Published in final edited form as:

Isr J Chem. 2013 June 1; 53(6-7): 417–430. doi:10.1002/ijch.201300022.

Electrical Current Signatures of DNA Base Modifications in Single Molecules Immobilized in the α -Hemolysin Ion Channel

Anna H. Wolna, Aaron M. Fleming, Na An, Lidong He, Henry S. White, and Cynthia J. Burrows*

Department of Chemistry University of Utah 315 S 1400 East Salt Lake City, UT 84112-0850

Abstract

Nanopore technology holds high potential for next-generation DNA sequencing. This method operates by drawing an individual single-stranded DNA molecule through a nanoscale pore while monitoring the current deflections that occur as the DNA passes through. Individual current levels for the four DNA nucleotides have been established by immobilization of an end biotinylated strand in the pore in which the nucleotide of interest is suspended at the most sensitive region of the ion channel. Due to the inherent reactivity of the DNA bases, many modified nucleotides in the genome exist resulting from oxidative and UV insults, among others. Herein, the current levels for the common DNA damages 8-oxo-7,8-dihydroguanine (OG), spiroiminodihydantoin (Sp), guanidinohydantoin (Gh), uridine (U), abasic sites (AP), thymine dimers (T=T), thymine glycol (Tg) and 5-iodocytosine (5-I-C) were assessed via immobilization experiments. In some cases, the current difference between the damaged and canonical nucleotides was not well resolved; therefore, we took advantage of the chemical reactivity of the new functional groups present to make amine adducts that shifted the current levels outside the range of the native nucleotides. Among adducts studied, only the 2-aminomethyl-18-crown-6 adduct was able to give a large current shift in the immobilization experiment, as well as to be observed in a translocation experiment. The results show potential in providing current level modulators for identification of some types of DNA damage. In principle, any DNA base modification that can be converted chemically or enzymatically to an abasic site could be identified in this way.

Keywords

-Hemolysin; Nanopore Sequencing; DNA Damage; Nucleic Acid Chemistry

1. Introduction

Emerging as a rapid and inexpensive single-molecule DNA sequencing platform, nanopore ion channel technology has been under intensive investigation.^[1-3] 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-stranded 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.^[4] 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:^[5]

- (1) The ideal system would require no sample amplification and minimal sample

* Phone: 801-585-7290 Fax: 801-585-0024 burrows@chem.utah.edu.

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.^[1]

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).^[6] The constriction zone is too narrow for double-stranded DNA (dsDNA, $d \sim 2$ nm) to be translocated; however, ssDNA ($d \sim 1$ 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.^[7] In a typical experiment, the DNA migrates from the *cis* compartment (negative electrode) to the *trans* compartment (positive electrode) 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 (40-100 mers) of homopolymer DNA.^[8-10] Translocation times typically range from 1-20 μ s/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.^[1] Existing 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,^[6] and as a result, 10-15 bases may simultaneously contribute to the blockage current level.^[11]

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,^[8] decreasing the pH,^[12] increasing the viscosity with glycerol,^[13, 14] or using organic salts;^[15] 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,^[16-18] 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^[19, 20] and others.^[21-27] 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.^[28-31] 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)^[32-35] have recently been shown to slow the DNA translocation rate to a level suitable for resolving the sequence of nucleotides by the Akesson laboratory using α -HL^[36] and by the Gundlach laboratory with the MspA^[35] 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.^[37] Lastly, mutant pores with incorporated molecular breaks (positive charges) have been shown to slow the translocation.^[38, 39] 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.^[40-43]

