UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mass Spectrometric Studies of Structurally Modified DNA

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Environmental Toxicology

by

Shuo Liu

December 2015

Dissertation Committee:

Dr. Yinsheng Wang, Chairperson
Dr. David A. Eastmond
Dr. Constance I. Nugent

The Dissertation of Shuo Liu is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

COPYRIGHT ACKNOWLEDGEMENTS

The text and figures in Chapter 1, in part or in full, are a reprint of the material as it appears in ***Chemical Society Reviews*** 2015, 44(21), p. 7829-7854. The coauthor (Dr. Yinsheng Wang) listed in that publication revised the manuscript which forms the basis of this chapter.

The text and figures in Chapter 2, in part or in full, are a reprint of the material as it appears in ***Analytical Chemistry*** 2013, 85(14), p. 6732-6739 and the supporting information therein. The coauthor (Dr. Yinsheng Wang) listed in this publication directed and supervised the research which forms the basis of this chapter.

The text and figures in Chapter 3, in part or in full, are a reprint of the material as it appears in ***Nucleic Acids Research*** 2013, 41(13), p. 6421-6429 and the supporting information there in. The coauthor (Dr. Yinsheng Wang) listed in this publication directed and supervised the research which forms the basis of this chapter.

The text and figures in Chapter 4, in part or in full, are a reprint of the material as it may appear in ***PLoS One*** 2013, 8(12), e84620. The coauthor (Dr. Yinsheng Wang)

listed in this publication directed and supervised the research which forms the basis of this chapter.

The text and figures in Chapter 5, in part or in full, are a reprint of the material as it appears in ***Environmental Science & Technology*** 2015, 49(19), p. 11923-11931 and the supporting information there in. The coauthor (Dr. Yinsheng Wang) listed in this publication directed and supervised the research which forms the basis of this chapter.

The text and figures in Chapter 6, in part or in full, are a reprint of the material as it appears in ***Journal of the American Society for Mass Spectrometry*** 2014, 25(10), p. 1763-1770 and the supporting information there in. The coauthor (Dr. Yinsheng Wang) listed in this publication directed and supervised the research which forms the basis of this chapter.

DEDICATION AND ACKNOWLEDGEMENTS

The research outlined in this dissertation would not be possible without the help and inspiration from many people at University of California, Riverside. First and foremost, I would like to thank my supervisor, Dr. Yinsheng Wang, for taking me as a graduate student over five years ago. As an undergraduate majored in environmental science, I began my doctoral study with very little experience in analytical chemistry. It was Dr. Wang who taught me almost everything required for my research. His breadth of knowledge, scope of teaching techniques, patience, encouragement, guidance and support enabled me to work in a wide range of bioanalytical fields and complete this dissertation. Dr. Wang also sets an excellent example for me in many essential skills that a researcher should have – great enthusiasm, attention to detail but also able to see the bigger picture, hard work, problem solving, multi-tasking and time management. His energy, enthusiasm, and interest in research have been conveyed to me, kindling my passion for science. I appreciate all his contributions of time, ideas and help on both my professional and personal growth.

Next, I would like to express my gratitude to the rest of my dissertation committee members: Dr. David Eastmond and Dr. Connie Nugent for their suggestions during my graduate career and discussion with my research. I truly appreciate their concerns, generous help and accessibility over the past five years. I am also grateful for the

technical supports from Ron New at UC Riverside who trained me how to use MALDI-TOF mass spectrometer. Additionally, I would like to thank my collaborators who provided me great opportunities for highly interdisciplinary training at the interface of chemistry and biology. They are Dr. Michael Seidman at National Institutes of Health, Dr. Hongjun Song at John Hopkins University, Dr. Gerd Pfeifer at City of Hope, Dr. Anjana Rao at La Jolla Institute for Allergy and Immunology, Dr. Xiaohua Peng at University of Wisconsin–Milwaukee, Dr. Robert Sabatini at University of Georgia, Dr. Kent Gates at University of Missouri, Dr. Guoping Fan at University of California, Los Angeles, and Dr. Jikui Song at University of California, Riverside.

The members of the Wang group have contributed greatly to my personal and professional time at UC Riverside, which became some of the best of my life. I would like to thank Dr. Lijuan Fu and Dr. Nisana Andersen for teaching me most of the basic techniques and for their priceless friendship since I first came to the United States. I appreciate Dr. Huachuan Cao, Dr. Candace Guerrero, Dr. Debin Ji, and Dr. Nicholas Amato for helping with the synthesis of internal standards used in this dissertation; to Dr. Renee William and Dr. Ashley Swanson who were always willing to help me improve my presentation skills; to Ming Huang and Dr. Lin Li who often cooked for me and kept me alive during the busy graduation season. My gratitude also extends to

other former and current group members: Dr. Jianshuang Wang, Dr. Hongxia Wang, Dr. Xiaoli Dong, Dr. Lei Xiong, Dr Bifeng Yuan, Dr. John Prins, Dr. Changjun You, Dr. Jin Wang, Dr. Yongsheng Xiao, Dr. Lei Guo, Dr. Xiaoxia Dai, Dr. Qianqian Zhai, Dr. Jun Wu, Dr. Nathan Price, Dr. Tao Bing, Dr. Xiaogang Jiang, Dr. Qian Cai, Dr. Fan Zhang, Preston Williams, Nicole Williams, Eric Stephens, Zi Wang, Pengcheng Wang, Yang Yu, Ji Jiang, Yuxiang Cui, Jiabin Wu, Weili Miao, Lok Tam, David Bade, Gwendolyn Gonzalez, Xuejiao Dong, and Jianan Sun.

Last but certainly not least, I would like to thank my family, including my grandparents, aunts, uncles, and cousins, for their endless support, encouragement and love; my boyfriend, Yang Yang for his love, understanding as well as his company no matter what we meet, well or ill. To my lovely parents, who gave me the best start in life and the support that I needed to build a dream to chase after. I love you from the bottom of my heart. Thank you for sharing your love of life, sense of humor, excitement for every experience, and courage to look fear directly in the face until it backs down. Thank you for being strict with your rules when I was led down to the wrong path and giving me the strength to stand up and expect nothing less than the very best. There is nothing I can do to repay you, but I hope to make you proud.

ABSTRACT OF THE DISSERTATION

Mass Spectrometric Studies of Structurally Modified DNA

by

Shuo Liu

Doctor of Philosophy, Graduate Program in Environmental Toxicology
University of California, Riverside, December 2015
Dr. Yinsheng Wang, Chairperson

The advances in mass spectrometry (MS) instrumentation have rendered MS an indispensable tool for structure elucidation, quantification, and assessment of the biological consequences of DNA modifications. DNA encodes genetic information that is used in the development and functioning of essentially all known organisms. Changes in DNA could give rise to compromised transmission of genetic information. Abnormal modifications are generally harmful to cells and will be converted to permanent genetic damage if they evade DNA repair, whereas epigenetic DNA modifications are regulators of cell function that influence chromatin structure and control the levels of gene expression. The emphasis of this dissertation is on the application of MS, together with biological techniques, for assessing the occurrence and biological consequences of these two major types of DNA modifications.

After an introduction presented in Chapter 1, Chapter 2 focuses on monitoring the

enzymatic removal of 8-methoxypsoralen-induced DNA interstrand cross-links (ICLs) in cultured mammalian cells by using an HPLC coupled with tandem-MS approach. With genetic depletion of DNA repair proteins, we also pinpointed the role of nucleotide excision repair pathway in removing DNA ICLs. Chapters 3 and 4 discuss the development of an HPLC-MS/MS/MS method, along with the isotope dilution technique, for accurate measurements of 5-hydroxymethyl, 5-formyl and 5-carboxyl derivatives of 2'-deoxycytidine and 5-hydroxymethyl-2'-deoxyuridine (5-hmdU) in genomic DNA of cultured human cells, mammalian tissues and plant tissues. This is the first direct quantitative comparison between the levels of multiple 5-mdC derivatives in mammalian tissues and the first report of their presence in the model plant Arabidopsis. Chapter 5 describes the combination of this method with other bioanalytical techniques for studying the mechanisms of arsenic epigenotoxicity. We found that arsenite could bind directly to the zinc fingers of Tet proteins and substantially impair the catalytic efficiency of Tet proteins in oxidizing 5-mC both *in vitro* and in cells. In Chapter 6, we report the application of LC-MS/MS for quantifying β-D-glucosyl-5-hydroxymethyl-2'-deoxyuridine (dJ) in *Trypanosoma brucei*. The results provide evidence that dJ binding proteins play an important role in oxidizing thymidine to form 5-hmdU, which then facilitates the generation of dJ.

Table of Contents

List of Schemes

List of Figures

List of Tables

Chapter 1: General Introduction

1.1. Overview

Faithful transmission of genetic information during cell differentiation and across generations plays a significant role in the development and survival of all the living organisms. Being a “bank” that stores most of the genetic instructions, double-stranded DNA consists of two sugar-phosphate backbones, which are coiled around each other and resistant to cleavage, as well as paired nucleobases. In addition to stabilizing the double-helical structure of DNA and permitting sequence-specific recognition by cellular proteins, the unique chemical properties of nucleobases render DNA susceptible to modifications. Among the many types of DNA modifications, methylation occurring at the C5 position of cytosine residues at CpG dinucleotide sites is the best-studied covalent modification of DNA and it is also indispensable for mammalian cells to regulate gene expression. However, opposite to the beneficial side of DNA modification, other DNA adducts induced by exogenous or endogenous sources are harmful and may compromise cellular functions by blocking DNA replication and/or inducing mutations. Given that the flow of genetic information through a modified DNA depends on its chemical structure, the biological significance and health risks may vary from one modification to another. Therefore, a robust

quantitative assessment of DNA modifications would be of great importance in biological, medical, and toxicological studies.

With the increasing interest in understanding the biological effects of DNA modifications on living organism, the investigation has proceeded using more sophisticated test systems, isolated DNA and cultured cells treated *in vitro* and experimental animals and humans. Since DNA modifications are often present at very low levels in biological systems, considerable emphasis has been placed on the development of sensitive and specific methods for the detection and quantification of these modifications. Especially for *in vivo* studies where a limited amount of tissue sample may be available, the analytical method of choice must have adequate sensitivity to measure DNA modifications.

Over the last three decades, a variety of analytical techniques have been used for DNA analysis, including the ^{32}P-postlabeling assay (1), immunoassays (2), and electrochemical detection (3). In this vein, the development of the ^{32}P-postlabeling method greatly enhanced the ability to detect chemically modified DNA, owing to its high sensitivity (1 modification in 10^{10} nucleotides) and low sample requirement (1-10 μg DNA) (1,4). However, ^{32}P-postlabeling, like any other analytical methods, has some disadvantages. The assay uses a radioactive isotope that can pose health risks; thus, the method requires extra caution and a dedicated, isolated experimental environment.

More importantly, the assay provides very limited structural information, which may lead to ambiguity in the identities of the DNA modifications under investigation. Recent developments in mass spectrometric methods for DNA analysis have overcome these shortcomings. Despite possessing poorer sensitivity than the ^{32}P-postlabeling method at the very beginning, the sensitivity of MS methods has improved rapidly as a consequence of sustained development in instrumentation and sample preparation methods. In light of these continuing advances, MS techniques have gained widespread application in the qualitative and quantitative analysis of DNA modifications by providing sufficient sensitivity, specificity, and detailed structural information.

The overall objective of this dissertation is to develop mass spectrometric methods for the structure elucidation, quantification, and the assessment of the biological consequences of DNA modifications, including both DNA damage and epigenetic marks. The main focus of this chapter will be placed on an introduction into reactive sites on nucleobases that tend to be modified, proteins interacting with various DNA modifications, sample processing prior to and during MS analysis, and applications of MS for understanding the biological consequences of DNA modifications.

1.2. Reactive Sites on Nucleobases

Being held by complementary base pairing, the stable structure of duplex DNA

relies mainly on the shape of nucleobases and the molecular electrostatic potentials of the nucleophilic sites on nucleobases and base pairs. The unique chemical properties of nucleobases also permit sequence-specific recognition of DNA by cellular proteins (5). Yet, for the same reason, the functional groups on nucleotides render DNA susceptible to modifications by electrophilic agents, which lie at the heart of understanding mutagenesis (6-8). Therefore, a better understanding of the reactive sites of DNA provides a strong reference for the identification of potential DNA modifications and builds a solid foundation for the assessment of toxicity and health risk. Figure 1.1 displays the reactive sites on nucleobases where modifications are preferably formed.

1.2.1 Alkylation of Nucleobases

The most extensively studied DNA modifications are generated through the covalent reactions between DNA bases and alkylating agents, the mode of which usually involves transferring alkyl groups from these agents to nucleophilic sites on nucleobases (9). Studies of these DNA alkylation-inducing chemicals uncovered most ring nitrogen atoms as well as exocyclic oxygen and nitrogen atoms of nucleobases as targets for alkylation. These include *N*7, *N*3, N^2, *N*1, and O^6 positions of guanine, the *N*7, N^6, *N*3, and *N*1 positions of adenine, the *N*3, N^4, and O^2 positions of cytosine, and the *N*3, O^2, and O^4 positions of thymine. In this context, alkylation may also occur at two nucleophilic sites in the same nucleobase. For instance, etheno derivatives of

guanine, adenine, and cytosine can arise from either metabolic activation of exogenous chemicals (10) or endogenous metabolic processes (e.g. lipid peroxidation) (11). Their targeting sites are almost all localized between the exocyclic nitrogen and one of its neighboring ring nitrogen atoms, which give rise to the formation of 1,N^2- and 3,N^2-ethenoguanine (εGua), 1,N^6-ethenoadenine (εAde), and 3,N^4-ethenocytosine (εCyt) (12-14).

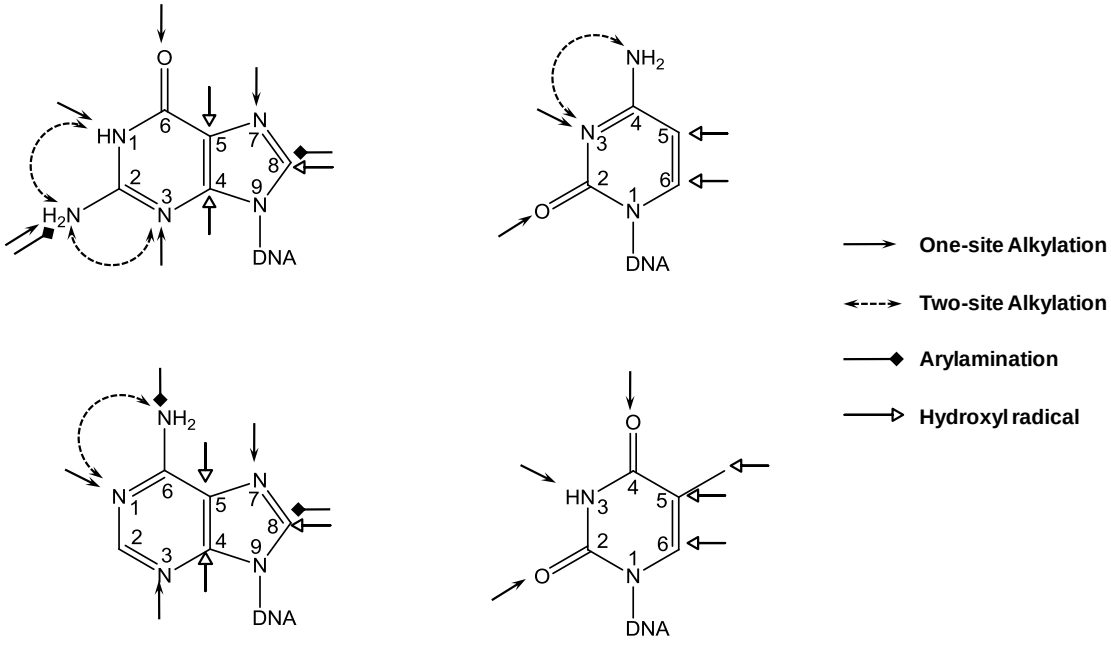


Figure 1.1. An illustration of the reactive sites on nucleobases.

Some therapeutic drugs act through their conjugation with reactive sites in DNA to produce monoadducts and/or cross-links that can inhibit critical cellular processes in abnormally proliferating cells (15,16). For example, nitrogen mustards bind to the *N*7

position of guanine to yield monoadducts (MAs), and a second reaction with the *N*7 of guanine on the opposing DNA strand yields interstrand cross-links (ICLs) (17). On the other hand, mitomycin C reacts almost exclusively with the N^2 of guanines to form either MAs or ICLs at CpG sites (18). More complex cross-links are produced upon the exposure of DNA to platinum derivatives, a group of bioreductive prodrugs that target the *N*7 positions of guanine and adenine (19). Another example for therapeutic DNA binding occurs between psoralen derivatives and thymine, whose C5 and C6 atoms frequently serve as the reactive sites in the formation of cross-links (20).

1.2.2 Arylamination of Nucleobases

Different from DNA alkylation, a heterogeneous group of carcinogens can modify DNA via arylamination. This group includes, for example, aromatic amines (e.g. 4-aminobiphenyl, 4-ABP) (21), nitroaromatic compounds (e.g. 2-nitrofluorene) (22), and heterocyclic aromatic amines found in cooked fish and meats (23-25). Although the sites of substitution on nucleobases vary substantially, it appears that the C8 atom and the exocyclic amino nitrogen of the purine bases, particularly of guanine, are the major targets for arylamination. Such a pattern is in stark contrast to that of alkylating agents which displayed lower efficiency in targeting the aforementioned two reactive sites.

1.2.3 Radical-Induced Modifications on Nucleobases

Hydroxyl radical is one of the most reactive species formed from aerobic

metabolism or from exposure to ionizing radiation. This radical can bind to pyrimidine bases with a preference for the C5 and C6 positions, generating C5-OH- and C6-OH-adduct radicals, respectively; it can also conjugate with purine bases, giving rise to C4-OH-, C5-OH-, and C8-OH-adduct radicals for both guanine and adenine (26,27) as well as C2-OH-adduct radical for adenine (28,29). Examples of hydroxyl radical-induced DNA adducts include 5-hydroxyuracil (5-OH-Ura), 5-hydroxycytosine (5-OH-Cyt), 5,6-dihydroxy-5,6-dihydrothymine (a.k.a. thymine glycol), and 8-oxo-7,8-dihydroguanine (8-oxo-Gua) (30-33). Additionally, hydroxyl radical can abstract a hydrogen atom from the 5-methyl group of 5-methylcytosine and thymine to yield the corresponding 5-hydoxymethyl, 5-formyl and 5-carboxyl derivatives of the two nucleobases (34-38). In this context, these oxidations of the 5-methyl group of 5-methylcytosine, mediated by the ten-eleven translocation (Tet) family of dioxygenases, have attracted great attention in recent years owing to their potential involvement in active cytosine demethylation in mammals (39-41).

1.3. Proteins Interacting with DNA Modifications in Mammals

1.3.1 Proteins Interacting with DNA Damage

Exogenous and endogenous chemical species can react, directly or after metabolic activation, with DNA, eventually leading to DNA damage. If not repaired correctly or

in a timely fashion, DNA lesions may result in myriads of alterations within the cell, including, but are not limited to, cell death, mutations in the genome, and aberrant cell cycle control (7). There are many DNA repair pathways - the major being nucleotide excision repair (NER), base excision repair (BER), homologous recombination (HR), non-homologous end joining (NHEJ), and mismatch repair (MMR) (Figure 1.2). To maintain the intact genome for a healthy organism, mammalian cells execute one or more repair processes to preserve genome integrity when damage arises.

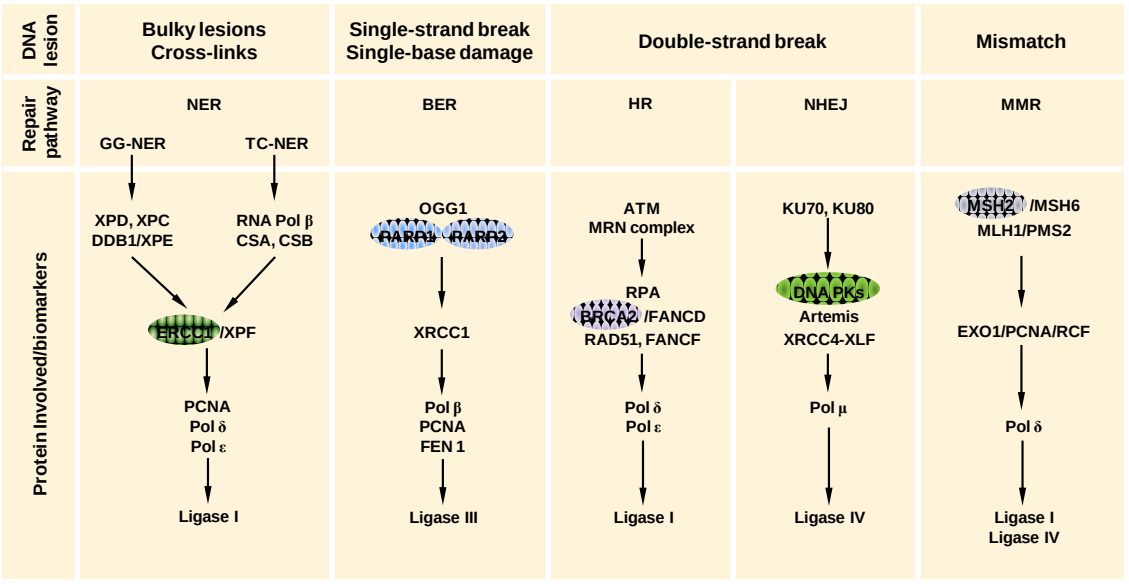


Figure 1.2. Major DNA repair pathways.

