown-6 (18c6), wild-type α-HL monomer, ethanol, acetonitrile, decane were purchased from commercial suppliers and used without further purification.

Oligodeoxynucleotide preparation and characterization

The oligodeoxynucleotides (ODN) were synthesized by the DNA-Peptide Core Facilty at the University of Utah with commercially available phosphoramidites (Glen Research, Sterling, VA). After being cleaved from the synthetic column and deprotected according to the manufacturer's protocols, ODNs were purified and characterized as described in Chapter 4. The ODN OG 18c6 adduct was synthesized by the method previously published by the Burrows laboratory.[18, 19] Briefly, the OG-containing ODN (10 μM) was thermally equilibrated with 18c6 (2 mM) in 75 mM NaPi buffer (pH 8.0) at 45 °C for 30 min, followed by the titration of 15 equivalents of Na_2IrCl_6 (150 μM). The reaction was kept for 30 min; then quenched with Na_2EDTA (pH 8.0, 1 mM). All the ODNs were characterized by negative ion electron spray (ESI-) on a Micromass Quattro II mass spectrometer equipped with Zspray API source in the mass spectrometry laboratory at the Department of Chemistry, University of Utah.

Nanopore measurements

The α-HL ion channel recording measurements and the data analysis were performed the same way as described in Chapter 4.

Results and Discussion

Stability of the OG-containing G-quadruplexes

Both OG-containing hybrid and basket folds presented faster translocation as they were driven through the α-HL ion channel, which was evaluated by fitting the event durations with a single exponential decay model with the decay constant τ. As discussed in Chapter 4, the undamaged hybrid folds were not able to pass the constriction of the protein for the time being studied, while the undamaged basket folds took a long time to unravel their structures. This is indicating that the presence of the oxidative damage decreased the stabilities of these G-quadruplexes.

Since the human telomeric G-quadruplexes have three tetrads, the location of the oxdative damage should also be a factor on its destabilization effect. We incorporated OG into various positions, and they presented rather different translocation decay constants (Figure 5.2). Placement of OG in a top or bottom tetrad results in destabilization and unfolding of that layer, whereas, formation of OG in a middle tetrad leads to complete unfolding of the G-quadruplex. Additionally, the translocation of different OG-containing quadruplexes from *cis* to *trans* of the protein channel have event durations that exhibit a linear correlation with the unfolding ΔG, calculated from T_m measurements. Surprisingly, breaking a single hydrogen bond (G→OG) could destabilize the quadruplex structure by as much as -20 °C (ΔT_m) and cause ~ 100-fold faster translocation duration (τ) in the nanopore experiments.

