

DEVELOPMENT AND APPLICATION OF A MODULAR ENZYMATIC LABELING
PROCESS FOR THE ANALYSIS OF DIVERSE DNA BASE MODIFICATIONS

BY

FANNY WANG

A Dissertation Submitted to the Graduate Faculty of

WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Biomedical Engineering

January 2021

Winston-Salem, North Carolina

Approved By:

Adam R. Hall, Ph.D., Advisor and Chair

Cristina M. Furdui, Ph.D.

Rahul M. Kohli, Ph.D.

Chang Lu, Ph.D.

Fred W. Perrino, Ph.D.

Elaheh Rahbar, Ph.D.

DEDICATION AND ACKNOWLEDGEMENTS

This dissertation is far more than the culmination of six years of work. Without the support of many other people, it would not have been possible.

First and foremost are my family. In search of a better life for himself and his future family in the aftermath of the Chinese Cultural Revolution, my father worked hard to earn the chance to come to America, the land of opportunity. With him, he brought my mother, who left behind a respected job as a teacher to do menial labor at a dry cleaner's to help support them as my father went through graduate school. Together, they were able to achieve the American Dream: not quite rags-to-riches, but making a comfortable life for themselves, my siblings, and me through their hard work. When I was young, they encouraged me to explore the world around me and ask questions, building the foundations of my interest in science. Since then, they have done their best to support me in all of my pursuits. My younger siblings have also supported me with their unwavering belief in me, even when I doubt myself. In particular, my sister has demonstrated an incredible love for me and pride in my achievements, even though I think that she is the more successful of the two of us.

With regard to the work itself, the person who I need to thank the most is Dr. Adam Hall, who provided me not only with this opportunity but also with immeasurable support in carrying it through to completion. From brainstorming ideas with me whenever I hit an obstacle to even helping me set up and run some of my experiments, in addition to his sincere concern about the mental (and physical) health of everyone in the lab, he has always been deeply invested in our successes, not just scientifically and professionally, but also

personally. It has been a great honor to be able to learn from and work under such a motivated, creative, and compassionate PI.

This also could not have been achieved without the assistance of Osama Zahid, who taught me so much when I first came to the lab (and throughout our time together), and Jan Ruzika, who helped me make the first incursions in my experiments. From that point on, many people came together in the spirit of scientific collaboration to make all of this possible, including Dr. Fred Perrino, who provided significant guidance on enzyme activity and behavior in our initial stages, Scott Harvey, who provided additional guidance in addition to teaching me specific laboratory techniques, and the Perrino Lab in general, which generously shared their laboratory space and equipment with us. Dr. Tom Hollis and his lab and Derek Parsonage of the Poole Lab were also instrumental in teaching me how to perform protein expression and lending me their space and materials to advance my research. In addition to having many thoughtful discussions as to the direction of my projects, Drs. Cristina Furdui and Rahul Kohli also provided us with materials that were particularly useful for the later part of my research. Furthermore, along with the rest of my committee, Drs. Chang Lu and Ellie Rahbar offered advice and suggestions on my work, including raising questions on points that required additional experimentation to provide clarity.

I have also received substantial support from many members of the lab, past and present. Besides Osama Zahid, there were also Shiny Rajan, Felipe Rivas Duarte, Sam Bearden, and Komal Sethi, all of whom made work a pleasure with the open environment in which

we could chat and bounce ideas off each other. Brandi Swain, one of our undergraduate researchers, contributed greatly to this work as well, particularly in the early stages.

Outside the lab, a small but irreplaceable support network kept me going even during the most difficult or frustrating times. Other than those already mentioned, these individuals include Berkan Guleyupoglu, another graduate student in the department; Connie Choi and Peter Bozzo, two of my best friends from college; and people who I have had the fortune to meet through the power of the Internet, the most significant being Adam Peterson (MI), Cory Carr (NC), and Joe Hunters (Australia).

Without all of these people, and many others who have helped me in smaller ways, I may have never reached this point. So I would like to take a moment to thank everyone who believed in me, supported me, and helped me to succeed.

Funding for this work was provided by the NIH, Wake Forest Comprehensive Cancer Center, Structural and Computational Biophysics T32 Grant, and Redox Biology and Medicine T32 Grant.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES	ix
LIST OF ABBREVIATIONS.....	xii
ABSTRACT.....	xiv
CHAPTER 1 – INTRODUCTION	1
1.1 Central Dogma of Molecular Biology.....	1
1.2 Sources of DNA Base Modifications.....	2
1.2.1 Replication Errors	2
1.2.2 DNA Oxidation.....	2
1.2.3 DNA Hydrolysis	3
1.2.4 DNA Methylation	3
1.2.5 Base Deamination	4
1.3 Common Methods of Analyzing Base Modifications.....	4
1.3.1 Bisulfite Sequencing.....	4
1.3.2 Liquid Chromatography-Mass Spectrometry	5
1.4 Base Excision Repair	7
CHAPTER 2 – MATERIALS AND METHODS	9
2.1 Labeling Procedures.....	9
2.2 Protein Expression.....	16
2.3 Analytical Techniques.....	19

CHAPTER 3 – DEVELOPMENT OF A MODULAR ENZYMATIC LABELING PROCESS FOR DNA BASE MODIFICATIONS	25
3.1 Introduction	25
3.2 Development of a method of labeling uracil and 8-oxoguanine	26
3.3 Development of a method of labeling T:G mismatches and ethenoadenine with APE1 displacement	30
3.4 Optimization of protocols to improve labeling accuracy	34
3.5 Discussion	36
CHAPTER 4 – LABELING OF PRODUCTS OF THE CYTOSINE DEMETHYLATION PATHWAY	38
4.1 Introduction	38
4.2 Development of protocols to selectively label the substrates of the cytosine demethylation pathway: methylcytosine, hydroxymethylcytosine, carboxylcytosine, and formylcytosine	41
4.3 Discussion	46
CHAPTER 5 - SOLID-STATE NANOPORE ANALYSIS OF DIVERSE DNA BASE MODIFICATIONS	49
5.1 Introduction	49
5.2 Nanopore analysis of uracil, 8-oxoguanine, T:G mismatches, and ethenoadenine with glycosylase specificity testing	51
5.3 Discussion	56

CHAPTER 6 – OPTIMIZATION OF LIGATION PROTOCOLS FOR LABELED	
MATERIALS.....	58
6.1 Introduction	58
6.2 Development of ligation protocols for bifunctional glycosylases (case study: Fpg)	59
6.3 Development of ligation protocols for monofunctional glycosylases (case study: UDG)	60
6.4 Development of ligation protocols for glycosylases requiring enzymatic displacement (case study: TDG)	62
6.5 Discussion	65
CHAPTER 7 – SIMPLE AND EFFICIENT ROOM-TEMPERATURE RELEASE OF	
BIOTINYLATED NUCLEIC ACIDS FROM STREPTAVIDIN.....	
7.1 Introduction	67
7.2 Release of biotinylated nucleic acid bound to free monovalent streptavidin.....	68
7.3 Release of biotinylated nucleic acid bound to streptavidin-coated magnetic beads	74
7.4 Isolation of biotinylated DNA from a mixture with non-biotinylated DNA and verification by solid-state nanopore analysis	75
7.5 Discussion	79
CHAPTER 8 - ANALYSIS OF 8-OXOGUANINE IN RADIATION-SENSITIVE AND	
RADIATION-RESISTANT CELL LINES	
	81

8.1	Introduction	81
8.2	Development of a protocol for labeling and quantifying oxidative damage in DNA extracted from cells with qPCR.....	82
8.3	Radiation dosage and time course study of oxidative damage.....	86
8.4	Discussion	90
CHAPTER 9 – FUTURE DIRECTIONS		93
REFERENCES		96
APPENDIX.....		118
CURRICULUM VITAE.....		121

LIST OF ILLUSTRATIONS AND TABLES

Figure 1. Schematic representation of the general labeling approach	27
Figure 2. Denaturing gel analyses of labeled uracil and oxoG-containing DNA constructs	29
Figure 3. Denaturing gel analyses of labeled T:G-mismatch-containing DNA constructs	30
Figure 4. WT APE1 and D308A APE1 exonuclease activity.....	32
Figure 5. Denaturing gel analyses of labeled 1,N ⁶ -ethenoadenine-containing DNA constructs	33
Figure 6. Limiting off-target T4 polymerase end-labeling with terminal transferase	35
Figure 7. Schematics of the cytosine methylation/demethylation pathway and the generalized labeling scheme.	40
Figure 8. Denaturing gel analyses of 34 nt DNA constructs featuring 5fC and 5caC.....	42
Figure 9. Denaturing gel analyses of 34 nt DNA constructs featuring 5mC and 5hmC...	44
Figure 10. Labeling scheme for differentiating the four bases of the cytosine demethylation pathway with modular glycosylase labeling.	46
Figure 11. Depiction of the selective SS-nanopore assay.....	50
Figure 12. SS-nanopore analyses of DNA oligonucleotides featuring either a single uracil or a single oxoG.....	52
Figure 13. SS-nanopore analyses of DNA oligonucleotides featuring either a single T:G mismatch or a single 1,N ⁶ -ethenoadenine.....	53
Figure 14. Additional SS-nanopore data for labeled 1,N ⁶ -ethenoadenine DNA.	55

Figure 15. Denaturing gel analyses of 40 nt DNA constructs featuring a single 8-oxoguanine at base position 27 following ligation.	59
Figure 16. Denaturing gel analyses of 40 nt DNA constructs featuring a single uracil at base position 33 following ligation.	61
Figure 17. Denaturing gel analyses of 34 nt DNA constructs featuring a single 5caC at base position 22 following ligation.	62
Figure 18. Denaturing gel analyses of 34 nt DNA constructs featuring a single 5caC at base position 22 following ligation.	64
Figure 19. Schematic of proposed biotin-streptavidin dissociation method.	70
Figure 20. Gel analysis of liquid-phase elution	72
Figure 21. Effect of phenol concentration on biotinylated DNA release	73
Figure 22. Gel analysis of 150 bp biotinylated DNA elution from streptavidin-conjugated magnetic beads	75
Figure 23. Isolation and analysis of biotinylated DNA from a mixture with non-biotinylated DNA.	77
Figure 24. DNA hairpin adapter schematic	83
Figure 25. Denaturing gel analysis of adapter ligation	84
Figure 26. Agarose gel of enzymatic fragmentation.	85
Figure 27. Effect of radiation dosage on 8-oxoG in SCC-61 and rSCC-61 cells 30 min after irradiation.	87
Figure 28. Change in net 8-oxoG content as a function of time for SCC-61 and rSCC-61 cells following 2 Gy irradiation	89
Figure 29. Cross-recognition for uracil modified bases.	118

Figure 30. Cross-recognition for oxoG modified bases	118
Figure 31. Cross-recognition for T:G mismatch.....	119
Figure 32. Cross-recognition for 1,N ⁶ -ethenoadenine modified bases	119

LIST OF ABBREVIATIONS

5caC: 5-carboxylcytosine

5fC: 5-formylcytosine

5hmC: 5-hydroxymethylcytosine

5mC: 5-methylcytosine

AP: abasic (apurinic/aprimidinic)

APE1: human AP endonuclease 1

BER: base excision repair

CpG: dinucleotide consisting of a cytosine nucleotide 5 prime to a guanine nucleotide (5'-C-phosphate-G-3')

ddATP: dideoxyadenosine triphosphate

DNA: deoxyribonucleic acid

DNMT: DNA methyltransferase

dNTP: deoxynucleotide triphosphate

dATP: deoxyadenosine triphosphate

dCTP: deoxycytidine triphosphate

dGTP: deoxyguanosine triphosphate

dUTP: deoxyuridine triphosphate

ds: double-stranded

ECD: event charge deficit

ELISA: enzyme-linked immunosorbent assay

EMSA: electromobility shift assay

EndoIV: endonuclease IV

EndoVIII: endonuclease VIII

Fpg: formamidopyrimidine DNA glycosylase

hAAG: human alkyladenine DNA glycosylase

hOGG1: human 8-oxoguanine DNA N-glycosylase 1

ILK: integrin-linked kinase

kbp: kilobase pair

LC-MS: liquid chromatography-mass spectrometry

MS: monovalent streptavidin

oxoG: 8-oxoguanine

PCR: polymerase chain reaction

qPCR: quantitative polymerase chain reaction

RNA: ribonucleic acid

ROS: reactive oxygen species

SBRT: stereotactic body radiation therapy

SS-: solid-state

T4(exo-): T4 polymerase lacking 3'-5' exonuclease activity

TET: ten-eleven translocase

TDG: thymine DNA glycosylase

UDG: uracil DNA glycosylase

ABSTRACT

There exists a large diversity of DNA base modifications, but the ability to study these modifications and their effects on disease development and progression has thus far been limited. Most techniques are only able to examine a few specific modifications, and some methods are destructive to the sample, precluding the possibility of additional analysis.

However, cells have evolved enzymes called glycosylases that are used during base excision repair in order to remove these modifications and replace them with appropriate unmodified bases. Thus, we sought to develop a modular enzymatic labeling process that takes advantage of these naturally occurring enzymes and pathways to expand the range of modifications that can be studied.

In this work, we have accomplished several aims. First, we have developed a methodology to successfully label a variety of modifications, including uracil, 8-oxoguanine, T:G mismatches, 1,N⁶-ethenoadenine, and the products of the cytosine demethylation pathway, with several glycosylases that display different types of activity (Chapters 3 and 4).

We then validated the success of this labeling method and applied it in two analytical techniques: solid-state nanopores and qPCR. With solid-state nanopores, we were able to measure DNA containing different base modifications by incorporating a biotinylated nucleotide during the labeling process, which produced positive results in our system with concentrations as low as 250 nM (Chapter 5).

Following additional optimizations in our protocols (Chapters 6 and 7), we then applied our labeling process to qPCR and examined a more complex system: oxidative damage generated following irradiation in head and neck cancer cell line models. We were able to determine differences in the trends in the accumulation and repair of oxidative damage in these cell lines (Chapter 8).

Through this work, we have developed a modular labeling method that can facilitate the study of a wide range of DNA base modifications without the limitations present in traditional methods. We have also demonstrated the applicability of this method to different analytical techniques. This may be of great use in determining the roles of less-studied modifications in organism growth and development, the development and progression of disease, and other genetic studies.

CHAPTER 1 – INTRODUCTION

1.1 Central Dogma of Molecular Biology

One of the pivotal discoveries of molecular biology was that of the structure of deoxyribonucleic acid (DNA) by James Watson and Francis Crick in 1953¹, which built upon the X-ray crystallography work of Rosalind Franklin and Maurice Wilkins². The key features of this proposed structure were the now widely recognized double-helix shape of DNA with its sugar-phosphate backbone and the complementarity of the strands that arose from the nitrogenous base pairing of adenine to thymine and cytosine to guanine. In this, a method by which genetic information could be reproduced with high fidelity was established, providing an explanation of how traits could be inherited between generations. This discovery subsequently led to the development of the central dogma of molecular biology, which states that, in general, the information encoded in DNA can be transcribed into ribonucleic acid (RNA) and subsequently translated into proteins³, and it is the expression of these proteins that contribute to the development of these inherited traits.

However, as with any process, particularly one so intrinsically linked with the evolutionary process that requires the presence of hereditary variation between individuals⁴, the possibility for error in each of these steps is present. DNA is particularly susceptible to changes, which may occur endogenously through oxidative damage, methylation, deamination, or various other methods^{5,6}; exogenously through environmental agents such as ultraviolet and ionizing radiation and genotoxic chemicals^{6,7}; or through other methods including psychological stressors⁸ or nutritional deficiencies and excesses⁹. These changes may occur in the form of single-nucleotide changes as a result of incorrect incorporation

of nucleotides, or even larger-scale insertions or deletions as a result of strand slippage in repetitive sequences⁶. Subsequently, as the DNA is transcribed into RNA, which is translated into protein, these modifications will result in the production of proteins that may display different functionalities than the original protein. In some cases, this new protein may provide the organism with some sort of evolutionary advantage, but in other cases, it may facilitate the development of disorders or diseases.

1.2 Sources of DNA Base Modifications

1.2.1 Replication Errors

During DNA replication or repair, it is possible for an incorrect base to be inserted, which, if not corrected, will often lead to a change in the DNA sequence with the error being carried through during subsequent replication events. This type of modification can occur as a result of the activity of DNA polymerases, enzymes that replicate or repair DNA, with low fidelity and are more subject to making these types of mistakes, or due to changes in the concentrations of deoxynucleotide triphosphates (dNTPs) in the cell environment⁶.

1.2.2 DNA Oxidation

Reactive oxygen species (ROS) are highly reactive molecules that contain oxygen and are naturally present in cells as a result of metabolic and other reactions^{5,6}. ROS can be found in many forms, including peroxides, superoxides, and hydroxyl radicals. ROS are naturally formed as a byproduct of oxygen metabolism and have important roles in cell signaling. They can also be generated by external sources such as ionizing radiation, which frequently produce hydroxyl radicals, in addition to the direct damage to DNA that can be induced by

the radiation⁶. A dramatic change in ROS levels can cause significant cell and tissue damage; this is often referred to as oxidative stress.

These ROS are capable of oxidizing DNA bases directly, as well as inducing other types of damage such as single- and double-stranded breaks. The most common type of lesion that occurs as a result of DNA oxidation is 8-oxoguanine (8-oxoG), which often incorrectly pairs with adenine instead of cytosine, which subsequently results in a change of sequence from G:C to T:A^{5,6}.

1.2.3 DNA Hydrolysis

Apurinic and apyrimidinic (AP, or abasic) sites can occur in DNA when the bond between the base and the DNA backbone is hydrolyzed or broken by a DNA glycosylase⁶. They can also be produced by reaction with reactive oxygen species (ROS)⁵. While many of these are repaired, those that are not can produce a change in the sequence, as adenine is preferentially incorporated opposite AP sites during DNA replication⁵.

1.2.4 DNA Methylation

DNA methylation, the process by which a methyl group is added to the DNA molecule, occurs most commonly via reaction with *S*-adenosylmethionine (SAM), another molecule with high reactivity^{5,6}. DNA methylation is an important method by which genes can be silenced, altering the properties of cells without actually changing the DNA sequence¹⁰, which most frequently occurs with 5-methylcytosine (5mC)¹¹. However, methylation of the other bases can also result in changes in the sequence, as with O⁶-methylguanine that

can alter a G:C base pair to an A:T or O⁴-methylthymine changing a T:A base pair to a C:G^{5,6}. Furthermore, the methylation of adenine can produce a cytotoxic modification due to the ability of methyladenine to inhibit DNA synthesis^{5,6}, which affects cell replication.

1.2.5 Base Deamination

Base deamination most commonly occurs among cytosine (C) and 5-methylcytosine (5mC) bases, which are changed into uracil (U) and thymine (T) respectively, though deamination can also occur on adenine (A) or guanine (G) bases. For cytosine, this results in a change from a C:G base pair to a U:A pairing, and ultimately T:A. For 5-methylcytosine, this results in a change from a C:G base pair to a mismatched T:G base pair, which may eventually be converted to T:A^{5,6}.

1.3 Common Methods of Analyzing Base Modifications

1.3.1 Bisulfite Sequencing

The most common method of analyzing the products of the cytosine demethylation pathway (5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine) is through bisulfite sequencing. In this method, DNA is treated with bisulfite to convert cytosine to uracil, which allows for 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) to be distinguished from cytosine, 5-formylcytosine (5-fC), or 5-carboxylcytosine (5-caC)¹². However, as is apparent, this method presents difficulties in distinguishing between 5-mC and 5-hmC, both of which remain unconverted following bisulfite treatment, and cytosine, 5-fC, and 5-caC, which are all converted to uracil following treatment.

Methods to address these issues have been developed, such as oxidative bisulfite sequencing, which oxidizes 5-hmC to 5-fC, and TET-assisted bisulfite sequencing, in which 5-hmC is protected from oxidation by TET, which converts 5-mC to 5-caC. These allow 5-mC and 5-hmC to be distinguished in a sample, as following treatment, either 5-mC or 5-hmC will not be converted to uracil so the differences between a sample in which both modifications remain and a sample in which one modification has been converted can provide information on each modification individually. Similarly, reductive bisulfite sequencing can be used to reduce 5-fC to 5-hmC, protecting it from conversion to uracil and allowing for differentiation between 5-fC and 5-caC.

However, a readily apparent limitation of bisulfite sequencing is its applicability to primarily these modifications. Furthermore, it results in the loss of pyrimidine bases in the DNA, which results in strand cleavage and loss of material, requiring several rounds of PCR prior to sequencing to produce enough material for sequencing¹². This need for amplification introduces the possibility of PCR biases that may produce less than accurate results.

1.3.2 Liquid Chromatography-Mass Spectrometry

Liquid chromatography-mass spectrometry is an analytical technique that incorporates physical separation and mass separation via liquid chromatography and electrospray ionization to identify the components of a mixture. In the liquid chromatography part of the process, the liquid mixture is first injected through a column under high pressure. Due

to differences in the weight of the molecules within the mixture, the components will separate, with smaller molecules traveling further than larger molecules¹³. When this liquid stream is exposed to a high voltage, it will form liquid droplets that contain a positive charge. As the liquid in the droplets gradually evaporates, the solution will aerosolize into small, ionized molecules^{14,15}. These ionized molecules can then be captured by an electric field and analyzed with mass spectrometry. Due to the charge contained on these ionized molecules, they can then be separated due to the ratio of the charge to the mass of the molecule, as lighter, more charged molecules will be more greatly influenced by electric fields and deflect further than heavier, less charged molecules¹³. By analyzing the mass to charge ratios collected from a sample, it is then possible to compare them to samples of known identity in order to determine what the sample is.

However, for LC-MS to be used for analyzing DNA bases, it is usually necessary for the DNA to be digested into its individual nucleotides¹⁶. As such, the original DNA cannot be recovered following LC-MS, and the sequence data will be lost. Furthermore, the necessity of ionizing the nucleotides presents the risk of inducing lesions such as oxidative damage where it was not present in the original DNA, artificially increasing the measurements of these base modifications¹⁷. The use of LC-MS is also limited by sample size requirements, due to the low prevalence of some DNA base modifications, particularly if the methodology is to be applied to human patients¹⁸.

1.4 Base Excision Repair

Base excision repair (BER) is a method by which base modifications that do not significantly distort the structure of the DNA are repaired⁶. In BER, the modification is recognized by a DNA glycosylase enzyme, which subsequently hydrolyzes the nitrogenous base to create an abasic (AP) site. If the DNA glycosylase has bifunctional activity, the glycosylase will then also cleave the DNA backbone after the abasic site. Monofunctional glycosylases do not display this additional behavior.

As such, following base excision, the natural repair process for monofunctional and bifunctional glycosylases differ. In monofunctional glycosylases, an additional enzyme, AP endonuclease, cleaves the DNA backbone at the front of the AP site. This gap is then primed for repair by the POL β enzyme, which acts to ensure that the ends of the gap have the proper chemical structures to allow for repair before inserting the correct nucleotide at the gap. Finally, a DNA ligase repairs the DNA backbone between the new nucleotide and the rest of the DNA strand, reconsolidating the structure of the DNA.

For bifunctional glycosylases, an AP endonuclease again cleaves the DNA backbone at the start of the AP site. Following this, a DNA polymerase (POL β or POL δ/ϵ) inserts the correct nucleotide at the gap before starting to synthesize the rest of the strand, partially replacing the original strand. The excess from the original strand is removed by a flap endonuclease, and then the partially synthesized strand is reconnected to the remaining segment of the original strand to complete the repair.

Other mechanisms of DNA repair for more severe damage are also present in the cell, such as nucleotide excision repair, crosslink repair, and strand break repair, but in our experiments, we focus exclusively on adapting the steps of the base excision repair pathway to develop a methodology for the labeling of DNA base modifications.

In order to achieve this, we developed a series of discrete steps to emulate each process that occurs during BER. First, we introduced a glycosylase specific to the modification we were attempting to label to cleave the modification and leave an AP site. Glycosylases that demonstrated bifunctional ability would also cleave the DNA backbone 3' to the new AP site. Then, we added an endonuclease to cleave 5' to the AP site and prime that end for nucleotide addition. Nucleotide addition was achieved by incubation of the DNA with a DNA polymerase and appropriate nucleotides. For analytical methods that required an unbroken DNA backbone, ligation was performed by an additional incubation step with a DNA ligase. In this way, we were able to develop a modular labeling protocol for a wide range of DNA modifications: by switching out the glycosylase being used, many different DNA modifications could be targeted with minimal changes to the procedure while preserving the DNA sequence.

CHAPTER 2 – MATERIALS AND METHODS

2.1 Labeling Procedures

Uracil labeling A custom 40 nt oligonucleotide with a 5' FAM (sequence: TCA CGA CTA GTG TTA ACA TGT GCA CCT GCA GAA UGA GAA T) was annealed to a complementary sequence by mixing both at an equimolar ratio, incubating in deionized water at 95°C for 10 minutes, and cooling to room temperature over 1 hour. To cap the 3' ends of the DNA, a 100 µL aliquot was prepared containing 385 pmol duplex DNA, 30 nmol 2',3'-dideoxyadenosine 5'-triphosphate (ddATP) (GE Healthcare), 500 U Terminal Transferase (New England Biolabs, Ipswich, MA), and 25 mmol CoCl₂ (New England Biolabs) and incubated in 1X Terminal Transferase Reaction Buffer (New England Biolabs) at 37°C for 1.5 hrs. The resulting material was purified with the QIAquick PCR purification kit (Qiagen, Valencia, CA) to allow for buffer exchange. To excise uracil, a 30 µL aliquot was prepared containing 100 pmol of capped duplex DNA, 20 U *E. coli* UDG (New England Biolabs), 40 U EndoIV (New England Biolabs), 3 µg bovine serum albumin (BSA, New England Biolabs), and incubated in 1X NEB2 buffer (New England Biolabs) at 37°C for 1 hr. Next, 1.5 nmol of biotinylated dUTP (Perkin Elmer, Waltham, MA) and 0.12 U T4(exo-) (Lucigen, Middleton, WI) were added to a final volume of 40 µL in 1X NEB2 buffer and the mixture was incubated at 37°C for 30 minutes. Finally, the mixture was subjected to purification by QIAquick PCR purification kit to remove proteins and excess nucleotides.

UDG Ligation 40 pmol of DNA was co-incubated with 16 U of Fpg (New England Biolabs) during the UDG incubation step, with the rest of the labeling and purification proceeding

as previously described. 400 U of T4 DNA Ligase (New England Biolabs) in T4 DNA Ligase buffer (50 mM tris-HCl 10 mM, MgCl₂, 1 mM ATP, 10 mM DTT, pH 7.5) overnight at room temperature. The DNA was then purified with the Nucleotide Removal Kit and eluted in deionized water.

OxoG labeling with hOGG1 A custom 40 nt oligonucleotide with a 5' FAM (sequence: TCA CGA CTA GTG TTA ACA TGT GCA CCT G^oCA GAA TGA GAA T, where G^o is oxoG) was annealed to a complementary sequence by mixing both at an equimolar ratio, incubating in deionized water at 95°C for 10 minutes, and cooling to room temperature over 1 hour. To excise oxoG, a 30 µL aliquot was prepared containing 100 pmol of duplex, 6.5 U hOGG1 (New England Biolabs), 40 U EndoIV, 3 µg BSA, and incubated in 1X NEB2 buffer at 37°C for 1 hr. Next, 1.5 nmol of biotinylated dGTP (Perkin Elmer) and 0.12 U T4(exo-) were added to a final volume of 40 µL in 1X NEB2 buffer and the mixture was incubated at 37°C for 30 minutes. Finally, the mixture was subjected to QIAquick PCR purification kit purification to remove proteins and excess nucleotides.

OxoG labeling with Fpg A custom 40 nt oligonucleotide with a 5' FAM (sequence: TCA CGA CTA GTG TTA ACA TGT GCA CCT G^oCA GAA TGA GAA T, where G^o is oxoG) was annealed to a complementary sequence by mixing both at an equimolar ratio, incubating in deionized water at 95°C for 10 minutes, and cooling to room temperature over 1 hour. To excise oxoG, a 30 µL aliquot was prepared containing 50 pmol of duplex, 16 U Fpg (New England Biolabs), 20 U EndoIV, 3 µg BSA, and incubated in 1X NEB2 buffer at 37°C for 1 hr. Next, 0.75 nmol of biotinylated dGTP (Perkin Elmer) and 0.06 U

T4(exo-) were added to a final volume of 40 μ L in 1X NEB2 buffer and the mixture was incubated at 37°C for 30 minutes. Finally, the mixture was subjected to QIAquick PCR purification kit purification to remove proteins and excess nucleotides.