1.3.1.1 Nucleotide Excision Repair (NER)

NER is the process that repairs large bulky lesions, particularly nucleobase modifications that alter the normal helical structure of duplex DNA. DNA damage of this type include cyclobutane pyrimidine dimers and pyrimidine(6-4)pyrimidone

photoproducts induced by ultraviolet light (42), base modifications created by exogenous chemical agents [e.g. cisplatin (43)] or by endogenous byproducts of lipid peroxidation (44), and tandem modifications such as purine cyclonucleosides that are generated by reactive oxygen species (45). The recognition by NER in response to DNA damage involves two subpathways: global-genome NER (GG-NER) and transcription-coupled NER (TC-NER). GG-NER is a process independent of transcription, which employs protein complexes comprised of damaged DNA-binding (DDB) protein and XPC-Rad23B to constantly screen for distortions in the genome; while TC-NER is initiated when RNA polymerase is blocked at a lesion site and serves as a signal for the recruitment of Cockayne syndrome proteins (CSA and CSB) instead of XPC-Rad23B (46). The two subpathways utilize the same machinery subsequently to execute the remaining steps for NER, which include excisions on both sides of the lesion by XPF-ERCC1 and XPG, gap-filling synthesis by DNA polymerases, and ligation by ligases to seal the nick (47,48). Because of the significance of NER in maintaining genome integrity, defects in NER are associated with a group of autosomal recessive human diseases, e.g. xeroderma pigmentosum (XP) and Cockayne syndrome (CS) (49).

1.3.1.2 Base Excision Repair (BER)

Different from the functions of NER, the BER pathway has evolved to protect

against the harmful consequences of small base lesions that don't totally alter the DNA double helical structure. The lesions of BER substrate include a variety of base modifications caused by oxidation, deamination, and alkylation (50). The participation of BER pathway requires a lesion-specific monofunctional DNA glycosylase (e.g. uracil-DNA glycosylase, UDG) or bifunctional DNA glycosylase such as 8-oxoguanine DNA glycosylase (OGG1) in the first step to recognize and excise a modified base (51). While bifunctional DNA glycosylases possess both glycosylase activity and an intrinsic 3' AP lyase activity, the abasic site left by the monofunctional DNA glycosylase is then incised by APE1, which belongs to class II AP endonucleases (51). When the necessary 3'-OH and 5'-phosphate termini are generated, BER typically recruits Pol β to the missing nucleotide and engages the XRCC1-LIGα complex to seal the remaining nick (50).

1.3.1.3 Homologous Recombination (HR)

HR pathway plays an essential role in repairing double strand breaks (DSBs) while maintaining replication fidelity, which requires a homologous DNA strand that has the correct coding information as reference (52). In this context, HR serves as a repair tool during the S and G2 phases of the cell cycle when copying is occurring. DSBs are formed upon replication fork collapse, for example at a polymerase blocking lesion, during duplication of chromosomal DNA in dividing cells. Although the

detailed molecular mechanisms for HR repair remain somewhat elusive, it has been proposed that this pathway is initiated by recognition of the DSBs by the MRN complex (53). As a break reader, the complex recruits the protein kinase and ATM (ataxia telangiectasia mutated) to DSB sites and facilitates the recombination step of HR. Following the recognition process, resection is carried out by C-terminal binding protein (CtBP)-interacting protein (CtIP) and exonucleases, and the resulting single-stranded DNA (ssDNA) is stabilized by binding with replication protein A (RPA) (52). Subsequently, BRCA2 and Rad51 mediate the assembly and replacement of RPA on the RPA-ssDNA complex, respectively (54). The Rad51-ssDNA complex then enables strand invasion of the intact homologous DNA region, which provides the genetic instruction for accurate repair.

1.3.1.4 Other Repair Pathways

In addition to NER and BER, DNA mismatch repair (MMR) constitutes the third pathway for removing single strand DNA damage. This system is responsible for recognizing and repairing erroneous base pairs and insertion-deletion loops that can arise during DNA replication and recombination, as well as some forms of DNA modifications (35). The major active components involved in the MMR pathway are broadly designated as MutS and MutL, which are highly conserved and essential for detecting the mismatch and directing repair machinery to it (55). Mutations in the

human homologues of the Mut proteins have been shown to exhibit genomic instability and result in microsatellite instability, a phenotypic evidence of abnormal MMR function that is implicated in most human cancers (55).

In comparison with HR, non-homologous end joining (NHEJ) is a major DSB repair system in higher eukaryotes especially when a homologous sister chromatid is absent (52). A number of proteins are utilized in the three main steps of NHEJ: 1) eukaryotic Ku70/Ku80 heterodimer forms a complex with the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and serves as a docking site for other NHEJ proteins (56); 2) nucleases process the DSB ends by removing non-ligatable termini or damaged nucleotides at the break and trimming the single-stranded overhangs to reveal complementary stretches; 3) the XRCC4/DNA ligase IV complex performs the ligation step of repair (57). While NHEJ can operate during all phases of the cell cycle, it is a relatively error-prone pathway due to the end processing step.

1.3.2 Proteins Interacting with Epigenetic DNA Modifications

Eukaryotic chromatin possesses a wealth of information that is embedded both genetically in the DNA sequence itself and epigenetically via DNA methylation and post-translational modifications of core histone proteins (58,59). In mammals, DNA methylation occurs at the C5 position of cytosine residues mainly at CpG dinucleotide sites (58). This methylation pattern is established during embryonic development by *de*

novo DNA methyltransferases DNMT3a and DNMT3b, and maintained during cell division by maintenance DNA methyltransferase DNMT1 (58). Studies in the past decade showed that DNA methylation pattern in mammals can be lost through passive or active mechanisms (60,61). Elimination or inhibition of DNMT1 activity can lead to loss of cytosine methylation during cell division, a process known as passive DNA cytosine demethylation. Active DNA cytosine demethylation is the conversion of 5-methylcytosine (5-mC) to unmethylated cytosine in a process that is independent of DNA replication (60).

Multiple lines of evidence support the existence of active cytosine demethylation in the entire genome or in specific genomic loci, and multiple mechanisms have been proposed for this process (61). In this vein, the ten-eleven translocation (Tet) family of Fe^{2+}- and 2-oxoglutarate-dependent dioxygenases have recently been identified as crucial enzymes involved in active DNA demethylation in mammals. The three subfamily members Tet1, Tet2, and Tet3 are capable of catalyzing the iterative oxidation of 5-mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-foC), and 5-carboxylcytosine (5-caC) (40,62). There is also evidence that AID (activation-induced deaminase)/APOBEC (apolipoprotein B mRNA-editing enzyme complex) families of cytidine deaminases mediated demethylation through converting 5-hmC to 5-hydroxymethyluracil (5-hmU). Except 5-hmC, all the other oxidation

products of 5-mC (i.e. 5-foC, 5-caC, and 5-hmU) are readily removed by thymine DNA glycosylase and replaced with unmethylated cytosine by base excision repair (BER) proteins.

1.4. Mass Spectrometry-Based Detection of DNA Modifications

Endogenous DNA modifications are often present at a level that is lower than $1/10^6$ canonical nucleosides. This necessitates the development of highly sensitive and specific methods that require only a small amount of DNA. As techniques have improved, MS has played an increasingly important role in taking up this analytical challenge. In comparison with earlier successful platforms such as ^{32}P-postlabeling, MS has advantages including the use of standards and capability of providing structural information for DNA modifications. To achieve highly sensitive detection, analytes are generally separated and/or enriched prior to MS analysis. The combination of MS with other techniques has also enabled sensitive quantification of trace amounts of DNA modifications in a relatively small amount of DNA isolated from living organisms. An illustration of sample processing for MS analysis of DNA modifications is shown in Figure 1.3.

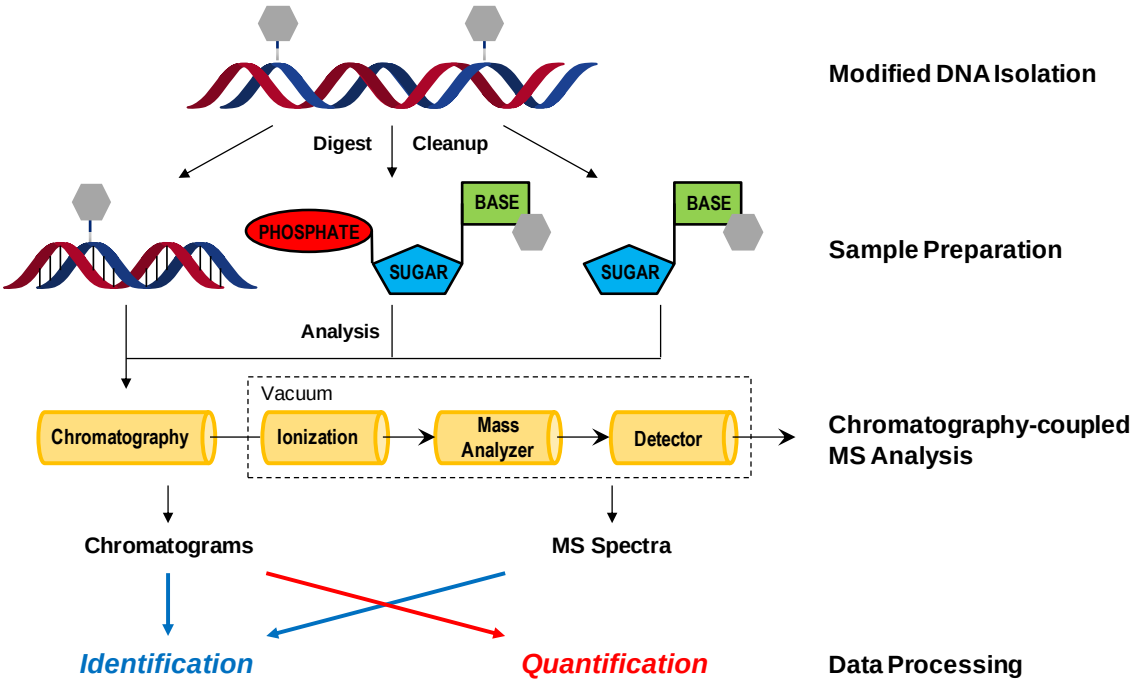


Figure 1.3. Sample processing scheme for MS analysis of DNA modifications.

1.4.1 Sample Preparation

1.4.1.1 DNA Hydrolysis

The analysis of DNA modifications has been performed for DNA samples isolated from cultured cells (63,64), tissues (65-70), and biofluids (14,71,72). Apart from DNA, these samples also contain large amounts of salts, RNA, and proteins in the matrices that may participate in the metabolism and/or interfere with the detection of targeted DNA modifications. Thus, it is crucial to lyse the cells or tissues, isolate DNA, and break up long DNA molecules to analytes that are amenable for MS analysis. In this respect, the analysis of modified DNA in the form of nucleobase, mononucleoside, or

mononucleotide by MS offers the highest sensitivity and accuracy for the quantification of the majority of DNA modifications.

For the majority of DNA modifications that are chemically stable, quantification is often conducted after complete digestion of DNA to 2'-deoxynucleosides. In general, modification-carrying DNA is first digested down to 2'-deoxynucleoside-5'-monophosphates. Among a number of digestion enzymes in this category, DNase I and nucleases are non-specific endonucleases that can cleave DNA into oligodeoxynucleotides (ODNs) and mononucleotides (24,73,74). The subsequent use of phosphodiesterases results in further cleavage of ODNs into mononucleotides. The commonly used phosphodiesterases include spleen phosphodiesterase (phosphodiesterase II) and snake venom phosphodiesterase (phosphodiesterase I), which are 5'- and 3'-exonucleases, respectively (75,76). To obtain better sensitivity through analysis by MS in the positive-ion mode, conversion of nucleotides to nucleosides is often necessary, where alkaline phosphatase is frequently used (77-79).

It should be noted that precautions are often needed during enzymatic hydrolysis to minimize errors if DNA lesions are unstable or can be produced artificially. For instance, nitric oxide ($NO^{\cdot}$), an important physiological messenger involved in cell signalling and an environmental pollutant, can induce deamination of nucleobases *in vivo* (80). The addition of radical scavengers, such as TEMPO (2,2,6,6-

tetramethylpiperidine 1-oxyl), appears to inhibit artifactual deamination throughout sample preparation processes. Problems are also encountered in LC-MS analysis of deamination or other modification products of 2'-deoxyadenosine when deaminase is present in DNA samples (81) or in commercial preparations of enzymes used for DNA hydrolysis (82). To avoid the adventitious deamination, a deaminase inhibitor [i.e. *erythro*-9-(2-hydroxy-3-nonyl)adenine] is often added during DNA digestion (68). Altogether, in cases where some modifications could form as artifacts during sample preparation, methods for measuring DNA modifications should be carefully validated, and the artificial formation or degradation of the analyte should be reduced to a level below the detection limit of the method for a successful analysis.

The efficiency of enzymes in hydrolyzing modified DNA is also affected by the structure of adduct-inducing agents and the conformation of modification-bearing DNA. Bulky DNA modifications that cannot fit into the active site of digestion enzymes may lead to incomplete hydrolysis. For instance, DNA modifications, including dimeric DNA photoproducts (83,84), thymidine glycol (85), and DNA interstrand cross-link lesions (86), cannot be fit into the active site of nuclease P1, which prevents the cleavage of the phosphodiester bond on the 3′ side of the modified nucleoside. While caution needs to be exerted to ensure complete digestion by enzymes during sample preparation, researchers also employed selective enzymatic digestion to produce

analytes suitable for more sensitive detection of DNA modifications by MS and for the assessment of isomeric DNA lesions produced at different sites. Wang and co-workers (87-90) also employed nuclease P1 to degrade ICL-containing DNA induced by psoralen derivatives to a tetranucleotide remnant for LC-MS and MS/MS analyses. Unlike releasing ICL-containing DNA as free base or nucleoside modifications that cannot be distinguished from intrastrand cross-links, the formation of the ICL-bearing tetranucleotide provides unequivocal chemical specificity for the subsequent quantitative analysis as well as information about the identities of the flanking nucleosides of the cross-linked nucleosides.

1.4.1.2 Enrichment

Sample enrichment has been widely employed for quantitative measurement of DNA modifications. Apart from modifications of interest, DNA hydrolysates also contain the bulk of unmodified nucleosides, proteins, inorganic salts, and other components which can interfere with MS analysis. As a result, extensive clean-up steps prior to MS analysis are typically required to further improve detection sensitivity. Meanwhile, following DNA hydrolysis, isolation could greatly influence the ability of MS methods to detect DNA modifications. Examples of sample enrichment methods for DNA modification analyses encompass liquid-liquid extraction (LLE), ultrafiltration, solid-phase extraction (SPE), on-line column switching, off-line

high-performance LC (HPLC), and immunoaffinity chromatography.

Column switching enables the trapping, enrichment and desalting of analytes prior to LC-MS/MS analysis. This is particularly important for nanoLC-MS analysis because column switching facilitates the injection of relatively large volume of sample for on-line LC-MS analysis in the nano flow range (91). Special attention often needs to be paid toward the choice of trapping materials for the trapping column, where the analyte of interest should be efficiently trapped and subsequently eluted for online LC-MS analysis, as shown previously for analyses of peptides and modified ribonucleosides (92,93).

Further separation of modifications from biological samples is also frequently achieved by off-line HPLC. When a DNA sample is digested to nucleosides and subjected to HPLC analysis, even if the modifications can escape detection by a conventional UV detector, the fractions known to contain specific modifications can be collected and analyzed by MS. A number of investigators have employed HPLC enrichment of DNA modifications [e.g. HE-DNA adducts (94,95) and cisplatin 1,2-intrastrand guanine-guanine adduct (96)] for the subsequent MS detection. Wang and co-workers (63,64,68) also employed HPLC enrichment to improve the MS detection sensitivity and specificity of a number of DNA modifications, including carboxyalkylated DNA lesions, oxidatively induced 8,5'-cyclopurine-2'-

deoxynucleosides, and oxidized derivatives of 5-methyl-2'-deoxycytidine and thymidine. This offline HPLC enrichment is considered one of the best options for isolating trace amounts of DNA modifications from DNA hydrolysates prior to MS analysis as it affords nearly quantitative analyte recovery, avoids the introduction of solid phase particles, which can clog nano-HPLC columns (97), and removes most unmodified canonical nucleosides as well as buffer salts employed in enzymatic digestion. However, when higher amounts of analytes are first injected to establish their elution times, the HPLC system can become contaminated. To avoid problems such as analyte carryover and cross-contamination, an HPLC enrichment blank must be included prior to sample enrichment.

1.4.2 Mass Spectrometry-based Quantitative Analysis

1.4.2.1 Standards for Quantitative Analysis

Reliable quantification of DNA modifications by MS requires accurate calibration of MS data using external or internal standards. For quantitation using external standards, the calibration normally involves a simple comparison of MS response (i.e. peak area or peak height) from a target DNA modification with that from an exactly measured quantity of the identical analyte (98,99). Simplicity and applicability to a wide variety of methods constitute the main benefits of external standard calibration; however, its use is severely restricted by the drawback that this calibration could be

greatly influenced by the stability of MS detector as well as the sample matrix that may affect chromatographic separation, analyte ionization and precursor ion selection.

The development of LC-MS has also facilitated the use of internal and co-chromatography standards, which is necessary for monitoring analyte recovery during the assay procedures. Additionally, the use of an internal standard removes the effects of sample matrix on signal suppression and variation in equipment response among runs. Among a variety of approaches, incorporation of a stable isotope-labeled form of the analyte as the internal standard provides accurate and precise quantification by MS. Because these standards are chromatographically identical to, but different in mass from the DNA modifications of interest, their retention time on chromatogram and fragmentation pattern in tandem MS offer unequivocal specificity for the identification of DNA modifications. Stable isotope-labeled internal standards of nucleosides containing ^{2}H, ^{15}N, and/or ^{13}C in nucleosides have been widely used for the quantification of DNA modifications (100-103). Overall, the use of stable isotope-labeled standard provides unambiguous identification and accurate quantification of DNA modifications, and the method should be used whenever possible.

1.4.2.2 Liquid Chromatography-Mass Spectrometry

LC-MS attempts to achieve three desirable attributes for DNA modification

analysis, namely, high sensitivity, low sample requirements, and structural identification. Since conventional-flow LC-MS was employed for the analysis of DNA modifications, a number of publications have demonstrated the increasing role that LC-MS would play in the characterization and detection of DNA modifications. Given that ESI produces ions in the gas-phase directly from a flowing liquid solution at a broad range of flow rates (typically in the range of 100-200,000 nL/min), it became an ideal ionization technique used to interface with liquid-phase separation techniques. In the early 1990s, the hyphenation of LC with ESI-MS was first introduced for the measurement of DNA modifications. The set-up of LC-ESI-MS methodology opened a new door to DNA modification analysis, leading to a dramatic increase in the utilization of MS for detection and quantification of DNA modifications.

Overall, the development of LC-MS in the field of DNA modification analyses has a trend towards a decline in sample consumption and an improvement in detection sensitivity. Tandem MS constitutes a powerful tool for providing specific structural information about DNA modifications via diagnostic fragment ions. Such specificity of modification identification further helps increase the sensitivity of the method by lowering the background noise. Consequently, LC-MS/MS has become the method of choice for the unambiguous and simultaneous structural identification during quantification. Earlier platforms used LC-MS/MS with conventional flow on 1-4.6 mm

internal diameter (i.d.) columns for *in vivo* animal studies. Although tandem MS is highly specific, sensitivity of the assay is sometimes sacrificed due to fragmentation of the parent ion to daughter ions. Moreover, the use of this technique in DNA modification analysis has been limited because the levels of some DNA modifications *in vivo* are far below the limits of detection attainable by standard LC-ESI-MS/MS. To increase assay sensitivity for analytes of low abundance, there is a trend to use LC with a small diameter column at a low flow rate along with micro- and nano-ESI (*μ*ESI and *n*ESI) sources (104,105). The recent emergence of these changes has resulted in profound decreases in the limits of detection for LC-MS/MS.

For instance, in analyses conducted on a conventional bore 2 mm i.d. column, a well-studied nucleoside modification, 4-aminobiphenyl (ABP)-C8-dG, was reported at levels of 4.9-30 modifications per 10^7 nucleosides in hepatic DNA isolated from mice treated with 4-ABP (21). By enzymatically hydrolyzing 100 μg of DNA, this assay achieved a detection limit of ~10 pg on-column, equivalent to 0.7 ABP-C8-dG in 10^7 normal nucleotides. Compared to columns with small i.d., the relatively large bore column used in this LC-MS method has limited analytical sensitivity. For the purpose of improving trace modification detection in biological DNA, a capillary LC-*μ*ESI-MS/MS technique was later carried out on a 320 μm i.d. column to investigate the presence of ABP-C8-dG modifications in human pancreatic DNA,

which yielded an LOQ approaching 1 modification per 10^8 nucleotides when using only 13.3 μg of DNA per analysis (106). This amount was further decreased to 2.5 μg of DNA per analysis in a subsequent study using a capillary column with a smaller i.d. (75 μm i.d.) and a lower flow rate of 200 nL/min compared to the previous 20 μL/min (107). Furthermore, Randall et al. (108) reported an integration of a 75 μm i.d. analytical column with online sample enrichment on a trapping column for sensitive quantification of ABP-C8-dG in 4-ABP-exposed human bladder cells and rat bladder tissues. This improved quantification method has a detection level of 5 modifications per 10^9 nucleosides with the use of 5 μg of DNA and the equivalent of only 1.25 μg of DNA per analysis.