Henrikson, *et al.* first demonstrated a way to immobilize the DNA in the α -HL using biotinylated DNA (Btn-DNA) and streptavidin.^[40] Streptavidin is too big to enter the vestibule of α -HL, which immobilizes the DNA inside of the α -barrel of the protein, causing a deep level current blockage. Figure 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 2). Next, capture of the DNA-protein complex brings the current to a deep blockage level (I , Figure 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 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 standard nucleotides;^[46] therefore, for the immobilization experiments described in this review 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, 47] In the confined space of the α -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 α -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.^[48] 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.^[49] In this review, 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 “ ” in front of the position number to differentiate these two numbering schemes. Figure 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 8), 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 14 that falls approximately in the middle of the α -barrel (Figure 3).^[49] Because this position gives the biggest current difference between 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.^[50] 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,^[51] water, food and a variety of chemotherapeutic drugs.^[52] 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.^[53] 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.^[54] Our laboratory has a long-standing history in studying oxidatively-derived DNA damage; particularly with respect to oxidation of guanosine (G).^[55-58] G has the lowest redox potential and is the site most prone to oxidative insults.^[59] 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).^[60] also, OG is mildly mutagenic causing G to T transversion mutations.^[54] Moreover, OG is the biomarker followed for assessing oxidative stress to the cell.^[60] 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 4A).^[55, 56] The yield of these two molecules is dependent on the context in which OG is oxidized;^[57] furthermore, these molecules are highly inhibitory to strand elongation by polymerases,^[61] and in vivo studies show them to be highly mutagenic causing G to T and G to C transversion mutations.^[62] 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).^[63] 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.^[64] 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), and in animals Tg has been used as a marker for oxidative stress (Figure 4, B).^[65] Furthermore, Tg is highly mutagenic due to its ability to stall DNA polymerases that leads to failed elongation of the DNA strand.^[66]

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 4, D).^[67, 68] 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^[69] that results from the fact that T=T lesions stall DNA polymerases.^[70] A single day spent in the sun can introduce up to 100,000 UV photoproducts per cell in the epidermis (Table 1).^[71]

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 glycosylic bond results in the formation of abasic sites (AP) that is observed at the purine nucleotides.^[72] The spontaneous base loss is thought to occur 10,000 times per cell per day (Table 1).^[73] AP sites are devoid of genetic information that causes them to be highly stalling to most DNA polymerases.^[74-76] 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 ($t_{1/2} \sim 19$ d)^[77] yielding uridine (U, Figure 4C), which is similar to T in its hydrogen-bonding properties.^[78] The fifth DNA base, 5-methylcytidine (5-mC), is also prone to deamination ($t_{1/2} \sim 9$ d)^[77]

yielding thymidine (T). If the resulting products U or T are not properly repaired, C to T transition mutations are observed.^[73] The deamination of C has been estimated to occur in 100-500 nucleotides per cell per day (Table 1).^[78] Although the overall percentage of damaged DNA bases is small (Table 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 review focuses on studies that have characterized the electrical current signatures of the common DNA base damages via immobilization experiments with the damage placed at 14 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.

2. Canonical nucleotide ion current levels

In the first set of studies, the ion current level histograms for the four canonical nucleotides at position 14 were measured in a poly-dC background sequence (C₄₀). Our studies utilized the glass nanopore membrane as the lipid bilayer support, developed in the White laboratory at the University of Utah.^[80, 81] The poly-dC 40mer was selected to be consistent with literature reports,^[49] 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 % I/I_0 . 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 5 displays the current level histograms for the canonical DNA nucleotides as observed in an immobilization experiment previously described in Figure 3.^[82] These distributions of currents are similar to those reported by Stoddart, *et al.*^[49] 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 °C).^[83]

3. Nucleotide Oxidation Product Current Levels

In the next set of studies, current level histograms were recorded for the G-oxidation products OG, Sp and Gh at position 14 in a poly-dC oligomer (Figure 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 non-planar 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 (which in this case is

caused by the conformation change between the diastereomers), 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 enantiomers of ibuprofen and thalidomide in a β -cyclodextrin adapted β -HL nanopore.^[84] 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 6A). In addition, one diastereomer of Sp would give a current level similar to C, while the other could potentially be identified (Figure 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 6A). It must be noted that this only holds true in sequence contexts that give current levels similar to poly-dC in WT- β -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 6B) for which point mutations have been shown to cause uncontrolled cell growth and loss of cell differentiation triggering various human adenocarcinomas.^[85] In lung cancer, the *K-ras* gene often undergoes a G to T transversion mutation that might result from unrepaired G oxidative damage.^[86] Therefore, the critical G was positioned at 14 (Figure 6B) to determine the current levels of the canonical nucleotide vs. its oxidatively-damaged forms. The histogram in Figure 6B shows that OG, Sp and Gh can be distinguished from G in this sequence.^[82] 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.^[82] This is a major limitation associated with the WT- β -HL nanopore; because of the three sensing zones, different sequence contexts can lead to overlapping current distributions.^[49] Consequently, the investigation of protein ion channel containing mutated sites that optimize the current difference between nucleotides is an important area of investigation.^[11]