Fpg Ligation 40 pmol of labeled DNA was incubated with 400 U of T4 DNA Ligase (New England Biolabs) in T4 DNA Ligase buffer (50 mM tris-HCl 10 mM, MgCl₂, 1 mM ATP, 10 mM DTT, pH 7.5) overnight at room temperature. The DNA was then purified with the Nucleotide Removal Kit and eluted in deionized water.

T:G mismatch labeling A custom 40 nt oligonucleotide with a 5' FAM (sequence: TCA CGA CTA GTG TTA ACA TGT CGA CCT TG A GAA TGA GAA T) was annealed to a complementary sequence (except with a guanine opposite the indicated thymine) by mixing both at an equimolar ratio, incubating in deionized water at 95°C for 10 minutes, and cooling to room temperature over 1 hour. To excise target thymine, a 30 μ L aliquot was prepared containing 100 pmol of duplex, 7.5 mg human TDG¹⁹, 40 fg APE1 (D308A mutant²⁰), 3 μ g BSA, and incubated in 1X HEMN.1 buffer (20 mM HEPES (pH 7.3), 100 mM NaCl, 2.5 mM MgCl₂, 0.2 mM EDTA) at 37°C for 1 hr. The mixture was purified with a QIAquick PCR purification kit. Then, 40 U EndoIV and 3 μ g BSA were added to a total volume of 30 μ L in 1X NEB2 buffer and incubated at 37°C for 30 min. 1.5 nmol biotinylated dCTP (Perkin Elmer, Waltham, MA) and 0.12 U T4(exo-) were added to a final volume of 40 μ L in 1X NEB2 buffer and the mixture was further incubated at 37°C for 30 minutes. Finally, the mixture was subjected to a second purification to remove proteins and excess nucleotides.

TDG Ligation 40 pmol of labeled DNA with the phosphate flap removed (i.e. treated with Endonuclease VIII) was incubated with 400 U of T4 DNA Ligase (New England Biolabs) in T4 DNA Ligase buffer (50 mM tris-HCl 10 mM, MgCl₂, 1 mM ATP, 10 mM DTT, pH 7.5) overnight at room temperature. The DNA was then purified with the Nucleotide Removal Kit and eluted in deionized water.

1,N⁶-ethenoadenine labeling A custom 34 nt oligonucleotide with a 5' FAM (sequence: CAG TTG AGG ATC CCC ATA A^e TG CGG CTG TTT TCT G, where A^e is 1,N⁶-ethenoadenine) was annealed to a complementary sequence by mixing both at an equimolar ratio, incubating in deionized water at 95°C for 10 minutes, and cooling to room temperature over 1 hour. To cap the 3' ends of the DNA, a 100 µL aliquot containing 385 pmol duplex DNA, 30 nmol ddATP, 500 U of Terminal Transferase, and 25 mmol CoCl₂ was incubated in 1X Terminal Transferase reaction buffer at 37°C for 1.5 hrs. The resulting material was purified with a QIAquick PCR purification kit to allow for buffer exchange. To excise target 1,N⁶-ethenoadenine, a 80 µL aliquot was prepared containing 100 pmol of duplex, 425 U hAAG (New England Biolabs), 200 fg APE1 D308A mutant, 8 µg BSA, and 1X Thermopol buffer (New England Biolabs) and incubated at 37°C for 1 hr. The mixture was purified with a QIAquick PCR purification kit. Then, 40 U EndoIV and 3 µg BSA were added to a total volume of 30 µL in 1X NEB2 buffer and incubated at 37°C for 30 min. 1.5 nmol biotinylated dATP (Perkin Elmer, Waltham, MA) and 0.12 U T4(exo-) were added to a final volume of 40 µL in 1X NEB2 buffer and the mixture was further

incubated at 37°C for 30 minutes. Finally, the mixture was subjected to a second QIAquick purification to remove proteins and excess nucleotides.

DNA constructs for cytosine demethylation labeling Four sets of 34 nt-long DNA oligonucleotides featuring a fluorescent 5' FAM label were purchased commercially (Integrated DNA Technologies, Coralville, IA) with the sequence 5'-CAG TTG AGG ATC CCC ATA ATG CGG CTG TTT TCT G-3', in which the highlighted nucleotide (C) was replaced with 5mC, 5hmC, 5fC, or 5caC, respectively. Duplex constructs were formed by mixing 10 µM of each with its unmodified complementary sequence at a ratio of 1:1.2 in deionized water, incubating at 95°C for 10 minutes, and gradually cooling to room temperature over two hours. Products were confirmed by gel electrophoresis.

Dual labeling 5fC and 5caC 40 pmol DNA was incubated with 3 µg wild-type TDG, 13.3 fg APE1 D308A, and 4 µg bovine serum albumin (BSA, New England Biolabs, Ipswich, MA) in 20 µL HEMN.1 Buffer (200 mM HEPES, 1M NaCl, 2 mM EDTA, 25 mM MgCl₂) at 37°C for 1 hour to excise target bases and detach the TDG from the resulting AP site. After purifying the DNA with a Nucleotide Removal Kit (Qiagen), it was incubated with 20 U of Endonuclease IV (New England Biolabs), 100 U of Endonuclease VIII (New England Biolabs), and 4 µg BSA in 20 µL NEB2 buffer (50 mM NaCl, 10 mM tris-HCl, 10 mM MgCl₂, 1 mM DTT, New England Biolabs) at 37°C for 30 minutes to prime the gap for base incorporation. Then, 1.5 nmol of biotinylated dCTP (Perkin Elmer, Waltham, MA) and 0.12 U of T4 polymerase having no exonuclease activity (Lucigen, Middleton,

WI) were added and the mixture and incubated at 37°C for an additional 30 minutes. The DNA was again purified with the Nucleotide Removal Kit and eluted in deionized water.

Selective labeling of 5fC An identical protocol was used as that described above for 5fC and 5caC, but substituting the TDG-N191A mutant for the WT TDG.

Dual labeling 5mC and 5hmC 12.5 pmol DNA was incubated for 2 hours at 37°C with 1.5 µg of TET2-CS, 5mM adenosine triphosphate (New England Biolabs), and 75 nM Fe(NH₄)₂(SO₄)₂ in 50 µL of reaction buffer containing 50 mM HEPES, 50 mM NaCl, 1 mM α-ketoglutarate, 2 mM L-ascorbic acid, and 1 mM DTT (pH 7.5) to fully oxidize both 5mC and 5hmC. The treated DNA was purified with the Nucleotide Removal Kit and eluted in deionized water. The above protocol for dual 5fC and 5caC was then followed for labeling.

Selective labeling of 5mC 40 pmol DNA construct was incubated for 1 hour with 10 pmol of UDP-Glucose (New England Biolabs) and 50 U of T4 phage βGT (New England Biolabs) in NEB4 buffer (50 mM potassium acetate, 20 mM tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9, New England Biolabs) at 37°C. Then, the protected DNA was purified with the Nucleotide Removal Kit and eluted in deionized water. The above protocol for dual 5mC and 5hmC was then followed, resulting in labeling of 5mC alone.

TDG release via phenol treatment Where phenol was used to release TDG from the AP site, the protocol described above was employed with two exceptions. First, no APE1 was included in the base excision mixture (i.e. 40 pmol DNA, 3 mg WT TDG, and 4 μ g BSA in HEMN.1 buffer). Second, directly following the excision step, an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) saturated with tris buffer (pH 8.0) was added and mixed by vortexing for 1 min, segregating the DNA construct into the aqueous (buffer) phase and the protein constituents (TDG, BSA) into the inorganic (phenol-chloroform) phase. The mixture was loaded into a phase-lock tube (5Prime, QuantaBio, Beverly, MA) and centrifuged at 14,000 $\times g$ for 25 min and then an equal volume of pure chloroform was added and centrifuged at the same speed for an additional 20 min to remove any remnant phenol. Finally, the aqueous phase containing DNA was aspirated, purified with the Nucleotide Removal Kit, and eluted in deionized water.

Genomic DNA Fragmentation DNA was extracted from SCC-61 and rSCC-61 cells following irradiation using the DNeasy Blood and Tissue Kit (Qiagen) and resuspended in 200 μ L of deionized water. 2 μ g of each sample was then incubated with 4 μ L of NEBNext dsDNA Fragmentase (NEB) in 20 μ L of 1X NEBNext dsDNA Fragmentase Buffer (NEB) at 37°C for 2 hours. 5 μ L of 0.5 M EDTA was added immediately after incubation and vortexed to stop the reaction. Then, the fragmented DNA was purified with the Nucleotide Removal Kit and eluted in deionized water.

Following fragmentation, 1.5 μ g of DNA was blunted by incubating the DNA with 1 nmol each of dNTPs (dATP, dCTP, dGTP, dTTP) (NEB) and 5 U DNA Polymerase I, Large

(Klenow) Fragment (NEB) in 1X NEB2 Buffer at 25°C for 15 minutes. Afterwards, the blunted DNA was purified with the Nucleotide Removal Kit and eluted in deionized water.

Adapter Ligation To allow for the analysis of any blunt-ended DNA with qPCR, adapters were ligated to the ends of blunt-ended DNA. To do this, 1.5 µg of blunted DNA was incubated with 10 U Terminal Transferase (NEB) in 1X Terminal Transferase Buffer (NEB) supplemented with 250 µM CoCl₂ and 100 µM dATP at 37°C for 30 minutes to attach a single dATP overhang to the 3' ends of the DNA. The single A-tailed DNA was then purified with the Nucleotide Removal Kit and eluted in deionized water.

Then, the A-tailed DNA was incubated with 500 µM of a hairpin adapter (5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCdUACACTCTTTCCCTACACGACGCTCTTCCGATC*T-3', dU = deoxyribouracil, * = phosphorothiolate bond) (IDT) and 600 U T4 DNA Ligase (NEB) in T4 DNA Ligase Buffer (NEB) at room temperature overnight before purification with the Nucleotide Removal Kit and elution of the adapter-labeled DNA in deionized water. The uracil within the hairpin adapter was excised as per the uracil labeling protocol presented above to regenerate double-stranded DNA from the DNA loop produced after adapter ligation.

2.2 Protein Expression

APE1 D308A protein expression APE1 D308A plasmid (provided by the Demple Lab, Stony Brook University) was transformed into BL21*(DE3) cells and grown in 1 L LB broth at 37°C. After bacterial cell cultures reached OD₆₀₀=0.6, expression was induced with

0.5 mM isopropyl β -D-thiogalactopyranoside (IPTG). The cultures were then incubated for another 90 minutes before being harvested by centrifugation, resuspended in 50 mM HEPES-KOH (pH 7.5), 100 mM KCl, 1 mM EDTA, 0.1 mM DTT, and 10% (v/v) glycerol, and lysed by two passes through an EmulsiFlex-C5 (Avestin, Ottawa, Canada). The lysate was cleared by centrifugation at $20,000\times g$ for 20 minutes, loaded onto a 15 mL SP Sepharose column (GE Healthcare, Pittsburgh, PA), and eluted with a linear gradient of 100-750 mM KCl. Elutions were analyzed by SDS-PAGE and fractions containing the protein were pooled and dialyzed overnight at 4°C against APE1 storage buffer (50 mM HEPES-KOH (pH 7.5), 200 mM KCl, 1 mM EDTA, 0.1 mM DTT, 10% (v/v) glycerol) and concentrated using 10 kDa MWCO centrifugal spin filter columns (EMD Millipore, Billerica, MA). Final protein concentration was determined with the Bio-Rad Protein Assay (Bio-Rad) and aliquots were stored at -20°C prior to use.

TDG protein expression from E. coli For expression of WT TDG, we followed an existing protocol²¹ adapted from prior work¹⁹ with minor modifications. A plasmid for human TDG based on pET28 was transformed into BL21(DE3) cells and grown in 1 L LB broth at 37°C until the cultures reached an OD600 of 0.6. Then, they were gradually cooled to 16°C , induced with 0.25 mM IPTG and incubated overnight. Harvesting was performed by centrifugation and retrieved cells were resuspended in 20 mL of TDG lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 25 mM imidazole) with protease inhibitors and then lysed by two passes through an EmulsiFlex-C5 homogeniser. The lysate was cleared by a 20 min centrifugation at $20,000\times g$, loaded onto a 1 mL column of HisPur cobalt resin (Fisher Scientific, Hampton, NH) equilibrated with TDG lysis buffer, and then bound by

two applications of the lysate to the column under gravity flow. The column was washed with 20 mL of TDG lysis buffer and subsequently eluted by a linear gradient of imidazole (100-500 mM) into 1 mL aliquots that were then analyzed by SDS-PAGE. Fractions containing TDG were pooled and dialyzed overnight at 4 °C against TDG storage buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM DTT, 0.5 mM EDTA, 1% v/v glycerol). Dialyzed proteins were concentrated using 10 kDa molecular weight cutoff centrifugal spin filter columns. Final protein concentration was determined analytically by Bradford protein assay and aliquots were stored at -80°C prior to use.

A mutant TDG²² with no recognition for 5caC (TDG-N191A) was expressed in an identical fashion to WT TDG but using the mutant plasmid.

Human TET2-CS, the crystal structure variant of the enzyme (1129-1936 Δ 1481-1843), was purified from insect cells as previously described²³. Briefly, the construct, with an N-terminal FLAG tag, was subcloned into a pFastBac1 vector. After generation of baculovirus, 1L of Sf9 cells were infected and cells were collected after 24 h and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, and 0.2% (v/v) NP-40) containing complete, EDTA-free Protease Inhibitor Cocktail (Roche, 1 tablet/10 mL). Cells were lysed by three passes through a microfluidizer at 15,000 psi and the lysate was cleared by centrifugation at 20,000 $\times g$ for 30 min. The supernatant was then passed three times over a 1 mL packed column of anti-FLAG M2 affinity resin (Sigma). The column was washed three times with 10 mL of wash buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, and 15% (v/v) glycerol). 1 column volume of elution buffer (wash buffer with 100

$\mu\text{g/mL}$ 3 \times FLAG peptide (Sigma) added) was then incubated on the column for 10 min followed by collection of the elution fraction. Serial elutions were similarly collected until no more protein was detected by the Bio-Rad Protein Assay. The three most concentrated fractions were pooled, aliquoted, and stored at $-80\text{ }^{\circ}\text{C}$.

2.3 Analytical Techniques

Gel Electrophoresis The DNA denaturing gel was prepared by mixing thoroughly 70 mL of 23% gel matrix (22% acrylamide, 1% bis-acrylamide, 7 M urea in 1X 3:1:1 tris/borate/EDTA (TBE) buffer), 240 μL of 25% ammonium persulfate, and 42 μL tetramethylethylenediamine. The gel mixture was cast and allowed to polymerize for 30 minutes before running samples with dye in 1X 3:1:1 TBE at 55 W for 90 minutes. Yields were approximated by measuring product band intensity relative to intermediates in the final lane using ImageJ analysis software²⁴. For electromobility shift assays (EMSA) gels (see Supplementary Fig. S3), 3.5% agarose gels were prepared in 1X TBE buffer with GelRed nucleic acid stain (Phenix Research Products, Candler, NC). Gel images were acquired using a Gel DocTM system (Bio-Rad, Hercules, CA).

SS-Nanopore measurements Fabricated silicon chips (4 \times 4 mm), each supporting a 10-20 μm thin film silicon nitride window (20 nm thickness) were obtained commercially (Norcada, Inc., Alberta, Canada). A single SS-nanopore (diameter 7.5-9.0 nm, as determined from resistance measurement²⁵) was produced in each membrane using a helium ion milling technique described elsewhere²⁶. Prior to measurement, a chip was rinsed with deionized water and ethanol, dried under filtered air flow, and then exposed to

air plasma (30 W) for 2 min on each side before being placed into a custom Ultem 1000 flow cell that enabled introduction of measurement buffer (1 M NaCl, 10 mM PBS buffer) to independent reservoirs on each side of the device. Ionic current measurements were performed with a patch clamp amplifier (Axopatch 200B) through Ag/AgCl electrodes and used to verify pore diameter. After introducing biomolecules in measurement buffer to the cathode chamber, current was recorded at a bandwidth of 200 kHz with a 100 kHz four-pole Bessel filter. Analysis was performed with custom software and an additional low-pass filter of 25 kHz. The event threshold for analysis was set at 4.5 standard deviations above the RMS noise level and only events with durations between 12.5 μ s and 2.5 ms were considered. Each rate measurement was determined by considering at least 3.5 min of uninterrupted trace recording, broken into segments of 3.2 s. The standard deviation between segments was taken as the measurement error.

Preparation of synthetic DNA by PCR 150 bp ds-DNA oligonucleotides containing a single biotin modification were prepared by PCR using λ -phage DNA (New England Biolabs) as a template. The unmodified forward primer (5'-AAC AAC TGT TTC AGC CAC TGC TTC-3') and the biotinylated reverse primer (5'-CAG TTG AGG ATC CCC ATA ATG CG -3', where T is a biotinylated base) were synthesized commercially (Integrated DNA Technologies, Coralville, IA). The PCR product was subsequently purified using a QIAquick PCR Purification Kit (Qiagen) and eluted in deionized water (EMD Millipore, Billerica, MA). The resulting stock solution was determined to have a concentration of 0.6 μ M by spectrophotometry (Nanodrop 2000c, ThermoScientific, Waltham, MA).

Binding of biotinylated DNA to free monovalent streptavidin For measurements involving free streptavidin, a mutant variant of streptavidin, MS²⁷, was employed consisting of a single active high-affinity biotin-binding region. The binding reaction was performed by mixing 100 nM MS with 50 nM biotinylated 150 bp dsDNA in 1x PBS and incubating for at least 10 minutes at room temperature. The single biotin-binding site of MS ensured purely symmetric binding of one protein to one DNA.

Binding of biotinylated DNA to streptavidin-coated magnetic beads For bead capture measurements, 10 µL of streptavidin conjugated beads (1 µm diameter Dynabeads MyOne Streptavidin C1 beads, Invitrogen, Carlsbad, CA) were washed three times in 1X binding/washing buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl) before being resuspended in 10 µL of 2X binding/washing buffer. 10 µL of biotinylated 150 bp dsDNA at a concentration of 50 nM were then added to the beads and agitated for 30 minutes.

Preparation of binary biotinylated and non-biotinylated DNA mixture 4.8 µg λ-phage DNA (New England Biolabs) were incubated at room temperature for 30 minutes with 3.3 µM dGTP (New England Biolabs), 3.3 µM biotinylated dCTP (Trilink Biotech, San Diego, CA), 6.7 µM ddATP (Trilink Biotech), and 5 units of Klenow Fragment (New England Biolabs) in 30 µL of 1X NEBuffer 2. Given the *cos* overhang sequences of λ-phage DNA (5' GGGCGGCGACCT 3' and 5' AGGTCGCCGCC 3'), this results in one end of the DNA being biotinylated and the other being capped to prevent further incorporation. The product was purified by conventional phenol-chloroform extraction and 1 µg of the labeled DNA was incubated at 37°C for 1 hour with 5 units of PspXI (New England Biolabs) in 50

μ L of 1X CutSmart Buffer to cut the DNA into fragments of approximately 33.5 kbp and 15 kbp. The resulting product was purified again by conventional phenol-chloroform extraction. For bead capture, 500 ng of the resulting mixture was incubated with 25 μ L of streptavidin beads resuspended in 2X binding/washing buffer, with the rest reserved for SS-nanopore measurements. Note that this the amount required according to the bead capacity supplied by the manufacturer, and which we have determined to be sufficient in capturing all the material. The binding reaction was incubated at room temperature for 1 hour under constant agitation. After agitation, the beads were immobilized magnetically and the supernatant (containing non-biotinylated DNA) was discarded. Finally, the beads were washed three times with 1X binding/washing buffer to remove any non-specifically bound DNA and re-suspended in 25 μ l of deionized water. The bound material was then eluted from the beads at room temperature using 12.5% (v/v) phenol in chloroform for SS-nanopore measurements.

Dissociation by phenol Phenol dilutions were prepared by adding pure chloroform (Ricca Chemical Company, Arlington, TX) to 25:24:1 phenol:chloroform:isoamyl alcohol (Acros Organics, Morris, NJ) until the desired proportion of phenol was obtained. Phenol-chloroform was added to an equal volume of sample (10 or 25 μ L) in binding/washing buffer and thoroughly mixed by vortexing for 1 minute. In the case of bead-bound streptavidin, beads were removed by placing the sample tube on a magnet and decanting. For both processes, the bi-phasic mixture was then transferred into a phase-lock tube (5Prime, QuantaBio, Beverly, MA) and centrifuged at 14,000g for 25 min. An additional sample volume of chloroform was added to the aqueous phase and the mixture was

centrifuged at the same speed for an additional 20 min. Finally, the aqueous phase containing eluted DNA was aspirated for subsequent use or analysis. All steps were performed at room temperature. Control experiments were performed with pure chloroform or deionized water instead of phenol-chloroform.

Streptavidin bead capture and elution for the SCC-61 and rSCC-61 experiments For bead capture measurements, 25 μL of streptavidin conjugated beads (Dynabeads M-280 Streptavidin beads, Invitrogen, Carlsbad, CA) were prepared for each sample. The beads were washed three times in 1X binding/washing buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl) before being resuspended in 8 μL of 2X binding/washing buffer. 8 μL of purified, labeled DNA were then added to the beads and agitated overnight.

Following bead capture, the beads were washed twice with 25 μL of 1X binding and washing buffer, washed once again with 25 μL of 100 mM NaCl, and resuspended in 25 μL of 100 mM NaCl. Then, 25 μL of water-saturated 25:24:1 phenol:chloroform:isoamyl alcohol (Acros Organics, Morris, NJ) was added to the resuspended beads, vortexed for 30 s, and left to sit for 30 minutes. After allowing time for dissociation, the sample was spun down in a centrifuge at 13,000 rpm for 25 minutes, and 20 μL of the aqueous layer was decanted into a new PCR tube. 20 μL of pure chloroform was added to the removed aqueous layer, vortexed for 30 s, and spun down at 13,000 rpm for 20 minutes to isolate any remaining phenol contamination. Subsequently, 12 μL of the aqueous layer was removed and stored in another clean PCR tube.

Finally, the isolated DNA was cleaned with ethanol precipitation to remove any remaining organic contaminants. To the recovered material, 0.1 volume of 5M NH₄OAc (Invitrogen), 0.05 volume of glycogen (20 mg/mL) (Molecular Biology Grade, Thermo Fisher Scientific, Waltham, MA), and 1 volume of isopropanol (Thermo Fisher) was added before incubating at -20°C overnight. The DNA and glycogen was pelleted by spinning the mixture at 13,000 rpm for 30 minutes at 4°C, after which the supernatant was removed without disturbing the pellet. Then, 150 µL of ice-cold 70% ethanol was added and vortexed for 10 s before spinning down again at 13,000 rpm for 20 minutes at 4°C, with the supernatant removed again afterward. The recovered material was then allowed to air-dry at room temperature before being resuspended in 12 µL RT-PCR grade water (Thermo Fisher Scientific).

qPCR Protocol The qPCR master mix was prepared for each sample by mixing 5 µL 2X SYBR Green PowerUp Master Mix (Applied Biosystems, Foster City, CA), 0.4 µL of forward primer (5'-GACGTGTGCTCTTCCGA-3') at 5 µM (IDT), 0.4 µL of reverse primer (5'-CACGACGCTCTTCCGAT-3') at 5 µM (IDT), and 2.2 µL PCR-grade water (Thermo Fisher Scientific) and scaling up to the required number of samples. In a 96-well qPCR plate, 2 µL of no template control, standards, and samples were added in triplicate, to which 8 µL of the master mix was added. The qPCR protocol consisted of an initial enzyme activation stage of 95°C for 50 s, 57°C for 40 s, and 72°C for two minutes. This was followed by 50 cycles of 95°C for 35 s, 57°C for 30 s, and 72°C for two minutes to replicate the sample DNA, and a final extension step of 72°C for two minutes. The results from these experiments were compiled and analyzed.

CHAPTER 3 – DEVELOPMENT OF A MODULAR ENZYMATIC LABELING PROCESS FOR DNA BASE MODIFICATIONS

3.1 Introduction

A variety of non-canonical bases are prevalent in genomic DNA and play crucial roles in cell functions that include gene expression and suppression²⁸, transposon expression^{29,30}, stem cell differentiation³¹, and chromosomal inactivation³². For example, the abundance and position of epigenetic modifications are tightly regulated and errors in this regulation have been linked to a wide range of diseases³³ including cancer. In addition, DNA base damage elements generated both endogenously and exogenously are a major source of point mutations if not correctly repaired by cellular processes. The locations of these elements can be random or could be linked to sequence accessibility in chromatin structures. While the impact of modified DNA bases is clear, their detection can be challenging due to a variety of issues. Conventional detection technologies such as sequencing, chromatography, and immunochemical assays lack the ability to identify the wide variety of non-canonical bases that may be present in DNA³⁴. Furthermore, cross-reactivity of antibodies³⁵ and non-specific induction of additional lesions¹⁷ can result in an overestimation of the amount of damage present in the DNA. Thus, there is a need for a technique that can recognize and accurately quantify specific DNA modifications. One approach that addresses these concerns has been pioneered recently by Song, *et al*³⁶ in which a single, high-affinity tag was attached enzymatically to 5-hydroxymethylcytosine bases, permitting downstream analysis, enrichment, or sequencing. While the process has been adapted to access some additional elements of the demethylation pathway³⁷⁻³⁹, only a limited suite of modifications are suitable for such tagging. As the ability to probe a wide

variety of base modifications would be of significant value, we set out to develop a new labeling strategy that would be permissive for a host of different DNA modifications. In our methodology, we exploit the enzymatic machinery of the DNA base excision repair (BER) pathway, which identifies and restores base lesions *in vivo*.

3.2 Development of a method of labeling uracil and 8-oxoguanine

In our approach (Figure 1), a modified base element was first excised from the DNA using a DNA glycosylase, which removed the target base from the phosphate backbone, leaving an abasic (AP) site. If the glycosylase was bifunctional (i.e. had AP lyase activity), the phosphodiester bond was also cleaved 3' to the modification, leaving a single strand nick. Ensuing steps were not affected by this activity. Next, an AP endonuclease was used to cleave the phosphodiester bond 5' to the abasic site and remove the exposed 3' phosphate, leaving a hydroxyl group that was amenable to the final step: treatment with a gap-filling polymerase to incorporate a biotin-conjugated nucleotide into the DNA structure. For this, we used a mutant polymerase lacking 3'-5' exonuclease activity (T4(exo-)) and provided it with only of the cognate biotin-dNTP, resulting in the insertion of a single affinity label at the precise location of the modified base.

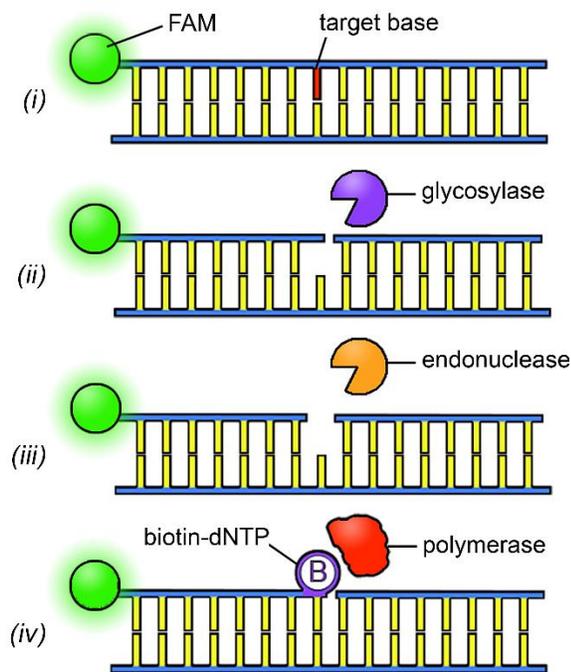


Figure 1. Schematic representation of the general labeling approach

(i) A duplex DNA molecule featuring a target base element (red). (ii) A glycosylase recognizes and excises the base element (diagram shows activity of a bifunctional glycosylase that nicks the phosphate backbone 3' to the excision). (iii) An AP endonuclease cuts the backbone 5' to the excision. (iv) A gap-filling polymerase incorporates a single biotinylated nucleotide at the modification position.