Impressive improvement in overall sensitivity for the detection of trace levels of DNA modifications has also been achieved by coupling *n*LC-based separation methods with multi-stage MS (MS^n). The MS^n scan mode affords better signal-to-noise ratio owing to its high specificity, especially when there are co-eluting isobaric interferences. Meanwhile, the method allows for unambiguous identification of analytes of interest as detailed structural information is obtained from additional fragmentation pathways. Attracted by the analytical merits of multi-stage MS, LC-ESI-MS^3 has been employed for the quantification of a number of DNA lesions including tobacco carcinogen- and cooked-meat carcinogen-induced DNA modifications, purine cyclonucleoside,

5-HmdU, 5-FodU, and N^2-(1-carboxyethyl)-2'-deoxyguanosine in cellular and tissue DNA, and has been shown to be more reliable than the MS^2 mode for the analyses of these types of lesions (63,67,68,109,110).

It is of note that, while LC-MS is advantageous over GC-MS in detecting DNA modifications directly, derivatization can still be employed to enhance the ionization efficiency of analytes during LC-MS analysis. In this regard, Hong et al. (110) employed LC-ESI-MS/MS together with chemical derivatization for the highly sensitive detection of an oxidatively induced thymidine lesion, 5-formyl-2'-deoxyuridine (5-fodU). Given that the detection limit of 5-fodU was relatively poor compared with other types of oxidative DNA base damage, Hong et al. (110) first derivatized 5-fodU with Girard's reagent T (GirT) to form a hydrazone conjugate harboring a precharged quaternary ammonium moiety. This enabled the facile detection of the resulting conjugate by positive-ion ESI-MS with a pronounced increase in detection sensitivity (by ~20-fold compared with direct analysis of the underivatized compound). Although chemical derivatization has not been widely employed for the detection of DNA modifications by LC-MS, we expect to see more applications of this method for the quantitative assessment of certain DNA modifications, which may dramatically improve the detection sensitivity of trace-level analytes in limited amounts of DNA.

1.5. Applications of MS in Biological Studies of DNA Modifications

1.5.1 Structure Elucidation

The last few decades have witnessed significant progress in our understanding of the biological consequences of DNA modifications. In addition to advances in molecular biology, this may also have arisen, at least partly, from the impressive improvement in analytical tools aimed at characterizing DNA at nucleoside or nucleotide levels and in DNA segments. Structure elucidation of an unknown nucleoside or nucleotide modification achieved by chromatography coupled with tandem MS provides a solid foundation for further quantitative analysis (24,100,111). In this vein, stable isotope-labeled standards often play an instrumental role in tandem MS-based qualitative analysis by possessing identical elution time with the analyte on chromatography as well as similar fragmentation pattern in MS/MS (24,101,112). On the other hand, selective stable isotope incorporation to unique sites in the modified nucleosides, together with multi-stage MS analysis, facilitates the unambiguous elucidation of fragmentation pathways of modified nucleosides (32).

1.5.2 Sequence Verification

Driven by the optimal sensitivity and simple fragmentation pattern, DNA modifications are frequently detected as nucleosides or nucleotides. However, sequence

information that is closely correlated with DNA modification formation is lost upon digestion to individual nucleosides or nucleotides. It has been revealed that reaction of carcinogens with DNA often exhibits sequence selectivity, the revelation of which could be important for understanding their mutagenic potentials (17,113). The improved mass range in modern mass spectrometers and the capability in forming multiply charged ions during ESI have made possible the detection and sequencing of larger modified base-bearing ODN fragments. Under the collisional activation conditions, multiply charged ODNs are shown to have a strong tendency to cleave at the *N*-glycosidic bond between the nucleobase and 2-deoxyribose and the 3' C-O bond of the same nucleoside to form [a_n – Base] and w_n series of fragment ions (Figure 1.4) (114). Apart from sequence confirmation and verification of site-specific incorporation of modifications into ODNs via MS and MS/MS, the combination of MS/MS-based ODN sequencing with molecular biology tools also allows for the examination of the biological consequences of DNA modifications, particularly how they compromise the efficiency and fidelity of DNA replication and transcription. For instance, LC-ESI-MS/MS was employed for studying *in vitro* translesion synthesis across various DNA lesions mediated by purified DNA polymerases (115-118).

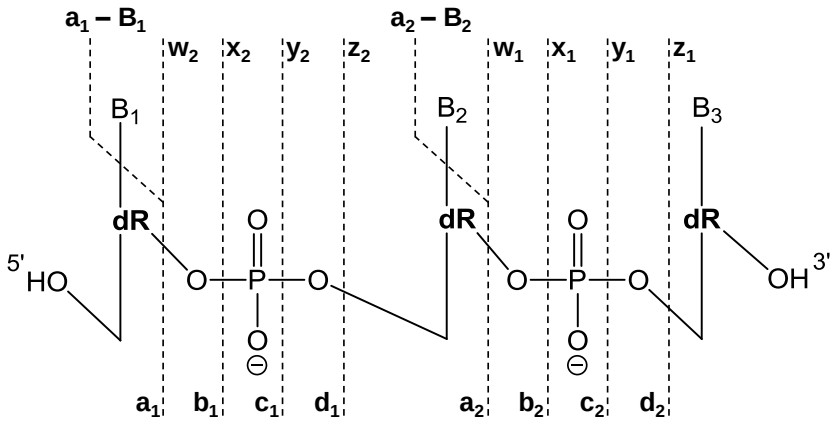


Figure 1.4. Nomenclature of fragment ions for oligodeoxyribonucleotides. 'B1', 'B2', and 'B3' represent nucleobases, and 'dR' designates 2-deoxyribose.

1.5.3 Biomarker Discovery

Chemical modifications in DNA followed by cell proliferation can lead to mutations that may eventually result in cancer development. Biomarkers of carcinogenesis are essential for assessing the human health consequences of exogenous or endogenous exposure to DNA damaging agents. Such exogenous exposure is hardly determined by external doses owing to the potential disposition of toxicants within an organism. Modulated by genetic and environmental factors, the patterns of absorption, distribution, metabolism, and excretion vary among different toxicants. To better predict the mutagenic and carcinogenic potentials of genotoxic agents, more suitable biomarkers are needed. In this vein, the analysis of modified DNA can elucidate mechanisms of action of chemical carcinogens in humans and identify risk-associated biomarkers for prediction of cancer development. Additionally, DNA modifications

formed from endogenous sources are generally considered end points of reaction; thus, monitoring their formation may provide a useful index of exposure to endogenous DNA damaging agents (119,120).

1.5.4 DNA Repair Studies

There are a variety of DNA repair pathways for removing DNA modifications from the genome, as noted above (7). Failure in repair or inaccurate repair may result in adverse biological consequences. A predominant pathway leading to mutation is structural modifications resulting in aberrant base pairing during DNA replication. Structural modifications of DNA bases (e.g. DNA lesions caused by ROS) bring not only immediate, but also long-lasting impacts on the cell. Slow but constant accrual of non-lethal genetic changes in DNA has been considered as a key event leading to chronic genetic diseases, including cancer and aging. While formation of only a few unrepaired DNA modifications among millions of normal DNA bases can result in significant biological consequences, measuring such small differences require analytical methods with high accuracy and sensitivity.

Improvements in MS-based analytical methods for measuring trace levels of DNA modifications, together with genetic manipulation, have offered an opportunity to assess DNA repair process by monitoring directly the levels of DNA modifications. For instance, Wang and co-workers (65,67) found that deficiency in ERCC1 led to

markedly elevated accumulation of 8,5'-cyclo-2'-deoxyadenosine (cdA) and 8,5'-cyclo-2'-deoxyguanosine (cdG) in mouse tissues, thereby providing direct evidence to support the involvement of ERCC1, and thus the NER pathway, in repairing the endogenously induced purine cyclonucleosides *in vivo*. Likewise, Malayappan et al. (121) used LC-MS/MS to follow the formation and repair of cyclophosphamide-induced *N*,*N*-bis[2-(*N*7-guanyl)ethyl]amine cross-link in human blood. The use of an LC-MS/MS assay also led to the observation that ICLs induced by mitomycin C could be partially removed in mouse mammary tumor cells (122).

1.5.5 Epigenetic Studies

1.5.5.1 Quantification for 5-Methylcytosine and Related Epigenetic Modifications

DNA methylation is one of the key epigenetic mechanisms that is involved in many physiological processes, including transcriptional silencing of retrotransposons, host defence, genomic imprinting, and X chromosome inactivation (123). 5-Methylcytosine (5-mC) in the symmetric CpG dinucleotide sites replaces about 4% of total cytosines in the mammalian genome (124,125). However, the mechanism of demethylation has long been elusive until the recent discovery of several related epigenetic modifications of DNA in addition to 5-mC as noted above. Because of the rapid removal of 5-foC, 5-caC, and 5-hmU by thymine DNA glycosylase (TDG) and the replacement with unmethylated C through BER pathway, these modifications are

proposed to be important in active DNA cytosine demethylation in mammals. Moreover, aside from being an intermediate in DNA demethylation, 5-hmC can be recognized by specific cellular proteins (126,127), and thus acts as an epigenetic mark on its own right and regulates gene transcription (128). Changes in the levels of 5-hmC are often accompanied with human diseases, especially cancer (128,129).

Assessing the involvement of these newly discovered DNA modifications in epigenetic regulation necessitates the measurement of the levels of these modifications in cells and tissues. Among these modified bases, 5-mC and 5-hmC are the first two that have been widely quantified using mass spectrometric methods owing to their relatively high abundance in the global genome. For example, by using an LC-ESI-MS/MS method to measure the levels of 5-mC and 5-hmC in cells, Le et al. (130) observed a dramatic increase in 5-mC and 5-hmC levels in the DNA isolated from mouse embryonic stem cells as compared with parental fibroblast cells. This result indicated the existence of DNA methylation and hydroxymethylation regulation during cellular reprogramming. Likewise, Chen et al. (131) developed a mass spectrometric method for quantifying 5-mC and 5-hmC in genomic DNA from hepatocellular carcinoma (HCC) tumor tissues to assess the effects of their presence on epigenetic regulation in HCC. In addition to cells and tissues, 5-mC and 5-hmC have been quantified in DNA from blood and saliva in healthy volunteers (132). More

recently, mass spectrometry has also played an increasingly important role in identifying and quantifying less abundant epigenetically modified bases, and achieved higher detection sensitivity in comparison with immunological, enzymatic, and ^{32}P-postlabeling 1D or 2D thin layer chromatography methods. In this respect, Ito et al. (40) employed HPLC-ESI-MS/MS to quantify 5-foC and 5-caC in genomic DNA of mouse embryonic stem cells and confirmed the two previously unknown cytosine derivatives as the products of Tet proteins. However, their method failed to obtain reliable quantitative data for 5-caC in several mouse organs due to its lower abundance in genomic DNA. Further developments in sample preparation methods and MS instrumentation are expected to improve the detection sensitivity and specificity while reducing the amount of sample required, especially for extreme trace analysis of DNA modifications *in vivo*.

1.5.5.2 Examination of Epigenetic Modifications Occurring in DNA sequences

Owing to the advantage of tandem-MS methods in charactering ODNs, MS also provides a powerful tool for the identification and quantification of epigenetically modified DNA sequences, instead of single nucleoside or nucleotide modifications. This alternative approach is particularly useful for the investigation of DNA modification-interacting proteins, whose activity is normally regulated by the length and/or sequence of DNA. For instance, Ji et al. (133) developed an LC-tandem MS

method to assess how the substitution of 5mC at a hemi-methylated CpG site with a 5-hmC, 5-foC, 5-caC or 5-hmU within synthetic duplex DNA alters the efficiencies of mouse Dnmt1 and human DNMT3a in methylating the cytosine residue in the opposing and adjacent 5' CpG sites. The results from this study provided important insights into the role of 5-mC oxidation products in maintenance cytosine methylation. In another approach, tandem MS was employed for identifying the replication products arising from DNA substrates with site-specific incorporation of cytosine modifications (134). It turned out that the replication of 5-hmC, 5-foC, and 5-caC displayed high fidelity and relatively high efficiency in human cells, which is in agreement with the roles of these 5-mC derivatives in epigenetic regulation.

1.6. Scope of the Dissertation

In Chapter 2, we describe an LC-MS/MS method, together with the isotope dilution technique, for assessing the repair of 8-methoxypsoralen (8-MOP)-induced DNA interstrand crosslinks (ICLs), as well as monoadducts (MAs), in cultured mammalian cells. We found that, while there were substantial decreases in the levels of ICL and MAs in repair-competent cells at 24 hr after 8-MOP/UVA treatment, there was little repair of 8-MOP-ICL and MAs in XPA-deficient human skin fibroblasts and ERCC1-deficient Chinese hamster ovary (CHO) cells over a 24-hr period. This result

provided unequivocal evidence to support that the 8-MOP photoadducts are substrates for nucleotide excision repair (NER) in mammalian cells. This is one of the first reports about the application of LC-MS/MS for assessing the repair of DNA ICLs.

In Chapters 3 and 4, we describe the use of a reversed-phase HPLC-MS/MS/MS method, along with the use of stable isotope-labeled standards, for accurate measurements of 5-hydroxymethyl, 5-formyl and 5-carboxyl derivatives of 2'-deoxycytidine (5-hmdC, 5-fodC and 5-cadC) and 5-hydroxymethyl-2'-deoxyuridine (5-hmdU) in genomic DNA of cultured human cells, mammalian tissues and plant tissues. We found that overexpression of the catalytic domain of human Tet1 led to marked increases in the levels of 5-hmdC, 5-fodC, and 5-cadC, but only a modest increase in 5-hmdU, in genomic DNA of HEK293T cells. Moreover, 5-hmdC is present at a level that is approximately 2-3 and 3-4 orders of magnitude greater than 5-fodC and 5-cadC, respectively, and 35-400 times greater than 5-hmdU in the mouse brain and skin, and in the human brain. To our knowledge, this is the first direct quantitative comparison between the levels of multiple 5-mdC derivatives in mammalian tissues. In addition, we found that, in Arabidopsis DNA, the levels of 5-hmdC, 5-fodC, and 5-cadC are approximately 0.8 modifications per 10^6 nucleosides, with the frequency of 5-hmdC (per 5-mdC) being comparable to that of 5-hmdU (per thymidine). The relatively low levels of the 5-mdC oxidation products suggest that they arise likely

from reactive oxygen species present in plant cells, which is in line with the lack of homologous Tet-family dioxygenase enzymes in Arabidopsis.

In Chapter 5, we report the use of a set of analytical and biological techniques to test our hypothesis that the highly conserved C_3H-type zinc fingers situated in Tet proteins may constitute potential targets for arsenic binding owing to the high binding affinity of As(III) toward cysteine residues. Our results showed that arsenite could bind directly to the zinc fingers of Tet proteins *in vitro* and in cells, and this interaction substantially impaired the catalytic efficiency of Tet proteins in oxidizing 5-mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-foC), and 5-carboxylcytosine (5-caC). Treatments with arsenite also led to a dose-dependent decrease in the level of 5-hmC, but not 5-mC, in DNA isolated from HEK293T cells overexpressing the catalytic domain of each of the three Tet proteins and from mouse embryonic stem cells. The study unveiled, for the first time, that arsenite could alter epigenetic signaling by targeting the zinc fingers of Tet proteins and perturbing the Tet-mediated oxidation of 5-mC *in vitro* and in cells.

In Chapter 6, we describe the development of a reversed-phase HPLC coupled with tandem mass MS method, together with the use of a surrogate internal standard (β-D-glucosyl-5-hydroxymethyl-2'-deoxycytidine), for the accurate detection of β-D-glucosyl-5-hydroxymethyl-2'-deoxyuridine (dJ) in *Trypanosoma brucei* DNA. For

comparison, we also measured the level of the precursor for dJ synthesis, i.e. 5-hmdU. We found that dJ was not detectable in the JBP-null cells while it replaced approximately 0.5% thymine in wild-type cells, which was accompanied with a markedly decreased level of 5-hmdU in JBP1/JBP2-null strain relative to the wild-type strain. These results provided direct evidence that JBP proteins play an important role in oxidizing thymidine to form 5-HmdU, which facilitated the generation of dJ. This is also the first report about the application of LC-MS/MS for the quantification of d J.

References

1. Randerath K, Reddy MV, Gupta RC (1981) ^{32}P-labeling test for DNA damage. *Proc Natl Acad Sci USA* 78: 6126-6129.

2. Hsu IC, Poirier MC, Yuspa SH, Grunberger D, Weinstein IB, et al. (1981) Measurement of benzo(*a*)pyrene-DNA adducts by enzyme immunoassays and radioimmunoassay. *Cancer Res* 41: 1091-1095.

3. Park JW, Cundy KC, Ames BN (1989) Detection of DNA adducts by high-performance liquid chromatography with electrochemical detection. *Carcinogenesis* 10: 827-832.

4. Phillips DH, Arlt VM (2007) The 32P-postlabeling assay for DNA adducts. *Nat Protoc* 2: 2772-2781.

5. Rohs R, Jin X, West SM, Joshi R, Honig B, et al. (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* 79: 233-269.

6. Burcham PC (1998) Genotoxic lipid peroxidation products: their DNA damaging properties and role in formation of endogenous DNA adducts. *Mutagenesis* 13: 287-305.

7. Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, et al. (2006) DNA Repair and Mutagenesis. Washington, D.C.: ASM Press.

8. Marnett LJ (2000) Oxyradicals and DNA damage. *Carcinogenesis* 21: 361-370.

9. Sedgwick B, Bates PA, Paik J, Jacobs SC, Lindahl T (2007) Repair of alkylated DNA: recent advances. *DNA Repair (Amst)* 6: 429-442.

10. Morinello EJ, Ham AJL, Ranasinghe A, Nakamura J, Upton PB, et al. (2002) Molecular dosimetry and repair of N(2),3-ethenoguanine in rats exposed to vinyl chloride. *Cancer Res* 62: 5189-5195.

11. Sun X, Nair J, Linseisen J, Owen RW, Bartsch H (2012) Lipid peroxidation and DNA adduct formation in lymphocytes of premenopausal women: Role of estrogen metabolites and fatty acid intake. *Int J Cancer* 131: 1983-1990.

12. Morinello EJ, Ham AJL, Ranasinghe A, Sangaiah R, Swenberg JA (2001) Simultaneous quantitation of N-2,3-ethenoguanine and 1,N-2-ethenoguanine with an immunoaffinity/gas chromatography/high-resolution mass spectrometry assay. *Chem*

Res Toxicol 14: 327-334.

13. Yen TY, Holt S, Sangaiah R, Gold A, Swenberg JA (1998) Quantitation of 1,N6-ethenoadenine in rat urine by immunoaffinity extraction combined with liquid chromatography/electrospray ionization mass spectrometry. *Chem Res Toxicol* 11: 810-815.

14. Chen HJC, Lin TC, Hong CL, Chiang LC (2001) Analysis of 3,N-4-ethenocytosine in DNA and in human urine by isotope dilution gas chromatography/negative ion chemical ionization/mass spectrometry. *Chem Res Toxicol* 14: 1612-1619.

15. Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA (2008) DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* 8: 193-204.

16. Gerson SL (2004) MGMT: its role in cancer aetiology and cancer therapeutics. *Nat Rev Cancer* 4: 296-307.

17. Hartley JA, Bingham JP, Souhami RL (1992) DNA sequence selectivity of guanine-N7 alkylation by nitrogen mustards is preserved in intact cells. *Nucleic Acids Res* 20: 3175-3178.

18. Warren AJ, Hamilton JW (1996) Synthesis and structural characterization of the N2G-mitomycin C-N2G interstrand cross-link in a model synthetic 23 base pair oligonucleotide DNA duplex. *Chem Res Toxicol* 9: 1063-1071.

19. Jamieson ER, Lippard SJ (1999) Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chem Rev* 99: 2467-2498.

20. Stern RS (2007) Psoralen and ultraviolet a light therapy for psoriasis. *N Engl J Med* 357: 682-690.

21. Doerge DR, Churchwell MI, Marques MM, Beland FA (1999) Quantitative analysis of 4-aminobiphenyl-C8-deoxyguanosyl DNA adducts produced in vitro and in vivo using HPLC-ES-MS. *Carcinogenesis* 20: 1055-1061.

22. Malejka-Giganti D, Parkin DR, Decker RW, Niehans GA, Bliss RL, et al. (2008) Tumorigenicity and genotoxicity of an environmental pollutant 2,7-dinitrofluorene after systemic administration at a low dose level to female rats. *Int J Cancer* 122: 1958-1965.