Another type of oxidative damage caused by ionizing radiation is thymine glycol, Tg (Figure 7). Tg occurs as a pair of diastereomers and the most common one detected is the *cis*-(5R, 6S).^[66] Tg, unlike the native nucleotides, is also non-planar.^[66] Figure 7 shows the structure and the current level blockages for Tg at either position 13, 14, 15, or 16 within a 40mer poly-dC DNA strand. The current level of Tg falls within the standard nucleotide current levels and would likely be incorrectly assigned in a sequencing experiment with WT- β -HL, although Tg can be correctly distinguished from a T nucleotide at any of the four positions studied (Figure 7B). In this experiment, the biggest difference between the modified and unmodified DNA was observed at position 15. Furthermore, when Tg was placed at 15, 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,^[66] 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 (15, inset Figure 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 6A and 6B).

4. Current levels for hydrolytic DNA damage

Aside from oxidative DNA damage, DNA can undergo spontaneous deamination or depurination. The structures and current level histograms for the U and AP residues are presented in Figure 8. Taking into account how much smaller the AP site is comparing to the native nucleotide, it is surprising to see that the current level blockage of this damage is the same as for G (Figure 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 (spontaneous depurination alone causes formation of 10,000 damaged sites per genome, Table 1), 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 14 position of the DNA. Unlike the previous studies^[82], incorporation of U and AP in a heterosequence (K-*ras*) leads to a similar pattern as in the poly-dC sequence context.^[87]

In Figure 9 the structure of *cis, syn*-cyclobutane thymine dimer (T=T) is shown together with its blockage current level histogram in an immobilization experiment. The T=T lesion blocks the current ~1% more than a DNA containing an undamaged TT sequence at positions 14, 15 (Figure 9A) making it feasible to distinguish the damage from the canonical nucleotides that block less than C₄₀. 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.^[67] 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 in Figure 9A show that the TT and T=T sequences can readily be resolved in an immobilization experiment, and a further experiment was conducted to monitor the photoreversal of T=T back to TT. In Figure 9B, the photoreversal^[88] 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, the reversal rate was considerably slower than predicted,^[89, 90] due to the absorbance of UV light by the added proteins. 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 base pK_a values for G and T.^[91]

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 6A and 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. This is the method adopted by Bayley and colleagues.^[11, 36] 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.

5. 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 that 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 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.^[92-94] 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 14) in the presence of a primary amine gives a spirocyclic product with a pendent amine (Figure 10).^[82] In these studies, the primary amines benzylamine (Bz, MW = 108), *N*-acetyl-*O*-methyllysine (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 10). 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 Spd 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) Spd is a polyamine that has an overall net positive charge, and (2) Spd 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 γ -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.

Apurinic/aprimidinic (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 11).^[87] Seven primary amines were tested as adducts to AP: 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 C₃₉GlcN₁₄). 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 11) or the Sp site (Figure 10). 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.^[87] All of the adducts presented in Figure 11 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 12).^[17] 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 α -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,^[95, 96] 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.^[97] Figure 13 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 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.

6. Alternative approach for determination of current level differences for non-native 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^[98] and it is non-randomly distributed throughout the genome.⁶⁴ Recent studies have shown that oxidation of 5-methylcytidine (5mC) to 5-hydroxymethylcytidine (5hC) may provide a further refinement in epigenetic regulation.^[99] 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.^[100] 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.^[101-103] A mutated MspA ion channel protein has also been shown to distinguish current levels of the native nucleotides in immobilization experiments.^[104] 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.^[35] 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.^[49, 102]

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. [32, 36, 105] 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.^[35] The ability to have tight control over the DNA polymerase has made this such 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 in this review. 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.

7. 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), 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.

Acknowledgments

We are grateful to the NIH, via grant numbers GM-093099 and HG-005095, for financial support. Furthermore, we appreciate thoughtful conversations with Dr. John Watkins from the University of Utah and Drs. Geoffrey Barrall and Anna Schibel at Electronic BioSciences (EBS), and the donation of EBS ion channel recording instrumentation used in the experiments described.

