SINGLE-MOLECULE ANALYSIS OF DNA CROSS-LINKS USING NANOPORE TECHNOLOGY

by

Anna H. Wolna

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 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Anna H. Wolna**

has been approved by the following supervisory committee members:

Cynthia J. Burrows	, Chair	**07.21.14** Date Approved
Jennifer M. Heemstra	, Member	**07.21.14** Date Approved
Bethany A. Buck-Koehntop	, Member	**07.21.14** Date Approved
Marc D. Porter	, Member	**07.21.14** Date Approved
David F. Blair	, Member	**07.21.14** Date Approved

and by **Cynthia J. Burrows**, Chair/Dean of

the Department/College/School of **Chemistry**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The alpha-hemolysin (α-HL) protein ion channel is a potential next-generation sequencing platform that has been extensively used to study nucleic acids at a single-molecule level. After applying a potential across a lipid bilayer, the imbedded α-HL allows monitoring of the duration and current levels of DNA translocation and immobilization. Because this method does not require DNA amplification prior to sequencing, all the DNA damage present in the cell at any given time will be present during the sequencing experiment. The goal of this research is to determine if these damage sites give distinguishable current levels beyond those observed for the canonical nucleobases. Because DNA cross-links are one of the most prevalent types of DNA damage occurring *in vivo,* the blockage current levels were determined for thymine-dimers, guanine(C8)-thymine(N3) crosslinks and platinum adducts. All of these cross-links give a different blockage current level compared to the undamaged strands when immobilized in the ion channel, and they all can easily translocate across the α-HL channel.

Additionally, the α-HL nanopore technique presents a unique opportunity to study the effects of DNA cross-links, such as thymine-dimers, on the secondary structure of DNA G-quadruplexes folded from the human telomere sequence. Using this single-molecule nanopore technique we can detect subtle structural differences that cannot be

easily addressed using conventional methods.

The human telomere plays crucial roles in maintaining genome stability. In the presence of suitable cations, the repetitive 5'-TTAGGG human telomere sequence can fold into G-quadruplexes that adopt the hybrid fold *in vivo*. The telomere sequence is hypersensitive to UV-induced thymine-dimer (T=T) formation, and yet the presence of thymine dimers does not cause telomere shortening. The potential structural disruption and thermodynamic stability of the T=T-containing natural telomere sequences were studied to understand how this damage is tolerated in telomeric DNA. The α-HL experiments determined that T=Ts disrupt double-chain reversal loop formation but are well tolerated in edgewise and diagonal loops of the hybrid G-quadruplexes. These studies demonstrated the power of the α-HL ion channel to analyze DNA modifications and secondary structures at a single-molecule level.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

5-mC: 5-methylcytosine

8-oxo-A: 8-oxo-7,8-dihydro-2'-deoxyadenosine

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

DNA: deoxyribonucleic acid

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

DSB: double-strand break

EBS: electronic biosciences

EDTA: ethylenediamminetetraacetic acid

EMSA: electrophoretic mobility shift assays

ESI: electrospray ionization

Gh: guanidinohydantoin

GlcN: D-(+)-glucosamine

GNM: glass nanopore membrane

GPRP: Gly-Pro-Arg-Pro amide

HPLC: high performance liquid chromatography

i-t: current-time

MALDI: matrix-assisted laser desorption/ionization

MS: mass spectrometry

MspA: *Mycobacterium smegmatis* porin A

NER: nucleotide-excision repair

NMR: nuclear magnetic resonance

ODN: oligodeoxynucleotide

OG: 8-oxo-7,8-dihydroguanine

RH: Arg-His

ROS: reactive oxygen species

ssDNA: single-stranded DNA

Sp: spiroiminodihydantoin

STM: streptomycin

Strep: streptavidin

TBA: thrombin-binding aptamer

$\mathbf{t_D}$: duration

Tg: thymine glycol

THF: tetrahydrofuran

$\mathbf{T_m}$: melting temperature

TOF: time of flight

Trn: taurine

T=T: [*cis-syn*] thymine dimer

U: uracil

UDG: uracil-DNA glycosylase

UV: ultraviolet

ACKNOWLEDGMENTS

The body of work presented here would not be possible without constant support, encouragement, and constructive critique by many people to whom I would like to offer my deepest thanks. First, I would like to acknowledge Prof. Cynthia J. Burrows for allowing me to join her laboratory, and then supporting, guiding and inspiring me through the duration of my Ph.D. studies. Next, I would like to thank Prof. Aaron M. Fleming who was always there to offer help. His dedication and enthusiasm toward science was inspiring, and I'm grateful for his friendship. My thanks also go to our nanopore project collaborator, Prof. Henry S. White, and my committee members, Prof. Jennifer M. Heemstra, Prof. Bethany A. Buck-Koehntop, Prof. Marc D. Porter, and Prof. David F. Blair.

I think that during my five years at the University of Utah I was very fortunate to work with a great team of people that I will always call my friends. They have made my life in the laboratory and outside of it much more exciting. They were always there to help me learn new techniques, listen to my practice talks, proof-read my manuscripts, and other written documents. I would like to give my thanks to many of the Burrows lab members and especially to: Yun Ding, Dr. Pranjali Ghude, Dr. Na An, Judy Zhu, Dr. Jan Riedl, Xibo Li, Omar Alshykhly, and Anton Alenko. I would also like to acknowledge the White lab research group members that were collaborating on the nanopore project

with me.

Furthermore, my thanks go to my friends: Katarzyna Kaczmarek, Joanna Gajewiak, Grzegorz Gajewiak, Katarzyna Rzepecka, Piotr Rzepecki, Gonzalo Baraza, Agnieszka Szarzec-Larsen, and Michael Larsen who were always there to support and encourage me. They have made my stay in Salt Lake City enjoyable, and thanks to them I really felt “at home" here. Finally, I’m eternally grateful to my family: my husband, Sebastian M. Wolny, who was always there for me; my parents, Halina Jedrzkiewicz and Kazimierz Jedrzkiewicz, who believed I could always do whatever it was I set out to do; my sister, Jolanta D. Jedrzkiewicz, who is my best friend and always there to listen; and finally my grandmothers, Wiktoria Jedrzkiewicz and Zofia Szczotka, who encouraged me to finish what I started. Without their support and encouragements this work would not be possible. Thank you.

CHAPTER 1

INTRODUCTION

Nanopore technology

Emerging as a rapid and inexpensive single-molecule DNA sequencing platform, nanopore ion channel technology has been under intensive investigation.[1-4] The bacterial ion-channel protein α-hemolysin (α-HL) has been extensively studied as a potential next-generation sequencing platform. This protein can self-assemble into a heptamer across a lipid bilayer, producing an aqueous channel that is large enough to accommodate single-strand DNA (ssDNA). Application of an electrical potential across the protein drives the negatively charged ssDNA to translocate through the channel while significantly reducing the current flow.[5] The resulting blockage in the current level, and the duration of the translocation event, is envisioned to be utilized for identification of the sequence of nucleotides. The nanopore sequencing platform offers advantages that the current methods do not provide:[6] (1) The ideal system would require no sample amplification and minimal sample preparation, because analysis would occur on a single molecule. (2) The ideal instrumentation would be small and portable. These features combined would deliver a small and inexpensive platform for sequencing DNA. The outlook for nanopore-sequencing technology is promising; however, challenges still remain that need to be resolved before this method is a viable technology.[2]

The α-HL ion channel is secreted by *Staphylococcus aureus,* and it consists of a larger vestibule with an interior dimension of ~4.5 nm that in turn leads to a β-barrel that penetrates the lipid bilayer with its narrowest constriction point being ~1.4 nm in diameter (Figure 1.1).[7] The constriction zone is too narrow for double-stranded DNA (dsDNA, d ~ 2 nm) to be translocated; however, ssDNA (d ~ 1 nm) can readily pass through the narrowest point. This poses a challenge associated with nanopore sequencing, because the DNA must be single stranded in order to translocate through the ion channel. One method to deal with this problem is to use high concentrations of urea; however, this decreases the conductance.[8] In a typical experiment, the negative electrode is placed with the DNA on the *cis* side of α-HL and the positive electrode is placed on the *trans* side. This arrangement causes DNA to migrate from the *cis* compartment to the *trans* compartment when an electrical field is applied across the channel. When the channel is not occluded by DNA, an open channel current is recorded (I_o), and during the translocation of DNA the current is blocked up to ~85% of I_o. To date, the blockage current levels for the various DNA nucleotides have been established with long single strands (40-100mers) of homopolymer DNA.[9-11] Translocation times typically range from 1-20 μs/nucleotide when operating above the critical voltage for DNA capture at room temperature, and these times are so short that the acquisition rate must be in the MHz range to observe single nucleotides.[2] Current electronics are not capable of detecting the small differences in current between adjacent nucleotides at this acquisition rate due to the inherent noise in the system. Furthermore, the β-barrel of the wild-type α-HL (WT-α-HL) is ~5 nm long,[7] and as a result, 10-15 bases may simultaneously contribute to the blockage current level.[12]

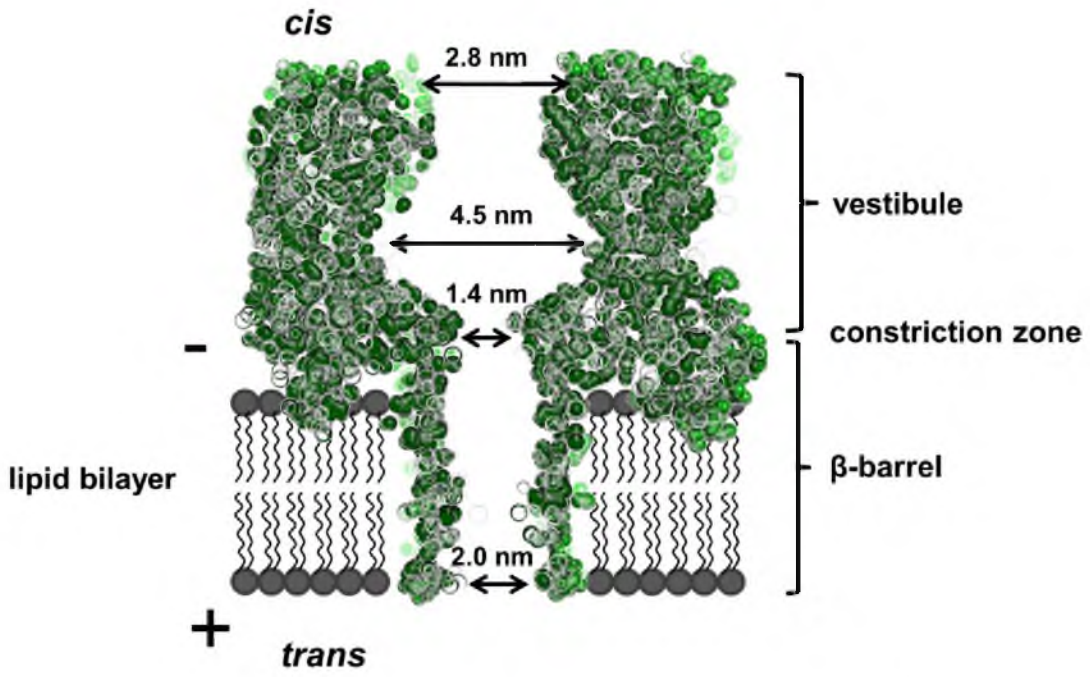


Figure 1.1. Structure of the heptameric bacterial toxin α-HL(pdb 7AHL)[7] embedded in a lipid bilayer. It consists of a vestibule on the *cis* side and a β-barrel on the *trans* side that are separated by a ~1.4 nm constriction zone. In most experiments, the negative electrode is on the *cis* side and the positive electrode is on the *trans* side, allowing DNA to travel from the *cis* to the *trans* compartment.

Consequently, one of the major challenges of nanopore sequencing is to decrease the translocation speed to a level appropriate for current electronics. Reported methods for slowing translocation have focused on either modifying the analysis conditions or utilizing mechanical means of slowing the movement of a DNA strand. The analysis conditions can be physically modified by decreasing the temperature,[9] decreasing the pH,[13] increasing the viscosity with glycerol,[14,15] or using organic salts;[16] however, these methods have not yet achieved the goal of slowing the DNA translocation to a rate that single-nucleotide resolution can be achieved. Other methods have capitalized on using chemical tags that sterically occlude the β-barrel during translocation,[17-19] thus slowing the movement of DNA. Short complementary probes to the single strand have also been used to slow down the translocation through duplex unzipping. This method has been utilized by our laboratory[20,21] and others.[22-28] Bayley's laboratory has shown through incorporation of a β-cyclodextrin adaptor to WT-α-HL that the identity of individual nucleotide-monophosphates can be resolved.[29-32] This method is proposed to allow DNA sequencing when an exonuclease is covalently attached at the vestibule mouth, such that it releases the nucleotides into the vestibule, thus allowing DNA sequencing to be achieved. Molecular motors (e.g., polymerases)[33-36] have recently been shown to slow the DNA translocation rate to a level suitable for resolving the sequence of nucleotides by the Akeson laboratory using α-HL[37] and by the Gundlach laboratory with the MspA[36] ion channel. Further, the Kasianowicz and Ju laboratories have utilized α-HL to conduct a sequencing-by-synthesis protocol, in which a pendant polymerase at the vestibule mouth releases a current-modulating pyrophosphate into the channel as the DNA strand is elongated.[38] Lastly, mutant pores with incorporated molecular breaks (positive charges)

have been shown to slow the translocation.[39,40] These methods all show great potential for solving the nanopore DNA sequencing challenge, but for any of these methods to be successful, the current signatures for the individual nucleotides must be established. Currently, the best method for establishing the current signatures has been achieved by appending a biotin on the end of a DNA strand, and then adding streptavidin that acts as a stopper to suspend the DNA strand in the protein channel. Immobilization of a DNA strand in the pore allows a relatively long time in which to record the current signature for the nucleotide of interest.[41-44]

Henrikson, *et al.* first demonstrated a way to immobilize the DNA in the α-HL using biotinylated DNA (Btn-DNA) and streptavidin.[41] Streptavidin is too big to enter the vestibule of α-HL, which immobilizes the DNA inside of the β-barrel of the protein, causing a deep level current blockage. Figure 1.2 shows the typical current-time (*i-t*) trace for an immobilization event. Such an event initiates with an open channel current (I_o, Figure 1.2). Next, capture of the DNA-protein complex brings the current to a deep blockage level (I, Figure 1.2), and after recording the current for 1 s, the polarity is reversed to release the molecule bringing the current back to the open channel value, and the system is ready to capture another DNA strand (Figure 1.2). The capture/release cycle is typically repeated ~200 times to collect a population of events, and the percentage residual current ($\%I/I_o$) of each event is calculated to plot the current level histogram.

The current level of DNA/protein complex depends not only on the DNA sequence but also on whether the 3` end or the 5` end of the ssDNA enters the ion channel first.[45,46] It was shown that 5` entry gives better discrimination between the

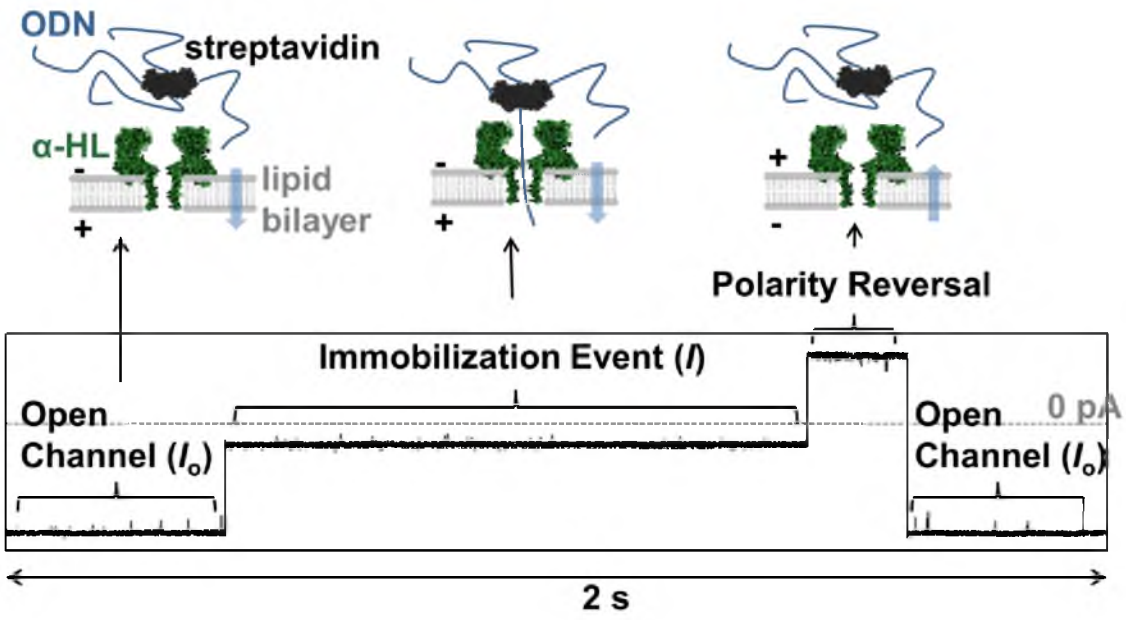


Figure 1.2. Schematic representation of the immobilization experiment. Top of the figure shows 3` Biotinylated DNA that is bound to streptavidin (pdb 1MK5).[47] At first only an open channel current (I_o) is recorded. Once the DNA is captured it causes a deep level current blockage (I) that lasts until the polarity is reversed (1 s), at which point the DNA is released, and the open channel current is restored.

standard nucleotides;[46] therefore, for the immobilization experiments described the biotin was attached on the 3` end to force the preferred 5` entry. The specific interactions that determine the blockage current level of the DNA inside the α-HL are not fully understood, but MD simulations have shown that they involve base stretching and tilting.[45,48] In the confined space of the β-barrel the ssDNA straightens and the nucleotides assume an asymmetric conformation by tilting toward the 5` end; consequently, 3` tethered DNA (5` entry) experiences a larger effective friction that gives better current discrimination. The regions of the β-barrel that are responsible for establishing the current level differences have been extensively studied. Ashkenasy, *et al.* showed that it is possible to distinguish a single adenosine (A) in a polycytidine (poly-dC) background (position 20 counting from the duplex) using a fishhook hairpin to immobilize the DNA inside the pore.[49] The pore was then more precisely mapped using streptavidin to immobilize a poly-dC 40mer with a single A substituted site specifically to reveal three sensing zones.[50] In this body of work, we have elected to be consistent with the nanopore literature, numbering all nucleotides from the 3` end; however, to avoid confusion with the traditional numbering of DNA sequences (i.e., from the 5` end) we have prefixed the label with an "ω" in front of the position number to differentiate these two numbering schemes. Figure 1.3 shows the regions of WT-α-HL corresponding to each sensing zone, along with the positions on the DNA strand. Surprisingly, the constriction zone (approximately position ω8), which is the narrowest portion of the pore, does not give the largest current resolution between C and A. The most resolving region in the protein is around position ω14 which falls approximately in the middle of the β-barrel (Figure 1.3).[50] Because this position gives the biggest current difference between

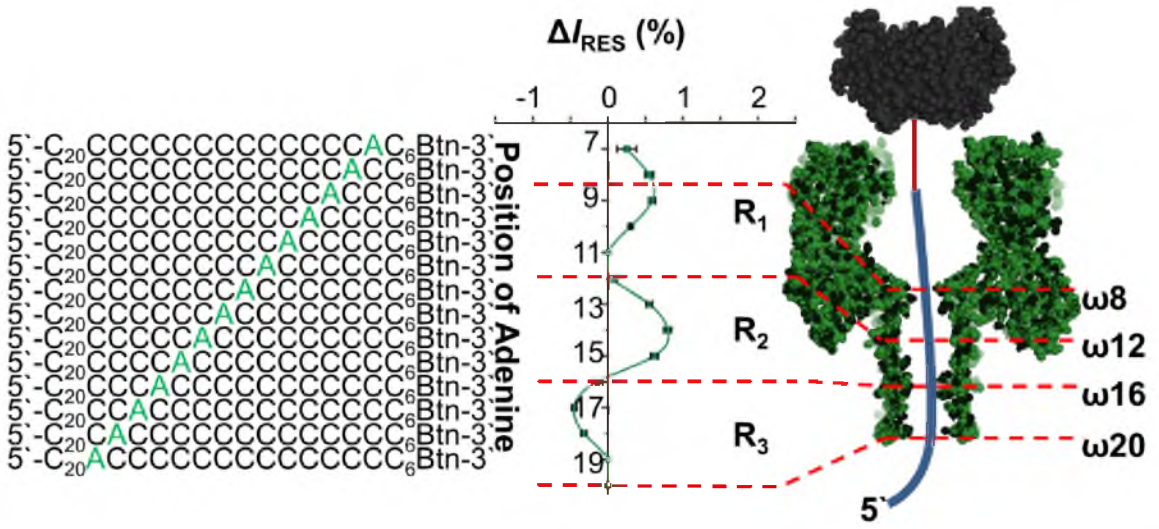


Figure 1.3. Sensing zone mapping experiments for WT-α-HL. Left: 3`-Biotinylated poly-dC DNAs with a single A nucleotide at position ω7 to ω19. Middle: residual current level differences ($\Delta I_{RES}(\%)$) between the DNAs. Right: representation of α-HL (pdb 7AHL)[7]/streptavidin (pdb 1MK5)[47] Btn-DNA complex with the approximate locations of the single A base in the protein channel. R_1, R_2, R_3 are the observed sensing zones along the β-barrel of the protein. The figure was adapted from Stoddard, *et al.*[50]

the native nucleotides it was chosen for the studies discussed below. In subsequent mapping studies conducted by Purnell, *et al.* in a polythymidine background, the four nucleotides gave different sensing regions.[51] The most sensitive region was found to be near the pore constriction zone. These results highlight the overall sequence-dependency in establishing a unique current level at one position in the ion channel. Utilizing the immobilization experiment the current levels for the native DNA nucleotides can be established in all possible sequence contexts that will be required for identifying the nucleotides in a nanopore sequencing experiment. However, the human genome is known to have many more nucleotides than A, T, G or C, and establishing the current levels for these "modified" nucleotides will be necessary for any single molecule sequencing application to avoid errors at these sites.

DNA is susceptible to damage by a variety of exogenous and endogenous agents that can introduce a wide range of modifications to the genome. Additionally, an assortment of DNA alkylating agents can be found in the air,[52] water, food and a variety of chemotherapeutic drugs.[53] Aside from the exogenous agents mentioned above, DNA damage arises continuously within living cells due to metabolic processes, the inflammatory response, mitochondrial respiration and other biological reactions.[54] Due to these processes, reactive oxygen species (ROS) can induce DNA damage leading to oxidized bases, abasic sites, cross-links, bulky adducts, single- and double-strand breaks.[55] Our laboratory has a long-standing history in studying oxidatively-derived DNA damage; particularly with respect to oxidation of guanosine (G).[56-59] G has the lowest redox potential and is the site most prone to oxidative insults.[60] The two-electron oxidation of G yields the most common damaged base, 8-oxo-7,8-dihydroguanosine

(OG), which ~6400 are thought to exist in the genome (Table 1.1);[61] also, OG is mildly mutagenic causing G to T transversion mutations.[55] Moreover, OG is the biomarker followed for assessing oxidative stress to the cell.[61] OG is prone to a second two-electron oxidation that yields a pair of hydantoin compounds, spiroiminodihydantoin (Sp) and guanidinohydantoin (Gh) that both exist as a pair of diastereomers (Figure 1.4A).[56,57] The yield of these two molecules is dependent on the context in which OG is oxidized;[58] furthermore, these molecules are highly inhibitory to strand elongation by polymerases,[62] and *in vivo* studies show them to be highly mutagenic causing G to T and G to C transversion mutations.[63] Recent studies have observed these molecules in mouse models of chronic inflammation, in which they are present at levels 100 times below that of OG (Table 1.1).[64] Ionizing radiation is another exogenous agent that produces an assortment of DNA damages including double- and single-strand breaks, abasic sites (AP) and base lesions.[65] Ionizing radiation gives high levels of damage at T nucleotides that yields thymine glycol (Tg). Tg is estimated to be formed 400 times per day in a cell (Table 1.1), and in animals Tg has been used as a marker for oxidative stress (Figure 1.4, B).[66] Furthermore, Tg is highly mutagenic due to its ability to stall DNA polymerases which leads to failed elongation of the DNA strand.[67]

Another form of DNA damage results from UV-induced photochemical reactions forming mutagenic cyclobutane-pyrimidine dimers (CPDs), 6-4 photoproducts and their Dewar valence isomers, and these products are typically observed at adjacent thymidine (T) nucleotides to yield a thymine dimer (T=T, Figure 1.4, D).[68,69] The T=T yield is highest in skin cells exposed to UV light, for which this form of DNA damage has been strongly correlated with skin cancer[70] that results from the fact that T=T lesions stall

Table 1.1. Frequency of occurrence for the damaged DNA bases found *in vivo*.

DNA modification	Approximate number in human genome
UV photoproducts	100,000[72]
Tg	400 [67]
Spontaneous depurination (AP sites)	10,000 [80]
OG	6,400[61]
Sp	60[a64]
Gh	60[a64]
U	500[79]

[a] Values were calculated from data determined in a mouse model

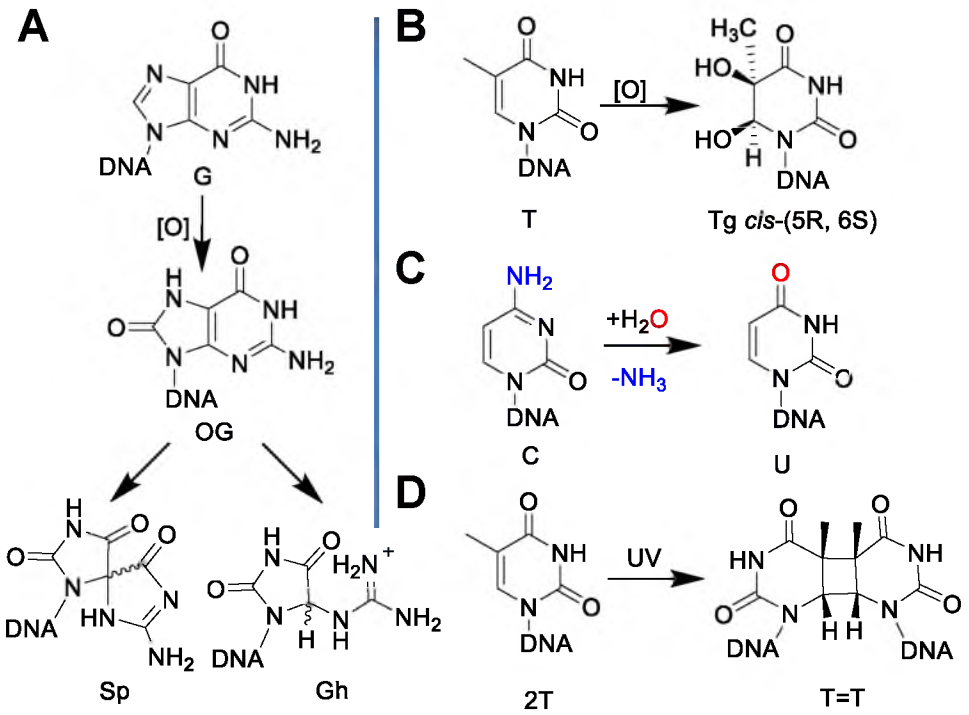


Figure 1.4. DNA damage reaction pathways. **A.** Oxidation of G to OG then to Sp or Gh. **B.** T oxidation pathway to Tg. **C.** Deamination of C to U. **D.** Irradiation of 2T sites to T=T.

DNA polymerases.[71] A single day spent in the sun can introduce up to 100,000 UV photoproducts per cell in the epidermis (Table 1.1).[72]

In addition to the exogenous and endogenous agents that cause DNA-base modifications, DNA itself is also inherently reactive, and these reactions contribute to genomic modifications that have been observed *in vivo*. Spontaneous hydrolysis of the glycosidic bond results in the formation of abasic sites (AP) that are observed at the purine nucleotides.[73] The spontaneous base loss is thought to occur 10,000 times per cell per day (Table 1.1).[74] AP sites are devoid of genetic information that causes them to be highly stalling to most DNA polymerases.[75-77] Considering all the sources of the AP sites, it is one of the most frequently occurring DNA damages; furthermore, the exocyclic amino groups found on the heterocyclic rings of the DNA bases are prone to deamination reactions under biological conditions. Cytidine is the base most prone to deamination ($t_{1/2}$ ~ 19 d)[78] yielding uridine (U, Figure 1.4C), which is similar to T in its hydrogen-bonding properties.[79] The fifth DNA base, 5-methylcytidine (5-mC), is also prone to deamination ($t_{1/2}$ ~ 9 d)[78] yielding thymidine (T). If the resulting products U or T are not properly repaired, C to T transition mutations are observed.[74] The deamination of C has been estimated to occur in 100-500 nucleotides per cell per day (Table 1.1).[79] Although the overall percentage of damaged DNA bases is small (Table 1.1) compared to the size of the genome, nanopore sequencing of unamplified DNA will encounter these damaged nucleotides. Therefore, it is crucial to establish the current signatures for the common forms of DNA damage that will be observed in any nanopore sequencing method. This information will be most beneficial for increasing the accuracy of making base calls, especially if the damaged bases give current levels similar to the native DNA bases. In

addition, the ability to determine the precise location and frequency of base damage from minute tissue samples would be a tremendous boon to understanding the occurrence of DNA damage as a function of diet, exposure to environmental toxins, drug metabolism, etc.

The remainder of this chapter focuses on studies that have characterized the electrical current signatures of the common DNA base damages via immobilization experiments with the damage placed at ω14 in either a polycytidine (40mer) or biologically relevant sequence context. Because some forms of damage overlapped in current levels with the native nucleotides, the inherent reactivity of some of these nucleotide damages was capitalized upon to add chemical tags that increase the current contrast between the modified and native nucleotides.