23. Choi JY, Stover JS, Angel KC, Chowdhury G, Rizzo CJ, et al. (2006) Biochemical

basis of genotoxicity of heterocyclic arylamine food mutagens: Human DNA polymerase eta selectively produces a two-base deletion in copying the N2-guanyl adduct of 2-amino-3-methylimidazo[4,5-f]quinoline but not the C8 adduct at the NarI G3 site. *J Biol Chem* 281: 25297-25306.

24. Singh R, Arlt VM, Henderson CJ, Phillips DH, Farmer PB, et al. (2010) Detection and quantitation of N-(deoxyguanosin-8-yl)-2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine adducts in DNA using online column-switching liquid chromatography tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 878: 2155-2162.

25. Bendaly J, Zhao S, Neale JR, Metry KJ, Doll MA, et al. (2007) 2-Amino-3,8-dimethylimidazo-[4,5-f]quinoxaline-induced DNA adduct formation and mutagenesis in DNA repair-deficient Chinese hamster ovary cells expressing human cytochrome P4501A1 and rapid or slow acetylator N-acetyltransferase 2. *Cancer Epidemiol Biomarkers Prev* 16: 1503-1509.

26. Steenken S (1989) Purine-Bases, Nucleosides, and Nucleotides - Aqueous-Solution Redox Chemistry and Transformation Reactions of Their Radical Cations and E- and Oh Adducts. *Chem Rev* 89: 503-520.

27. Breen AP, Murphy JA (1995) Reactions of oxyl radicals with DNA. *Free Radic Biol Med* 18: 1033-1077.

28. Cheng Q, Gu J, Compaan KR, Schaefer HF, 3rd (2012) Isoguanine formation from adenine. *Chemistry* 18: 4877-4886.

29. Kamiya H (2003) Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: approaches using synthetic oligonucleotides and nucleotides: survey and summary. *Nucleic Acids Res* 31: 517-531.

30. Dizdaroglu M, Bergtold DS (1986) Characterization of free radical-induced base damage in DNA at biologically relevant levels. *Anal Biochem* 156: 182-188.

31. D'Ham C, Romieu A, Jaquinod M, Gasparutto D, Cadet J (1999) Excision of 5,6-dihydroxy-5,6-dihydrothymine, 5,6-dihydrothymine, and 5-hydroxycytosine from defined sequence oligonucleotides by Escherichia coli endonuclease III and Fpg proteins: kinetic and mechanistic aspects. *Biochemistry* 38: 3335-3344.

32. Cao H, Wang Y (2006) Collisionally activated dissociation of protonated 2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives. *J Am Soc*

Mass Spectrom 17: 1335-1341.

33. Hua YS, Wainhaus SB, Yang YN, Shen LX, Xiong YS, et al. (2001) Comparison of negative and positive ion electrospray tandem mass spectrometry for the liquid chromatography tandem mass spectrometry analysis of oxidized deoxynucleosides. *J Am Soc Mass Spectrom* 12: 80-87.

34. Cao H, Wang Y (2007) Quantification of oxidative single-base and intrastrand cross-link lesions in unmethylated and CpG-methylated DNA induced by Fenton-type reagents. *Nucleic Acids Res* 35: 4833-4844.

35. Brierley DJ, Martin SA (2013) Oxidative stress and the DNA mismatch repair pathway. *Antioxid Redox Signal* 18: 2420-2428.

36. Madugundu GS, Cadet J, Wagner JR (2014) Hydroxyl-radical-induced oxidation of 5-methylcytosine in isolated and cellular DNA. *Nucleic Acids Res* 42: 7450-7460.

37. Zhang Q, Wang Y (2003) Independent generation of 5-(2'-deoxycytidinyl)methyl radical and the formation of a novel crosslink lesion between 5-methylcytosine and guanine. *J Am Chem Soc* 125: 12795-12802.

38. Romieu A, Bellon S, Gasparutto D, Cadet J (2000) Synthesis and UV photolysis of oligodeoxynucleotides that contain 5- (phenylthiomethyl)-2'-deoxyuridine: a specific photolabile precursor of 5-(2'-deoxyuridilyl)methyl radical. *Org Lett* 2: 1085-1088.

39. He YF, Li BZ, Li Z, Liu P, Wang Y, et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333: 1303-1307.

40. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333: 1300-1303.

41. Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 11: 607-620.

42. Rechkunova NI, Krasikova YS, Lavrik OI (2011) Nucleotide excision repair: DNA damage recognition and preincision complex assembly. *Biochemistry (Mosc)* 76: 24-35.

43. Jamieson ER, Lippard SJ (1999) Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chem Rev* 99: 2467-2498.

44. Niedernhofer LJ, Daniels JS, Rouzer CA, Greene RE, Marnett LJ (2003)

Malondialdehyde, a product of lipid peroxidation, is mutagenic in human cells. *J Biol Chem* 278: 31426-31433.

45. Brooks PJ (2008) The 8,5'-cyclopurine-2'-deoxynucleosides: candidate neurodegenerative DNA lesions in xeroderma pigmentosum, and unique probes of transcription and nucleotide excision repair. *DNA Repair (Amst)* 7: 1168-1179.

46. Wood RD (2010) Mammalian nucleotide excision repair proteins and interstrand crosslink repair. *Environ Mol Mutagen* 51: 520-526.

47. Rahn JJ, Adair GM, Nairn RS (2010) Multiple roles of ERCC1-XPF in mammalian interstrand crosslink repair. *Environ Mol Mutagen* 51: 567-581.

48. Evans E, Moggs JG, Hwang JR, Egly JM, Wood RD (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J* 16: 6559-6573.

49. Kraemer KH, Lee MM, Scotto J (1987) Xeroderma pigmentosum. Cutaneous, ocular, and neurologic abnormalities in 830 published cases. *Arch Dermatol* 123: 241-250.

50. Iyama T, Wilson DM, 3rd (2013) DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)* 12: 620-636.

51. Barnes DE, Lindahl T (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu Rev Genet* 38: 445-476.

52. Helleday T, Lo J, van Gent DC, Engelward BP (2007) DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair (Amst)* 6: 923-935.

53. Lamarche BJ, Orazio NI, Weitzman MD (2010) The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett* 584: 3682-3695.

54. Yuan SSF, Lee SY, Chen G, Song MH, Tomlinson GE, et al. (1999) BRCA2 is required for ionizing radiation-induced assembly of rad51 complex in vivo. *Cancer Res* 59: 3547-3551.

55. Sameer AS, Nissar S, Fatima K (2014) Mismatch repair pathway: molecules, functions, and role in colorectal carcinogenesis. *Eur J Cancer Prev* 23: 246-257.

56. Grundy GJ, Moulding HA, Caldecott KW, Rulten SL (2014) One ring to bring them

all--the role of Ku in mammalian non-homologous end joining. *DNA Repair (Amst)* 17: 30-38.

57. Radhakrishnan SK, Jette N, Lees-Miller SP (2014) Non-homologous end joining: emerging themes and unanswered questions. *DNA Repair (Amst)* 17: 2-8.

58. Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33: 245-254.

59. Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293: 1074-1080.

60. Ooi SK, Bestor TH (2008) The colorful history of active DNA demethylation. *Cell* 133: 1145-1148.

61. Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 11: 607-620.

62. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466: 1129-1133.

63. Yuan B, Cao H, Jiang Y, Hong H, Wang Y (2008) Efficient and accurate bypass of N2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase in vitro and in vivo. *Proc Natl Acad Sci U S A* 105: 8679-8684.

64. Liu S, Wang J, Su Y, Guerrero C, Zeng Y, et al. (2013) Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res* 41: 6421-6429.

65. Tilstra JS, Robinson AR, Wang J, Gregg SQ, Clauson CL, et al. (2012) NF-kappaB inhibition delays DNA damage-induced senescence and aging in mice. *J Clin Invest* 122: 2601-2612.

66. Wang J, Cao H, You C, Yuan B, Bahde R, et al. (2012) Endogenous formation and repair of oxidatively induced G[8-5 m]T intrastrand cross-link lesion. *Nucleic Acids Res* 40: 7368-7374.

67. Wang J, Clauson CL, Robbins PD, Niedernhofer LJ, Wang Y (2012) The oxidative DNA lesions 8,5'-cyclopurines accumulate with aging in a tissue-specific manner. *Aging Cell* 11: 714-716.

68. Wang J, Yuan B, Guerrero C, Bahde R, Gupta S, et al. (2011) Quantification of

oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal Chem* 83: 2201-2209.

69. Monien BH, Schumacher F, Herrmann K, Glatt H, Turesky RJ, et al. (2014) Simultaneous detection of multiple DNA adducts in human lung samples by isotope-dilution UPLC-MS/MS. *Anal Chem* 87: 641-648.

70. Yun BH, Rosenquist TA, Nikolic J, Dragicevic D, Tomic K, et al. (2013) Human formalin-fixed paraffin-embedded tissues: an untapped specimen for biomonitoring of carcinogen DNA adducts by mass spectrometry. *Anal Chem* 85: 4251-4258.

71. Chen HJ, Lin WP (2011) Quantitative analysis of multiple exocyclic DNA adducts in human salivary DNA by stable isotope dilution nanoflow liquid chromatography-nanospray ionization tandem mass spectrometry. *Anal Chem* 83: 8543-8551.

72. Chung YT, Chen CL, Wu CC, Chan SA, Chi CW, et al. (2008) Safrole-DNA adduct in hepatocellular carcinoma associated with betel quid chewing. *Toxicol Lett* 183: 21-27.

73. Lao Y, Villalta PW, Sturla SJ, Wang M, Hecht SS (2006) Quantitation of pyridyloxobutyl DNA adducts of tobacco-specific nitrosamines in rat tissue DNA by high-performance liquid chromatography-electrospray ionization-tandem mass spectrometry. *Chem Res Toxicol* 19: 674-682.

74. Doerge DR, Churchwell MI, Fang JL, Beland FA (2000) Quantification of etheno-DNA adducts using liquid chromatography, on-line sample processing, and electrospray tandem mass spectrometry. *Chem Res Toxicol* 13: 1259-1264.

75. Delatour T, Mally A, Richoz J, Ozden S, Dekant W, et al. (2008) Absence of 2'-deoxyguanosine-carbon 8-bound ochratoxin A adduct in rat kidney DNA monitored by isotope dilution LC-MS/MS. *Mol Nutr Food Res* 52: 472-482.

76. Paini A, Punt A, Viton F, Scholz G, Delatour T, et al. (2010) A physiologically based biodynamic (PBBD) model for estragole DNA binding in rat liver based on in vitro kinetic data and estragole DNA adduct formation in primary hepatocytes. *Toxicol Appl Pharmacol* 245: 57-66.

77. Pan SS, Iracki T (1988) Metabolites and DNA adduct formation from flavoenzyme-activated porfiromycin. *Mol Pharmacol* 34: 223-228.

78. Marsch GA, Mundkowski RG, Morris BJ, Manier ML, Hartman MK, et al. (2001) Characterization of nucleoside and DNA adducts formed by S-(1-acetoxymethyl)glutathione and implications for dihalomethane-glutathione conjugates. *Chem Res Toxicol* 14: 600-608.

79. Frandsen H, Grivas S, Turesky RJ, Andersson R, Dragsted LO, et al. (1994) Formation of DNA adducts by the food mutagen 2-amino-3,4,8-trimethyl-3H-imidazo[4,5-f]quinoxaline (4,8-DiMeIQx) in vitro and in vivo. Identification of a N2-(2'-deoxyguanosin-8-yl)-4,8-DiMeIQx adduct. *Carcinogenesis* 15: 2553-2558.

80. Wink DA, Kasprzak KS, Maragos CM, Elespuru RK, Misra M, et al. (1991) DNA deaminating ability and genotoxicity of nitric oxide and its progenitors. *Science* 254: 1001-1003.

81. Van der Weyden MB, Kelley WN (1976) Human adenosine deaminase. Distribution and properties. *J Biol Chem* 251: 5448-5456.

82. Dong M, Wang C, Deen WM, Dedon PC (2003) Absence of 2'-deoxyoxanosine and presence of abasic sites in DNA exposed to nitric oxide at controlled physiological concentrations. *Chem Res Toxicol* 16: 1044-1055.

83. Wang Y, Taylor JS, Gross ML (2001) Isolation and mass spectrometric characterization of dimeric adenine photoproducts in oligodeoxynucleotides. *Chem Res Toxicol* 14: 738-745.

84. Wang Y, Taylor JS, Gross ML (1999) Nuclease P1 digestion combined with tandem mass spectrometry for the structure determination of DNA photoproducts. *Chem Res Toxicol* 12: 1077-1082.

85. Wang Y (2002) HPLC isolation and mass spectrometric characterization of two isomers of thymine glycols in oligodeoxynucleotides. *Chem Res Toxicol* 15: 671-676.

86. Wang Y, Wang Y (2003) Structure elucidation of DNA interstrand cross-link by a combination of nuclease P1 digestion with mass spectrometry. *Anal Chem* 75: 6306-6313.

87. Lai C, Cao H, Hearst JE, Corash L, Luo H, et al. (2008) Quantitative analysis of DNA interstrand cross-links and monoadducts formed in human cells induced by psoralens and UVA irradiation. *Anal Chem* 80: 8790-8798.

88. Cao H, Hearst JE, Corash L, Wang Y (2008) LC-MS/MS for the detection of DNA interstrand cross-links formed by 8-methoxypsoralen and UVA irradiation in human cells. *Anal Chem* 80: 2932-2938.

89. Wang Y (2003) Structure elucidation of DNA interstrand cross-link by a combination of nuclease P1 digestion with mass spectrometry. *Anal Chem* 75: 6306-6313.

90. Liu S, Wang Y (2013) A quantitative mass spectrometry-based approach for assessing the repair of 8-methoxypsoralen-induced DNA interstrand cross-links and monoadducts in mammalian cells. *Anal Chem* 85: 6732-6739.

91. Vanhoutte K, Van Dongen W, Hoes I, Lemiere F, Esmans EL, et al. (1997) Development of a nanoscale liquid chromatography/electrospray mass spectrometry methodology for the detection and identification of DNA adducts. *Anal Chem* 69: 3161-3168.

92. Tang H, Fang H, Yin E, Brasier AR, Sowers LC, et al. (2014) Multiplexed parallel reaction monitoring targeting histone modifications on the QExactive mass spectrometer. *Anal Chem* 86: 5526-5534.

93. Fu L, Amato NJ, Wang P, McGowan SJ, Niedernhofer LJ, et al. (2015) Simultaneous quantification of methylated cytidine and adenosine in cellular and tissue RNA by nano-flow liquid chromatography-tandem mass spectrometry coupled with the stable isotope-dilution method. *Anal Chem* in press.

94. Kao CY, Giese RW (2005) Measurement of N7-(2'-hydroxyethyl)guanine in human DNA by gas chromatography electron capture mass spectrometry. *Chem Res Toxicol* 18: 70-75.

95. Tompkins EM, Jones DJ, Lamb JH, Marsden DA, Farmer PB, et al. (2008) Simultaneous detection of five different 2-hydroxyethyl-DNA adducts formed by ethylene oxide exposure, using a high-performance liquid chromatography/electrospray ionisation tandem mass spectrometry assay. *Rapid Commun Mass Spectrom* 22: 19-28.

96. Baskerville-Abraham IM, Boysen G, Troutman JM, Mutlu E, Collins L, et al. (2009) Development of an ultraperformance liquid chromatography/mass spectrometry method to quantify cisplatin 1,2 intrastrand guanine-guanine adducts. *Chem Res Toxicol* 22: 905-912.

97. Tretyakova N, Goggin M, Sangaraju D, Janis G (2012) Quantitation of DNA adducts by stable isotope dilution mass spectrometry. *Chem Res Toxicol* 25: 2007-2035.

98. Inagaki S, Esaka Y, Goto M, Deyashiki Y, Sako M (2004) LC-MS study on the formation of cyclic 1,N2-propano guanine adduct in the reactions of DNA with acetaldehyde in the presence of histone. *Biol Pharm Bull* 27: 273-276.

99. Douki T, Odin F, Caillat S, Favier A, Cadet J (2004) Predominance of the 1,N2-propano 2'-deoxyguanosine adduct among 4-hydroxy-2-nonenal-induced DNA lesions. *Free Radic Biol Med* 37: 62-70.

100. Wang M, Yu N, Chen L, Villalta PW, Hochalter JB, et al. (2006) Identification of an acetaldehyde adduct in human liver DNA and quantitation as N2-ethyldeoxyguanosine. *Chem Res Toxicol* 19: 319-324.

101. Lao Y, Villalta PW, Sturla SJ, Wang M, Hecht SS (2006) Quantitation of pyridyloxobutyl DNA adducts of tobacco-specific nitrosamines in rat tissue DNA by high-performance liquid chromatography-electrospray ionization-tandem mass spectrometry. *Chem Res Toxicol* 19: 674-682.

102. Stout MD, Jeong YC, Boysen G, Li Y, Sangaiah R, et al. (2006) LC/MS/MS method for the quantitation of trans-2-hexenal-derived exocyclic 1,N(2)-propanodeoxyguanosine in DNA. *Chem Res Toxicol* 19: 563-570.

103. Balbo S, Villalta PW, Hecht SS (2011) Quantitation of 7-ethylguanine in leukocyte DNA from smokers and nonsmokers by liquid chromatography-nanoelectrospray-high resolution tandem mass spectrometry. *Chem Res Toxicol* 24: 1729-1734.

104. Emmett MR, Caprioli RM (1994) Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins. *J Am Soc Mass Spectrom* 5: 605-613.

105. Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68: 850-858.

106. Ricicki EM, Soglia JR, Teitel C, Kane R, Kadlubar F, et al. (2005) Detection and quantification of N-(deoxyguanosin-8-yl)-4-aminobiphenyl adducts in human pancreas tissue using capillary liquid chromatography-microelectrospray mass spectrometry. *Chem Res Toxicol* 18: 692-699.

107. Ricicki EM, Luo W, Fan WH, Zhao LP, Zarbl H, et al. (2006) Quantification of N-(deoxyguanosin-8-yl)-4-aminobiphenyl adducts in human lymphoblastoid TK6 cells dosed with N-hydroxy-4-acetylaminobiphenyl and their relationship to mutation, toxicity, and gene expression profiling. *Anal Chem* 78: 6422-6432.

108. Randall KL, Argoti D, Paonessa JD, Ding Y, Oaks Z, et al. (2010) An improved liquid chromatography-tandem mass spectrometry method for the quantification of 4-aminobiphenyl DNA adducts in urinary bladder cells and tissues. *J Chromatogr A* 1217: 4135-4143.

109. Bessette EE, Goodenough AK, Langouet S, Yasa I, Kozekov ID, et al. (2009) Screening for DNA adducts by data-dependent constant neutral loss-triple stage mass spectrometry with a linear quadrupole ion trap mass spectrometer. *Anal Chem* 81: 809-819.

110. Hong H, Wang Y (2007) Derivatization with Girard reagent T combined with LC-MS/MS for the sensitive detection of 5-formyl-2'-deoxyuridine in cellular DNA. *Anal Chem* 79: 322-326.

111. Chen HJC, Chiang LC, Tseng MC, Zhang LL, Ni JS, et al. (1999) Detection and quantification of 1,N-6-ethenoadenine in human placental DNA by mass spectrometry. *Chem Res Toxicol* 12: 1119-1126.

112. Thomson NM, Mijal RS, Ziegel R, Fleischer NL, Pegg AE, et al. (2004) Development of a quantitative liquid chromatography/electrospray mass spectrometric assay for a mutagenic tobacco specific nitrosamine-derived DNA adduct, O6-[4-Oxo-4-(3-pyridyl)butyl]-2'-deoxyguanosine. *Chem Res Toxicol* 17: 1600-1606.

113. Koehl P, Valladier P, Lefevre JF, Fuchs RP (1989) Strong structural effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Nucleic Acids Res* 17: 9531-9541.

114. Mcluckey SA, Vanberkel GJ, Glish GL (1992) Tandem Mass-Spectrometry of Small, Multiply Charged Oligonucleotides. *J Am Soc Mass Spectrom* 3: 60-70.

115. Andersen N, Wang P, Wang Y (2013) Replication across regioisomeric ethylated thymidine lesions by purified DNA polymerases. *Chem Res Toxicol* 26: 1730-1738.

116. Andersen N, Wang J, Wang P, Jiang Y, Wang Y (2012) In-vitro replication studies on O(2)-methylthymidine and O(4)-methylthymidine. *Chem Res Toxicol* 25: 2523-2531.

117. Gu C, Wang Y (2007) In vitro replication and thermodynamic studies of methylation and oxidation modifications of 6-thioguanine. *Nucleic Acids Res* 35: 3693-3704.

118. Choi JY, Chowdhury G, Zang H, Angel KC, Vu CC, et al. (2006) Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J Biol Chem* 281: 38244-38256.

119. Otteneder MB, Knutson CG, Daniels JS, Hashim M, Crews BC, et al. (2006) In vivo oxidative metabolism of a major peroxidation-derived DNA adduct, M1dG. *Proc Natl Acad Sci U S A* 103: 6665-6669.

120. Chen HJ, Chang CM (2004) Quantification of urinary excretion of 1,N6-ethenoadenine, a potential biomarker of lipid peroxidation, in humans by stable isotope dilution liquid chromatography-electrospray ionization-tandem mass spectrometry: comparison with gas chromatography-mass spectrometry. *Chem Res Toxicol* 17: 963-971.