References

1. 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, X, Ling S, Mastrangelo C, H, Meller A, Oliver J, S, Pershin Y, V, Ramsey JM, Riehn R, Soni G, V, Tabard-Cossa V, Wanunu M, Wiggin M, Schloss J, A. *Nat. Biotechnol.* 2008; 26:1146. [PubMed: 18846088]
2. Deamer DW, Branton D. *Acc. Chem. Res.* 2002; 35:817. [PubMed: 12379134]
3. Rhee M, Burns MA. *Trends in Biotechnology.* 2006; 24:580. [PubMed: 17055093]
4. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. *Proc. Natl. Acad. Sci. U.S.A.* 1996; 93:13770. [PubMed: 8943010]
5. Bayley H. *Curr. Opinion Chem. Biol.* 2006; 10:628.
6. Song L, Haobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux EJ. *Science.* 1996; 274:1859. [PubMed: 8943190]

7. Japrun D, Henricus M, Li Q, Maglia G, Bayley H. *Biophys. J.* 2010; 98:1856. [PubMed: 20441749]
8. Meller A, Nivon L, Brandin E, Golovchenko J, Branton D. *Proc. Natl. Acad. Sci. U.S.A.* 2000; 97:1097.
9. Butler TZ, Gundlach JH, Troll M. *Biophys. J.* 2007; 93:3229. [PubMed: 17675346]
10. Meller A, Branton D. *Electrophoresis.* 2002; 23:2583. [PubMed: 12210161]
11. Stoddart D, Maglia G, Mikhailova E, Heron AJ, Bayley H. *Angew. Chem., Int. Ed.* 2010; 49:556.
12. de Zoysa RSS, Krishantha DMM, Zhao Q, Gupta J, Guan X. *Electrophoresis.* 2011; 32:3034. [PubMed: 21997574]
13. Fologea D, Uplinger J, Thomas B, McNabb DS, Li J. *Nano Lett.* 2005; 5:1734. [PubMed: 16159215]
14. Kawano R, Schibel AEP, Cauley C, White HS. *Langmuir.* 2008; 25:1233. [PubMed: 19138164]
15. de Zoysa RSS, Jayawardhana DA, Zhao Q, Wang D, Armstrong DW, Guan X. *J. Phys. Chem. B.* 2009; 113:13332. [PubMed: 19736966]
16. Mitchell N, Howorka S. *Angew. Chem., Int. Ed.* 2008; 47:5565.
17. An N, Fleming AM, White HS, Burrows CJ. *Proc. Natl. Acad. Sci. U.S.A.* 2012; 109:11504. [PubMed: 22711805]
18. Borsenberger V, Mitchell N, Howorka S. *J. Am. Chem. Soc.* 2009; 131:7530. [PubMed: 19441786]
19. Jin Q, Fleming AM, Burrows CJ, White HS. *J. Am. Chem. Soc.* 2012; 134:11006. [PubMed: 22690806]