Canonical nucleotide ion current levels

In the first set of studies performed by Schibel, *et al.* (White research group),[81] the ion current level histograms for the four canonical nucleotides at position ω14 were measured in a poly-dC background sequence (C_{40}). Our studies utilized the glass nanopore membrane as the lipid bilayer support, developed in the White laboratory at the University of Utah.[82,83] The poly-dC 40mer was selected to be consistent with literature reports;[50] as well as being the representative strand for the C nucleotide current level, this sequence was also utilized as an internal standard to plot all current histograms as $\Delta\%I/I_O$. These measurements established current levels for the native nucleotides that allowed the determination of whether the damaged nucleotides gave current signatures similar in value to A, T, G or C. Figure 1.5 displays the current level histograms for the canonical

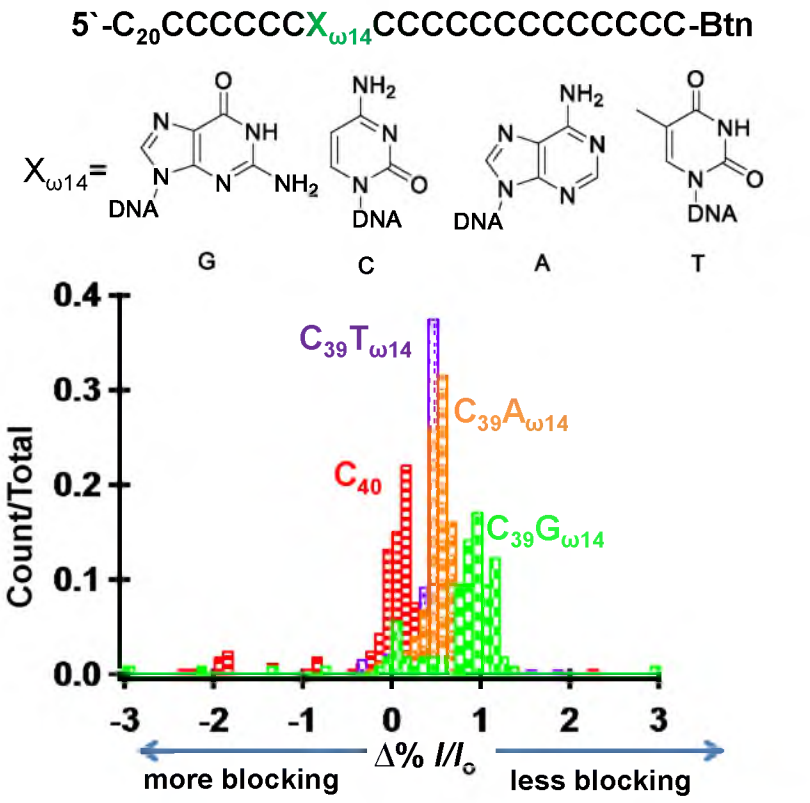


Figure 1.5. $\Delta\%I/I_o$ histograms representing the current level blockage of G, C, A, and T. The base of interest was placed at position ω14 in a C_{40} DNA that was biotinylated at the 3` end. The experiments were conducted in 1 M KCl, 25 mM Tris-HCl, and 1.0 mM EDTA (pH 7.9). The structures for the four canonical nucleobases are shown at top. The figure was adapted from Schibel, *et al.*[81]

DNA nucleotides as observed in an immobilization experiment previously described in Figure 1.3.[81] These distributions of currents are similar to those reported by Stoddart, *et al.*[50] In this sequence context, the pyrimidine nucleotides block the pore more than the purine nucleotides, which was surprising considering that the pyrimidines are smaller in size than the purines. This suggests that there are additional factors that determine the current level differences beyond the size of the nucleotide. The order of most blocking to least blocking was C > T > A > G, which, interestingly, follows the same trend as their water solubility: 65.8, 27.8, 19.8, and 1.8 mM (unbuffered water at 25 ^{o}C).[84]

Nucleotide oxidation product current levels

In the next set of studies performed by Schibel, *et al.* (White laboratory),[81] current level histograms were recorded for the G-oxidation products OG, Sp and Gh at position ω14 in a poly-dC oligomer (Figure 1.6A). For this data, the current histograms for G and OG display significantly overlapping signals. The hydantoins Sp and Gh gave significantly broader current level distributions than G and OG. Furthermore, their current histograms presented with multiple population subtypes that may be best explained by the fact these nucleotides have a stereocenter in the heterocycle that exists as two diastereomers. Furthermore, Sp and Gh provide an interesting example of how nonplanar nucleotides impede the current in an immobilization nanopore experiment. The intriguing observation of different current levels for the diastereomers of Sp and Gh suggests that the current level is also established by the shape of the nucleotide, or a shape effect on the local structure of the DNA strand. This observation is consistent with measurements made by Bayley's laboratory, in which they could distinguish the

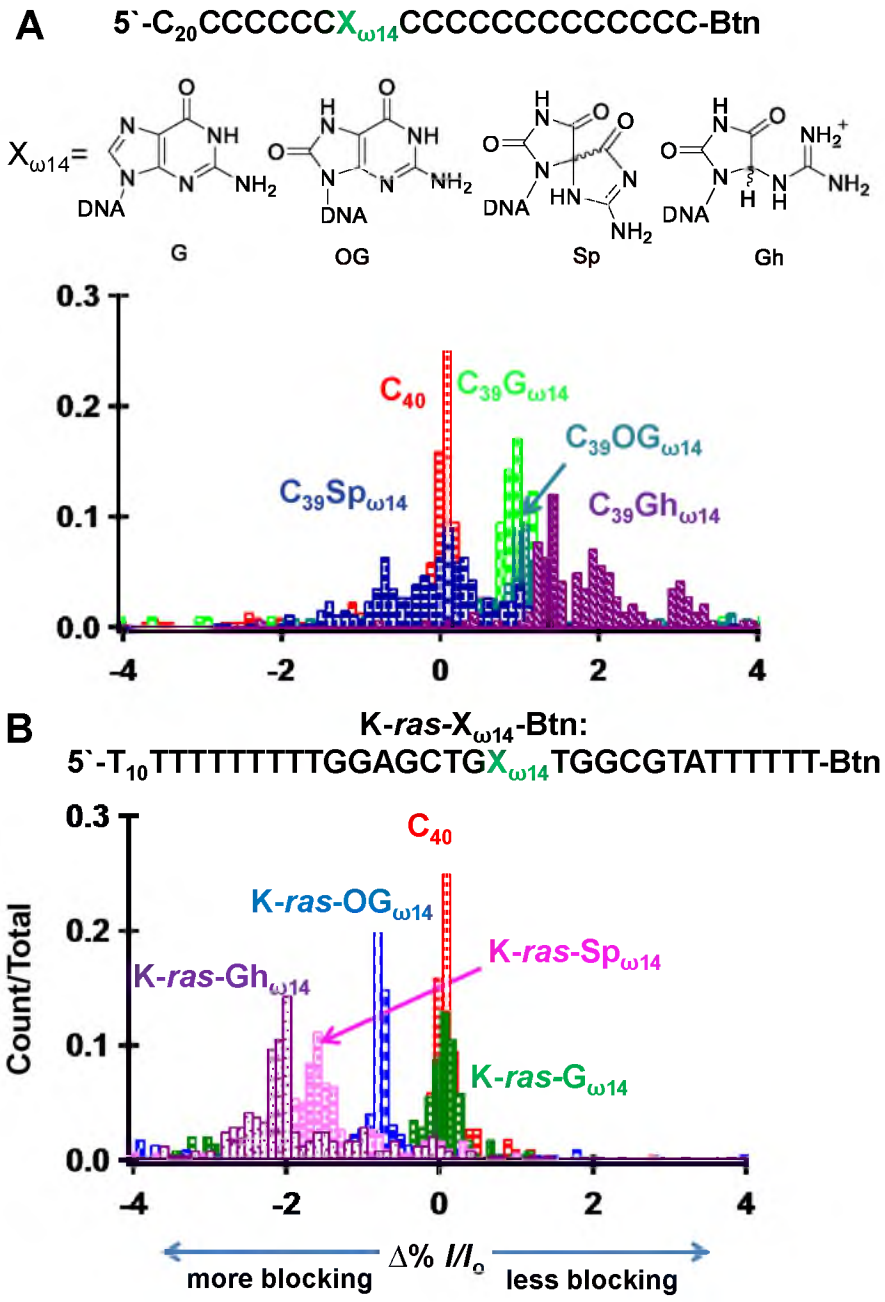


Figure 1.6. Relative percent residual current histograms for G-oxidation products. **A.** Top: structures of G, OG, Sp, Gh. Bottom: $\Delta\%I/I_o$ histograms, representing the current level blockages for G, Gh, OG, Sp at a position ω14 (biotinylated at the 3` end) in a poly-dC DNA, in which C_{40} was the internal standard. **B.** $\Delta\%I/I_o$ histograms, representing the current level blockage of G, Gh, OG, Sp at position ω14 in the K-*ras* sequence in a poly-dT background. The internal standard was C_{40} that was set to 0. The experiments were conducted in 1 M KCl, 25 mM Tris-HCl, and 1.0 mM EDTA (pH 7.9). The figure was adapted from Schibel, *et al.*[81]

enantiomers of ibuprofen and thalidomide in a β-cyclodextrin adapted α-HL nanopore.[85] This again suggests that molecular weight alone is not the sole explanation for the observed differences in current levels. From these data, the current levels for G and OG show significant overlap and would be hard to distinguish in a nanopore sequencing experiment (Figure 1.6A). In addition, one diastereomer of Sp would give a current level similar to C, while the other could potentially be identified (Figure 1.6A). In other words, an oxidation product of G might, ~50% of the time, be incorrectly interpreted as a G to C single nucleotide polymorphism (SNP) in the sequence. In contrast, the Gh diastereomers led to current signatures that differ from the background sequence and could possibly be correctly identified (Figure 1.6A). It must be noted that this only holds true in sequence contexts that give current levels similar to poly-dC in WT-α-HL.

The current histograms change dramatically when a different background sequence was used for the measurements. The heterosequence chosen for study was a part of the K-*ras* protooncogene surrounding codon 12 (Figure 1.6B) for which point mutations have been shown to cause uncontrolled cell growth and loss of cell differentiation triggering various human adenocarcinomas.[86] In lung cancer, the K-*ras* gene often undergoes a G to T transversion mutation that might result from unrepaired G oxidative damage.[87] Therefore, the critical G was positioned at ω14 (Figure 1.6B) to determine the current levels of the canonical nucleotide vs. its oxidatively-damaged forms. The histogram in Figure 1.6B shows that OG, Sp and Gh can be distinguished from G in this sequence.[81] Interestingly, although Gh and Sp gave broader distributions, the diastereomers of the hydantoins were not as well resolved as was observed in the poly-dC context. This further supports the idea that sequence context is crucial for

determining the current levels.[81] This is a major limitation associated with the WT-α-HL nanopore; because of the three sensing zones, different sequence contexts can lead to overlapping current distributions.[50] Consequently, the investigation of protein ion channel containing mutated sites that optimize the current difference between nucleotides is an important area of investigation.[12]

Another type of oxidative damage caused by ionizing radiation is thymine glycol, Tg (Figure 1.7). Tg occurs as a pair of diastereomers and the most common one formed *in vivo* is *cis*-(5R, 6S) in a 83% yield.[67] Tg, unlike the native nucleotides, is also non-planar.[67] Figure 1.7 shows the structure and the current level blockages for Tg at either position ω13, ω14, ω15, or ω16 within a 40mer poly-dC DNA strand (this set of studies was conducted by Lidong He, Burrows laboratory).[88] The current level of Tg falls within the standard nucleotide current levels and would likely be incorrectly assigned in a sequencing experiment with WT-α-HL, although Tg can be correctly distinguished from a T nucleotide at any of the four positions studied (Figure 1.7B). In this experiment, the biggest difference between the modified and unmodified DNA was observed at position ω15. Furthermore, when Tg was placed at ω15, two populations of currents were observed that had a ratio of approximately 85:15. Previously, it has been shown that solid-phase synthesis of Tg gives the *cis*-(5R, 6S) and *trans*-(5R, 6R) in a ratio of 87:13,[67] respectively, which nicely correlates to the observed subpopulations observed in our experiment. Furthermore, this agreement between prediction and observation of two current levels for Tg (ω15, inset Figure 1.7) adds further support for our hypothesis that the nanopore can distinguish the shape, or its effect on local DNA structure, for diastereomers of damaged nucleotides including Sp, Gh and now Tg (Figure 1.6).

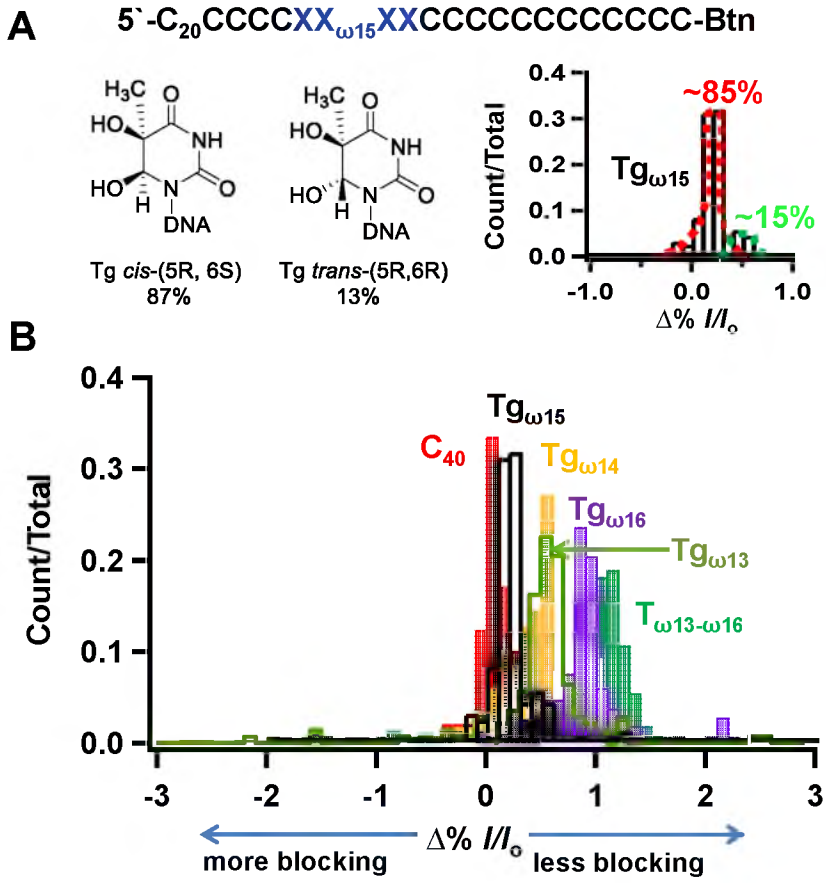


Figure 1.7. Current level histograms for Tg-containing DNA. **A.** Two diastereomers of Tg that were present in the DNA strand studied. $\Delta\%I/I_o$ histogram of Tg at position ω15 showing two current populations. **B.** $\Delta\%I/I_o$ histograms, representing the blockage current level of thymine glycol (Tg) at position ω13 through ω16. The Tg phosphoramidite was introduced at positions ω13 to ω16 in a DNA that was poly-dC except of T at those positions labeled X (sequence above). $C_{36}T_{\omega 13\text{-}\omega 16}$ was the standard that did not contain Tg. The experiments were conducted in 1 M KCl, 25 mM KP_i and 1 mM EDTA (pH 7.4).[88]

Current levels for hydrolytic DNA damage

The structures and current level histograms for the AP and U residues are presented in Figure 1.8, work conducted by An, *et al.* (Burrows laboratory).[89] Taking into account how much smaller the AP site is compared to the native nucleotide, it is surprising to see that the current level blockage of this damage is the same as for G (Figure 1.5), which suggests that the presence of an AP site may have an impact on the surrounding DNA secondary structure and its interaction with the α-HL. Given the abundance of AP sites in the human genome, unmodified AP sites could produce a false readout of G base under these circumstances. The current level of U falls in the same place as that of T, and this is an indication that the methyl group does not have a distinguishable effect on the interaction between the nucleotide and protein near the ω14 position of the DNA. Unlike the previous studies[81], incorporation of U and AP in a heterosequence (K-*ras*) leads to a similar pattern as in the poly-dC sequence context.[89]

From the data presented above, the current levels for known forms of DNA damage that will be found in the genome were established. It is anticipated that these lesions will be found during any sequencing application that does not preamplify the sample (i.e., nanopore sequencing), and their current levels fall into two types: (1) Those that give unique current level histograms, and (2) those that give current level histograms similar to the native DNA nucleotides. Moreover, each modified base might fall into both classes depending on sequence context, as was shown for the G-oxidation products between a poly-dC background and the K-*ras* sequence (Figure 1.6A and 1.6B). One approach to deal with this current level redundancy is to engineer an α-HL pore that reduces the sensing regions, as well as gives higher contrast between the nucleotides.

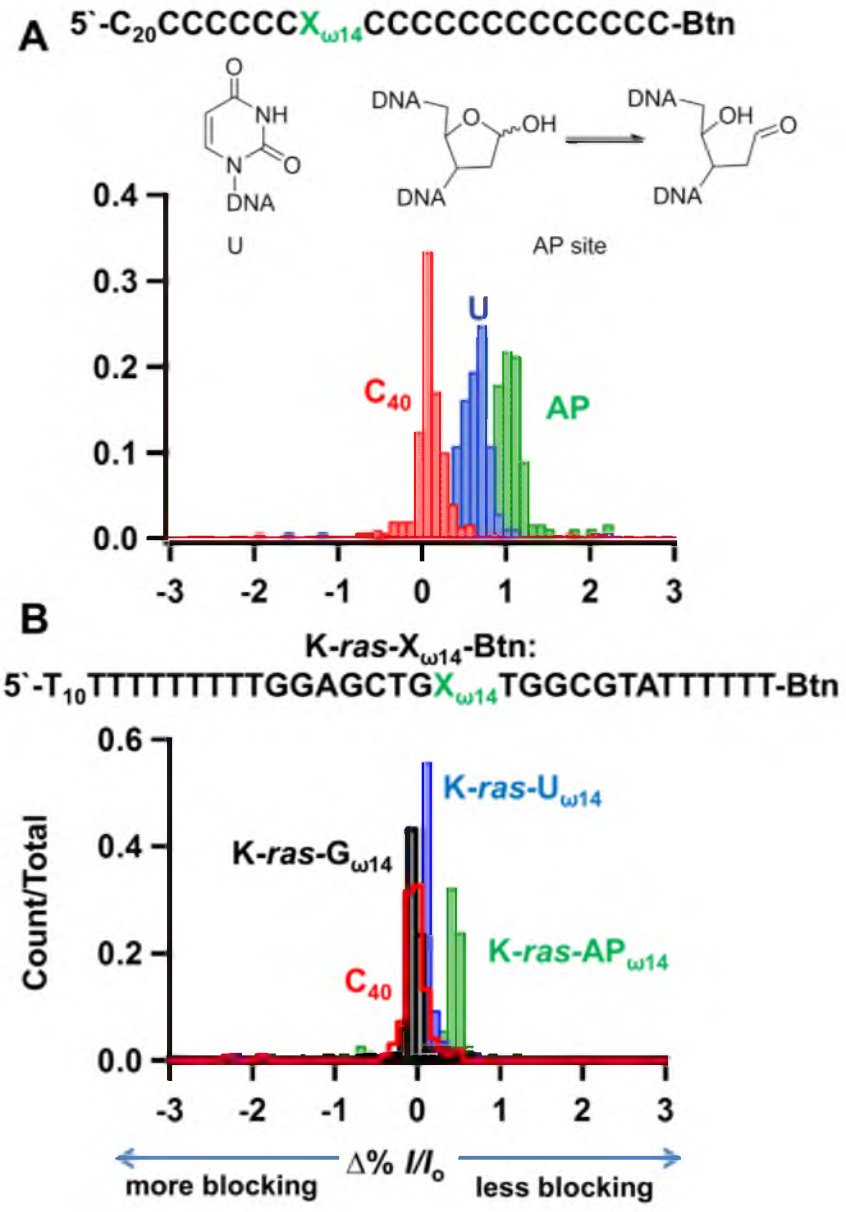


Figure 1.8. $\Delta\%I/I_o$ histograms representing the current level blockage of U and AP. The modified bases were placed at the position ω14 where C_{40} was the internal standard. **A.** In a homopolymer sequence. **B.** In a heteropolymer sequence. The experiments were conducted in 1 M KCl, 25 mM Tris-HCl, and 1.0 mM EDTA (pH 7.9).[89]

This is the method adopted by Bayley and colleagues.[12,37] Herein, we have elected another method to deal with the tight distributions of currents between the native and damaged DNA nucleotides. Damage to DNA can give new chemical functional groups that have orthogonal reactivities compared to native DNA; therefore, we have harnessed the reactivity of these new functional groups to add chemical tags that shift the currents of damaged DNA nucleotides outside the range of the native nucleotides.

DNA adduct formation to tag damage sites

While the use of immobilization experiments demonstrates that native DNA bases as well as damaged bases can be distinguished, the current level difference between them is too small to be observed during a translocation experiment, because the rapid transit of the bases through the sensing zone does not permit enough signal averaging. As a consequence, we turned to the formation of DNA adducts via chemical modification because that approach capitalizes on the unique chemistry of the various heterocycles or AP sites. Ideally the current signature introduced by the chemical modification should be used to label the nucleotide of interest in a fashion that makes sequencing by nanopore ion channel current recordings possible. In order to determine if a chemical tag increased the current level difference, immobilization experiments were first conducted. The Burrows and White laboratories (Schibel, *et al.*)[81] first introduced chemical modifications to OG, because OG has a lower redox potential than the other nucleotides that allows specific chemical modification under mildly oxidizing conditions in the presence of virtually any primary amine.[90-92] Amines provide a versatile chemical tagging system because there are so many commercially available options that are water soluble, which is

a necessity when conducting reactions on DNA. Oxidation of an OG-containing DNA strand (OG was positioned at ω14) in the presence of a primary amine gives a spirocyclic product with a pendent amine (Figure 1.9).[81] In these studies, the primary amines benzylamine (Bz, MW = 108), N^{α}-acetyl-*O*-methyl-lysine (Lys, MW = 202), glucosamine (GlcN, MW = 179), spermidine (Spd, MW = 147), spermine (Spm, MW = 205), and the tetrapeptide Gly-Pro-Arg-Pro carboxamide (GPRP, MW = 397) were selected to give a range of sizes and different charge states for study (Figure 1.9). It is apparent that chemical modifications have a much bigger influence over the current level blockage than the DNA damage to OG. The Sp adducts have more than one current population, or very broad current distributions, that are best explained by the two diastereomers of the Sp adducts. In an attempt to understand the current level distributions for adducts there is, to a first approximation, a correlation between size and the current level (i.e., bigger adducts are more blocking). However, closer inspection of the data highlights some nuances to this trend. For example, the Spm adduct is much more blocking than the GlcN adduct, even though GlcN has a larger molecular weight and is clearly bulkier in cross-section. There are two key differences between these adducts: (1) Spm is a polyamine that has an overall net positive charge, and (2) Spm is linear in shape while GlcN is not. Therefore, adduct charge and shape likely provide further refinements to consider when predicting the current blocking level of an adduct. In a subsequent set of experiments, all of these adducts were translocated through α-HL (i.e., allowed to pass from the *cis* side to *trans* side of the pore without streptavidin), and none of them gave a current signature that could be distinguished from the current level of the background sequence due to the fast translocation speed.

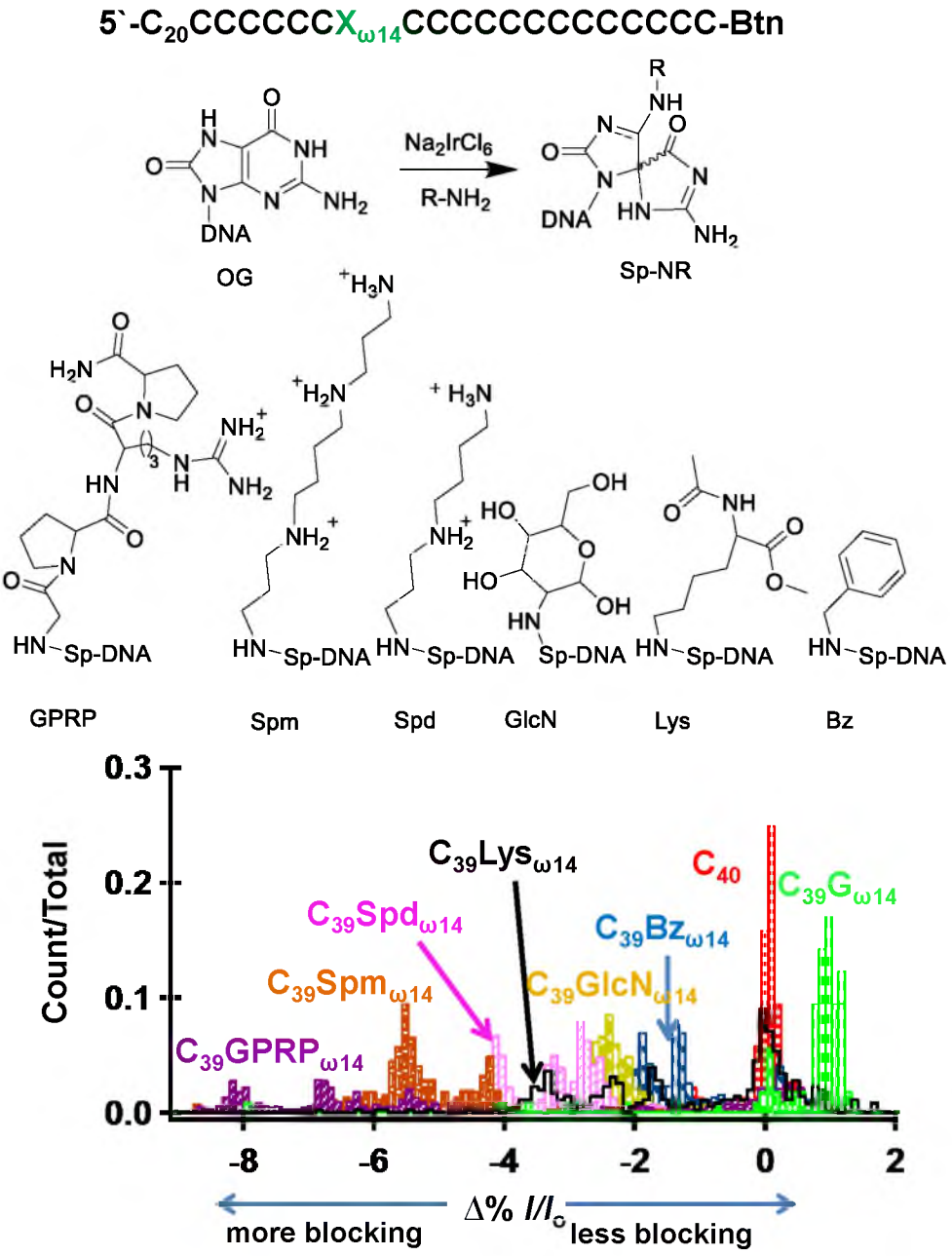


Figure 1.9. Spirocyclic adducts to OG. Top: the structures of the spirocyclic adducts: Bz, Lys, GlcN, Spd, Spm, and GPRP. Bottom: $\Delta\%I/I_o$ histograms showing all the previously mentioned adducts in a poly-dC background at position ω14 where the internal standard was C_{40}. The experiments were conducted in 1 M KCl, 25 mM Tris-HCl, and 1.0 mM EDTA (pH 7.9). The figure was adapted from Schibel, *et al.* [81]

Apurinic/apyrimidinic (AP) sites in DNA present a functional group that shows reactivity orthogonal to native DNA. Hydrolysis of the glycosidic bond yields a hemiacetyl that is in equilibrium with the ring opened aldehyde that can be functionalized by a primary amine under reductive conditions (Figure 1.10).[89] Seven primary amines were tested as adducts to AP in the work done by An, *et al.*[89]: taurine (Trn), glucosamine (GlcN), Arg-His carboxamide (RH), Gly-Pro-Arg-Pro carboxamide (GPRP), streptomycin (STM), 2-aminomethyl-15-crown-5 (15c5) and 2-aminomethyl-18-crown-6 (18c6). The modifications to AP sites gave predominantly one distribution of current levels unlike the modifications to OG, which gave more than one distribution (with the exception of $C_{39}GlcN_{\omega 14}$). The tighter distribution of current levels for the AP sites adducts compared to the OG adducts may be a result of there not being an additional stereocenter in the newly formed adduct. Interestingly, the GlcN adduct gives the same current level distribution regardless of whether it is attached to the AP site (Figure 1.10) or the Sp site (Figure 1.9). This again hints that the size of the molecule is not necessarily a determining factor of the current level blockage, but more of the underlying chemical nature of the adduct that is establishing the blockage current level. Imbedding the adduct in the K-*ras* sequence yielded similar results.[89] All of the adducts presented in Figure 1.10 can translocate through the nanopore, and again, disappointingly, none of them gave a distinct modulation in the current level except for the 18c6 adduct in NaCl solution. Translocation of an 18c6 adduct embedded in the middle of an 87mer poly-dC DNA strand led to a modulation of the current when passing through the pore (Figure 1.11).[18] The 18c6 adduct is proposed to give this distinct current level blockage because it has to release the bound alkali metal cation before passing through the restrictive β-

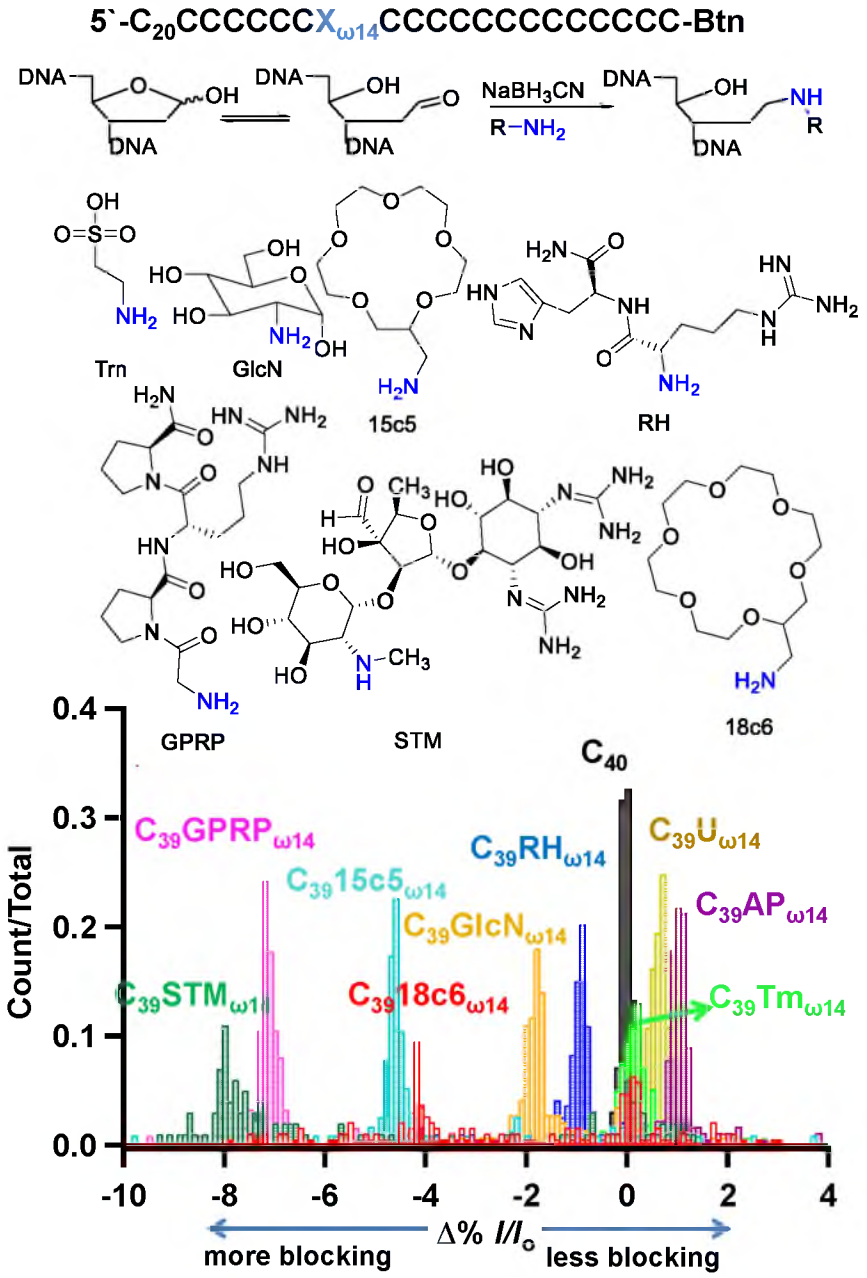


Figure 1.10. DNA abasic site adducts. Top: the structures of the AP site adducts: Trn, GlcN, RH, GPRP, STM, 15c5 and 18c6. Bottom: $\Delta\%I/I_o$ histograms showing all the previously mentioned adducts in a poly-dC background at position $\omega 14$, in which the internal standard was C_{40}. The experiments were conducted in 1 M KCl, 25 mM Tris-HCl, and 1.0 mM EDTA (pH 7.9). The figure was adapted from An, *et al.*[89]

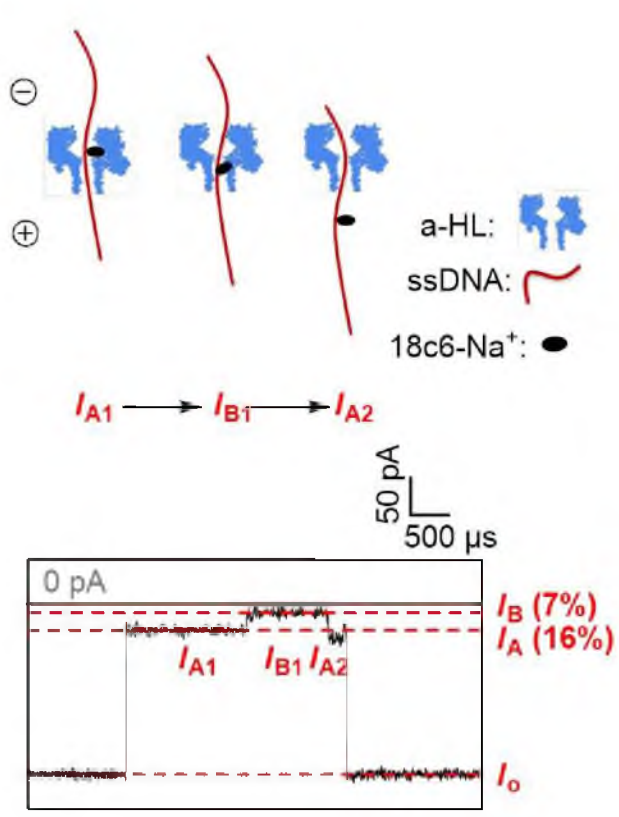


Figure 1.11. Translocation current vs. time trace for the AP 18c6 adduct showing modulation of the current in 3 M NaCl, 25 mM Tris-HCl, and 1.0 mM EDTA (pH 7.9). There were two current levels observed: I_A and I_B. When ssDNA entered the vestibule, 18c6-Na^+ complex hesitated above the constriction, while the sensing β-barrel was recognizing the poly-dC background of the molecule, generating I_{A1}. After a certain period of time, the complex changed its conformation, likely followed by the dissociation of the ion, and was able to pass the narrow constriction. This produced a deeper blocking current level I_B due to the presence of the bulky adduct. Once the adduct exited the protein β-barrel to its *trans* side, the current level returned back to I_{A2}, representing the translocation of the rest of the ssDNA. The figure was adapted from An, *et al.*[18]

barrel. Given the fact that AP sites serve as an intermediate in the base excision repair pathway and that various types of DNA base modifications can be converted to AP sites by the corresponding glycosylases,[93,94] one can envision this approach being readily applied to detect other DNA damages as well, including DNA methylation sites important in epigenetic sequencing.