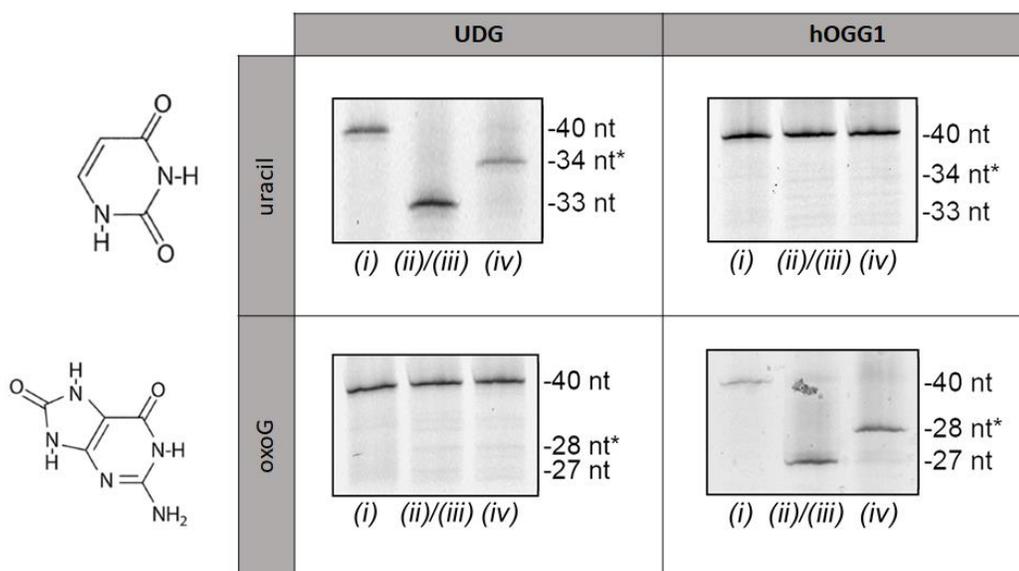
We have found the absence of exonuclease activity to be particularly important as processive cleavage of nucleotides from the modification site can result in prevention or misincorporation of the biotinylated nucleotide label. We note that this methodology ultimately resulted in a nick 3' to the inserted biotin-dNTP. While it is possible to repair this nick through ligation, we did not include such a step because of the potential for

reduced product yield. The presence of the nick did not negatively impact subsequent measurements.

Crucially, this general approach could be used to target a variety of distinct modified bases through variation of two central components: the DNA glycosylase, selected for recognition of a particular lesion, and the biotin-conjugated nucleotide, selected to match the canonical identity of the target modified base (or in the case of a mismatch target, the appropriate nucleotide for Watson-Crick base-pairing with the opposite strand). As an initial demonstration of this modularity, we first showed selective detection of uracil and oxoguanine (oxoG) bases. Uracils arise in DNA upon deamination of cytosine, resulting in a mutagenic U:G mismatch, or upon misincorporation of dUTP, resulting in a genotoxic U:A pair⁴⁰. Meanwhile, oxoG is the major oxidative base damage associated with reactive oxygen species (ROS) due to the low redox potential of guanine and has known mutagenic potential via transversion during DNA replication⁴¹. For these measurements, we used synthetic 40 bp double-strand (ds-) DNA oligonucleotides, with one strand containing the target modified base at a known position and a fluorescent FAM label at the 5' end. We utilized endonuclease IV (EndoIV) to prime the excised gap for T4(exo-) incorporation of a biotin-dNTP.

Denaturing gel analysis of each sequential step for the two bases using an appropriate glycosylase/nucleotide combination showed excision of the modified base and incorporation of the biotin-dNTP. Labeling of uracil was achieved using a combination of uracil DNA glycosylase (UDG) and biotin-dUTP while oxoG labeling employed human

oxoG DNA glycosylase (hOGG1) and biotin-dGTP. Notably, we made use of a “one-pot” treatment for each of these targets that minimized material loss and enabled high product yields of ~92% and ~83%, respectively. In addition, UDG is a monofunctional glycosylase while hOGG1 is bifunctional, showing that the approach was not affected significantly by either absence or presence of AP lyase activity in the glycosylase. Identical treatments of each base modification with non-target components showed no detectable labeling, highlighting process selectivity that was facilitated by the low cross-recognition of each glycosylase (Figure 2).



*Figure 2. Denaturing gel analyses of labeled uracil and oxoG-containing DNA constructs (Steps numbered as in Figure 1) Lane 1: annealed oligonucleotide; lane 2: following glycosylase/endonuclease treatment; lane 3: following T4(exo-) fill-in. * indicates DNA length plus biotin tag. Left: molecular structures of the target bases.*

3.3 Development of a method of labeling T:G mismatches and ethenoadenine with APE1 displacement

While these data clearly demonstrated a flexible approach that could in principle be extended to a broad range of base targets⁴², glycosylases can also have additional activities that could interfere with the labeling procedure as described. For example, thymine DNA glycosylase (TDG) is a major component of the cytosine demethylation process, recognizing T:G mismatches⁴³ among other elements⁴⁴, but it also recruits additional enzymes like histone acetyltransferases⁴⁵. Because of this latter role, TDG has a high affinity for the AP site resulting from base excision, making it difficult to detach for subsequent labeling steps (Figure 3a).

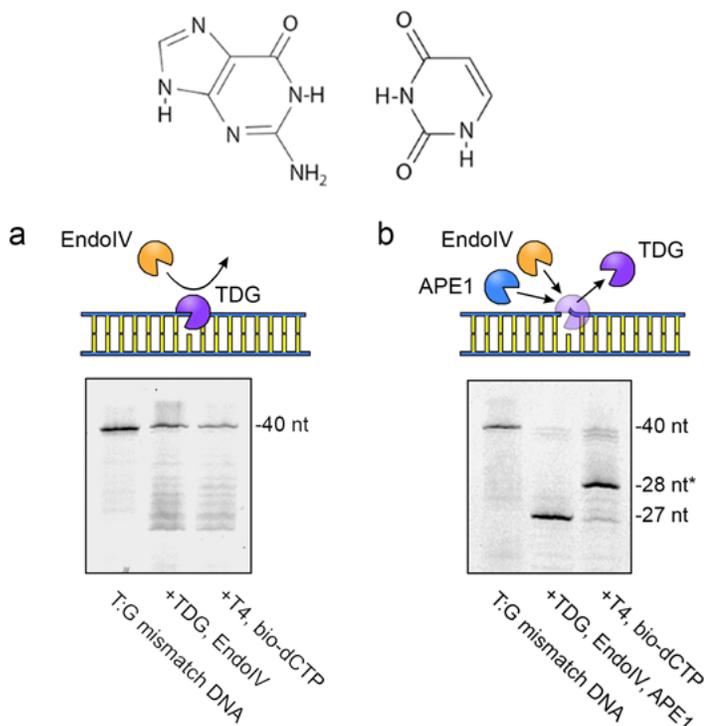


Figure 3. Denaturing gel analyses of labeled T:G-mismatch-containing DNA constructs

Top: molecular structure of T:G mismatch

(a) Top: Schematic showing inaccessibility of AP site by EndoIV caused by TDG binding. Bottom: denaturing gel of labeling steps for a T:G mismatch oligonucleotide using EndoIV only. No significant labeling is observed. (b) Top: schematic showing release of TDG by APE1, leaving DNA accessible by EndoIV. Bottom: denaturing gel of labeling steps for a T:G mismatch oligonucleotide using both EndoIV and APE1 (D308A mutant), indicating recovery of high yield labeling.

To address this, we sought to promote enzyme disengagement through the incorporation into the protocol of an additional endonuclease, AP endonuclease 1 (APE1). The extensive dsDNA binding surface of APE1 and the prominent kinking it induces in the DNA helix⁴⁶ have been suggested as means to promote displacement of glycosylases more efficiently than EndoIV⁴⁷. However, the improved activity of APE1 comes at the expense of 3'-5' exonuclease activity not found in the other enzyme, especially under key buffer conditions⁴⁸. To partially mitigate this effect, we used the APE1 D308A mutant²⁰, which features reduced 3'-5' exonuclease activity. This inclusion improved yield significantly over wild type APE1 (Figure 4), but the remaining nucleotide digestion activity still necessitated a supplementary purification step prior to polymerase gap-filling to limit decomposition of the DNA. While this increased the number of steps and decreased overall product yield somewhat, the resulting material showed successful incorporation of biotin-labeled nucleotides on gel (Figure 3b) at a high yield (~73%).

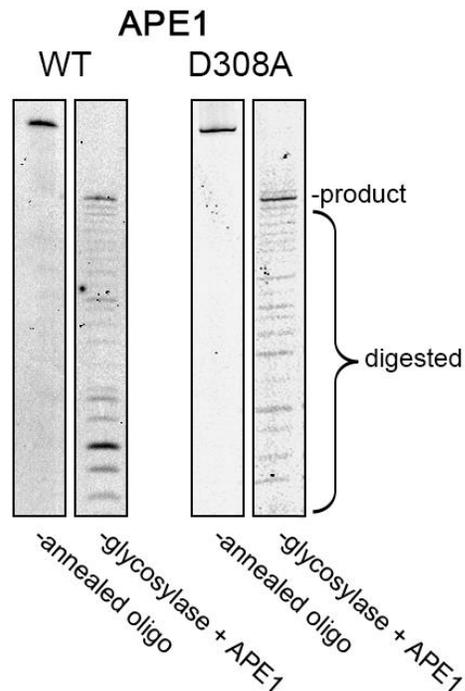


Figure 4. WT APE1 and D308A APE1 exonuclease activity.

Denaturing gel analysis of a FAM-labeled 40 nt T:G mismatch DNA oligonucleotide before and after treatment with TDG and APE1. Left gels show WT APE1 behavior, resulting in massive loss of product due to 3'-5' exonuclease activity. Right gels show same treatment with the APE1 D308A mutant. While some exonuclease activity is observed in our buffer conditions, it is significantly reduced compared to WT.

Notably, this alternative method could be used to incorporate other glycosylases with similar behavior as well. As an example, we utilized human alkyladenine DNA glycosylase (hAAG), which excises alkylated bases from DNA, but has also been observed to bind tightly to its DNA template⁴⁹. The major target of hAAG is the important epigenetic element methyladenine⁵⁰, but this base is known to be unstable for *in vitro* measurements.

Consequently, we instead used for our demonstration a synthetic oligonucleotide featuring the methyladenine analog 1,N⁶-ethenoadenine, and employed hAAG and biotin-dATP for labeling. Subsequent analyses of the product again indicated efficient (~74%) labeling on gel (Figure 5). Thus, we are able to demonstrate the broad modularity of the labeling scheme.

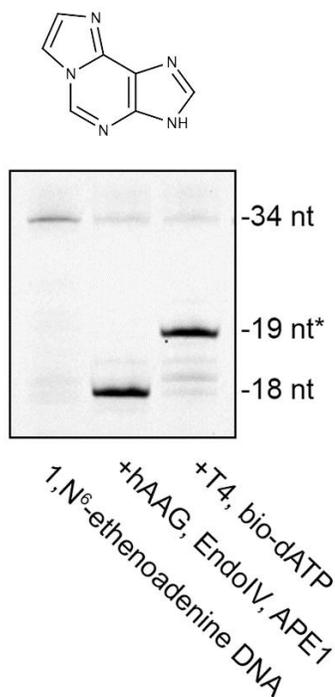


Figure 5. Denaturing gel analyses of labeled 1,N⁶-ethenoadenine-containing DNA constructs

Top: molecular structure of 1,N⁶-ethenoadenine. Bottom: Denaturing gel of labeling steps for a 1,N⁶-ethenoadenine oligonucleotide using EndoIV and APE1 (D308A mutant), showing high yield biotin labeling.

Unlike the labeling methods developed for uracil and oxoG, “one-pot” reactions were not viable for the labeling of T:G mismatches and 1,N⁶-ethenoadenine due to the requisite

buffer compositions for the TDG and hAAG glycosylases, which differed greatly from the buffers that could be utilized by the subsequent EndoIV and T4(exo-) reactions. This may partially account for the lower yield in the labeling of these modifications in comparison due to the removal of the glycosylase during additional purification steps during the labeling process to facilitate buffer exchange, thus preventing the glycosylase from continuing to act on any oligo that retained the modification of interest. However, it may be possible to further increase the yields of these reactions by increasing the incubation time with the glycosylase or increasing the concentration of glycosylase within the reaction in order to fully excise the modification of interest prior to the addition of endonuclease.

3.4 Optimization of protocols to improve labeling accuracy

Upon the performance of electromobility shift assay (EMSA) gels by conjugating monovalent streptavidin (MS) to biotin-labeled nucleotides, we observed the presence of secondary shifts in some gels, with a prevalence of 7.6% in uracil-containing oligos and 4.0% in 1,N⁶-ethenoadenine-containing oligos. We suspected that these secondary shifts, indicative of the presence of a second biotinylated nucleotide in the labeled oligo, could be the result of non-specific end-labeling of the oligos by the T4 polymerase. To verify whether this was the case and to correct this erroneous labeling, which could result in false positives in applications of this labeling method, we performed end-labeling reactions on oligos prior to glycosylase labeling with terminal transferase and the chain-terminating dideoxyadenosine triphosphate (ddATP), the latter of which lacks 2' and 3' hydroxyl groups. This structure of the ddATP renders the resulting 3' end of the DNA oligo incapable of receiving further deoxynucleotide additions⁵¹, including that of biotinylated

deoxynucleotides. We found that this change in the protocol was capable of eliminating the occurrence of secondary shifts, indicating an inhibition of the undesired incorporation of an additional biotinylated nucleotide (Figure 6).

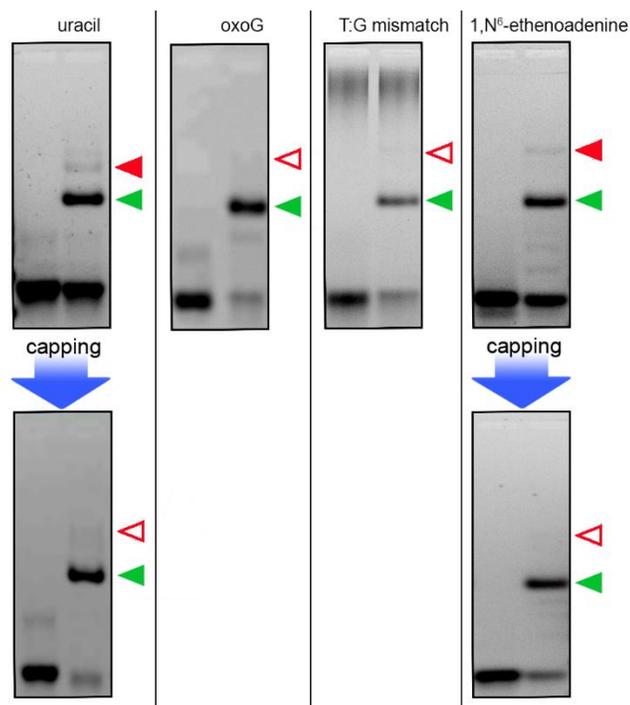


Figure 6. Limiting off-target T4 polymerase end-labeling with terminal transferase

Electromobility shift assay (EMSA) gels for all specifically-labeled DNA constructs. In each image, labeled DNA is visualized without (lane 1) and with (lane 2) MS. Green arrows indicate target DNA-MS construct. For uracil (left) and 1,N⁶-ethenoadenine (right), an additional band was observed (red arrow), indicating superfluous end-labeling of the construct by the T4 polymerase, possibly driven by incorporation kinetics of dUTP and dATP, respectively. By capping the 3' ends of these oligonucleotides with terminal transferase and ddATP (see Materials and Methods) prior to labeling (blue arrows), we find that multiple biotin-labeling is inhibited, with results similar to those of oxoG and T:G mismatch (open red arrows).

3.5 Discussion

We have shown that a variety of single-base modifications can be assessed by incorporating an efficient and targeted affinity-labeling technique that exploited the physiological activities of enzymes involved in the BER pathway to install a single biotin tag at the precise location of a given base element. We first showed selective recognition of uracil and oxoG bases with the glycosylases UDG and hOGG1, respectively. Next, we sought to utilize other glycosylases by integrating a mixture of endonucleases designed to promote enzyme release and limit DNA digestion. While the alternative procedure entailed some loss of material due to increased exonuclease activity, it enabled the use of glycosylases that are specifically challenging to incorporate in the labeling approach due to strong AP binding capacity. As a demonstration, we showed that this approach could be used for the study of T:G mismatch bases with TDG and the methyladenine analog 1,N⁶-ethenoadenine with hAAG. Therefore, with the flexible protocols established here, nearly any glycosylase could be integrated, facilitating the labeling and analysis of a broad range of bases that they target, including the widely studied methylcytosine¹⁰. The central limiting factor for this capacity is in the specificities of the glycosylases themselves, since many have recognition for multiple elements. However, the affinity for specific targets can vary wildly, offering a potential pathway to high selectivity. We expect that the use of point mutations in the glycosylases may also be able to tailor their specificity and enable high certainty in recognition. In total, this work opens new avenues to study base modifications that may have important impacts on biology and disease, but are challenging to probe with conventional techniques. In addition, the modular labeling approach alone could also be

employed in applications like affinity enrichment and genomic analyses³⁶⁻³⁸, and is amenable to the integration of any label that can be incorporated by polymerase activity, including fluorescent tags.

Peer-Reviewed Publication: F. Wang, O. K. Zahid, B. Swain, D. Parsonage, T. Hollis, S. Harvey, F. Perrino, R. M. Kohli, E. W. Taylor, A. R. Hall. "Solid-state nanopore analysis of diverse DNA base modifications using a modular enzymatic labeling process." *Nano Letters*, 17 (11), 7110–7116, 2017.

CHAPTER 4 – LABELING OF PRODUCTS OF THE CYTOSINE

DEMETHYLATION PATHWAY

4.1 Introduction

Composed structurally of a cytosine nucleobase with a methyl group at the fifth carbon atom, the epigenetic modification 5-methylcytosine (5mC) has an overall prevalence of ~4% (5mC/C) in the human genome⁵². It is the most widely studied DNA base variant, largely because of the early advent of a technique with which it could be probed; it was demonstrated^{53,54} as early as 1970 that exposure to sodium bisulfite is capable of deaminating cytosines and converting them to uracils, but that this chemical reaction is blocked by methylation. In combination with the growing availability of sequencing technologies, this simple treatment has enabled a large number of studies that have been able to determine the genomic positions of 5mC as well as highlight its importance in diverse biological processes. For example, physiologically, 5mC has been shown to occur primarily in symmetric CpG dinucleotides⁵⁵, where it plays an important role in the regulation of gene expression⁵⁶ and has consequently been implicated in a variety of diseases⁵⁷ including cancer⁵⁸.

While bisulfite treatment is the gold standard for DNA epigenetic analysis, it has two significant drawbacks. First, the procedure induces widespread damage to DNA in general. Bisulfite conversion of cytosines requires a single-strand target, so the process is typically carried out at elevated temperature. This, combined with the chemical reactivity of sodium bisulfite itself, results in substantial fragmentation of the DNA⁵⁹ that can reduce its viability for downstream analyses and places practical limitations on the minimum starting DNA

mass. Second, bisulfite conversion is limited in its intrinsic ability to resolve multiple cytosine modifications. The recent identification of the ten-eleven translocase (TET) family of enzymes^{60,61} has elucidated the pathway by which cytosine demethylation is achieved physiologically (Figure 7a): 5mC is oxidized in a stepwise fashion by TET to each of the three additional modified bases 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), the final two of which can be excised by thymine DNA glycosylase (TDG) and replaced with canonical cytosine upon completion of base excision repair (BER). Each of the three additional modified bases represents a potentially independent regulatory element, but bisulfite treatment has variable effects on them⁶², with 5mC and 5hmC each blocking conversion and 5fC and 5caC each able to be deaminated. Consequently, analyses incorporating conventional bisulfite treatment are inadequate to probe all components of the demethylation pathway. Innovative and effective strategies have been developed to expand possible base targets, but most still employ bisulfite^{63–65,37,66} (and thus still encounter the challenge of DNA damage above).

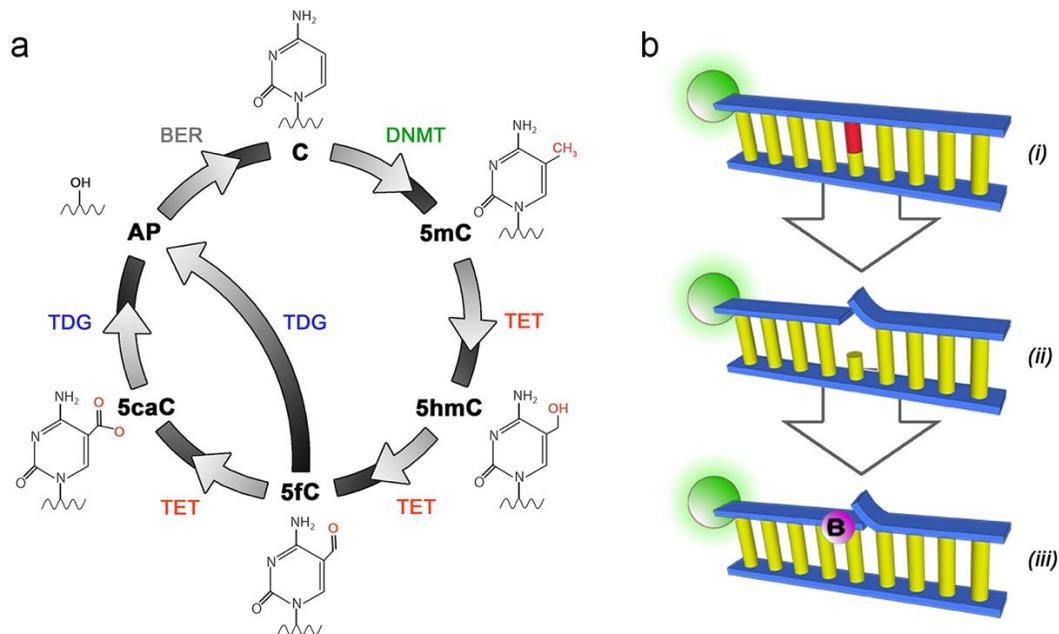


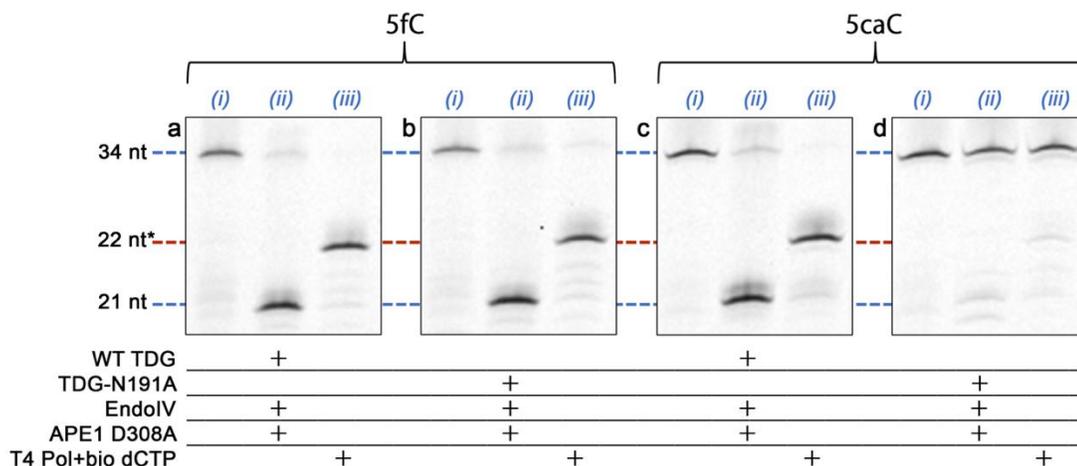
Figure 7. Schematics of the cytosine methylation/demethylation pathway and the generalized labeling scheme.

(a) The cytosine methylation/demethylation pathway. Canonical cytosine (C) is methylated to 5mC by DNA methyltransferase (DNMT), which can then undergo TET oxidation to 5hmC, 5fC, and 5caC sequentially. Both 5fC and 5caC are recognized by TDG, leaving an AP site and enabling the BER pathway to install a canonical C. (b) Steps of the generalized labeling scheme. DNA (i) containing a single modified base (red) is treated with a targeting glycosylase (a monofunctional glycosylase for illustration) to excise the base and endonuclease and produce a site for polymerase activity (ii). A gap-filling polymerase is then used with a matched dNTP containing a biotin ('B') to install an affinity moiety at the precise location of the original modified base (iii). Illustration shows the fluorescent FAM label (green) employed in our constructs.

Driven partially by the interest generated by recent non-bisulfite approaches to the analysis of DNA epigenetics^{67,68}, we present a modular approach for installing a single affinity label at the precise locations of cytosine modifications and demonstrate adaptations to the process that enable all four elements of cytosine demethylation – 5mC, 5hmC, 5fC, and 5caC – to be assessed. We show that modified bases can be replaced by a biotinylated nucleotide with high efficiency, providing a mechanism for selective isolation by e.g. streptavidin-driven affinity precipitation.

4.2 Development of protocols to selectively label the substrates of the cytosine demethylation pathway: methylcytosine, hydroxymethylcytosine, carboxylcytosine, and formylcytosine

First, we exploit the capability of wild-type (WT) TDG to excise both 5fC and 5caC. As a demonstration, we perform our full labeling procedure on test double-strand (ds) DNA oligonucleotides 34 bp in length that feature a single base modification positioned 22 nt from a fluorescent 5' reporter. Figure 8a,c shows the results of the sequential process for both of the modifications, as demonstrated by a denaturing gel that follows the single DNA strand featuring the 5' fluorescent label (Figure 7b).



*Figure 8. Denaturing gel analyses of 34 nt DNA constructs featuring 5fC and 5caC. Gels show 5fC (a-b) or 5caC (c-d) at base position 22 and labeled using either WT TDG (a, c) or TDG-N191A mutant (b, d). Steps numbered (i-iii) as in Figure 7. Lane (i): annealed oligonucleotide; lane (ii): following glycosylase/endonuclease treatment; lane (iii): following polymerase fill-in with a biotinylated nucleotide to yield a labeled product (red). Construct lengths at left apply to all gels and * indicates DNA length plus biotin tag.*

The initial 34 nt construct (lane 1) is first exposed to WT TDG glycosylase, along with AP endonuclease 1 (APE1 mutant D308A²⁰ with reduced exonuclease activity) to displace the glycosylase⁶⁹, which is known to bind tightly to the DNA substrate⁴⁹. After a subsequent treatment with EndoIV to nick the DNA 5' to the remnant AP site, we observe a shorter 21 nt product (lane 2), consistent with the position of the modification at base 22. After incubation with T4 DNA polymerase and biotinylated dCTP to fill the gap, the product increases in molecular weight to greater than 22 nt (lane 3); note that the shift appears larger than 1 nt because of the added mass of the attached biotin.

Our results show high labeling yield with WT TDG for both 5fC and 5caC (87% and 79%, respectively), but also demonstrate a lack of differentiation with the procedure. To rectify this, we employ a mutant TDG (TDG-N191A) that has been shown²² to have selectivity for 5fC in particular. Repeating our procedure with this alternative glycosylase, we find that the 5fC construct yields the same characteristic shifts observed for WT TDG (Figure 8b), indicating a comparable high labeling yield (80%). In contrast, the 5caC construct results in minimal shifts (<10%) through the same process (Figure 8d), confirming the lack of 5caC recognition by the mutant TDG and indicating that no label is inserted. Consequently, the combined use of WT TDG and TDG-N191A in separate treatments can be used to deliver information about both modified bases through differential analysis. We expect that utilization of another recently discovered⁷⁰ mutant TDG (N157D) with specific recognition for 5caC only could also be used for completely independent analyses.