20. Schibel AEP, Fleming AM, Jin Q, An N, Liu J, Blakemore CP, White HS, Burrows CJ. *J. Am. Chem. Soc.* 2011; 133:14778. [PubMed: 21875081]
21. Nakane J, Wiggins M, Marziali A. *Biophys. J.* 2004; 87:615. [PubMed: 15240494]
22. Lee, JW.; Meller, A. *Book Rapid DNA sequencing by direct nanoscale reading of nucleotide bases on individual DNA chains. Vol. 2. Elsevier; City: 2007. p. 245* Chap. Chapter
23. Sauer-Budge AF, Nyamwanda JA, Lubensky DK, Branton D. *Phys. Rev. Lett.* 2003; 90:238101. [PubMed: 12857290]
24. Viasnoff V, Chiaruttini N, Bockelmann U. *Eur. Biophys. J.* 2009; 38:263. [PubMed: 18836709]
25. Sutherland TC, Dinsmore MJ, Kraatz H-B, Lee JS. *Biochem. Cell Biol.* 2004; 82:407. [PubMed: 15181475]
26. Muzard J, Martinho M, Mathé J, Bockelmann U, Viasnoff V. *Biophys. J.* 2010; 98:2170. [PubMed: 20483325]
27. Liu A, Zhao Q, Krishantha DMM, Guan X. *J. Phys. Chem. Lett.* 2011; 2:1372. [PubMed: 21709813]
28. Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S, Bayley H. *Nat. Nano.* 2009; 4:265.
29. Wu H-C, Astier Y, Maglia G, Mikhailova E, Bayley H. *J. Am. Chem. Soc.* 2007; 129:16142. [PubMed: 18047341]
30. Astier Y, Braha O, Bayley H. *J. Am. Chem. Soc.* 2006; 128:1705. [PubMed: 16448145]
31. Banerjee A, Mikhailova E, Cheley S, Gu L-Q, Montoya M, Nagaoka Y, Gouaux E, Bayley H. *Proc. Natl. Acad. Sci. U.S.A.* 2010; 107:8165. [PubMed: 20400691]
32. Cockroft SL, Chu J, Amorin M, Ghadiri MR. *J. Am. Chem. Soc.* 2008; 130:818. [PubMed: 18166054]
33. Hornblower B, Coombs A, Whitaker RD, Kolomeisky A, Picone SJ, Meller A, Akeson M. *Nat. Meth.* 2007; 4:315.
34. Benner S, Chen RJA, Wilson NA, Abu-Shumays R, Hurt N, Lieberman KR, Deamer DW, Dunbar WB, Akeson M. *Nat. Nano.* 2007; 2:718.
35. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH. *Nat. Biotechnol.* 2012; 30:349. [PubMed: 22446694]
36. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. *Nat. Biotechnol.* 2012; 30:344. [PubMed: 22334048]