As a final test of formation of a DNA base adduct that might significantly alter the current level, halogenation of C followed by adduct formation was conducted. Iodination of C in the presence of $KHSO_5$ and KI readily occurs at the C5 position of cytidine, and this activates the base toward nucleophilic aromatic substitution by 2-aminomethyl-18-crown-6.[95] Figure 1.12 provides the current level histograms for 5-iodocytidine (5-I-C) as well as the 18c6 adduct to the base (5-18c6-C). Interestingly, 5-I-C gave the same current level as the C_{40} standard; however, the 5-18c6-C adduct was observed to be much more blocking. The translocation of the 5-18c6-C adduct also produced current signatures similar to the one in Figure 1.12. The two examples of 18c6 adducts, where 18c6 was attached to either AP sites or C, confirm that the polyether moiety is responsible for the desired current modulation, and now point to its use as a universal label for damage detection and sequencing efforts.

Alternative approach for determination of current level differences for nonnative nucleotides

5-Methylcytidine is one of the most common modifications found in the genome that was biologically selected for epigenetic regulation. This modification resides on ~5% of all C nucleotides[96] and it is nonrandomly distributed throughout the genome.[64]

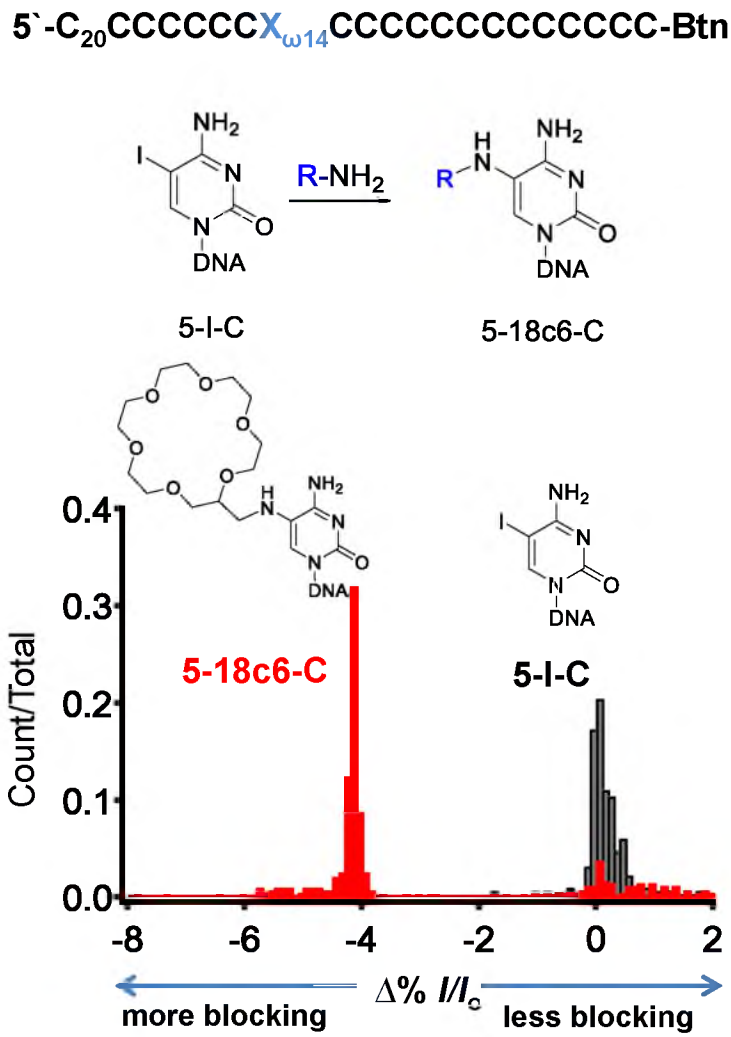


Figure 1.12. Current blockage levels for the 18c6 adduct (5-18c6-C) and 5-iodocytidine (5-I-C). Top: reaction sheme for the 5-18c6-C formation followed by the structures of the 5-18c6-C and 5-I-C. Bottom: $\Delta\%I/I_o$ histograms showing mentioned adducts in a poly-dC background at position $\omega 14$, in which the internal standard was C_{40} that was set to 0 $\%I/I_o$. The experiment was conducted in 1.00 M NaCl, 25 mM Tris, and 1 mM EDTA (pH 7.9).

Recent studies have shown that oxidation of 5-methylcytidine (5mC) to 5-hydroxymethylcytidine (5hC) may provide a further refinement in epigenetic regulation.[97] Thus, sequencing for these modifications has great potential in answering many biological questions, especially if this can be conducted on a single-molecule platform, such as the nanopore apparatus. Bayley's laboratory monitored the current levels for 5mC and 5hC in an immobilization experiment using an engineered pore (NNY mutant) that has one sensing region and gives high contrast between the nucleotides.[98] This example demonstrates that in the future, nanopore sequencing of cellular DNA may help in addressing many biologically relevant modification states of DNA.

The ion channel protein α-HL is only one protein among many others that has potential as a DNA-sequencing platform.[99-101] A mutated MspA ion channel protein has also been shown to distinguish current levels of the native nucleotides in immobilization experiments.[102] The current level differences were much bigger than those observed with α-HL pores. Similarly, Manrao, *et al.* demonstrated that a mutant MspA ion channel can readily distinguish the 5mC from C.[36] The bigger difference in the current levels observed for the nucleotides has been attributed to MspA having one sensing zone that spans fewer nucleotides than α-HL.[50,100]

The use of molecular motors has been gaining traction for decreasing the speed at which the DNA moves through the nanopore to a level that existing electronics can readily record. [33,37,103] The methods currently reported utilize a DNA polymerase (phi29) to ratchet the DNA up through the pore in the presence of an electrical bias, as each nucleotide passes through the one sensing head its current signature is recorded for milliseconds, which readily allows base to base discrimination.[36] The ability to have

tight control over the DNA polymerase has made this approach an attractive alternative for controlling the translocation rate. However, the DNA polymerase presents a few problems on its own, particularly when sequencing genomic DNA that has nucleotide modifications as discussed herein. DNA damage lesions such as Sp, Gh, AP sites, Tg and T=T cause DNA polymerases to stall; that is to say, DNA polymerases cannot read through these types of DNA damage and would cause a stop in reading of the sequence in molecular-motor based sequencing methods when conducted on a single-molecule level on unamplified DNA.

Summary and outlook

Immobilization experiments have allowed current signatures to be established for native, damaged and chemically modified nucleotides. When considering all the established current levels, an overlap is observed between some of these nucleotides; for example, U and T, and an AP site and G give the same blocking current levels. Chemical modification via adduct formation to OG or AP sites is a promising way to shift the current redundancy away from the current range of the other nucleotides. With the chemical adducts studied, there is some degree of correlation to the size of the adduct with its blocking current level; though, it is prudent to point out that this parameter alone is not the only determining factor for predicting the blocking current level. The current levels also depend on hydration, rigidity of the strand (TT vs. T=T, 5'-GCT vs. 5'-G*CT*), shape of the adduct (diastereomer resolution for Sp, Gh and Tg), as well as interaction of the adduct with the protein and electrolyte, as was observed with the 18c6 adducts. Further, the 18c6 adduct has the added feature of being observed in translocation

experiments, and it is the only adduct that can not only generate much deeper blockage current levels, but also significantly slow the translocation speed via the ion dissociation mechanism proposed. Both factors proved to be necessary to develop a successful chemical tag for the purpose of electrical current modulation.

References

(1) Wolna, A. H.; Fleming, A. M.; An, N.; He, L.; White, H. S.; Burrows, C. J. Electrical current signatures of DNA base modifications in single molecules immobilized in the α-hemolysin ion channel. *Isr. J. Chem.* **2013**, *53*, 417-430.

(2) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; 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.

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

(4) Rhee, M.; Burns, M. A. Nanopore sequencing technology: research trends and applications. *Trends in Biotechnology* **2006**, *24*, 580-586.

(5) 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.

(6) Bayley, H. Sequencing single molecules of DNA. *Curr. Opinion Chem. Biol.* **2006**, *10*, 628-637.

(7) Song, L.; Haobaugh, M. R.; Shustak C.; Cheley, S.; Bayley, H.; Gouaux, E. J. Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **1996**, *274*, 1859-1866.

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

(9) 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*, 1097-1084.

(10) Butler, T. Z.; Gundlach, J. H.; Troll, M. Ionic current blockades from DNA and RNA molecules in the alpha-hemolysin nanopore. *Biophys. J.* **2007**, *93*, 3229-3240.

(11) Meller, A.; Branton, D. Single molecule measurements of DNA transport through a nanopore. *Electrophoresis* **2002**, *23*, 2583-2591.

(12) 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.

(13) de Zoysa, R. S. S.; Krishantha, D. M. M.; Zhao, Q.; Gupta, J.; Guan, X. Translocation of single-stranded DNA through the α-hemolysin protein nanopore in acidic solutions. *Electrophoresis* **2011**, *32*, 3034-3041.

(14) Fologea, D.; Uplinger, J.; Thomas, B.; McNabb, D. S.; Li, J. Slowing DNA translocation in a solid-state nanopore. *Nano Lett.* **2005**, *5*, 1734-1737.

(15) 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* **2008**, *25*, 1233-1237.

(16) de Zoysa, R. S. S.; Jayawardhana, D. A.; Zhao, Q.; Wang, D.; Armstrong, D. W.; Guan, X. Slowing DNA translocation through nanopores using a solution containing organic salts. *J. Phys. Chem. B* **2009**, *113*, 13332-13336.

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

(18) 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.

(19) Borsenberger, V.; Mitchell, N.; Howorka, S. Chemically labeled nucleotides and oligonucleotides encode DNA for sensing with nanopores. *J. Am. Chem. Soc.* **2009**, *131*, 7530-7531.

(20) 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.

(21) 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.

(22) Nakane, J.; Wiggin, M.; Marziali, A. A nanosensor for transmembrane capture and identification of single nucleic acid molecules. *Biophys. J.* **2004**, *87*, 615-621.

(23) Lee, J. W.; Meller, A. Rapid DNA sequencing by direct nanoscale reading of nucleotide bases on individual DNA chains. In *Perspectives in Bioanalysis*; Mitchelson, K., Ed.; Elsevier: Oxford, UK, 2007; 2, pp 245-263.

(24) Sauer-Budge, A. F.; Nyamwanda, J. A.; Lubensky, D. K.; Branton, D. Unzipping kinetics of double-stranded DNA in a nanopore. *Phys. Rev. Let.* **2003**, *90*, 238101.

(25) Viasnoff, V.; Chiaruttini, N.; Bockelmann, U. Probing DNA base pairing energy profiles using a nanopore. *Eur. Biophys. J.* **2009**, *38*, 263-269.

(26) Sutherland, T. C.; Dinsmore, M. J.; Kraatz, H.-B.; Lee, J. S. An analysis of mismatched duplex DNA unzipping through a bacterial nanopore. *Biochem. Cell Biol.* **2004**, *82*, 407-412.

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

(28) Liu, A.; Zhao, Q.; Krishantha, D. M. M.; Guan, X. Unzipping of double-stranded DNA in engineered α-hemolysin pores. *J. Phys. Chem. Lett.* **2011**, *2*, 1372-1376.

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

(30) Wu, H.-C.; Astier, Y.; Maglia, G.; Mikhailova, E.; Bayley, H. Protein nanopores with covalently attached molecular adapters. *J. Am. Chem. Soc.* **2007**, *129*, 16142-16148.

(31) Astier, Y.; Braha, O.; Bayley, H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *J. Am. Chem. Soc.* **2006**, *128*, 1705-1710.

(32) Banerjee, A.; Mikhailova, E.; Cheley, S.; Gu, L.-Q.; 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.

(33) Cockroft, S. L.; Chu, J.; Amorin, M.; Ghadiri, M. R. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *J. Am. Chem. Soc.* **2008**, *130*, 818-820.

(34) Hornblower, B.; Coombs, A.; Whitaker, R. D.; Kolomeisky, A.; Picone, S. J.; Meller, A.; Akeson, M. Single-molecule analysis of DNA-protein complexes using nanopores. *Nat. Meth.* **2007**, *4*, 315-317.

(35) Benner, S.; Chen, R. J. A.; Wilson, N. A.; Abu-Shumays, R.; Hurt, N.; Lieberman, K. R.; Deamer, D. W.; Dunbar, W. B.; Akeson, M. Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat. Nano* **2007**, *2*, 718-724.

(36) 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.

(37) 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-A precision. *Nat. Biotechnol.* **2012**, *30*, 344-348.

(38) Kumar, S.; Tao, C.; Chien, M.; Hellner, B.; Balijepalli, A.; Robertson, J. W. F.; Li, Z.; Russo, J. J.; Reiner, J. E.; Kasianowicz, J. J.; Ju, J. PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Sci. Rep.* **2012**, *2*.

(39) 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.

(40) 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.

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

(42) Purnell, R.; Schmidt, J. Measurements of DNA immobilized in the alpha-hemolysin nanopore. In *Nanopore-Based Technology*; Gracheva, M. E., Ed.; Humana Press, 2012; 870; pp 39-53.

(43) Howorka, S.; Bayley, H. Probing distance and electrical potential within a protein pore with tethered DNA. *Biophys. J.* **2002**, *83*, 3202-3210.

(44) Ayub, M.; Bayley, H. Individual RNA base recognition in immobilized oligonucleotides using a protein nanopore. *Nano Lett.* **2012**, *12*, 5637-5643.

(45) Mathé, J.; Aksimentiev, A.; Nelson, D. R.; Schulten, K.; Meller, A. Orientation discrimination of single-stranded DNA inside the α-hemolysin membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12377-12382.

(46) 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.

(47) Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S. Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein Sci.* **2006**, *15*, 459-467.

(48) Martin, H. S. C.; Jha, S.; Howorka, S.; Coveney, P. V. Determination of free energy profiles for the translocation of polynucleotides through α-hemolysin nanopores using non-equilibrium molecular dynamics simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2135-2148.

(49) Ashkenasy, N.; Sánchez-Quesada, J.; Bayley, H.; Ghadiri, M. R. Recognizing a single base in an individual DNA strand: a step toward DNA sequencing in nanopores. *Angew. Chem., Int. Ed.* **2005**, *44*, 1401-1404.

(50) 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.

(51) 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.

(52) Møller, P.; Folkmann, J. K.; Forchhammer, L.; Bräuner, E. V.; Danielsen, P. H.; Risom, L.; Loft, S. Air pollution, oxidative damage to DNA, and carcinogenesis. *Cancer Lett.* **2008**, *266*, 84-97.

(53) 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.

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

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

(56) Luo, W.; Muller, J. G.; Rachlin, E. M.; Burrows, C. J. Characterization of spiroiminodihydantoin as a product of one-electron oxidation of 8-oxo-7,8-dihydroguanosine. *Org. Lett.* **2000**, *2*, 613-616.

(57) Ye, Y.; Muller, J. G.; Luo, W.; Mayne, C. L.; Shallop, A. J.; Jones, R. A.; Burrows, C. J. Formation of 13C-, 15N-, and 18O-labeled guanidinohydantoin from guanosine oxidation with singlet oxygen. Implications for structure and mechanism. *J. Am. Chem. Soc.* **2003**, *125*, 13926-13927.

(58) Fleming, A. M.; Muller, J. G.; Dlouhy, A. C.; Burrows, C. J. Structural context effects in the oxidation of 8-oxo-7,8-dihydro-2′-deoxyguanosine to hydantoin products: electrostatics, base stacking, and base pairing. *J. Am. Chem. Soc.* **2012**, *134*, 15091-15102.

(59) Burrows, C. J.; Muller, J. G. Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **1998**, *98*, 1109-1152.

(60) Neeley, W. L.; Essigmann, J. M. Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem. Res. Toxicol.* **2006**, *19*, 491-505.

(61) Gedik, C. M.; Collins, A. Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study. *FASEB J.* **2004**, *19*, 82-84.

(62) Kornyushyna, O.; Berges, A. M.; Muller, J. G.; Burrows, C. J. In vitro nucleotide misinsertion opposite the oxidized guanosine lesions spiroiminodihydantoin and guanidinohydantoin and DNA synthesis past the lesions using *Escherichia coli* DNA polymerase I (Klenow fragment). *Biochemistry* **2002**, *41*, 15304-15314.

(63) Henderson, P. T.; Delaney, J. C.; Muller, J. G.; Neeley, W. L.; Tannenbaum, S. R.; Burrows, C. J.; Essigmann, J. M. The hydantoin lesions formed from oxidation of 7,8-dihydro-8-oxoguanine are potent sources of replication errors in vivo. *Biochemistry* **2003**, *42*, 9257-9262.

(64) 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. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E1820-E1829.

(65) 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.

(66) Cathcart, R.; Schwiers, E.; Saul, R. L.; Ames, B. N. Thymine glycol and thymidine glycol in human and rat urine: a possible assay for oxidative DNA damage. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 5633-5637.

(67) Dolinnaya, N. G.; Kubareva, E. A.; Romanova, E. A.; Trikin, R. M.; Oretskaya, T. S. Thymidine glycol: the effect on DNA molecular structure and enzymatic processing. *Biochimie* **2013**, *95*, 134-147.

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

(69) 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.

(70) Norval, M.; Lucas, R. M.; Cullen, A. P.; de Gruijl, F. R.; Longstreth, J.; Takizawa, Y.; van der Leun, J. C. The human health effects of ozone depletion and interactions with climate change. *Photochem. Photobiol. Sci.* **2011**, *10*, 199-225.

(71) Cordonnier, A. M.; Fuchs, R. P. P. Replication of damaged DNA: molecular defect in Xeroderma pigmentosum variant cells. *Mutat. Res.* **1999**, *435*, 111-119.

(72) Garinis, G. A.; van der Horst, G. T. J.; 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.

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

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

(75) Hübscher, U.; Maga, G.; Spadari, S. Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **2002**, *71*, 133-163.

(76) Goodman, F. M. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **2002**, *71*, 17-50.

(77) Kroeger, K. M.; Goodman, M. F.; Greenberg, M. M. A comprehensive comparison of DNA replication past 2-deoxyribose and its tetrahydrofuran analog in Escherichia coli. *Nucl. Acids Res.* **2004**, *32*, 5480-5485.

(78) House, C. H.; Miller, S. L. Hydrolysis of dihydrouridine and related compounds. *Biochemistry* **1996**, *35*, 315-320.

(79) Krokan, E. K.; Drabløs, F.; Slupphaug, G. Uracil in DNA-occurrence, consequencess and repair. *Oncogene* **2002**, *21*, 8935-8948.

(80) 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.

(81) 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.

(82) 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 nanopore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **2007**, *79*, 4778-4787.

(83) 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.

(84) Herskovits, T. T.; Bowen, J. J. Solution studies of the nucleic acid bases and related model compounds. Solubility in aqueous urea and amide solutions. *Biochemistry* **1974**, *13*, 5474-5483.

(85) Kang, X.-f.; Cheley, S.; Guan, X.; Bayley, H. Stochastic detection of enantiomers. *J. Am. Chem. Soc.* **2006**, *128*, 10684-10685.

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

(87) David, S. S.; O'Shea, V. L. Base-excision repair of oxidative DNA damage. *Nature* **2007**, *447* 941-950.

(88) He, L. Single-molecule study of thymidine glycol and i-motif through the alpha-hemolysin ion channel. University of Utah, 2013.

(89) An, N.; White, H. S.; Burrows, C. J. Modulation of the current signatures of DNA abasic site adducts in the alpha-hemolysin ion channel. *Chem. Commun.* **2012**, *48*, 11410-11412.

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

(91) 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.* **2007**, *130*, 703-709.

(92) Xue, L.; Greenberg, M. M. Facile quantification of lesions derived from 2'-deoxyguanosine in DNA. *J. Am. Chem. Soc.* **2007**, *129*, 7010-7011.

(93) Jacobs, A.; Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **2012**, *121*, 1-20.

(94) David, S. S.; Williams, S. D. Chemistry of glycosylases and endonucleases involved in base-excision repair. *Chem. Rev.* **1998**, *98*, 1221-1262.

(95) An, N.; Fleming, A. M.; Rosecrans, N. C.; Liao, Y.; Burrows, C. J. Synthesis of site-specific crown ether adducts to DNA abasic sites, 8-oxoguanosine and cytidine. *Methods Mol. Biol.* **2013**, *Submitted.*

(96) Beck, S.; Rakyan, V. K. The methylome: approaches for global DNA methylation profiling. *Trends in Genetics* **2008**, *24*, 231-237.

(97) Fu, Y.; He, C. Nucleic acid modifications with epigenetic significance. *Curr. Opinion Chem. Biol.* **2012**, *16*, 516-524.

(98) 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.

(99) Robertson, J. W. F.; Kasianowicz, J. J.; Banerjee, S. Analytical approaches for studying transporters, channels and porins. *Chem. Rev.* **2012**, *112*, 6227-6249.

(100) Butler, T. Z.; Pavlenok, M.; Derrington, I. M.; Niederweis, M.; Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20647-20652.

(101) Chen, M.; Khalid, S.; Sansom, M. S. P.; Bayley, H. Outer membrane protein G: Engineering a quiet pore for biosensing. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6272-6277.

(102) Manrao, E. A.; Derrington, I. M.; Pavlenok, M.; Niederweis, M.; H., G. J. Nucleotide discrimination with DNA immobilized in the MspA nanopore. *PLoS ONE* **2011**, *6*, e25723.

(103) 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.

CHAPTER 2

SINGLE-MOLECULE DETECTION OF GUANINE(C8) – THYMINE(N3) CROSS-LINK USING ION CHANNEL RECORDING

Introduction

Inflammation and oxidative stress are linked with the initiation and progression of many diseases such as cancer[1] and neurological disorders.[2] Inflammation produces a variety of oxidizing species that includes carbonate radical anion ($CO_3^{\bullet-}$). This one-electron oxidant is derived from $ONOO^-$ and HCO_3^- precursors that react to form nitrosoperoxycarbonate ($ONOOCO_2^-$) that undergoes spontaneous homolysis to yield the $CO_3^{\bullet-}$ radical under biological reaction conditions.[3,4] The DNA nucleobase guanine (G) has the lowest redox potential, thus it is the site most susceptible to oxidative insults.[5,6] G oxidation by $CO_3^{\bullet-}$ yields a wide range of products, shown in Figure 2.1, and their yields are dependent on the reaction context.[7,8] *In vitro* studies have shown that photochemically generated $CO_3^{\bullet-}$ selectively oxidizes G in single-stranded oligodeoxynucleotide (ODN) contexts to produce a cross-link in high yield. In a ssODN the cross-link was found in the sequence context 5'-GCT-3', in which the C8 position of G was covalently linked with the N3 position of T (5'-G*CT*-3').[9] Interestingly, the cross-link yield is highest when the G and T bases are separated by C, and G is at the 5'

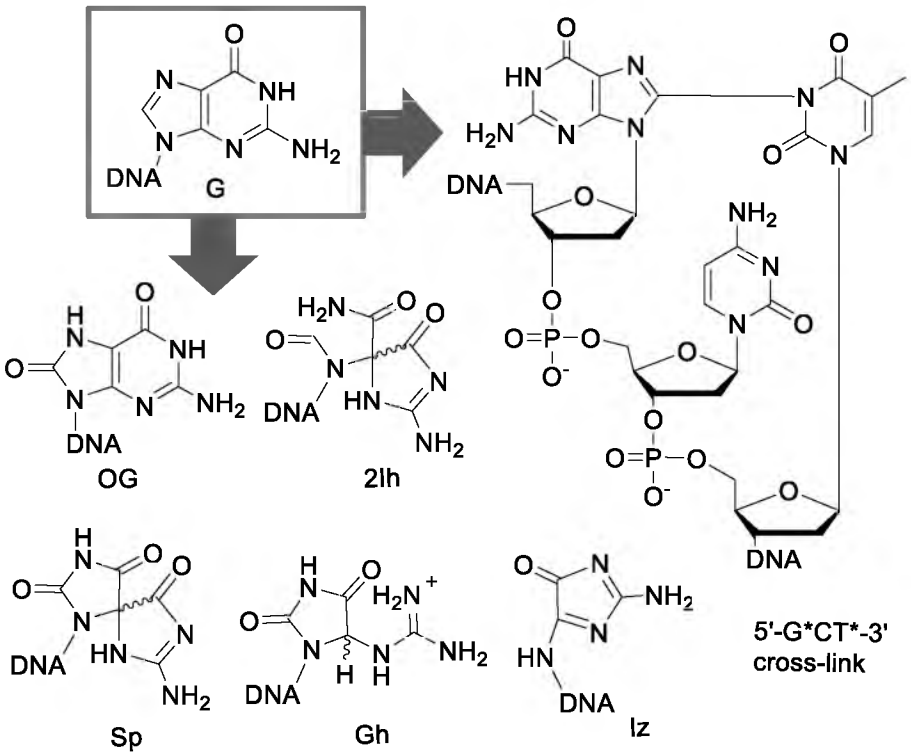


Figure 2.1. Guanine (G) lesions observed from $CO_3^{\bullet-}$ oxidation. 8-oxo-7,8-dihydroguanine (OG), 5-carboxamido-5-formamido-2-iminohydantoin (2Ih), spiroimino-dihydantoin (Sp), guanidinohydantoin (Gh), 2,5-diamino-4*H*-imidazolone (Iz), and the intrastrand 5'-G*CT*-3'cross-link.[10]

end of the sequence context (Figure 2.1).[9] Recently, the G*T* cross-link, the detectable form of the 5'-G*CT*-3' cross-link postnuclease digestion, was detected in human cells under conditions of inflammation.[11]

Methods currently used to quantify damaged bases in genomic DNA samples employ either the comet assay,[12,13] or enzymatic digestion followed by HPLC-MS/MS analysis;[14] the later method was used to quantify the G*T cross-link in human cells that was derived from the 5'-G*CT*-3' cross-link.[11] However, neither of these methods provides sequence information. Recent advances in single-molecule DNA sequencing with α-hemolysin (α-HL) nanopore offers the potential to obtain the sequence context and identify the type of damaged DNA.[15-19] This method harnesses the size limiting properties of the α-HL nanopore to give ion current signatures for the DNA bases when single-strand DNA (ssDNA) is allowed to pass through the pore under an electrophoretic potential.[20-22] However, sequencing based on free translocation of ssDNA proves to be problematic due to the fast rate at which the DNA passes through the pore.[16] To overcome this problem, and determine the ion current signature for the bases, immobilization experiments have been employed.[19,23-26] In this experiment, a biotin tag is added to the ssDNA terminus, upon binding streptavidin the complex is too big to pass through the pore, therefore allowing the complex to be captured for long times. Previously it was shown that placement of biotin 14 bases 3' to the site of interest (lesion) on ssDNA positions that sit in the most sensitive zone of α-HL, which is situated just past the 1.4 nm constriction zone at the top of the β-barrel.[27,28] Additionally, this trapped complex is stationary in the pore allowing the current level to be recorded for long times and signal averaged. Our laboratories have used this method to differentiate damaged

bases from the canonical ones based on their unique current levels.[19,29] Specifically, the G oxidation products spiroiminodihydantoin (Sp), guanidinohydantoin (Gh) and 8-oxo-7,8-dihydroguanine (OG; Figure 2.1), have been characterized using the immobilization experiment,[30] as well as an abasic site, thymine dimer and thymine glycol.[19] The unique chemical reactivity of abasic sites was harnessed to add chemical tags under appropriate reaction conditions that modulates the current signature well beyond that of all other bases, both conanical and lesion bearing; chemical modification to modulate the current signature has also been adopted in labeling 5-hydroxymethylcytosine, an epigenetic marker. [31] In the present study, the current-level signature for the 5'-G*CT*-3' cross-link was determined in an immobilization experiment (Figure 2.2) when the lesion was placed in the most sensitive region of α-HL, as well as dangling it deeper into the β-barrel to analyze positional effects in the ion current flow. Then the cross-link was synthesized in the middle of a 53-mer strand, which did not have biotin, and then free translocation experiments were conducted to determine if this lesion could be detected during a simulated translocation experiment.

Experimental

Materials and ODN preparation

All chemicals were purchased from commercial suppliers and used without further purification, except for the DNA strands. The DNA strands were obtained from the DNA-peptide core facility at the University of Utah and were purified by ion-exchange HPLC prior to their use. The 5'-G*CT*-3' cross-link was synthesized according to published procedures and purified by ion-exchange HPLC.[9,34,35] For the

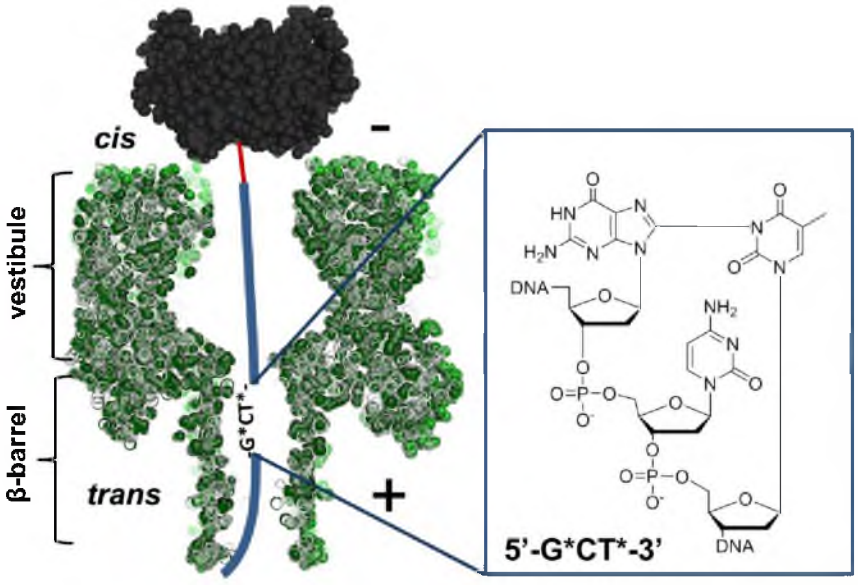


Figure 2.2. Representation of α-HL (pdb 7AHL)[32]/streptavidin (pdb 1MK5)[33] Btn-DNA complex that contains 5'-G*CT*-3' cross-link.

immobilization experiments a 40-mer 5'-CCCCCCCCCC CCCCCCCCCC CCCAACGCTA CCCCCCCCCC-Btn-3' and 5'-CCCCCCCCCC CCCCCCCCCA ACGCTACCCC CCCCCCCCCC-Btn-3' sequence were studied, and to avoid any unwanted oxidation reactions on the biotin tag, desthiobiotin was used. For the free translocation experiments, a 53-mer 5'-CCCCCCCCCC CCCCCCCCCC CCAACGCTAC CCCAAAAAAA AAAAAAAAAA AAA-3' sequence was studied.

The 5'-G*CT*-3' cross-link was introduced by irradiating for 2 h (254 nm wavelength, the shorter strand was irradiated for 30 min) a 100 μL solution containing 1 nmol of ODN, 20 mM of $Na_2S_2O_8$, 300 mM $NaHCO_3$, pH 10.5. The reaction mixture was than HPLC purified running a method that consisted of a linear gradient of 15% to 100% B over 30 min where the mobile phases were: B: 1.0 M NaCl, 25 mM Tris, 10% ACN, pH 8, A: 10% ACN, with a flow rate of 1 mL/min while monitoring the absorbance at 260 nm. The thus obtained pure 5'-G*CT*-3' cross-link containing ODN was then dialyzed to remove excess purification salts. Product purity was determined by reinjection of the sample onto an analytical ion-exchange column running the previously described method, and the product identity was determined by ESI-MS.

Ion channel recordings

The ion channel recordings were conducted with a custom built amplifier and data acquisition system designed by Eletronic BioSciences (EBS), San Diego, CA. The glass nanopore membrane (GNM) was fabricated using previously established procedures. [36,37] The immobilization and translocation events were collected in 1 M KCl, 1 mM EDTA, 25 mM Tris, pH 7.9 electrolyte solution under a 120 mV bias (*trans* vs. *cis*) with a 10

kHz filter, and data acquisition rate of 50 kHz. For the immobilization experiments the DNA-streptavidin (DNA-Strep) complex was obtained by mixing a 200 nM solution of biotinylated DNA with a 50 nM solution of streptavidin in the analysis electrolyte solution for 10 min at 22 ^{o}C prior to adding this sample to the analysis chamber. In the immobilization experiments a 200 nM DNA-streptavidin complex containing the lesion was added to the chamber. Then, after recording >200 events the same amount of a non-lesion bearing DNA-streptavidin complex was added to determine the current modulation of the cross-link relative to the native-DNA sequence. Lastly, the 3'-biotinylated 40-mer poly-C sequence was added as an internal standard for referencing the new ion currents to those obtained in our previously published results.[29,30] The data were analyzed using software donated by EBS. The events were extracted using QUB 2.0.0.20, and the immobilization experiments were plotted using Igor Pro 6.1.

In the translocation studies, a 5 μM DNA solution of ssDNA was added to the chamber and >1000 translocation events were collected at a 100, 120 and 140 mV (*trans* vs. *cis*) bias using a 100 kHz filter and a 500 kHz data acquisition rate. In these studies, both lesion-containing and nonlesion containing DNA strands were analyzed in a 1.00 M KCl, 10 mM NaP_i (pH 7.4), 1 mM EDTA electrolyte solution. The data were analyzed using the same software as previously mentioned.