Having established protocols to assess 5fC and 5caC, we next investigate 5mC and 5hmC as base targets. For recognition of these two modifications collectively, we first employ TET to oxidize them and then subsequently carry out labeling with WT TDG as above. While TET oxidation converts these bases sequentially through each successive derivative, 5caC is the terminal element in the process. Consequently, the treatment can be performed to completion rather than requiring scheduled cessation to capture a particular base modification, in contrast to some existing applications of TET to demethylation analysis³⁸. The results of this overall strategy using oligonucleotides with 5mC and 5hmC are shown in Figure 9.

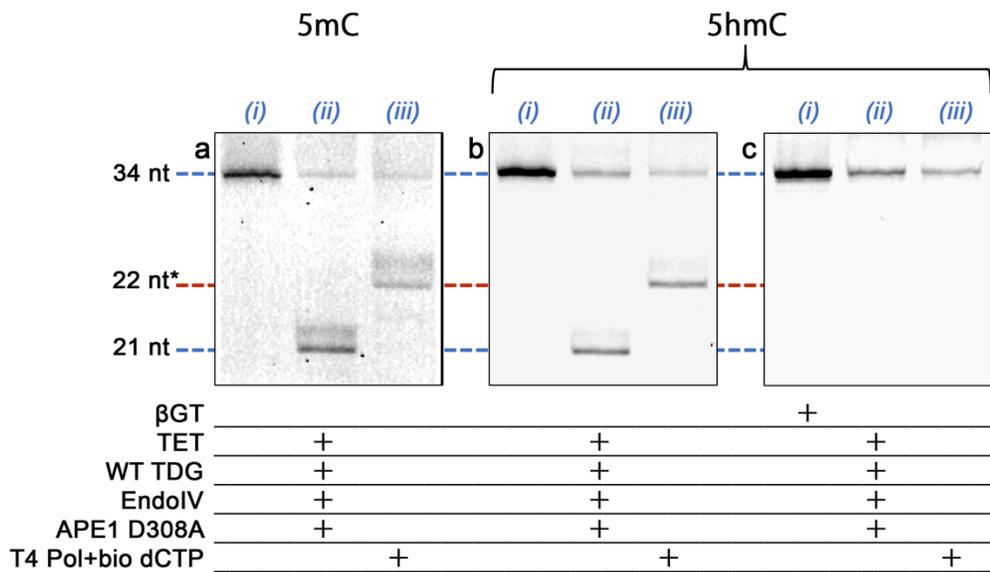


Figure 9. Denaturing gel analyses of 34 nt DNA constructs featuring 5mC and 5hmC.

Gels show 5mC (a) or 5hmC (b-c) at base position 22 and labeled using WT TDG following oxidation of each base with TET. In (c), a treatment with β GT prevents labeling of 5hmC specifically. Steps numbered (i-iii) as in Fig. 1b. Lane (i): annealed oligonucleotide \pm β GT; lane (ii): following glycosylase/endonuclease treatment; lane (iii): following polymerase fill-in with a biotinylated nucleotide to yield a labeled product (red). Construct lengths at left apply to all gels and * indicates DNA length plus biotin tag.

For both, an identical protocol results in effective insertion of biotinylated bases (62% and 65% yield, respectively), demonstrating the effectiveness of WT TDG on the TET-oxidized substrates. No labeling was observed for either base without TET treatment, confirming that WT TDG has no intrinsic recognition for 5mC or 5hmC⁷¹.

As with 5fC and 5caC above, this procedure labels two base modifications simultaneously and so additional steps must be taken to discriminate 5mC and 5hmC. To achieve differentiation, we incorporate an additional treatment with β -glucosyltransferase (β GT), an enzyme that affixes a glucose moiety to 5hmC bases selectively. The presence of this bulky sugar disrupts the target recognition of TET and inhibits oxidation of 5hmC, thus preventing labeling with WT TDG. The effectiveness of this strategy is demonstrated in Figure 9c, showing that β GT-treated 5hmC yields no measurable product with the same treatment as above. In this way, the combination of TET treatments with and without β GT in independent treatments enables analysis of both 5mC and 5hmC.

Because TET oxidizes 5mC, 5hmC, and 5fC bases in DNA to 5caC, the treatment renders all cytosine variant considered here susceptible to WT TDG recognition and labeling. This produces a potential complication in comprehensive analysis of all four demethylation elements independently. In practical terms, given the abundance of 5mC and 5hmC over 5fC and 5caC, the protocols are likely to be used for different profiling goals. A simple differential comparison between protocols with and without TET could be used to assign labeled DNA to either the 5mC/5hmC grouping or the 5fC/5caC grouping before further analysis. However, a more precise assessment could also be achieved by incorporating into the 5mC and 5hmC protocols an additional pretreatment with WT TDG in which canonical dCTP is incorporated rather than biotinylated nucleotides. This would preclude labeling of 5fC and 5caC selectively in subsequent steps and ultimately enable assessment of all four cytosine demethylation pathway base elements (Figure 10).

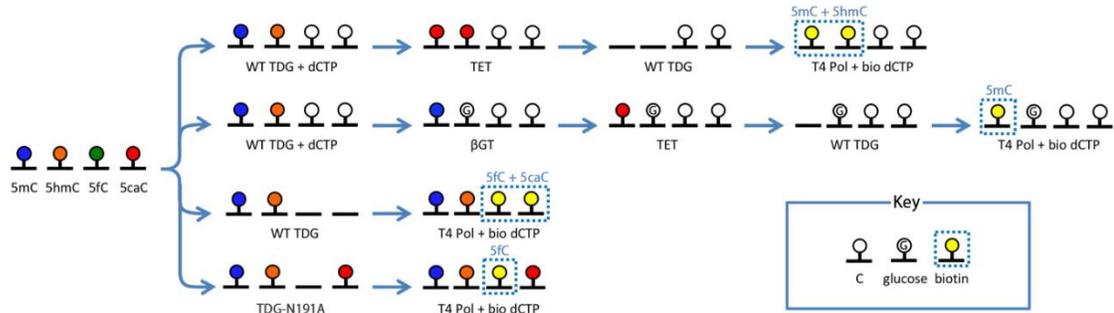


Figure 10. Labeling scheme for differentiating the four bases of the cytosine demethylation pathway with modular glycosylase labeling.

4.3 Discussion

We report a method for affinity labeling the four components of the cytosine demethylation pathway in DNA, comprising 5mC, 5hmC, 5fC, and 5caC. Our approach builds on a modular labeling strategy employing the enzymatic constituents of the BER⁷² in which (i) a glycosylase is used to excise a target base, (ii) an endonuclease is used to hydroxylate the 3' DNA end at the gap, and (iii) a polymerase is used to introduce a biotinylated base at the same position. Here, we exploit the recognition of TDG for some cytosine variants (5fC and 5caC) and enact a series of additional modifications to the general protocol to permit the assessment of all four independent modifications: first, a TDG mutant (TDG-N191A) is employed to differentiate 5fC from 5caC; second, TET enzymes are used to oxidize 5mC and 5hmC and enable their joint recognition by WT TDG; and third, β GT is used to preferentially block 5hmC recognition and distinguish it from 5mC. Consequently, information about each variant can be attained by performing pairwise comparisons across the four closely related protocols.

The incorporation of biotin tags enables the enrichment and isolation of DNA fragments containing the modification or modifications of interest in a manner similar to immunoprecipitation^{73,74}. Isolated products can subsequently be assessed by a broad range of analytical approaches including quantitative PCR or sequencing. In addition, the generalized method can also be applied easily to alternative labels like fluorophores or chemical linkers, provided that nucleotides synthesized to contain them are viable for polymerase incorporation. While modularity and diversity of base recognition are major advantages of our approach, another potential benefit is its directedness. In contrast to the widespread DNA damage induced by bisulfite exposure, the enzymatic activity employed is limited only to the base targets themselves. Thus, our methodology could enable improved analyses of small amounts of DNA, including those derived from inherently limited samples like liquid biopsies⁷⁵.

There are key challenges that remain with implementing our approach. Due to the base excision step in our process, we envision potential challenges with assessing symmetric modifications, i.e. modifications that are present on both strands of DNA. Critically, 5mC is often⁵⁵ (though not always⁷⁶) found in symmetric CpG dinucleotides. It is unclear how TDG will act on symmetric modifications that have been oxidized by TET, however there is a theoretical risk of generated breaks on both strands of DNA. One potential solution could be to purposefully employ lower amounts of TET or TDG to limit excision efficiency, but another possibility could also include performing a single cycle of amplification prior to processing, thereby forming hemimethylated target sites that would not be prone to breakage.

In conclusion, we have described adaptations to an enzymatic procedure for affinity labeling that can be used to tag the four base modifications involved in cytosine demethylation. Overall, our approach adds to the epigenetics analytical toolbox by providing modularity and extended target recognition, thereby progressing towards more comprehensive characterization of DNA modifications.

Peer-Reviewed Publication: F. Wang, O. K. Zahid, U. Ghanty, R. M. Kohli, A. R. Hall. “A modular approach for affinity-labeling the base elements of the cytosine demethylation pathway in DNA.” *Scientific Reports*, 2020.

CHAPTER 5 - SOLID-STATE NANOPORE ANALYSIS OF DIVERSE DNA BASE MODIFICATIONS

5.1 Introduction

SS-nanopores^{77,78} have been widely studied as a means to assess biological molecules like DNA^{79,80}, RNA^{81,82}, and proteins^{83–85} using the principle of resistive pulse sensing. The platform consists of an insulating thin-film membrane that contains a nanometer-scale pore, positioned in an electrolyte solution. Application of an electrical bias across the membrane generates an electric field through the pore, and as charged molecules are threaded electrophoretically one-by-one, they temporarily occlude the aperture and interrupt the measured ionic current. These brief electrical disruptions are designated as “events”, and their properties have been used to study molecular attributes⁸⁶, probe intermolecular interactions^{87,88}, and determine analyte concentration^{25,89}. Historically, a significant limitation of this measurement approach has been a lack of selectivity: all molecules of like-charge will translocate and contribute to the overall signal, thus requiring differentiation *ex post facto* via often subtle differences in event characteristics. We have developed a SS-nanopore assay that enables nearly binary detection and quantification of DNA featuring a single biotin affinity tag⁹⁰. Briefly, when target DNA fragments (below ~250 bp) or a key chaperone protein (monovalent streptavidin⁹¹, MS) are introduced individually to a SS-nanopore of appropriate diameter, their rapid translocations prevent events from being resolved by conventional electronics (Figure 11 a-b). However, when the two molecules bind, the larger nucleoprotein complex interacts with the walls of the nanopore during passage, slowing its translocation to a resolvable speed and yielding events (Figure 11c).

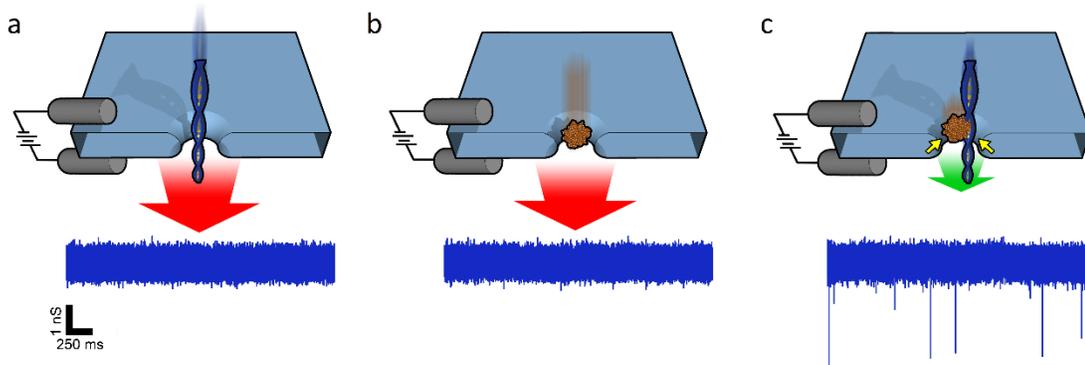


Figure 11. Depiction of the selective SS-nanopore assay.

Individual passage of (a) a short DNA or (b) a chaperone protein (MS) yields no events due to the high translocation speed (red arrows); (c) a DNA-protein complex interacts with the pore walls (yellow arrows), resulting in slower translocation speed (green arrow) and resolvable events. Sample conductance traces at bottom were measured at 300 mV using 75 bp DNA at 500 nM with a synthetic biotin.

Recently, we expanded this basic approach to assess hydroxymethylcytosine epigenetic modifications⁸⁹ by employing an established method for specific biotin labeling of the base³⁶, enabling direct assessment of a base modification with physiological relevance. However, the scope of possible targets for the labeling approach was limited intrinsically by enzymatic recognition. Here, we enhance our SS-nanopore measurement scheme significantly by integrating it with an alternative, modular labeling technique that enables the targeted detection of diverse base modifications.

5.2 Nanopore analysis of uracil, 8-oxoguanine, T:G mismatches, and ethenoadenine with glycosylase specificity testing

SS-nanopore analyses of the uracil-containing and oxoG-containing oligos labeled in Chapter 3 demonstrated clear specificity in the resulting electrical signal as well. For the appropriate combinations of base modification and enzymes, we observed exponential voltage-dependent event rates (Figure 12a,d), characteristic of the assay^{89,92}. Provided with the same total DNA concentrations (250 nM), the nearly identical event rate trends for both cases further indicated not only the similarity of the yields for the two labeling protocols, but also the reproducibility of the assay. In contrast, mismatched components yielded negligible event rates that were indistinguishable from negative controls across the entire investigated range of applied voltage (Figure 12b,c). These results suggested that non-specific labeling of DNA was insignificant and confirmed intrinsic discrimination for an intended base element.

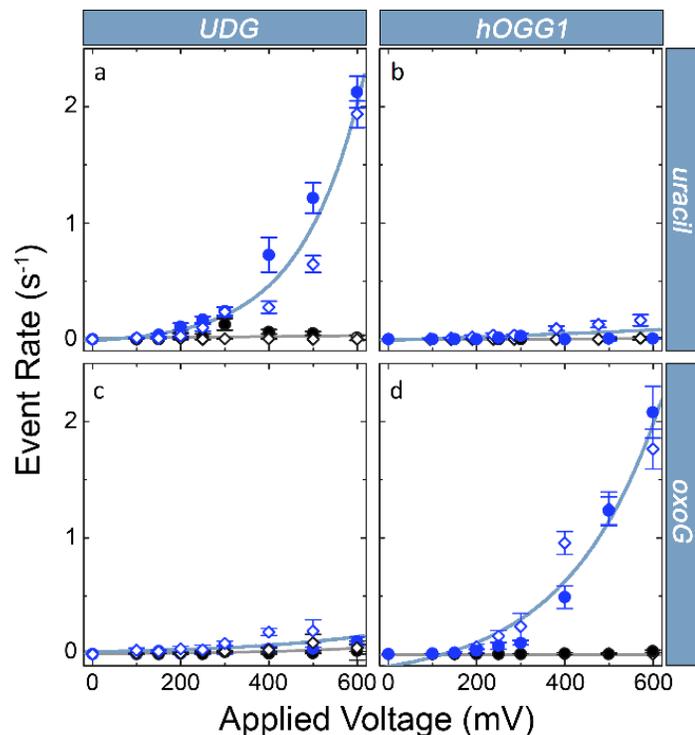


Figure 12. SS-nanopore analyses of DNA oligonucleotides featuring either a single uracil or a single oxoG.

Data points indicate measurements on treated DNA at 250 nM with (blue) and without (black) MS. Filled circles and open diamonds are independent measurements on different SS-nanopore devices and all lines are exponential fits to the data. Dramatic increases in event rate are measured for DNA-MS when a glycosylase specific for the target base is used (blue data, upper left and lower right). Almost no effect is observed for mismatched glycosylase (blue data, upper right and lower left).

Additional nanopore analysis of the labeled T:G-containing oligos also demonstrated selective detection by our SS-nanopore assay (Figure 13a). Indeed, we recovered the same exponential trend in measured SS-nanopore event rate and the same selectivity over a negative control as found for uracil and oxoG. The event rate dependence was slightly

higher for TDG labeling than for previous examples, which could be due to minor residual enzyme binding or small differences in pore attributes (diameter, shape, etc.). Similar trends were observed in the nanopore analysis of the labeled 1,N⁶-ethenoadenine-containing oligos. However, this modification did present a lower maximum rate at 600 mV, which we attributed to the smaller length⁸⁹ of this DNA as compared to the other constructs described in this report. However, the selective rate difference is easily resolved, demonstrating the broad modularity of both the labeling scheme and the measurement approach.

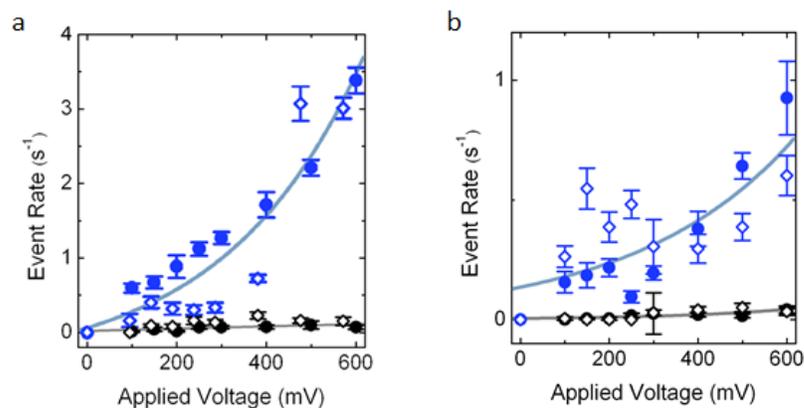


Figure 13. SS-nanopore analyses of DNA oligonucleotides featuring either a single T:G mismatch or a single 1,N⁶-ethenoadenine.

(a) T:G mismatch; (b) 1,N⁶-ethenoadenine. Data points indicate measurements on treated DNA at 250 nM with (blue) and without (black) MS. Filled circles and open diamonds are independent measurements on different SS-nanopore devices and all lines are exponential fits to the data.

In some preparations of the labeled 1,N⁶-ethenoadenine oligos (Figure 13b), we observed a reproducible artifact for this material wherein an anomalously high event rate was

measured at low voltages (<400 mV) that returned to the expected exponential relationship (dashed line) at higher voltages (Figure 14), which we suggest may have been due to structural irregularities associated with the modified base itself. This contributes to a greater variability in the data, and one possible explanation for this behavior may be in the structure of 1,N⁶-ethenoadenine, which is known to cause kinking in DNA⁹³ and may potentially alter interactions with the nanopore during translocation. While capping DNA reduced the apparent emergence of end labeling (Figure 6), gel analysis could not confirm the location of the single biotin tag. It is possible that some population of monobiotinylated DNA could be end labeled while maintaining the 1,N⁶-ethenoadenine. As a result, a kinked structure may be maintained. We hypothesize that the high event rate at low voltages could be due to orientation-specific interactions of the kinked DNA-MS construct threading through the pore, which are not present at high voltage due to a greater electrophoretic force. Future measurements will focus on elucidating the basis of this effect and optimizing procedures as necessary, both by considering the use of other sizes of pores and improving the efficacy of the labeling process.

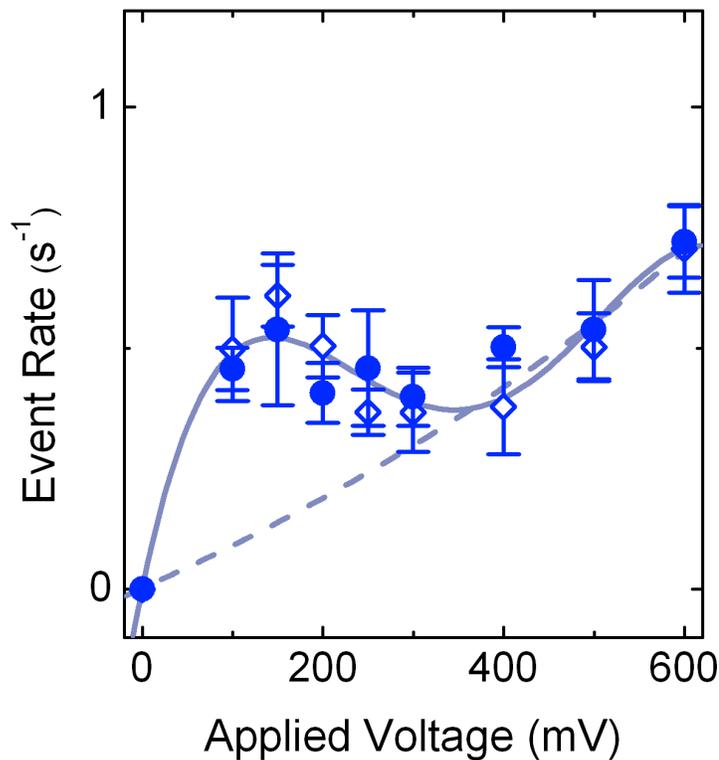


Figure 14. Additional SS-nanopore data for labeled 1,N⁶-ethenoadenine DNA.

Data points indicate measurements on treated DNA at 250 nM with MS. Filled circles and open diamonds are independent measurements on different SS-nanopore devices.

The solid line is a polynomial fit to the data to serve as a guide to the eye, and the dotted line is an exponential fit to the data.

The 250 nM concentration used in these experiments is sufficient as a proof of concept of the viability of this approach. In previous work with other DNA base modifications, we have demonstrated the ability to detect concentrations of labeled DNA as low as 10 nM⁸⁹, a more physiologically relevant value. We have also previously shown that a calibration curve can be prepared to calculate the concentration of the targeted material in solution from the event rate⁹², allowing for precise quantification of difficult-to-detect

modifications in DNA. Coupled with our previous reports showing assay viability among a background of non-target components^{90,92}, these data established a highly selective SS-nanopore technique for assessing physiologically-relevant base modifications in DNA.

5.3 Discussion

We have shown that a variety of single-base modifications can be assessed with a selective SS-nanopore assay in conjunction with our targeted affinity-labeling technique. This provides an additional analysis tool in the investigation of DNA base modifications due to the high sensitivity of nanopores that allows the rapid analysis of low-concentration samples and the flexibility provided by this labeling method. Notably, the presence of a nick in the DNA backbone, which would greatly hinder analysis methods such as sequencing and qPCR, does not appear to have a negative effect on the ability to gather nanopore data, as events are generated by the interaction of the bulky biotin-streptavidin interaction with the nanopore rather than through any intrinsic character of the DNA structure itself.

However, this process can still be refined further, as the 1,N⁶-ethenoadenine results illustrate. In comparing the labeling efficiency of this modification as presented in Chapter 3 and the nanopore analysis of the labeled material, it seems probable that the incomplete labeling of the bulky base modification may be a cause, at least in part, of the artifact observed in low-voltage nanopore measurements. Thus, by taking steps to improve the labeling efficiency of this modification, it may be possible to obtain nanopore results that are more comparable to those of the other modifications within this study.

This work demonstrates that this labeling method can be incorporated into our already existing nanopore system to expand the applications of this analysis method, with the potential for further expansions to recognize additional analytes by incorporating the use of other glycosylases that target other DNA base modifications. This provides a flexible and useful tool to identify and quantify the presence and prevalence of different modifications within a DNA sample and may help to better elucidate the profiles of different diseases.

Peer-Reviewed Publication: F. Wang, O. K. Zahid, B. Swain, D. Parsonage, T. Hollis, S. Harvey, F. Perrino, R. M. Kohli, E. W. Taylor, A. R. Hall. "Solid-state nanopore analysis of diverse DNA base modifications using a modular enzymatic labeling process." *Nano Letters*, 17 (11), 7110–7116, 2017.

CHAPTER 6 – OPTIMIZATION OF LIGATION PROTOCOLS FOR LABELED MATERIALS

6.1 Introduction

Our labeling strategy as we have demonstrated it thus far leaves a nick in the DNA backbone. This defect has no apparent negative effect on many applications including immunoisolation and single-molecule detection and quantification by solid-state nanopore⁶⁹, but would be disruptive to other important analytical techniques like quantitative PCR or sequencing. Therefore, we next develop a ligation step to repair the nick and restore the DNA structure.

There are two families of glycosylase: bifunctional and monofunctional⁴². Bifunctional glycosylases (i.e. those having AP lyase activity, like formamidopyrimidine-DNA glycosylase⁹⁴) leave the labeled DNA strand primed for phosphate ester linkage. However, monofunctional glycosylases like UDG result in a phosphate flap that renders the nick a poor substrate for ligation. In principle, the inclusion of an additional enzyme with independent AP lyase activity would remove the flap and enable subsequent ligation. Furthermore, glycosylases that display a high binding affinity for the AP site such as TDG may have additional difficulties in displacing the enzyme and repairing the DNA backbone. Due to the difference in the activities of these glycosylases, we have performed independent ligation experiments to develop protocols for the ligation of DNA containing modifications targeted by these different glycosylases.

6.2 Development of ligation protocols for bifunctional glycosylases (case study: Fpg)

To perform ligation with a bifunctional glycosylase, we chose to use formamidopyrimidine-DNA glycosylase (Fpg). Following our previously described labeling procedure with Fpg, the purified, labeled DNA was subsequently incubated with T4 DNA Ligase in appropriate buffer overnight. Analysis of the result on a DNA denaturing gel indicated a good yield of ligated construct (51%, Figure 15)

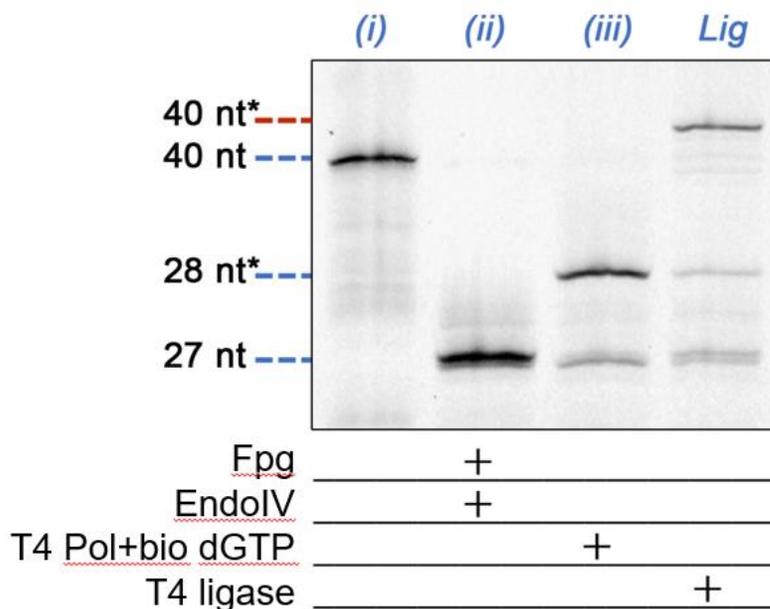


Figure 15. Denaturing gel analyses of 40 nt DNA constructs featuring a single 8-oxoguanine at base position 27 following ligation.

The base is excised with the bifunctional glycosylase formamidopyrimidine-DNA glycosylase (Fpg) and the construct is treated with EndoIV to prepare the 3' end of the gap, T4 polymerase and biotinylated dGTP to label, and T4 ligase to repair the remaining nick. Lane (i): annealed oligonucleotide; lane (ii): following glycosylase/endonuclease treatment; lane (iii): following polymerase fill-in with a

*biotinylated nucleotide; lane “Lig” is post ligation yielding a biotin-labeled construct with a repaired backbone (red). Construct lengths at left apply to both gels and * indicates DNA length plus biotin tag.*

6.3 Development of ligation protocols for monofunctional glycosylases (case study: UDG)

Subsequently, we attempted to perform ligation with the monofunctional glycosylase uracil DNA glycosylase (UDG). Earlier attempts to simply include a DNA ligase following our previously described labeling protocol demonstrated very low efficiencies of ligation, which we attributed to structural artifacts remaining after the labeling process, most significantly the presence of a small flap of the DNA backbone 3' to the gap following cleavage 5' to the AP site by EndoIV. To remove this flap, we subsequently incorporated the AP lyase endonuclease EndoVIII before proceeding to the nucleotide incorporation. As with the bifunctional glycosylase, this labeled material was then ligated by incubation with T4 DNA Ligase overnight. This protocol allows the nick to be ligated with good yield (50%, Figure 16), comparable to that of the bifunctional glycosylase Fpg.