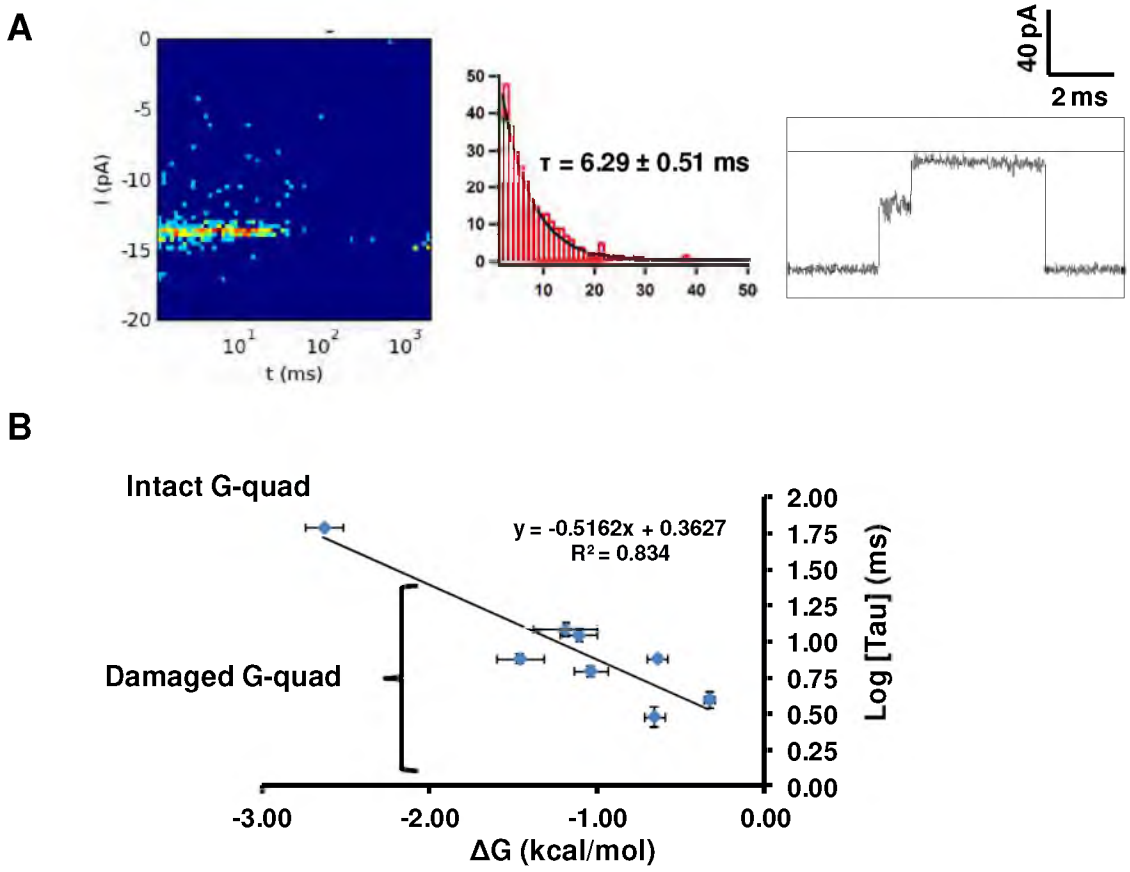


Figure 5.2. The α-HL nanopore studies of the OG-containing G-quadruplexes. (A) sample density plot, duration distribution, and *i-t* trace. (B) plot of log (τ) vs. ΔG for various OG locations.

Detection of OG in the G-quadruplex

We took advantage of the detectable label 18-crown-6 (18c6) and the site-specific chemistry developed in our laboratory, to form the OG 18c6 adduct at position 3 (Figure 5.3).[18, 19] The resulting bulky adduct yielded distinct current signatures ($I_A \rightarrow I_B \rightarrow I_A$) in 1 M NaCl as the electrolyte, indicating the presence of this specific lesion. This provided a potential opportunity to better understand the location and identity of oxidative damage to the telomere, and its correlation to telomere shortening.

Future Studies

Future studies will be focused on the application of this method to a larger biological sample. Normal cells and the ones under oxidative stress will be evaluated in terms of the translocation duration of their telomeres. And the site-specific chemistry will be performed to attach 18c6 as a way to detect the formation of OG under these circumstances and its location. Furthermore, recent unpublished data from Dr. Aaron Fleming in our laboratory has identified hotspots for G oxidation to occur. One of the most common products observed was spiroiminodihydantoin, a four-electron oxidation product of G. Further studies conducted with this oxidative lesion will be of interest.