Results and discussion

The most commonly occurring G*T* cross-link observed from $CO_3^{\bullet-}$ oxidation has the G and T separated by a C (5'-G*CT*-3').[9] More specifically, the Geacintov laboratory has studied this cross-link in the sequence context 5'-CCATCGCTACC-3'.[38]

Therefore, we elected to embed this sequence (substituting the 5' T with A to prevent formation of the 5'-T*CG*-3' cross-link) in a 3'-desthiobiotinylated 40-mer poly-C homopolymer background that placed the cross-link at the most sensitive region of α-HL, positions $\omega_{12\text{-}14}$ (this nomenclature was chosen to identify that counting started from the nontraditional 3'-end). This strand will be noted throughout the text with the 5'-G*CT*-3' nomenclature, while the native sequence will be designated by the 5'-GCT-3' nomenclature. Next, ion channel recordings were commenced to determine the ion-current signature for the 5'-G*CT*-3' cross-link in a wild-type α-HL that was imbedded in a lipid bilayer suspended across a GNM.[36,37] When an electrical potential was applied across the channel, the Strep-Btn DNA complex was eletrophoretically driven through the pore causing a deep current level blockage (I) to the open channel current (I_o). The deep-current level blockage was recorded for 1 s, followed by the reversal of polarity that drove the Strep-Btn DNA complex into the bulk solution, after which the open channel current was restored. A typical current-time (i-t) trace is shown in Figure 1.2 in Chapter 1. The capture/release cycle was repeated >200 times to collect a population of events. The percentage residual current ($\%I/I_o$) for each event was calculated and plotted into a current-level histogram.

The current blockages for the 5'-G*CT*-3' cross-link, the native sequence (5'-GCT-3'), both at $\omega_{12\text{-}14}$, and the internal standard (C_{40}) were then evaluated. The C_{40} current-blockage level, which was used as a reference, was assigned to a value of $\Delta\%I/I_o = 0$. Figure 2.3A shows a histogram of the percent blockage current level when the G of the 5'-GCT-3' sequence was placed at position ω_{14} for a lesion and native strands. In this

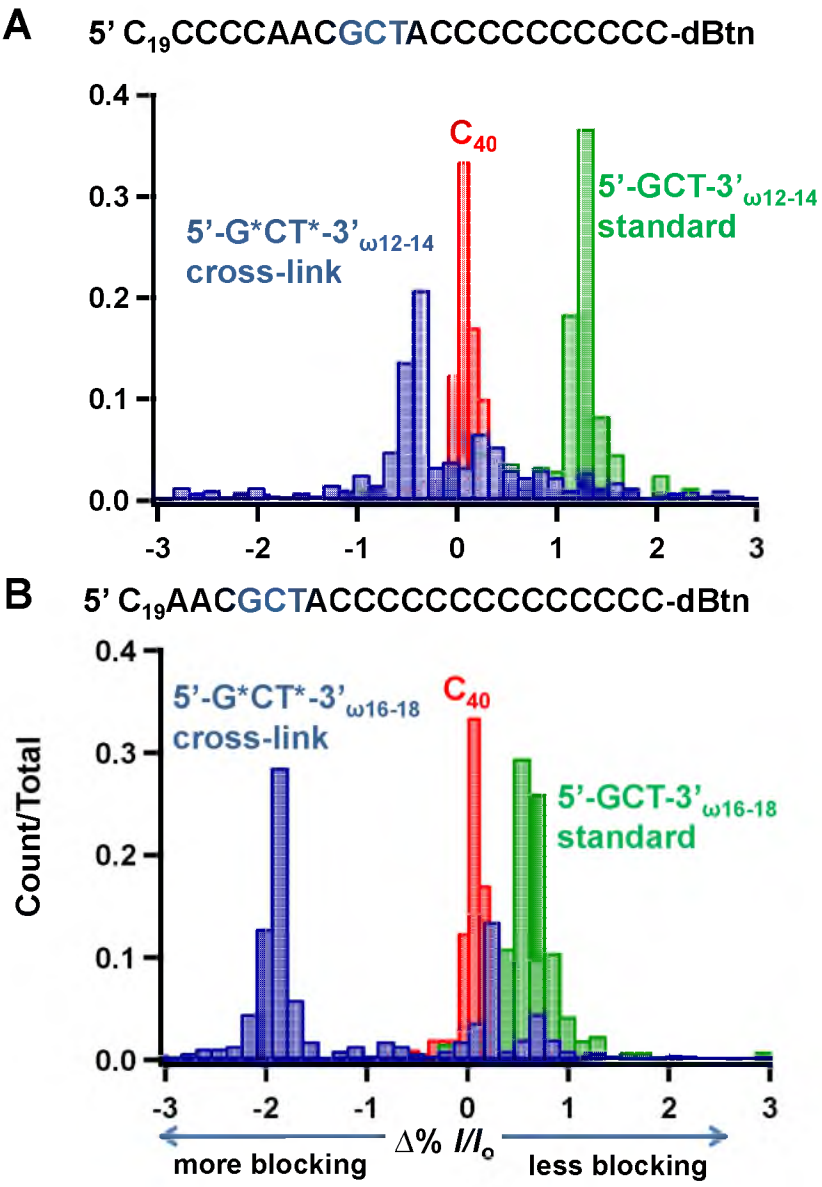
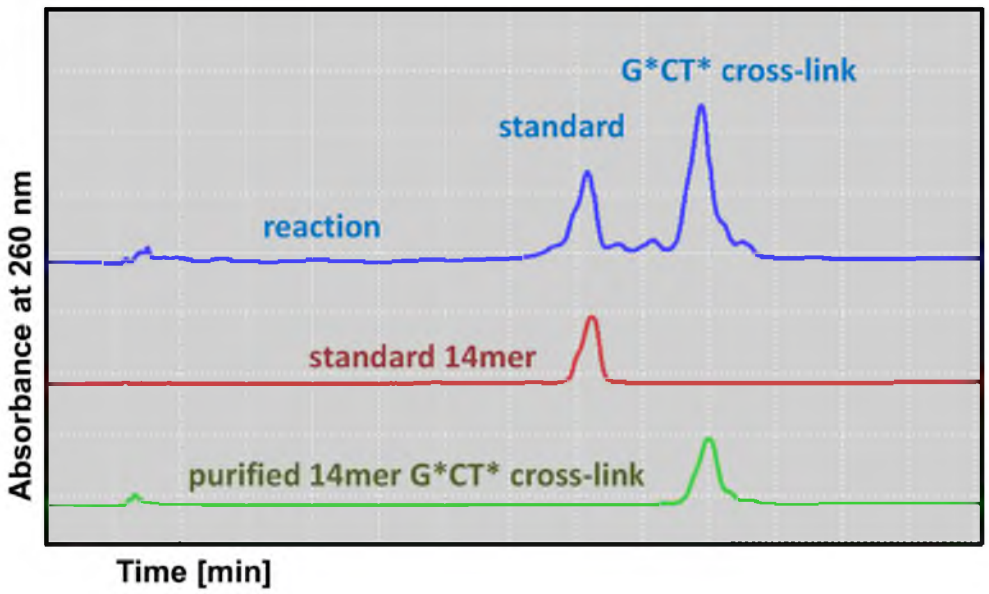


Figure 2.3. Current level histograms for the 5'G*CT*-3' cross-link, 5'-GCT-3' and the C_{40} reference strand. **A.** Histograms recorded when the G of the 5'GCT-3'sequence was positioned at ω_{14}. **B.** Histograms recorded when the G of the 5'GCT-3' sequence was positioned at ω_{18}. In both histograms the 5'-G*CT*-3' cross-link is shown in blue, the unmodified 5'-GCT-3' sequence is shown in green, and the C_{40} internal standard is shown in red.[10]

sequence context the difference between the unmodified- and the 5'-G*CT*-3' cross-link containing strand was $\Delta\%I/I_o$ = 1.7%, with the cross-link being the more blocking to the current level. Additionally, this places the 5'- G*CT*-3' strand at a value of $\Delta\%I/I_o$ = 0.5% more blocking than the C_{40} reference (Figure 2.3A). Figure 2.3B shows a histogram of the percent blockage current when the reactive G was placed deeper into the β-barrel at the position ω_{18}. This increased the difference that was observed between the cross-link and the unreacted standard. Placement of the cross-link deeper into the β-barrel gave a $\Delta\%I/I_o$ = 2.4%, again the crosslink gave a deeper blockage to the current in comparison to the unmodified strand. Additionally, by placing the damaged site deeper into the channel (4 bases) the blocking current increased by a $\Delta\%I/I_o$ = 0.7%. These results demonstrate that the unencumbered 5'-G*C*T-3' cross-link is more blocking to the current level, and this effect is enhanced as the lesion is dangled deeper into the β-barrel. The difference in the current-blockage level measured between the unmodified and the cross-link strands is hypothesized to result from the increased rigidity and bulk of the 5'-G*CT*-3' cross-link. Consistent with this hypothesis is a previous molecular dynamics (MD) simulation conducted by the Broyde and Geacintov laboratories that revealed that the 5'-G*CT*-3' cross-link greatly disturb base stacking interactions;[38] furthermore, this cross-link severely unwinds double-strand DNA based on MD analysis, by introducing a sever kink in the sugar-phosphate backbone, that ejects the C base out toward solvent. It is anticipated that these distorting effects will exist in ssDNA as well, especially when the strand is confined in the sterically-demanding nanopore channel. Thus, the ion flow through the channel is perturbed more with the cross-link, and this is reflected in the greater ion current blockage.

In both current-level histograms (Figure 2.3A and B) the 5'-G*CT*-3' strand had a second peak that was close in $\Delta\%I/I_o$ current blockage to the C_{40} reference. We hypothesize that this minor peak (~20% of the recorded events) might be due to the presence of the other G-oxidation product, 2Ih that has been observed from $CO_3^{\bullet-}$ oxidations (Figure 2.1).[35] Consistent with this hypothesis is a peak in the ESI-MS for this sample that has the same mass as 2-Ih (Figure 2.4) and a similar relative concentration (~20%) as was detected in the immobilization experiment. Unfortunately, due to the long length of this strand (40-mer), the cross-link and 2-Ih were not resolvable, so further studies to confirm this hypothesis were not conducted.

The immobilization experiments presented above show that the 5'-G*CT*-3' cross-link can be differentiated from the unmodified 5'-GCT-3' strand based on current-level modulation (Figure 2.3A and B). Additionally, the 5'-G*CT*-3' cross-link was considerably more blocking to the ion current. Because the same C_{40} internal standard has been used in all of our previous studies, it is also known that the cross-link is >2.0% more blocking than any of the natural base within a poly-C background.[28,30] The next step was to examine the behavior of the 5'-G*CT*-3' cross-link in a single stranded DNA construct as it translocated through the α-HL ion channel under an electrophoretic potential. During the translocation experiments the DNA was allowed to freely pass through the protein channel from the *cis* to *trans* side of the pore, under a 100, 120 and 140 mV bias (*trans* vs. *cis*). The sequence 5'- C_{20}CCAAC<u>GCT</u>ACCCCA$_{20}$-3' was used for these experiments, and it was designed to enhance the 5'-capture events at lower bias. Capture of the 5'-end has previously been demonstrated to be a minor event type compared to 3'-capture at low bias.[39-41] Additionally, in our previous studies it was



Damage in 14mer	Calculated mass [M+H]	Experimental mass [M+H]
5'-G*CT*-3'	4160.79	4160.32
2Ih	4194.80	4195.38

Figure 2.4. Representative HPLC traces and ESI MS characterization of the 5'-G*CT*-3' cross-link in a 14-mer ODN (5'-CCAAC**GCT**ACCACA-3'). HPLC mobile phases, B: 1.5M NaOAc, 10% ACN, pH 7, A: 10% ACN. HPLC method: linear gradient starting at 30% B to 100% B over 30 min on an analytical ion-exchange column.

determined that current modulation of adducts (18-crown-6 adducted to an abasic site) only modulated the current upon capture of the 5'-end;[39] therefore, a stretch of polyC was placed 5'-end to the cross-link, and a polyA stretch was placed 3' to the cross-link, because it was previously demonstrated that 5'-entry is maximal with polyC while being minimal with polyA.[39,40]

Figure 2.5A shows an *i-t* trace for a single DNA molecule transloction event for an unmodified 53-mer ODN. The data collected from the translocation experiments were investigated with respect to directionality effects on entry and the time duration (t_D) for the translocation to occur. Consistent with the diffusional broadening of the translocation times, the t_D histogram of each population was described by a Gaussian curve with a mean peak value t_{max}. The majority of the events (75%) displayed a constant deep blockage current level throughout the translocation (Figure 2.5A), with a small fraction of the events (25%) featuring a shoulder current level that preceded the deep blockage current level. The shoulder signature has previously been shown to result from the polyA tail interacting with the vestibule above the constriction zone.[21] The individual *i-t* traces were then plotted into density plots that are shown in Figure 2.5B, in which two populations of events were observed that differed in the deep blockage current level. Based on the literature reports, these two event populations represent the highest blockage to the current (-16 pA; 0.13 I/I_o) to be the 3' polyA entry first, and the lower blockage to the current (-22 pA; 0.18 I/I_o) to be the 5' polyC entry first.[39,40] Next, analysis was conducted to determine if there were any detectable differences between the reacted and unreacted strands.

First, inspection of the *i-t* traces for the cross-link and unmodified strands were

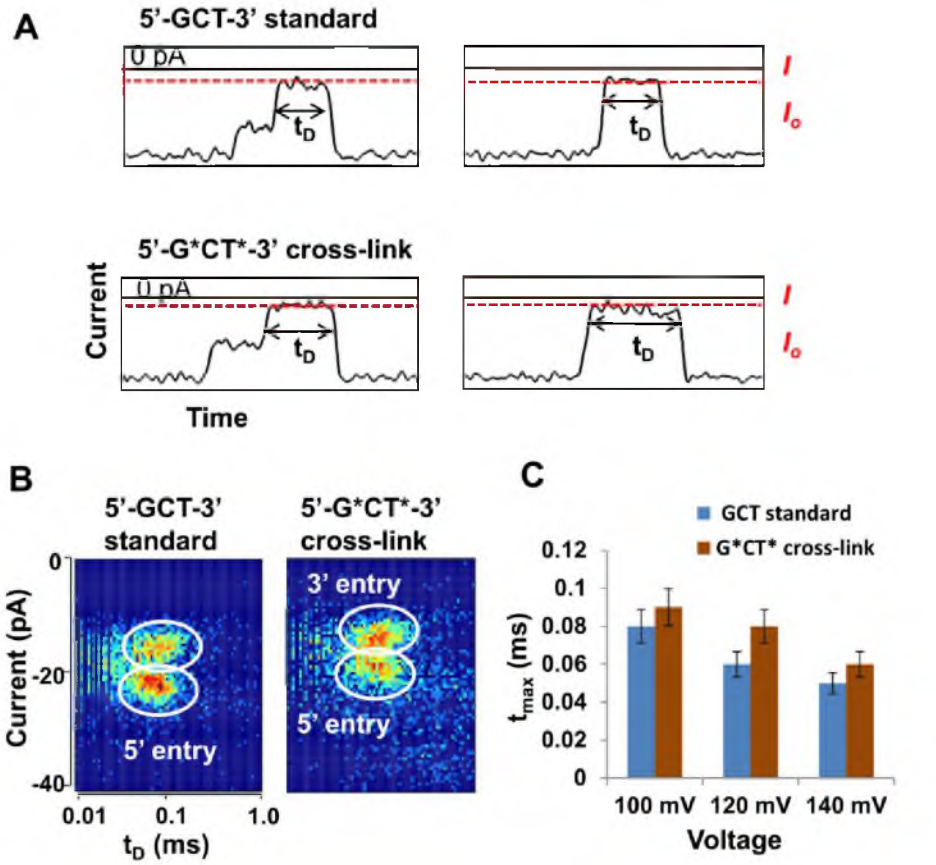


Figure 2.5. Translocation studies of the 5'-G*CT*-3' cross-link in a 5' C_{20}CCAAC**GCT**ACCCCA$_{20}$ sequence. **A.** Typical *i-t* trace of a single DNA molecule translocating through the pore. **B.** Density plots of the control and cross-link strands at 120 mV (*trans* vs. *cis*). **C.** Voltage dependence study showing the t_{max} vs. voltage plot.[10]

looked at, and from Figure 2.5A, there did not appear to be any detectable differences in the *i-t* trace signature. Analysis of the *i-t* traces was looked at for both 3' and 5' entry, and no observable differences were recorded. Next, analysis of the translocation time, t_{max}, revealed that at 100 mV bias (*trans* vs. *cis*) both strands gave similar 3' vs. 5' entries, and t_{max} values (Figure 2.5C). Finally, a voltage dependence study was performed to determine if the cross-link was small enough to fit and translocate across the α-HL ion channel during free translocation. From these data (Figure 5C), it was determined that for both the standard and cross-link strands the translocation time decreased as the voltage increased. This is a diagnostic signature of a strand being able to translocate from the *cis* to *trans* side of α-HL.[42] These results are interpreted to mean that the 5'G*CT*-3' cross-link is small enough to fit and translocate across the pore, and it does not slow down the translocation speed; thus, this cross-link cannot be differentiated from an undamaged strand during free translocation based on either translocation current or time.

Conclusions

The studies described herein have shown that the α-HL ion channel has the ability to detect the 5'-G*CT*-3' cross-link in an immobilization experiment; however, during a free translocation experiment the cross-link could not be differentiated from an unmodified strand. The immobilization results are promising, because they provide a unique current-level signature for this damage. Additionally, the bigger current difference can be observed when the cross-link was placed deeper inside the β-barrel. Once single-molecule DNA sequencing is perfected, this study will provide the signature

for this cross-link that is likely to be observed during any DNA sequencing experiment. Furthermore, because it has been shown that cross-links are highly toxic to cellular processes that require an intact genome (i.e., replication and transcription),[43] finding their specific location will aid in better understanding how oxidative and inflammatory stress damage the genome.

References

(1) Lonkar, P.; Dedon, P. C. Reactive species and DNA damage in chronic inflammation: reconciling chemical mechanisms and biological fates. *Int. J. Cancer* **2011**, *128*, 1999-2009.

(2) Mangialasche, F.; Polidori, M. C.; Monastero, R.; Ercolani, S.; Camarda, C.; Cecchetti, R.; Mecocci, P. Biomarkers of oxidative and nitrosative damage in Alzheimer's disease and mild cognitive impairment. *Ageing Res. Rev.* **2009**, *8*, 285-305.

(3) Pacher, P.; Beckman, J. S.; Liaudet, L. Nitric oxide and peroxynitrite in health and disease. *Physiol. Rev.* **2007**, *87*, 315-424.

(4) Goldstein, S.; Czapski, G. Viscosity effects on the reaction of peroxynitrite with CO_2: evidence for radical formation in a solvent cage. *J. Am. Chem. Soc.* **1999**, *121*, 2444-2447.

(5) Neeley, W. L.; Essigmann, J. M. Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem. Res. Toxicol.* **2006**, *19*, 491-505.

(6) Burrows, C. J.; Muller, J. G. Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **1998**, *98*, 1109-1152.

(7) Misiaszek, R.; Uvaydov, Y.; Crean, C.; Geacintov, N. E.; Shafirovich, V. Combination reactions of superoxide with 8-oxo-7,8-dihydroguanine radicals in DNA: kinetics and end products. *J. Biol. Chem.* **2005**, *280*, 6293-6300.

(8) Fleming, A. M.; Muller, J. G.; Ji, I.; Burrows, C. J. Characterization of 2'-deoxyguanosine oxidation products observed in the Fenton-like system Cu(II)/H_2O_2/reductant in nucleoside and oligodeoxynucleotide contexts. *Org. Biomol. Chem.* **2011**, *9*, 3338-3348.

(9) Crean, C.; Uvaydov, Y.; Geacintov, N. E.; Shafirovich, V. Oxidation of single-stranded oligonucleotides by carbonate radical anions: generating intrastrand

cross-links between guanine and thymine bases separated by cytosines. *Nucleic Acids Res.* **2008**, *36*, 742-755.

(10) Wolna, A. H.; Fleming, A. M.; Burrows, C. J. Single-molecule detection of a guanine(C8)-thymine(N3) cross-link using ion channel recording. *J. Phys. Org. Chem.* **2014**, *27*, 247-251.

(11) Madugundu, G. S.; Wagner, J. R.; Cadet, J.; Kropachev, K.; Yun, B. H.; Geacintov, N. E.; Shafirovich, V. Generation of guanine-thymine cross-links in human cells by one-electron oxidation mechanisms. *Chem. Res. Toxicol.* **2013**, *26*, 1031-1033.

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

(13) Liao, W.; McNutt, M. A.; Zhu, W. G. The comet assay: a sensitive method for detecting DNA damage in individual cells. *Methods* **2009**, *48*, 46-53.

(14) Shigenaga, M. K.; Aboujaoude, E. N.; Chen, Q.; Ames, B. N. Assays of oxidative DNA damage biomarkers 8-oxo-2'-deoxyguanosine and 8-oxoguanine in nuclear DNA and biological fluids by high-performance liquid chromatography with electrochemical detection. *Methods Enzymol.* **1994**, *234*, 16-33.

(15) Bayley, H. Sequencing single molecules of DNA. *Curr. Opin. Chem. Biol.* **2006**, *10*, 628-637.

(16) Branton, D.; Deamer, D., W; Marziali, A.; Bayley, H.; Benner, S., A; Butler, T.; Di Ventra, M.; 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.

(17) 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-A precision. *Nat. Biotechnol.* **2012**, *30*, 344-348.

(18) Larkin, J.; Carson, S.; Stoloff, D. H.; Wanunu, M. Nanopore-based analysis of chemically modified DNA and nucleic acid drug targets. *Isr. J. Chem.* **2013**, *53*, 431-441.

(19) Wolna, A. H.; Fleming, A. M.; An, N.; He, L.; White, H. S.; Burrows, C. J. Electrical current signatures of DNA base modifications in single molecules immobilized in the α-hemolysin ion channel. *Isr. J. Chem.* **2013**, *53*, 417-430.

(20) 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.

(21) Butler, T. Z.; Gundlach, J. H.; Troll, M. Ionic current blockades from DNA and RNA molecules in the α-hemolysin nanopore. *Biophys. J.* **2007**, *93*, 3229-3240.

(22) Meller, A.; Branton, D. Single molecule measurements of DNA transport through a nanopore. *Electrophoresis* **2002**, *23*, 2583-2591.

(23) 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.

(24) Howorka, S.; Bayley, H. Probing distance and electrical potential within a protein pore with tethered DNA. *Biophys. J.* **2002**, *83*, 3202-3210.

(25) Ayub, M.; Bayley, H. Individual RNA base recognition in immobilized oligonucleotides using a protein nanopore. *Nano Lett.* **2012**, *12*, 5637-5643.

(26) Purnell, R.; Schmidt, J. Measurements of DNA immobilized in the alpha-hemolysin nanopore. In *Nanopore-Based Technology*; Gracheva, M. E., Ed.; Humana Press, 2012; Vol. 870; pp 39-53.

(27) Ashkenasy, N.; Sánchez-Quesada, J.; Bayley, H.; Ghadiri, M. R. Recognizing a single base in an individual DNA strand: a step toward DNA sequencing in nanopores. *Angew. Chem. Int. Ed.* **2005**, *44*, 1401-1404.

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

(29) An, N.; White, H. S.; Burrows, C. J. Modulation of the current signatures of DNA abasic site adducts in the [small alpha]-hemolysin ion channel. *Chem. Commun.* **2012**, *48*, 11410-11412.

(30) 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.

(31) Li, W.-W.; Gong, L.; Bayley, H. Single-molecule detection of 5-hydroxymethylcytosine in DNA through chemical modification and nanopore analysis. *Angew. Chem. Int. Ed.* **2013**, *52*, 4350-4355.

(32) Song, L.; Haobaugh, M. R.; Shustak C.; Cheley, S.; Bayley, H.; Gouaux, E. J. Structure of staphylococcal α-hemolysin, heptameric transmembrane pore. *Science* **1996**, *274*, 1859-1866.

(33) Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S. Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein Sci.* **2006**, *15*, 459-467.

(34) Crean, C.; Lee, Y. A.; Yun, B. H.; Geacintov, N. E.; Shafirovich, V. Oxidation of guanine by carbonate radicals derived from photolysis of carbonatotetramminecobalt(III) complexes and the pH dependence of intrastrand DNA cross-links mediated by guanine radical reactions. *ChemBioChem* **2008**, *9*, 1985-1991.

(35) Rokhlenko, Y.; Geacintov, N. E.; Shafirovich, V. Lifetimes and reaction pathways of guanine radical cations and neutral guanine radicals in an oligonucleotide in aqueous solutions. *J. Am. Chem. Soc.* **2012**, *134*, 4955-4962.

(36) 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 nanopore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **2007**, *79*, 4778-4787.

(37) 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.

(38) Ding, S.; Kropachev, K.; Cai, Y.; Kolbanovskiy, M.; Durandina, S. A.; Liu, Z.; Shafirovich, V.; Broyde, S.; Geacintov, N. E. Structural, energetic and dynamic properties of guanine(C8)–thymine(N3) cross-links in DNA provide insights on susceptibility to nucleotide excision repair. *Nucleic Acids Res.* **2012**, *40*, 2506-2517.

(39) 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.

(40) 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.

(41) Mathé, J.; Aksimentiev, A.; Nelson, D. R.; Schulten, K.; Meller, A. Orientation discrimination of single-stranded DNA inside the α-hemolysin membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12377-12382.

(42) Meller, A.; Nivon, L.; Branton, D. Voltage-driven DNA translocations through a nanopore. *Phys. Rev. Lett.* **2001**, *86*, 3435-3438.

(43) Kanuri, M.; Nechev, L. V.; Kiehna, S. E.; Tamura, P. J.; Harris, C. M.; Harris, T. M.; Lloyd, R. S. Evidence for *Escherichia coli* polymerase II mutagenic bypass of intrastrand DNA crosslinks. *DNA Repair* **2005**, *4*, 1374-1380.

CHAPTER 3

SINGLE-MOLECULE DETECTION OF DNA-PLATINUM CROSS-LINKS USING ION CHANNEL RECORDINGS

Introduction

DNA damage is known to cause mutations that may eventually lead to the development of cancer. Ironically, one of the most successful cancer treatments currently available, platinum-based anticancer drugs, causes even more DNA damage, leading to the cell no longer being able to process the genetic information and thus causing apoptosis.[1,2] Platinum-based anticancer drugs target DNA to form intrastrand and interstrand DNA-platinum cross-links that link adjacent GG bases (Figure 3.1) or GA bases at the N7 positions.[3,4] *cis*-Diamminedichloroplatinum (II) or cisplatin, (Figure 3.1) was the first platinum based chemotherapeutic agent that was serendipitously discovered by Rosenberg in 1965,[5] and FDA approved in 1978.[1] Cisplatin is especially successful in curing testicular germ-cell cancer (90% overall cure rate, 100% if diagnosed early),[6] but is also used in the treatment of several other forms of cancer such as nonsmall-cell lung, cervical, ovarian, esophageal, and head and neck cancers.[7-10] Cisplatin, although successful in some cancer chemotherapies, cannot be used in many cases due to its toxicity and side effects,[7] as well as acquired or inherent resistance to the drug.[11]

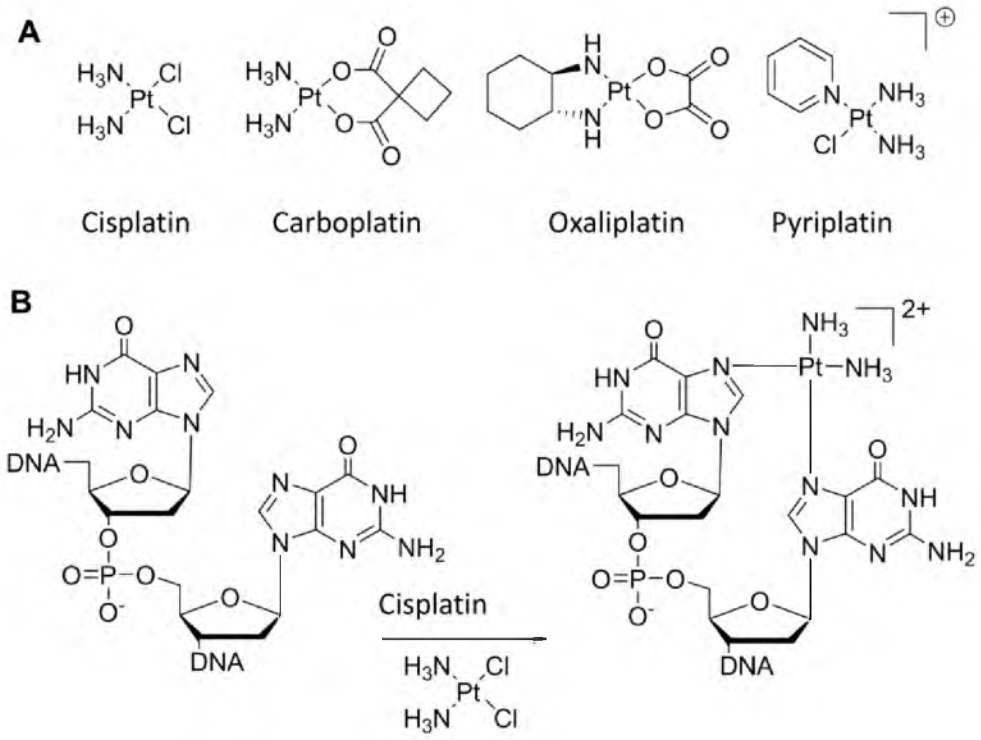


Figure 3.1. Structures of selected platinum anticancer drugs. **A.** Cisplatin, carboplatin and oxaliplatin are bifunctional DNA platinating drugs that are FDA approved. Pyriplatin is a monofunctional DNA adduct. **B.** Platinum binds the DNA at the N7 positions of adjacent G residues.

Subsequently, carboplatin and oxaliplatin (Figure 3.1A) have been approved by the FDA. Carboplatin and cisplatin lead to the same DNA cross-link after activation,[3,12] and they are also cross-resistant.[13] Oxaliplatin (Figure 3.1A) was shown to have a different spectrum of activity than cisplatin or carboplatin, and has become a preferred chemotherapeutic for colorectal cancer.[14] Many other platinum-based drugs were also shown to destroy cancerous cells,[15-18] among which there are the nonclassical, monofunctional platinum adducts such as pyriplatin (Figure 3.1A).[19]

Platinum-based compounds trigger cell death by a mechanism that involves four initial stages:[3,12] (1) The uptake of the drug into the cell is accomplished by using both passive and active mechanisms.[20-22] (2) The activation of the compound involves replacement of the chloride ligands or other labile leaving groups with water molecules.[2] (3) Once inside the cell, the platinum drug can bind to RNA, proteins, phospholipids or filaments; however, its main target has been established to be DNA (Figure 3.1B).[23,24] (4) The platinum DNA-cross links activate the DNA damage response pathway, leading to the cell undergoing apoptosis if the damage is too severe to repair.[1,12] The majority of the DNA-cisplatin adducts (~65% of the total products) are GG intrastrand cross-links; however, others such as AG (~25%) and GNG (~5-10%) intrastrand cross-links were also observed *in vitro* and *in vivo*.[25,26] Although the platinum-based drugs have been in clinical use for over 35 years, many aspects of their mechanism of action are still not clear; for example, it is unknown whether there are certain hot spots along the genome that are more predisposed to platinum adduct formation.

In order to be able to determine possible hot spots for the platinum adduct formation, a single-molecule sequencing platform needs to be utilized. Ever since the

first human genome was sequenced, many novel sequencing technologies were developed,[27,28] among which alpha-hemolysin (α-HL, Figure 1.1, Chapter 1) ion channel recordings offer many advantages, one of them being able to distinguish oxidative DNA damages from the canonical bases (Chapter 1 and 2).[29-34] The α-HL is a protein excreted by *Staphylococcus aureus* that can self-assemble into a lipid bilayer to form a heptameric ion channel consisting of a vestibule and a β-barrel with a constriction zone of 1.4 nm (Figure 1.1).[35] The dimensions of the protein channel allow only single-strand DNA (ssDNA) to be able to translocate across the pore. Currently, due to the fast translocation time and low signal-to-noise ratio, the best method to establish a current level for a particular damage is to implement an immobilization experiment.[29,36-39] During this experiment a DNA molecule is immobilized inside the α-HL ion channel allowing the signal to be averaged for a precise current determination. The damage is introduced at position ω14 along the DNA strand, which places it in the most sensitive sensing zone inside the β-barrel of the α-HL.[39] The current study focuses on establishing the current levels for the DNA damage that were caused by chemotherapeutic drugs such as cisplatin, oxaliplatin and pryiplatin (Figure 3.2). It was shown that during the immobilization experiments those cross-links gave significantly different current levels than the canonical DNA bases.