121. Malayappan B, Johnson L, Nie B, Panchal D, Matter B, et al. (2010) Quantitative High-Performance Liquid Chromatography-Electrospray Ionization Tandem Mass Spectrometry Analysis of Bis-N7-Guanine DNA-DNA Cross-Links in White Blood Cells of Cancer Patients Receiving Cyclophosphamide Therapy. *Anal Chem* 82: 3650-3658.

122. Paz MM, Ladwa S, Champeil E, Liu Y, Rockwell S, et al. (2008) Mapping DNA adducts of mitomycin C and decarbamoyl mitomycin C in cell lines using liquid chromatography/ electrospray tandem mass spectrometry. *Chem Res Toxicol* 21: 2370-2378.

123. Riggs AD, Jones PA (1983) 5-methylcytosine, gene regulation, and cancer. *Adv Cancer Res* 40: 1-30.

124. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, et al. (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* 10: 2709-2721.

125. Cooper DN, Youssoufian H (1988) The CpG dinucleotide and human genetic disease. *Hum Genet* 78: 151-155.

126. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system.

Cell 151: 1417-1430.

127. Yildirim O, Li R, Hung JH, Chen PB, Dong X, et al. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147: 1498-1510.

128. Wang J, Tang J, Lai M, Zhang H (2014) 5-Hydroxymethylcytosine and disease. *Mutat Res Rev Mutat Res* 762: 167-175.

129. Ye C, Li L (2014) 5-hydroxymethylcytosine: a new insight into epigenetics in cancer. *Cancer Biol Ther* 15: 10-15.

130. Le T, Kim KP, Fan G, Faull KF (2011) A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Anal Biochem* 412: 203-209.

131. Chen ML, Shen F, Huang W, Qi JH, Wang Y, et al. (2013) Quantification of 5-methylcytosine and 5-hydroxymethylcytosine in genomic DNA from hepatocellular carcinoma tissues by capillary hydrophilic-interaction liquid chromatography/quadrupole TOF mass spectrometry. *Clin Chem* 59: 824-832.

132. Godderis L, Schouteden C, Tabish A, Poels K, Hoet P, et al. (2015) Global Methylation and Hydroxymethylation in DNA from Blood and Saliva in Healthy Volunteers. *Biomed Res Int* 2015: 845041.

133. Ji D, Lin K, Song J, Wang Y (2014) Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol Biosyst* 10: 1749-1752.

134. Ji D, You C, Wang P, Wang Y (2014) Effects of tet-induced oxidation products of 5-methylcytosine on DNA replication in mammalian cells. *Chem Res Toxicol* 27: 1304-1309.

Chapter 2: Quantitative Mass Spectrometry-Based Approach for Assessing the Repair of 8-Methoxypsoralen-Induced DNA Interstrand Crosslink and Monoadducts in Mammalian Cells

Introduction

DNA in eukaryotic cells is constantly damaged through exposure to a variety of reactive chemicals derived from external and internal sources (1,2). It was estimated that tens of thousands of DNA lesions can be induced in each of the ~10^{13} cells in the human body per day (1,3). Because damage to DNA may be incompatible with its essential roles in preservation and transmission of genetic information, multiple DNA repair pathways have evolved in organisms to maintain genomic stability and cell viability. These include, but are not limited to, nucleotide excision repair (NER), mismatch repair (MMR), base excision repair (BER) and recombinational repair (2). If left unrepaired, DNA damage may perturb the flow of genetic information by inhibiting DNA replication and transcription and by inducing mutations during these processes, which may ultimately lead to the development of cancer, neurodegeneration, or other human diseases (2).

DNA interstrand crosslink (ICL) is among the most toxic type of DNA lesions (4). It has been estimated that a single ICL can kill a repair-deficient bacterial or yeast cell,

and as few as 20-40 ICLs can be lethal to a mammalian cell lacking the ability to remove the crosslink (5,6). Because of the presence of covalent linkages between the opposing strands of DNA, ICLs may completely block replication and transcription, and consequently pose a formidable challenge to cell survival and to the maintenance of genetic information (7,8). On the other hand, it is also this cytotoxic effect that constitutes the mechanistic basis of action for many currently used cancer chemotherapeutic agents (9).

Psoralen and its derivatives, a group of furocoumarins naturally found in leafy vegetables and other plants, are well-known bifunctional photoreactive DNA crosslinking agents. Upon long-wavelength UV irradiation, psoralen reacts preferentially with thymine residue at 5'-TpA site to form monoadducts (MAs) or interstrand crosslink via [2+2] cycloaddition (10). Psoralen plus UVA (PUVA) is a widely used and effective therapy for several hyper-proliferative skin disorders such as psoriasis and vitiligo (11) and has been used for the treatment of cutaneous T-cell lymphoma (12). The effectiveness of PUVA treatment is largely attributed to the formation of DNA ICLs. However, the benefit of PUVA treatment is accompanied with the risk of developing non-melanoma skin cancer (13). Thus, a better understanding of the repair of psoralen-photoactivated ICL will provide important knowledge for developing improved PUVA therapies for the treatment of cancer and other human

diseases.

Owing to the different properties of ICLs and depending on the cell cycle phase, ICLs are most likely processed through multiple repair mechanisms *in vivo* (9). A conceptual model consisting of a two-cycle scheme has been proposed for ICL repair. The first cycle is thought to begin with the cleavage of one strand on both the 5' and 3' sides of the damaged base by NER enzymes, resulting in the generation of a gapped intermediate (14). Gap filling is executed by translesion synthesis in non-dividing cells or by recombination repair in actively dividing cells using the undamaged homologous chromosome as the template (15). The gap-filling process converts an ICL into an MA, after which the second cycle of excision repair removes the MA presumably through the conventional NER pathway (5). The above scheme has delineated a general mechanism for ICL repair in *E. coli* and yeast. However, our understanding about the repair of ICLs in mammalian cells is far from complete.

Assessing the repair of DNA ICLs often necessitates the measurement of the levels of these lesions in cells and tissues. Several biophysical and biochemical techniques have been used for measuring ICLs, which include agarose gel-based method (16), alkaline filter elution (17), alkaline comet assay (18), and chromatographic techniques (19). A drawback of these methods is their relatively low sensitivity or a requirement for a high level of ICL, which is lethal to almost all cells

and may saturate cellular DNA repair capacity. Additionally, these methods do not provide structural information about the ICLs. With the development and application of mass spectrometry techniques, the detection sensitivity of DNA damage analysis has been greatly improved. By using HPLC coupled with tandem mass spectrometry (LC-MS/MS), Courdavault et al. (20) studied the repair of three types of bipyrimidine photoproducts in keratinocytes, where they observed a rapid elimination of TT-(6-4) photoproduct from cellular DNA 72 hr post-irradiation. With the use of inductively coupled plasma mass spectrometry (ICP-MS), Harrington et al. (21) discovered a faster removal of cisplatin 1,2-intrastrand guanine-guanine adducts (1,2-GG) in cisplatin-resistant human lung tumor cells relative to drug-sensitive cells. LC-MS/MS/MS with the isotope-dilution technique has also been employed to examine the role of NER in repairing endogenously induced intrastrand crosslink d(G[8-5m]T), 8,5'-cyclo-2'-deoxyadenosine, and 8,5'-cyclo-2'-deoxyguanosine in mammalian tissues (22,23).

We recently reported the application of LC-MS/MS in assessing the levels of ICLs and MAs formed in human cells upon exposure to different psoralen derivatives (24,25). LC-MS/MS, however, has been rarely employed for examining the repair of DNA ICLs. In this vein, considering the high cytotoxicity of DNA ICLs, it is potentially challenging to identify an experimental condition that can allow for the

generation of adequate level of DNA ICLs for LC-MS/MS detection without compromising the repair capacity or survival of the host cells. Malayappan et al. (26) used LC-MS/MS to follow the formation and repair of cyclophosphamide-induced *N*,*N*-bis[2-(*N*7-guanyl)ethyl]amine crosslink in human blood, but intrastrand and interstrand crosslinks were not distinguished in that study because both types of lesions were released as free base adducts after neutral thermal hydrolysis. It was also reported that ICLs introduced by mitomycin C could be partially removed in mouse mammary tumor cells (27). However, a crosslinked dinucleoside standard was used for the quantitative measurement. Given that the release of mitomycin C interstrand crosslink from genomic DNA may not be complete, the accuracy for the quantification remains uncertain.

Herein, we demonstrated the capability of LC-MS/MS in assessing directly the repair of 8-MOP-induced DNA ICL, as well as MAs, in mammalian cells. Moreover, our results provided unequivocal evidence to support that NER functions in repairing the 8-MOP-induced ICL and MAs in mammalian cells.

Materials and Methods

Materials and Cell Lines

All chemicals and enzymes, unless otherwise specified, were from Sigma-Aldrich

(St. Louis, MO) and New England Biolabs (Ipswich, WA). All oligodeoxyribonucleotides (ODNs) used in this study were purchased from Integrated DNA Technologies (Coralville, IA). [8-D_3]-8-methoxypsoralen (D_3-8-MOP) was synthesized previously (24).

Human embryonic kidney 293T cells (HEK293T) were obtained from the American Type Culture Collection (Manassas, VA). Human skin fibroblasts (HFs) that are repair-competent (GM00637) or deficient in xeroderma pigmentosum complementation group A (XPA, GM04429) were provided by Prof. Gerd P. Pfeifer (City of Hope). Repair-proficient AA8 Chinese hamster ovary (CHO) cells and the isogenic CHO cells that are deficient in excision repair cross-complementing rodent repair deficiency, complementation group 1 (ERCC1, CHO-7-27) were provided by Prof. Michael M. Seidman from National Institute of Aging (28).

Preparation of Standard ICL- and MA-Containing ODNs

ODNs containing an ICL or MA introduced by 8-methoxypsoralen (8-MOP) or D_3-8-MOP were prepared following the previously reported procedures with modifications (24). Structures of unlabeled and labeled 8-MOP are depicted in Scheme 2.1a. A 10-nmol self-complementary ODN d(CGCGCTAGCGCG) was annealed in a 40-μL reaction buffer that contained 5 mM Tris (pH 7.6), 50 mM NaCl and 0.2 mM EDTA. The ODN solution was diluted to 1 mL with the reaction buffer, and a solution

of 8-MOP or D_3-8-MOP in ethanol was added until its final concentration reached 1 mM. The reaction mixture was subsequently dispersed in a 3.5-cm-i.d. Petri dish and incubated in the dark for 1 hr. The resulting solution was irradiated on ice for 45 min with UVA light, which was delivered by two 15-W Spectroline light tubes with emitting wavelength centered at 365 nm (Spectronics Corp., Westbury, NY). The irradiation mixture, which contained both ICLs and unreacted starting material, was immediately irradiated on ice with short-wavelength UV light centered at 254 nm (Spectronics Corp.) for 15 sec to facilitate the photo-reversal of ICL-containing ODNs (25). After irradiation, the solution was extracted with chloroform, and the ODNs were recovered by ethanol precipitation, separated by HPLC, and used as standards for LC-MS/MS analysis.

HPLC

The purification of ICL- and MA-containing ODNs was performed on an Agilent 1100 HPLC system with a UV detector monitoring at 260 nm. A 4.6×150 mm Aeris WIDEPORE column (3.6 μm in particle size and 200 Å in pore size, Torrance, CA) was used, and the column was heated to 50 ℃ by using an SP8790 column heater (Spectra Physics, San Jose, CA). A solution of 50 mM triethylammonium acetate (TEAA, pH 6.8) and a solution of 30% (v/v) acetonitrile in 50 mM TEAA were used as mobile phases A and B, respectively. A gradient of 5 min 0-25% B and 40 min 25-50%

B was employed, and the flow rate was 0.8 mL/min. The purities of the ICL- and MA-containing ODNs were verified by LC-MS/MS analysis.

Cell Culture and Drug Treatment

Cells were routinely cultured at 37℃ in 5% CO_2 atmosphere. Human skin fibroblasts and HEK293T cells were grown in Dulbecco's Modified Eagle Medium, and CHO cells were grown in Alpha Minimum Essential Medium without ribonucleosides or deoxyribonucleosides. All culture media were supplemented with 10% fetal bovine serum and 100 IU/ml penicillin. Cells were seeded in 6-well plates at a density of 250,000 cells per well in complete medium, and 24 hr later, a stock solution of 8-MOP was added to the culture medium to render its final concentration 2.5 μM. The mixture was incubated at 37℃ in dark for 30 min, and subsequently irradiated with UVA light at a dose of 2 J/cm^2, which was measured by a Mannix UV-340 light meter (Mannix Instrument Inc., New York, NY). The cells were then cultured at 37℃ in complete medium for different time intervals to allow for lesion repair. Cells were then harvested and washed with PBS. Trypan blue exclusion assay showed that cell viability was greater than 90% at the end of the entire experiment. It is worth noting that 8-MOP could be rapidly absorbed after oral administration or penetrates into the epidermis and dermis following topical (cream or bath) application. The peak concentrations of 8-MOP in patient plasma and dermis are in the range of 45-970

ng/mL (29-31), and the UVA dose employed for PUVA treatment is typically within the range of 1-10 J/cm^2 (32,33). Thus, the concentration of 8-MOP (2.5 μM, or 540 ng/mL) and the dose of UVA irradiation (2 J/cm^2) used in this study were pharmacologically relevant.

DNA Extraction and Enzymatic Digestion

Genomic DNA was isolated from the resulting cell lysate using a high-salt method (34) and desalted by ethanol precipitation. The isolated DNA (10 μg in 100 μL ddH_2O) was subsequently mixed with 0.5 pmol of authentic D_3-8-MOP-ICL-carrying duplex ODN and 0.2 pmol of D_3-8-MOP-MA-containing ODN. To the mixture were added nuclease P1 (0.5 unit) and a 5-μL solution containing 300 mM sodium acetate (pH 5.6) and 10 mM zinc acetate. Nuclease P1 has combined endo- and exonuclease activities that can hydrolyze nucleic acid to 5'-mononucleotides. To prevent the potential loss of 5' phosphate in the resulting nucleotides, a 5-μL phosphate buffer (10 mM, pH 5.7) was also added to the sample prior to the digestion. The digestion mixture was incubated at 37 ℃ for 1 hr and extracted twice with chloroform to remove the enzyme. The aqueous layer was dried in a Speed-Vac concentrator, reconstituted in 20 μL H_2O, and a 5-15 μL aliquot was injected for LC-MS/MS analysis.

LC-MS/MS Analysis

The LC-MS/MS experiments were conducted on an LTQ linear ion-trap mass

spectrometer (Thermo, San Jose, CA), and a 0.5×250 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies, Santa Clara, CA) was employed, as described previously (24,25). The flow rate was 8.0 $\mu L \cdot min^{-1}$, and a solution of 400 mM 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, pH was adjusted to 7.0 by addition of triethylamine, solution A) and methanol (solution B) were used as mobile phases for the separation of nucleotide mixtures arising from nuclease P1 digestion. A gradient of 5 min at 0% B, 5 min of 0-20% B, and 40 min of 20-50% B was employed for the separation. The mass spectrometer was set up to acquire the tandem mass spectra (MS/MS) for the fragmentation of the $[M-2H]^{2-}$ ions of tetranucleotides harboring ICL induced by 8-MOP or D_3-8-MOP and the $[M-H]^-$ ions of dinucleotides containing MAs formed from 8-MOP or D_3-8-MOP.

Calibration curves for the quantification of 8-MOP-ICL and 8-MOP-MAs were constructed by LC-MS/MS analysis of the nuclease P1 digestion products of standard ODN mixtures, which contained known amounts of the unlabeled 8-MOP-ICL- and -MA1-housing 12mer ODNs, together with 0.5 pmol D_3-8-MOP-ICL-bearing 12mer duplex ODN and 0.2 pmol 8-MOP-MA1-bearing single-stranded ODN. The LC-MS/MS experiments were conducted under identical conditions as described above for the nuclease P1-produced nucleotide mixture of cellular DNA.

Results and Discussion

Preparation of Standard ICL- and MA-containing ODNs

To examine the repair of 8-MOP-induced lesions in cells by LC-MS/MS, we first prepared ODNs bearing an ICL or MA induced by 8-MOP or D_3-8-MOP (Scheme 2.1a), the latter of which was synthesized previously (24). After UVA exposure and photoreversal, we isolated the ICL-carrying duplex ODN and MA-bearing single-stranded ODNs from the mixture by HPLC. The 32.9-min fraction shown in Figure A.1 was identified as 8-MOP-ICL in d(CGCGCTAGCGCG), and the fractions eluting at 27.5, 29.1, and 29.7 min were found to be the ODN carrying isomeric 8-MOP-MAs. The identities of the ICL- and MA-containing ODNs were further confirmed via mass spectrometric characterization after nuclease P1 digestion, and the amounts of the ODNs were quantified by UV absorption measurements at 260 nm.

Unlike a previously described method where mitomycin C crosslink-containing dinucleoside was employed as standard (27), we chose to use stable isotope-labeled, lesion-carrying ODNs as standards because the standard ODNs are digested together with DNA isolated from 8-MOP-treated cells. Therefore, the assay allows for more accurate quantification of 8-MOP-ICL and -MAs in cellular DNA by accounting for potential incomplete release of ICL and MAs from DNA during the enzymatic

digestion and potential loss of target analytes during other stages of sample preparation process.

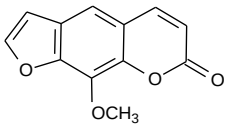
Identification and Quantification of DNA ICL and MAs Induced by 8-MOP/UVA in Mammalian Cells

Previous studies revealed that ICL and MAs induced by different psoralen derivatives prevent the nuclease P1-mediated hydrolysis of the phosphodiester linkage between the crosslinked dT and its neighboring 3' dA (24,35). Thus, nuclease P1 digestion facilitates the release of 8-MOP-induced ICL and MA from DNA as lesion-bearing tetranucleotide and dinucleotide, respectively (Scheme 2.1b). Indeed, negative-ion ESI-MS of the nuclease P1 digestion products of 8-MOP- and D_3-8-MOP-ICL-containing ODNs showed the ions of *m/z* 742 and 743.5, corresponding to the $[M-2H]^{2-}$ ions of 8-MOP- and D_3-8-MOP-ICL-bearing tetranucleotide. Likewise, the MS of digestion mixtures of ODNs carrying both 8-MOP- and D_3-8-MOP-MA revealed the *m/z* 850 and 853 ions for the $[M-H]^-$ ions of 8-MOP- and D_3-8-MOP-MA-containing dinucleotide (data not shown).

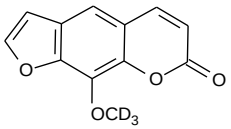
By employing MS/MS, along with the use of authentic ODNs carrying the D_3-8-MOP lesions, we demonstrated that 8-MOP-induced ICL and MAs are readily detectable in human skin fibroblasts and CHO cells treated with 8-MOP and UVA light (representative LC-MS/MS results are shown in Figures 2.1 and 2.2, and the

Scheme 2.1. (a) The chemical structures of 8-MOP and D_3-8-MOP; (b) Experimental outline of the enzymatic digestion of 8-MOP-ICL and 8-MOP-MAs in ODNs or cellular DNA; (c) The chemical structures of the tetranucleotide and dinucleotide generated from nuclease P1 digestion and the cleavages for the formation of major fragment ions observed in MS/MS.

(a)

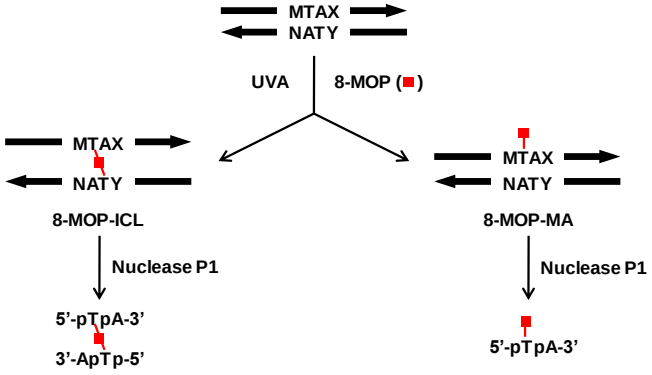


8-MOP

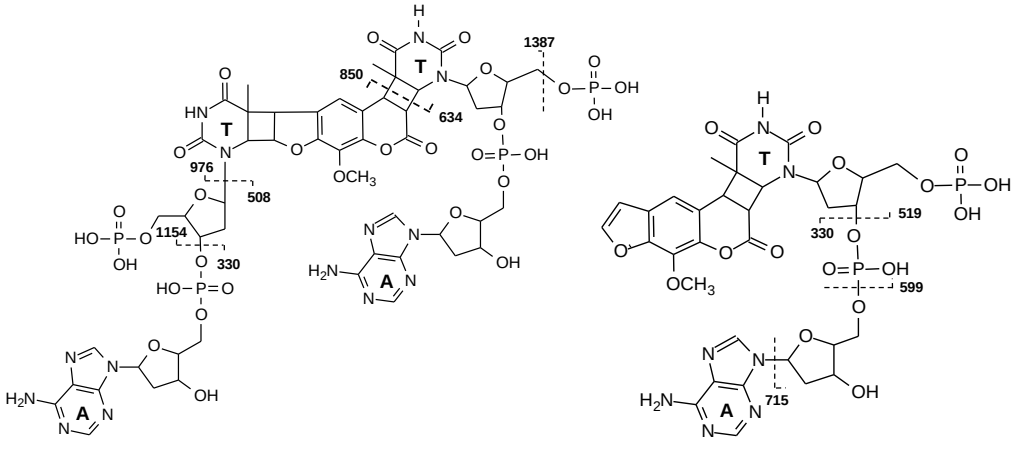


D_3-8-MOP

(b)



(c)



8-MOP-Tetranucleotide **8-MOP-Dinucleotide**

assignments for major fragment ions are depicted in Scheme 2.1c). Our LC-MS/MS method also allowed for the accurate quantification of 8-MOP-induced ICL and MAs in mammalian cells. In this respect, three peaks at 25.4, 28.2 and 30.5 min were observed in the SIC for the *m/z* 850 → 519 transition (Figure 2.2a), and according to their elution orders, these three MAs were designated as 8-MOP-MA1, -MA2 and -MA3, respectively. We assumed that the dinucleotides carrying the three isomeric MAs share the same ionization efficiency. This should be a reasonable assumption in the viewpoint that the deprotonated ions were monitored and the three MAs are structural isomers. Thus, we constructed only one calibration curve by analyzing the nuclease P1 digestion products of ODN mixtures containing D_3-8-MOP MA1 and its unlabeled counterpart at different molar ratios, and used this curve to quantify all three 8-MOP MAs formed in cells (Figure A.2)

The limits of quantitation (LOQ) and limits of detection (LOD), defined as the amounts of analytes giving rise to signal-to-noise ratios of 10 and 3, were 33.3 fmol and 10.9 fmol for 8-MOP-ICL and 8-MOP-MA, respectively, and 10 fmol and 3.3 fmol for the ICL and MA, respectively. Although immunoassays can offer a sensitivity of detecting 17 fmol MAs in cells treated with 8-MOP (36), the LC-MS/MS method that we developed here can facilitate the simultaneous quantification of 8-MOP ICL and

MAs with the use of a relatively small amount (5 μg) of DNA. This level of sensitivity renders the method useful for assessing the repair of DNA adducts in mammalian cells.