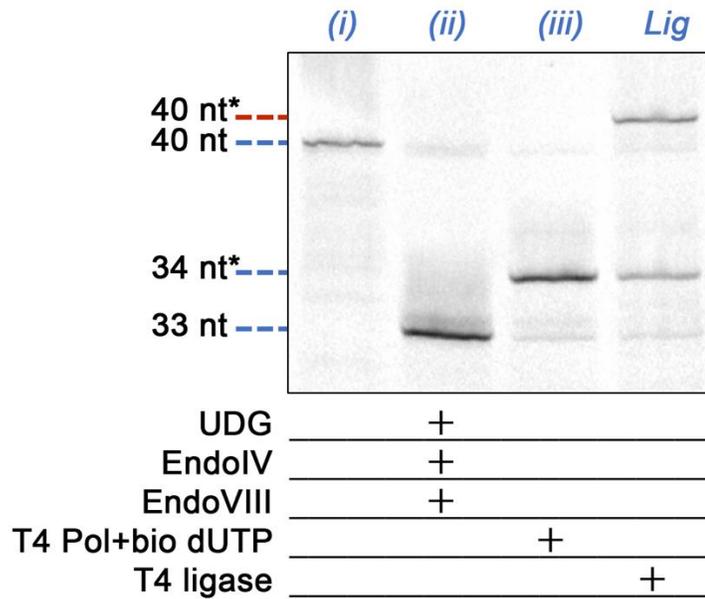


Figure 16. Denaturing gel analyses of 40 nt DNA constructs featuring a single uracil at base position 33 following ligation.

In each, the base is excised with the monofunctional glycosylase uracil DNA glycosylase (UDG) and the construct is treated with EndoIV to prepare the 3' end of the gap, EndoVIII to remove the phosphate flap, T4 polymerase and biotinylated dUTP to label, and T4 ligase to repair the remaining nick. Lane (i): annealed oligonucleotide; lane (ii): following glycosylase/endonuclease treatment; lane (iii): following polymerase fill-in with a biotinylated nucleotide; lane "Lig" is post ligation yielding a biotin-labeled construct with a repaired backbone (red). Construct lengths at left apply to both gels and * indicates DNA length plus biotin tag.

6.4 Development of ligation protocols for glycosylases requiring enzymatic displacement (case study: TDG)

Due to the use of thymine DNA glycosylase (TDG) in several of our labeling experiments, we considered it vital to examine the ligation of substrates following treatment with TDG, as TDG has the characteristic of maintaining strong binding affinity to the AP site after base excision⁹⁵; this factor necessitated⁶⁹ the use of an active displacement element in our protocol in the form of AP endonuclease 1 (APE1), which we considered may have a deleterious effect on any ligation attempts. As we suspected, either the specific activity of TDG binding to the DNA or its forcible removal appears to induce damage to the proximal substrate because we find a very low yield (~12%) of ligated construct and observe additional bands using the same protocol as for UDG (Figure 17).

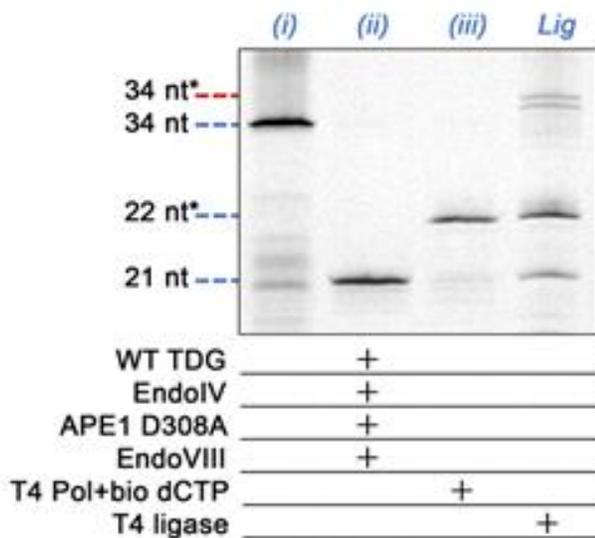


Figure 17. Denaturing gel analyses of 34 nt DNA constructs featuring a single 5caC at base position 22 following ligation.

In each, the base is excised with WT TDG and the construct is treated with EndoIV to prepare the 3' end of the gap, EndoVIII to remove the phosphate flap, T4 polymerase and

*biotinylated dCTP to label, and T4 ligase to repair the remaining nick. Lane (i): annealed oligonucleotide; lane (ii): following glycosylase, APE1 treatment, and endonuclease; lane (iii): following polymerase fill-in with a biotinylated nucleotide; lane “Lig” is post ligation yielding a biotin-labeled construct with a repaired backbone (red). Construct lengths at left apply to both gels and * indicates DNA length plus biotin tag.*

Increasing the EndoVIII concentration by up to 50% does not improve this yield, possibly indicating that the low efficiency may be a consequence of damage induced to the DNA itself rather than incomplete removal of the DNA flap. We note that while the precise nature of the damage is unclear, the observation that efficient base incorporation is achieved at the available 3' end in the gap with T4 polymerase suggests that it is localized predominantly at the flap or at the base directly after the AP site. This could be related in part to the unusual binding conformation of TDG to DNA⁹⁶.

To address this challenge, we finally investigate an alternative mechanism for TDG release intended to improve ligation yield by avoiding structural complications known to accompany APE1, including extensive DNA kinking⁴⁶. For this, we use a phenol incubation following base excision by TDG. The low polarity of phenol makes it capable of inducing conformational changes in proteins exposed to the solvent⁹⁷, driving hydrophilic residues into a more interior position while drawing hydrophobic residues to the surface in an inversion of the aqueous conformation. As such, we hypothesize that treatment of the TDG-bound DNA with phenol would result in release of the DNA with reduced substrate damage and sequestration of the TDG in the organic layer. To validate

this, following TDG incubation, we introduce to the bound DNA a phenol solution to a final concentration of 25% (v/v). We then decant the aqueous layer to recover the released DNA, purify it via column purification, and continue labeling and ligation as with UDG. The results of this procedure demonstrate a significant improvement over the use of APE1 for TDG removal (Figure 18), achieving a yield of ~26%.

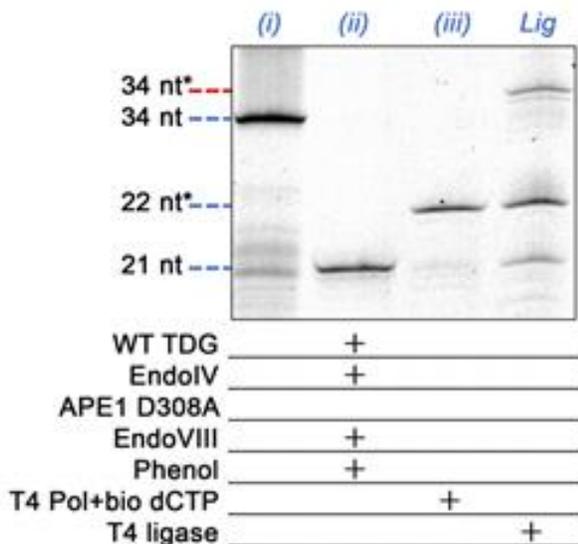


Figure 18. Denaturing gel analyses of 34 nt DNA constructs featuring a single 5caC at base position 22 following ligation.

In each, the base is excised with WT TDG and the construct is treated with EndoIV to prepare the 3' end of the gap, EndoVIII to remove the phosphate flap, T4 polymerase and biotinylated dCTP to label, and T4 ligase to repair the remaining nick. Lane (i): annealed oligonucleotide; lane (ii): following glycosylase, phenol treatment, and endonuclease; lane (iii): following polymerase fill-in with a biotinylated nucleotide; lane "Lig" is post ligation yielding a biotin-labeled construct with a repaired backbone (red). Construct lengths at left apply to both gels and * indicates DNA length plus biotin tag.

While this approach is not as effective as the protocol for glycosylases that do not demonstrate high binding affinity to AP sites, additional improvements may be instituted in the future to realize higher yields.

6.5 Discussion

Through these experiments, we have demonstrated some success in ligating the nicked DNA products produced through our labeling processes. However, there is still room for improvement with regard to improving the efficiency of the ligation. One way in which this may be achieved is by further improving the efficiency of the labeling process to ensure the incorporation of the labeled nucleotide into more of the molecules; the absence of this replacement for the excised modification would leave a 1 nt gap at the site of the excision, assuming that the cleanup of the DNA backbone 3' to the gap was entirely efficient, which would understandably hinder ligation attempts. Another method of increasing the ligation efficiency would likely be the aforementioned cleanup of the DNA backbone, particularly for materials treated with TDG or another glycosylase that demonstrates high binding affinity to the AP site. We also note that in the experiments we conducted, the thermal stabilities of the short DNA strands remaining after the nick may limit the overall yields and that these could improve with long constructs or genomic DNA fragments.

However, although our ligation efficiencies are less than ideal, this process can still be used in the preparation of labeled DNA for techniques that require an un-nicked DNA backbone. Since a sample of DNA will usually contain multiple copies of fragments of interest, even a 26% efficiency of ligation, the lowest efficiency we were able to achieve in these

experiments, should not significantly affect the representation of any particular fragment or sequence. As such, this additional adaptation to our labeling process will expand the potential applications of this work to include technologies such as qPCR and sequencing.

Peer-Reviewed Publication: F. Wang, O. K. Zahid, U. Ghanty, R. M. Kohli, A. R. Hall.

“A modular approach for affinity-labeling the base elements of the cytosine demethylation pathway in DNA.” (Submitted)

CHAPTER 7 – SIMPLE AND EFFICIENT ROOM-TEMPERATURE RELEASE OF BIOTINYLATED NUCLEIC ACIDS FROM STREPTAVIDIN

7.1 Introduction

The strong non-covalent bond ($K=10^{-15}$ M) between biotin and streptavidin allows for robust biochemical labeling in competitive chemical environments, where the biotin moiety interacts with the highly electronegative binding site of streptavidin⁹⁸. Because of this stability, biotin capture has become a foundational technique for the biochemical selection of target molecules, particularly with nucleic acids where it can play a central role in diverse analytical approaches like single nucleotide polymorphism detection⁹⁹, DNA sequencing¹⁰⁰, and epigenetic labeling^{39,73,101}. However, a major disadvantage of the bond strength is that biotinylated molecules are often difficult to recover for downstream isolation, amplification, and other processes. Existing elution methods have significant challenges. For example, modified forms^{102–104} of biotin and streptavidin have been developed that enable improved release kinetics or enzymatic cleavage. However, the enhanced elution is achieved by reducing binding strength, and cleavage of the biotin tag precludes the possibility of iterative use, both of which diminish the advantages of the system. Biotin-streptavidin can also be disrupted chemically, but with significant drawbacks. For example, ammonium hydroxide at elevated temperatures has been reported¹⁰⁵ to yield up to 96% recovery of bound nucleic acids, but the chemical also damages the nucleic acids themselves. Alternatively, Holmberg *et al.*¹⁰⁶ have reported that the biotin-streptavidin bond can be dissociated with over 95% yield in pure water at temperatures above 70°C. While dissociation in water is an improvement in terms of chemical biocompatibility, the practical yield can vary significantly due to the presence of

buffer salts¹⁰⁶. In addition, high temperatures are not compatible with all constructs, particularly small nucleic acids with low melting temperatures or DNA nanostructures composed of short staple strands¹⁰⁷.

Here, an alternative approach to disrupting the biotin-streptavidin interaction is demonstrated that is highly compatible with nucleic acids and operates at room temperature. The low-polarity solvent phenol is used to reconfigure the structure of streptavidin and release bound nucleic acids. By creating a bi-phasic system between a binding buffer and a phenol-chloroform mixture¹⁰⁸, as is done widely in conventional biomolecular preparations^{109–111}, synthetic biotinylated DNA constructs can be recovered from both free streptavidin and streptavidin-coated beads with reproducible yields approaching 100%. While the process may render the streptavidin non-functional, the DNA-conjugated biotin tag remains intact and active for downstream uses. As an application for this release approach, the technique is incorporated into an isolation strategy to enable target selection for solid-state nanopore analysis.

7.2 Release of biotinylated nucleic acid bound to free monovalent streptavidin

The phenol dissociation technique was first demonstrated for solution-phase nucleoprotein complexes of nucleic acids with free monovalent streptavidin (MS). MS is a variant of tetravalent wild-type streptavidin in which only a single binding region retains its high affinity for biotin²⁷. The nucleic acid portion of the nucleoprotein complex is a model 150 bp double-strand (ds-) DNA featuring a single biotin tag. The low polarity of phenol relative to water is known to induce conformational changes in proteins^{97,112} that result in

less polar residues on the surface and increased solubility in non-polar solvent; nucleic acids remain soluble only in the aqueous phase¹¹³. The organic nature of phenol creates a biphasic mixture with aqueous solutions, and so in conventional phenol extraction¹⁰⁸, proteins and nucleic acids are effectively partitioned into separate phases. In the case of a nucleoprotein complex, the dual nature of the construct could cause it to be sequestered to the liquid-liquid interface. However, the biotin-streptavidin interaction is critically dependent on both a sterically-defined binding pocket and the highly polar residues within it¹¹⁴, as is clear from the free energy surface¹¹⁵ of the wild-type (WT) streptavidin-biotin complex displayed¹¹⁶ in Figure 19a. As a consequence, the protein structural changes induced by phenol exposure should both significantly disrupt the bond and segregate the constituents.

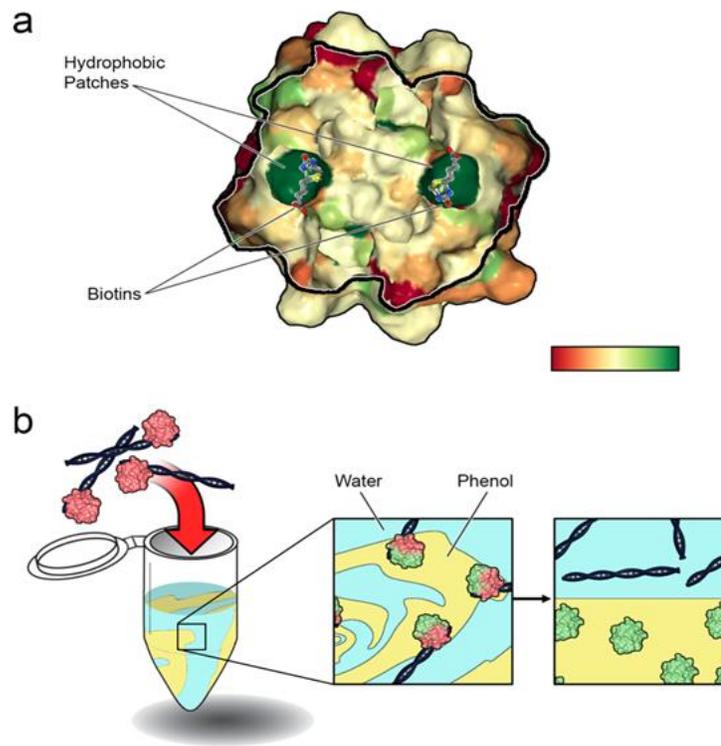


Figure 19. Schematic of proposed biotin-streptavidin dissociation method

(a) Cut-away diagram of the crystal structure of WT streptavidin (Protein Database structure 1MK5) showing that its biotin-binding regions are highly hydrophobic in their active state. Scale goes from hydrophilic (red) to hydrophobic (green). (b) Liquid-phase elution. Biotinylated DNA-MS complexes (in water) are added to a combination of phenol/chloroform/isoamyl alcohol (yellow) with water (blue) and mixed. As proteins are exposed to phenol (center), they dissociate from the DNA and then segregate into the organic when the phases are allowed to separate (right).

Separation of the biotin-streptavidin bond was initially demonstrated by performing conventional phenol extraction (Figure 19b) using a 25:24:1 mixture of phenol:chloroform:isoamyl alcohol on 150 bp biotinylated dsDNA-MS complex.

Chloroform is an organic solvent in which phenol is soluble but is less effective in rearranging protein conformation independently; its central role in the extraction process is to increase the density of the non-polar solution to aid phase separation. Likewise, isoamyl alcohol is an additive to prevent foaming that does not otherwise impact the protocol.

An electromobility shift assay (EMSA) with the synthesized 150 bp construct confirmed its capacity to bind MS with high affinity, yielding a clear shift to a higher molecular weight upon incubation with MS (Figure 20c, lanes 1-2). The lack of higher-order bands after complex formation is indicative of the 1:1 binding symmetry with MS. The unshifted band remaining at the 150 bp position (typically <5% of total lane intensity) was minor and likely due to incomplete biotin incorporation. Analyzing the extracted aqueous phase material (Figure 20c, lane 3), a single band was observed at the unbound 150 bp position. The absence of a significant population remaining at higher molecular weight indicates that MS was removed from the dsDNA with high efficiency. Incubation of the recovered DNA with a fresh aliquot of MS (Figure 20c, lane 4) yielded a single, high molecular weight band indistinguishable from the initial nucleoprotein complex, demonstrating that biotin is not modified irreversibly by the procedure and there is no measurable effect on MS binding. A control measurement of the nucleoprotein complex treated identically as above but with 100% chloroform in the incubation (Figure 20c, lane 5) confirmed that phenol is critical for separation while incubation in pure water at room temperature resulted in only a modest release of DNA (~20%, Figure 20c, lane 6), consistent with past reports¹⁰⁶.

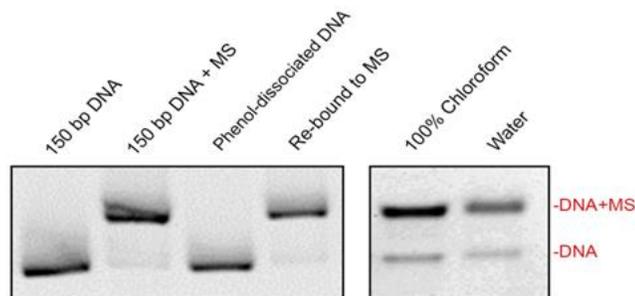


Figure 20. Gel analysis of liquid-phase elution

In the first panel, lane 1 is the 150 bp DNA construct alone, lane 2 is the construct bound to MS, lane 3 is the recovered DNA following elution, and lane 4 is the recovered DNA bound to fresh MS. In the second panel, the left lane shows elution with pure chloroform (no phenol) and the right lane shows elution with pure water at room temperature.

The low but non-zero solubility of phenol in water can result in trace quantities remaining in the aqueous phase following extraction. Since phenol contamination can have deleterious effects on proteins in downstream processes, avoiding remnant phenol is critical. While subsequent washing steps can be used to remove most of the residual chemical, reducing total phenol in solution may be an important preventative measure. Moreover, the toxic, corrosive, and flammable nature of phenol make limiting it in experimental protocols valuable. Therefore, the minimum phenol concentration required for efficient dissociation was determined next. Employing the same construct as above (biotinylated 150 bp dsDNA bound to MS), phenol was titrated against chloroform in the separation protocol and the recovered aqueous phase analyzed on gel (Figure 21, top). Qualitatively, there was a transition from ~10% release of biotinylated dsDNA at 0% (v/v) phenol to total release at 25% (v/v). By plotting the intensity ratios of shifted to non-shifted bands (Figure 21, bottom), a sigmoidal relationship was observed from which it was

determined that DNA constructs are fully released from the complex with as little as 12% (v/v) phenol.

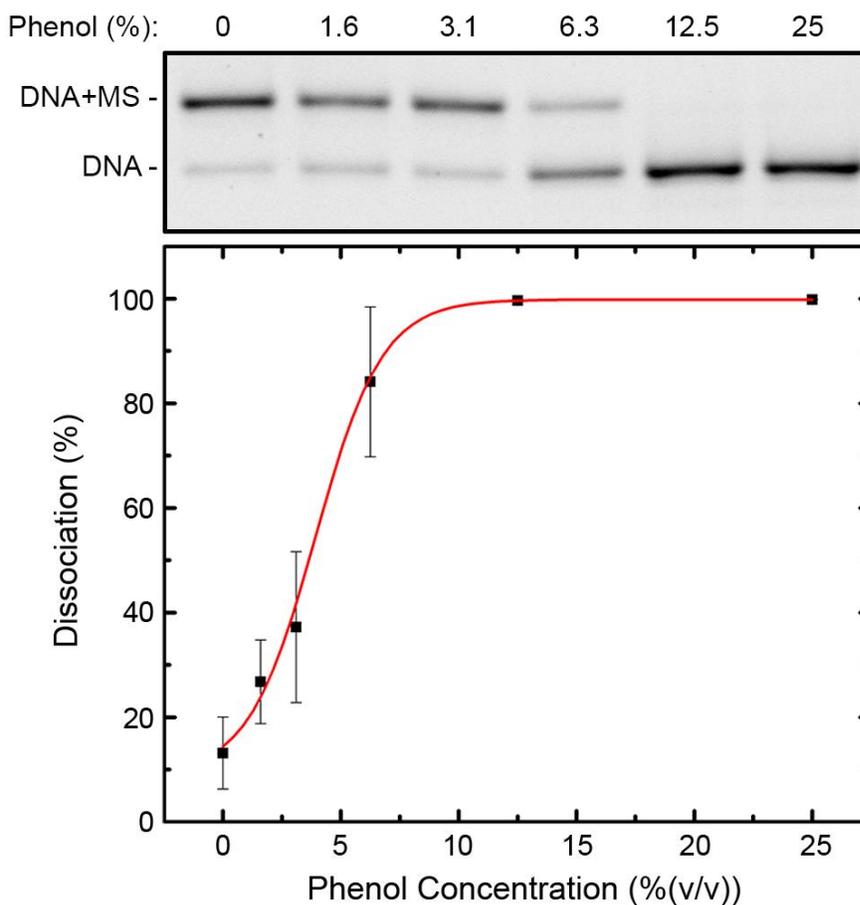


Figure 21. Effect of phenol concentration on biotinylated DNA release

Top: Gel analysis of 150 bp DNA-MS complex treated with a phenol titration. Bottom: DNA dissociation as determined from intensity of the DNA band relative to that of the nucleoprotein complex in each lane across three replicate experiments. Red line is a sigmoidal fit to the data from which it is determined that complete dissociation occurs at 12% (v/v) phenol. Each data point represents $n=3$. Note that certainty in the apparent 100% dissociation data points is limited by sensitivity of the imaging system.

7.3 Release of biotinylated nucleic acid bound to streptavidin-coated magnetic beads

Because streptavidin-coated surfaces are often used for isolation of biotinylated species, the use of phenol to release biotinylated dsDNA from streptavidin-coated magnetic beads was also investigated. For quantitation, samples were incubated with sufficient beads such that all biotinylated dsDNA in solution could be captured. Following elution with phenol above the minimal concentration for dissociation (12.5% v/v in chloroform), the aqueous phase was loaded onto a gel and imaged. From band intensity analysis of the initial and recovered DNA (Figure 22, lanes 1-2), ~90% of bound material was recovered; the minor losses are theorized to have resulted from non-specific binding with the beads and plasticware. Consistent with results using free MS, eluted biotinylated dsDNA remained active and could be bound to fresh streptavidin beads successively (Figure 22, lane 3). Since the beads themselves are capable of being isolated magnetically from the organic phase, this approach also provided the opportunity to investigate the potential for their reuse as well. Unfortunately, biotinylated dsDNA incubated with used streptavidin beads resulted in very poor (<3%) yield using either recovered (Figure 22, lane 4) or fresh DNA (Figure 22, lane 5). Given that the attachment of streptavidin to the beads is covalent in nature, it is unlikely that phenol-induced conformational changes could cause protein loss. These observations therefore suggest that streptavidin is altered irreversibly to some degree by the phenol, permanently adopting a non-functional conformation even after reintroduction into an aqueous environment. Consequently, while it would be possible in principle to extend the applications of this procedure to retrieve streptavidin and detached

biotinylated proteins from the organic phase following phenol extraction¹¹⁷, it is unclear whether their activity would be recovered.

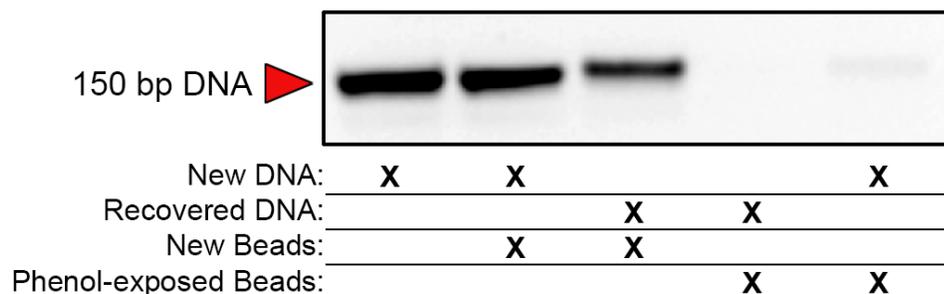


Figure 22. Gel analysis of 150 bp biotinylated DNA elution from streptavidin-conjugated magnetic beads

Lanes from L-R show: unprocessed DNA; fresh DNA bound to and eluted from fresh beads; recovered DNA bound to and eluted from fresh beads; recovered DNA bound to and eluted from used beads; and fresh DNA bound to and eluted from used beads. All elutions performed with 12.5% (v/v) phenol at room temperature.

7.4 Isolation of biotinylated DNA from a mixture with non-biotinylated DNA and verification by solid-state nanopore analysis

Biotin-streptavidin linkages are used broadly in diverse single-molecule assays for anchoring nucleic acids to substrates¹¹⁸⁻¹²⁰, and solid-state (SS-) nanopores^{78,121} comprise one technique for which the interaction has been particularly beneficial in recent years¹²²⁻¹²⁴. In the fundamental platform, translocations of charged molecules through a nanometer-scale pore in a thin film membrane are probed electrically through resistive pulse sensing¹²⁵, producing a signal that is critically dependent on molecular structure^{126,127}.

However, because each translocating biomolecule typically yields a signal, a remaining challenge in the field has been the general absence of intrinsic discrimination by signal generation. Conventionally, any differentiation has been accomplished through signal analysis. However, while significantly different molecules can typically be resolved¹²⁸⁻¹³⁰, those with size or structure similarity can be difficult to distinguish. One selective SS-nanopore assay has demonstrated¹³¹ that it can be applied to the detection and quantification of diverse molecular biomarkers like DNA base modifications^{69,89} and sequence motifs¹³², but it is viable for short (<250 bp) DNA fragments only and so additional capabilities are needed.

A potential solution to this challenge is the isolation or enrichment of target molecules prior to measurement for which phenol elution can be an enabling factor. Therefore, the optimized protocol was finally used to add selectivity to SS-nanopore analysis. As a model, λ -phage DNA was digested by restriction enzyme into two fragments approximately 33.5 and 15 kbp in length, respectively, with the latter containing a single biotin moiety at its end.