Experimental

Materials and oligodeoxynucleotide (ODN) preparation

All chemicals and reagents were purchased from commercial suppliers and were not further purified unless otherwise stated. ODNs were synthesized at the DNA-peptide

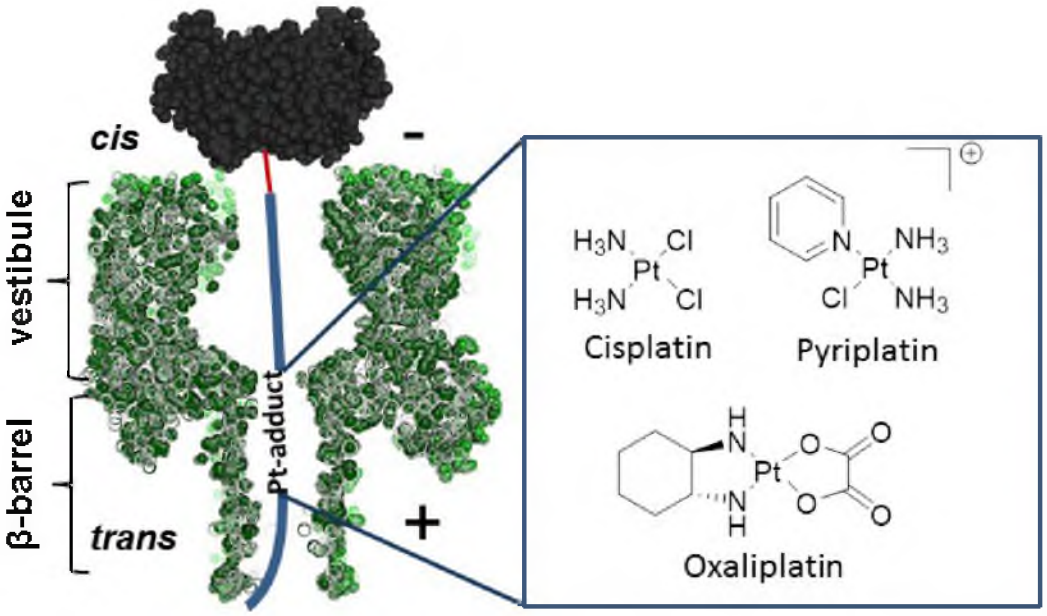


Figure 3.2. Representation of α-HL (pdb 7AHL)[35] /streptavidin (pdb 1MK5)[40] Btn-DNA complex that contains cisplatin, oxaliplatin or pytriplatin cross-link.

core facility at the University of Utah and were purified on preparative ion exchange HPLC before use. The cisplatin, oxaliplatin and pyriplatin adducts were synthesized according to the previously established protocols.[41-43] For the immobilization experiment the sequences used for the platination reactions were 40-mers: 5'-CCCCCCCCCC CCCCCCCCCC CCCCCGGCCC CCCCCCCCCC-Btn-3' for the bifunctional adducts (cisplatin and oxaliplatin) and 5'-CCCCCCCCCC CCCCCCCCCC CCCCCCGCCC CCCCCCCCCC-Btn-3' for the monofunctional adduct (pyriplatin), also a polyC strand was used as an internal standard. For the translocation experiments the sequences were: 5'-CCCCCCCCCC CCCCCCCCCC GG CCCCCCCCCC CCCCCCCC-3' for the bifunctional adduct and 5'-CCCCCCCCCC CCCCCCCCCC G CCCCCCCCCC CCCCCCCCC-3' for a monofunctional adduct.

The platination reaction was conducted by first incubating the platinum compound for 2 days in the dark with 2 equivalents of $AgNO_3$ in water, for pyriplatin and cisplatin.[41] Oxaliplatin was incubated in phosphate buffer at pH 7.4 with the rest of the conditions the same as the other compounds.[44] This step yields aquated derivatives of the platinum drugs that were then centrifuged to remove the AgCl precipitate. The above mentioned ODNs were then incubated with the aquated compound for 12 h at 37 ^{o}C in phosphate buffer, pH 7.4. The platinated ODNs were HPLC purified on an ion-exchange column running a method that consisted of a linear gradient of 20% to 100% B over 30 min where the mobile phases were: B: 1.0 M NaCl, 25 mM Tris (pH 8), 10% ACN, pH 8, A: 10% ACN, with a flow rate of 1 mL/min while monitoring the absorbance at 260 nm. After purification the pure platinum-DNA cross-links were dialyzed to remove excess purification salts. Product purity was determined by reinjection of the sample onto an

analytical ion-exchange column running the previously described method, and the product identity was determined by mass spectrometry.

Ion channel recordings

The translocation and immobilization experiments were conducted on custom made instrumentation donated to our laboratory by Eletronic BioSciences (EBS), San Diego, CA. The glass nanopore membrane (GNM) was fabricated according to the previously established procedures.[45] The surface of the GNM was modified with 3-cyanopropyldimehtylchlorosilane before use.[46] The phospholipid used during the nanopore experiments to form the lipid bilayer was 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC). The nanopore set up was as follows: (1) an Ag/AgCl electrode was positioned inside the GNM (filled with the electrolyte solution) and attached to a pressure gauge and a 10-mL gastight syringe. (2) Another Ag/AgCl electrode was placed in the nanopore holder chamber that was than filled with electrolyte solution. Upon the addition of DPhPC, a resistance increase, associated with the suspended lipid bilayer formation, was observed. The protein channel was inserted into the lipid bilayer by applying a pressure to the back of the GNM, which resulted in the decrease of the resistance to ~1 GΩ.

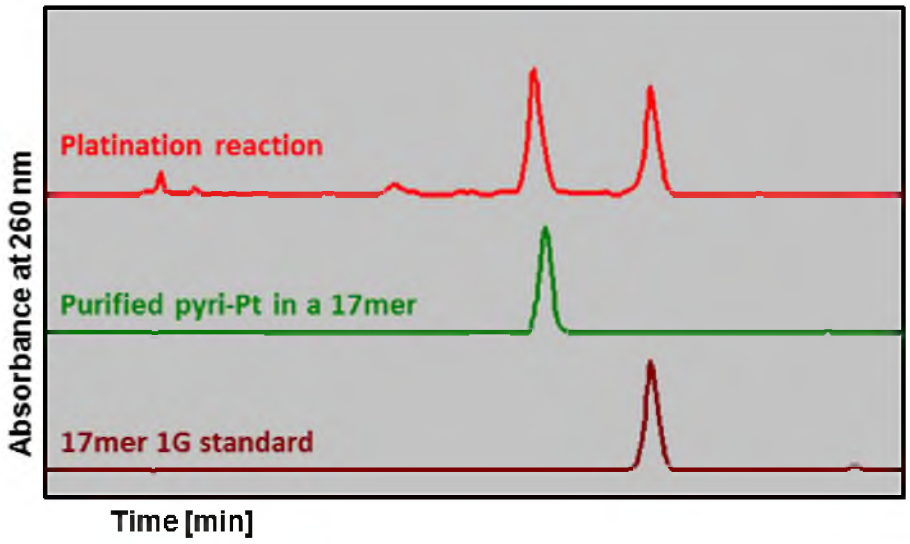
For solution mixtures, ultrapure water (> 18 MΩ·cm) was used; additionally, all electrolyte solutions were filtered using a sterile 0.22 mm Millipore vacuum filter. For the immobilization experiments, 1.00 M KCl, 1 mM EDTA, 25 mM Tris, pH 7.9 electrolyte solution was used and the data were collected under a 120 mV bias (*trans* vs. *cis*) with a 10 kHz filter, and data acquisition rate of 50 kHz. The Strep-Btn ODN

complex (Figure 3.2) was added at a final concentration of 200 nM and ~200 capture/release events (Figure 1.2, Chapter1) were recorded with platinated ODN alone, and after the addition of the undamaged strand and internal standard (polyC).[31,47] For the translocation experiments, 3.00 M NaCl, 10 mM PBS, 1 mM EDTA, pH 7.4 electrolyte solution was used and the data were collected under 100, 120 and 140 mV with a 100 kHz filter and data acquisition rate of 500 kHz. The concentration of the ODN inside the GNM was 10 μM and ~1,000 events were collected for each voltage. Placing the platinated ODN on the inside of the GNM greatly decreased the amount of material required for each experiment. The data were analyzed using software donated to our laboratory by EBS. Events were extracted using QUB 2.0.0.20 software.

Results and discussion

Immobilization experiment

Aquated platinum compounds bind DNA predominantly at the N7 position of G; therefore, to introduce the damage at a specific site, we designed ODNs that contained only one GG or G site where the platination could occur. For the nanopore studies a longer ODN needs to be used due to the fast translocation rate,[28] however, in order to characterize the cross-link a shorter 17-mer strand was used. The representative HPLC and MS analysis for the platination reaction is shown in Figure 3.3; the reactions for cisplatin, oxaliplatin and pyriplatin were consistent with the previously described results.[48,49] For the immobilization experiments, a 3'-biotinylated 40-mer poly-C homopolymer was used with the single G or double G placed at the most sensitive region of α-HL, position ω_{14} and $\omega_{14\&15}$. This nontraditional nomenclature for counting DNA



17mer	Calculated mass $[M+H]^+$	Experimental mass $[M+H]^+$
Standard	5014.3	5016.1
Pyriplatin	5322.4	5321.8

Figure 3.3. Representative HPLC traces and MS (MALDI-TOF) data for the platination reaction. A shorter ODN was used to obtain MS data, with a sequence of: 5'-TCTCCTCTTGTCTCCTC-3'. The HPLC mobile phase was B: 1.00 M NaCl, 25 mM Tris (pH 8), 10% ACN, pH 8, A: 10% ACN, running a method with a linear gradient starting at 20% B to 100% B over 30 min on an analytical ion-exchange HPLC column.

bases was used to identify that counting started from the 3'-biotinylated end. Next, ion channel recordings were initiated to define the current differences between the platinum-bearing strands and the canonical bases. First, the lipid bilayer was painted across the GNM, followed by insertion of wild-type α-HL into the suspended bilayer, leading to the open channel current.[45,46] Second, Strep-Btn DNA complex was added to the nanopore chamber, producing deep current level blockages (I) to the open channel current (I_o), caused by the complex being electrophoretically driven through the pore. Because streptavidin is significantly bigger than the opening to the α-HL, the complex was immobilized inside the ion channel until the polarity was reversed, allowing it to be driven back out, into the bulk solution. A typical current-time (i-t) trace is shown in Figure 1.2 in Chapter 1. For each immobilization experiment approximately 200 capture/release cycles were collected for statistical analysis and plotted into a current-level histogram.

The current blockage for the cisplatin, oxaliplatin and pyriplatin adducts, as well as the native strand and the internal standard (C_{40}) were then evaluated. The C_{40} internal standard current-blockage level was used as a reference and therefore assigned to a value of $\Delta\%I/I_o = 0$. Figure 3.4 shows a histogram of the percent blockage current level when the G and GG platinum adducts were placed at positions ω_{14} and $\omega_{14\&15}$, respectively. In this sequence context the difference between the unmodified GG-containing standard and the cisplatin-containing strand was $\Delta\%I/I_o = 5.2\%$, with the cisplatin adduct being the more blocking to the current level. In addition, this places the cisplatin adduct strand at a value of $\Delta\%I/I_o = 3.1\%$ more blocking than the internal standard, C_{40} (Figure 3.4). The oxaliplatin (309.27 g/mol, without the oxalate group) that is more bulky than cisplatin

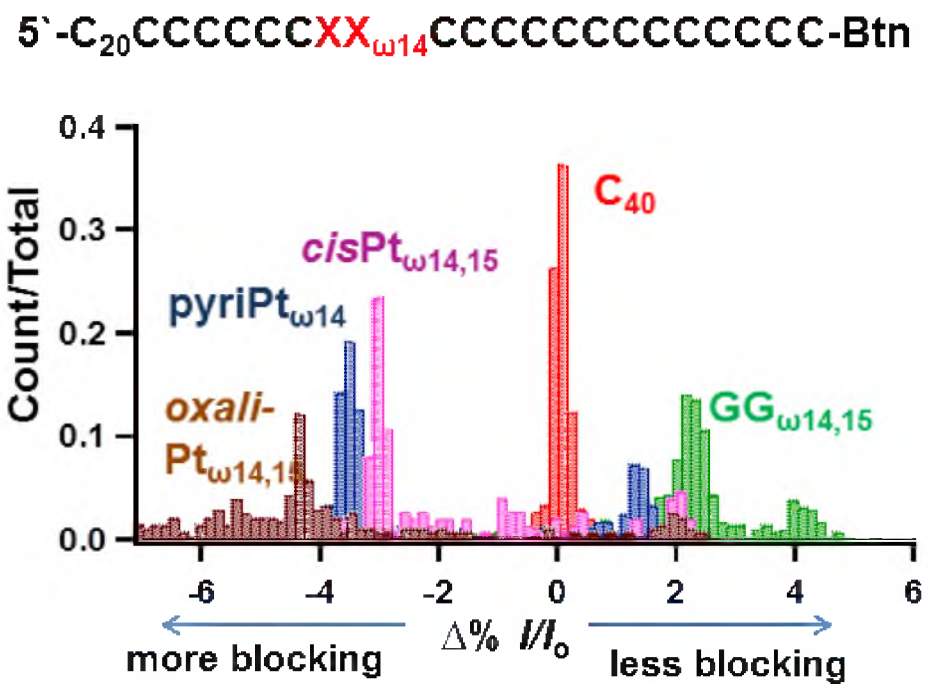


Figure 3.4. Current level histograms for the platinated ODNs. Histograms recorded in the polyC background (sequence shown on top) with cisplatin (*cis*Pt $\omega_{14\&15}$), and oxaliplatin (oxaliPt $\omega_{14\&15}$) positioned 14 and 15 bases counting from the 3' biotinylated end, and pyriplatin (pyriPt ω_{14}) positioned 14 bases from the 3' end.

adduct (229.10 g/mol, without the chlorides), but also bifunctional, increased the difference that was observed between the cross-link and the unreacted strand. Placement of the oxaliplatin in the α-HL ion channel, gave a $\Delta\%I/I_o$ = 6.4% more blocking current than the unmodified GG-containing strand, and $\Delta\%I/I_o$ = 4.3% more blocking current than the C_{40} reference (Figure 3.4). The final platinum adduct studied was pyriplatin (343.69 g/mol, without the chloride) a monofunctional platinum adduct that is also more bulky than cisplatin; however, it does not create a cross-link between two bases, and therefore does not introduce the rigidity that is observed with bifunctional adducts. Pyriplatin gave a $\Delta\%I/I_o$ = 3.5% more blocking than the internal standard, C_{40} (Figure 3.4) and $\Delta\%I/I_o$ = 4.8% more blocking than the unreacted strand containing a single G at position ω_{14} (the $\Delta\%I/I_o$ for the canonical bases at this position is further described by Schibel, *et al.*, and Stoddart, *et al.*)[39,47] These results demonstrate that the bulky platinum adducts are more blocking to the current level than the canonical bases, also, the blockage level appears to be dependent on the rigidity introduced to the strand. Pyriplatin was the adduct with the highest molecular weight (343.69 g/mol), however, it was less blocking than oxaliplatin (309.27 g/mol). Subsequently, the rigidity caused by the oxaliplatin linking two G bases has an effect on the current level blockage. Previous NMR and X-ray diffraction studies of the oxaliplatin adduct revealed that the platinum adduct severely unwinds and bends DNA double helix (global bend angle 31°).[42,50] It is anticipated that these distorting effects will also exist in ssDNA, especially in the sterically restricted β-barrel of α-HL. Pyriplatin, on the other hand, does not cause any significant distortion of the double helix as it is a monofunctional adduct that does not form a cross-link between the bases.[43] Although cisplatin was the least blocking adduct out of the platinum

compounds studied, it was significantly more blocking than any of the canonical bases. This is due to the fact that aside from being bulky (229.1 g/mol) it also greatly distorts the structure of the DNA double helix (global bend angle 80^{o}) and that will consequently affect the rigidity of the ssDNA.[43,51,52] Therefore, the significant increase in the current-blockage level measured for the platinated strands (cisplatin, pyriplatin, and oxaliplatin) is hypothesized to result from the increased rigidity (pyriplatin vs. oxaliplatin) and bulk of the adduct (cisplatin vs. oxaliplatin). Consistent with this hypothesis is a previous immobilization experiment suggesting that cross-linked bases are more blocking to the ion flow even without a significant change in the molecular weight like a thymine dimer (T=T) or 5'-G*CT*-3' cross-link (Chapter 2 and 4).[29,30]

Based on the immobilization experiments presented, cisplatin and oxaliplatin cross-links, as well as the pyriplatin adduct can be differentiated from the unmodified GG- or G-containing strands based on current-level variations (Figure 3.4). Furthermore, all of the platinated strands were considerably more blocking to the ion current flow than any of the canonical bases.[47] Due to the same C_{40} reference strand being used in all of our previous studies, it is also known that all the platinated strands are more blocking than other cross-links such as T=T (Chapter 4) and 5'-G*CT*-3' cross-link (Chapter 2). Also, the current level blockages for the platinated strands were higher than the oxidative damage-containing strands (such as 8-oxo-7,8-dihydro-2'deoxyguanosine (OG), spiroiminodihydantoin (Sp), or guanidinohydantoin (Gh)).[33]

Translocation experiment

The next step was to examine the behavior of the platinated strands in a translocation experiment, during which the DNA was allowed to freely pass through the protein channel from the *cis* to *trans* side of the pore, under -100, -120 and -140 mV bias (*trans* vs. *cis*). Figure 3.5A shows an *i-t* trace for a single molecule translocation event for an unmodified and the platinum-containing strand. The individual *i-t* traces were the same between the cisplatin-, oxaliplatin-, pyriplatin-containing strands and the unreacted standard, indicating no current signature for the platinated ODN. The individual *i-t* traces were then plotted into density plots that are shown in Figure 3.5B. For the unreacted strand the density plot reveals two populations of events with difference in the deep blockage current level (Figure 3.5B). Based on previous studies conducted in our laboratory,[30,32] the two event populations represent the higher blockage to the current (17 pA) to be the 3' entry, and the lower blockage to the current (24 pA) to be the 5' entry. For the platinated strands the directionality effect was not observed (as shown for pyriplatin in Figure 3.5B), and the distribution of events was much broader than for the standard. The data that were collected for the translocation experiments were then examined with respect to the time it took (t_D) for the strand to translocate from the *cis* side to the *trans* side of the pore. The t_D histogram for each population was plotted and fit with a Gaussian-like curve (diffusional broadening of the translocation times) with a mean peak value of t_{max}. The translocation time (t_{max}) was than analyzed for three different voltages to establish if the adducts were small enough to fit and translocate across the α-HL ion channel during free translocation (voltage dependence study). As shown in Figure 3.5C, as the voltage was increased (for all of the platinum adducts and

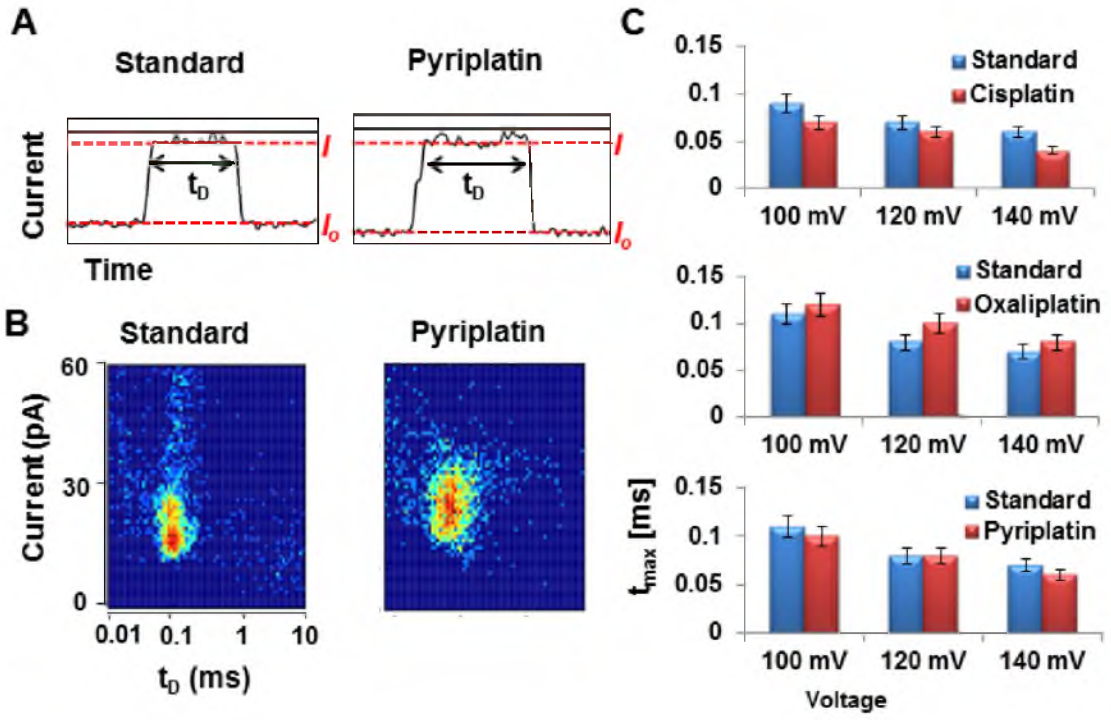


Figure 3.5. Translocation studies of the ODN containing cisplatin, oxaliplatin and pryiplatin adduct. **A.** Typical *i-t* traces of a single ODN molecule translocating through the pore. **B.** Density plots of the control and pyriplatin-containing strands at -120 mV (*trans* vs. *cis*) with the electrolyte 3.00 M NaCl, 10 mM PBS, 1 mM EDTA, pH 7.4. **C.** Voltage dependence study showing the t_{max} vs. voltage plot.

the unreacted strands) the time of the translocation decreased. This is indicative of the ODN being able to translocate from the *cis* to *trans* side of the protein ion channel.[53] The data presented for the translocation experiments suggest that cisplatin, oxaliplatin, and pryiplatin adducts are small enough to fit and translocate across α-HL and do not slow down the translocation speed.

Conclusion

The data presented suggest that the wild-type α-HL ion channel has the ability to detect platinum adducts such as cisplatin, oxaliplatin and pyriplatin in an immobilization experiment. These results are promising because they provide a unique current-level for these damages that is significantly different than the unmodified strand, native bases and many other naturally occurring cross-links (T=T, 5'-G*CT*-3') and base damages (OG, Sp, Gh). During the free translocation experiments the platinum adducts did not significantly differ from an unmodified strand. In order to be able to sequence the platinum cross-links during a free translocation experiment, higher base to base resolution is required, which may entail site-directed mutagenesis of α-HL. Also, because the platinum cross-links significantly bend the double-stranded DNA (cisplatin 80^{o} and oxaliplatin 31^{o}), a method might be developed to monitor adducts through detecting the bend in the DNA structure through unzipping experiments,[33,54] or using a bigger protein ion channel such as γ-HL.[55,56]

References

(1) Todd, R. C.; Lippard, S. J. Inhibition of transcription by platinum antitumor compounds. *Metallomics* **2009**, *1*, 280-291.

(2) Bloemink, M. J.; Reedijk, J. Cisplatin and derived anticancer drugs: mechanism and current status of DNA binding. *Met. Ions Biol. Syst.* **1996**, *32*, 641-685.

(3) Wang, D.; Lippard, S. J. Cellular processing of platinum anticancer drugs. *Nat. Rev. Drug Discov.* **2005**, *4*, 307-320.

(4) Ho, Y. P.; Au-Yeung, S. C.; To, K. K. Platinum-based anticancer agents: innovative design strategies and biological perspectives. *Med. Res. Rev.* **2003**, *23*, 633-655.

(5) Rosenberg, B.; Vancamp, L.; Krigas, T. Inhibition of cell division in *Escherichia coli* by electrolysis products from a platinum electrode. *Nature* **1965**, *205*, 698-699.

(6) Horwich, A.; Shipley, J.; Huddart, R. Testicular germ-cell cancer. *The Lancet* **2006**, *367*, 754-765.

(7) Loehrer, P. J.; Einhorn, L. H. Drugs five years later. Cisplatin. *Ann. Intern. Med.* **1984**, *100*, 704-713.

(8) Keys, H. M.; Bundy, B. N.; Stehman, F. B.; Muderspach, L. I.; Chafe, W. E.; Suggs, C. L., 3rd; Walker, J. L.; Gersell, D. Cisplatin, radiation, and adjuvant hysterectomy compared with radiation and adjuvant hysterectomy for bulky stage IB cervical carcinoma. *N. Engl. J. Med.* **1999**, *340*, 1154-1161.

(9) Morris, M.; Eifel, P. J.; Lu, J.; Grigsby, P. W.; Levenback, C.; Stevens, R. E.; Rotman, M.; Gershenson, D. M.; Mutch, D. G. Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aortic radiation for high-risk cervical cancer. *N. Engl. J. Med.* **1999**, *340*, 1137-1143.

(10) Muggia, F. Platinum compounds 30 years after the introduction of cisplatin: implications for the treatment of ovarian cancer. *Gynecol. Oncol.* **2009**, *112*, 275-281.

(11) Kartalou, M.; Essigmann, J. M. Mechanisms of resistance to cisplatin. *Mutat. Res.* **2001**, *478*, 23-43.

(12) Jung, Y.; Lippard, S. J. Direct cellular responses to platinum-induced DNA damage. *Chem. Rev.* **2007**, *107*, 1387-1407.

(13) Canetta, R.; Bragman, K.; Smaldone, L.; Rozencweig, M. Carboplatin: current status and future prospects. *Cancer Treatment Rev.* **1988**, *15, Supplement B*, 17-32.

(14) Raymond, E.; Faivre, S.; Chaney, S.; Woynarowski, J.; Cvitkovic, E. Cellular and molecular pharmacology of oxaliplatin. *Mol. Cancer Ther.* **2002**, *1*, 227-235.

(15) Kelland, L. R.; Abel, G.; McKeage, M. J.; Jones, M.; Goddard, P. M.; Valenti, M.; Murrer, B. A.; Harrap, K. R. Preclinical antitumor evaluation of bis-acetato-ammine-dichloro-cyclohexylamine platinum(IV): an orally active platinum drug. *Cancer Res.* **1993**, *53*, 2581-2586.

(16) Perez, J. M.; Fuertes, M. A.; Alonso, C.; Navarro-Ranninger, C. Current status of the development of trans-platinum antitumor drugs. *Crit. Rev. Oncol. Hematol.* **2000**, *35*, 109-120.

(17) Aris, S. M.; Farrell, N. P. Towards antitumor active trans-platinum compounds. *Eur. J. Inorg. Chem.* **2009**, *2009*, 1293-1302.

(18) Johnstone, T. C.; Park, G. Y.; Lippard, S. J. Understanding and improving platinum anticancer drugs-phenanthriplatin. *Anticancer Res.* **2014**, *34*, 471-476.

(19) Hollis, L. S.; Amundsen, A. R.; Stern, E. W. Chemical and biological properties of a new series of cis-diammineplatinum(II) antitumor agents containing three nitrogen donors: cis-$[Pt(NH_3)_2(N\text{-donor})Cl]^+$. *J. Med. Chem.* **1989**, *32*, 128-136.

(20) Ishida, S.; Lee, J.; Thiele, D. J.; Herskowitz, I. Uptake of the anticancer drug cisplatin mediated by the copper transporter Ctr1 in yeast and mammals. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14298-14302.

(21) Holzer, A. K.; Samimi, G.; Katano, K.; Naerdemann, W.; Lin, X.; Safaei, R.; Howell, S. B. The copper influx transporter human copper transport protein 1 regulates the uptake of cisplatin in human ovarian carcinoma cells. *Mol. Pharmacol.* **2004**, *66*, 817-823.

(22) Hermann, G.; Heffeter, P.; Falta, T.; Berger, W.; Hann, S.; Koellensperger, G. In vitro studies on cisplatin focusing on kinetic aspects of intracellular chemistry by LC-ICP-MS. *Metallomics* **2013**, *5*, 636-647.

(23) Jamieson, E. R.; Lippard, S. J. Structure, recognition, and processing of cisplatin-DNA adducts. *Chem. Rev.* **1999**, *99*, 2467-2498.

(24) Pinato, O.; Musetti, C.; Farrell, N. P.; Sissi, C. Platinum-based drugs and proteins: reactivity and relevance to DNA adduct formation. *J. Inorg. Biochem.* **2013**, *122*, 27-37.

(25) Terheggen, P. M.; Floot, B. G.; Scherer, E.; Begg, A. C.; Fichtinger-Schepman, A. M.; den Engelse, L. Immunocytochemical detection of interaction products

of cis-diamminedichloroplatinum(II) and cis-diammine(1,1-cyclobutanedicarboxylato) platinum(II) with DNA in rodent tissue sections. *Cancer Res.* **1987**, *47*, 6719-6725.

(26) Fichtinger-Schepman, A. M.; van der Veer, J. L.; den Hartog, J. H.; Lohman, P. H.; Reedijk, J. Adducts of the antitumor drug cis-diamminedichloroplatinum(II) with DNA: formation, identification, and quantitation. *Biochem.* **1985**, *24*, 707-713.

(27) Bayley, H. Sequencing single molecules of DNA. *Curr. Opin. Chem. Biol.* **2006**, *10*, 628-637.

(28) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; 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.

(29) Wolna, A. H.; Fleming, A. M.; An, N.; He, L.; White, H. S.; Burrows, C. J. Electrical current signatures of DNA base modifications in single molecules immobilized in the α-hemolysin ion channel. *Isr. J. Chem.* **2013**, *53*, 417-430.

(30) Wolna, A. H.; Fleming, A. M.; Burrows, C. J. Single-molecule detection of a guanine(C8)-thymine(N3) cross-link using ion channel recording. *J. Phys. Org. Chem.* **2014**, *27*, 247-251.

(31) An, N.; White, H. S.; Burrows, C. J. Modulation of the current signatures of DNA abasic site adducts in the alpha-hemolysin ion channel. *Chem. Commun. (Camb)* **2012**, *48*, 11410-11412.

(32) 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.

(33) Schibel, A. E.; 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.

(34) Larkin, J.; Carson, S.; Stoloff, D. H.; Wanunu, M. Nanopore-based analysis of chemically modified DNA and nucleic acid drug targets. *Isr. J. Chem.* **2013**, *53*, 431-441.

(35) Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **1996**, *274*, 1859-1866.

(36) 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.

(37) Howorka, S.; Bayley, H. Probing distance and electrical potential within a protein pore with tethered DNA. *Biophys. J.* **2002**, *83*, 3202-3210.

(38) Ayub, M.; Bayley, H. Individual RNA base recognition in immobilized oligonucleotides using a protein nanopore. *Nano Lett.* **2012**, *12*, 5637-5643.

(39) 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.

(40) Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S. Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein Sci.* **2006**, *15*, 459-467.

(41) Takahara, P. M.; Rosenzweig, A. C.; Frederick, C. A.; Lippard, S. J. Crystal structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin. *Nature* **1995**, *377*, 649-652.

(42) Wu, Y.; Pradhan, P.; Havener, J.; Boysen, G.; Swenberg, J. A.; Campbell, S. L.; Chaney, S. G. NMR solution structure of an oxaliplatin 1,2-d(GG) intrastrand cross-link in a DNA dodecamer duplex. *J. Mol. Biol.* **2004**, *341*, 1251-1269.

(43) Lovejoy, K. S.; Todd, R. C.; Zhang, S.; McCormick, M. S.; D'Aquino, J. A.; Reardon, J. T.; Sancar, A.; Giacomini, K. M.; Lippard, S. J. *cis*-Diammine(pyridine)chloroplatinum(II), a monofunctional platinum(II) antitumor agent: uptake, structure, function, and prospects. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8902-8907.

(44) Verstraete, S.; Heudi, O.; Cailleux, A.; Allain, P. Comparison of the reactivity of oxaliplatin, Pt(diaminocyclohexane)Cl_2 and Pt(diaminocyclohexane^1) $(OH_2)_2^{2+}$ with guanosine and L-methionine. *J. Inorg. Biochem.* **2001**, *84*, 129-135.

(45) 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 nanopore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **2007**, *79*, 4778-4787.

(46) 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.

(47) Schibel, A. E.; 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.

(48) Jangir, D. K.; Mehrotra, R. Raman spectroscopic evaluation of DNA adducts of a platinum containing anticancer drug. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2014**, *130*, 386-389.

(49) Ang, W. H.; Brown, W. W.; Lippard, S. J. Preparation of mammalian expression vectors incorporating site-specifically platinated-DNA lesions. *Bioconjug. Chem.* **2009**, *20*, 1058-1063.

(50) Spingler, B.; Whittington, D. A.; Lippard, S. J. 2.4 A crystal structure of an oxaliplatin 1,2-d(GpG) intrastrand cross-link in a DNA dodecamer duplex. *Inorg. Chem.* **2001**, *40*, 5596-5602.

(51) Yang, D.; van Boom, S. S.; Reedijk, J.; van Boom, J. H.; Wang, A. H. Structure and isomerization of an intrastrand cisplatin-cross-linked octamer DNA duplex by NMR analysis. *Biochem.* **1995**, *34*, 12912-12920.

(52) Gelasco, A.; Lippard, S. J. NMR solution structure of a DNA dodecamer duplex containing a *cis*-diammineplatinum(II) d(GpG) intrastrand cross-link, the major adduct of the anticancer drug cisplatin. *Biochem.* **1998**, *37*, 9230-9239.

(53) Meller, A.; Nivon, L.; Branton, D. Voltage-driven DNA translocations through a nanopore. *Phys. Rev. Lett.* **2001**, *86*, 3435-3438.

(54) Jin, Q.; Fleming, A. M.; Ding, Y.; Burrows, C. J.; White, H. S. Structural destabilization of DNA duplexes containing single-base lesions investigated by nanopore measurements. *Biochem.* **2013**, *52*, 7870-7877.