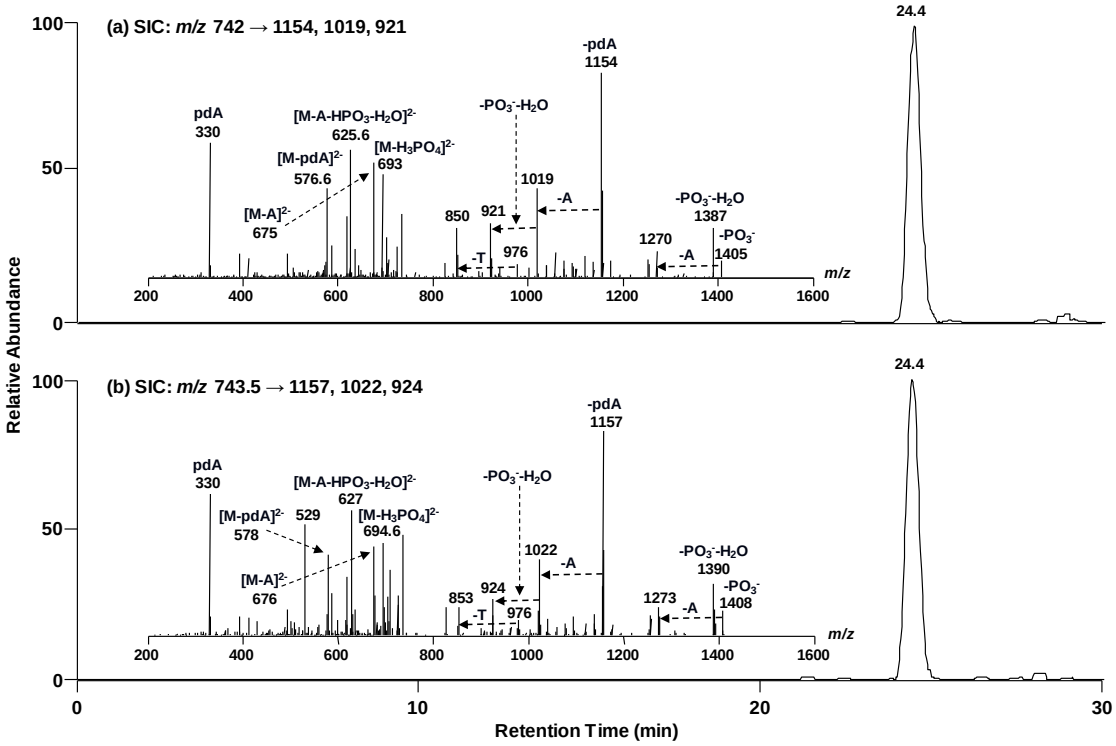


Figure 2.1. Selected-ion chromatograms (SICs) for monitoring the *m/z* 742.0→1154, 1019, 921 transitions for 8-MOP-ICL-containing tetranucleotide (a) and the *m/z* 743.5→1157, 1022, 924 transitions for D_3-8-MOP-ICL-containing tetranucleotide (b) resulting from nuclease P1 digestion of DNA extracted from wild-type CHO cells immediately after photoirradiation (i.e., at 0 hr). Shown in the insets are the product-ion spectra of the ESI-produced $[M-2H]^{2-}$ ions of the tetranucleotide housing unlabeled and labeled 8-MOP-ICL.

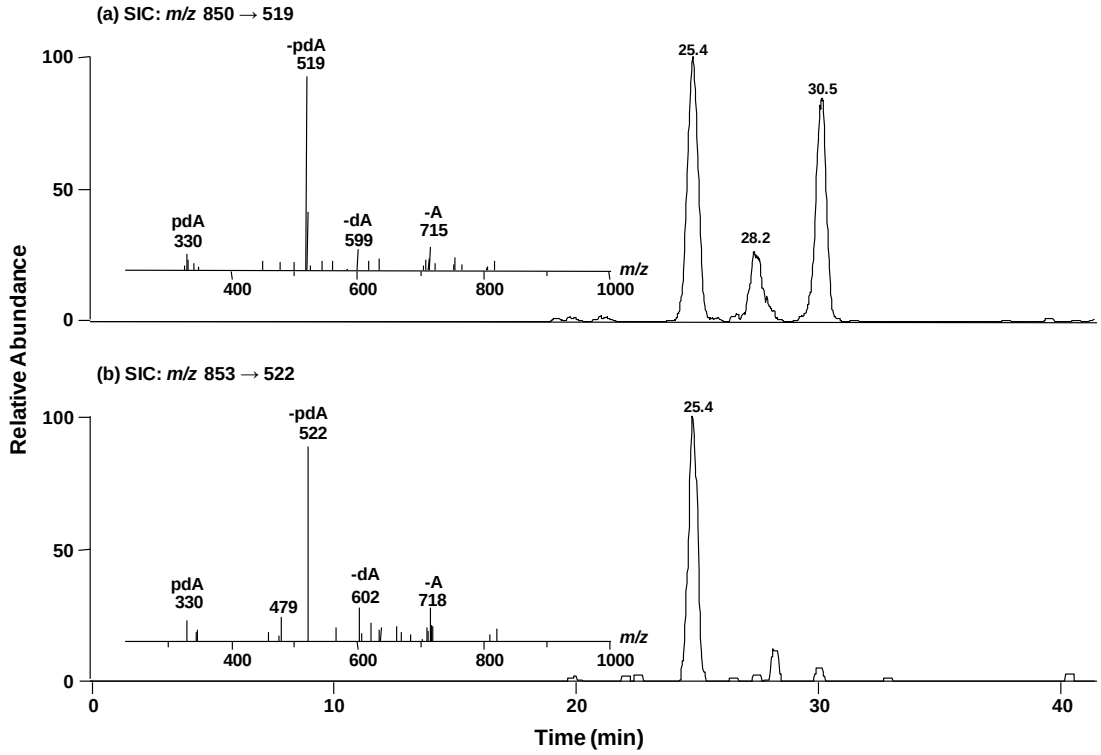


Figure 2.2. Selected-ion chromatograms (SICs) for monitoring the *m/z* 850→519 transition for 8-MOP-MA-containing dinucleotides (a) and the *m/z* 853→522 transition for D_3-8-MOP-MA-containing dinucleotide (b) resulting from nuclease P1 digestion of DNA extracted from wild-type CHO cells immediately after photoirradiation (i.e., at 0 hr). Shown in the insets are the product-ion spectra of the ESI-produced $[M-H]^-$ ions of the dinucleotide containing unlabeled and labeled 8-MOP-MA.

Role of NER in the Cellular Repair of 8-MOP Photolesions

After demonstrating that ICL and MAs could be readily detected in HF and CHO cells after exposure to 8-MOP and UVA light, we next assessed their repair in these two types of cells. For these experiments, we employed a drug treatment and UVA irradiation condition that minimally perturbs the physiology and DNA repair capacity

of the host cells. In particular, instead of dispersing the cells in PBS in previous experiments (24,25), we seeded the cells in the culture media during drug uptake and UVA light exposure. Additionally, the cells were not placed on ice during UVA exposure. With the use of these conditions, we found more than 90% viability of the cells at the end of the experimental period.

These experimental conditions, together with the use of LC-MS/MS, also allowed for examining the repair of 8-MOP-ICL and -MAs in mammalian cells. Our quantification results showed that, at 8 hr after photo-irradiation, the levels of 8-MOP-ICL in repair-proficient HF (GM00637) and CHO cells (CHO-AA8) decreased from 8.7 to 3.2 lesions and from 4.4 to 0 lesions per 10^6 nucleotides (Figure 2.3). However, cells deficient in XPA or ERCC1 failed to efficiently remove ICLs at 24 hr after 8-MOP had been photo-activated in these cells. For instance, ICL was detected at the level of 3.2 per 10^6 nucleotides in repair-competent GM00637 cells at 24 hr after 8-MOP/UVA treatment, while almost twice of that (5.7 ICLs per 10^6 nucleotides) was observed in XPA-deficient GM04429 cells (Figure 2.3a). The level of 8-MOP-ICL in ERCC1-deficient CHO cells also exhibited no significant change at 24 hr after irradiation, where the lesion remains at the level of 7.4 ICLs per 10^6 nucleotides (CHO-7-27, Figure 2.3b). These results also suggest that the dose of 8-MOP (2.5 μM)

and UVA exposure (2.0 J/cm^2) did not compromise the repair capacity of the HF and CHO cells.

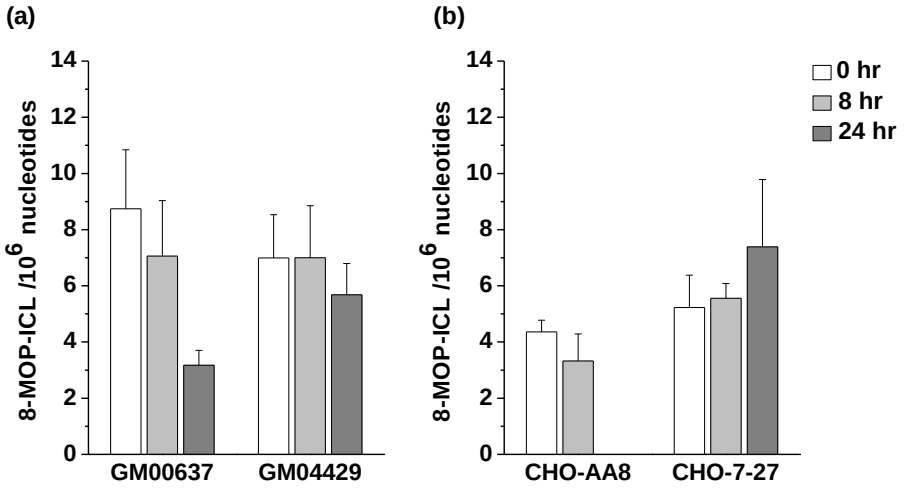


Figure 2.3. LC-MS/MS quantification results for 8-MOP-ICLs in DNA samples from human skin fibroblast (a) and Chinese hamster ovary cells (b). Data represent the means and standard deviations of results obtained from three independent experiments. *, $p < 0.05$; **, $p < 0.01$; #, $p > 0.05$. The *p*-values were calculated by using unpaired two-tailed *t*-test.

In the repair-proficient GM00637 cells, the level of 8-MOP-MA3 substantially declined from 47.7 to 12.3 adducts per 10^6 nucleotides at 24 hr after UVA exposure, which was accompanied with slight decreases of 8-MOP-MA1 and -MA2 (by 26% and 29%, respectively, Figure 2.4a-c). On the other hand, repair-proficient CHO-AA8 cells exhibited a more efficient removal of the 8-MOP-MA lesions. The levels of all three MAs in AA8 cells decreased dramatically with time, with MA1 being hardly detectable in cells at 24 hr after UVA irradiation (CHO-AA8, Figure 2.4). In contrast, the XPA-deficient GM04429 cells and ERCC1-deficient CHO-7-27 cells displayed no

apparent repair of any of the three MAs (Figure 2.4). These results demonstrated that the sensitivity, accuracy, and precision of the LC-MS/MS coupled with stable isotope dilution method facilitated the assessment of the repair of 8-MOP-induced ICL and MAs in mammalian cells. Although others and we previously employed LC-MS/MS for detecting DNA ICLs in cells or tissues (24-26), to our knowledge, this is one of the first few reports about the application of LC-MS/MS for examining the repair of any DNA ICLs (27).

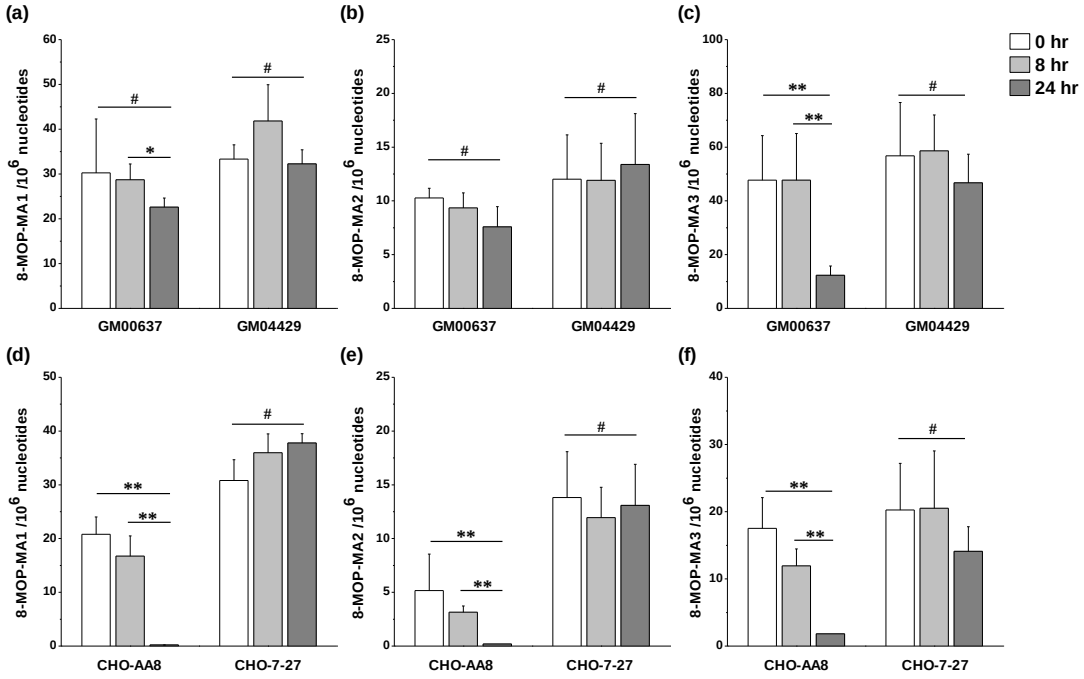


Figure 2.4. LC-MS/MS quantification results for 8-MOP-MAs in DNA samples. The results for MA1, MA2, and MA3 in human skin fibroblast cells are shown in (a), (b), and (c), respectively. The corresponding data for Chinese hamster ovary cells are displayed in (d), (e), and (f), respectively. The data represent the means and standard deviations of results obtained from three independent experiments. *, $p < 0.05$; **, $p < 0.01$; #, $p > 0.05$. The *p*-values were calculated by using unpaired two-tailed *t*-test.

The ERCC1-XPF complex is a structure-specific endonuclease that functions in NER by cleaving damaged strand at the 5' side to the lesion (37). Previous studies showed that, unlike yeast where loss of any of the several NER proteins confers sensitivity to crosslinking agents, only elimination of ERCC1 or XPF leads to exquisite sensitivity of mammalian cells toward these agents (38,39). In comparison, the increase in sensitivity to crosslinking agents was observed to be more modest with mutations in XPA, another NER protein involved in damage recognition (40). However, even if XPA is not required for the repair of an ICL substrate *in vitro* (41), failure to complete the repair step by XPA renders the ICLs persistent in cells and is at least cytostatic (40). Additionally, Niedernhofer et al. (42) found that XPA-mutant primary HFs are more sensitive than normal controls to mitomycin C, a potent DNA crosslinker. In support of these previous notions, our results provided direct evidence to demonstrate that XPA and ERCC1 are both required for the efficient repair of 8-MOP-ICL in mammalian cells, as reflected by the elevated accumulation of 8-MOP-ICL in mammalian cells deficient in one of these two genes. A similar finding was made for 8-MOP-MAs, underscoring the involvement of NER pathway in the repair of these psoralen-induced MAs.

It is of note that there is substantial functional overlap among the proteins in the various pathways for complete processing of bulky DNA damage. For example, recent

studies suggested that NEIL1-mediated base excision repair (BER) may also be involved in the repair of ICLs induced by 8-MOP/UVA (43,44). Therefore, it is important to assess, in the future, the roles of other DNA repair pathways in the removal of psoralen ICLs in mammalian cells. It can be envisaged that the analytical method developed here can be readily employed for such studies.

Removal Efficiency of 8-MOP-Induced DNA Lesions in Different Cell Types

The LC-MS/MS results also allowed us to compare the relative repair efficiencies of 8-MOP photoproducts in different cell types. The overall level of 8-MOP-ICL was found to be approximately 1.6-fold higher in HFs than in CHO cells obtained after UVA photoactivation. In addition, we observed that, at 8 hr post irradiation, 19% and 40% of the 8-MOP-ICL were removed in repair-proficient HF (GM00637) and CHO (CHO-AA8) cells, respectively (Figure 2.3). We also examined, by using the same method, the repair of 8-MOP photoadducts in 293T human embryonic kidney epithelial (HEK293T) cells. Interestingly, the yield of 8-MOP-ICL in HEK293T cells was below the detection limit, indicating that the lesion is much more rapidly repaired in this type of cells than in HFs and CHO cells (Figure A.3). Similarly, the rates of removal of 8-MOP-introduced MAs in three types of cells followed the same trend as that of the ICL, where the MA processing rate in repair-competent CHO cells was 1.5-2.0-fold higher than that in HFs, but markedly lower than that in HEK293T cells. This result

shows that the removal of psoralen photoadducts of DNA in mammalian cells is cell type-specific. Similar observation was made for other types of DNA lesions. For instance, D'Errico et al. (45) showed that lower levels of UV photoproducts are induced in keratinocytes relative to fibroblasts under the same UV dose regimen.

References

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

2. Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, et al. (2006) DNA Repair And Mutagenesis. Washington D.C.: American Society for Microbiology Press.

3. Lindahl T, Barnes DE (2000) Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* 65: 127-133.

4. Dronkert ML, Kanaar R (2001) Repair of DNA interstrand cross-links. *Mutat Res* 486: 217-247.

5. Muniandy PA, Liu J, Majumdar A, Liu ST, Seidman MM (2010) DNA interstrand crosslink repair in mammalian cells: step by step. *Crit Rev Biochem Mol Biol* 45: 23-49.

6. Deans AJ, West SC (2011) DNA interstrand crosslink repair and cancer. *Nat Rev Cancer* 11: 467-480.

7. McCabe KM, Olson SB, Moses RE (2009) DNA interstrand crosslink repair in mammalian cells. *J Cell Physiol* 220: 569-573.

8. Wood RD (2010) Mammalian nucleotide excision repair proteins and interstrand crosslink repair. *Environ Mol Mutagen* 51: 520-526.

9. Noll DM, Mason TM, Miller PS (2006) Formation and repair of interstrand cross-links in DNA. *Chem Rev* 106: 277-301.

10. Hearst JE (1989) Photochemistry of the psoralens. *Chem Res Toxicol* 2: 69-75.

11. Stern RS (2007) Psoralen and ultraviolet a light therapy for psoriasis. *N Engl J Med* 357: 682-690.

12. Dalla Via L, Marciani Magno S (2001) Photochemotherapy in the treatment of cancer. *Curr Med Chem* 8: 1405-1418.

13. Nijsten TE, Stern RS (2003) The increased risk of skin cancer is persistent after discontinuation of psoralen+ultraviolet A: a cohort study. *J Invest Dermatol* 121: 252-258.

14. Hlavin EM, Smeaton MB, Miller PS (2010) Initiation of DNA interstrand cross-link repair in mammalian cells. *Environ Mol Mutagen* 51: 604-624.

15. Zheng H, Wang X, Legerski RJ, Glazer PM, Li L (2006) Repair of DNA interstrand cross-links: interactions between homology-dependent and homology-independent pathways. *DNA Repair* 5: 566-574.