37. Kumar S, Tao C, Chien M, Hellner B, Balijepalli A, Robertson JWF, Li Z, Russo JJ, Reiner JE, Kasianowicz JJ, Ju J. *Sci. Rep.* 2012;2.
38. Rincon-Restrepo M, Mikhailova E, Bayley H, Maglia G. *Nano Lett.* 2011; 11:746. [PubMed: 2122450]
39. Maglia G, Restrepo MR, Mikhailova E, Bayley H. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:19720. [PubMed: 19060213]
40. Henrickson SE, Misakian M, Robertson B, Kasianowicz JJ. *Phys. Rev. Lett.* 2000; 85:3057. [PubMed: 11006002]
41. Purnell, R.; Schmidt, J. *Nanopore-Based Technology*. Gracheva, ME., editor. Vol. 870. Humana Press; 2012. p. 39Chap. 3
42. Howorka S, Bayley H. *Biophys. J.* 2002; 83:3202. [PubMed: 12496089]
43. Ayub M, Bayley H. *Nano Lett.* 2012; 12:5637. [PubMed: 23043363]
44. Hyre DE, Le Trong I, Merritt EA, Eccleston JF, Green NM, Stenkamp RE, Stayton PS. *Protein Sci.* 2006; 15:459. [PubMed: 16452627]
45. Mathé J, Aksimentiev A, Nelson DR, Schulten K, Meller A. *Proc. Natl. Acad. Sci. U.S.A.* 2005; 102:12377. [PubMed: 16113083]
46. Purnell RF, Mehta KK, Schmidt JJ. *Nano Lett.* 2008; 8:3029. [PubMed: 18698831]
47. Martin HSC, Jha S, Howorka S, Coveney PV, Chem J. *Theory Comput.* 2009; 5:2135.
48. Ashkenasy N, Sánchez-Quesada J, Bayley H, Ghadiri MR. *Angew. Chem., Int. Ed.* 2005; 44:1401.
49. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106:7702. [PubMed: 19380741]
50. Purnell RF, Schmidt JJ. *ACS Nano.* 2009; 3:2533. [PubMed: 19694456]
51. Møller P, Folkmann JK, Forchhammer L, Bräuner EV, Danielsen PH, Risom L, Loft S. *Cancer Lett.* 2008; 266:84. [PubMed: 18367322]
52. Fu D, Calvo JA, Samson LD. *Nat. Rev. Cancer.* 2012; 12:104. [PubMed: 22237395]
53. Pluskota-Karwatka D. *Bioorg. Chem.* 2008; 36:198. [PubMed: 18561974]
54. Cadet J, Douki T, Ravanat J-L. *Acc. Chem. Res.* 2008; 41:1075. [PubMed: 18666785]
55. Luo W, Muller JG, Rachlin EM, Burrows CJ. *Org. Lett.* 2000; 2:613. [PubMed: 10814391]
56. Ye Y, Muller JG, Luo W, Mayne CL, Shalloo AJ, Jones RA, Burrows CJ. *J. Am. Chem. Soc.* 2003; 125:13926. [PubMed: 14611206]
57. Fleming AM, Muller JG, Dlouhy AC, Burrows CJ. *J. Am. Chem. Soc.* 2012; 134:15091. [PubMed: 22880947]
58. Burrows CJ, Muller JG. *Chem. Rev.* 1998; 98:1109. [PubMed: 11848927]
59. Neeley WL, Essigmann JM. *Chem. Res. Toxicol.* 2006; 19:491. [PubMed: 16608160]
60. Gedik CM, Collins A. *FASEB J.* 2004; 19:82. [PubMed: 15533950]
61. Korniyushyna O, Berges AM, Muller JG, Burrows CJ. *Biochemistry.* 2002; 41:15304. [PubMed: 12484769]
62. Henderson PT, Delaney JC, Muller JG, Neeley WL, Tannenbaum SR, Burrows CJ, Essigmann JM. *Biochemistry.* 2003; 42:9257. [PubMed: 12899611]
63. Mangerich A, Knutson CG, Parry NM, Muthupalani S, Ye W, Prestwich E, Cui L, McFaline JL, Mobley M, Ge Z, Taghizadeh K, Wishnok JS, Wogan GN, Fox JG, Tannenbaum SR, Dedon PC. *Proc. Natl. Acad. Sci. U.S.A.* 2012; 109:E1820. [PubMed: 22689960]
64. Shikazono N, Noguchi M, Fujii K, Urushibara A, Yokoya A. *J. Radiat. Res.* 2009; 50:27. [PubMed: 19218779]
65. Cathcart R, Schwiers E, Saul RL, Ames BN. *Proc. Natl. Acad. Sci. U.S.A.* 1984; 81:5633. [PubMed: 6592579]
66. Dolinnaya NG, Kubareva EA, Romanova EA, Trikin RM, Oretskaya TS. *Biochimie.* 2013; 95:134. [PubMed: 23000318]
67. Sinha RP, Hader DP. *Photochem. Photobiol. Sci.* 2002; 1:225. [PubMed: 12661961]
68. Batista LFZ, Kaina B, Meneghini R, Menck CFM. *Mutat. Res.* 2009; 681:197. [PubMed: 18845270]