(55) Menestrina, G.; Dalla Serra, M.; Comai, M.; Coraiola, M.; Viero, G.; Werner, S.; Colin, D. A.; Monteil, H.; Prevost, G. Ion channels and bacterial infection: the case of beta-barrel pore-forming protein toxins of *Staphylococcus Aureus*. *FEBS Lett.* **2003**, *552*, 54-60.

(56) Yamashita, K.; Kawai, Y.; Tanaka, Y.; Hirano, N.; Kaneko, J.; Tomita, N.; Ohta, M.; Kamio, Y.; Yao, M.; Tanaka, I. Crystal structure of the octameric pore of staphylococcal gamma-hemolysin reveals the beta-barrel pore formation mechanism by two components. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17314-17319.

CHAPTER 4

SINGLE-MOLECULE ANALYSIS OF THYMINE-DIMER CONTAINING OLIGODEOXYNUCLEOTIDES USING NANOPORE TECHNOLOGY

Introduction

In recent years, ozone layer depletion has caused increased exposure to the sun's harmful UVB and UVC radiation which are known to cause direct photoinduced DNA damage. There is also evidence that UVA light (~95% of the UV solar irradiation reaching the earth, as well as many tanning lamps) causes direct and indirect photoinduced DNA damage.[1,2] The predominant photoproducts from UV irradiation are interstrand and intrastrand cyclobutane-pyrimidine dimers (CPDs), and more specifically, thymine dimers (T=T), and to a much lesser extent 6-4 photoproducts and their Dewar valence isomers, that are formed when two adjacent thymines (TT) are photoirradiated (Figure 4.1).[3,4] Unsurprisingly, the T=T yield is highest in skin cells exposed to UV light, for which this form of DNA damage has been strongly correlated with skin cancer.[5,6] A single day spent in the sun can introduce up to 100,000 UV photoproducts per cell in the epidermis.[7] UV irradiation causes damage that, if left unrepaired results in stalling of DNA polymerases[8] and subsequently leads to the development of mutations and tumors. The cell's ability to repair DNA cross-links is crucial for its survival. It has

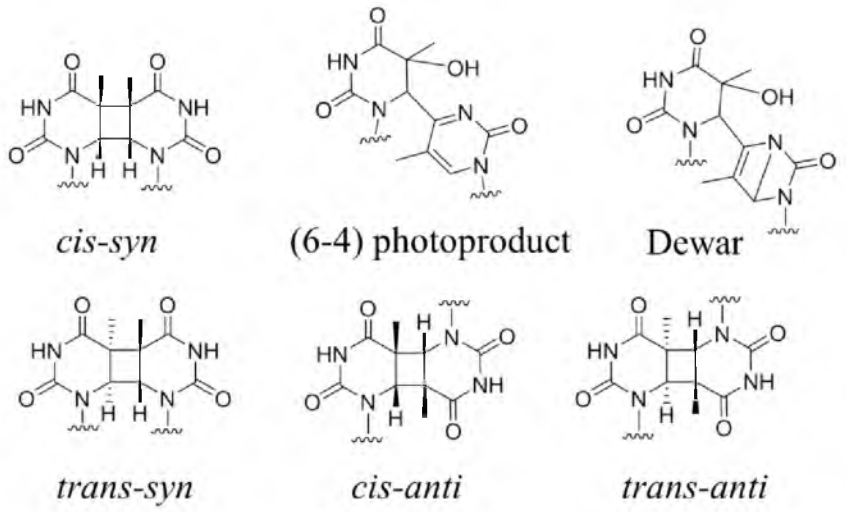


Figure 4.1. Structures of the major thymine dimer (T=T) photoproducts.

been reported that even a single interstrand cross-link left unrepaired can cause death in bacteria or yeast cells.[9,10] Therefore, cells have evolved elaborate pathways to repair DNA cross-links. One of the most efficient CPD repair mechanisms is direct photoreactivation by photolyase.[11-14] Cross-links and other DNA damages are also repaired by the nucleotide excision repair (NER),[15,16] base excision repair (BER) or mismatch repair (MMR) pathways.[17] If these repair mechanisms fail, hereditary diseases such as xeroderma pigmentosum may occur.[18,19] Other nonhereditary diseases, such as melanoma and nonmelanoma skin cancer can also occur, and are thought to be the most common types of cancers worldwide.[5] Due to the frequency of occurrence of these types of diseases, there is a need to develop rapid screening methods for early detection of the cross-links associated with these cancers.

In order to map or identify the DNA damage and cross-link sites, many different methods have been developed, including: the comet assay,[20] PCR-based methods[21] or HPLC-electrospray tandem mass spectrometry.[22,23] Nevertheless, many of these methods cannot specifically identify the cross-link or do not provide sequence information. For instance, the comet assay can detect the location of damage, but it cannot provide damage identity or multiple damage sites. Analysis by HPLC-electrospray tandem mass spectrometry is the only way to distinguish all of the CPD isomers;[24] however, this method is not cost efficient, and requires prior DNA digestion or glycosidic bond hydrolysis, and does not provide damage location or sequence information. All of the currently available DNA cross-link detection methods lack one or more of the following attributes: single-molecule detection, sequencing information, multiple-cross-link

detection on an individual DNA strand, high sensitivity, specificity, low cost, or mild DNA handling conditions.

Electrochemical methods for DNA damage detection, namely the alpha-hemolysin (α-HL) ion channel nanopore,[25,26] have been proposed as a possibly low-cost and high-speed DNA analysis tool.[27] Assembled α-HL creates a water-filled channel that is composed of a vestibule with an opening of about 2.8 nm and a lipid-embedded stem with an interior constriction zone of about 1.4 nm in diameter (Figure 1.1, Chapter 1).[28] The interior dimension of the pore allows only a single-strand DNA (ssDNA) to translocate across the pore,[29] which makes α-HL a suitable tool for nucleic acid analysis. The translocation of ssDNA is driven by an electrophoretic force. Once an electric potential is applied, the background electrolyte (KCl) creates an open-channel current that is recorded. When a ssDNA is occupying the α-HL ion channel, the open-channel current will be reduced to a blocking current. In this study, the blocking current level of a *cis-syn* T=T was measured by an immobilization experiment (Figure 4.2)[30] and the time of translocation was measured in subsequent experiments.

Experimental

Oligodeoxynucleotide preparation

Oligodeoxynucleotides (ODNs) containing the *cis-syn* T=T were synthesized at the DNA/Peptide Core Facility at the University of Utah. The ODNs were deprotected using previously established techniques,[31] avoiding any light exposure to ensure that the T=T did not degrade back to TT. Next, the ODNs were purified using an ion-exchange HPLC on a Dionex DNA Pac PA-100 column with a method that consisted of a linear

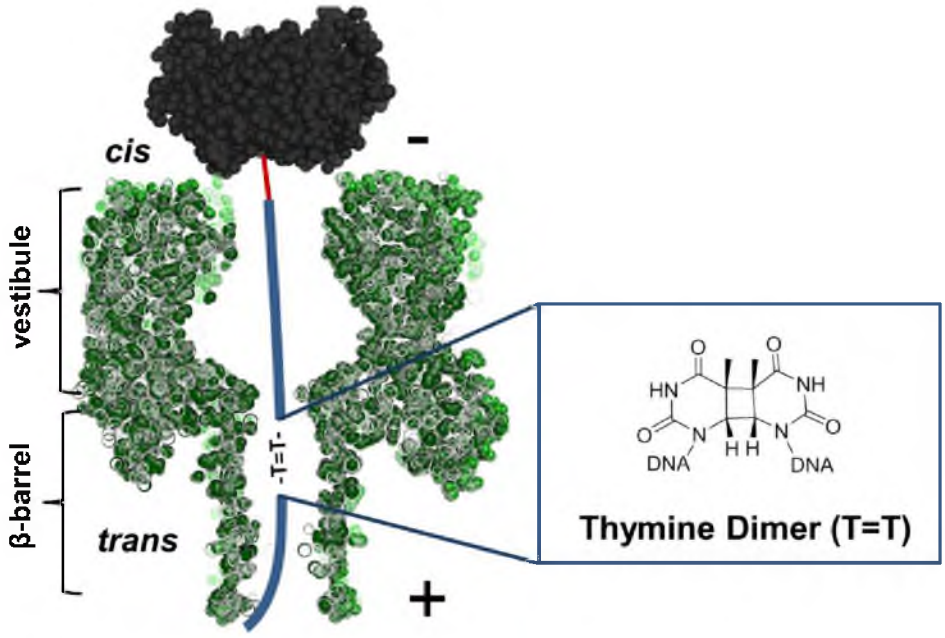


Figure 4.2. Representation of α-HL (pdb 7AHL)[28] /streptavidin (pdb 1MK5)[32] Btn-DNA complex that contains a thymine dimer.

gradient from 5% to 40% B over 30 min, while monitoring the absorbance at 260 nm. The mobile phases were: A: 10% MeCN and 90% ddH_2O, B: 1 M NaCl, 25 mM Tris,10% MeCN, pH 8. Dialysis was used to remove excess salts from the purification. The purity was confirmed by reinjecting the sample onto an analytical HPLC running the previously described method. Through all the experimental procedures, the ODNs containing the T=T were handled with caution to prevent any light exposure.

Nanopore experiments

Ultrapure water (>18 MΩ·cm) was used to prevent any contamination, and electrolyte solutions were filtered using a sterile 0.22 μm Millipore vacuum filter. The electrolyte solution used for the nanopore experiments consisted of 1.00 M KCl, 1 mM EDTA, 25 mM Tris, pH 7.9 for the immobilization experiments and 1.00 M KCl, 10 mM NaP_i (pH 7.4), 1 mM EDTA for the translocation experiments. The surface of the glass nanopore membrane (GNM) was modified with 3-cyanopropyldimethylchlorosilane before the measurements.[33,34] The phospholipid used to form the lipid bilayer was 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC). The nanopore setup was as follows: (1) a Ag/AgCl electrode was positioned inside the GNM (filled with the electrolyte solution), which was attached to a pressure gauge and a 10-mL gastight syringe. (2) Another Ag/AgCl electrode was placed in the nanopore holder chamber that was then filled with electrolyte solution. The instrument was donated by the Electronic BioSciences (EBS). Once the lipid bilayer was painted across the GNM, the protein channel was inserted into the lipid bilayer by applying a pressure to the GNM, resulting in the protein insertion associated with an increased conductance. For the immobilization

experiments 200 nM of the streptavidin-ODN complex was added to the chamber. For each experiment roughly 200 capture-release cycles (Figure 1.2, Chapter 1) were collected, followed by the addition of the same amount of the internal polyC standard. For the translocation experiments 5 μM DNA solution of ssDNA was used and three voltages were studied: 120, 140 and 160 mV. For each voltage roughly 1000 events were collected. The data were collected using a 10 kHz filter and 50 kHz acquisition rate for the immobilization experiments and a 100 kHz filter and 500 kHz data acquisition rate for the translocation experiments. For the purposes of the presentation, the translocation data were then re-filtered to 10 kHz. The data were extracted using QUB 2.0.0.20. OriginPro 9.1 and Event States Analysis (EBS software) was used for statistical analysis of the events.

Results and discussion

Immobilization experiments

It has been determined that the most sensitive detection zone of the nanopore β-barrel is at position 14 relative to the 3' terminus (ω14) of the DNA when it is immobilized in the α-HL channel using 3' biotinylation (Btn).[35,36] In order to better understand how the thymine dimer influences the current blockage level in α-HL, immobilization experiments were conducted in which a biotinylated 40-mer poly-dC with a *cis-syn* T=T at positions ω14-15 was analyzed. To be consistent with the previous results, a poly-C background sequence was used.[35,37,38] To immobilize the ODN at the specific position in the nanopore, streptavidin was added to bind the biotin that was

incorporated at the 3' end. Once the 5' end enters the channel it threads through until it encounters the streptavidin, which is too big to enter the pore (Figure 4.2).

In Figure 4.3 the structure of *cis-syn* T=T is shown together with its blockage current level histogram in an immobilization experiment. The relative positions of C_{40} and $C_{38}TT_{\omega 14,15}$ standards are consistent with the literature values.[35,39] The T=T lesion blocks the current ~1% more than a DNA containing an undamaged TT sequence at positions ω14, 15 (Figure 4.3A) making it feasible to distinguish the damage from the canonical nucleotides that block less than C_{40}. We suspect that the more blocking current level of the T=T-containing DNA is due to the increased rigidity of the DNA at this point.[3] The T=T lesion in a sequencing experiment will likely give a distinct current level for interpretation, although, this lesion stalls polymerases, and will be problematic for long reads using molecular motors relying on complementary strand synthesis for controlling the translocation speed.

The data presented in Figure 4.3A show that the TT and T=T sequences can readily be resolved in an immobilization experiment, therefore a further experiment was conducted to monitor the photoreversal of T=T back to TT. In Figure 4.3B, the photoreversal[40] of the T=T with 254 nm light was plotted from data collected in a time-dependent nanopore immobilization experiment. Because the photoreversal experiment was conducted in the presence of the lipid, α-HL and streptavidin (in the nanopore chamber), the reversal rate was considerably slower than predicted,[41,42] due to the absorbance of UV light by the added proteins and lipid. This experiment illustrates an example in which the α-HL nanopore can be used to monitor a chemical reaction; another example, which was demonstrated by Bayley's laboratory, was the determination of the

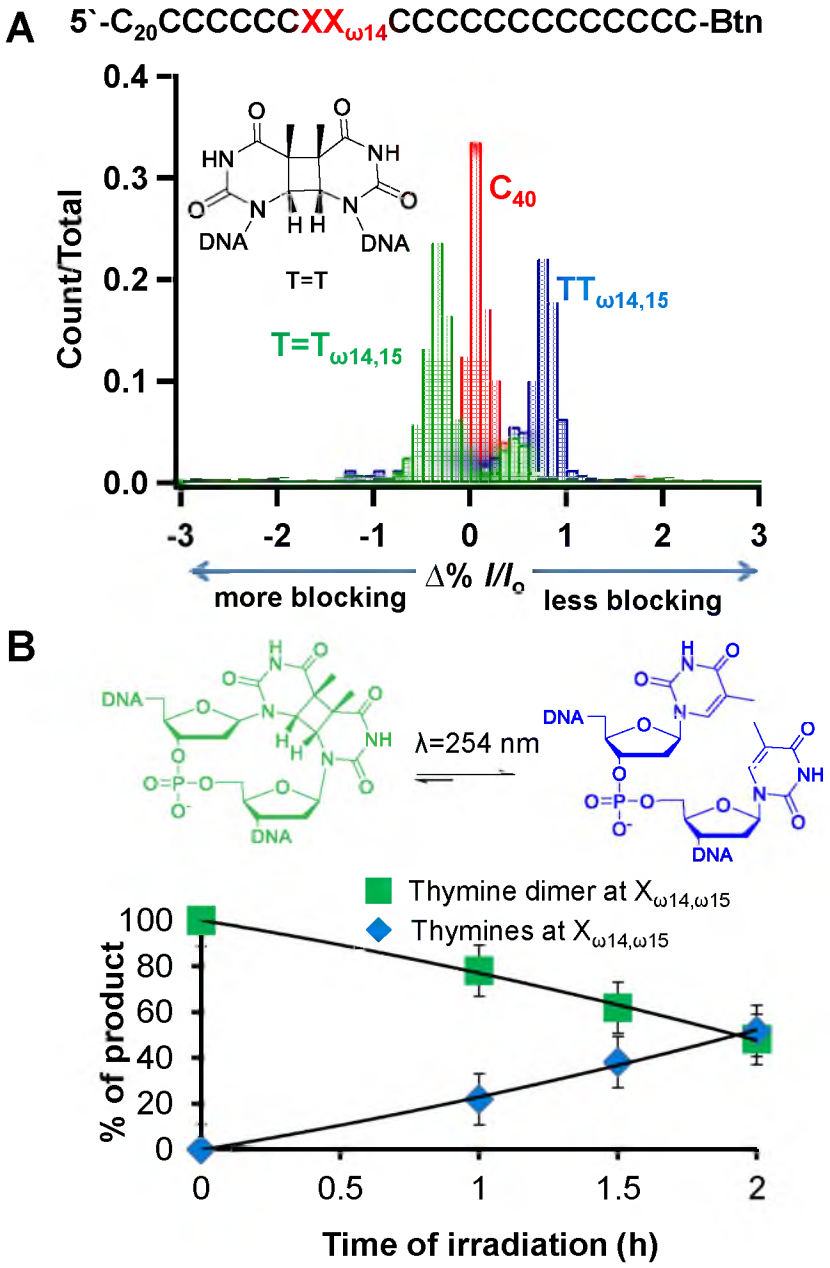


Figure 4.3. Current level histograms for thymine dimer containing DNA. **A.** $\Delta\%I/I_o$ histograms representing the current level blockage of TT and T=T at a position ω14 and 15. **B**. Photoreversal experiment. The starting material was 3` biotinylated C_{38} with a thymine dimer at positions ω14 and 15 (green square) that was irradiated with UV light (254 nm) to convert the DNA damage back to thymines at those positions. The irradiation experiment was conducted within the nanopore cell holder with streptavidin, α-HL and lipid present in 1.00 M KCl, 25 mM Tris-HCl, and 1.0 mM EDTA (pH 7.9). The nanopore experiments were conducted after irradiation for 0.5, 1, 1.5 and 2 h.

base pK_a values for G and T.[43]

Translocation experiments

Next, an ODN bearing a synthetic *cis-syn* T=T was designed for conducting translocation experiments. The representative *i-t* trace of a translocation event and the voltage dependence study are shown in Figure 4.4. The two ODNs, a 32-mer standard and 32-mer T=T dimer containing strands (sequence shown in Figure 4.4A), were shown to translocate across the pore with the time of the translocation decreasing as the voltage increases. Unfortunately, there was not a distinguishable difference in the current levels, or translocation times. These data can be interpreted to suggest that although the T=T increases the rigidity of the ssDNA,[44] it does not cause enough obstruction to the current flow to change the *i-t* signature during the free translocation experiment, which can be also explained by the fact that there is no mass increase from the undamaged strand to the T=T containing strand. Similar results were obtained during the translocation experiment for the much bigger 5'-G*CT*-3' cross-link (Chapter 2)[45] and the platinum adducts (Chapter 3). The combination of these results highlights the difficulty in detecting DNA damage via free translocation experiments using wild type α-HL.

Future direction

Developing methods that allow DNA damage interrogation in the sequence context is of great interest; therefore, future studies will focus on developing a better ion channel that would allow base-to-base resolution. Site-directed mutagenesis can be used to decrease the constriction zone to a smaller size that may lead to better sensitivity for

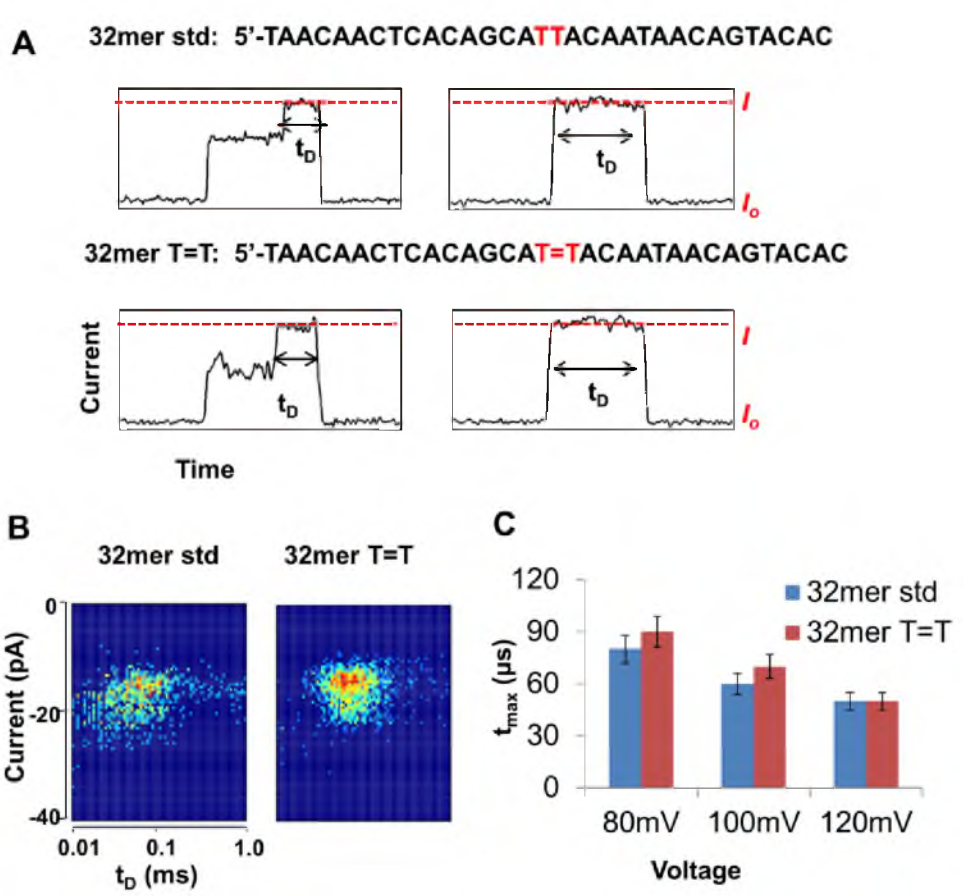


Figure 4.4. Translocation studies of the T=T-containing 32mer. **A.** Typical *i-t* trace of a single DNA molecule translocating through the pore. **B.** Density plots of the control and T=T-containing strands at 120 mV (*trans* vs. *cis*). **C.** Voltage dependence study showing the t_{max} vs. voltage.

the canonical bases, as well as the damaged bases, such as T=Ts, 5'-G*CT* cross-links and platinum adducts. Also, instead of direct sequencing of the damaged base, a marker may be introduced using BER enzymes. For example, T=T might be treated with EndoV to make a strand break and then a marker base might be inserted opposite the abasic site.[46-48] The method to incorporate nonnative bases opposite the abasic site is currently being developed in our laboratory.

Conclusions

In conclusion, the studies described herein show that the α-HL ion channel can distinguish between the UV-induced T=T and the undamaged strand in the immobilization experiment. The free translocation of the T=T-containing strands did not show a significant difference in the translocation time or current level from the undamaged ODN. The immobilization results are promising, because similarly to the case of 5'- G*CT*-3' cross-links (Chapter 2) and platinum adducts (Chapter 3), the T=T gives a unique current level that can be distinguished from the undamaged strand and the canonical bases. Additionally, due to the current difference between undamaged and damage-containing strands, a photoreversal of the T=T back to TT could be observed using the nanopore set up. Currently, there are no known hot spots for UV-induced damage; therefore, once perfected, nanopore sequencing of damaged DNA could provide insights into understanding the progression of skin cancer and related diseases resulting from DNA damage.

References

(1) Mouret, S.; Baudouin, C.; Charveron, M.; Favier, A.; Cadet, J.; Douki, T. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13765-13770.

(2) Jiang, Y.; Rabbi, M.; Kim, M.; Ke, C.; Lee, W.; Clark, R. L.; Mieczkowski, P. A.; Marszalek, P. E. UVA generates pyrimidine dimers in DNA directly. *Biophys. J.* **2009**, *96*, 1151-1158.

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

(4) 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.

(5) Norval, M.; Lucas, R. M.; Cullen, A. P.; de Gruijl, F. R.; Longstreth, J.; Takizawa, Y.; van der Leun, J. C. The human health effects of ozone depletion and interactions with climate change. *Photochem. Photobiol. Sci.* **2011**, *10*, 199-225.

(6) Ikehata, H.; Ono, T. The mechanisms of UV mutagenesis. *J. Radiat. Res.* **2011**, *52*, 115-125.

(7) Garinis, G. A.; van der Horst, G. T. J.; 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.

(8) Cordonnier, A. M.; Fuchs, R. P. P. Replication of damaged DNA: molecular defect in Xeroderma pigmentosum variant cells. *Mutat. Res.* **1999**, *435*, 111-119.

(9) McVey, M. Strategies for DNA interstrand crosslink repair: insights from worms, flies, frogs, and slime molds. *Environ. Mol. Mutagen.* **2010**, *51*, 646-658.

(10) McHugh, P. J.; Spanswick, V. J.; Hartley, J. A. Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. *Lancet. Oncol.* **2001**, *2*, 483-490.

(11) Komori, H.; Masui, R.; Kuramitsu, S.; Yokoyama, S.; Shibata, T.; Inoue, Y.; Miki, K. Crystal structure of thermostable DNA photolyase: pyrimidine-dimer recognition mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13560-13565.

(12) Lima-Bessa, K. M. d.; Armelini, M. G.; Chigançcas, V.; Jacysyn, J. F.; Amarante-Mendes, G. P.; Sarasin, A.; Menck, C. F. M. CPDs and 6-4PPs play different

roles in UV-induced cell death in normal and NER-deficient human cells. *DNA Repair* **2008**, *7*, 303-312.

(13) Sancar, A. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem. Rev.* **2003**, *103*, 2203-2238.

(14) Essen, L. O.; Klar, T. Light-driven DNA repair by photolyases. *Cell. Mol. Life Sci.* **2006**, *63*, 1266-1277.

(15) Costa, R. M. A.; Chiganças, V.; da Silva Galhardo, R.; Carvalho, H.; Menck, C. F. M. The eukaryotic nucleotide excision repair pathway. *Biochimie* **2003**, *85*, 1083-1099.

(16) Marteijn, J. A.; Lans, H.; Vermeulen, W.; Hoeijmakers, J. H. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 465-481.

(17) Hoffman, P. D.; Wang, H.; Lawrence, C. W.; Iwai, S.; Hanaoka, F.; Hays, J. B. Binding of MutS mismatch repair protein to DNA containing UV photoproducts, "mismatched" opposite Watson-Crick and novel nucleotides, in different DNA sequence contexts. *DNA Repair* **2005**, *4*, 983-993.

(18) van Steeg, H.; Kraemer, K. H. Xeroderma pigmentosum and the role of UV-induced DNA damage in skin cancer. *Mol. Med. Today* **1999**, *5*, 86-94.

(19) Menck, C. F.; Munford, V. DNA repair diseases: what do they tell us about cancer and aging? *Genet. Mol. Biol.* **2014**, *37*, 220-233.

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

(21) Rochette, P. J.; Bastien, N.; Todo, T.; Drouin, R. Pyrimidine (6-4) pyrimidone photoproduct mapping after sublethal UVC doses: nucleotide resolution using terminal transferase-dependent PCR. *Photochem. Photobiol.* **2006**, *82*, 1370-1376.

(22) Douki, T.; Court, M.; Sauvaigo, S.; Odin, F.; Cadet, J. Formation of the main UV-induced thymine dimeric lesions within isolated and cellular DNA as measured by high performance liquid chromatography-tandem mass spectrometry. *J. Biol. Chem.* **2000**, *275*, 11678-11685.

(23) Kumari, S.; Rastogi, R. P.; Singh, K. L.; Singh, S. P.; Sinha, R. P. DNA damage: detection strategies. *EXCLI J.* **2008**, *7*, 44-62.

(24) Su, D. G.; Kao, J. L.; Gross, M. L.; Taylor, J. S. Structure determination of an interstrand-type cis-anti cyclobutane thymine dimer produced in high yield by UVB

light in an oligodeoxynucleotide at acidic pH. *J. Am. Chem. Soc.* **2008**, *130*, 11328-11337.

(25) Ervin, E. N.; Kawano, R.; White, R. J.; White, H. S. Simultaneous alternating and direct current readout of protein ion channel blocking events using glass nanopore membranes. *Anal. Chem.* **2008**, *80*, 2069-2076.

(26) 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.

(27) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; 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.

(28) Song, L.; Haobaugh, M. R.; Shustak C.; Cheley, S.; Bayley, H.; Gouaux, E. J. Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **1996**, *274*, 1859-1866.

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

(30) Wolna, A. H.; Fleming, A. M.; An, N.; He, L.; White, H. S.; Burrows, C. J. Electrical current signatures of DNA base modifications in single molecules immobilized in the α-hemolysin ion channel. *Isr. J. Chem.* **2013**, *53*, 417-430.

(31) Nguyen, K. V.; Burrows, C. J. A prebiotic role for 8-oxoguanosine as a flavin mimic in pyrimidine dimer photorepair. *J. Am. Chem. Soc.* **2011**, *133*, 14586-14589.

(32) Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S. Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein Sci.* **2006**, *15*, 459-467.

(33) 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 nanopore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **2007**, *79*, 4778-4787.

(34) 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.

(35) 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.

(36) 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.

(37) An, N.; White, H. S.; Burrows, C. J. Modulation of the current signatures of DNA abasic site adducts in the alpha-hemolysin ion channel. *Chem. Commun.* **2012**, *48*, 11410-11412.

(38) 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.

(39) Mouret, S.; Baudouin, C.; Charveron, M.; Favier, A.; Cadet, J.; Douki, T. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proc. Natl. Acad. Sci.* **2006**, *103*, 13765-13770.

(40) Pan, Z.; Chen, J.; Schreier, W. J.; Kohler, B.; Lewis, F. D. Thymine dimer photoreversal in purine-containing trinucleotides. *J. Phys. Chem. B* **2011**, *116*, 698-704.

(41) Johns, H. E.; Rapaport, S. A.; Delbrueck, M. Photochemistry of thymine dimers. *J. Mol. Biol.* **1962**, *4*, 104-114.

(42) Garcés, F.; Dávila, C. A. Alterations in DNA irradiated with ultraviolet radiation - I. The formation process of cyclobutylpyrimidine dimers: cross sections, action spectra and quantum yields. *Photochem. Photobiol.* **1982**, *35*, 9-16.

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

(44) Park, H.; Zhang, K.; Ren, Y.; Nadji, S.; Sinha, N.; Taylor, J. S.; Kang, C. Crystal structure of a DNA decamer containing a *cis-syn* thymine dimer. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15965-15970.

(45) Wolna, A. H.; Fleming, A. M.; Burrows, C. J. Single-molecule detection of a guanine(C8)-thymine(N3) cross-link using ion channel recording. *J. Phys. Org. Chem.* **2014**, *27*, 247-251.

(46) Yamashige, R.; Kimoto, M.; Takezawa, Y.; Sato, A.; Mitsui, T.; Yokoyama, S.; Hirao, I. Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res.* **2012**, *40*, 2793-806.

(47) Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E. Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12005-10.

(48) Kimoto, M.; Cox, R. S., 3rd; Hirao, I. Unnatural base pair systems for sensing and diagnostic applications. *Expert Rev. Mol. Diagn.* **2011**, *11*, 321-31.

CHAPTER 5

SINGLE-MOLECULE ANALYSIS OF THYMINE-DIMER CONTAINING G-QUADRUPLEXES FORMED FROM THE HUMAN TELOMERE SEQUENCE

Introduction

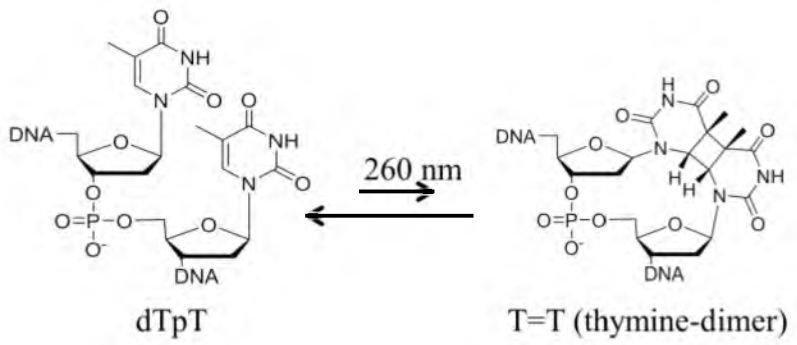
Telomeres are DNA-protein complexes that are essential to maintain genome integrity.[1] The human telomere DNA consists of a 5'-$(TTAGGG)_n$-3' repetitive sequence that is 5,000 to 20,000 base pairs long with a 3' single-stranded overhang of 100 to 250 nucleotides.[2,3] The guanine (G) rich single-stranded DNA (ssDNA) can form G-quadruplex structures that were recently confirmed in human cells.[4] The building unit of a G-quadruplex is a G-quartet that consists of four planar Gs that are hydrogen-bonded through the Hoogsteen face.[5] These alternative DNA-secondary structures are proposed to play an important role in DNA recombination, transcription and replication.[6-9] The ends of chromosomes are capped by telomeres to prevent fusion of the uncapped end of a chromosome with another telomere or a double-strand break, and to prevent nucleolytic resection.[10,11] Understanding the secondary structure of G-quadruplexes is a key step to addressing their role in biological processes.