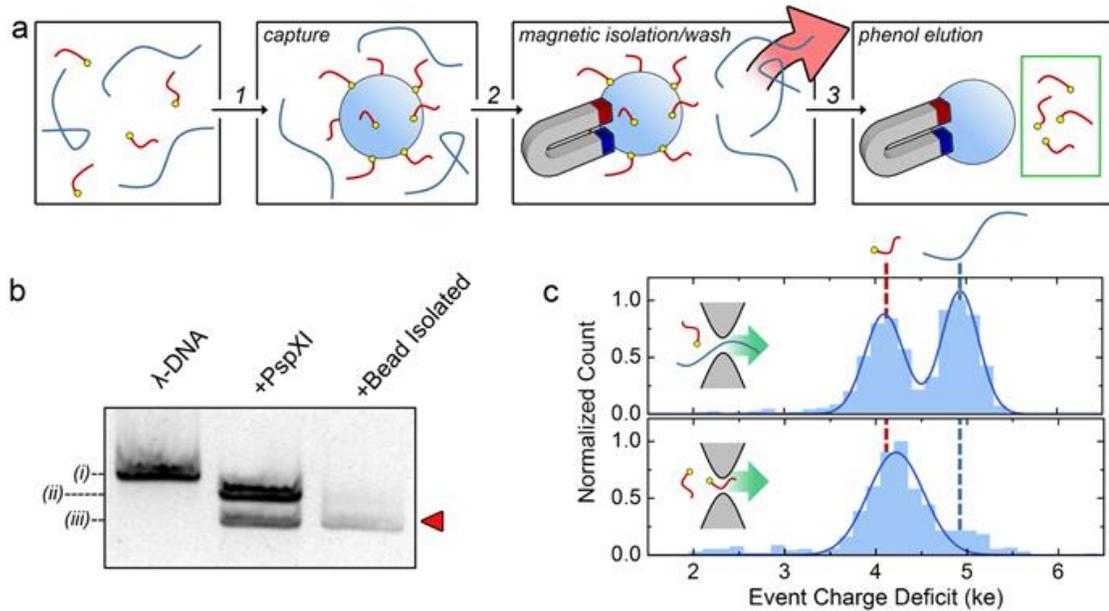


Figure 23. Isolation and analysis of biotinylated DNA from a mixture with non-biotinylated DNA

(a) Schematic showing isolation of biotinylated DNA. A mixture of biotinylated (red) and non-biotinylated (blue) DNA fragments are incubated with streptavidin magnetic beads (1). The beads with bound DNA are collected magnetically and non-biotinylated DNA is washed away (2). Biotinylated DNA is eluted with 12.5% (v/v) phenol (3). (b) Gel analysis showing: unprocessed 48.5 kbp (i) λ -phage DNA (lane 1); λ -phage DNA biotinylated at one end digested with PspXI to produce fragments approximately 33.5 (ii) and 15 kbp (iii) in length (lane 2); and the fragments after isolation and phenol elution from streptavidin beads (lane 3). The red arrow indicates the biotinylated ~15 kbp fragment. (c) Normalized SS-nanopore event histograms of ECD for the initial admixture (top, $n=1,158$) and the product of bead isolation (bottom, $n=519$). Lower ECD corresponds to lower molecular weight (i.e. the 15 kbp biotinylated DNA). Insets: schematics of SS-nanopore translocation.

Figure 23a shows the isolation procedure in which a mixture (1:1) of the two fragments was first added to streptavidin beads for capture of biotinylated fragments (1), then the beads were held magnetically while unbound (i.e. non-biotinylated) fragments were washed away (2), and finally the biotinylated fragments were eluted with phenol for collection and analysis (3). Gel analysis (Figure 23b) confirmed the selective capture of the 15 kbp fragment with this method. SS-nanopore translocations of both the original mixture and the isolate were then measured. We observed no additional contamination, noise (Figure 23c), or increased tendency towards clogging of the pore with the sample eluted with phenol. It is possible that trace amounts of phenol remain in the solution, but these did not affect nanopore performance in any measurable way. The event charge deficit (ECD), or the area defined by each translocation signal^{133,134}, was used as a primary metric because it gives an accurate assessment of fragment molecular weight (i.e. length). Indeed, an ECD histogram of the initial mixture of fragments (Figure 23d, top) showed a clear bimodal distribution that indicated the two distinct DNA sizes present. In contrast, the same analysis of the isolated and phenol dissociated material (Figure 23d, bottom) yielded a single population that matched the position of the smaller (15 kbp) fragment from the mixture, in agreement with the prepared biotin position. This result demonstrates that biotinylated nucleic acids can be isolated selectively prior to measurement, allowing for simple discrimination by SS-nanopore. While the fragments investigated here were considerably different in size for clarity, the approach is equally viable for molecules very close in size or structure that could be difficult to discriminate on subtle differences in signal alone.

7.5 Discussion

In conclusion, the biotin-streptavidin bond can be disrupted efficiently through incubation with phenol at room temperature. Isolation and recovery were demonstrated with biotinylated dsDNA constructs bound first to solution-phase MS and then to substrate-bound streptavidin in the form of conjugated magnetic beads. Extraction with phenol leaves incorporated biotin intact and capable of subsequent binding. Streptavidin was not reusable, although it is not clear from this study whether the protein was denatured in phenol, competitively bound to phenol, or deactivated by some other mechanism. Limited extraction was observed in pure water or pure chloroform at room temperature, showing that phenol is responsible for the high efficiency of extraction. A minimal 12% (v/v) phenol content was sufficient to fully dissociate the biotin-streptavidin bond.

Though affinity labeling and purification are commonly used techniques, the strong biotin-streptavidin bond often results in low yield, substrate damage, or other challenges that produce difficulty in performing sequential experiments on tagged molecules. As such, the capability to reliably and reversibly dissociate biotin-labeled DNA is expected to open up new possibilities in the downstream analysis of biotinylated constructs. The method reported here for breaking the biotin-streptavidin bond is unique for the absence of high temperatures and harsh chemicals while maintaining the integrity of the biotin and biotinylated construct. Consequently, the approach will be valuable not only to SS-nanopore measurements, but also in a broad range of diverse nucleic acid preparations and analyses.

Peer-Reviewed Publication: S. Bearden, F. Wang, A. R. Hall. "Simple and efficient room-temperature release of biotinylated nucleic acids from streptavidin and its application to selective molecular detection." *Analytical Chemistry*, 91 (13), 7996-8001, 2019.

CHAPTER 8 - ANALYSIS OF 8-OXOGUANINE IN RADIATION-SENSITIVE AND RADIATION-RESISTANT CELL LINES

8.1 Introduction

To synthesize our work to date and demonstrate the applicability of the ligation process, we subsequently used our labeling methodology to examine the formation and repair of 8-oxoguanine in radiation-sensitive and radiation-resistant head and neck cancer cell lines. From a parental radiation-sensitive SCC-61 cell line, the radiation-resistant rSCC-61 cell line was generated by exposing the SCC-61 cell line to 2 Grays (Gy) of ^{137}Cs radiation, culturing the surviving cells, and repeating the exposure and culture to a cumulative total of 16 Gy¹³⁵. In this way, cells that demonstrated a higher resistance to radiation exposure were selected for. Analysis of these two cell lines following this treatment has indicated a number of phenotypic differences between the parental SCC-61 and the resulting rSCC-61 cell line, including in their protein regulation, response to a small-molecule inhibitor of the epidermal growth factor receptor, and energy metabolism^{135,136}. Some of these phenotypic changes may be the result of alterations in DNA methylation and subsequent gene expression, with the rSCC-61 cells demonstrating altered ILK signaling, glucocorticoid receptor signaling, fatty acid oxidation, and cell cycle regulation¹³⁷. However, our interest lay in the immediate aftereffects of the exposure to radiation, which we believed would induce oxidative damage, including that of 8-oxoguanine, and whether the ability of the cells to repair this damage affected their viability.

8.2 Development of a protocol for labeling and quantifying oxidative damage in DNA extracted from cells with qPCR

As an analytical assessment to quantify the levels of oxidative damage generated by radiation exposure in the two cell lines, we chose to use qPCR due to its high sensitivity. Initial studies were performed with synthetic DNA oligos. In developing our protocol, we primarily adapted next-gen sequencing techniques through which adapters of known sequence could be ligated to the ends of the DNA¹³⁸. This would allow us to design primers that could be used regardless of the actual sequence of the DNA. In order to allow for accurate ligation of this adapter to the DNA, we sought to first end-label the target DNA with a single adenine overhang and then use designed adapters with a single thymine overhang and a stem-loop structure containing a single uracil (Figure 24). This design was intended to minimize the possibility of adapter dimers or multiple pieces of dsDNA ligating together, which could occur with blunt-ended DNA or adapters. Following successful ligation of the adapters, the loop can be cut through the excision of the uracil to result in a linear dsDNA construct.

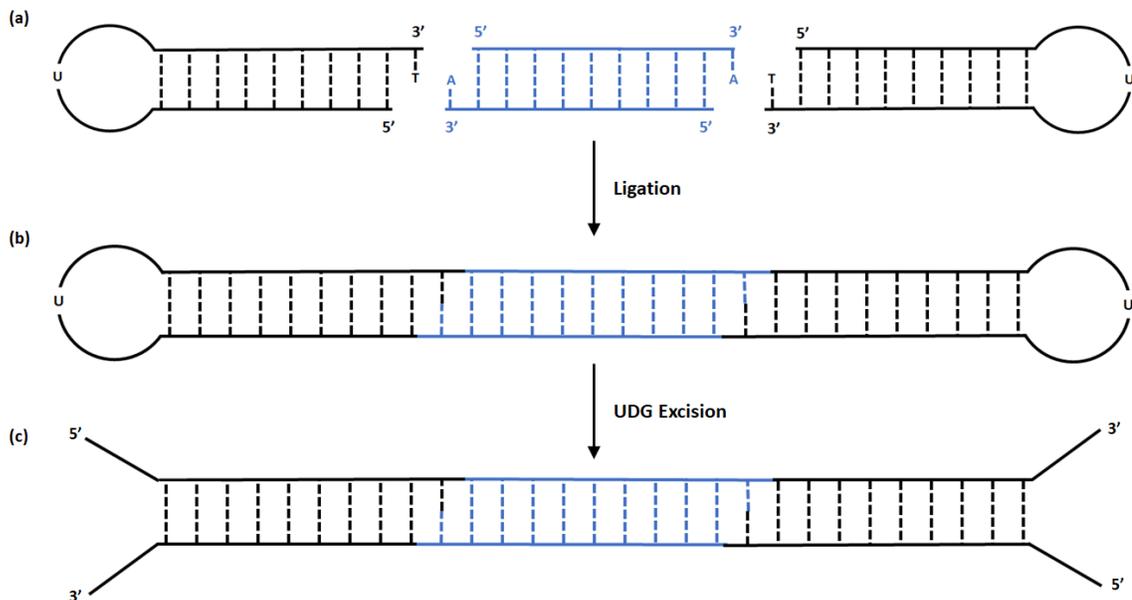


Figure 24. DNA hairpin adapter schematic

(a) DNA adapters designed to form a stem-loop structure, containing a 3'-thymine and a uracil in the loop (left, right) are aligned with DNA to which a 3'-adenine overhang has been added (center). (b) The adapters are ligated to the DNA. (c) Following ligation of the adapters, the uracil is excised.

This labeling approach was validated on synthetic DNA oligos containing a single 8-oxoG and analyzed by DNA denaturing gel electrophoresis (Figure 25, lanes 1-4). Due to the presence of a 5'-FAM fluorophore on one strand of the DNA oligo, an adapter was unable to be incorporated onto that end of the oligo, resulting in the characteristic labeling shifts associated with adapter attachment and ligation on the remaining end.

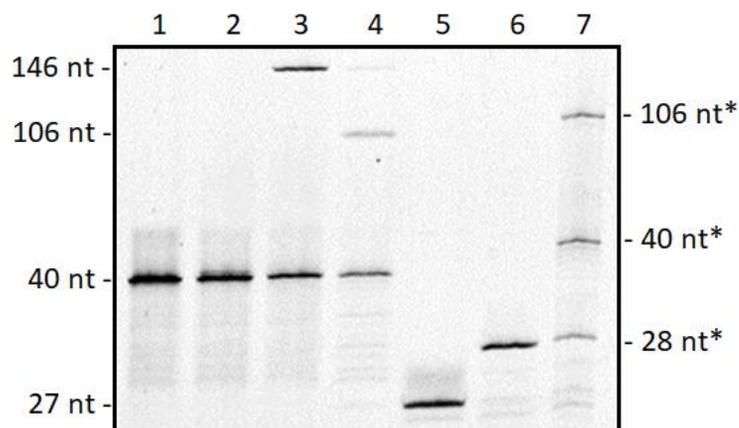


Figure 25. Denaturing gel analysis of adapter ligation

*Lane 1: annealed oligos; Lane 2: single A-tailed oligo; Lane 3: adapter ligation; Lane 4: adapter uracil excision; Lane 5: 8-oxoG excision; Lane 6: bio-dGTP incorporation; Lane 7: ligation. * indicates DNA length plus biotin tag.*

In our assessment, we estimated the yield for adapter attachment and ligation to be 37% (calculated from Figure 25, Lane 3). However, to fully establish the protocol, we also performed labeling of 8-oxoG in the product (Figure 25, lanes 5-7). Following the final ligation step, we observed a yield of 33% for adapter-ligated, labeled DNA. Since only one DNA end was available here, we therefore estimated that, when using a construct with two viable ends, a ~12% yield of fully labeled and adapter-ligated DNA could be expected. While this can be optimized further, it was sufficient for the analysis of genomic DNA.

Given the unwieldy length of genomic DNA, we then performed enzymatic fragmentation experiments to obtain fragments of a useable length. Our goal was to create DNA fragments that were as small as possible to allow for more accuracy of quantification but still large

enough to be adapter labeled and efficiently amplified by qPCR. For these reasons, we aimed for mean fragment lengths of approximately 150 bp.

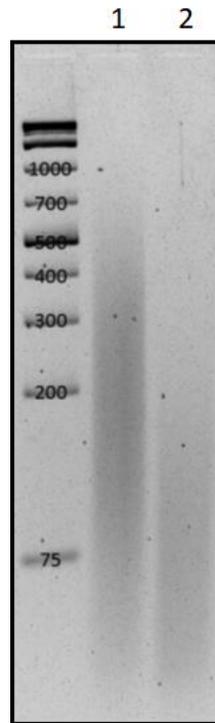


Figure 26. Agarose gel of enzymatic fragmentation

Enzymatic fragmentation of 2 µg genomic DNA after 35 minutes (Lane 1) and 1 hour (Lane 2) incubation with 4 µL of dsDNA fragmentase. Ladder is GeneRuler 1kb Plus DNA Ladder.

The results of fragmentation under two conditions are shown in Figure 26. Although the 1 hour fragmentation produced smaller fragment lengths (mean ~85 bp), it was also accompanied by a significant loss of material (~40%). Consequently, we chose to employ 35 min incubation which yielded a mean fragment length of ~170 bp. The enzymatically-fragmented DNA was subsequently blunt-ended by treating with DNA Polymerase I, Large (Klenow) Fragment and then processed with the adapter ligation protocol described above.

The ability of the resulting product to be amplified and measured by qPCR was confirmed by performing qPCR on a serial dilution of the material and comparing to a no-adapter control.

Finally, adapter-labeled samples were affinity-labeled for 8-oxoG through the insertion of a biotinylated-dGTP to enable the isolation of oxidized fragments and assessment of the fraction of oxidative damage within the two cell lines. Using the incorporated biotin tag, labeled fragments could be captured selectively with streptavidin-conjugated magnetic beads and then eluted with phenol-chloroform¹³⁹. Because trace amounts of organic solvents present following phenol-chloroform extraction were found to impact quantification, an additional ethanol precipitation step was employed and followed by resuspension in deionized water for increased purity.

8.3 Radiation dosage and time course study of oxidative damage

Following the development of our isolation and quantification protocol, we first investigated the effects of radiation dosage on SCC-61 and rSCC-61 cell lines. We hypothesized that the radiation-resistant rSCC-61 cells would either accumulate less oxidative damage through protective mechanisms or that they would repair the damage more quickly. We first tested dose response by plating SCC-61 and rSCC-61 cells and exposing each cell type to four conditions of single-dose radiation: 0 (control), 1, 2, and 4 Gy, respectively. 30 minutes after irradiation, surviving cells were harvested and their DNA was extracted with a commercial kit (Qiagen DNeasy Blood and Tissue Kit). Isolated DNA was subsequently fragmented, labeled for 8-oxoG, affinity-isolated, and quantified

by qPCR as described above. To determine the net amount of oxidative DNA damage in each sample, the quantity of affinity-isolated DNA (i.e. fragments containing 8-oxoG) was compared to the total amount of extracted DNA. The results of this measurement, corrected for material loss during the isolation protocol, are shown in Figure 27.

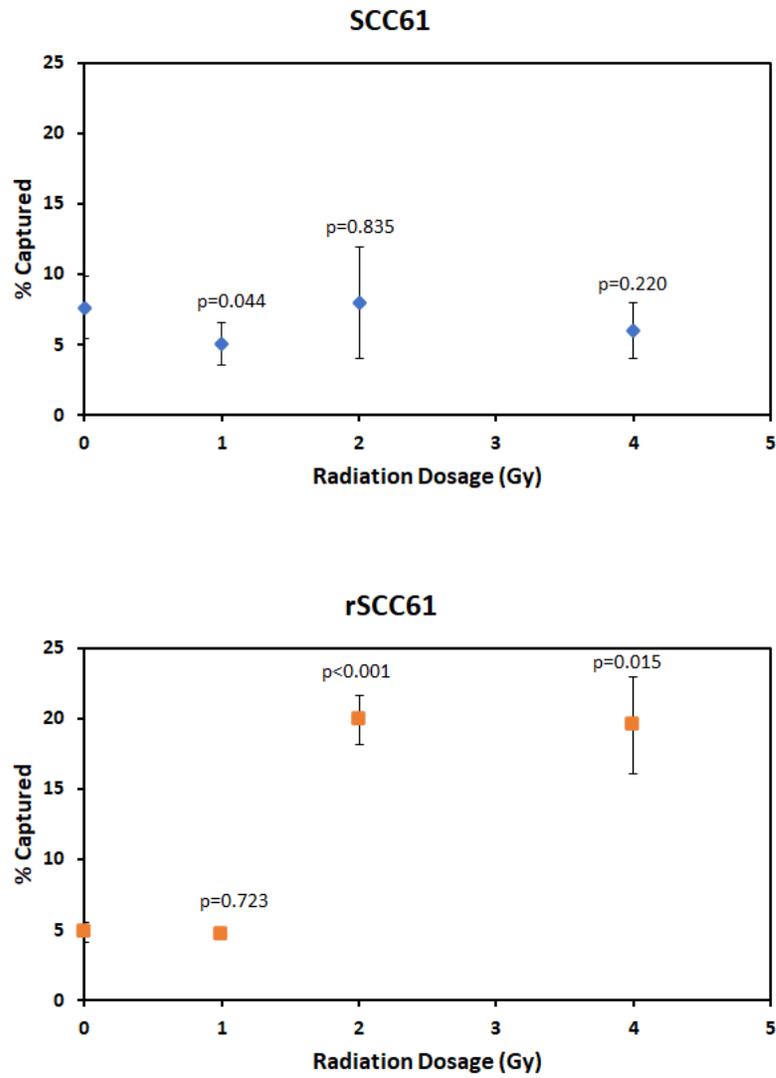


Figure 27. Effect of radiation dosage on 8-oxoG in SCC-61 and rSCC-61 cells 30 min after irradiation

Each data point represents $n=3$.

Interestingly, the surviving rSCC-61 cells showed accumulation of more oxidative damage than the surviving SCC-61 cells. This result suggested that rSCC-61 cells are capable of tolerating larger quantities of damage than SCC-61, an observation that coincides with the method by which rSCC-61 were originally identified¹³⁵. Due to our method of cell collection following irradiation, dead cells would lose their adherence to the plate and be rinsed away, preventing their contents from being quantified. Thus, if the SCC-61 cells with large oxidative damage accumulations died, those higher quantities of damage would not be included in the data.

Having investigated oxidative damage formation, we next studied its repair in each cell line by quantifying 8-oxoG at different timepoints following irradiation at a single dose. For this, we chose 2 Gy because *(i)* it is a physiologically relevant dose that is used in conventional radiotherapy for a variety of cancers^{140–142} and *(ii)* a significant difference in response was observed at this dose between the SCC-61 and rSCC-61 in our previous measurements above. For these measurements, plated SCC-61 and rSCC-61 cells were irradiated in matched pairs at 2 Gy and then placed back into the incubator. Pairs of plates (i.e. each cell line) were removed at 1, 2, 6 and 24 hrs post-irradiation and DNA was extracted and processed as above. Again, one pair of plates was not irradiated to serve as a baseline control group for comparison.

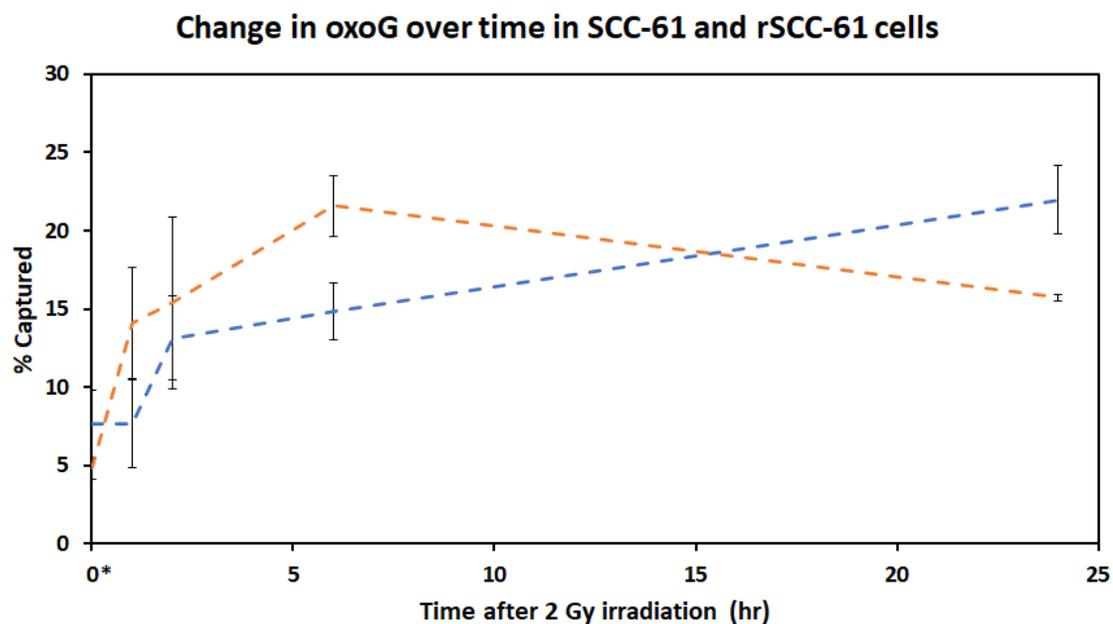


Figure 28. Change in net 8-oxoG content as a function of time for SCC-61 and rSCC-61 cells following 2 Gy irradiation

Blue data refers to SCC-61; orange data refers to rSCC-61. 0 = non-irradiated comparison samples. Each data point represents n=3.*

The results of the time course experiment are shown in Figure 28. For the first six hours, both of the cell lines demonstrated an overall increase in the presence of oxidative damage, likely due to increasing reactive oxygen species (ROS) in the media following irradiation. The rate of damage accumulation appeared greater in the rSCC-61 cells, potentially related again to their higher survivability in response to radiation damage. However, by 24 hrs, there was a marked change in the oxidative damage present in the cell lines. Significantly, the SCC-61 cells continued to accumulate damage while the rSCC-61 cells began to show a clear decrease in the amount of oxidative damage, with total content falling below that of the SCC-61 cells. This suggested a strong difference in the BER mechanisms within the

cells, indicating that rSCC-61 cells may have a higher capacity for efficient damage repair. We therefore anticipate that rSCC-61 cells may have higher expression of hOGG1, the glycosylase responsible for 8-oxoG repair in humans¹⁴³. While 8-oxoG levels did not return to pre-irradiation levels for the rSCC-61 cells within the time period being examined, the trend suggested that this would occur. We anticipate that the SCC-61 cells would also recover over a longer time course; given that the rSCC-61 cells are in fact a subset of the SCC-61 cells, irradiation is likely to simply select for the population of resistant cells within that group.

8.4 Discussion

Through direct measurement of 8-oxoG content in DNA extracted from cell culture, we have successfully demonstrated the application of our labeling protocol to the analysis of genomic samples via qPCR. By incorporating the usage of streptavidin-coated magnetic beads to capture selective fragments and phenol to elute them, we effectively isolated DNA containing 8-oxoG from a background of other DNA fragments. Although more optimization could improve the efficiency of the adapter ligation process and thus reduce the amount of required input, our yields were sufficient to perform the experiments and analysis.

These experiments have also provided a preliminary investigation into the effects of radiation on the formation of oxidative damage in two head and neck cancer cell lines. We first observed that following irradiation, surviving radiation-resistant rSCC-61 cells appear to have a greater accumulation of 8-oxoG than SCC-61 cells. We interpreted this to suggest

that rSCC-61 cells are able to tolerate a higher level of oxidative damage and remain viable, whereas SCC-61 cells are more prone to die from the exposure and thus not contribute to quantified damage. Over a time course following irradiation, both rSCC-61 cells and SCC-61 cells displayed sharp increases in 8-oxoG content. However, SCC-61 cells were found to both accumulate the damage slower and continued to accumulate throughout the entire measurement time, whereas rSCC-61 cells showed clear signs of repair by 24 hrs. We interpreted this to indicate a significant difference in the efficiency of the BER processes in the rSCC-61 cells.

Decisive conclusions were challenging to draw from our results in part because of the confounding factor of cell viability and the loss of dead cell DNA for quantification. Thus, future experiments may include examining 8-oxoG content in living and dead cells collectively. In addition, time course experiments could be performed at multiple additional points, including beyond the 24-hour window examined in our experiments so far. Finally, coupled analyses of the cells at different time points would illuminate the involved processes further, including ROS content and expression of BER enzymes, either directly through ELISA analyses or by quantifying mRNA.

Our general approach may also serve to support other studies. For example, it could enable additional screening for cancer patients in combination with stereotactic body radiation therapy (SBRT). In SBRT, patient tumors are imaged and used to generate 3D models to allow for accurate targeting of the tumors with high doses of radiation (typically between 7.5-20 Gys per dose) while minimizing damage to normal tissues^{144,145}. We hypothesize

that these high doses of radiation would induce elevated oxidative damage profiles within the targeted tissues, which subsequently shed their DNA into the blood following apoptosis or necrosis, potentially allowing for the performance of liquid biopsies¹⁴⁶. In conjunction with our method of labeling modifications such as 8-oxoguanine with an affinity tag, this could facilitate the isolation and examination of cell-free tumor DNA from among a background of other cell-free DNA. For patients with inaccessible tumors (e.g. brain tumors), this could provide a new means of performing tumor genetic screening and personalized medicine that could not be achieved otherwise.

CHAPTER 9 – FUTURE DIRECTIONS

Thus far, we have been able to demonstrate the applicability of our labeling method with several different glycosylases that have different functionalities: monofunctional glycosylases like UDG, bifunctional glycosylases like Fpg and hOGG1, and monofunctional glycosylases with a high affinity for AP sites such as hAAG and TDG. Therefore, it is highly probable that the techniques we have developed can be used for the labeling of other DNA modifications by simply using an appropriate glycosylase that can target the modification of interest.

However, many glycosylases are capable of identifying multiple modifications, as we have demonstrated with the use of TDG to excise not only T:G mismatches but also 5-fC, 5-caC, and to a lesser extent, uracil. This may be cause for concern during attempts to analyze less frequently occurring modifications such as 5-fC and 5-caC, where off-target labeling of other modifications would result in false positives that could affect the accuracy of the analysis. We have also seen, though, that introducing mutations into glycosylases can alter their recognition of the targets; with TDG-N191A, the ability to excise 5-caC was lost, allowing for differentiation between 5-fC and 5-caC. Experiments to determine the ability of TDG-N191A to recognize and remove T:G mismatches and uracil have not been performed as of this time. As such, there may be ways to improve the specificity of other promiscuous glycosylases through mutation in order to obtain more reliable results. It may also be possible to introduce other mutations that decrease the affinity of glycosylases to AP sites after base excision, which we believe to be a major factor in the lesser labeling yield for those types of glycosylases compared to other glycosylases. As such, research

into altering the recognition of glycosylases to improve specificity and decrease AP site binding would be valuable in increasing the accuracy and efficiency of labeling for specific modifications.

Another step that can be taken is improving the efficiency of the labeling process itself. Although we were able to successfully apply the labeling process to nanopore analysis and qPCR, improving the efficiency of labeling by further optimizing reaction conditions would correlate to a smaller required input of DNA. Clearly, requiring a smaller sample size would be an immense benefit regardless of the work being done, especially in cases of patient samples, in which the amount of material that can be extracted can vary. Though many steps of the labeling process can be improved, the one that would provide the greatest benefit to explore first would be ligation. As ligation is used in several steps within the process and has relatively poor yields, improving its efficiency, particularly for analytical techniques such as qPCR that require an unbroken backbone, would greatly improve the applicability of this technique. For bifunctional glycosylases, this may be able to be achieved through optimization of the reaction conditions, as both sides of the AP site have been sufficiently primed for ligation. However, for monofunctional glycosylases, we believe that for an undetermined proportion of the excised material, the 5' end of the gap may retain a fragment of the original DNA backbone, which would result in steric hindrance during attempts at ligation.