16. Pospisilova S, Kypr J (1997) UV light-induced crosslinking of the complementary strands of plasmid pUC19 DNA restriction fragments. *Photochem Photobiol* 65: 945-948.

17. Lee CS, Gibson NW (1991) DNA damage and differential cytotoxicity produced in human carcinoma cells by CC-1065 analogues, U-73,975 and U-77,779. *Cancer Res* 51: 6586-6591.

18. Hartley JM, Spanswick VJ, Gander M, Giacomini G, Whelan J, et al. (1999) Measurement of DNA cross-linking in patients on ifosfamide therapy using the single cell gel electrophoresis (comet) assay. *Clin Cancer Res* 5: 507-512.

19. Hartley JA, Souhami RL, Berardini MD (1993) Electrophoretic and chromatographic separation methods used to reveal interstrand crosslinking of nucleic acids. *J Chromatogr* 618: 277-288.

20. Courdavault S, Baudouin C, Charveron M, Canguilhem B, Favier A, et al. (2005) Repair of the three main types of bipyrimidine DNA photoproducts in human keratinocytes exposed to UVB and UVA radiations. *DNA Repair* 4: 836-844.

21. Harrington CF, Le Pla RC, Jones GDD, Thomas AL, Farmer PB (2010) Determination of Cisplatin 1,2-Intrastrand Guanine-Guanine DNA Adducts in Human Leukocytes by High-Performance Liquid Chromatography Coupled to Inductively Coupled Plasma Mass Spectrometry. *Chem Res Toxicol* 23: 1313-1321.

22. Wang J, Cao H, You C, Yuan B, Bahde R, et al. (2012) Endogenous formation and repair of oxidatively induced G[8-5 m]T intrastrand cross-link lesion. *Nucleic Acids Res* 40: 7368-7374.

23. Wang J, Clauson CL, Robbins PD, Niedernhofer LJ, Wang Y (2012) The oxidative DNA lesions 8,5'-cyclopurines accumulate with aging in a tissue-specific manner. *Aging Cell* 11: 714-716.

24. Cao H, Hearst JE, Corash L, Wang Y (2008) LC-MS/MS for the detection of DNA

interstrand cross-links formed by 8-methoxypsoralen and UVA irradiation in human cells. *Anal Chem* 80: 2932-2938.

25. Lai C, Cao H, Hearst JE, Corash L, Luo H, et al. (2008) Quantitative analysis of DNA interstrand cross-links and monoadducts formed in human cells induced by psoralens and UVA irradiation. *Anal Chem* 80: 8790-8798.

26. Malayappan B, Johnson L, Nie B, Panchal D, Matter B, et al. (2010) Quantitative High-Performance Liquid Chromatography-Electrospray Ionization Tandem Mass Spectrometry Analysis of Bis-N7-Guanine DNA-DNA Cross-Links in White Blood Cells of Cancer Patients Receiving Cyclophosphamide Therapy. *Anal Chem* 82: 3650-3658.

27. Paz MM, Ladwa S, Champeil E, Liu Y, Rockwell S, et al. (2008) Mapping DNA adducts of mitomycin C and decarbamoyl mitomycin C in cell lines using liquid chromatography/ electrospray tandem mass spectrometry. *Chem Res Toxicol* 21: 2370-2378.

28. Rolig RL, Layher SK, Santi B, Adair GM, Gu F, et al. (1997) Survival, mutagenesis, and host cell reactivation in a Chinese hamster ovary cell ERCC1 knock-out mutant. *Mutagenesis* 12: 277-283.

29. von Kobyletzki G, Hoffmann K, Kerscher M, Altmeyer P (1998) Plasma levels of 8-methoxypsoralen following PUVA-bath photochemotherapy. *Photodermatol Photoimmunol Photomed* 14: 136-138.

30. Thomas SE, O'Sullivan J, Balac N (1991) Plasma levels of 8-methoxypsoralen following oral or bath-water treatment. *Br J Dermatol* 125: 56-58.

31. Tegeder I, Brautigam L, Podda M, Meier S, Kaufmann R, et al. (2002) Time course of 8-methoxypsoralen concentrations in skin and plasma after topical (bath and cream) and oral administration of 8-methoxypsoralen. *Clin Pharmacol Ther* 71: 153-161.

32. Tanew A, Kipfelsperger T, Seeber A, Radakovic-Fijan S, Honigsmann H (2001) Correlation between 8-methoxypsoralen bath-water concentration and photosensitivity in bath-PUVA treatment. *J Am Acad Dermatol* 44: 638-642.

33. Yeo UC, Shin JH, Yang JM, Park KB, Kim MM, et al. (2000) Psoralen-ultraviolet A-induced erythema: sensitivity correlates with the concentrations of psoralen in suction blister fluid. *Br J Dermatol* 142: 733-739.

34. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16: 1215.

35. Wang Y, Wang Y (2003) Structure elucidation of DNA interstrand cross-link by a combination of nuclease P1 digestion with mass spectrometry. *Anal Chem* 75: 6306-6313.

36. Gasparro FP (1988) Immunological assay of 8-MOP photoadducts in DNA. *J Photochem Photobiol B* 2: 286-288.

37. Evans E, Moggs JG, Hwang JR, Egly JM, Wood RD (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J* 16: 6559-6573.

38. De Silva IU, McHugh PJ, Clingen PH, Hartley JA (2000) Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. *Mol Cell Biol* 20: 7980-7990.

39. Usanova S, Piee-Staffa A, Sied U, Thomale J, Schneider A, et al. (2010) Cisplatin sensitivity of testis tumour cells is due to deficiency in interstrand-crosslink repair and low ERCC1-XPF expression. *Mol Cancer* 9: 248.

40. Bergstralh DT, Sekelsky J (2008) Interstrand crosslink repair: can XPF-ERCC1 be let off the hook? *Trends Genet* 24: 70-76.

41. Cipak L, Watanabe N, Bessho T (2006) The role of BRCA2 in replication-coupled DNA interstrand cross-link repair in vitro. *Nat Struct Mol Biol* 13: 729-733.

42. Niedernhofer LJ, Garinis GA, Raams A, Lalai AS, Robinson AR, et al. (2006) A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature* 444: 1038-1043.

43. Couve S, Mace-Aime G, Rosselli F, Saparbaev MK (2009) The human oxidative DNA glycosylase NEIL1 excises psoralen-induced interstrand DNA cross-links in a three-stranded DNA structure. *J Biol Chem* 284: 11963-11970.

44. Couve-Privat S, Mace G, Rosselli F, Saparbaev MK (2007) Psoralen-induced DNA adducts are substrates for the base excision repair pathway in human cells. *Nucleic Acids Res* 35: 5672-5682.

45. D'Errico M, Teson M, Calcagnile A, Nardo T, De Luca N, et al. (2005) Differential

role of transcription-coupled repair in UVB-induced response of human fibroblasts and keratinocytes. *Cancer Res* 65: 432-438.

Chapter 3: Quantitative Assessment of Tet-induced Oxidation Products of 5-methylcytosine in Cellular and Tissue DNA

Introduction

Eukaryotic chromatin possesses a wealth of information that is required for the development and growth of a multi-cellular organism. Such information is embedded both genetically in the DNA sequence itself and epigenetically via DNA methylation and post-translational modifications of core histone proteins (1,2). In mammals, DNA methylation occurs at the C5 position of cytosine residues mainly at CpG dinucleotide sites (1). This methylation pattern is established during embryonic development by *de novo* DNA methyltransferases DNMT3a and DNMT3b, and maintained during cell division by maintenance DNA methyltransferase DNMT1 (1). The importance of cytosine methylation in mammalian development is manifested by the observations that mice deficient in DNMT3a die at 4 weeks of age (3), and mice lacking DNMT1 or DNMT3b are embryonically lethal (3,4).

Studies in the past decade showed that DNA methylation pattern in mammals can be lost through passive or active mechanisms (5,6). Elimination or inhibition of DNMT1 activity can lead to loss of cytosine methylation during cell division, a process known as passive cytosine DNA demethylation. Active cytosine DNA demethylation is

the conversion of 5-methylcytosine (5-mC) to unmethylated cytosine in a process that is independent of DNA replication (5). Multiple lines of evidence support the existence of active cytosine demethylation in the entire genome or in specific genomic loci, and multiple mechanisms have been proposed for this process (6).

Recent discovery of Ten-eleven translocation (Tet) family dioxygenases, including Tet1, Tet2 and Tet3, in the oxidation of 5-mC sheds significant new light on active cytosine demethylation in mammals. Tet1 was first observed to catalyze the oxidation of 5-mC to 5-hydroxymethylcytosine (5-HmC) in DNA *in vitro* and in human cells (7), and all three members of mouse Tet family were later found to carry the same enzymatic activity (8). Aside from the generation of 5-HmC, Tet family proteins are also capable of oxidizing 5-mC to 5-formylcytosine (5-FoC) and 5-carboxylcytosine (5-CaC) (9,10). 5-FoC and 5-CaC in DNA can be readily excised by thymine DNA glycosylase (TDG) *in vitro* (9,11), and genetic depletion of TDG in mouse embryonic stem cells leads to substantial accumulation of 5-CaC (9). These results suggest that Tet-mediated oxidation of 5-mC to 5-FoC and 5-CaC, followed by the excision of the latter two by TDG and restoration of an unmethylated cytosine by base-excision repair (BER) machinery, constitutes a new pathway for active cytosine demethylation in mammals (Figure 3.1).

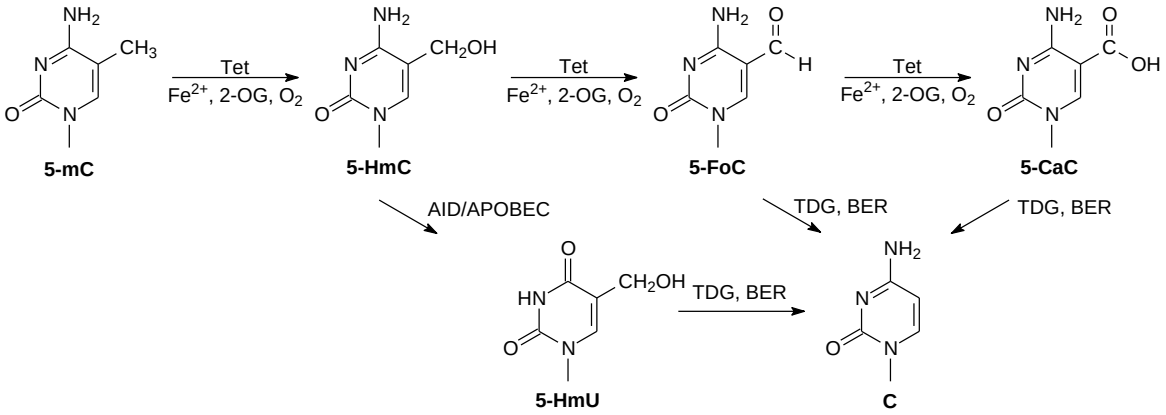


Figure 3.1. Proposed mechanisms for Tet-mediated oxidative demethylation of 5-methylcytosine in DNA. In the presence of Fe^{2+}, 2-oxoglutarate (2-OG) and O_2, Tet family proteins can catalyze the sequential oxidation of 5-mC in DNA to give 5-HmC, 5-FoC and 5-CaC. 5-CaC can then be recognized and cleaved by TDG and subsequent action of BER machinery leads to the restoration of unmethylated cytosine. Alternatively, the Tet-produced 5-HmC can be deaminated by AID/APOBEC deaminases to give 5-HmU, which can be cleaved and replaced with unmethylated cytosine by TDG/SMUG1 and BER machinery.

Apart from the direct removal of Tet-produced oxidation products (i.e., 5-FoC and 5-CaC) by TDG, it was also proposed that 5-HmC in DNA can be deaminated, by AID (activation-induced deaminase)/APOBEC (apolipoprotein B mRNA-editing enzyme complex) families of cytidine deaminases, to yield 5-hydroxymethyluracil (5-HmU) (12,13). While 5-HmC in DNA is a poor substrate for TDG, 5-HmU, when paired with a guanine, can be readily excised by DNA glycosylases (e.g. TDG and SMUG1) (11,12,14). Thus, the oxidation of 5-mC to 5-HmC by Tet proteins, deamination of the latter nucleobase by AID/APOBECs, and TDG-induced BER of the resulting 5-HmU

may also give rise to active cytosine demethylation in mammals (Figure 3.1) (12,13). The involvement of this sequential oxidation-deamination mechanism in active cytosine demethylation was recently challenged by the apparent lack of significant biochemical activity of recombinant AID or APOBEC toward 5-HmC deamination *in vitro* or in cultured cells because of lack of detection of 5-HmU (15). Nevertheless, it remains possible that such deamination may occur in specific cellular context(s).

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) method has played an important role in the identification of the Tet-induced oxidation products of 5-mC (7,9,10). The method, along with the use of external standards, has also been used for assessing the levels of 5-HmC, 5-FoC and 5-CaC in cells or tissues (10,15). However, we reason that the use of stable isotope-labeled internal standards will offer unambiguous identification and more accurate measurements of levels of intermediates that are proposed to be involved in active cytosine demethylation, which, along with genetic manipulation, may provide important insights into the mechanisms of active cytosine demethylation.

Aberrant epigenetic regulation of gene expression has long been known to be associated with cancer development (16). Along this line, Tet1 is a fusion partner of the mixed lineage leukemia (MLL) gene in acute myeloid leukemia (17,18). Lower levels of 5-HmC has been observed in myeloid leukemia and a number of solid tumors

(including carcinomas of the breast, liver, lung, pancreas and prostate), as shown by immunohistochemistry or dot blot analysis (19,20), and in lung and brain tumors, as revealed by both immunohistochemistry and LC-MS/MS analysis (21). Thus, a robust analytical method for the accurate quantification of the aforementioned 5-mC oxidation intermediates may also result in the discovery of biomarkers for the diagnosis of cancer and perhaps other human diseases.

Herein, we developed an LC-MS/MS/MS coupled with the stable isotope-dilution method for the sensitive, accurate and simultaneous quantification of 5-hydroxymethyl-2'-deoxycytidine (5-HmdC), 5-formyl-2'-deoxycytidine (5-FodC), 5-carboxyl-2'-deoxycytidine (5-CadC), and 5-hydroxymethyl-2'-deoxyuridine (5-HmdU) in genomic DNA of cultured human cells and multiple mammalian tissues.

Materials and Methods

Materials

All chemicals and enzymes, unless otherwise specified, were from Sigma-Aldrich (St. Louis, MO). [1,3-$^{15}N_2$-2'-D]-5-hydroxymethyl-2'-deoxycytidine ([1,3-$^{15}N_2$-2'-D]-HmdC), [1,3-$^{15}N_2$-2'-D]-5-hydroxymethyl-2'-deoxyuridine ([1,3-$^{15}N_2$-2'-D]-HmdU), and [1,3-$^{15}N_2$-2'-D]-5-formyl-2'-deoxycytidine ([1,3-$^{15}N_2$-2'-D]-FodC) were synthesized previously (22,23). [4-*amino*-1,3-$^{15}N_3$]-2'-deoxycytidine ([4-*amino*-1,3-

$^{15}N_3$]-dC) was purchased from Cambridge Isotope Laboratories (Andover, MA). Proteinase K was obtained from New England Biolabs (Ipswich, WA). The HEK293T human embryonic kidney cells, HeLa human cervical cancer cells, WM-266-4 human melanoma cells, and cell culture reagents were purchased from ATCC (Manassas, VA, USA). Expression vectors for the catalytic domain of human Tet1 (amino acids 1418-2136) and its corresponding mutant Tet1m (H1672Y/D1674A) were previously described (12) and were deposited to Addgene (#39454 and #39455). The DNA of mouse embryonic stem (ES) cells, which were derived from mouse blastocysts, was provided by Prof. Gerd P. Pfeifer (City of Hope).

Synthesis and Characterization of [4-amino-1,3-$^{15}N_3$]-5-Carboxyl-2'-deoxycytidine ([4-amino-1,3-$^{15}N_3$]-5-CadC)

The title compound was prepared at a microscale according to previously published methods for the synthesis of the corresponding unlabeled compounds with some modifications (24). [4-*amino*-1,3-$^{15}N_3$]-dC (3.0 mg, 13.0 μmol) was dissolved in DMF (100 μL) and allowed to stir at room temperature (rt) for 10 min. Slight excess of iodine (2.13 mg), followed by *m*-chloroperbenzoic acid (2.97 mg), was added and allowed to stir at rt for 1.5 h. The reaction mixture was then dried by Speed-vac. The sample was reconstituted with water and filtered. The resulting [4-*amino*-1,3-$^{15}N_3$]-5-iodo-2'-deoxycytidine ([4-*amino*-1,3-$^{15}N_3$]-5-IdC) was purified

by HPLC.

The purified [4-*amino*-1,3-$^{15}N_3$]-5-IdC was dissolved in methanol (500 μL) and CO was bubbled through the solution for 30 min, to which solution were subsequently added triphenylphosphine (1.37 mg), triethylamine (2 μL) and tris(dibenzylidene-acetone)dipalladium(0) (0.81 mg). The reaction was stirred at 50°C under CO atmosphere overnight. The mixture was then cooled to rt and the solid removed by filtration. The resulting [4-*amino*-1,3-$^{15}N_3$]-5-methoxycarbonyl-2'-deoxycytidine was readily converted to the desired [4-*amino*-1,3-$^{15}N_3$]-5-CadC by treating with 0.1 M K_2CO_3 in MeOH/H_2O at 40°C overnight. The latter compound was isolated from the reaction mixture by HPLC. The identity of [4-*amino*-1,3-$^{15}N_3$]-5-CadC was confirmed by LC-MS/MS analysis and by high-resolution MS analysis on an ESI-TOF instrument. The latter measurement gave *m/z* 273.0618 for the $[M - H]^-$ ion, whose calculated *m/z* was 273.0637.

Cell Culture, Transfection and DNA Extraction

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (ATCC). HeLa human cervical cancer cells and WM-266-4 human melanoma cells were cultured in Eagle's Minimum Essential Medium (ATCC). All cells were incubated at 37℃ in 5% CO_2 atmosphere. The culture medium was supplemented with 10% fetal bovine serum and 100 IU/mL penicillin. The HEK293T cells were seeded in 6-well

plates at 50-60% confluence and, 24 h later, the cells were transfected with approximately 1.5 μg Tet1 or Tet1m expression plasmids using Lipofectamine 2000 (Invitrogen). The cells were harvested for DNA extraction 48 h after plasmid transfection.

Genomic DNA was extracted from the cells and mouse brain tissues using a high-salt method, following published procedures (25). With this method, we could typically obtain approximately 30 μg of DNA from 5 million cells or 100 μg of DNA from 100 mg of tissues. The whole brain tissues were dissected from 8-weeks old C57/B6 strain of male mice, snap-frozen in liquid nitrogen and stored in –80 ℃ freezer prior to DNA extraction. Genomic DNA samples from human brain (cerebellum, Table B.1 gives the patient information) and mouse skin tissues were previously isolated (26,27). The DNA pellet was redissolved in doubly distilled water and its concentration measured by UV absorbance at 260 nm.

Enzymatic Digestion

For the enzymatic digestion of DNA, nuclease P1 (0.1 U/μg DNA), phosphodiesterase 2 (0.00025 U/μg DNA), and erythro-9-(2-hydroxy-3-nonyl)adenine (EHNA, 1 mM in final concentration) were added to 30-80 μg of cellular or tissue DNA in a solution containing 30 mM sodium acetate (pH 5.6) and 1.0 mM zinc acetate. The digestion mixture was incubated at 37 ℃ for 48 h and continued for another 2 h

after alkaline phosphatase (0.05 U/μg DNA), phosphodiesterase 1 (0.0005 U/μg DNA) and Tris-HCl (pH 8.9, final concentration 50 mM) were added. To the mixture were then added 30 fmol of [1,3-$^{15}N_2$-2'-D]-5-FodC, 25 fmol of [4-*amino*-1,3-$^{15}N_3$]-5-CadC, and 2 pmol of [1,3-$^{15}N_2$-2'-D]-5-HmdU. The enzymes were removed by chloroform extraction. The aqueous layer was dried, reconstituted in doubly distilled water, and subjected to off-line HPLC for the enrichment of target nucleosides.

For LC-MS/MS/MS quantification of 5-HmdC, 50 fmol of [1,3-$^{15}N_2$-2'-D]-5-HmdC was added to the enzymatic digestion mixture of 10-50 ng genomic DNA. The samples were subsequently extracted with chloroform to remove the enzymes and the aqueous layer subjected to LC-MS/MS/MS analysis.

HPLC

HPLC purification of [4-*amino*-1,3-$^{15}N_3$]-5-IdC was performed on an Agilent 1100 series HPLC system equipped with a binary pump and a UV detector. A 4.6×250 mm Apollo C18 column (5 μm in particle size and 300 Å in pore size, Grace Inc., Deerfield, IL) was used and the wavelength was set at 260 nm. A solution of 10 mM ammonium formate (pH 6.3, solution A) and a mixture of 10 mM ammonium formate and acetonitrile (70:30, v/v, solution B) were employed as mobile phases. The flow rate was 0.50 mL/min, and a gradient of 5 min 0-20% B, 40 min 20-55% B and 5 min 55-98% B was used. The purification of [4-*amino*-1,3-$^{15}N_3$]-5-CadC was conducted under the

same conditions except that a 4.6×250 mm Alltima C18 column (5 μm in particle size and 300 Å in pore size, Grace Inc., Deerfield, IL) with a gradient of 10 min 0-2% B, 15 min 2-15% B and 5 min 15-98% B was employed.