G-quadruplexes can assume different structures that are dependent upon the coordinated alkali metal ion (K^+ or Na^+), and the physical conditions of the solution. In

KCl solution, the human telomere sequence adopts two interchangeable structures (termed hybrid 1 and hybrid 2) with two edgewise loops and one double-chain reversal loop.[12-14] The main difference between these hybrid folds is the order of the loops, for which hybrid 1 has the double-chain reversal loop closest to the 5' end and hybrid 2 has it closest to the 3' end. The basket fold occurs in NaCl solution and it contains two edgewise loops and one central diagonal loop.[15] A third topology has been characterized and termed the propeller fold; it has been observed in highly dehydrating solutions or in solid state experiments.[16-18] The propeller fold is formed in the presence of K^+ and contains only double-chain reversal loops.[16,17,19] Based on the higher concentration of K^+ inside the cell,[20] and the higher G-quadruplex binding constant to K^+ over Na^+, the hybrid folds are the G-quadruplex conformations that were proposed to exist *in vivo*.[13,21,22] The folding pattern that the human telomere sequence assumes is crucial for its proper function in the cell.[23,24]

The telomere sequence is hypersensitive to oxidative and UV-induced DNA damage that can alter its folding pattern.[25-27] Oxidative stress is a known contributor to telomere shortening, which is directly associated with cell mortality and aging.[28-30] Guanine is the most oxidation-prone DNA base due to its low redox potential; [31,32] consequently, the G-rich telomere sequence is hypersensitive to oxidative DNA damage.[26,33] The majority of the G-oxidation products were found to be present in the human telomere sequence.[33] Surprisingly, even though the damage disturbs the core of G-quadruplex structure (G-tetrads), it still can maintain some degree of secondary structure, depending on the location of the damage.[33-35] In contrast to the core damage, DNA lesions found in the loops of G-quadruplex folds are well tolerated. Thymine

glycol and 8-oxo-7,8-dihydrodeoxyadenosine (8-oxo-A) present in the loops of hybrid and basket G-quadruplexes did not prevent folding and gave similar structures.[34,36] Interestingly, the stabilities of these structures were dependent on the nature of the damage.[34,36] Another type of persistent DNA damage is UV-induced photo-crosslinking of two adjacent thymine (T) residues resulting in the formation of thymine-dimers (T=T, Scheme 5.1).[37] The natural human telomere sequence is perfectly poised for T=T formation. This damage is readily formed in the epidermis and correlates with skin cancer;[38] in addition, model studies of UV damage to cells identified the telomere to be hypersensitive to yield T=Ts.[25] Interestingly, T=Ts in the telomeres were shown to not increase shortening, and these cells did not show increased rates of apoptosis.[25] These results suggest that T=Ts are not detrimental to telomere structure and function. Despite the fact that the telomere region is hypersensitive to UV light induced DNA damage,[25] few studies have been conducted on the effects of T=T in G-quadruplexes. An *in vitro* study conducted by Taylor's laboratory showed that the hybrid and basket folds can be UV-irradiated to form T=Ts.[39,40] In the present work, we investigated the stability and structure of T=T-containing G-quadruplexes to address the question how UV-induced damage affects human telomere G-quadruplex structure. These studies included established methods of analysis (circular dichroism, electrophoretic mobility shift assay, and thermal melting), as well as a new single-molecule method developed in our laboratory.[41,42] In this method, a single alpha-hemolysin (α-HL) ion channel was used to capture the different G-quadruplexes in the vestibule under an electrophoretic force. During the capture event, the recorded ion current gives characteristic patterns based on the G-quadruplex fold that was captured in the vestibule (Figure 5.1).



Scheme 5.1. Two adjacent Ts being converted to a thymine-dimer (T=T) upon UV irradiation.

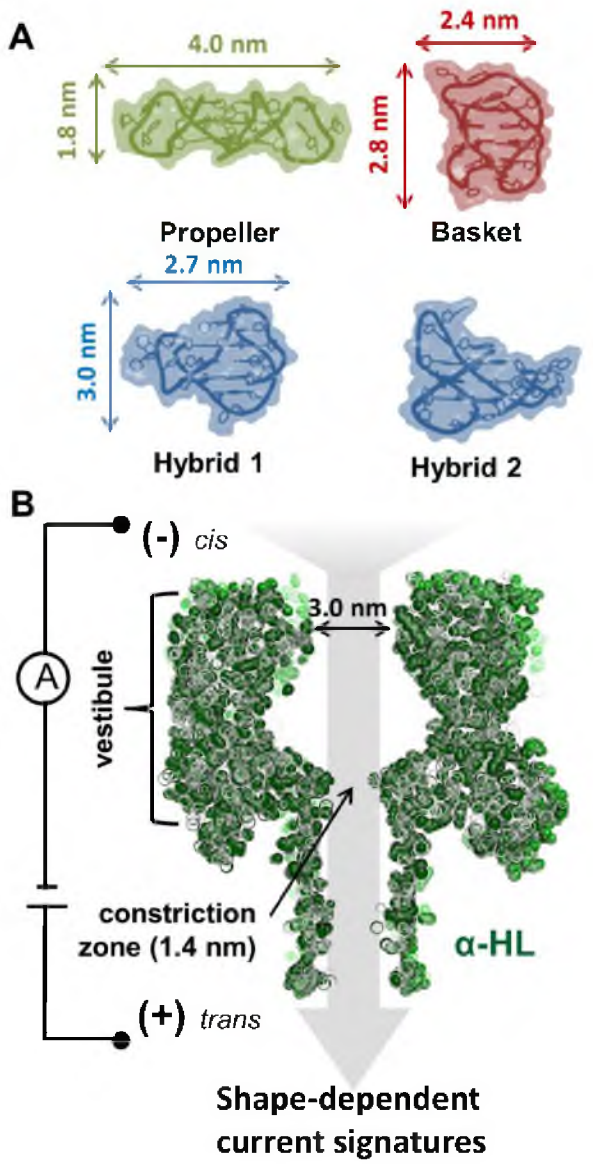


Figure 5.1. Nanopore measurements of G-quadruplexes in the α-hemolysin (α-HL) nanopore. **A.** G-quadruplex folds: propeller (pdb 1KF1),[16] basket (pdb 143D),[15] hybrid type 1 (pdb 2JSK)[12] and type 2 (pdb 2JSQ).[12] **B.** Experimental setup: the α-HL (pdb 7AHL)[43] was first assembled into a lipid bilayer, then a voltage was applied across the channel and current levels were measured.

Materials and methods

DNA preparation and purification

ODNs were synthesized at the DNA/Peptide Core Facility at the University of Utah using commercially available phosphoramidites (Glen Research). The ODNs were then deprotected[44] and HPLC purified following standard protocols.

CD and T_m study

To induce the hybrid folds, a 950 mM LiCl, 50 mM KCl, 25 mM Tris, 1 mM EDTA, pH 7.9 solution was used; the basket folds were induced with a 1.00 M NaCl, 25 mM Tris, 1 mM EDTA, pH 7.9; and, the propeller fold was induced with a 5.00 M LiCl, 20 mM KCl, 25 mM Tris, 1 mM EDTA, pH 7.9 solution. The ODNs were annealed by first heating the samples to 96 ^{o}C, and then cooling to room temperature over 4 h; next, the samples were stored at 4 ^{o}C over 2 days. The thermal denaturation studies were conducted by measuring the absorbance at a 295 nm wavelength for basket and hybrid folds and at 260 nm for hybrid and triplex folds.

Current time recordings

The ion channel recordings were conducted with a custom-built amplifier and data acquisition system designed by Electronic BioSciences (EBS), San Diego, CA. The glass nanopore membrane (GNM) was fabricated using previously established procedures.[45,46] The data were collected at 21 ± 1 ^{o}C using a 500 kHz sampling rate and 100 kHz low-pass filter; however, for presentation purposes, the data were refiltered to 20 kHz. The data were analyzed using QUB 1.5.0.31 and fitted using OriginPro 9.1. For

each ODN, three voltages were studied, 120 mV, 140 mV, and 160 mV (*trans* vs. *cis*). At each voltage, data for roughly 1000 events were collected.

Results

Thymine-dimer containing G-quadruplexes characterization using conventional methods

The T=T-containing human telomere sequences chosen for the study were synthesized from commercially available phosphoramidites. The T=T was introduced in three places along the truncated natural-form of the human telomere sequence that could fold into a single G-quadruplex and possessed a tail at both ends comprised of two nucleotides (Table 5.1). Circular dichroism (CD) spectroscopy, thermal denaturation studies (T_m) and electrophoretic mobility shift assays (EMSA) were conducted on each sequence under the following high ionic strength conditions that are required to conduct the α-HL ion channel recordings.[41,42] These conditions are: (1) 50 mM KCl, 950 mM LiCl (hybrid folds), (2) 1.00 M NaCl (basket folds) or (3) 20 mM KCl with 5.00 M LiCl (propeller folds) all samples were in 25 mM Tris (pH 7.9) with 1 mM EDTA. The 5.00 M LiCl solution was previously shown by our laboratory to induce the propeller fold because of its dehydrating effect.[42] These data were compared to those collected under low ionic strength conditions in the following salts: (1) 140 mM KCl and (2) 140 mM NaCl in phosphate buffer (pH 7.4).

First, CD measurements were conducted for the natural sequence under the three different high ionic strength salt conditions. The natural telomere sequence gave data that were consistent with literature reports (Figure 5.2).[47] Then, the T=T-containing G-

Table 5.1. The T=T-containing human telomere G-quadruplex sequences studied.

Name	Sequence
Natural sequence	5`-TAGGGTTAGGGTTAGGGTTAGGGTT
5' T=T	5`-TAGGGT=TAGGGTTAGGGTTAGGGTT
M T=T	5`-TAGGGTTAGGGT=TAGGGTTAGGGTT
3' T=T	5`-TAGGGTTAGGGTTAGGGT=TAGGGTT

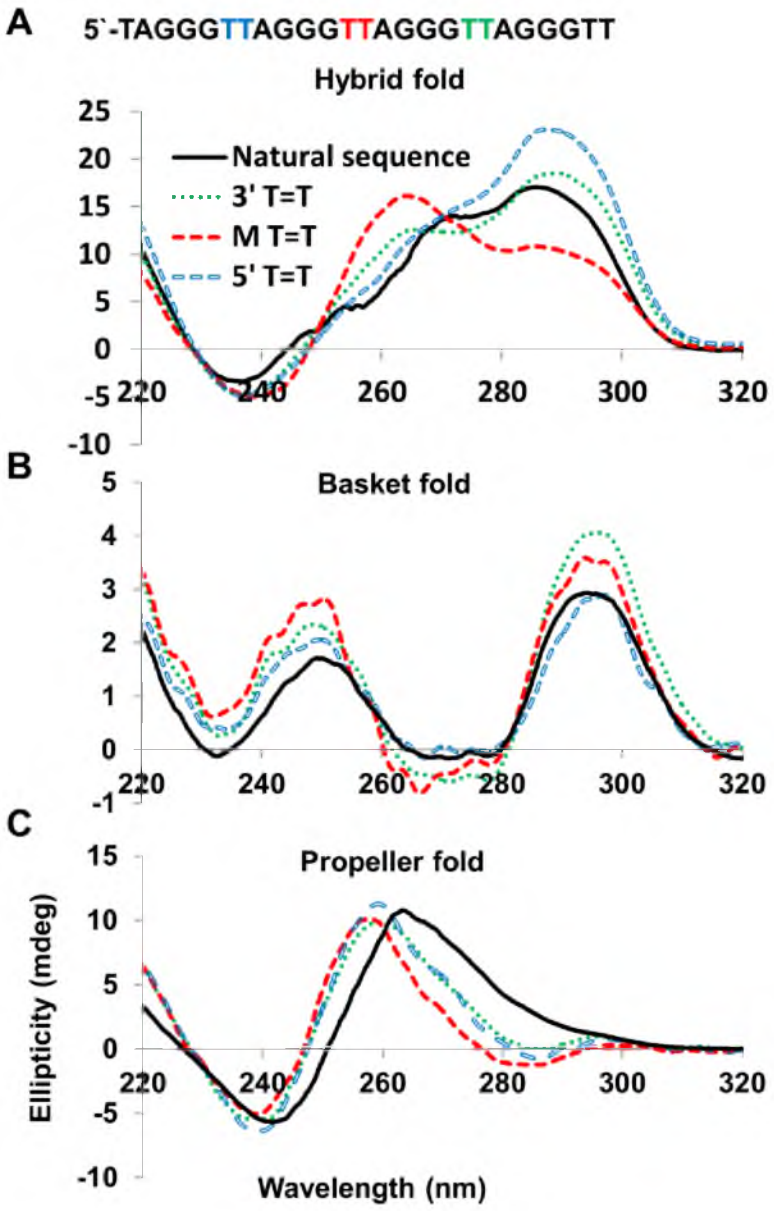
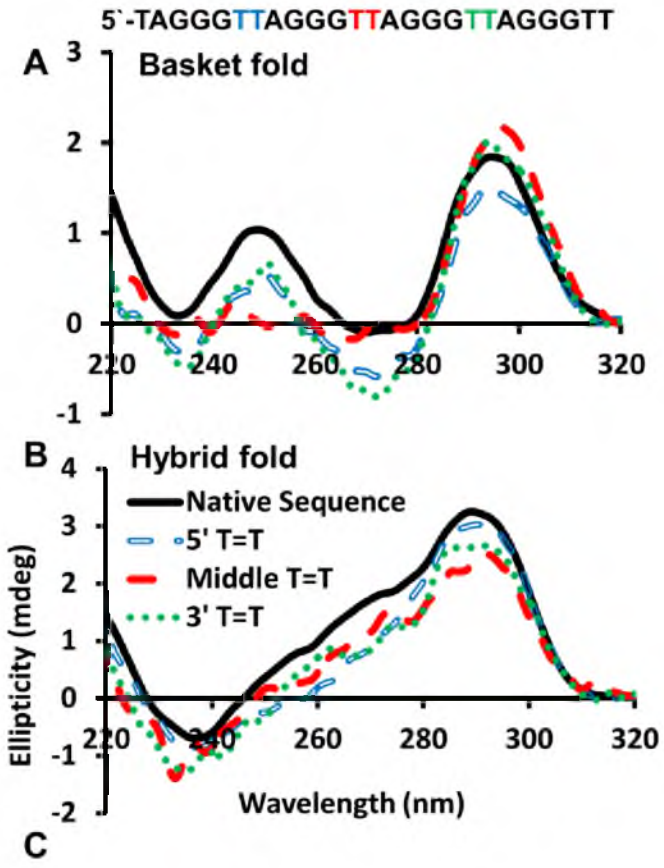


Figure 5.2. CD spectra of the human telomeric G-quadruplexes. The sequences are color coded as follows: blue – the T=T in the 5' loop, red – the T=T is introduced in the middle loop (M T=T), green – the T=T in the 3' loop of the quadruplex, black – natural sequence. **A.** Hybrid fold: annealed in 950 mM LiCl, 50 mM KCl, 25 mM Tris, 1 mM EDTA, pH 7.9. **B.** Basket fold: annealed in 1.00 M NaCl, 25 mM Tris, 1 mM EDTA, pH 7.9. **C.** Propeller fold: annealed in 5.00 M LiCl, 20 mM KCl, 25 mM Tris, 1 mM EDTA, pH 7.9.

quadruplexes were tested and gave characteristic antiparallel (hybrid and basket) signatures with positive peaks at ~290 nm (Figure 5.2A and B). Additionally, the hybrid folds featured a characteristic shoulder at 265 nm and a negative peak at 240 nm (Figure 5.2A). The middle T=T-containing strand had a bigger 265 nm peak that was shifted to the left, possibly signifying partial unfolding into a triplex. The basket fold featured an additional positive peak at around 240 nm and a negative peak at around 265 nm (Figure 5.2B). These data lead to the conclusion that the T=T-containing oligodeoxynucleotides (ODNs) fold to the hybrid and basket topologies in the presence of K^+ and Na^+, respectively. As verification that the high ionic strength conditions do not change the folding patterns, these data were compared to CD data that were obtained under biologically relevant ionic strength of 140 mM. The data look the same between the two different salt concentrations (Figures 5.3 and 5.2). This is interpreted to mean that the high ionic strength does not change the overall structure of the quadruplexes. In contrast to the antiparallel folds, the propeller fold (parallel stranded) featured an intense peak at around 264 nm, and a negative peak at 240 nm. The T=T-containing ODNs were shifted by around 5 nm toward smaller wavelengths compared to the natural sequence (Figure 5.2C). These data suggest that the T=T-containing human telomere sequence adopted a slightly different fold than propeller in high molar concentrations of LiCl (5.00 M). The recorded CD spectra for these lesion bearing G-quadruplexes were similar to the CD spectrum reported for the triplex folding sequence 5'-TTA$(GGGTTA)_3$-3'.[18,48,49] However, these data did not allow us to further confirm this observation.

Next, we examined the thermal stability of T=T-containing ODNs via monitoring the UV-Vis signature change as a function of increased temperature. In KCl solution that



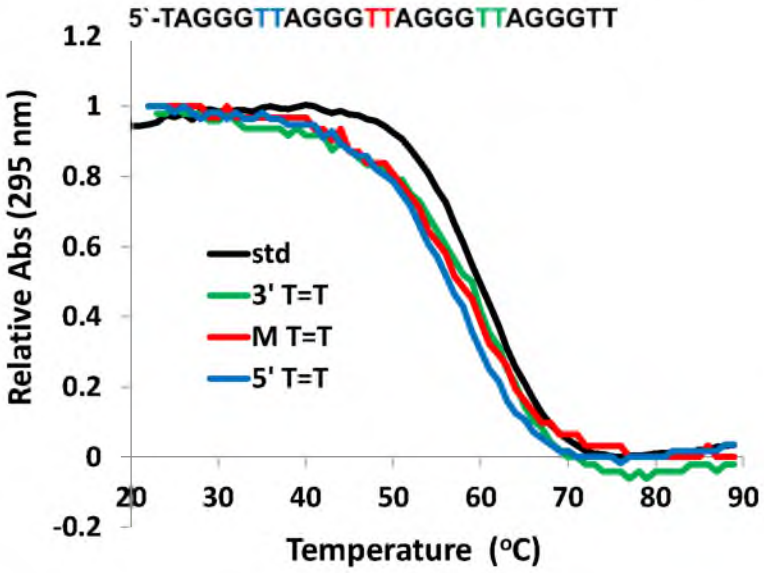
	Hybrid fold[a]	Basket fold[b]
	T_m (^oC)	T_m (^oC)
Natural sequence	65.4 +/- 0.7	56.3 +/- 0.8
5' T=T	60.6 +/- 1.0	52.6 +/- 0.6
3' T=T	62.0 +/- 0.7	52.4 +/- 0.7
M T=T	51.6 +/- 0.6	52.3 +/- 0.4

[a]120 mM KCl, 20 mM PBS, pH 7.4.
[b]120 mM NaCl, 20 mM PBS, pH 7.4.

Figure 5.3. CD and T_m data for the physiological salt concentrations of G-quadruplexes containing the T=T. **A.** CD spectra for the basket fold annealed at 120 mM NaCl, 20 mM PBS, pH 7.4. **B.** CD spectra for the hybrid fold annealed at 120 mM KCl, 20 mM PBS, pH 7.4. **C.** T_m data for the physiological salt concentrations.

induces the hybrid fold, the T=T-containing ODNs, showed slightly lower T_m values ($\Delta T_m \sim -4\ ^oC$). The representative T_m curves are included in Figure 5.4. The data indicate that the presence of T=T only slightly destabilized the structure in a KCl solution. Similarly, in NaCl solution that gives the basket folds, the T=T-containing folds had slightly lower T_m values ($\Delta T_m \sim -4\ ^oC$) than the native sequence, again indicating slightly lower stability of the damage-containing G-quadruplexes. In the solution that induces the propeller fold, the T_m values were dependent on the location of the T=T. The biggest T_m decrease ($\Delta T_m \sim -10\ ^oC$) was observed when the T=T was positioned in the middle loop of a propeller fold; when the T=T was positioned in either the 5' and 3' loops, the T_m was slightly lower than the native sequence ($\Delta T_m \sim -5\ ^oC$).

Finally, we conducted EMSA experiments in NaCl or KCl to further support the ability of T=T-containing telomere sequences to fold into G-quadruplex structures (Figure 5.5). All T=T-containing strands migrated approximately the same distance as the natural sequence. This experiment could not be performed for the propeller fold due to the 5.00 M LiCl that was required to induce this topology. The methods mentioned so far indicate that the T=T-containing human telomere sequences folded into G-quadruplexes are slightly less stable than the undamaged ODNs in the presence of K^+ and Na^+. The CD spectra for the damaged G-quadruplexes in KCl and NaCl indicate minor structural changes, suggesting that T=T incorporation in the loops does not dramatically alter the quadruplex folds. Lastly, in 5.00 M LiCl the human telomere sequence assumes a propeller fold; however, the damage-containing ODNs presented subtle perturbations to the CD spectra that are challenging to interpret by these methods. In order to gain more insight into the structural differences between the damage-containing strands, a single-



	Hybrid fold[a]	Basket fold[b]	Propeller fold[c]
	T_m (°C)	T_m (°C)	T_m (°C)
Natural sequence	59.9 +/- 0.8	74.5 +/- 0.3	54.2 +/- 0.9
5' T=T	54.4 +/- 0.5	69.2 +/- 0.8	51.0 +/- 0.6
3' T=T	57.3 +/- 0.4	68.8 +/- 0.5	49.4 +/- 0.4
M T=T	55.8 +/- 0.5	71.5 +/- 0.6	43.9 +/- 0.7

[a]50 mM KCl, 950 mM LiCl, 25 mM Tris, 1 mM EDTA, pH 7.9
[b]1.00 M NaCl, 25 mM Tris, 1 mM EDTA, pH 7.9.
[c]5.00 M LiCl, 20 mM KCl, 25 mM Tris, 1 mM EDTA, pH 7.9

Figure 5.4. Representative T_m curves for a hybrid fold and T_m values for all the G-quadruplex folds studied.

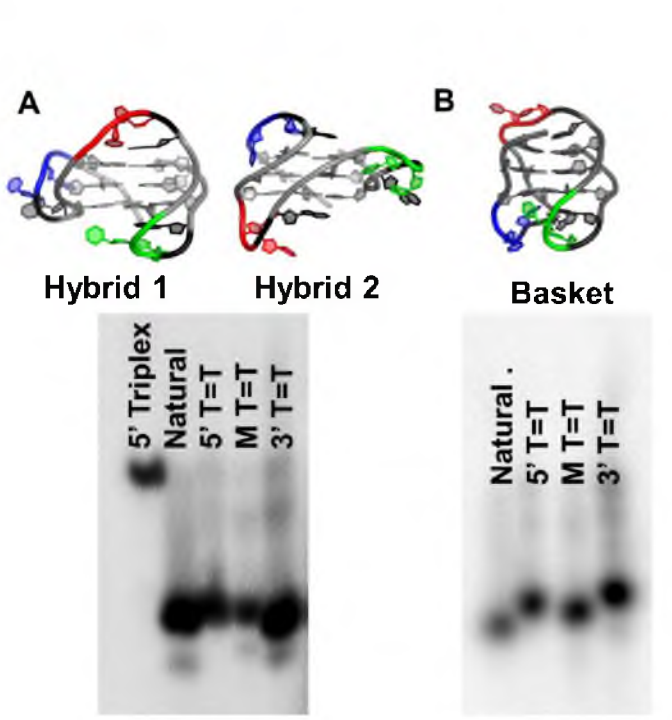


Figure 5.5. EMSA experiments to confirm the hybrid and basket folds for the T=T-containing strands. **A.** Hybrid fold: native gel (20%) run for 8h in 100 mM NaOAc, TBE buffer pH 7.9. **B.** Basket fold: Native gel (20%) run for 8h in 100 mM KOAc, TBE buffer pH 7.9

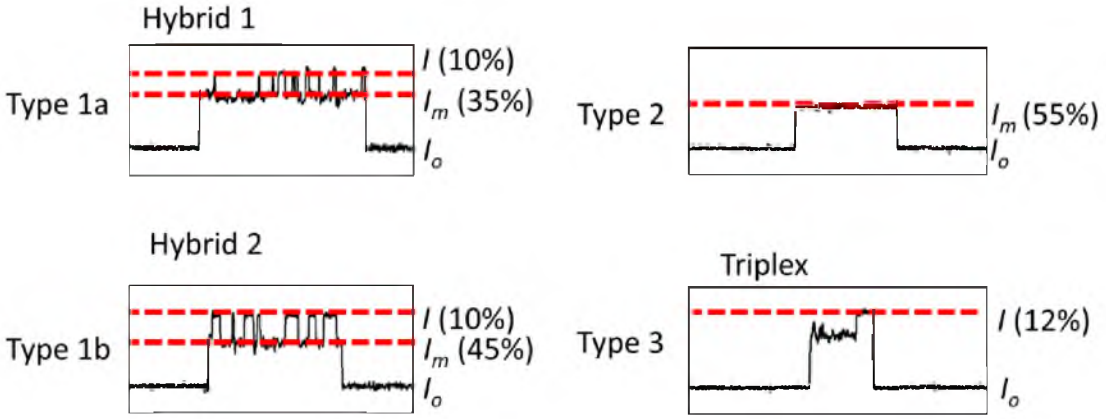
molecule experiment was conducted with the α-HL ion channel. This method has been recently developed in our laboratory to discriminate different topologies of G-quadruplexes.[41,42]

Single-molecule analysis of the thymine dimer-containing G-quadruplexes

The ion channel recordings were conducted according to previously established procedures from the White laboratory.[45,46] Briefly, the wild type α-HL was assembled from monomers into a lipid bilayer that was painted across a glass nanopore membrane. A voltage was applied across the membrane, and the DNA was electrophoretically driven toward the *cis* side of the α-HL (Figure 5.1B). Different G-quadruplex folds can yield unique current signatures by interacting differently with the α-HL nanocavity.[41,42] We took advantage of this system to understand the folding patterns of T=T-containing G-quadruplexes in the three different salt solutions.

Hybrid folds

First, the native sequence: 5`-TAGGGTTAGGGTTAGGGTTAGGGTT-3' was studied. In the presence of K^+, three different event types were observed. They represent the hybrid folds (type 1 or type 2), loop entry (shallow blockage only, no translocation) and triplex folds, similar to our previous report (Figure 5.6).[41] These event types were demonstrated to enter the nanopore with different frequencies based on their shape; therefore, correction factors for this event frequency bias were applied to the current data, in which the hybrid 1, hybrid 2 and triplex intermediate had relative event frequencies of



Event Types	Std [%]*	5'[%]*	M[%]*	3'[%]*
Type 1 (a + b)	74 ± 9	70 ± 8	61 ± 7	69 ± 8
Type 2	21 ± 3	21 ± 2	26 ± 4	25 ± 3
Type 3	5 ± 2	9 ± 1	12 ± 1	6 ± 1

Figure 5.6. Types and percent distribution of the current-time (*i-t*) traces (damaged ODNs and standard) for the nanopore experiments in KCl.

*The percent distribution was corrected according to the entry rate ratios previously established by our laboratory.[41]

1.1, 1.0 and 6.2, respectively.[41] To briefly reiterate, the major event types observed were characteristic of the hybrid folds (~75%, Figure 5.6). These events gave different current signatures that oscillated between a deep current level (I) and a midcurrent level (I_M, Figure 5.7A). Furthermore, the midcurrent levels were distinct and assigned as hybrid 1, $I_M/I_o = 37 \pm 3\%$, and hybrid 2, $I_M/I_o = 44 \pm 4\%$ based on our previous work.[41] Hybrid 1 was shown to be more abundant than hybrid 2 with a ratio of approximately 2:1. The statistical analysis of the event types corresponding to hybrid 1 and hybrid 2 is presented in Figure 5.7B. The error on the ratio between hybrid 1 and hybrid 2 is 10% of each value. Next, current signatures for the loop entry corresponded to ~20% of the event types, and event types for the triplex fold represented ~5% of the population.

The G-quadruplexes that contained UV-induced damage at the 3' loop, middle loop or at the 5' loop, presented three types of ion current events that were similar to the native sequence. The hybrid event types were slightly less frequent (60-70%, Figure 5.6) than observed for the undamaged G-quadruplex. Hybrid 1 and 2 distributions for the ODNs containing the T=T at the 3' loop and the middle loop gave a similar 2:1 ratio, respectively, as observed in the native sequence. This observation is interpreted to mean that incorporation of T=T in the 3' or middle loops of a hybrid quadruplex does not alter the hybrid structural distribution. A different trend was observed in the case of the T=T in the 5' loop, in which the predominant fold was hybrid 2 (Figure 5.7B). We hypothesize that the presence of T=T in the double-chain reversal loop caused the equilibrium between the hybrids to change in favor of hybrid 2, in which the damage was placed in the edgewise loop. This hypothesis was further explored by examining the basked fold that does not have a double-chain reversal, and the propeller fold that only

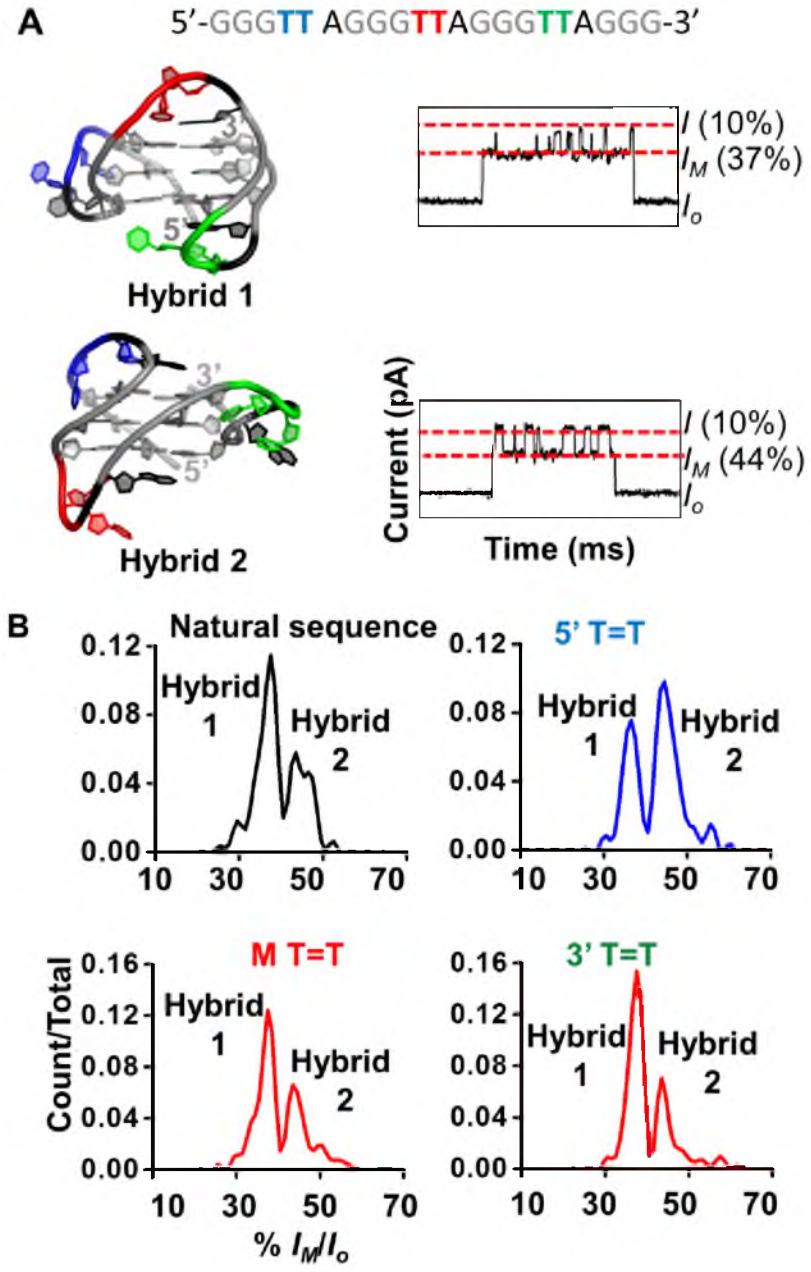


Figure 5.7. Analysis of hybrid 1 and hybrid 2 folds for G-quadruplexes containing the damaged base in different positions along the human telomere sequence at a single-molecule level. **A.** The human telomere sequence with the positions of the T=T color coded in blue, red and green followed by the structure of hybrid type 1 (pdb 2JSK)[12] and type 2 (pdb 2JSQ)[12] with their representative current time (*i-t*) trace signature. **B.** Current histograms representing % I_M/I_o of the events characteristic for the hybrid folds.

has double-chain reversal loops.

The remainder of the event types observed in the T=T-containing quadruplexes were characteristic of the loop entry and the triplex folding intermediate. The frequency of loop entry in all damaged quadruplexes was similar to that observed in the native sequence (~20%, Figure 5.6). Finally, the observed percentage of triplex-type events was slightly higher than the natural sequence (~5%, Figure 5.6). This observation was expected based on the lower T_m values for the damage-containing strands. In summary, the α-HL experiment allowed us to determine that T=T in the loops of G-quadruplexes show a favorability for edgewise loops over double-chain reversal loops that could not have been determined by CD, EMSA and T_m experiments.