69. Norval M, Lucas RM, Cullen AP, de Gruijl FR, Longstreth J, Takizawa Y, van der Leun JC. *Photochem. Photobiol. Sci.* 2011; 10:199. [PubMed: 21253670]
70. Cordonnier AM, Fuchs RPP. *Mutat. Res.* 1999; 435:111. [PubMed: 10556591]
71. Garinis GA, van der Horst GTJ, Vijg J, Hoeijmakers JHJ. *Nat. Cell Biol.* 2008; 10:1241. [PubMed: 18978832]
72. Boiteux S, Guillet M. *DNA Repair.* 2004; 3:1. [PubMed: 14697754]
73. Lindahl T. *Nature.* 1993; 362:709. [PubMed: 8469282]
74. Hübscher U, Maga G, Spadari S. *Annu. Rev. Biochem.* 2002; 71:133. [PubMed: 12045093]
75. Goodman FM. *Annu. Rev. Biochem.* 2002; 71:17. [PubMed: 12045089]
76. Kroeger KM, Goodman MF, Greenberg MM. *Nucl. Acids Res.* 2004; 32:5480. [PubMed: 15477395]
77. House CH, Miller SL. *Biochemistry.* 1996; 35:315. [PubMed: 8555190]
78. Krokan EK, Drabløs F, Slupphaug G. *Oncogene.* 2002; 21:8935. [PubMed: 12483510]
79. Jackson AL, Loeb LA. *Mutat. Res.* 2001; 477:7. [PubMed: 11376682]
80. Zhang B, Galusha J, Shiozawa PG, Wang G, Bergren AJ, Jones RM, White RJ, Ervin EN, Cauley CC, White HS. *Anal. Chem.* 2007; 79:4778. [PubMed: 17550232]
81. White RJ, Ervin EN, Yang T, Chen X, Daniel S, Cremer PS, White HS. *J. Am. Chem. Soc.* 2007; 129:11766. [PubMed: 17784758]
82. Schibel AEP, An N, Jin Q, Fleming AM, Burrows CJ, White HS. *J. Am. Chem. Soc.* 2010; 132:17992. [PubMed: 21138270]
83. Herskovits TT, Bowen JJ. *Biochemistry.* 1974; 13:5474. [PubMed: 4549016]
84. Kang, X.-f.; Cheley, S.; Guan, X.; Bayley, H. *J. Am. Chem. Soc.* 2006; 128:10684. [PubMed: 16910655]
85. Pfeifer G, Besaratinia A. *Human Genetics.* 2009; 125:493. [PubMed: 19308457]
86. David SS, O'Shea VL. *Nature.* 2007; 447:941. [PubMed: 17581577]
87. An N, White HS, Burrows CJ. *Chem. Commun.* 2012; 48:11410.
88. Pan Z, Chen J, Schreier WJ, Kohler B, Lewis FD. *J. Phys. Chem. B.* 2011; 116:698. [PubMed: 22103806]
89. Garcés F, Dávila CA. *Photochem. Photobiol.* 1982; 35:9. [PubMed: 7043501]
90. Johns HE, Rapaport SA, Delbruck M. *Mol J. Biol.* 1962; 4:104.
91. Franceschini L, Mikhailova E, Bayley H, Maglia G. *Chem. Commun.* 2012; 48:1520.
92. Hosford ME, Muller JG, Burrows CJ. *J. Am. Chem. Soc.* 2004; 126:9540. [PubMed: 15291548]
93. Xu X, Muller JG, Ye Y, Burrows CJ. *J. Am. Chem. Soc.* 2007; 130:703. [PubMed: 18081286]
94. Xue L, Greenberg MM. *J. Am. Chem. Soc.* 2007; 129:7010. [PubMed: 17497789]
95. Jacobs A, Schär P. *Chromosoma.* 2012; 121:1. [PubMed: 22048164]
96. David SS, Williams SD. *Chem. Rev.* 1998; 98:1221. [PubMed: 11848931]
97. An N, Fleming AM, Rosecrans NC, Liao Y, Burrows CJ. *Methods Mol. Biol.* 2013 Submitted.
98. Beck S, Rakyán VK. *Trends in Genetics.* 2008; 24:231. [PubMed: 18325624]
99. Fu Y, He C. *Curr. Opinion Chem. Biol.* 2012; 16:516.
100. Wallace EVB, Stoddart D, Heron AJ, Mikhailova E, Maglia G, Donohoe TJ, Bayley H. *Chem. Commun.* 2010; 46:8195.
101. Robertson JWF, Kasianowicz JJ, Banerjee S. *Chem. Rev.* 2012; 112:6227.
102. Butler TZ, Pavlenok M, Derrington IM, Niederweis M, Gundlach JH. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:20647. [PubMed: 19098105]
103. Chen M, Khalid S, Sansom MSP, Bayley H. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:6272. [PubMed: 18443290]
104. Manrao EA, Derrington IM, Pavlenok M, Niederweis M, G. J. H. *PLoS ONE.* 2011:6.
105. Lieberman KR, Cherf GM, Doody MJ, Olasagasti F, Kolodji Y, Akeson M. *J. Am. Chem. Soc.* 2010; 132:17961. [PubMed: 21121604]