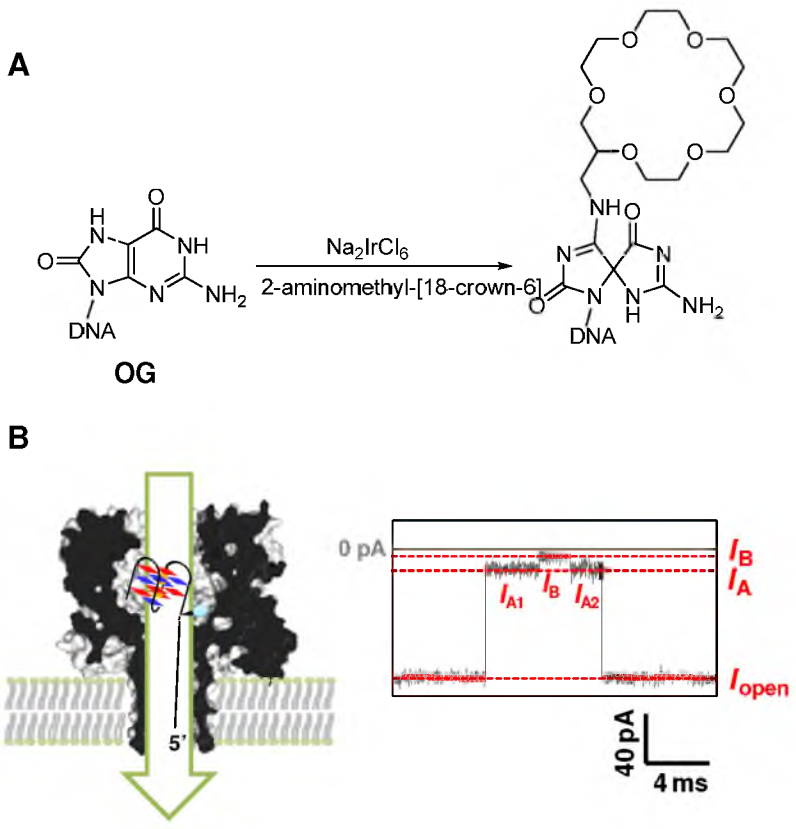


Figure 5.3. Detection of the presence of OG (position 3) in the G-quadruplex by site-specific attachment of 18c6 in 1 M NaCl. (A) adduct synthesis scheme (B) sample *i-t* trace.

References

1. Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **1988,** 85, 6622-6626.

2. Xu, Y. Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chem. Soc. Rev.* **2011,** 40, 2719-2740.

3. Samassekou, O.; Gadji, M.; Drouin, R.; Ju, Y. Sizing the ends: Normal length of human telomeres. *Ann. Anat.* **2010,** 192, 284-291.

4. Greider, C. W.; Blackburn, E. H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **1985,** 43, 405-413.

5. Dunham, M. A.; Neumann, A. A.; Fasching, C. L.; Reddel, R. R. Telomere maintenance by recombination in human cells. *Nat. Genet.* **2000,** 26, 447-450.

6. Blackburn, E. H. Telomere states and cell fates. *Nature* **2000,** 408, 53-56.

7. Cech, T. R. Beginning to understand the end of the chromosome. *Cell* **2004,** 116, 273-279.

8. Blasco, M. A. Telomeres and human disease: Aging, cancer and beyond. *Nat. Rev. Genet.* **2005,** 6, 611-622.

9. Verdun, R. E.; Karlseder, J. Replication and protection of telomeres. *Nature* **2007,** 447, 924-931.

10. de Lange, T. How telomeres solve the end-protection problem. *Science* **2009,** 326, 948-952.

11. Lantuejoul, S.; Salon, C.; Soria, J.-C.; Brambilla, E. Telomerase expression in lung preneoplais and neoplasia. *Int. J. Cancer* **2007,** 120, 1835-1841.

12. Epel, E. S.; Blackburn, E. H.; Lin, J.; Dhabhar, F. S.; Adler, N. E.; Morrow, J. D.; Cawthon, R. M. Accelerated telomere shortening in response to life stress. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, 17312-17315.

13. Wolkowitz, O. M.; Mellon, S. H.; Epel, E. S.; Lin, J.; Dhabhar, F. S.; Su, Y.; Reus, V. I.; Rosser, R.; Burke, H. M.; Kupferman, E.; Compagnone, M.; Nelson, J. C.; Blackburn, E. H. Leukocyte telomere length in major depression: correlations with chronicity, inflammation and oxidative stress - preliminary findings. *PLoS One* **2011,** 6, e17837.

14. Steenken, S.; Jovanovic, S. V. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J. Am. Chem. Soc* **1997,** 119, 617-618.

15. Cadet, J.; Douki, T.; Ravanat, J. Oxidatively generated damage to the guanine moiety of DNA: mechanistic aspects and formation in cells. *Acc. Chem. Res* **2008,** 41, 1075-1083.

16. Delaney, S.; Jarem, D. A.; Volle, C. B.; Yennie, C. J. Chemical and biological consequences of oxidatively damaged guanine in DNA. *Free Radical Res.* **2012,** 46, 420-441.

17. An, N.; Fleming, A. M.; White, H. S.; Burrows, C. J. Crown ether-electrolyte interactions permit nanopore detection of individual DNA abasic sites in single molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** 109, 11504-11509.

18. Xu, X.; Muller, J. G.; Ye, Y.; Burrows, C. J. DNA-protein cross-links between guanine and lysine depend on the mechanism of oxidation for formation of C5 vs C8 guanosine adducts. *J. Am. Chem. Soc* **2008,** 130, 703-709.

19. Hosford, M. E.; Muller, J. G.; Burrows, C. J. Spermine participates in oxidative damage of guanosine and 8-oxogunosine leading to deoxyribosylurea formation. . *J. Am. Chem. Soc* **2004,** 126, 9540-9541.