

UC Riverside

UC Riverside Previously Published Works

Title

Detection and Discrimination of DNA Adducts Differing in Size, Regiochemistry, and Functional Group by Nanopore Sequencing.

Permalink

<https://escholarship.org/uc/item/1jw947vf>

Journal

Chemical Research in Toxicology, 33(12)

Authors

Nookaew, Intawat
Jenjaroenpun, Piroon
Du, Hua
et al.

Publication Date

2020-12-21

DOI

10.1021/acs.chemrestox.0c00202

Peer reviewed



HHS Public Access

Author manuscript

Chem Res Toxicol. Author manuscript; available in PMC 2021 December 21.

Published in final edited form as:

Chem Res Toxicol. 2020 December 21; 33(12): 2944–2952. doi:10.1021/acs.chemrestox.0c00202.

Detection and Discrimination of DNA Adducts Differing in Size, Regiochemistry, and Functional Group by Nanopore Sequencing

Intawat Nookaew¹,

Department of Biomedical Informatics and Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, United States

Piroon Jenjaroenpun¹,

Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, United States

Hua Du,

Department of Chemistry, University of California, Riverside, California 92521-0403, United States

Pengcheng Wang,

Department of Chemistry, University of California, Riverside, California 92521-0403, United States

Jun Wu,

Department of Chemistry, University of California, Riverside, California 92521-0403, United States

Thidathip Wongsurawat,

Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, United States

Sun Hee Moon,

Environmental and Occupational Health, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, United States

En Huang,

Environmental and Occupational Health, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, United States

Yinsheng Wang,

Department of Chemistry, University of California, Riverside, California 92521-0403, United States

Corresponding Authors Intawat Nookaew inookaew@uams.edu, Gunnar Boysen gboysen@uams.edu.

¹These authors contributed equally to this work.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.chemrestox.0c00202>.

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.chemrestox.0c00202>

The authors declare no competing financial interest.

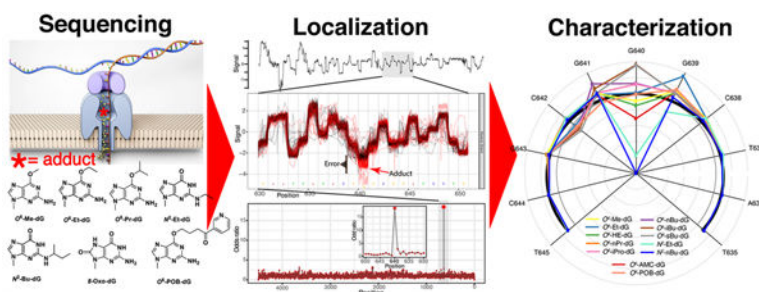
Gunnar Boysen

Environmental and Occupational Health and Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, United States

Abstract

Chemically induced DNA adducts can lead to mutations and cancer. Unfortunately, because common analytical methods (e.g., liquid chromatography-mass spectrometry) require adducts to be digested or liberated from DNA before quantification, information about their positions within the DNA sequence is lost. Advances in nanopore sequencing technologies allow individual DNA molecules to be analyzed at single-nucleobase resolution, enabling us to study the dynamic of epigenetic modifications and exposure-induced DNA adducts in their native forms on the DNA strand. We applied and evaluated the commercially available Oxford Nanopore Technology (ONT) sequencing platform for site-specific detection of DNA adducts and for distinguishing individual alkylated DNA adducts. Using ONT and the publicly available ELIGOS software, we analyzed a library of 15 plasmids containing site-specifically inserted O^6 - or N^2 -alkyl-2'-deoxyguanosine lesions differing in sizes and regiochemistries. Positions of DNA adducts were correctly located, and individual DNA adducts were clearly distinguished from each other.

Graphical Abstract



INTRODUCTION

Humans are constantly exposed to endogenous and exogenous sources of genotoxic compounds, which can give rise to adducts with DNA directly or after metabolic activation.¹⁻⁴ The resulting DNA adducts, if not repaired in time, can induce mutations during DNA replication and, ultimately, lead to cancer.⁵ Detection and characterization of DNA adducts provide an important foundation for elucidating the causal chain of events from exposure to DNA adducts and from DNA adducts to mutations.⁶⁻⁸

Recent advances in mass spectrometry instrumentation and methods have enabled highly specific and sensitive detection of DNA adducts, which allows for investigation of the formation and repair of DNA adducts.^{1,2,9} These methods, however, do not provide information about the specific locations of DNA adducts within the genome. To overcome this limitation, Burrows and colleagues pioneered the applications of nanopore technology to detect DNA modifications. Using α/γ -hemolysin-based nanopores, they demonstrated detection, at single-nucleotide resolution, of DNA adducts and other types of DNA

modifications, including *N*²-benzo[*a*]pyrene diolepoxide-2'-deoxyguanosine,¹⁰ 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxo-dG),¹¹ abasic sites,¹² 5-guanidinohydantoin,¹³ 2'-deoxyinosine,¹⁴ and DNA mismatch sites.¹⁵ Laszlo et al. subsequently reported the application of engineered nanopore porin protein to detect two epigenetic marks, 5-methyl-2'-deoxycytidine and 5-hydroxymethyl-2'-deoxycytidine.¹⁶ In addition, mutant forms of the nanopore porin proteins have been successfully used to discriminate signals from five C5-modified 2'-deoxycytidine derivatives.¹⁷ Moreover, Wang et al. used a *Mycobacterium smegmatis* porin A nanopore to correctly identify *O*⁶-carboxymethyl-2'-deoxyguanosine.¹⁸ Those studies provided concrete foundations for using protein-based nanopores to detect individual preselected DNA modifications. The reported studies used custom-engineered nanopores and mainly focused on short DNA molecules (synthetic oligodeoxyribonucleotides) or a single DNA adduct type within the genome.

The principle underlying nanopore technology is that electrochemical forces are used to pull single-stranded DNA in native form through the tiny pores (Scheme 1); the accompanying change in electric current indicates the physiochemical properties of the DNA bases transiting through the pore, revealing the DNA sequence and any DNA modifications, such as DNA adducts.¹⁹ Oxford Nanopore Technology (ONT) developed and commercialized an engineered protein nanopore for nucleic acid sequencing that has the capability to sequence long to ultralong DNA molecules (>2 Mb) in their native form, which preserves positional information on any DNA modifications.²⁰ Moreover, it was recently demonstrated that some epigenetic modifications, including 5-methyl-2'-deoxycytidine and *N*⁶-methyl-2'-deoxyadenosine, can be detected by ONT within the context of genome-wide investigation.

Herein, we report the application and evaluation of the commercially available ONT sequencing platform for site-specific detection of DNA adducts and for distinguishing alkylated 2'-deoxyguanosine adducts that differ in alkyl chain length, structure, and regiochemistry. With a library of synthetic DNA plasmids that contain site-specific and regioisomeric *O*⁶- or *N*²-alkyl-dG DNA adducts (Table 1), we demonstrated the suitability of the ONT platform and the ELIGOS tool for locating, identifying, and distinguishing DNA adducts at single-nucleotide resolution within a given sequence context.

MATERIAL AND METHODS

Construction of Synthetic Plasmid DNA.

The DNA adduct-containing plasmids used in this study were constructed previously (Table 1).^{4,25–32} The DNA adduct of interest was inserted at position G640 in the sequence TGGCGGGCTAT of the pTGFP-Hha10 shuttle vector, which includes an SV40 replication origin as described previously.³³ In addition, *O*⁶-Me-dG* was inserted at position G1414 (G*) in the sequence TTATA^{Me}G*CTATT of a pPtPBR11-derived vector (provided by Dr. Robert Fuchs, Inserm, France). As a representative epigenetic mark, 5-formyl-2'-deoxycytidine (5-Fm-dC) was inserted at position C638 in the sequence GCGGG^{Fm}CTATTC of a slightly modified pTGFP-Hha10 shuttle vector (provided by Dr. Natalia Tretyakova, University of Minnesota, Minnesota, USA).⁴ Lastly, representative of oxidative stress-derived DNA modification, 8-oxo-dG was inserted at position G6235 in the

sequence CTTAA^{oxo}GCTCGAG of an M13mp18 plasmid (provided by Dr. Colin Campbell, University of Minnesota, Minnesota, USA).³⁴ In this study, all adducts were located on the negative strand, so nucleotide positions are shown clockwise from the 5' to 3' direction in the figures.