Figure 1. Structure of the heptameric bacterial toxin α -HL(pdb 7AHL)^[6] 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.

Figure 2.

Schematic representation of the immobilization experiment, in which 3' biotinylated DNA is bound to streptavidin (pdb 1MK5).^[44] 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.

Figure 3.

Sensing zone mapping experiments for WT-βHL. Left: 3'-Biotinylated poly-dC DNAs with a single A nucleotide at position 7 to 19. Middle: residual current level differences ($\Delta I_{RES}(\%)$) between the DNAs. Right: representation of βHL (pdb 7AHL)^[6]/streptavidin (pdb 1MK5)^[44] Btn-DNA complex with the approximate locations of the single A base in the protein channel. R₁, R₂, R₃ are the observed sensing zones along the β-barrel of the protein. The figure was adapted from Stoddard, *et al.*^[49]

Figure 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.

Figure 5.

$\% I/I_0$ histograms representing the current level blockage of G, C, A, and T, at position 14 in a C₄₀ 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 nucleotides are shown on top. The figure was adapted from Schibel, *et al.*^[82]

Figure 6.

Relative percent residual current histograms for G-oxidation products. **A.** Top: structures of G, OG, Sp, Gh. Bottom: $\% I/I_0$ histograms, representing the current level blockages for G, Gh, OG, Sp at a position 14 (biotinylated at the 3' end) in a polydC DNA, in which C₄₀ was the internal standard. **B.** $\% I/I_0$ histograms, representing the current level blockage of G, Gh, OG, Sp at position 14 in the K-ras sequence in a poly-dT background. The internal standard was C₄₀ 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.*^[82]

Figure 7.

Current level histograms for Tg-containing DNA. **A.** Two diastereomers of Tg that were present in the DNA strand studied. % I/I_0 histogram of Tg at position 15 showing two current populations. **B.** % I/I_0 histograms, representing the blockage current level of thymine glycol (Tg) at position 13 through 16. The Tg phosphoramidite was introduced at positions 13 to 16 in a DNA that was poly-dC except of T at those positions labeled X (sequence above). C₃₆T₁₃₋₁₆ 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). (Unpublished results, L. He.)

Figure 8.

% I/I_0 histograms, representing the current level blockage of U and AP at the position 14 where C₄₀ 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). The figure was adapted from An, *et al.*^[87]

Figure 9.

Current level histograms for thymine dimer containing DNA. **A.** % I/I_0 histograms, representing the current level blockage of TT and T=T at a position 14 and 15. **B.** Photoreversal experiment. The starting material was 3' biotinylated C₃₈ with a thymine dimer at positions 14 and 15 (green square) that was irradiated with UV light to convert the DNA damage back to thymines at those positions. The irradiation experiment was conducted within the nanopore cell holder with streptavidin, -HL and lipid present in 1 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. (Unpublished results, A. H. Wolna.)

Figure 10.

Spirocyclic adducts to OG. Top: the structures of the spirocyclic adducts: Bz, Lys, GlcN, Spd, Spm, and GPRP. Bottom: $\%I/I_0$ histograms showing all the previously mentioned adducts in a poly-dC background at position 14 where the internal standard was C₄₀. 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.*^[82]

Figure 11.

DNA abasic site adducts. Top: the structures of the AP site adducts: Trn, GlcN, RH, GPRP, STM, 15c5 and 18c6. Bottom: %I/I₀ histograms showing all the previously mentioned adducts in a poly-dC background at position 14, in which the internal standard was C₄₀. 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.*^[87]

Figure 12.

Translocation current vs. time trace for two AP 18c6 adducts 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 α -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, which produced a deeper blocking current level I_{B1} due to the presence of the bulky adduct. Once the first adduct exited the protein α -barrel to its *trans* side, the current level returned back to I_{A2} , followed by another 18c6 translocation where the current level increased to I_{B2} , and finally, once the second adduct exited the α -barrel, the current returned to I_{A3} , representing the translocation of the rest of the ssDNA. The figure was adapted from An, *et al.*^[17]

Figure 13.

Current blockage levels for the 18c6 adduct (5-18c6-C) and 5-iodocytidine (5-I-C). Top: reaction scheme for the 5-18c6-C formation followed by the structures of the 5-18c6-C and 5-I-C. Bottom: $\%I/I_0$ histograms showing mentioned adducts in a poly-dC background at position 14, in which the internal standard was C₄₀ that was set to 0 $\%I/I_0$. The experiment was conducted in 1 M NaCl, 25 mM Tris, and 1 mM EDTA (pH 7.9). (Unpublished results, N. An.)

Table 1

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

DNA modification	Approximate Number in human genome
UV photoproducts	100,000 ^[71]
Tg	400 ^[66]
Spontaneous depurination (AP sites)	10,000 ^[79]
OG	6,400 ^[60]
Sp	60 ^{a[63]}
Gh	60 ^{a[63]}
U	500 ^[78]

^aValues were calculated from data determined in a mouse model