These improvements could subsequently allow for the capture and analysis of small, physiologically relevant quantities of patient DNA from large samples of a mixed

background. As in the potential application of the labeling and capture methodologies to analyzing liquid biopsies from patients who receive SBRT or the comparable Gamma Knife radiosurgery^{147,148}, which primarily targets tumors in the brain, the amount of DNA that can be collected from a sample of the patient's blood is expected to be relatively low and combined with various other components found in blood including other cellular DNA that is shed by normal, non-cancerous cells. As such, improving the efficiency of this process will facilitate the ability to conduct these types of less invasive diagnostics and analyses.

But perhaps more importantly, this labeling methodology allows for the examination of many different DNA modifications whose effects on gene expression, disease development and progression, and other processes have not been able to be determined due to difficulty in analyzing them with traditional genetic analysis techniques. As such, this may hold the key to a greater understanding of how certain illnesses arise or whether they have particular genetic profiles, which could have implications in this methodology's use as part of a potential diagnostic tool in addition to simple analytics.

REFERENCES

- (1) Watson, J. D.; Crick, F. H. C. The Structure of Dna. *Cold Spring Harb Symp Quant Biol* **1953**, *18*, 123–131. <https://doi.org/10.1101/SQB.1953.018.01.020>.
- (2) Henderson, R.; Krude, T. *DNA: Changing Science and Society*; Cambridge University Press, 2004.
- (3) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227* (5258), 561–563. <https://doi.org/10.1038/227561a0>.
- (4) Darwin, C. *On the Origin of the Species and The Voyage of the Beagle*; Graphic Arts Books, 2012.
- (5) De Bont, R.; van Larebeke, N. Endogenous DNA Damage in Humans: A Review of Quantitative Data. *Mutagenesis* **2004**, *19* (3), 169–185. <https://doi.org/10.1093/mutage/geh025>.
- (6) Chatterjee, N.; Walker, G. C. Mechanisms of DNA Damage, Repair, and Mutagenesis. *Environmental and Molecular Mutagenesis* **2017**, *58* (5), 235–263. <https://doi.org/10.1002/em.22087>.
- (7) Rodriguez-Rocha, H.; Garcia-Garcia, A.; Panayiotidis, M. I.; Franco, R. DNA Damage and Autophagy. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **2011**, *711* (1), 158–166. <https://doi.org/10.1016/j.mrfmmm.2011.03.007>.
- (8) Gidron, Y.; Russ, K.; Tissarchondou, H.; Warner, J. The Relation between Psychological Factors and DNA-Damage: A Critical Review. *Biological Psychology* **2006**, *72* (3), 291–304. <https://doi.org/10.1016/j.biopsycho.2005.11.011>.

- (9) Fenech, M. F. Nutriomes and Personalised Nutrition for DNA Damage Prevention, Telomere Integrity Maintenance and Cancer Growth Control. In *Advances in Nutrition and Cancer*; Zappia, V., Panico, S., Russo, G. L., Budillon, A., Della Ragione, F., Eds.; Cancer Treatment and Research; Springer: Berlin, Heidelberg, 2014; pp 427–441. https://doi.org/10.1007/978-3-642-38007-5_24.
- (10) Morales-Ruiz, T.; Ortega-Galisteo, A. P.; Ponferrada-Marín, M. I.; Martínez-Macías, M. I.; Ariza, R. R.; Roldán-Arjona, T. DEMETER and REPRESSOR OF SILENCING 1 Encode 5-Methylcytosine DNA Glycosylases. *PNAS* **2006**, *103* (18), 6853–6858. <https://doi.org/10.1073/pnas.0601109103>.
- (11) Holliday, R.; Ho, T. Gene Silencing and Endogenous DNA Methylation in Mammalian Cells. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **1998**, *400* (1), 361–368. [https://doi.org/10.1016/S0027-5107\(98\)00034-7](https://doi.org/10.1016/S0027-5107(98)00034-7).
- (12) Raiber, E.-A.; Hardisty, R.; van Delft, P.; Balasubramanian, S. Mapping and Elucidating the Function of Modified Bases in DNA. *Nature Reviews Chemistry* **2017**, *1* (9), 1–13. <https://doi.org/10.1038/s41570-017-0069>.
- (13) Niessen, W. M. A. *Liquid Chromatography-Mass Spectrometry*; CRC Press, 2006.
- (14) Electrospray ionization for mass spectrometry of large biomolecules | Science <https://science.sciencemag.org/content/246/4926/64.abstract> (accessed Aug 25, 2020).
- (15) Pitt, J. J. Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *Clin Biochem Rev* **2009**, *30* (1), 19–34.

- (16) Schumacher, F.; Herrmann, K.; Florian, S.; Engst, W.; Glatt, H. Optimized Enzymatic Hydrolysis of DNA for LC–MS/MS Analyses of Adducts of 1-Methoxy-3-Indolylmethyl Glucosinolate and Methyleugenol. *Analytical Biochemistry* **2013**, *434* (1), 4–11. <https://doi.org/10.1016/j.ab.2012.11.001>.
- (17) Cadet, J.; Douki, T.; Ravanat, J. L. Artifacts Associated with the Measurement of Oxidized DNA Bases. *Environmental Health Perspectives* **1997**, *105* (10), 1034–1039. <https://doi.org/10.2307/3433836>.
- (18) Singh, R.; Farmer, P. B. Liquid Chromatography-Electrospray Ionization-Mass Spectrometry: The Future of DNA Adduct Detection. *Carcinogenesis* **2006**, *27* (2), 178–196. <https://doi.org/10.1093/carcin/bgi260>.
- (19) Morgan, M. T.; Bennett, M. T.; Drohat, A. C. Excision of 5-Halogenated Uracils by Human Thymine DNA Glycosylase: Robust Activity for DNA Contexts Other than CpG. *J Biol Chem* **2007**, *282* (38), 27578. <https://doi.org/10.1074/jbc.M704253200>.
- (20) Masuda, Y.; Bennett, R. A. O.; Demple, B. Rapid Dissociation of Human Apurinic Endonuclease (Ape1) from Incised DNA Induced by Magnesium. *J. Biol. Chem.* **1998**, *273* (46), 30360–30365. <https://doi.org/10.1074/jbc.273.46.30360>.
- (21) Liu, M. Y.; Torabifard, H.; Crawford, D. J.; DeNizio, J. E.; Cao, X.-J.; Garcia, B. A.; Cisneros, G. A.; Kohli, R. M. Mutations along a TET2 Active Site Scaffold Stall Oxidation at 5-Hydroxymethylcytosine. *Nat Chem Biol* **2017**, *13* (2), 181–187. <https://doi.org/10.1038/nchembio.2250>.
- (22) Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C. Divergent Mechanisms for Enzymatic Excision of 5-Formylcytosine and 5-Carboxylcytosine

from DNA. *J Am Chem Soc* **2013**, *135* (42), 15813–15822.

<https://doi.org/10.1021/ja406444x>.

- (23) Liu, M. Y.; DeNizio, J. E.; Kohli, R. M. Quantification of Oxidized 5-Methylcytosine Bases and TET Enzyme Activity. *Methods Enzymol* **2016**, *573*, 365–385. <https://doi.org/10.1016/bs.mie.2015.12.006>.
- (24) Collins TJ. ImageJ for Microscopy. *BioTechniques* **2007**, *43* (1), 25–30.
- (25) Wanunu, M.; Dadosh, T.; Ray, V.; Jin, J.; McReynolds, L.; Drndic, M. Rapid Electronic Detection of Probe-Specific MicroRNAs Using Thin Nanopore Sensors. *Nature Nanotechnology* **2010**, *5* (11), 807–814.
- (26) Yang, J.; Ferranti, D. C.; Stern, L. A.; Sanford, C. A.; Huang, J.; Ren, Z.; Qin, L.-C.; Hall, A. R. Rapid and Precise Scanning Helium Ion Microscope Milling of Solid-State Nanopores for Biomolecule Detection. *Nanotechnology* **2011**, *22* (28), 285310.
- (27) Fairhead, M.; Krndija, D.; Lowe, E. D.; Howarth, M. Plug-and-Play Pairing via Defined Divalent Streptavidins. *Journal of Molecular Biology* **2014**, *426* (1), 199–214. <https://doi.org/10.1016/j.jmb.2013.09.016>.
- (28) Jaenisch, R.; Bird, A. Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals. *Nat Genet* **2003**, *33*, 245–254. <https://doi.org/10.1038/ng1089>.
- (29) Slotkin, R. K.; Martienssen, R. Transposable Elements and the Epigenetic Regulation of the Genome. *Nature Reviews. Genetics* **2007**, *8* (4), 272–285. <http://dx.doi.org.go.libproxy.wakehealth.edu/10.1038/nrg2072>.

- (30) Zhang, G.; Huang, H.; Liu, D.; Cheng, Y.; Liu, X.; Zhang, W.; Yin, R.; Zhang, D.; Zhang, P.; Liu, J.; Li, C.; Liu, B.; Luo, Y.; Zhu, Y.; Zhang, N.; He, S.; He, C.; Wang, H.; Chen, D. N6-Methyladenine DNA Modification in *Drosophila*. *Cell* **2015**, *161* (4), 893–906. <https://doi.org/10.1016/j.cell.2015.04.018>.
- (31) Reik, W.; Dean, W.; Walter, J. Epigenetic Reprogramming in Mammalian Development. *Science* **2001**, *293* (5532), 1089–1093. <https://doi.org/10.1126/science.1063443>.
- (32) Okamoto, I.; Otte, A. P.; Allis, C. D.; Reinberg, D.; Heard, E. Epigenetic Dynamics of Imprinted X Inactivation during Early Mouse Development. *Science* **2004**, *303* (5658), 644–649. <https://doi.org/10.1126/science.1092727>.
- (33) Egger, G.; Liang, G. N.; Aparicio, A.; Jones, P. A. Epigenetics in Human Disease and Prospects for Epigenetic Therapy. *Nature* **2004**, *429* (6990), 457–463. <https://doi.org/10.1038/nature02625>.
- (34) Korlach, J.; Turner, S. W. Going beyond Five Bases in DNA Sequencing. *Current Opinion in Structural Biology* **2012**, *22* (3), 251–261. <https://doi.org/10.1016/j.sbi.2012.04.002>.
- (35) Baker, M. Reproducibility Crisis: Blame It on the Antibodies. *Nature* **2015**, *521* (7552), 274–276. <https://doi.org/10.1038/521274a>.
- (36) Song, C.-X.; Clark, T. A.; Lu, X.-Y.; Kislyuk, A.; Dai, Q.; Turner, S. W.; He, C.; Korlach, J. Sensitive and Specific Single-Molecule Sequencing of 5-Hydroxymethylcytosine. *Nature Methods* **2012**, *9* (1), 75-U188. <https://doi.org/10.1038/nmeth.1779>.

- (37) Song, C.-X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S.-Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G.-L.; Jin, P.; He, C. Genome-Wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell* **2013**, *153* (3), 678–691. <https://doi.org/10.1016/j.cell.2013.04.001>.
- (38) Zhang, L.; Szulwach, K. E.; Hon, G. C.; Song, C. X.; Park, B.; Yu, M.; Lu, X. Y.; Dai, Q.; Wang, X.; Street, C. R.; Tan, H. P.; Min, J. H.; Ren, B.; Jin, P.; He, C. TET-Mediated Covalent Labelling of 5-Methylcytosine for Its Genome-Wide Detection and Sequencing. *Nature Communications* **2013**, *4*, 1517. <https://doi.org/10.1038/ncomms2527>.
- (39) Lu, X.; Han, D.; Zhao, B. S.; Song, C.-X.; Zhang, L.-S.; Doré, L. C.; He, C. Base-Resolution Maps of 5-Formylcytosine and 5-Carboxylcytosine Reveal Genome-Wide DNA Demethylation Dynamics. *Cell Res* **2015**, *25* (3), 386–389. <https://doi.org/10.1038/cr.2015.5>.
- (40) Kavli, B.; Otterlei, M.; Slupphaug, G.; Krokan, H. E. Uracil in DNA--General Mutagen, but Normal Intermediate in Acquired Immunity. *DNA Repair (Amst.)* **2007**, *6* (4), 505–516. <https://doi.org/10.1016/j.dnarep.2006.10.014>.
- (41) Grollman, A. P.; Moriya, M. Mutagenesis by 8-Oxoguanine: An Enemy Within. *Trends in Genetics* **1993**, *9* (7), 246–249. [https://doi.org/10.1016/0168-9525\(93\)90089-z](https://doi.org/10.1016/0168-9525(93)90089-z).
- (42) Jacobs, A. L.; Schär, P. DNA Glycosylases: In DNA Repair and Beyond. *Chromosoma* **2012**, *121* (1), 1–20. <https://doi.org/10.1007/s00412-011-0347-4>.
- (43) Wiebauer, K.; Jiricny, J. Mismatch-Specific Thymine Dna Glycosylase and Dna Polymerase-Beta Mediate the Correction of G.t Mispairs in Nuclear Extracts from

- Human-Cells. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, 87 (15), 5842–5845.
<https://doi.org/10.1073/pnas.87.15.5842>.
- (44) Maiti, A.; Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites. *J. Biol. Chem.* **2011**, 286 (41), 35334–35338.
<https://doi.org/10.1074/jbc.C111.284620>.
- (45) Tini, M.; Benecke, A.; Evans, R. M.; Chambon, P. Association of CBP/P300 Acetylase and Thymine DNA Glycosylase Links DNA Repair and Transcription. *Mol. Cell* **2002**, 9 (2), 265–277. [https://doi.org/10.1016/S1097-2765\(02\)00453-7](https://doi.org/10.1016/S1097-2765(02)00453-7).
- (46) Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. A. DNA-Bound Structures and Mutants Reveal Abasic DNA Binding by APE1 DNA Repair and Coordination. *Nature* **2000**, 403 (6768), 451–456. <https://doi.org/10.1038/35000249>.
- (47) Waters, T. R.; Gallinari, P.; Jiricny, J.; Swann, P. F. Human Thymine DNA Glycosylase Binds to Apurinic Sites in DNA but Is Displaced by Human Apurinic Endonuclease 1. *J. Biol. Chem.* **1999**, 274 (1), 67–74.
- (48) Chou, K.-M.; Cheng, Y.-C. The Exonuclease Activity of Human Apurinic/Apyrimidinic Endonuclease (APE1). Biochemical Properties and Inhibition by the Natural Dinucleotide Gp4G. *J. Biol. Chem.* **2003**, 278 (20), 18289–18296. <https://doi.org/10.1074/jbc.M212143200>.
- (49) Abner, C. W.; Lau, A. Y.; Ellenberger, T.; Bloom, L. B. Base Excision and DNA Binding Activities of Human Alkyladenine DNA Glycosylase Are Sensitive to the Base Paired with a Lesion. *J. Biol. Chem.* **2001**, 276 (16), 13379–13387.
<https://doi.org/10.1074/jbc.M010641200>.

- (50) Wyatt, M. D.; Allan, J. M.; Lau, A. Y.; Ellenberger, T. E.; Samson, L. D. 3-Methyladenine DNA Glycosylases: Structure, Function, and Biological Importance. *Bioessays* **1999**, *21* (8), 668–676. [https://doi.org/10.1002/\(SICI\)1521-1878\(199908\)21:8<668::AID-BIES6>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1521-1878(199908)21:8<668::AID-BIES6>3.0.CO;2-D).
- (51) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *PNAS* **1977**, *74* (12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>.
- (52) Breiling, A.; Lyko, F. Epigenetic Regulatory Functions of DNA Modifications: 5-Methylcytosine and Beyond. *Epigenetics & Chromatin* **2015**, *8* (1), 24. <https://doi.org/10.1186/s13072-015-0016-6>.
- (53) Shapiro, Robert.; Servis, R. E.; Welcher, Marvin. Reactions of Uracil and Cytosine Derivatives with Sodium Bisulfite. *J. Am. Chem. Soc.* **1970**, *92* (2), 422–424. <https://doi.org/10.1021/ja00705a626>.
- (54) Hayatsu, H.; Wataya, Y.; Kai, K. Addition of Sodium Bisulfite to Uracil and to Cytosine. *Journal of the American Chemical Society* **1970**, *92* (3), 724–726. <https://doi.org/10.1021/ja00706a062>.
- (55) Bird, A. The Essentials of DNA Methylation. *Cell* **1992**, *70* (1), 5–8. [https://doi.org/10.1016/0092-8674\(92\)90526-i](https://doi.org/10.1016/0092-8674(92)90526-i).
- (56) Razin, A.; Cedar, H. DNA Methylation and Gene Expression. *Microbiology and Molecular Biology Reviews* **1991**, *55* (3), 451–458.
- (57) Bergman, Y.; Cedar, H. DNA Methylation Dynamics in Health and Disease. *Nat. Struct. Mol. Biol.* **2013**, *20* (3), 274–281. <https://doi.org/10.1038/nsmb.2518>.

- (58) Das, P. M.; Singal, R. DNA Methylation and Cancer. *Journal of Clinical Oncology* **2004**, *22* (22), 4632–4642. <https://doi.org/10.1200/jco.2004.07.151>.
- (59) Tanaka, K.; Okamoto, A. Degradation of DNA by Bisulfite Treatment. *Bioorganic & Medicinal Chemistry Letters* **2007**, *17* (7), 1912–1915. <https://doi.org/10.1016/j.bmcl.2007.01.040>.
- (60) Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. Role of Tet Proteins in 5mC to 5hmC Conversion, ES Cell Self-Renewal, and ICM Specification. *Nature* **2010**, *466* (7310), 1129–1133. <https://doi.org/10.1038/nature09303>.
- (61) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **2011**, *333* (6047), 1300–1303. <https://doi.org/10.1126/science.1210597>.
- (62) Huang, J.; Wang, L. Cell-Free DNA Methylation Profiling Analysis—Technologies and Bioinformatics. *Cancers* **2019**, *11* (11), 1741. <https://doi.org/10.3390/cancers11111741>.
- (63) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C.-X.; Jin, P.; Ren, B.; He, C. Tet-Assisted Bisulfite Sequencing of 5-Hydroxymethylcytosine. *Nat Protoc* **2012**, *7* (12), 2159–2170. <https://doi.org/10.1038/nprot.2012.137>.
- (64) Booth, M. J.; Ost, T. W. B.; Beraldi, D.; Bell, N. M.; Branco, M. R.; Reik, W.; Balasubramanian, S. Oxidative Bisulfite Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine. *Nat Protoc* **2013**, *8* (10), 1841–1851. <https://doi.org/10.1038/nprot.2013.115>.

- (65) Lu, X.; Song, C.-X.; Szulwach, K.; Wang, Z.; Weidenbacher, P.; Jin, P.; He, C. Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA. *J Am Chem Soc* **2013**, *135* (25), 9315–9317. <https://doi.org/10.1021/ja4044856>.
- (66) Booth, M. J.; Marsico, G.; Bachman, M.; Beraldi, D.; Balasubramanian, S. Quantitative Sequencing of 5-Formylcytosine in DNA at Single-Base Resolution. *Nature Chemistry* **2014**, *6* (5), 435–440. <https://doi.org/10.1038/nchem.1893>.
- (67) Schutsky, E. K.; DeNizio, J. E.; Hu, P.; Liu, M. Y.; Nabel, C. S.; Fabyanic, E. B.; Hwang, Y.; Bushman, F. D.; Wu, H.; Kohli, R. M. Nondestructive, Base-Resolution Sequencing of 5-Hydroxymethylcytosine Using a DNA Deaminase. *Nat. Biotechnol.* **2018**. <https://doi.org/10.1038/nbt.4204>.
- (68) Liu, C.; Cui, X.; Zhao, B. S.; Narkhede, P.; Gao, Y.; Liu, J.; Dou, X.; Dai, Q.; Zhang, L.-S.; He, C. DNA 5-Methylcytosine-Specific Amplification and Sequencing. *Journal of the American Chemical Society* **2020**. <https://doi.org/10.1021/jacs.9b12707>.
- (69) Wang, F.; Zahid, O. K.; Swain, B. E.; Parsonage, D.; Hollis, T.; Harvey, S.; Perrino, F. W.; Kohli, R. M.; Taylor, E. W.; Hall, A. R. Solid-State Nanopore Analysis of Diverse DNA Base Modifications Using a Modular Enzymatic Labeling Process. *Nano Lett.* **2017**, *17* (11), 7110–7116. <https://doi.org/10.1021/acs.nanolett.7b03911>.
- (70) Hashimoto, H.; Zhang, X.; Cheng, X. Selective Excision of 5-Carboxylcytosine by a Thymine DNA Glycosylase Mutant. *J. Mol. Biol.* **2013**, *425* (6), 971–976. <https://doi.org/10.1016/j.jmb.2013.01.013>.

- (71) Maiti, A.; Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *J Biol Chem* **2011**, 286 (41), 35334–35338. <https://doi.org/10.1074/jbc.C111.284620>.
- (72) Krokan, H. E.; Bjørås, M. Base Excision Repair. *Cold Spring Harb Perspect Biol* **2013**, 5 (4). <https://doi.org/10.1101/cshperspect.a012583>.
- (73) Song, C. X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C. Q.; Li, X. K.; Li, Y. J.; Chen, C. H.; Zhang, W.; Jian, X.; Wang, J.; Zhang, L.; Looney, T. J.; Zhang, B. C.; Godley, L. A.; Hicks, L. M.; Lahn, B. T.; Jin, P.; He, C. A. Selective Chemical Labeling Reveals the Genome-Wide Distribution of 5-Hydroxymethylcytosine. *Nature Biotechnology* **2011**, 29 (1), 68–72. <https://doi.org/10.1038/nbt.1732>.
- (74) Dapprich, J.; Ferriola, D.; Mackiewicz, K.; Clark, P. M.; Rappaport, E.; D'Arcy, M.; Sasson, A.; Gai, X.; Schug, J.; Kaestner, K. H.; Monos, D. The next Generation of Target Capture Technologies - Large DNA Fragment Enrichment and Sequencing Determines Regional Genomic Variation of High Complexity. *BMC Genomics* **2016**, 17 (1), 486. <https://doi.org/10.1186/s12864-016-2836-6>.
- (75) Crowley, E.; Di Nicolantonio, F.; Loupakis, F.; Bardelli, A. Liquid Biopsy: Monitoring Cancer-Genetics in the Blood. *Nature Reviews Clinical Oncology* **2013**, 10 (8), 472–484. <https://doi.org/10.1038/nrclinonc.2013.110>.
- (76) Ramsahoye, B. H.; Biniszkiwicz, D.; Lyko, F.; Clark, V.; Bird, A. P.; Jaenisch, R. Non-CpG Methylation Is Prevalent in Embryonic Stem Cells and May Be Mediated by DNA Methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, 97 (10), 5237–5242. <https://doi.org/10.1073/pnas.97.10.5237>.

- (77) Dekker, C. Solid-State Nanopores. *Nature Nanotechnology* **2007**, 2 (4), 209–215. <https://doi.org/10.1038/nnano.2007.27>.
- (78) Wanunu, M. Nanopores: A Journey towards DNA Sequencing. *Physics of Life Reviews* **2012**, 9 (2), 125–158. <https://doi.org/10.1016/j.plrev.2012.05.010>.
- (79) Li, J. L.; Gershow, M.; Stein, D.; Brandin, E.; Golovchenko, J. A. DNA Molecules and Configurations in a Solid-State Nanopore Microscope. *Nature Materials* **2003**, 2 (9), 611–615. <https://doi.org/10.1038/nmat965>.
- (80) Storm, A. J.; Chen, J. H.; Zandbergen, H. W.; Dekker, C. Translocation of Double-Strand DNA through a Silicon Oxide Nanopore. *Physical Review E* **2005**, 71 (5), 051903.
- (81) Skinner, G. M.; van den Hout, M.; Broekmans, O.; Dekker, C.; Dekker, N. H. Distinguishing Single- and Double-Stranded Nucleic Acid Molecules Using Solid-State Nanopores. *Nano Letters* **2009**, 9 (8), 2953–2960. <https://doi.org/10.1021/nl901370w>.
- (82) van den Hout, M.; Skinner, G. M.; Klijnhout, S.; Krudde, V.; Dekker, N. H. The Passage of Homopolymeric RNA through Small Solid-State Nanopores. *Small* **2011**, 7 (15), 2217–2224. <https://doi.org/10.1002/sml.201100265>.
- (83) Firnkes, M.; Pedone, D.; Knezevic, J.; Doeblinger, M.; Rant, U. Electrically Facilitated Translocations of Proteins through Silicon Nitride Nanopores: Conjoint and Competitive Action of Diffusion, Electrophoresis, and Electroosmosis. *Nano Letters* **2010**, 10 (6), 2162–2167. <https://doi.org/10.1021/nl100861c>.

- (84) Plesa, C.; Kowalczyk, S. W.; Zinsmeister, R.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. Fast Translocation of Proteins through Solid State Nanopores. *Nano Letters* **2013**, *13* (2), 658–663. <https://doi.org/10.1021/nl3042678>.
- (85) Larkin, J.; Henley, R. Y.; Muthukumar, M.; Rosenstein, J. K.; Wanunu, M. High-Bandwidth Protein Analysis Using Solid-State Nanopores. *Biophysical Journal* **2014**, *106* (3), 696–704. <https://doi.org/10.1016/j.bpj.2013.12.025>.
- (86) Storm, A. J.; Storm, C.; Chen, J. H.; Zandbergen, H.; Joanny, J. F.; Dekker, C. Fast DNA Translocation through a Solid-State Nanopore. *Nano Letters* **2005**, *5* (7), 1193–1197. <https://doi.org/10.1021/nl048030d>.
- (87) Smeets, R. M. M.; Kowalczyk, S. W.; Hall, A. R.; Dekker, N. H.; Dekker, C. Translocation of RecA-Coated Double-Stranded DNA through Solid-State Nanopores. *Nano Letters* **2009**, *9* (9), 3089–3095. <https://doi.org/10.1021/nl803189k>.
- (88) Marshall, M. M.; Ruzicka, J.; Zahid, O. K.; Henrich, V. C.; Taylor, E. W.; Hall, A. R. Nanopore Analysis of Single-Stranded Binding Protein Interactions with DNA. *Langmuir* **2015**, *31* (15), 4582–4588. <https://doi.org/10.1021/acs.langmuir.5b00457>.
- (89) Zahid, O. K.; Zhao, B. S.; He, C.; Hall, A. R. Quantifying Mammalian Genomic DNA Hydroxymethylcytosine Content Using Solid-State Nanopores. *Scientific Reports* **2016**, *6*, 29565. <https://doi.org/10.1038/srep29565>.
- (90) Carlsen, A. T.; Zahid, O. K.; Ruzicka, J. A.; Taylor, E. W.; Hall, A. R. Selective Detection and Quantification of Modified DNA with Solid-State Nanopores. *Nano Letters* **2014**, *14* (10), 5488–5492. <https://doi.org/10.1021/nl501340d>.

- (91) Howarth, M.; Chinnapen, D. J. F.; Gerrow, K.; Dorrestein, P. C.; Grandy, M. R.; Kelleher, N. L.; El-Husseini, A.; Ting, A. Y. A Monovalent Streptavidin with a Single Femtomolar Biotin Binding Site. *Nature Methods* **2006**, *3* (4), 267–273. <https://doi.org/10.1038/nmeth861>.
- (92) Zahid, O. K.; Wang, F.; Ruzicka, J. A.; Taylor, E. W.; Hall, A. R. Sequence-Specific Recognition of MicroRNAs and Other Short Nucleic Acids with Solid-State Nanopores. *Nano Lett.* **2016**, *16* (3), 2033–2039. <https://doi.org/10.1021/acs.nanolett.6b00001>.
- (93) Guliaev, A. B.; Sági, J.; Singer, B. Sequence-Dependent Conformational Perturbation in DNA Duplexes Containing an EA•T Mismatch Using Molecular Dynamics Simulation. *Carcinogenesis* **2000**, *21* (9), 1727–1736. <https://doi.org/10.1093/carcin/21.9.1727>.
- (94) Boiteux, S.; Gajewski, E.; Laval, J.; Dizdaroglu, M. Substrate Specificity of the Escherichia Coli Fpg Protein Formamidopyrimidine-DNA Glycosylase: Excision of Purine Lesions in DNA Produced by Ionizing Radiation or Photosensitization. *Biochemistry* **1992**, *31* (1), 106–110. <https://doi.org/10.1021/bi00116a016>.
- (95) Steinacher, R.; Schär, P. Functionality of Human Thymine DNA Glycosylase Requires SUMO-Regulated Changes in Protein Conformation. *Current Biology* **2005**, *15* (7), 616–623. <https://doi.org/10.1016/j.cub.2005.02.054>.
- (96) Maiti, A.; Morgan, M. T.; Pozharski, E.; Drohat, A. C. Crystal Structure of Human Thymine DNA Glycosylase Bound to DNA Elucidates Sequence-Specific Mismatch Recognition. *Proc Natl Acad Sci U S A* **2008**, *105* (26), 8890–8895. <https://doi.org/10.1073/pnas.0711061105>.