DNA Sequencing and Data Acquisition.

Approximately 100 ng of adduct-containing plasmids was used as the input for preparing the DNA sequencing library with the Rapid Sequencing Kit (SQK-RBK004; Oxford Nanopore Technologies [ONT], Oxford Science Park, UK), following the manufacturer's protocol. The constructed libraries were loaded onto the R9.5/FLO-MIN107 flow cell equipped on the MinION Mk1B (ONT). The sequencing and data acquisition were performed under MinKNOW software v18.12.6 (ONT) to generate a .fast5 file containing raw ionic signal for individual molecules (i.e., reads). All data generated in this study were deposited in the NIH's Sequence Read Archive database under Bioproject PRJNA615636.

Data Analysis for Adduct Localization.

The workflow to analyze DNA adducts by using ELIGOS³⁵ is summarized in Figure S1. The first step in our analysis was to localize the adduct position by examining error frequencies of the base caller script at each position. The statistical analysis and plots were performed with R suite. We used the ELIGOS tool³⁵ to compute the sequencing errors of individual bases and to compare the differences in error fractions with Fisher's exact test, producing odds ratios for individual nucleotide positions. The odds ratios that were significantly different from those of the corresponding positions in the control DNA were calculated to identify the adduct locations (Figures S2 and S3). ELIGOS is available at <https://gitlab.com/piroonj/eligos2>.

Adduct Localization.

To identify the position of a DNA adduct, the error at specific base (ESB) profiles was first calculated using ELIGOS tool.³⁵ ESB is defined as the frequency of the sum of substitutions, insertions, and deletions of individual positions over the total mapped reads obtained from read alignment results based on the reference sequence.³⁵ The raw signal files in .fast5 format (obtained from Minknow software, v18.12.6; ONT) were base-called with the guppy software v3.2.4 (ONT) to generate raw .fastq files. The NanoFilt software v2.5³⁶ was used to filter out reads shorter than 200 bp. The resulting filtered .fastq files were aligned to the plasmid reference sequences with Minimap2 v2.16³⁷ and were converted to BAM files with Samtools software v1.6.³⁸ The BAM files were used to identify sequencing errors, which included substitutions, insertions, and deletions and to calculate the sequencing error rates of individual positions. Those in the adduct plasmid were compared with those in the control plasmid by using Fisher's exact test to generate odds ratios, adjusted *p* values (i.e., Benjamini–Hochberg method),³⁹ and ESB.

Adduct Characterization.

The second step was to characterize the adduct-specific ion signals. Because the ion signal is affected by all bases that pass through the pore at the time of measurement, we characterized

the adduct-specific distortion in a region identified by the previous step that included the five bases before and five bases after the DNA adduct. The raw squiggle signals were resquiggled with Tombo software v1.5 (ONT), and the noise of the resquiggled signal was subsequently reduced with Box-Cox transformation⁴⁰ of the resquiggled signal for individual positions at relevant loci. The transformed signals (denoised) derived from adduct-containing plasmids were compared with the control plasmid to identify differential ionic signals by using Student's *t* test and fold-change for calculation of π statistical value (following methods of Xiao et al.).⁴¹ Characteristic DNA adduct-induced distortions were visualized as radar plots of the π values and called differential ionic signal (DIS) plots.

Assessments of Adduct Detection by ONT.

The resquiggle signals of reads mapped on 11 nucleotides of positions 635–645 of pTGFP-Hha10 as used in the DIS plots were used for the assessments of adduct discrimination by ONT in three scenarios. (1) The discrimination between an individual adduct and its corresponding unmodified nucleotide (control) was examined on the basis of the pattern of signals. The results were then used to estimate false positive rate and false negative rate using receiver operating characteristic curve (ROC) analysis. (2) The assessment of discrimination between the pairs of regioisomeric *N*²- and *O*⁶-Et-dG and *N*²- and *O*⁶-*n*Bu-dG were performed at different mixtures of reads of adduct- and dG-containing plasmids. (3) We estimated the detection level of an individual DNA adduct using ELIGOS to calculate *p* value based on Fisher's exact test by comparing *in silico* mixtures (i.e., by mixing reads for adduct-containing plasmids with those of the control) at different percentages of DNA adduct in mixture with control reads. The analysis was performed based on 10,000 reads per mixture. We conducted the analysis of the *in silico* mixtures for 30 random samplings for each individual mixture, and the mean of them is presented Figure 5C.

RESULTS

Locations of DNA Adducts Are Revealed by Disturbances in Ionic Signal.

Based on the ONT sequencing principle, double-stranded DNA is first unzipped by a DNA helicase and is then translocated at a controlled speed through the nanopore. The translocation of the DNA strand is driven by an ionic current passing through the nanopore (Scheme 1).¹⁹ The transient DNA strand reduced, and sometimes temporarily blocked, the ion current as it simultaneously passed through the pore. These changes in ion current lead to alterations, known as squiggles, from the baseline signal (illustrated in Figure 1A), and they are influenced by the size and chemical properties of the transitioning DNA. As a result, the ion current is affected by modified nucleotide and two to three of its neighboring nucleotides that are all present in the pore at the time of measurement (Scheme 1).¹⁹ In Figure 1B, the resquiggled signals of an individual read derived from either control plasmid (black lines) and an adduct-containing plasmid (red lines) are overlaid for comparison.

The signals obtained from ONT sequencing normally are fuzzy and even after resquiggle preprocessing (Figure 1B). For this particular example, the reads obtained from the synthetic *O*⁶-Me-dG-containing plasmid had lower signal levels (red squiggle line) at G1414 than the reads from the corresponding control plasmid. Furthermore, a distinctively different signal

was observed for vicinal nucleotides of the \mathcal{O}^6 -Me-dG adduct. The distinct signals typically gave rise to a sequencing error during the standard base-calling algorithm, which is generally derived from unmodified nucleotide sequences.

The ELIGOS tool^{35,42} (see Supporting Information for details and summary in Figure S1) was used to calculate the sequencing error at specific base (ESB) from the profiles of reads obtained at each position from the \mathcal{O}^6 -Me-dG*-containing plasmids and the corresponding unmodified dG-bearing control plasmids. Subsequently, odds ratios for all nucleotide positions were computed with Fisher's exact test, comparing the error of reads derived from adduct-containing plasmid with that derived from the corresponding control plasmid (Figure 1C). The odds ratio for the error significantly increases at the position of the DNA adduct, indicating that the signal diverges from that of the non-adducted bases, thereby correctly locating the DNA adduct within the model sequence studied (see Supporting Information, Figures S2 and S3).

Characterizing DNA Adduct-Specific Disturbances to Ionic Signal.

The positions of DNA adducts can be identified by examining the mean ESB profile across the 11 nucleotide sequence (i.e., the adducted nucleotide and 5 nucleotides each on its 5' and 3' sides) in the region with high odds ratio (Figure 1C). The signal shown in the radar plot around the site of \mathcal{O}^6 -Me-dG in the adduct-containing plasmid is higher than that in the control plasmid, especially at the \mathcal{O}^6 -Me-dG position and one or two neighboring 3' bases (Figure 1D). To investigate the pattern of signal alterations at the loci that elicit sequencing errors, we first transformed the resquiggle signals of individual positions by using the Box-Cox method to reduce the noisy signal behaviors obtained from ONT.⁴⁰ We then used the π statistics⁴¹ to evaluate differential ionic signal (DIS), that is, differential changes in the direction of ionic signal between adducted and control plasmids (Figure 1E).