Basket fold

In NaCl solutions the human telomere sequence assumes a basket fold that consists of two edgewise loops and one diagonal loop (Figure 5.8A). The basket fold can interact with the α-HL in two different manners: (1) it enters the vestibule, unravels itself, threads into the narrow constriction zone with either one of the two-nucleotide overhang tails and translocates across the pore, or (2) it enters the vestibule from the loop side with both overhangs away from the constriction; thus, it is unable to thread and translocate until the molecule escapes back to the *cis* side of the protein channel in a diffusion controlled process.[42,50] Similar to the natural sequence, the damage-containing ODNs featured the same event distributions: one that translocated across the α-HL presenting deep current blockage and one that did not translocate through the channel presenting shallower current blockage (Figure 5.8B). Moreover, the deep current blockage event

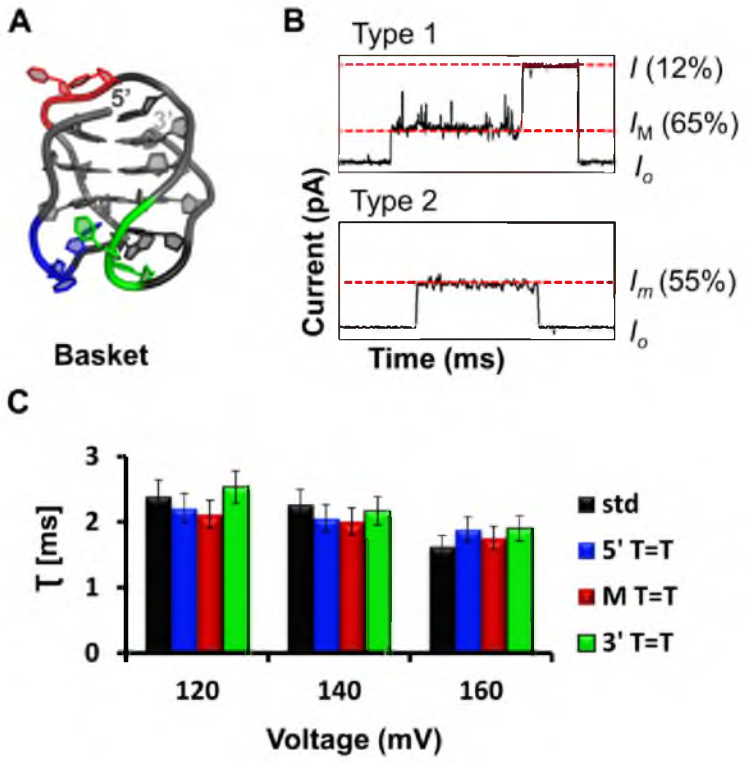


Figure 5.8. Nanopore analysis of the basket fold. **A.** Structure of a basket (pdb 143D),[15] with the different positions of the T=T color coded. **B.** Representative *i-t* traces for the two different event types of events observed. **C.** Voltage-dependence studies in which the translocation time (t_D) of ~500 events was fitted to an exponential decay model.

type was modestly sensitive to increased voltage giving an inverse correlation, indicating that it unravels and translocates to the *trans* side of the pore (Figure 5.8C). The shallow current blockage gave a time distribution with change of the voltage that was shallow, and was interpreted to indicate a diffusion controlled process back out the *cis* opening of the pore. These results are interpreted to say that the T=T in a basket fold does not affect the structure of the G-quadruplex significantly. Further, this single-molecule experiment supports the conclusions drawn from the T_m and CD experiments, in which incorporation of a T=T in an edgewise or diagonal loop is not that destabilizing to the structure.

Propeller fold

Under dehydrating conditions (5.00 M LiCl), the human telomere sequence can assume the propeller fold (all double-chain reversal loops) in the presence of K^+.[17,18,51] The intact propeller fold was too large (~4 nm, Figure 5.1) to enter the vestibule of the α-HL ion channel (~3.0 nm, Figure 5.1); therefore, the only events detected during the nanopore study were very short spikes caused by random interactions of the G-quadruplex with the mouth of the vestibule (Figure 5.9A). This observation was previously reported by our laboratory.[42] When UV-induced damage was introduced into the sequence, deep blockage events were recorded as shown in the bottom *i-t* trace in Figure 5.9A. Two distinct current vs. time populations were observed in heat plots of the data (Figure 5.9B and Figure 5.10). We interpret these two populations to represent different entry orientations of the ODN. Based on our previous knowledge from the hybrid studies, T=T is not preferred in a double-chain reversal loop. Therefore, the T=T can unwind the propeller fold to a triplex structure that has a head with a loop and a tail

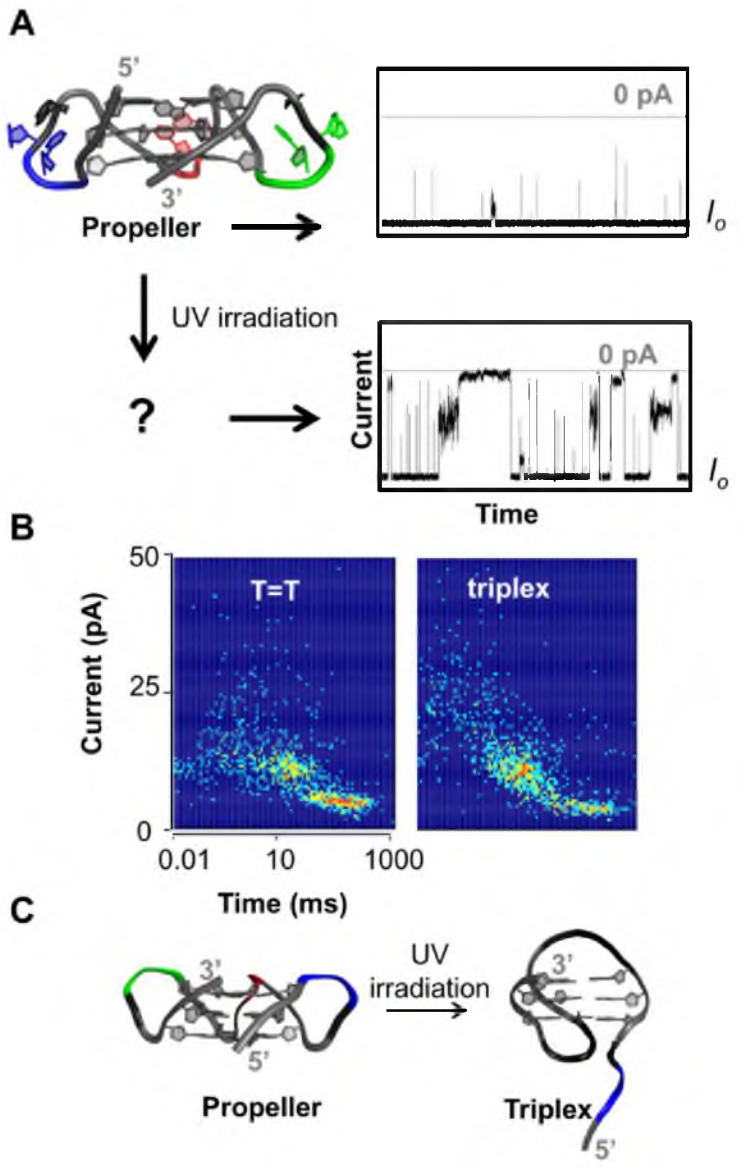


Figure 5.9. Nanopore analysis of T=T-containing human telomere sequence in 5.00 M LiCl. **A.** Undamaged propeller fold (pdb 1KF1)[16] structure and the typical nanopore experiment *i-t* traces. The intact propeller fold is too large to enter the vestibule of the α-HL, therefore, there are no translocation events. Once the damage is introduced long, deep blockage events are observed that might correspond to a triplex-like structure caused by the partial unfolding of the propeller due to the damage. **B.** Representative heat plot of the T=T containing telomere sequence and a triplex forming sequence where Gs have been substituted with Ts at either 5' or 3' loop. **C.** Proposed structure based on the report from Koirala, *et al.*[48]

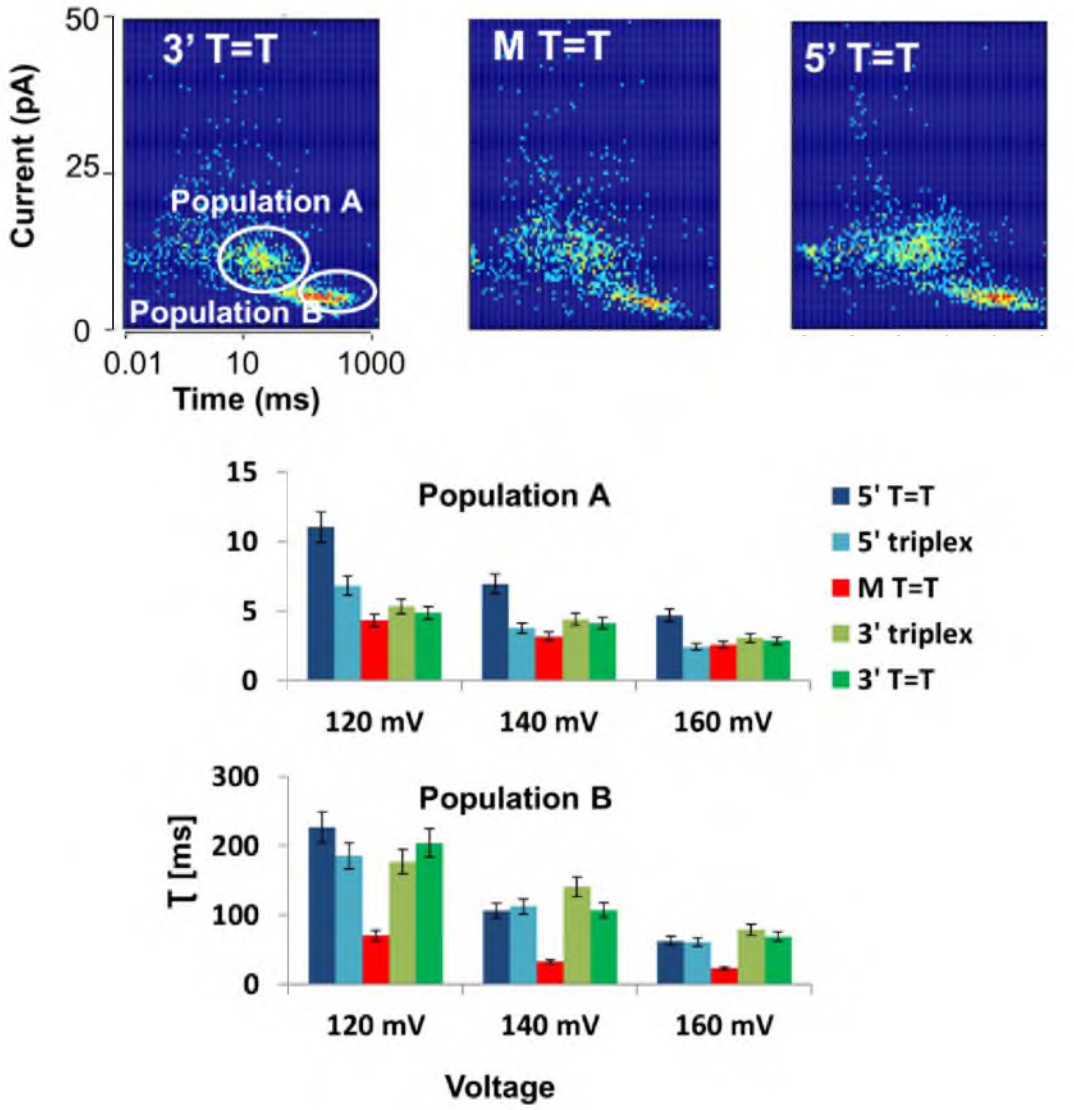


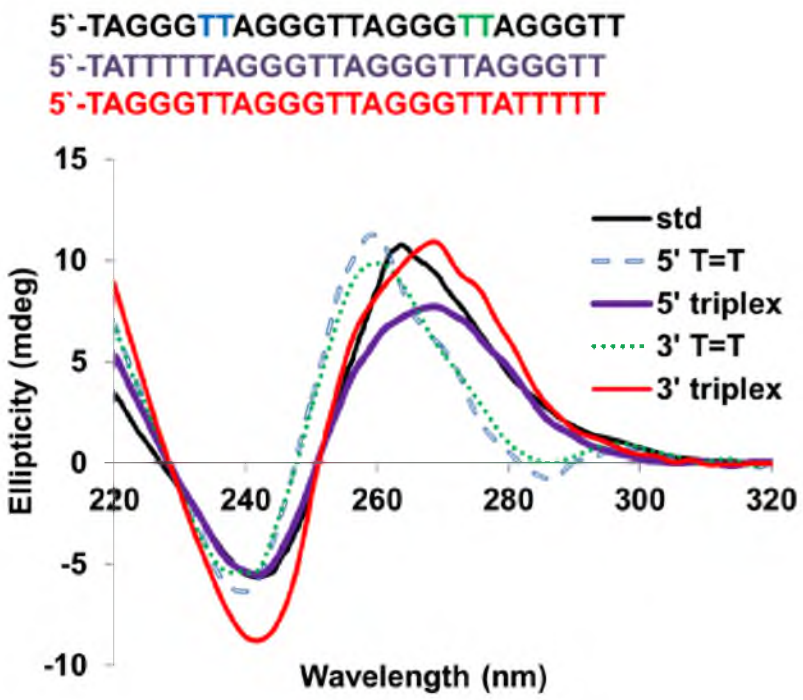
Figure 5.10. Color plots and statistical analysis of the event durations in 5.00 M LiCl during the nanopore experiments of the T=T-containing human telomere sequence and a 5' and 3' triplex sequence. The data were plotted using a single exponential decay model. As shown in the color plots there are two populations that show voltage dependence.

comprised of eight nucleotides (Figure 5.9C). Coinciding with the T_m data, the shortest translocation time was observed for the ODN that contained the T=T in the middle of the structure (Figure 5.4 and 5.10). The T=T in the middle loop would have the biggest effect on the ability of the ODN to fold, explaining its lower T_m value and the shortest translocation time (Figure 5.10).

Two lines of evidence support the hypothesis that T=T incorporation into a propeller G-quadruplex yields a triplex-like structure. First, the CD spectra for the T=T-containing quadruplexes in 5.00 M LiCl give similar spectra (λ_{max} = 264 nm and λ_{min} = 240 nm) as reported by the Sugiyama laboratory for the triplex-forming sequence.[48] Second, a series of control strands was designed to model the triplex species induced by the T=T, in which the terminal 5' or 3' G run was converted to a 5' or 3' T run, respectively. These controls had an eight-nucleotide tail on the 5' or 3' end (Figure 5.11). Nanopore analysis of these control strands also gave two current time populations in the heat plots shown in Figure 5.9B. These control studies were similar to the data collected for T=T-containing quadruplexes, which are interpreted to mean that T=T cause the double-chain reversal loop to unwind giving a triplex-like structure. This observation further supports our hypothesis that T=Ts do not facilitate double-chain reversal loop formation.

Discussion

In the present work, established methods were used to study how UV-induced damage affects the structure of the human telomere G-quadruplex folds. Utilizing CD and EMSA techniques, it was found that the hybrid and basket folds are formed



	T_m (^oC)
3' T=T	**49.4 +/- 0.4**
3' triplex	**47.2 +/- 0.8**
5' T=T	**51.0 +/- 0.6**
5' triplex	**50.8 +/- 0.5**

Figure 5.11. T_m and CD spectra of the human telomere folds and triplex folds in a 5.00 M LiCl solution. CD spectra: The comparison between a propeller fold (black), triplex folds (5'-purple and 3'-red) and T=T-containing ODNs (5'-blue and 3'-green). T_m data: comparison between triplex folds and T=T-containing folds.

regardless of the loop in which the T=T was placed (Figure 5.2 to 5.5). The T_m results suggest that T=Ts cause a moderate decrease (~5 ^oC) in the thermal stability of hybrid and basket folds (Figure 5.11). Similar to the current results, the introduction of thymine glycol[34] or 8-oxo-A,[36] into a loop was not detrimental to the hybrid or basket structure based on CD and EMSA experiments. Furthermore, introduction of T=T or thymine glycol into a loop of a G-quadruplex induces a moderated decrease in the thermal stability. Contrary to these results, the introduction of 8-oxo-A into the loops causes a significant increase in the T_m.[36] These data are interpreted to mean that the cyclobutane ring of the T=T restricts the conformational flexibility of the loops that leads to subtle decrease in stability, without altering the overall fold for hybrid and basket G-quadruplexes. In contrast to these observations, T=T in the human telomere sequence under propeller folding conditions gave the largest effect on the thermal stabilities (Figure 5.11). Also, the CD showed that the fold resembles a triplex and not a propeller fold.[47,48] The conventional studies performed, yielded interesting results. However, they were inadequate to predict the changes that introduction of the T=T causes to the G-quadruplex fold. Therefore, we have utilized the single-molecule method that was recently shown to be a good tool in observing the different folds of the G-quadruplex.[41] The nanopore measurements gave us a unique opportunity to further explore how the UV-induced DNA damage affects folding of the human telomeric G-quadruplexes under different physical conditions.

The α-HL ion channel has been extensively used to study nucleic acids.[52-54] A simple G-quadruplex containing only two tetrads (thrombin binding aptamer, TBA) was first studied using α-HL regarding its cation affinities, followed by the studies of

thrombin-TBA interactions.[55-59] Later, the human telomere sequence and the different folds that G-quadruplex can assume were studied by our laboratory.[41,42] The ability of the α-HL to distinguish between different G-quadruplex folds is attributed, among others, to the recently discovered sensing zone that is located in the vestibule and not in the β-barrel of the protein ion channel.[42,60] The vestibule of the α-HL (~3.0 nm) was shown to exclude the propeller fold (1.8 nm x 4.0 nm), since its dimensions are too big for it to enter (Figure 5.1).[42] The hybrid folds were barely able to enter (2.7 nm x 3.0 nm), and therefore even the slight difference in shape between hybrid 1 and hybrid 2 gave different signals.[41] Finally, the basket fold (2.4 nm x 2.8 nm) was small enough to easily enter and translocate across the pore.[42] In this study, we have shown using the nanopore method, that when the T=T was introduced into a double-chain reversal loop of the hybrid fold the equilibrium between the hybrids changed to favor the conformation with T=T positioned in an edgewise loop. This hypothesis was further supported by examining the other folds of the human telomere sequence: propeller (all double-chain reversal loops) and basket (no double-chain reversal loops). The T=T-containing basket fold showed the same type of events as for the natural sequence, whereas the propeller fold showed events that were similar to a triplex fold (not a propeller fold); thus, supporting our hypothesis that T=Ts are not well accommodated in the double-chain reversal loop.

The hybrid fold is the proposed topology of the G-quadruplex that the telomere region can assume *in vivo*.[13,20,22] Hybrids can adopt two different folds positioning the double-chain reversal loop either at the 3' or 5' termini of the G-quadruplex. The ability of the hybrid to change from hybrid 1 to hybrid 2 allows the fold to accommodate the T=T damage without significant structural perturbations. This quality of the telomere is

crucial for its function since the region is hypersensitive to the UV-induced damage.[25] Telomere damage that causes a change in the folding pattern (G-oxidation products),[33] is proposed to induce telomere shortening that ultimately leads to cell death.[28] The single-stranded portion of the telomere is protected from degradation and RPA recognition by the overhang binding proteins, POT1 and TTP1.[24,61] It was recently shown by Ray, *et al.*,[23] that the hybrid fold is crucial for POT1 binding. Therefore, DNA damage that significantly affects the G-quadruplex structure will decrease the binding of the POT1/TTP1 complex *in vivo*, such as G-oxidation products that significantly impact the structural fold.[27] However, based on the current results T=T does not affect the topology of the G-quadruplex hybrid fold; therefore, it does not lose protection by POT1/TTP1 complex binding, and ultimately does not cause telomere shortening, which is in contrast to damage from G oxidation. The current data demonstrate that T=Ts are accommodated in the G-quadruplex folds of the human telomere sequence, and therefore, can accumulate without causing significant structural perturbations leading to telomere truncation.

Conclusions

Following the traditional procedures to analyze G-quadruplexes (CD, T_m and EMSA) we demonstrated that the presence of T=T damage does not significantly distort the overall secondary structure, and only slightly decreases the stabilities of hybrid and basket folds (Figures 5.2 to 5.5 and 5.11). In order to gain more insight into the structural changes that can arise from T=T formation in telomeres, we used the α-HL nanopore to analyze the damaged folds at a single-molecule level. From these experiments, we established that the presence of a T=T in the hybrid folds changed the ratio of the hybrid

types to favor the conformation that positions the damaged site at an edgewise loop instead of a double-chain reversal loop. This observation was further verified by examining the basket fold (edgewise and diagonal loops) and the propeller fold (all double-chain reversal loops), in which the basket fold was not altered by T=Ts and the structure for the propeller topology was significantly impacted.

In cells, the predominant G-quadruplex structure for the natural human telomere sequence is proposed to be the hybrid fold, based on the K^+ concentration inside the cell and its binding affinity to the G-quadruplex.[13,20,22] Once the T=T is present, the shape of the structure remains predominantly the same, aside from the hybrid 1 to hybrid 2 ratio if the damage is located in the double-chain reversal loop. This observation explains why the telomere can still maintain its functions and does not undergo shortening even after extensive UV irradiation.[25] Even if the T=T is present in the hybrid fold the damage does not affect the overall shape of the fold making it possible for the proteins POT1/TPP1 to bind and maintain proper function. Therefore, hybrid G-quadruplexes can mask any detrimental structural consequences of T=Ts in the cell.

The current study focused on telomeres that are thought to fold into hybrids *in vivo*; however, the hybrid folds are hypothesized to be the predominant fold only in telomeres. There are many other regions along the genome that can fold into the G-quadruplexes that have other folds than hybrid.[62] For instance, most promoter G-quadruples are thought to fold into the propeller fold, in which case the presence of a T=T would greatly destabilize the fold based on the current data. Therefore, future studies will focus on examining how T=T influences different DNA sequences that can fold into G-qudruplexes. Also, other damages aside from T=T, such as platinum

adducts,[63] Sp, Gh, Tg[33] or 8-oxoA,[36] might be studied using the nanopore technology to determine how the damage affects the structure of the G-quadruplexes.

References

(1) Roderick, J. O. S.; Jan, K. Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 171-181.

(2) Osterhage, J. L.; Friedman, K. L. Chromosome end maintenance by telomerase. *J. Biol. Chem.* **2009**, *284*, 16061-16065.

(3) 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.

(4) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013**, *5*, 182-183.

(5) Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988**, *334*, 364-366.

(6) Cahoon, L. A.; Seifert, H. S. An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science* **2009**, *325*, 764-767.

(7) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **2012**, *8*, 301-310.

(8) Cheung, I.; Schertzer, M.; Rose, A.; Lansdorp, P. M. Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat. Genet.* **2002**, *31*, 405-409.

(9) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593-11598.

(10) Osterhage, J. L.; Friedman, K. L. Chromosome end maintenance by telomerase. *J. Biol. Chem.* **2009**, *284*, 16061-5.

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

(12) 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.

(13) 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.

(14) 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.

(15) Wang, Y.; Patel, D. J. Solution structure of the human telomeric repeat d[AG$_3$(T$_2$AG$_3$)$_3$] G-tetraplex. *Structure* **1993**, *1*, 263-282.

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

(17) Heddi, B.; Phan, A. T. Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.* **2011**, *133*, 9824-9833.

(18) Gray, R. D.; Buscaglia, R.; Chaires, J. B. Populated intermediates in the thermal unfolding of the human telomeric quadruplex. *J. Am. Chem. Soc.* **2012**, *134*, 16834-16844.

(19) 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.

(20) Yang, D.; Okamoto, K. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.* **2010**, *2*, 619-646.

(21) 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.

(22) Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.* **2011**, 1-8.

(23) Ray, S.; Bandaria, J. N.; Qureshi, M. H.; Yildiz, A.; Balci, H. G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, doi: 10.1073/pnas.1321436111.

(24) He, Q.; Zeng, P.; Tan, J.-H.; Ou, T.-M.; Gu, L.-Q.; Huang, Z.-S.; Li, D. G-quadruplex-mediated regulation of telomere binding protein POT1 gene expression. *Biochim. Biophys. Acta* **2014**, *1840*, 2222-2233.

(25) Rochette, P. J.; Brash, D. E. Human telomeres are hypersensitive to UV-induced DNA damage and refractory to repair. *PLoS Genetics* **2010**, *6*, e1000926.

(26) Henle, E. S.; Han, Z.; Tang, N.; Rai, P.; Luo, Y.; Linn, S. Sequence-specific DNA cleavage by Fe_2^+-mediated Fenton reactions has possible biological implications. *J. Biol. Chem.* **1999**, *274*, 962-971.

(27) Oikawa, S.; Tada-Oikawa, S.; Kawanishi, S. Site-specific DNA damage at the GGG sequence by UVA involves acceleration of telomere shortening. *Biochem.* **2001**, *40*, 4763-4768.

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

(29) Samassekou, O.; Gadji, M.; Drouin, R.; Yan, J. Sizing the ends: normal length of human telomeres. *Annals of Anatomy - Anatomischer Anzeiger* **2010**, *192*, 284-291.

(30) Owen, M. W.; Synthia, H. M.; Elissa, S. E.; Jue, L.; Firdaus, S. D.; Yali, S.; Victor, I. R.; Rebecca, R.; Heather, M. B.; Eve, K.; Mariana, C.; Nelson, J. C.; Elizabeth, H. B. Leukocyte telomere length in major depression: correlations with chronicity, inflammation and oxidative stress - preliminary findings. *PLoS ONE* **2011**, *6*.

(31) Steenken, S.; Jovanovic, S. 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.

(32) Burrows, C. J.; Muller, J. G. Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **1998**, *98*, 1109-1152.

(33) Fleming, A. M.; Burrows, C. J. G-quadruplex folds of the human telomere sequence alter the site reactivity and reaction pathway of guanine oxidation compared to duplex DNA. *Chem. Res. Toxicol.* **2013**, *26*, 593-607.

(34) Zhou, J.; Liu, M.; Fleming, A. M.; Burrows, C. J.; Wallace, S. S. Neil3 and NEIL1 DNA glycosylases remove oxidative damages from quadruplex DNA and exhibit preferences for lesions in the telomeric sequence context. *J. Biol. Chem.* **2013**, *288*, 27263-27272.

(35) Fleming, A. M.; Orendt, A. M.; He, Y.; Zhu, J.; Dukor, R. K.; Burrows, C. J. Reconciliation of chemical, enzymatic, spectroscopic and computational data to assign the absolute configuration of the DNA base lesion spiroiminodihydantoin. *J. Am. Chem. Soc.* **2013**, *135*, 18191-18204.

(36) Aggrawal, M.; Joo, H.; Liu, W.; Tsai, J.; Xue, L. 8-Oxo-7,8-dihydrodeoxyadenosine: the first example of a native DNA lesion that stabilizes human telomeric G-quadruplex DNA. *Biochem. Biophys. Res. Commun.* **2012**, *421*, 671-677.

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

(38) Norval, M.; Lucas, R. M.; Cullen, A. P.; de Gruijl, F. R.; Longstreth, J.; Takizawa, Y.; van der Leun, J. C. The human health effects of ozone depletion and interactions with climate change. *Photochem. Photobiol. Sci.* **2011**, *10*, 199-225.

(39) Su, D. G. T.; Fang, H.; Gross, M. L.; Taylor, J.-S. A. Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12861-12866.

(40) Smith, J. E.; Lu, C.; Taylor, J.-S. Effect of sequence and metal ions on UVB-induced anti cyclobutane pyrimidine dimer formation in human telomeric DNA sequences. *Nucleic Acids Res.* **2014**, *42*, 5007-5019.

(41) An, N.; Fleming, A. M.; Burrows, C. J. Interactions of the human telomere sequence with the nanocavity of the α-hemolysin ion channel reveal structure-dependent electrical signatures for hybrid folds. *J. Am. Chem. Soc.* **2013**, *135*, 8562-8570.

(42) An, N.; Fleming, A. M.; Middleton, E. G.; Burrows, C. J. Size-selective property of α-hemolysin provides signatures for DNA nanostructures formed by the human telomere sequence. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *Accepted.*

(43) Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of *Staphylococcal* α-hemolysin, a heptameric transmembrane pore. *Science* **1996**, *274*, 1859-1865.

(44) Nguyen, K. V.; Burrows, C. J. A prebiotic role for 8-oxoguanosine as a flavin mimic in pyrimidine dimer photorepair. *J. Am. Chem. Soc.* **2011**, *133*, 14586-14589.

(45) 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 nanopore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **2007**, *79*, 4778-4787.

(46) 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.

(47) 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.

(48) Koirala, D.; Mashimo, T.; Sannohe, Y.; Yu, Z.; Mao, H.; Sugiyama, H. Intramolecular folding in three tandem guanine repeats of human telomeric DNA. *Chem. Commun.* **2012**, *48*, 2006-8.

(49) Rajendran, A.; Endo, M.; Hidaka, K.; Sugiyama, H. Direct and single-molecule visualization of the solution-state structures of G-hairpin and G-triplex intermediates. *Angew. Chem., Int. Ed.* **2014**, *53*, 4107-4112.

(50) 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.

(51) 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**, *134*, 15324-15330.

(52) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; 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. Biotech.* **2008**, *26*, 1146-1158.

(53) 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-A precision. *Nat. Biotech.* **2012**, *30*, 344-348.

(54) 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.

(55) Shim, J.; Gu, L.-Q. Single-molecule investigation of G-quadruplex using a nanopore sensor. *Methods* **2012**, *57*, 40-46.

(56) 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.* **2009**, *37*, 972-982.

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

(58) Renner, S.; Geltinger, S.; Simmel, F. C. Nanopore translocation and force spectroscopy experiments in microemulsion droplets. *Small* **2010**, *6*, 190-4.

(59) Gu, L.-Q.; Shim, J. W. Single molecule sensing by nanopores and nanopore devices. *Analyst* **2010**, *135*, 441-451.

(60) Jin, Q.; Fleming, A. M.; Johnson, R. P.; Ding, Y.; Burrows, C. J.; White, H. S. Base-excision repair activity of uracil-DNA glycosylase monitored using the latch zone of α-hemolysin. *J. Am. Chem. Soc.* **2013**, *135*, 19347-19353.

(61) Blasco, M. A. Telomere length, stem cells and aging. *Nat. Chem. Biol.* **2007**, *3*, 640-649.

(62) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **2011**, *10*, 261-275.

(63) Redon, S.; Bombard, S.; Elizondo-Riojas, M. A.; Chottard, J. C. Platinum cross-linking of adenines and guanines on the quadruplex structures of the $AG_3(T_2AG_3)_3$ and $(T_2AG_3)_4$ human telomere sequences in Na^+ and K^+ solutions. *Nucleic Acids Res.* **2003**, *31*, 1605-1613.

CHAPTER 6

SUMMARY AND OUTLOOK

Nanopore technology has emerged as a next-generation, single-molecule sequencing platform that can potentially sequence not only the four native DNA bases, but also DNA modifications, such as epigenetic markers and damage sites. The immobilization experiments shown in Chapters 1 through 4 established current signatures for canonical, damaged, and chemically modified nucleotides at the most sensitive region of the β-barrel of α-HL. When considering the established current levels, it was observed that DNA cross-links have more blocking current levels than all four canonical DNA bases. There is some degree of correlation to the size of the cross-link with its blocking current level, as in the case of T=T vs. 5'-G*CT* vs. platinum adducts (least blocking to most blocking). Additionally, the current levels also depend on hydration, rigidity of the strand (TT vs. T=T), shape of the adduct (diastereomer resolution for Sp, Gh and Tg), as well as interaction of the adduct with the protein and electrolyte, as was observed with the 18c6 adducts. Currently, the 18c6 adduct is the only modification that can be observed in translocation experiments, with a characteristic *i-t* signature and much slower translocation time. Future studies will focus on developing a means to observe other DNA damage sites (such as the cross-links) using either different protein ion channels or mutant α-HL pores. So far, the cross-links have been studied in ssDNA in either

immobilization or translocation experiments. Recently, α-HL was shown to have another sensing zone in the latch region of the vestibule; therefore, studies will be performed to determine whether DNA cross-links can be detected during an unzipping experiment when the cross-links are placed in the latch sensing zone.

Aside from sequencing applications, nanopore technology can be used to detect subtle structural changes in the secondary structure of the DNA such as in G-quadruplexes. The telomere sequence, which is thought to fold into a hybrid fold G-quadruplex, is hypersensitive to UV-induced thymine-dimer (T=T) formation and yet it does not cause telomere shortening. In Chapter 5, the potential structural disruption and thermodynamic stability of the T=T-containing natural telomere sequences were studied to understand why this damage is tolerated in telomeres. First, established methods, such as T_m measurements, electrophoretic mobility shift assays, and circular dichroism spectroscopy were utilized, to determine the effects of damage on these structures. Second, a single-molecule ion channel recording technique with α-HL was used to further examine the structural differences between the damaged sequences. It was observed that the damage caused slightly lower thermal stabilities and subtle changes in the CD spectra for hybrid and basket folds. The α-HL experiments determined that T=Ts disrupt double-chain reversal loop formation but are tolerated in edgewise and diagonal loops. The biggest change was observed for the T=T-containing natural telomere sequence when the propeller fold (all double-chain reversal loops) was studied. Future studies will focus on other regions of the genome that are known to fold into a propeller folds (promoter G-quadruplexes). A T=T that is introduced into a propeller fold was shown to greatly destabilize the secondary structure of the human telomere sequence; therefore, the

damage should also influence other DNA sequences that can fold into a propeller. Additionally, other damages aside from T=T, such as platinum adducts, Sp, Gh, Tg or 8-oxoA, and 8-oxo-G, might be studied using the α-HL to determine how the damage affects the different folds of the G-quadruplexes.