- (97) Tan, S. C.; Yiap, B. C. DNA, RNA, and Protein Extraction: The Past and The Present. *J Biomed Biotechnol* **2009**, *2009*. <https://doi.org/10.1155/2009/574398>.
- (98) Green, N. M. Avidin and Streptavidin. *Methods in Enzymology* **1990**, *184*, 51–67.
- (99) Li, S.; Liu, H.; Jia, Y.; Deng, Y.; Zhang, L.; Lu, Z.; He, N. A Novel SNPs Detection Method Based on Gold Magnetic Nanoparticles Array and Single Base Extension. *Theranostics* **2012**, *2* (10), 967–975. <https://doi.org/10.7150/thno.5032>.
- (100) Royo, J. L.; Hidalgo, M.; Ruiz, A. Pyrosequencing Protocol Using a Universal Biotinylated Primer for Mutation Detection and SNP Genotyping. *Nature Protocol* **2007**, *2* (7), 1734–1739. <https://doi.org/10.1038/nprot.2007.244>.
- (101) Song, C.-X.; Yi, C.; He, C. Mapping Recently Identified Nucleotide Variants in the Genome and Transcriptome. *Nat Biotech* **2012**, *30* (11), 1107–1116. <https://doi.org/10.1038/nbt.2398>.
- (102) Szychowski, J.; Mahdavi, A.; Hodas, J. J. L.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. Cleavable Biotin Probes for Labeling of Biomolecules via the Azide – Alkyne Cycloaddition. *Journal of the American Chemical Society* **2010**, *132* (51), 18351–18360. <https://doi.org/10.1021/ja1083909.Cleavable>.
- (103) Rudolf, G. C.; Heydenreuter, W.; Sieber, S. A. Chemical Proteomics : Ligation and Cleavage of Protein Modifications. *Current Opinion in Chemical Biology* **2013**, *17* (1), 110–117. <https://doi.org/10.1016/j.cbpa.2012.11.007>.
- (104) Leriche, G.; Chisholm, L.; Wagner, A. Bioorganic & Medicinal Chemistry Cleavable Linkers in Chemical Biology. *Bioorganic & Medicinal Chemistry* **2012**, *20* (2), 571–582. <https://doi.org/10.1016/j.bmc.2011.07.048>.

- (105) Jurinke, C.; van den Boom, D.; Collazo, V.; Lüchow, A.; Jacob, A.; Köster, H. Recovery of Nucleic Acids from Immobilized Biotin-Streptavidin Complexes Using Ammonium Hydroxide and Applications in MALDI-TOF Mass Spectrometry. *Anal. Chem.* **1997**, *69* (5), 904–910.
- (106) Holmberg, A.; Blomstergren, A.; Nord, O.; Lukacs, M.; Lundeberg, J.; Uhlén, M. The Biotin-Streptavidin Interaction Can Be Reversibly Broken Using Water at Elevated Temperatures. *Electrophoresis* **2005**, *26*, 501–510.
<https://doi.org/10.1002/elps.200410070>.
- (107) Rothmund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440* (7082), 297. <https://doi.org/10.1038/nature04586>.
- (108) Sambrook, J.; Russell, D. W. Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harb Protoc* **2006**, *2006* (1), pdb.prot4455.
<https://doi.org/10.1101/pdb.prot4455>.
- (109) Adams, M. D.; Soares, M. B.; Kerlavage, A. R.; Fields, C.; Venter, J. C. Rapid CDNA Sequencing (Expressed Sequence Tags) from a Directionally Cloned Human Infant Brain CDNA Library. *Nat. Genet.* **1993**, *4* (4), 373–380.
<https://doi.org/10.1038/ng0893-373>.
- (110) Cheng, H.-R.; Jiang, N. Extremely Rapid Extraction of DNA from Bacteria and Yeasts. *Biotechnol Lett* **2006**, *28* (1), 55–59. <https://doi.org/10.1007/s10529-005-4688-z>.
- (111) Barnett, R.; Larson, G. A Phenol–Chloroform Protocol for Extracting DNA from Ancient Samples. In *Ancient DNA: Methods and Protocols*; Shapiro, B., Hofreiter,

- M., Eds.; *Methods in Molecular Biology*; Humana Press: Totowa, NJ, 2012; pp 13–19. https://doi.org/10.1007/978-1-61779-516-9_2.
- (112) Kirby, K. S. A New Method for the Isolation of Deoxyribonucleic Acids: Evidence on the Nature of Bonds between Deoxyribonucleic Acid and Protein. *Biochem J* **1957**, *66* (3), 495–504.
- (113) Green, M. R.; Sambrook, J. Isolation of High-Molecular-Weight DNA Using Organic Solvents. *Cold Spring Harb Protoc* **2017**, *2017* (4), pdb.prot093450. <https://doi.org/10.1101/pdb.prot093450>.
- (114) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **1996**, *16* (271), 997–999.
- (115) Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S. Cooperative Hydrogen Bond Interactions in the Streptavidin-Biotin System. *Protein Sci.* **2006**, *15* (3), 459–467. <https://doi.org/10.1110/ps.051970306>.
- (116) Rose, A. S.; Bradley, A. R.; Valasatava, Y.; Duarte, J. M.; Prlic, A.; Rose, P. W. NGL Viewer: Web-Based Molecular Graphics for Large Complexes. *Bioinformatics* **2018**, *34* (21), 3755–3758. <https://doi.org/10.1093/bioinformatics/bty419>.
- (117) Chen, Z.; FUJIWARA, S.; BHAT, N. K.; LAUTENBERGER, J. A.; FISHER, R. J. Immunologic Detection of Protein Antigens After Phenol- Chloroform Denaturation. *Gene Analysis Techniques* **1989**, *6* (2), 44–46. [https://doi.org/10.1016/0735-0651\(89\)90026-5](https://doi.org/10.1016/0735-0651(89)90026-5).

- (118) Strick, T. R.; Allemand, J. F.; Bensimon, D.; Bensimon, A.; Croquette, V. The Elasticity of a Single Supercoiled DNA Molecule. *Science* **1996**, *271* (5257), 1835–1837. <https://doi.org/10.1126/science.271.5257.1835>.
- (119) Smith, S. B.; Cui, Y. J.; Bustamante, C. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* **1996**, *271* (5250), 795–799. <https://doi.org/10.1126/science.271.5250.795>.
- (120) Smith, D. E.; Tans, S. J.; Smith, S. B.; Grimes, S.; Anderson, D. L.; Bustamante, C. The Bacteriophage Phi 29 Portal Motor Can Package DNA against a Large Internal Force. *Nature* **2001**, *413* (6857), 748–752. <https://doi.org/10.1038/35099581>.
- (121) Dekker, C. Solid-State Nanopores. *Nature Nanotechnology* **2007**, *2* (4), 209–215. <https://doi.org/10.1038/nnano.2007.27>.
- (122) Tabard-Cossa, V.; Wiggin, M.; Trivedi, D.; Jetha, N. N.; Dwyer, J. R.; Marziali, A. Single-Molecule Bonds Characterized by Solid-State Nanopore Force Spectroscopy. *ACS Nano* **2009**, *3* (10), 3009–3014. <https://doi.org/10.1021/nn900713a>.
- (123) Bell, N. A. W.; Keyser, U. F. Digitally Encoded DNA Nanostructures for Multiplexed, Single-Molecule Protein Sensing with Nanopores. *Nat Nanotechnol* **2016**, *11* (7), 645–651. <https://doi.org/10.1038/nnano.2016.50>.
- (124) Chen, K.; Juhasz, M.; Gularek, F.; Weinhold, E.; Tian, Y.; Keyser, U. F.; Bell, N. A. W. Ionic Current-Based Mapping of Short Sequence Motifs in Single DNA Molecules Using Solid-State Nanopores. *Nano Lett.* **2017**, *17* (9), 5199–5205. <https://doi.org/10.1021/acs.nanolett.7b01009>.

- (125) Bayley, H.; Martin, C. R. Resistive-Pulse Sensing-From Microbes to Molecules. *Chem. Rev.* **2000**, *100* (7), 2575–2594.
- (126) Sischka, A.; Galla, L.; Meyer, A. J.; Spiering, A.; Knust, S.; Mayer, M.; Hall, A. R.; Beyer, A.; Reimann, P. Controlled Translocation of DNA through Lipid-Coated Membranes. **2015**, 4843–4847. <https://doi.org/10.1039/c4an02319f>.
- (127) Marshall, M. M.; Ruzicka, J. A.; Taylor, E. W.; Hall, A. R. Detecting DNA Depurination with Solid-State Nanopores. **2014**, *9* (7), 1–6. <https://doi.org/10.1371/journal.pone.0101632>.
- (128) Wanunu, M.; Sutin, J.; McNally, B.; Chow, A.; Meller, A. DNA Translocation Governed by Interactions with Solid-State Nanopores. *Biophysical Journal* **2008**, *95* (10), 4716–4725. <https://doi.org/10.1529/biophysj.108.140475>.
- (129) Li, J.; Talaga, D. S. The Distribution of DNA Translocation Times in Solid-State Nanopores. *J Phys Condens Matter* **2010**, *22* (45), 454129. <https://doi.org/10.1088/0953-8984/22/45/454129>.
- (130) Bell, N. A. W.; Muthukumar, M.; Keyser, U. F. Translocation Frequency of Double-Stranded DNA through a Solid-State Nanopore. *Phys. Rev. E* **2016**, *93* (2), 022401. <https://doi.org/10.1103/PhysRevE.93.022401>.
- (131) Carlsen, A. T.; Zahid, O. K.; Ruzicka, J. A.; Taylor, E. W.; Hall, A. R. Selective Detection and Quantification of Modified DNA with Solid-State Nanopores. *Nano Lett.* **2014**, *14* (10), 5488–5492. <https://doi.org/10.1021/nl501340d>.
- (132) Zahid, O. K.; Wang, F.; Ruzicka, J. A.; Taylor, E. W.; Hall, A. R. Sequence-Specific Recognition of MicroRNAs and Other Short Nucleic Acids with Solid-

State Nanopores. *Nano letters* **2016**, *16* (3), 2033–2039.

<https://doi.org/10.1021/acs.nanolett.6b00001>.

- (133) Fologea, D.; Gershow, M.; Ledden, B.; McNabb, D. S.; Golovchenko, J. A.; Li, J. L. Detecting Single Stranded DNA with a Solid State Nanopore. *Nano Lett.* **2005**, *5* (10), 1905–1909. <https://doi.org/10.1021/nl051199m>.
- (134) Rivas, F.; Zahid, O. K.; Reesink, H. L.; Peal, B. T.; Nixon, A. J.; DeAngelis, P. L.; Skardal, A.; Rahbar, E.; Hall, A. R. Label-Free Analysis of Physiological Hyaluronan Size Distribution with a Solid-State Nanopore Sensor. *Nat Commun* **2018**, *9* (1), 1037. <https://doi.org/10.1038/s41467-018-03439-x>.
- (135) Bansal, N.; Mims, J.; Kuremsky, J. G.; Olex, A. L.; Zhao, W.; Yin, L.; Wani, R.; Qian, J.; Center, B.; Marrs, G. S.; Porosnicu, M.; Fetrow, J. S.; Tsang, A. W.; Furdui, C. M. Broad Phenotypic Changes Associated with Gain of Radiation Resistance in Head and Neck Squamous Cell Cancer. *Antioxidants & Redox Signaling* **2014**, *21* (2), 221–236. <https://doi.org/10.1089/ars.2013.5690>.
- (136) Mims, J.; Bansal, N.; Bharadwaj, M. S.; Chen, X.; Molina, A. J.; Tsang, A. W.; Furdui, C. M. Energy Metabolism in a Matched Model of Radiation Resistance for Head and Neck Squamous Cell Cancer. *Radiat Res* **2015**, *183* (3), 291–304. <https://doi.org/10.1667/RR13828.1>.
- (137) Chen, X.; Liu, L.; Mims, J.; Punska, E. C.; Williams, K. E.; Zhao, W.; Arcaro, K. F.; Tsang, A. W.; Zhou, X.; Furdui, C. M. Analysis of DNA Methylation and Gene Expression in Radiation-Resistant Head and Neck Tumors. *Epigenetics* **2015**, *10* (6), 545–561. <https://doi.org/10.1080/15592294.2015.1048953>.

- (138) van Dijk, E. L.; Auger, H.; Jaszczyszyn, Y.; Thermes, C. Ten Years of Next-Generation Sequencing Technology. *Trends in Genetics* **2014**, *30* (9), 418–426. <https://doi.org/10.1016/j.tig.2014.07.001>.
- (139) Bearden, S.; Wang, F.; Hall, A. R. Simple and Efficient Room-Temperature Release of Biotinylated Nucleic Acids from Streptavidin and Its Application to Selective Molecular Detection. *Anal. Chem.* **2019**, *91* (13), 7996–8001. <https://doi.org/10.1021/acs.analchem.9b01873>.
- (140) Marcial, V. A.; Komaki, R. Altered Fractionation and Extended-Field Irradiation of Carcinoma of the Cervix. *Cancer* **1995**, *76* (10 Suppl), 2152–2158. [https://doi.org/10.1002/1097-0142\(19951115\)76:10+<2152::aid-cncr2820761340>3.0.co;2-w](https://doi.org/10.1002/1097-0142(19951115)76:10+<2152::aid-cncr2820761340>3.0.co;2-w).
- (141) Lawrence, Y. R.; Li, X. A.; el Naqa, I.; Hahn, C. A.; Marks, L. B.; Merchant, T. E.; Dicker, A. P. RADIATION DOSE–VOLUME EFFECTS IN THE BRAIN. *Int J Radiat Oncol Biol Phys* **2010**, *76* (3 Suppl), S20–S27. <https://doi.org/10.1016/j.ijrobp.2009.02.091>.
- (142) Ma, L.; Men, Y.; Feng, L.; Kang, J.; Sun, X.; Yuan, M.; Jiang, W.; Hui, Z. A Current Review of Dose-Escalated Radiotherapy in Locally Advanced Non-Small Cell Lung Cancer. *Radiol Oncol* **2019**, *53* (1), 6–14. <https://doi.org/10.2478/raon-2019-0006>.
- (143) Radicella, J. P.; Dherin, C.; Desmaze, C.; Fox, M. S.; Boiteux, S. Cloning and Characterization of HOGG1, a Human Homolog of the OGG1 Gene of *Saccharomyces Cerevisiae*. *PNAS* **1997**, *94* (15), 8010–8015. <https://doi.org/10.1073/pnas.94.15.8010>.

- (144) Kestin, L.; Grills, I.; Guckenberger, M.; Belderbos, J.; Hope, A. J.; Werner-Wasik, M.; Sonke, J.-J.; Bissonnette, J.-P.; Xiao, Y.; Yan, D. Dose–Response Relationship with Clinical Outcome for Lung Stereotactic Body Radiotherapy (SBRT) Delivered via Online Image Guidance. *Radiotherapy and Oncology* **2014**, *110* (3), 499–504. <https://doi.org/10.1016/j.radonc.2014.02.002>.
- (145) FACR, D. E. H., MD, MBA, FACRO; FInstP, M. S. H., PhD, DABR, FAAPM; MSc, J. M. H., MD. *Stereotactic Radiosurgery and Stereotactic Body Radiation Therapy (SBRT)*; Springer Publishing Company, 2018.
- (146) Diaz, L. A.; Bardelli, A. Liquid Biopsies: Genotyping Circulating Tumor DNA. *J Clin Oncol* **2014**, *32* (6), 579–586. <https://doi.org/10.1200/JCO.2012.45.2011>.
- (147) Goodman, M. Gamma Knife Radiosurgery: Current Status and Review. *Southern Medical Journal* **1990**, *83* (5), 551–554.
- (148) Yamamoto, M. Gamma Knife Radiosurgery: Technology, Applications, and Future Directions. *Neurosurgery Clinics of North America* **1999**, *10* (2), 181–202. [https://doi.org/10.1016/S1042-3680\(18\)30189-X](https://doi.org/10.1016/S1042-3680(18)30189-X).

APPENDIX

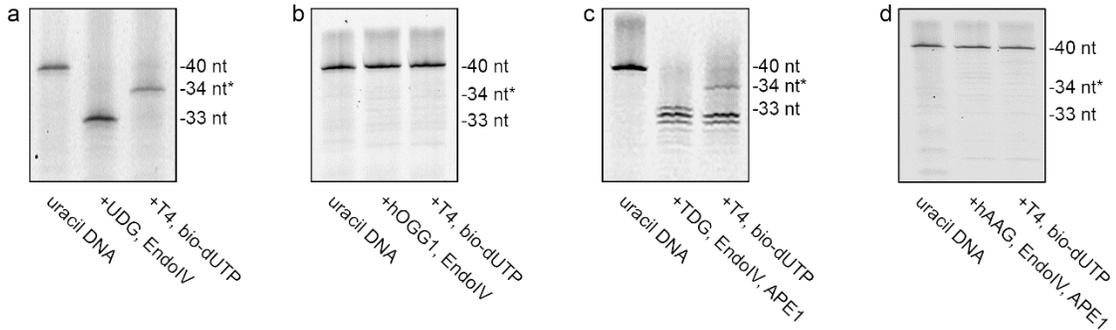


Figure 29. Cross-recognition for uracil modified bases

*Denaturing gel analyses for an oligonucleotide containing a single uracil treated with (a) UDG, (b) hOGG1, (c) TDG, and (d) hAAG. Lanes are labeled with treatment steps and * indicates 34 nt length plus the incorporated biotin. Aside from UDG, only TDG shows recognition, though yield of biotin-labeled product is very low due to additional exonuclease activity (lower bands) under our conditions.*

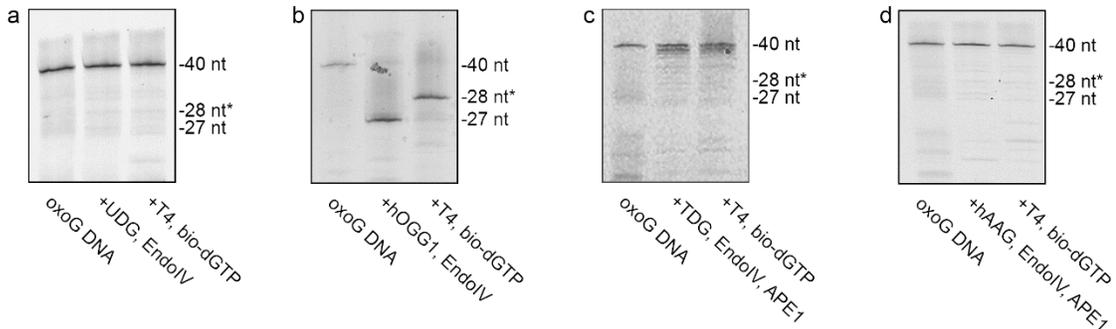


Figure 30. Cross-recognition for oxoG modified bases

*Denaturing gel analyses for an oligonucleotide containing a single oxoG treated with (a) UDG, (b) hOGG1, (c) TDG, and (d) hAAG. Lanes are labeled with treatment steps and * indicates 28 nt length plus the incorporated biotin.*

indicates 28 nt length plus the incorporated biotin. Only *hOGG1* shows any significant yield of biotin-labeled product.

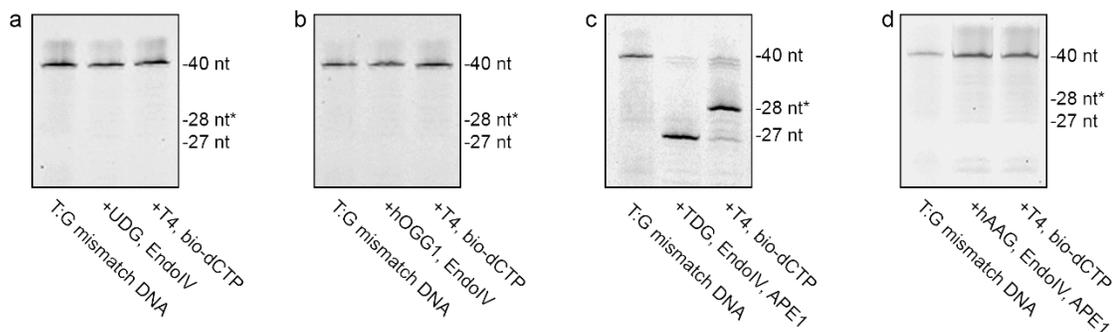


Figure 31. Cross-recognition for T:G mismatch

Denaturing gel analyses for an oligonucleotide containing a single T:G mismatch treated with (a) UDG, (b) *hOGG1*, (c) TDG, and (d) hAAG. Lanes are labeled with treatment steps and * indicates 28 nt length plus the incorporated biotin. Only TDG shows any significant yield of biotin-labeled product.

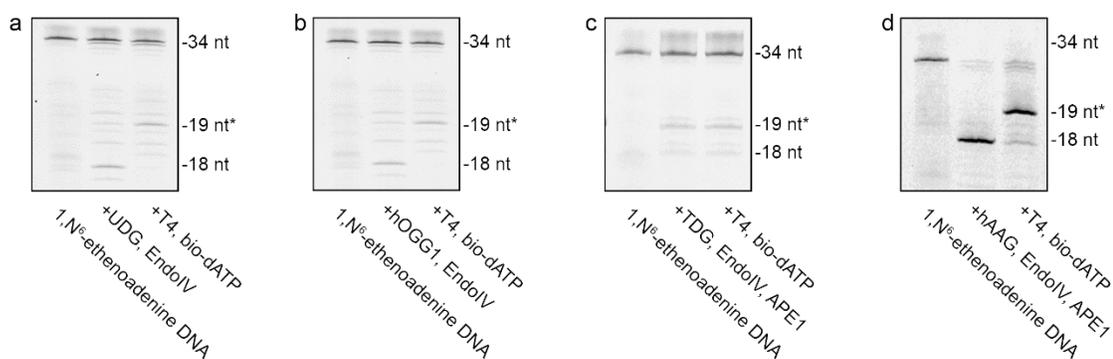


Figure 32. Cross-recognition for 1,*N*⁶-ethenoadenine modified bases

Denaturing gel analyses for an oligonucleotide containing a single 1,*N*⁶-ethenoadenine treated with (a) UDG, (b) *hOGG1*, (c) TDG, and (d) hAAG. Lanes are labeled with

*treatment steps and * indicates 19 nt length plus the incorporated biotin. Some minor recognition is observed for off-target glycosylases, possibly due to DNA kinks induced by 1,N⁶-ethenoadenine, but only hAAG shows a high yield of biotin-labeled product.*

CURRICULUM VITAE

FANNY WANG

fwang35@gmail.com

cell: (917) 535-0628

EDUCATION

Virginia Tech-Wake Forest School of Biomedical Engineering, Winston-Salem, NC

December 2020

Anticipated Ph.D. in Biomedical Engineering

Harvard University, Cambridge, MA

May 2012

B.A. with concentration in Chemistry, citation in Spanish

RESEARCH EXPERIENCE

Adam Hall Lab, Winston-Salem, NC

September 2014-Present

Used solid-state nanopores for biological detection with an eye toward diagnostic purposes of diseases caused by epigenetic modifications (particularly cancer). Developed a novel enzymatic labeling system for labeling and isolating specific epigenetic modifications.

Experimental Physical Chemistry Lab, Cambridge, MA

January 2012-May 2012

Adsorbed chemical dyes onto nanoparticles to test their viability for breaking down bisphenol-A (BPA) contamination in water sources under natural light. Prepared a research paper and poster to present results.

George Church Lab, Boston, MA

January 2011-May 2012

Incorporated artificial amino acids into peptide chains with the flexizyme system, produced a specialized PURE enzymatic system for in-house use.

Biogen IDEC, Cambridge, MA: Intern

May 2011-August 2011

Synthesized a naphthalene-core based library of organic molecules for testing in the autotoxin (ATX) assay to find potential inhibitors of ATX for the treatment of multiple sclerosis. Analyzed data and prepared procedures for future papers and participated in group meetings to update the Medicinal Chemistry department on the project's progress.

Experimental Chemistry and Chemical Biology Lab, Cambridge, MA

September-December 2010

Synthesized artificial amino acids and incorporated them into peptide chains with the flexizyme system. Wrote a research paper summarizing procedures and results.

New York City Public Health Laboratory, New York, NY: Supervisor's Assistant

June-August 2010

Performed viral testing and identification assays with PCR. As a supervisor's assistant, designed and implemented Levy-Jennings charts in Excel, assisted in creating a proposal for state approval, wrote assay validation summaries for the implementation of new assays, developed improved Standard Operating Procedures process flows, organized unit meetings and prepared memos.

PUBLICATIONS

OK Zahid, **F Wang**, JA Ruzicka, EW Taylor, AR Hall. "Sequence-specific recognition of microRNAs and other short nucleic acids with solid-state nanopores." Nano Letters, 2016.

F Wang, OK Zahid, B Swain, D Parsonage, T Hollis, S Harvey, F Perrino, RM Kohli, EW Taylor, AR Hall. “Solid-state nanopore analysis of diverse DNA base modifications using a modular enzymatic labeling process.” *Nano Letters*, 2017.

S Bearden*, **F Wang***, AR Hall. “Simple and efficient room-temperature release of biotinylated nucleic acids from streptavidin and its application to selective molecular detection.” *Analytical Chemistry*, 2019.

F Wang, OK Zahid, U Ghanty, RM Kohli, AR Hall. “A modular approach for affinity-labelling the base elements of the cytosine demethylation pathway in DNA.” (Submitted)

OK Zahid, F Rivas, **F Wang**, S Bearden, K Sethi, K Reiss, AR Hall. “Solid-state nanopore analysis of global 5-hydroxymethylcytosine content in healthy breast tissue and early-stage breast tumor tissue.” (In preparation)

CONFERENCES

F Wang, OK Zahid, AR Hall. Detection of Specific Nucleic Acid Sequences in a Mixed Solution with Solid-State Nanopores. *2015 BMES Annual Meeting*. Tampa, FL. October 7-10, 2015.

F Wang, OK Zahid, AR Hall. Highly specific and modular affinity labeling of DNA damage elements. *Wake Forest University Annual Redox Retreat*, Winston Salem, NC. May 20, 2016.

F Wang, OK Zahid, AR Hall. Highly Specific and Modular Affinity Labeling of Epigenetic Modifications. *2016 BMES Annual Meeting*. Minneapolis, MN. October 5-8, 2016.

F Wang, OK Zahid, AR Hall. Solid-state nanopore analysis of diverse DNA base modifications using a modular enzymatic labeling process. *Structural and Computational Biology Symposium*, Winston Salem, NC. June 5, 2018.

F Wang, OK Zahid, AR Hall. Solid-state nanopore analysis of diverse DNA base modifications using a modular enzymatic labeling process. *2019 Biophysical Society Annual Meeting*. Baltimore, MD. March 2-5, 2019.

F Wang, OK Zahid, AR Hall. Solid-state nanopore analysis of diverse DNA base modifications using a modular enzymatic labeling process. *Wake Forest University Center for Redox Biology and Medicine Retreat*. Winston Salem, NC. April 5, 2019.

F Wang, OK Zahid, AR Hall. Analysis of diverse DNA base modifications using a modular enzymatic labeling process. *Virginia Tech-Wake Forest University School of Biomedical Engineering and Sciences Research Symposium*, Blacksburg, VA. May 15, 2019.

FUNDING

NIH Ruth L. Kirchstein National Research Service Award, 2016-2019