Disturbances in Ionic Signal by Adducts Are Dependent on Sequence Context.

To determine the potential effects of the DNA sequence context, we used the ONT/ELIGOS platform,^{35,42} to analyze plasmids containing the simplest \mathcal{O}^6 -alkyl-dG adduct, \mathcal{O}^6 -Me-dG, in two different sequences, TTATA^{Me}G*CTATT and TGGCG^{me}GGCTAT. Interestingly, even though the two synthetic plasmids contained the same \mathcal{O}^6 -Me-dG, the characteristic disturbances in signal had different patterns and also were in opposite directions (Figures 1D and E). The patterns of DIS plots and ESB profiles were not exactly at the same position due to the base-calling algorithm, which harnesses the information on neighboring nucleotides for interpretation.

Structural Isomers of Adducts Had Similar ESB Profiles and DIS Plots.

The ESB profile and the DIS plot shown as radar plots of the three larger Et, Pr, and Bu adducts were compared, as illustrated in Figure 2, to evaluate the impact of alkyl chain lengths and structural isomers on the alterations to ion signal of the ONT platform. The ESB profile of the \mathcal{O}^6 -Et-dG adduct (Figure 2A, left panel) shows an error that spans three nucleotides, including the adduct position (G640) and the two neighboring nucleotides in the 3' direction (G639 and C638). In contrast, \mathcal{O}^6 -hydroxyethyl-dG (\mathcal{O}^6 -HE-dG) showed a higher ESB fraction only at the G639 position. As DNA containing the \mathcal{O}^6 -Et-dG adduct

moved through the pore, the ionic current increased at G639 and eventually decreased at G640 and G642 (Figure 2A, right panel). In comparison, however, the O^6 -HE-dG adduct elicited a much more pronounced reduction in ion signal at G640. The difference in ionic currents might be attributed to the presence of the hydrophilic hydroxyl group in the O^6 -Et-dG adduct, which modulates the ion current as the lesion transits through the nanopore (Scheme 1).

We next investigated the effects of structural isomers of O^6 -alkyl-dG adducts that have longer alkyl chains, such as O^6 -Pr-dG and O^6 -Bu-dG. Introduction of O^6 -*n*Pr-dG led to an ESB fraction (Figure 2B, left panel) and ion blockage at C638 and C643 (Figure 2B, right panel) that were greater than those for O^6 -*i*Pr-dG, possibly because *n*Pr has a longer alkyl chain than *i*Pr. Similarly, different structural isomers of O^6 -Bu-dG adducts (i.e., O^6 -*n*Bu-dG, O^6 -*i*Bu-dG, and O^6 -*s*Bu-dG) exhibited unique signal profiles. The results demonstrated a strong adduct-dependent blockage at C642 and, to a lesser extent, at C638 (Figure 2C, left panel). At the other two positions, G639 and G640, the ESB fractions were similar for the different butyl isomers. The changes in ionic signal pattern for O^6 -*n*Bu-dG were distinct from those for the other two structural isomers, O^6 -*i*Bu-dG and O^6 -*s*Bu-dG, at G640 and G641 (Figure 2C, right panel). Notably, the latter two branched chain lesions demonstrated a common ionic blockage signature at G642 (Figure 2C, right panel).

Differential Impacts of Regioisomeric Lesions on the Ionic Current through the Nanopore.

Pairs of O^6 - and N^2 -conjugated Et and *n*Bu adducts were used to investigate the ways in which regioisomeric alkylation of guanine may affect ONT sequencing. The ESB profiles for the two regioisomers are clearly different (Figure 3A and B, left panels), and compared to the corresponding O^6 -alkyl-dG adducts, the N^2 -alkyl-dG adducts exhibited much larger ESB fractions at the two neighboring 3' nucleotides (C638 and G639). Furthermore, the O^6 -alkyl-dG adducts displayed a wider sequencing error than the corresponding N^2 -alkyl-dG adducts by including the adduct position (G640). These characteristics were evident for both pairs of Et-dG and *n*Bu-dG adducts.

The DIS plots (Figure 3A,B, right panels) indicated that ion flow through the nanopore was impeded more strongly by the N^2 -alkyl-dG adducts than by the corresponding O^6 -alkyl-dG adducts, especially at the position of the adduct itself (G640). The DIS plots show that, relative to the N^2 -alkyl-dG adducts, the O^6 -alkyl-dG lesions elicited a broader alteration in the signals by increasing the ion current through the pore at G639. For the Et adduct, this was accompanied by a reduction in current at G640 and C642 (Figure 3A, right panel) and, for the *n*Bu adduct, an augmented ion flow through the pore at G641 and a diminution at C642 (Figure 3B, right panel). These results are consistent with those demonstrated by ESB profiles.

We also investigated the effects of O^6 -aminocarbonylmethyl-2'-deoxyguanosine (O^6 -AMC-dG) and a bulkier adduct, O^6 -pyridyloxobutyl-2'-deoxyguanosine (O^6 -POB-dG). As illustrated in Figure 4A, O^6 -AMC-dG and O^6 -POB-dG can be clearly distinguished by both ESB profile and DIS plot. The largest adduct in the present study, O^6 -POB-dG modified the ionic blockage at three nucleotides, G640, G641, and C642. In addition, we analyzed plasmids containing epigenetic mark 5-formyl-dC (Figure 4B) and oxidative stress-derived

8-oxo-dG (Figure 4C). The radar plots suggest that, based on characteristic disturbances of the ONT signal, these common DNA modifications can be easily detected and distinguished from the other exposure-derived alkyl-dG adducts within a particular sequence context as discussed above.

Lastly, we performed analyses to assess the discrimination and detection of dG adducts by ONT based on the pTGFp-Hha10 plasmid. Following the DIS plot, the signals of the 11 nucleotides at and near the DNA adduct site (i.e., the adduct and the five flanking nucleotides each on the 5' and 3' sides) were used. The ROC analysis of discrimination between individual dG adduct and control (Figure 5A) showed that we can discriminate most dG adducts very well with AUC > 0.96, except that *O*⁶-*n*Pr-dG and *O*⁶-*s*Bu-dG had a lower AUC < 0.91. We then extended our analysis to different in silico mixtures of *O*⁶-alkyl-dG and *N*²-alkyl-dG adducts by ONT and found clear discrimination of both pairs of regioisomeric *N*²-/*O*⁶-Et-dG and *n*Bu-dG adducts (Figure 5B). The limit of detection was assessed using ELIGOS to calculate the *p* value on the different in silico mixtures by comparing the individual in silico mixture of reads of an individual DNA adduct-containing plasmid in the presence of reads of the control plasmid as reference (Figure 5C). We can detect an adduct as low as 5% at the *p* value cutoff 0.05 for most adducts. We observed a poorer limit of detection of *O*⁶-Me-dG, possibly due to the smallest alkyl group that confers less impact on signal alteration in ONT sequencing.

DISCUSSION

The open source ELIGOS tool^{35,42} was developed to identify modifications to RNA throughout the transcriptome by translating sequencing errors during ONT sequencing into the corresponding RNA modifications.^{35,42} Here, we report the first step toward extending the ELIGOS tool to decode alkylated DNA adducts in native DNA sequences. Double-stranded DNA plasmids containing a variety of alkylated guanine adducts at known positions were used to provide proof of principle for detection and characterization. Sequencing and base-calling results were compared to the expected sequence and between the adduct-containing plasmids and control (i.e., parent plasmid). The ELIGOS tool specifically uncovered each known adduct site as a result of increased (relative to control) base-calling errors at the modified nucleotide and its surrounding nucleotides. The call error rate was computed for each nucleotide position and was used to localize the DNA adduct and to distinguish it from noise. Pairwise comparisons for the site-specific call error rate identified sites in the adduct-containing plasmids that differed from corresponding sites in the unmodified plasmid. After these analyses, the DNA adduct positions were correctly identified in all plasmids (Supporting Information, Figures S2 and S3). These results demonstrate that nanopore sequencing has the capability in identifying unknown biomolecules.⁴³

Furthermore, we characterized the signal disturbance relative to the standard canonical signal of individual adducts. A DNA modification or DNA adduct also may alter the signal of its neighboring 3' and 5' nucleotides because the ionic current signal results from all nucleotides in the pore during the measurement (Scheme 1); a pore can accommodate approximately seven nucleotides. The complex alterations to ion signals that are elicited by

DNA adducts and the corresponding unmodified nucleotides are visualized in radar plots of DIS, with the adduct position at the top vertical position. The plots clearly illustrate that the DNA adducts significantly altered ion signals for up to three nucleotides on each side of the adduct (Figures 1–4). Unfortunately, at this time our library of plasmids containing DNA adducts is limited and insufficient to build models to predict structures of other DNA adducts within the model sequence or within other sequences.

Both the ESB profile and the DIS pattern can be used to discriminate the adduct-containing plasmids from each other and from control plasmids (Supporting Information, Figure S4). The DIS pattern seems to explain the behavior of ionic current derived from an adducted nucleotide within a certain sequence transiting through the nanopore; however, this requires processing of resquiggle signals from the raw signals, which is computationally intensive. In contrast, calculation of ESB profiles across all nucleotides is less intensive and more feasible. Therefore, a genome-wide analysis is expected to be more efficiently achieved by first using ESB profiles to flag, with high confidence, the locations of possibly modified bases throughout the genome and then using DIS patterns to refine analysis and identification of epigenetic DNA modification or exposure-induced DNA adduct at these positions.

The sequence context surrounding the DNA adduct site had a strong impact on the patterns of ESB profiles in both the synthetic adduct plasmid and control, indicating the sequence dependence of the nanopore signal (Figure 1D and E). The radar plots are simplified representations of the complex structure and fluid dynamics measured as ion current while the DNA transits through the nanopore (Scheme 1). Effects on the ion current may occur when the modified base enters (Scheme 1A), passes through (Scheme 1B), or exits (Scheme 1C) the pore. Notably, we observed the characteristic differences in ESB profiles and DIS plots among the different DNA adduct-containing plasmids studied. One challenging aspect of the ONT sequencing method is that the sequence context of the adduct significantly affected the readout and subsequent identification of the adduct. Development of a predictive model based on a statistical model or machine learning will require a comprehensive training set that covers all possible sequence compositions. Theoretically, each DNA adduct of interest requires a set of DNA standards comprised of all possible sequences of the three adjacent 3' and 5' nucleotides ($4^6 = 4096$ standards per adduct of interest). This seems to be an astronomical task unless a prediction algorithm can be established based on structure-fluid dynamic models. An alternative would be to focus on DNA sequences that are important in disease development. For example, preparing standards that contain the DNA adduct of interest within the sequence context of known hotspots of cancer driver mutations would enable us to specifically examine formation and repair of DNA adducts at positions already known to be clinically important. Such an approach, while very important, will, however, limit the analysis to preselected regions and prohibits unbiased genome-wide investigations.

The ONT/ELIGOS platform detects signal disturbances while the DNA transits through the pore. The measurement of the interaction between the nanopore channel/motor protein and the molecule of DNA sequence is in contrast to other detection approaches that measure specific chemical characteristics of the DNA adduct itself. Unlike conventional adduct

detection methods, for example, LC-MS/MS, that measure DNA adduct concentration, expressed as DNA adduct per total nucleosides, ONT employs count of reads as a proxy, representing a population of sequencing reads containing the DNA adduct per total reads sequenced at same base position. Notably, the limit of detection of a particular DNA adduct by ONT sequencing may not be comparable with traditional values due to site specificity of the ONT measurement. Future investigation of the association between site-specific detection by ONT and quantification by LC-MS/MS is warranted. Furthermore, there were some DNA adducts, within a certain sequence context, that showed a strong effect on the ion current, like the N^2 -alkyl-dG studied herein, and others with minor effect on the ion current (e.g., O^6 -Me-dG, O^6 -*n*Pr-dG, and O^6 -*s*Bu-dG). This leads to the different levels of false positive rate between DNA adducts and between the same DNA adduct at different positions (Figure 5). It will be important to systematically examine how sequence context affects ONT sequencing of DNA adducts in the future.

In general, the DNA adduct increased error calling on neighboring 3' and 5' nucleotides. Qualitative comparison of ion signal perturbations of O^6 -Me-dG, O^6 -*n*Pr-dG, and O^6 -*n*Bu-dG within the sequence context of the pTGFH-Hha10 plasmid revealed more pronounced effects on G640, G639, and C638 with increasing lengths of the alkyl chain (Figure 1). In contrast, O^6 -HE-dG has a much less pronounced effect on G640 and C639. As illustrated by the counterclockwise radar plots, the adduct at position G640 affected the signal of the preceding nucleotide read. This suggests that the increase in error calling may be due to hydrophilic interactions between the DNA adduct and the helicase, possibly slowing unwinding of the DNA, thereby perturbing the signal in a way that depends on alkyl chain length. This is supported by the observation of less disturbance by O^6 -HE-dG, which is more polar than, albeit similar in size to, O^6 -*n*Pr-dG. The N^2 -Et-dG and N^2 -*n*Bu-dG showed even more disturbance than the corresponding O^6 -alkyl-dG adducts. In simple terms, alkylation at the O^6 -position elongates the nucleotide, while alkylation at the N^2 position widens the nucleotide and is more effective in blocking the ion current traveling through the nanopore, which leads to stronger signal disturbance.

In summary, we demonstrated that the commercially available ONT, combined with our ELIGOS software, is a suitable platform to qualitatively localize DNA adducts at nucleotide resolution. In addition, the signal disturbances can be characterized and used to distinguish DNA adducts that differ in size, regiochemistry, and functional group. The current limitation is the sequence specificity of the read, preventing genome-wide localization and characterization of DNA modifications. Current efforts by us and others are underway to expand ONT sequencing to include detection of other DNA modifications, and it is expected that the ONT/ELIGOS platform will be suitable for genome-wide analyses of epigenetic modifications and exposure-induced DNA adducts in the near future. This will enable us to better understand the mechanism of mutagenesis and carcinogenesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Dr. Natalia Tretyakova and Dr. Colin Campbell from the University of Minnesota for providing the plasmids containing 5-Fm-dG and 8-oxo-dG, respectively. We are also grateful to Dr. Robert Fuchs from Inserm, France, for providing the plasmid containing O^6 -Me-dG* and to Robert Carlson, Masonic Cancer Center, University of Minnesota, for help with figure preparation. The manuscript was edited by the Science Communication Group at the University of Arkansas for Medical Sciences, Little Rock, AR. Financial support was received from Arkansas Biosciences Institute (to G.B.), National Institutes of Environmental Health Sciences (ES029749 to Y.W.), and National Institute of General Medical Sciences of the National Institutes of Health (award P20GM125503 to I.N.).

REFERENCES

- (1). Balbo S, Turesky RJ, and Villalta PW (2014) DNA adductomics. *Chem. Res. Toxicol* 27, 356–366. [PubMed: 24437709]
- (2). Hwa Yun B, Guo J, Bellamri M, and Turesky RJ (2020) DNA adducts: Formation, biological effects, and new biospecimens for mass spectrometric measurements in humans. *Mass Spectrom. Rev* 39, 55. [PubMed: 29889312]
- (3). Jackson SP, and Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* 461, 1071–1078. [PubMed: 19847258]
- (4). Ji S, Park D, Kropachev K, Kolbanovskiy M, Fu I, Broyde S, Essawy M, Geacintov NE, and Tretyakova NY (2019) 5-Formylcytosine-induced DNA-peptide cross-links reduce transcription efficiency, but do not cause transcription errors in human cells. *J. Biol. Chem* 294, 18387–18397. [PubMed: 31597704]
- (5). Ciccia A, and Elledge SJ (2010) The DNA damage response: making it safe to play with knives. *Mol. Cell* 40, 179–204. [PubMed: 20965415]
- (6). Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, and Linn S (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem* 73, 39–85. [PubMed: 15189136]
- (7). Delaney JC, and Essigmann JM (2008) Biological properties of single chemical-DNA adducts: a twenty year perspective. *Chem. Res. Toxicol* 21, 232–252. [PubMed: 18072751]
- (8). Chatterjee N, and Walker GC (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen* 58, 235–263. [PubMed: 28485537]
- (9). Pottenger LH, Boysen G, Brown K, Cadet J, Fuchs RP, Johnson GE, and Swenberg JA (2019) Understanding the importance of low-molecular weight (ethylene oxide- and propylene oxide-induced) DNA adducts and mutations in risk assessment: Insights from 15 years of research and collaborative discussions. *Environ. Mol. Mutagen* 60, 100–121. [PubMed: 30536466]
- (10). Perera RT, Fleming AM, Johnson RP, Burrows CJ, and White HS (2015) Detection of benzo[a]pyrene-guanine adducts in single-stranded DNA using the alpha-hemolysin nanopore. *Nanotechnology* 26, 074002. [PubMed: 25629967]
- (11). An N, Fleming AM, White HS, and Burrows CJ (2015) Nanopore detection of 8-oxoguanine in the human telomere repeat sequence. *ACS Nano* 9, 4296–4307. [PubMed: 25768204]
- (12). An N, Fleming AM, White HS, and Burrows CJ (2012) Crown ether-electrolyte interactions permit nanopore detection of individual DNA abasic sites in single molecules. *Proc. Natl. Acad. Sci. U. S. A* 109, 11504–11509. [PubMed: 22711805]
- (13). Zeng T, Fleming AM, Ding Y, Ren H, White HS, and Burrows CJ (2018) Nanopore Analysis of the 5-Guanidinohydantoin to Iminoallantoin Isomerization in Duplex DNA. *J. Org. Chem* 83, 3973–3978. [PubMed: 29490132]
- (14). Tan CS, Fleming AM, Ren H, Burrows CJ, and White HS (2018) gamma-Hemolysin Nanopore Is Sensitive to Guanine-to-Inosine Substitutions in Double-Stranded DNA at the Single-Molecule Level. *J. Am. Chem. Soc* 140, 14224–14234. [PubMed: 30269492]
- (15). Johnson RP, Fleming AM, Perera RT, Burrows CJ, and White HS (2017) Dynamics of a DNA Mismatch Site Held in Confinement Discriminate Epigenetic Modifications of Cytosine. *J. Am. Chem. Soc* 139, 2750–2756. [PubMed: 28125225]

- (16). Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM, Bartlett JJ, Pavlenok M, and Gundlach JH (2013) Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl. Acad. Sci. U. S. A* 110, 18904–18909. [PubMed: 24167255]
- (17). Wescoe ZL, Schreiber J, and Akeson M (2014) Nanopores discriminate among five C5-cytosine variants in DNA. *J. Am. Chem. Soc* 136, 16582–16587. [PubMed: 25347819]
- (18). Wang Y, Patil KM, Yan S, Zhang P, Guo W, Wang Y, Chen HY, Gillingham D, and Huang S (2019) Nanopore Sequencing Accurately Identifies the Mutagenic DNA Lesion O(6)-Carboxymethyl Guanine and Reveals Its Behavior in Replication. *Angew. Chem., Int. Ed* 58, 8432–8436.
- (19). Wick RR, Judd LM, and Holt KE (2019) Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20, 129. [PubMed: 31234903]
- (20). Payne A, Holmes N, Rakyan V, and Loose M (2019) BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35, 2193–2198. [PubMed: 30462145]
- (21). Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, and Paten B (2017) Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* 14, 411–413. [PubMed: 28218897]
- (22). Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, and Timp W (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* 14, 407–410. [PubMed: 28218898]
- (23). Liu Q, Fang L, Yu G, Wang D, Xiao CL, and Wang K (2019) Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat. Commun* 10, 2449. [PubMed: 31164644]
- (24). Jenjaroenpun P, Wongsurawat T, Pereira R, Patumcharoenpol P, Ussery DW, Nielsen J, and Nookaew I (2018) Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113–7D. *Nucleic Acids Res* 46, No. e38.
- (25). You C, Swanson AL, Dai X, Yuan B, Wang J, and Wang Y (2013) Translesion synthesis of 8,5'-cyclopurine-2'-deoxynucleosides by DNA polymerases eta, iota, and zeta. *J. Biol. Chem* 288, 28548–28556. [PubMed: 23965998]
- (26). Wu J, Li L, Wang P, You C, Williams NL, and Wang Y (2016) Translesion synthesis of O4-alkylthymidine lesions in human cells. *Nucleic Acids Res* 44, 9256–9265. [PubMed: 27466394]
- (27). Price NE, Li L, Gates KS, and Wang Y (2017) Replication and repair of a reduced 2-deoxyguanosine-abasic site interstrand cross-link in human cells. *Nucleic Acids Res* 45, 6486–6493. [PubMed: 28431012]
- (28). Wu J, Wang P, Li L, Williams NL, Ji D, Zahurancik WJ, You C, Wang J, Suo Z, and Wang Y (2017) Replication studies of carboxymethylated DNA lesions in human cells. *Nucleic Acids Res* 45, 7276–7284. [PubMed: 28531304]
- (29). Liang Q, Dexheimer TS, Zhang P, Rosenthal AS, Villamil MA, You C, Zhang Q, Chen J, Ott CA, Sun H, Luci DK, Yuan B, Simeonov A, Jadhav A, Xiao H, Wang Y, Maloney DJ, and Zhuang Z (2014) A selective USP1-UAF1 inhibitor links deubiquitination to DNA damage responses. *Nat. Chem. Biol* 10, 298–304. [PubMed: 24531842]
- (30). Wu J, Du H, Li L, Price NE, Liu X, and Wang Y (2019) The Impact of Minor-Groove N(2)-Alkyl-2'-deoxyguanosine Lesions on DNA Replication in Human Cells. *ACS Chem. Biol* 14, 1708–1716. [PubMed: 31347832]
- (31). Du H, Leng J, Wang P, Li L, and Wang Y (2018) Impact of tobacco-specific nitrosamine-derived DNA adducts on the efficiency and fidelity of DNA replication in human cells. *J. Biol. Chem* 293, 11100–11108. [PubMed: 29789427]
- (32). Du H, Wang P, Li L, and Wang Y (2019) Repair and translesion synthesis of O(6)-alkylguanine DNA lesions in human cells. *J. Biol. Chem* 294, 11144–11153. [PubMed: 31167778]
- (33). Wang P, and Wang Y (2018) Cytotoxic and mutagenic properties of O(6)-alkyl-2'-deoxyguanosine lesions in *Escherichia coli* cells. *J. Biol. Chem* 293, 15033–15042. [PubMed: 30068548]

- (34). Chesner LN, and Campbell C (2018) A quantitative PCR-based assay reveals that nucleotide excision repair plays a predominant role in the removal of DNA-protein crosslinks from plasmids transfected into mammalian cells. *DNA Repair* 62, 18–27. [PubMed: 29413806]
- (35). Jenjaroenpun P, Wongsurawat T, Wadley TD, Wassenaar TM, Liu J, Dai Q, Wanchai V, Akel NS, Jamshidi-Parsian A, Franco AT, Boysen G, Jennings ML, Ussery DW, He C, and Nookaew I (2020) Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res*, gkaa620. [PubMed: 32710622]
- (36). De Coster W, D’Hert S, Schultz DT, Cruts M, and Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. [PubMed: 29547981]
- (37). Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. [PubMed: 29750242]
- (38). Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. [PubMed: 19505943]
- (39). Benjamini Y, and Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 57, 289–300.
- (40). Box GEP, and Cox DR (1964) An Analysis of Transformations. *J. Roy Stat Soc. B* 26, 211–252.
- (41). Xiao Y, Hsiao TH, Suresh U, Chen HI, Wu X, Wolf SE, and Chen Y (2014) A novel significance score for gene selection and ranking. *Bioinformatics* 30, 801–807. [PubMed: 22321699]
- (42). Wongsurawat T, Jenjaroenpun P, Wassenaar TM, Wadley TD, Wanchai V, Akel NN, Franco AT, Jennings ML, Ussery DW, and Nookaew I (2018) Decoding the Epitranscriptional Landscape from Native RNA Sequences. *bioRxiv* 2018
- (43). Ying YL, and Long YT (2019) Nanopore-Based Single-Biomolecule Interfaces: From Information to Knowledge. *J. Am. Chem. Soc* 141, 15720–15729. [PubMed: 31509414]

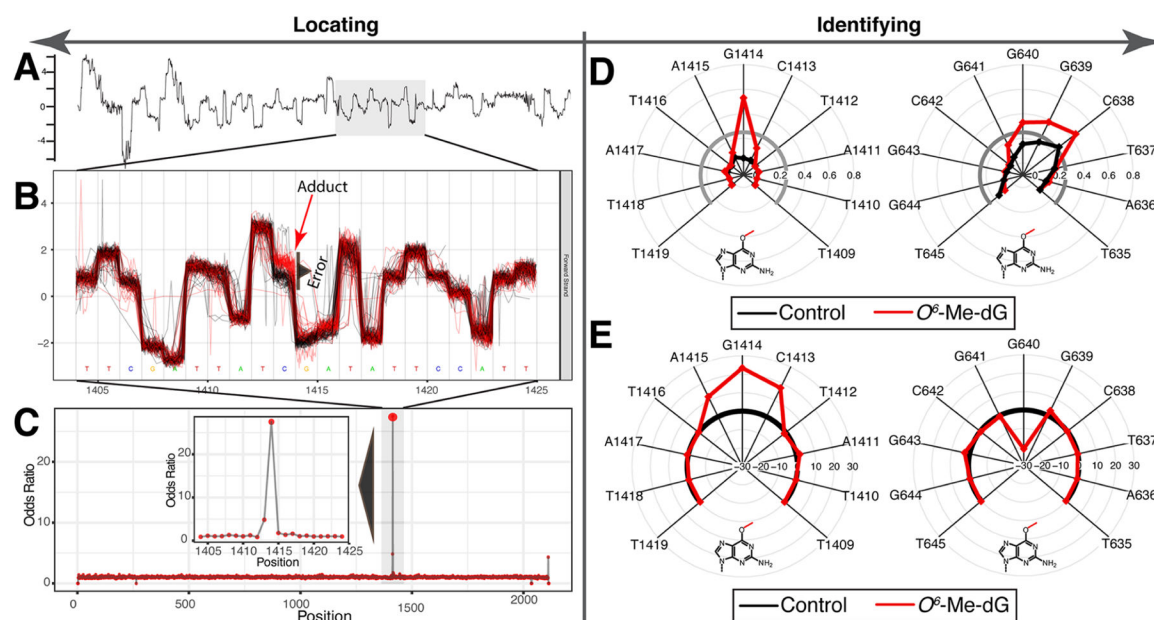


Figure 1.

Overview of stepwise data processing and intermittent data output with ELIGOS software. (A) Representative ONT raw squiggle signal of a read. (B) Comparison of the raw signal and region of a dG adduct-containing plasmid (red) with the control, unmodified dG-containing plasmid (black). Each squiggle line represents an individual read on the (-) strand from right to left. (C) ELIGOS-computed odds ratios at each base position determined the position with the highest probability for harboring the DNA adduct. (D) Radar plots display error at specific base (ESB) profiles, comparing sequences of two *O*⁶-Me-dG-containing plasmids (TTATA^{Me}G*CTATT, left; TGGCG^{Me}GGCTAT, right; unpublished) with the corresponding control plasmid (pTGFH-Hha10). (E) Radar plots display DIS plots for sequences of the two *O*⁶-Me-dG-containing plasmids and control plasmid.

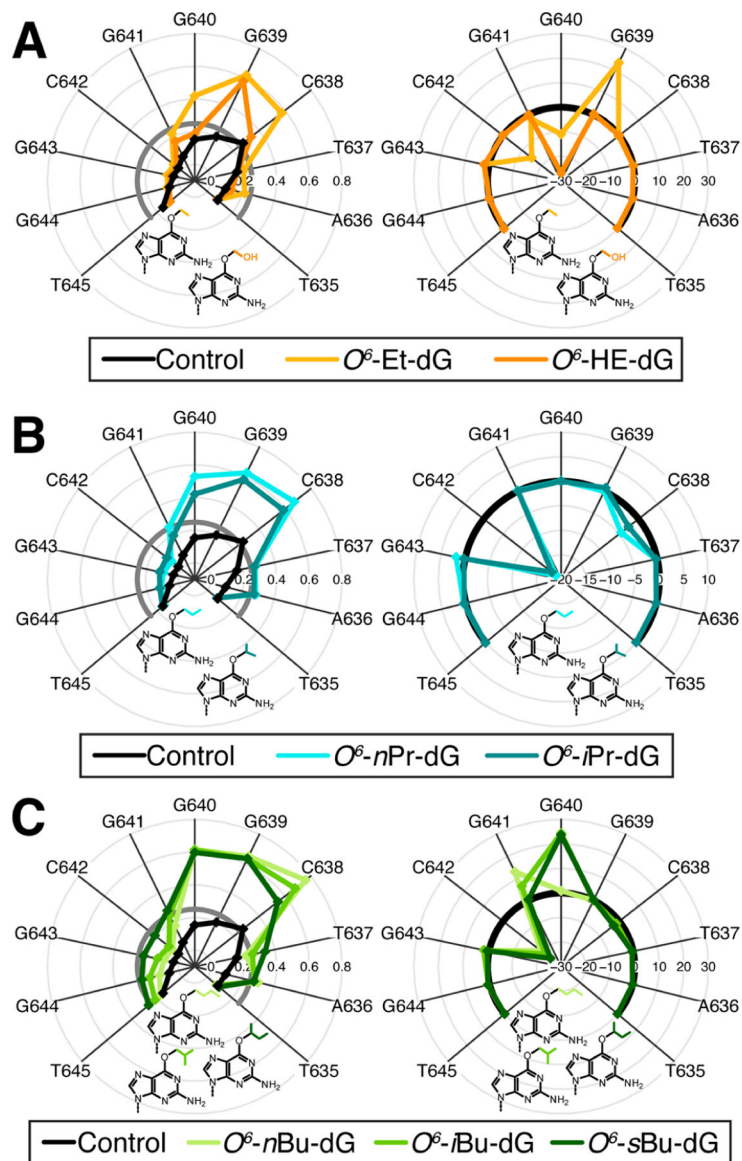


Figure 2. Radar plots display adduct-specific ESB profiles (left) and DIS plots (right) of the adduct-containing plasmids and control plasmid in counterclockwise direction (i.e., (-) strand). Shown are characteristic plots of O^6 -alkyl-dG adducts: (A) O^6 -Et-dG, (B) O^6 -Pr-dG, and (C) O^6 -Bu-dG. All adducts were at the same sequence position (G640) in the same parent plasmid. The plots display signal disturbance for five bases before and after the adduct position. Different scales are used for visualization purposes.

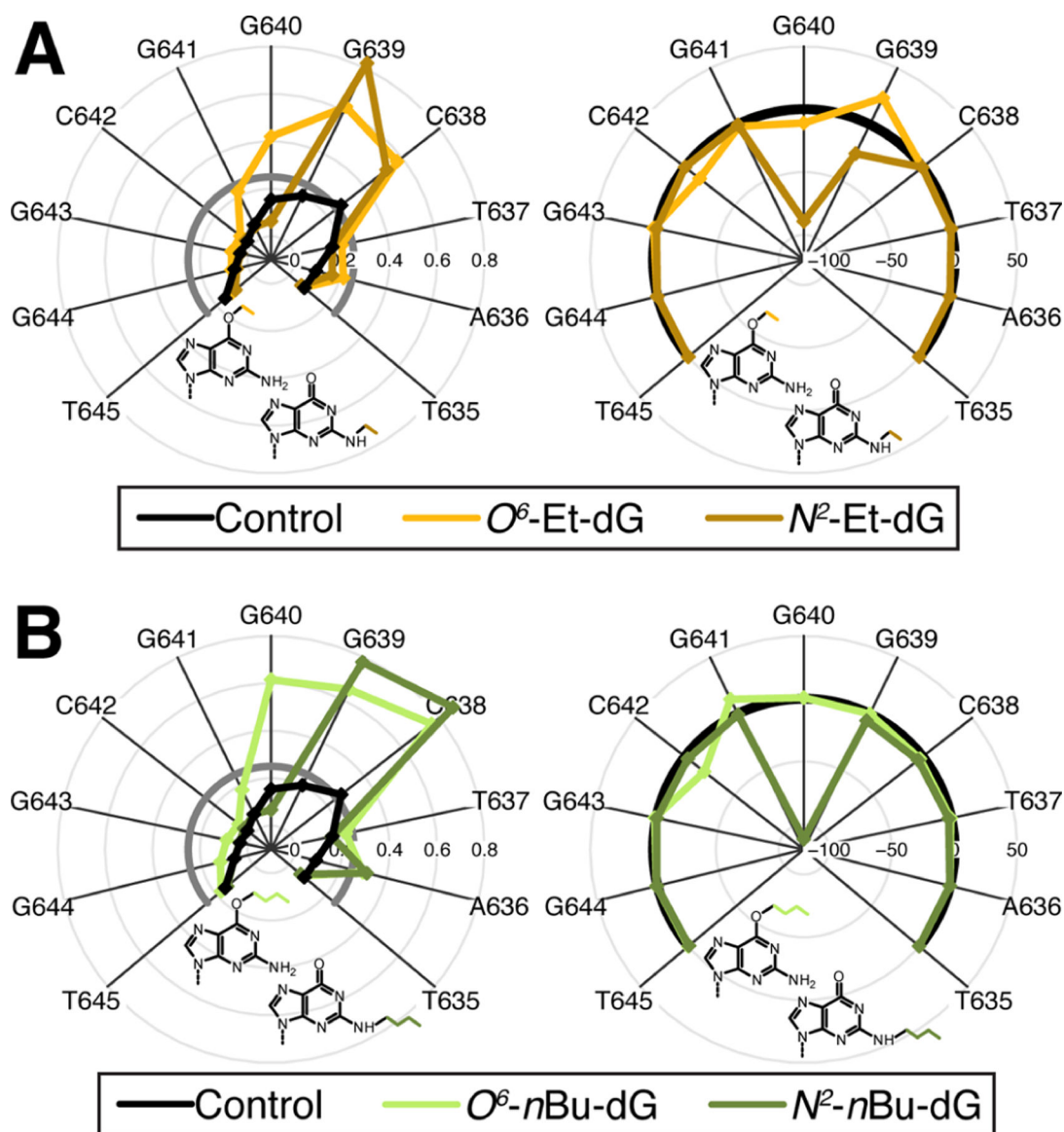


Figure 3. Radar plots display adduct-specific ESB profiles (left) and DIS plots (right) of the adduct-containing plasmids and control plasmid in counterclockwise direction (i.e., (-) strand). Shown are characteristic plots of (A) Et-dG and (B) Bu-dG variants of O^6 -alkyl-dG and N^2 -alkyl-dG adducts. All adducts were at the same sequence position (G640) in the same parent plasmid. The plots show signal disturbance for five bases before and after the adduct position. Different scales are used for visualization purposes.

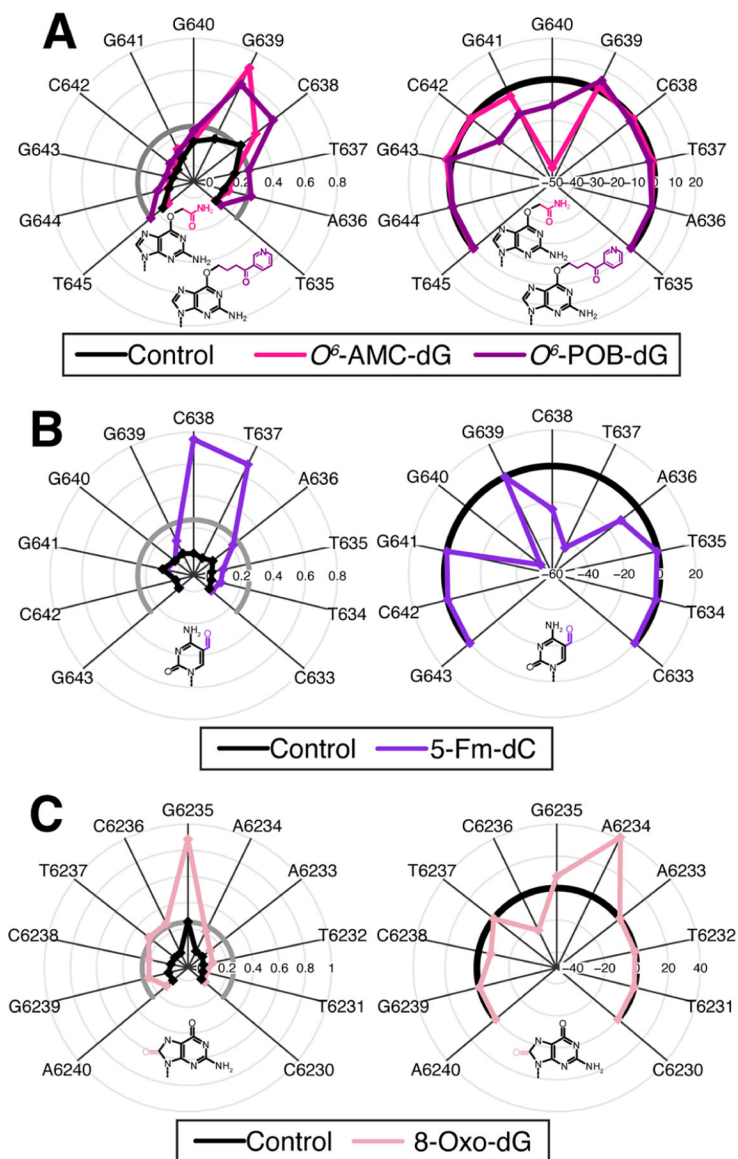
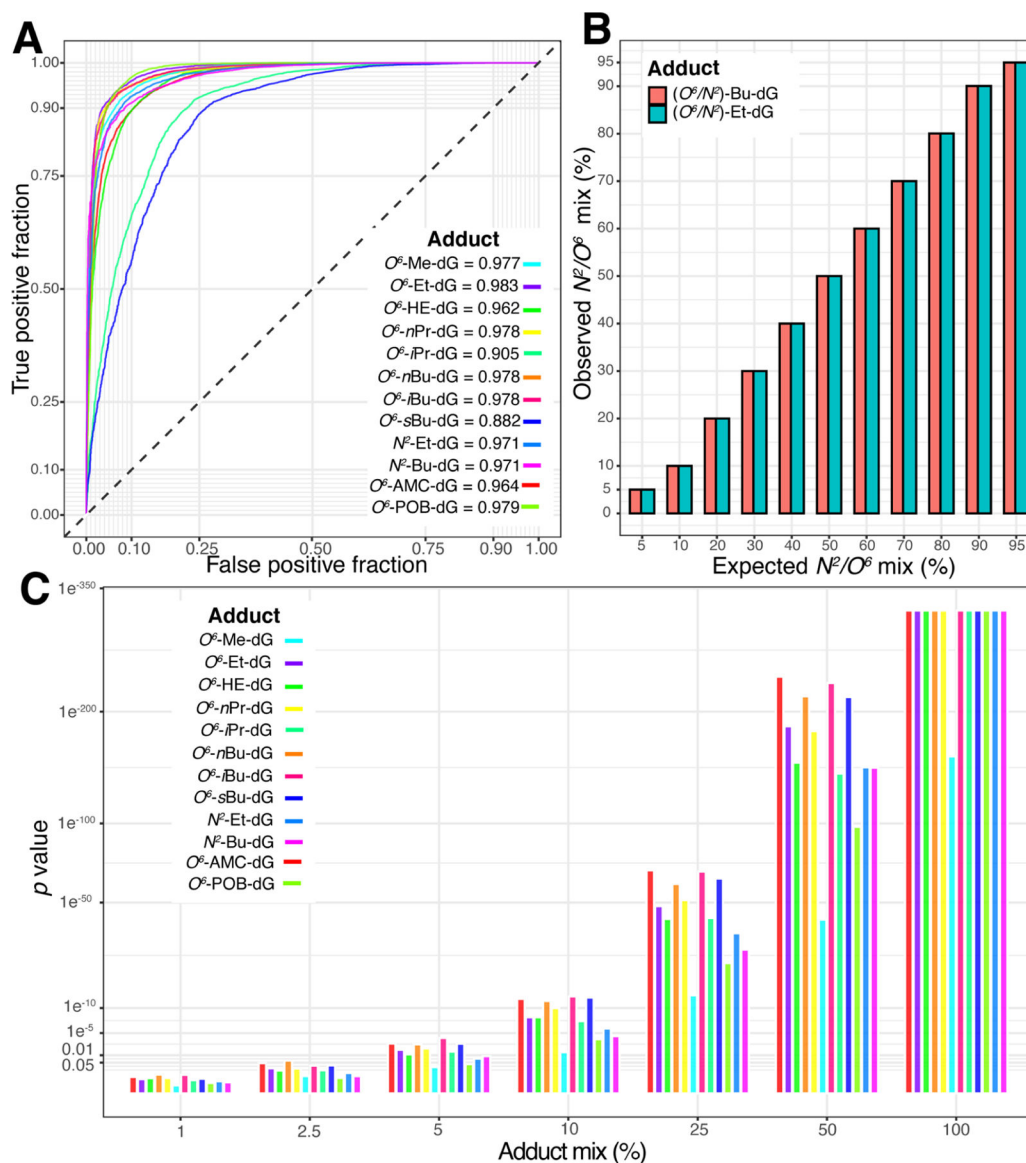
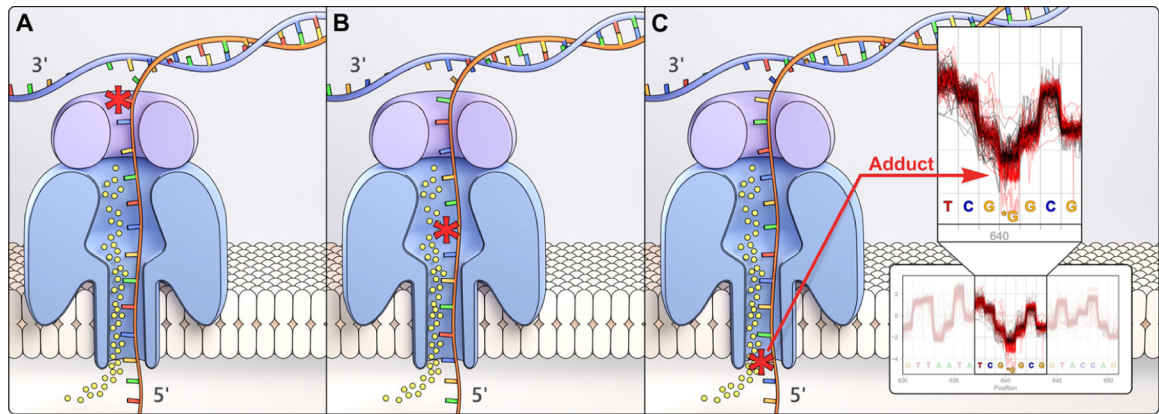


Figure 4. Radar plots display adduct-specific ESB profiles (left) and DIS plots (right) of the adduct-containing plasmids and control plasmid in counterclockwise direction (i.e., (–) strand). Shown are characteristic plots for (A) O⁶-AMC-dG and O⁶-POB-dG, (B) epigenetic mark 5-Fm-dC, and (C) oxidative stress-induced 8-oxo-dG. Adducts shown herein are in different plasmids and sequence context, see Table 1 for details. The plots show signal disturbance for five bases before and after the adduct position. Different scales are used for visualization purposes.

**Figure 5.**

Assessment of adduct detection and discrimination by ONT. Shown are the: (A) ROC curve displaying the ability to discriminate between the adducts and control sequence based on the 11 position as the DIS plot. (B) Bar chart showing the results of analysis of an in silico mixtures of the two regioisomeric N^2 -/ O^6 -ethyl-dG (cyan) and N^2 -/ O^6 -nbutyl-dG (red). The observed values are derived from the mean of 30 samplings (y -axis) of various mixtures plotted against that of the expected ratio (x -axis). The analysis was performed based on the mixture of 2000 reads. (C) The p value derived from ELIGOS for different in silico mixtures of various percentages of reads from adduct-containing plasmids in the presence of 10,000 reads from the control dG-containing plasmid. The p values are derived from the mean of 30 samplings of individual percentage.



Scheme 1.
 DNA Strand, Containing a DNA Adduct (*) Passing through a Nanopore and Potentially
 Blocking or Altering the Ion Current (Yellow Dots) at or near the Adduct Site: (A)
 Preceding, (B) at, or (C) Trailing the DNA Adduct Position

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

DNA Adducts Considered in This Study

Adduct type	Abbreviation	Plasmid
<i>O</i> ⁶ -methyl-dG	<i>O</i> ⁶ -Me-dG*	unpublished ^a
<i>O</i> ⁶ -methyl-dG	<i>O</i> ⁶ -Me-dG	pTGFP-Hha10
<i>O</i> ⁶ -ethyl-dG	<i>O</i> ⁶ -Et-dG	pTGFP-Hha10
<i>O</i> ⁶ - <i>n</i> -propyl-dG	<i>O</i> ⁶ - <i>n</i> Pr-dG	pTGFP-Hha10
<i>O</i> ⁶ - <i>i</i> -propyl-dG	<i>O</i> ⁶ - <i>i</i> Pr-dG	pTGFP-Hha10
<i>O</i> ⁶ - <i>n</i> -butyl-dG	<i>O</i> ⁶ - <i>n</i> Bu-dG	pTGFP-Hha10
<i>O</i> ⁶ - <i>i</i> -butyl-dG	<i>O</i> ⁶ - <i>i</i> Bu-dG	pTGFP-Hha10
<i>O</i> ⁶ - <i>s</i> -butyl-dG	<i>O</i> ⁶ - <i>s</i> Bu-dG	pTGFP-Hha10
<i>O</i> ⁶ -hydroxyethyl-dG	<i>O</i> ⁶ -HE-dG	pTGFP-Hha10
<i>O</i> ⁶ -aminocarbonyl-methyl-2'-dG	<i>O</i> ⁶ -AMC-dG	pTGFP-Hha10
<i>O</i> ⁶ -4-oxo-4-(3-pyridyl)butyl-dG	<i>O</i> ⁶ -POB-dG	pTGFP-Hha10
<i>N</i> ² -ethyl-dG	<i>N</i> ² -Et-dG	pTGFP-Hha10
<i>N</i> ² - <i>n</i> -butyl-dG	<i>N</i> ² - <i>n</i> Bu-dG	pTGFP-Hha10
5-formyl-dC	5-Fm-dC	pTGFP-Hha10 ^b
8-oxo-2'-dG	8-oxo-dG	M13mp18

^aUnpublished plasmid provided by Dr. Robert Fuchs.

^bThe plasmid is based on pTGFP-Hha with minor sequence modifications, as described in ref 4.