The enrichment of 5-FodC, 5-CadC, and 5-HmdU from the enzymatic digestion products of DNA was carried out on a Beckman HPLC system with pump module 125 and a UV detector (module 126). A 4.6 × 250 mm Aeris Widepore C18 column (3.6 μm in particle size, Phenomenex, Torrance, CA) was used. An isocratic elution at a flow rate of 0.8 mL/min and with a solution of 10 mM ammonium formate (pH 8.5) as mobile phase was employed. A typical HPLC trace is depicted in Figure B.1 in the Supporting Information. From the enrichment HPLC trace, we often observed a low level (< 5%) of RNA contamination for the DNA samples, and the DNA amount was corrected based on the ratio of peak areas found for guanosine (rG) to 2'-deoxyguanosine (dG) in the HPLC trace (Figure B.1). The HPLC fractions eluting at 3.0-4.0, 9.5-10.5, and 26.0-27.5 min were collected for 5-CadC, 5-HmdU, and 5-FodC, respectively. The collected fractions were dried, redissolved in water, and injected for LC-MS/MS/MS analysis.

During the above enrichment, 5-mdC could also be resolved from other nucleosides. The levels of 5-mdC, expressed as percentage of 2'-deoxyguanosine (dG), were quantified based on the peak areas of 5-mdC and dG found in the chromatogram

with the consideration of the extinction coefficients of the two nucleosides at 260 nm (5,020 and 11,700 L mol^{-1} cm^{-1} for 5-mdC and dG, respectively) (28,29).

Quantification with LC-MS/MS/MS

The LC-MS/MS/MS experiments were conducted using a 0.5 × 250 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies, Santa Clara, CA) and an Agilent 1200 capillary HPLC pump. The flow rate was 8.0 μL/min. The effluent from the LC column was directed to an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific), which was set up for monitoring the labeled and unlabeled 5-HmdC, 5-FodC, and 5-CadC in the positive-ion mode, and the labeled and unlabeled 5-HmdU in the negative-ion mode. A solution of 0.1% (v/v) formic acid in water (solution A) and a solution of 0.1% (v/v) formic acid in methanol (solution B) were used as mobile phases for 5-HmdC, 5-FodC, and 5-CadC analyses. A solution of 2 mM ammonium formate in water (solution A) and a solution of 2 mM ammonium formate in methanol (solution B) were used as mobile phases for 5-HmdU analysis. A gradient of 5 min 0-20% B and 25 min 20-70% B was employed for the separation of all the modified nucleosides. The temperature for the ion transport tube was maintained at 275 ℃ and 300 ℃ in the positive- and negative-ion modes, respectively. The sheath gas flow rate was 15 arbitrary units, and no auxiliary gas was used. The electrospray, capillary, and tube lens voltages for the positive-ion mode analyses were 5 kV, 14 V, and 20 V,

respectively, and these values for the negative-ion mode analyses were 4.5 kV, −10 V, and −100 V, respectively. The sensitivities for detecting all these modified nucleosides were optimized by varying two working parameters of the LTQ mass spectrometer, i.e., normalized collision energy and activation Q (Table B.2).

The numbers of moles of 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU in the nucleoside mixtures were calculated from the area ratios of peaks found in the selected-ion chromatograms (SICs) for the analytes over their corresponding isotope-labeled standards, the amounts of the labeled standards added, and the equations derived from the calibration curves. The final data, expressed in terms of numbers of modifications per 10^6 nucleosides (or per 10^3 5-mdC), were calculated by dividing the number of moles of the modified nucleosides with the number of moles of total nucleosides (or 5-mdC) in the digested DNA (Table B.3).

Results

Quantification of 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU in DNA of HEK293T Cells with Tet1 Overexpression and in DNA of Cultured Human Cancer Cells

We set out to develop an LC-MS/MS/MS combined with stable isotope-dilution method for the accurate assessment of levels of 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU in genomic DNA of cultured cells and tissues. We first examined how overexpression of Tet1 affects the levels of the four modified pyrimidine nucleosides in

genomic DNA of cultured human cells. We expressed the catalytic domain of Tet1 or its corresponding inactive mutant in HEK293T cells, isolated the genomic DNA from the cells, and digested the DNA to nucleosides with a cocktail of four enzymes (See *Materials and Methods*). Three of the aforementioned pyrimidine nucleosides, 5-HmdU, 5-FodC, and 5-CadC, were subsequently enriched from the digestion mixture with reversed-phase HPLC, and the enriched fractions were subjected to LC-MS^3 analyses using an LTQ linear ion trap mass spectrometer. As displayed in Figure B.1, the three modified pyrimidine nucleosides can be readily resolved from each other and from the four canonical nucleosides by reversed-phase HPLC with isocratic elution. Owing to the high level of 5-HmdC in cellular and tissue DNA, the digestion mixture of much lower amount of genomic DNA (i.e., 10 ng of tissue DNA or 10-50 ng of cellular DNA) was directly subjected for LC-MS^3 analysis of 5-HmdC.

The fragmentation of the protonated ions of unlabeled and labeled 5-HmdC and 5-FodC was previously described (23). Briefly, upon collisional activation, the $[M+H]^+$ ions of unlabeled 5-HmdC and 5-FodC can readily eliminate a 2-deoxyribose moiety to yield the protonated ions of the nucleobase portion (*m/z* 142 and 140 for 5-HmdC and 5-FodC, respectively) as the dominant fragment ions in MS/MS. Further collisional activation of the ions of *m/z* 142 and 140 leads to facile loss of H_2O and HNCO moieties to yield fragment ions of *m/z* 124 and 97 in MS^3 for 5-HmdC and 5-FodC,

respectively (Figure 3.2a&b, inset). Collisional activation of the protonated ion of 5-CadC again resulted in facile formation of the protonated ion of 5-CaC (*m/z* 156) in MS/MS (spectrum not shown), the latter fragment can lose a H_2O molecule to give the ion of *m/z* 138 in MS^3 (Figure 3.2c, inset). As we found previously, 5-HmdU exhibits better sensitivity when analyzed in the negative-ion mode (22). The deprotonated 5-HmdU loses readily an HNCO moiety to yield an ion of *m/z* 214 in MS/MS, and further collisional activation of the latter leads to the loss of part of the 2-deoxyribose component to produce the ion of *m/z* 124 in MS^3 (Figure 3.2d, inset). The proposed structures for the major fragment ions observed in MS/MS and MS^3 are displayed in Figure B.2.

The nearly identical elution time and similar MS^3 spectra for the analytes and their stable isotope-labeled counterparts allowed for the unambiguous identification and reliable quantification of the four pyrimidine nucleosides in the digestion mixture of DNA isolated from the HEK293T cells (Figure 3.2). The calibration curves for 5-HmdC, 5-FodC and 5-CadC are depicted in Figure B.3, and the calibration curve for 5-HmdU was shown previously (30). In this vein, because the levels of 5-HmdC, 5-FodC, and 5-CadC cover a wide range in different samples, calibration curves in two concentration ranges were used for determining accurately the levels of these mdC modifications. Our quantification results revealed that the levels of 5-HmdC, 5-FodC,

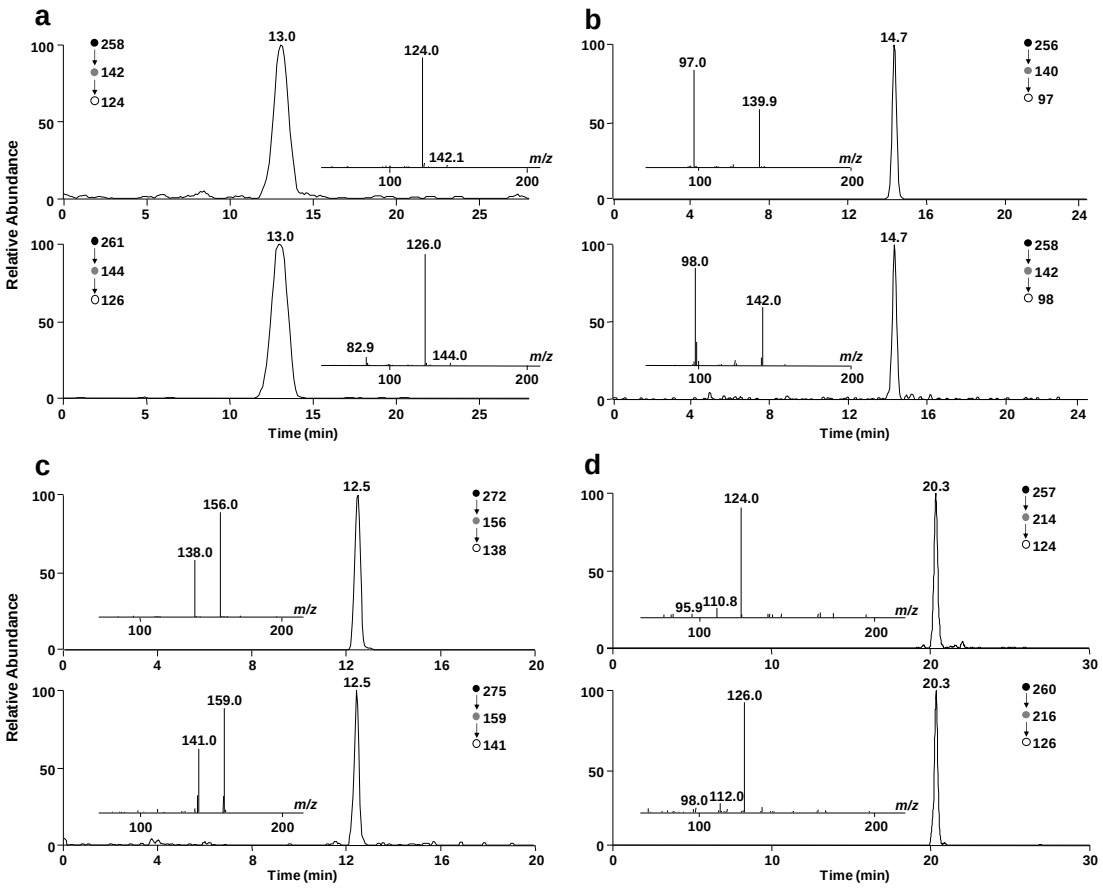


Figure 3.2. Representative LC-MS/MS/MS results for the quantification of 5-HmdC (a), 5-FodC (b), 5-CadC (c), and 5-HmdU (d) in HEK293T cells expressing the catalytic domain of Tet1. Shown are the selected-ion chromatograms for monitoring the indicated transitions for the analytes (top trace) and the isotope-labeled standards (bottom trace), and the inset gives the corresponding MS/MS/MS for the analytes and internal standards.

5-CadC, and 5-HmdU in HEK293T cells overexpressing the catalytically active form of Tet1 were 960, 76, 102, and 7.9 modifications per 10^6 nucleosides, respectively (Figure 3.3a-d and Table B.3). The levels of the four nucleosides were 32.5, 0.23, 0.18 and 6.3 modifications per 10^6 nucleosides, respectively, in DNA of HEK293T cells

overexpressing the catalytically inactive form of Tet1 (Figure 3.3a-d and Table B.3). The catalytic activity of Tet1 led to markedly higher levels (by ~29.5-, 330-, and 570-fold for 5-HmdC, 5-FodC and 5-CadC, respectively) of the three 5-mdC oxidation products. Thus, we confirmed the previous findings that 5-HmdC, 5-FodC and 5-CadC can be induced from Tet1-mediated oxidation of 5-mdC in genomic DNA (7-10). Our results also provided, for the first time, unambiguous evidence to support that, in contrast to the substantial elevation of the levels of the three 5-mdC oxidation products described above, Tet1 catalytic activity alone only gave rise to a modest (by ~25%) increase in the level of 5-HmdU in genomic DNA of cultured cells (Figure 3.3d).

We next assessed whether these oxidation products of 5-mdC are present in genomic DNA of cultured human cancer cells. Our results showed that, in genomic DNA of HeLa cells, the levels of 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU were 31.2, 0.67, 0.27, and 3.0 per 10^6 nucleosides, respectively (Figure 3.3a-d). In addition, these four nucleosides were present in genomic DNA of WM-266-4 human melanoma cells at the levels of 12.2, 0.69, 0.29, and 3.4 modifications per 10^6 nucleosides, respectively (Figure 3.3a-d). Thus, the levels of these modifications are similar as what were found for HEK293T cells expressing catalytically inactive form of Tet1, except that the level of 5-HmdC in WM-266-4 cells is approximately 2.5-fold lower than that in HEK293T or HeLa cells.

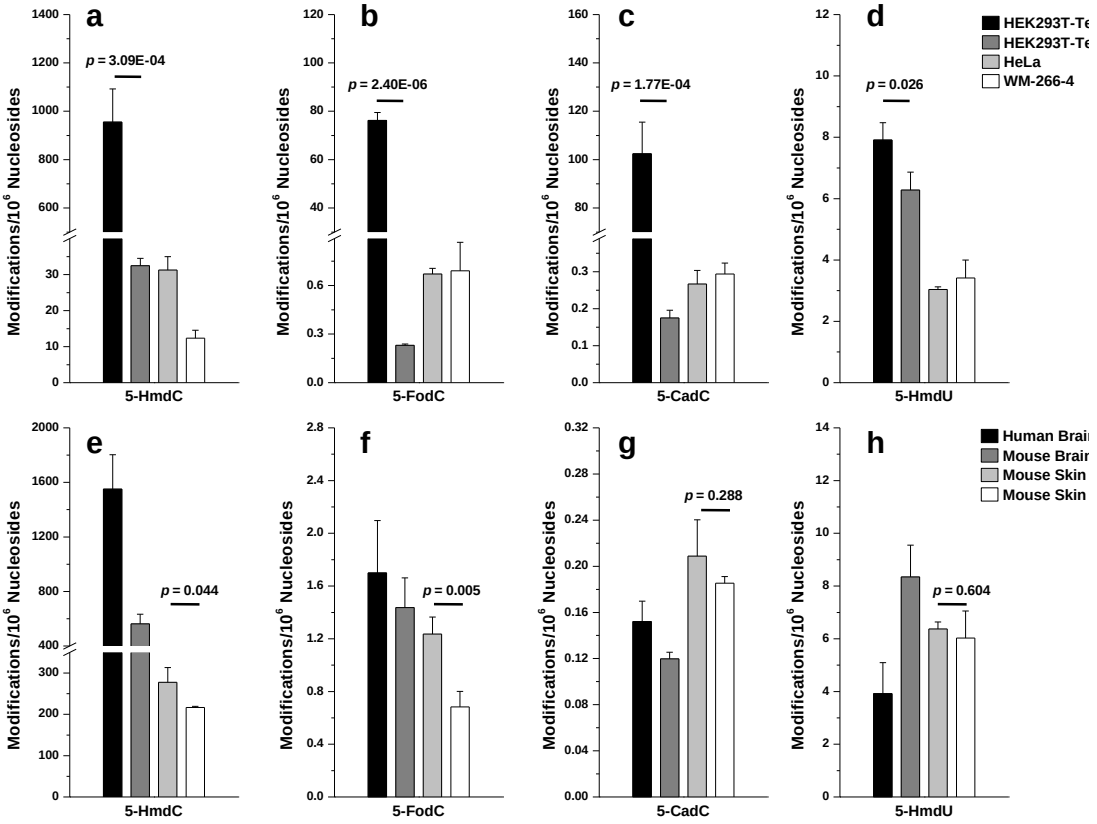


Figure 3.3. Quantification results for the levels of 5-HmdC (a), 5-FodC (b), 5-CadC (c), and 5-HmdU (d) in HEK293T cells expressing the active or inactive forms of Tet1 (n = 3) and in HeLa and WM-266-4 cells (n = 3). Quantification results for the levels of 5-HmdC (e), 5-FodC (f), 5-CadC (g), and 5-HmdU (h) in human brain tissues (black bar, n = 4), mouse brain tissues (dark grey bar, n = 3), and skin tissues of redhead (light grey bar, n = 3) and albino (open bar, n = 3) mice. The data represent the mean and standard deviation of the measurement results. The *p* values were calculated using unpaired two-tailed *t* test.

We further measured the levels of 5-HmdC, 5-FodC, and 5-CadC in DNA of mouse embryonic stem (ES) cells (Table B.3). For comparison with previous LC-MS/MS measurement results obtained from the use of the external standards (10),

we converted the data to the number of modified nucleosides per 10^6 dC. Our results showed that the levels of 5-HmdC, 5-FodC and 5-CadC are 680±50, 14.5±2.9, and 3.4±0.3 modifications per 10^6 dC, respectively. Our data for 5-FodC and 5-CadC are largely consistent with those reported previously (10), where the levels of the two nucleosides from duplicate measurements were 17.9/19.8 and 3.6/3.1 modifications per 10^6 dC, respectively. However, the level of 5-HmdC from the previous measurement, i.e., 1120/1380 modifications per 10^6 dC (10), was approximately two-fold higher than what we measured here.

Quantification of 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU in Genomic DNA from Human and Mouse Brain

We next quantified the levels of the modified nucleosides in human and mouse brain tissues. In human brain (cerebellum) DNA, the average levels of 5-HmdC, 5-FodC and 5-CadC are 1550, 1.7, and 0.15 modifications per 10^6 nucleosides, respectively, with the three nucleosides exhibiting a ratio of ~10000:11:1 (Figure 3.3e-g and Table B.3). Given these brain tissues were from four individuals at different ages (Table B.1), the relatively small standard deviations associated with measurement results suggested a lack of substantial age-dependent effect on the levels of these modifications in human brain. In mouse brain genomic DNA, the contents of 5-HmdC, 5-FodC and 5-CadC display a ratio of ~4700:12:1, at levels of 560, 1.4, and 0.12

modifications per 10^6 nucleosides, respectively (Figure 3.3e-g and Table B.3). The levels of 5-HmdU in DNA from human and mouse brain were 3.9 and 8.3 modifications per 10^6 nucleosides, respectively, which are drastically (by ~400- and 70-fold, respectively) lower than those of 5-HmdC (Figure 3.3h), but higher than 5-FodC (by 2.3- and 5.9-fold, respectively) and markedly higher than 5-CadC (by 26- and 69-fold, respectively).

Quantification of 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU in DNA from Mouse Skin Tissues

Mitra et al. (27) recently studied genomic consequences of functional deficiency in the melanocortin 1 receptor gene (*Mc1r*$^{e/e}$), the "redhead-fairskin" gene in numerous species including mouse and man. The redhair-fairskin *Mc1r* allele results in elevated production of pheomelanin (red pigment), and elevations in endogenous reactive oxygen species (ROS) were suggested to contribute to UV-independent generation of melanoma in "redhair-fairskin" (redhead) mice (27). We reasoned that the increased generation of ROS may also result in a rise in levels of 5-mdC modifications within skin tissues of redhead mice as compared to albino mice, the latter of which carry a dysfunctional tyrosinase gene (*Tyr*$^{c/c}$) that abolishes all melanin (including pheomelanin) biosynthesis. Thus, we also analyzed the levels of the four pyrimidine nucleosides in DNA isolated from the skin tissues of these two strains of mice (*Mc1r*$^{e/e}$ vs.

Mc1r$^{e/e}$;*Tyr*$^{c/c}$). Our results showed that indeed the skin of redhead mice contains significantly higher levels of 5-HmdC (280 vs. 217 per 10^6 nucleosides) and 5-FodC (1.2 vs. 0.7 per 10^6 nucleosides) as compared to skin from *Mc1r*$^{e/e}$ albino mice (Figure 3.3e&f, representative LC-MS/MS/MS results are shown in Figure B.4). The levels of 5-CadC (0.21 vs. 0.19 per 10^6 nucleosides) and 5-HmdU (6.4 vs. 6.0 per 10^6 nucleosides, Figure 3.3g&h) in these two strains of mice are, however, very similar.

Discussion

Our LC-MS/MS/MS coupled with stable isotope-dilution method facilitates the accurate and reproducible detection of 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU in DNA isolated from cultured human cells and multiple mammalian tissues. In this context, it is worth emphasizing the advantages of the isotope-dilution method for the measurements of these mdC oxidation products. In this method, we added the stable isotope-labeled standards to the nucleoside mixture; thus, the analytes and the corresponding internal standards are enriched by HPLC and analyzed by LC-MS/MS/MS under identical conditions. Any variation in experimental conditions during the HPLC enrichment and LC-MS/MS/MS analysis will not influence the measurement of the molar ratios of the analytes over their corresponding internal standards. Therefore, the determination of the levels of the analytes will not be affected by alterations in these experimental conditions. Additionally, the co-elution of the

analyte and its isotope-labeled standard, along with the similar fragmentation pattern of the analyte and internal standard, offers unequivocal chemical specificity for analyte identification.

Relative to the previously reported LC-MS/MS method with the use of external standards, our method also displays superior sensitivity and reproducibility. The limits of detection for 5-HmdC, 5-FodC, 5-CadC (10), and 5-HmdU were reported to be 2.5 fmol, 5 fmol, 10 fmol, and 2-3 modifications per 10^5 dC, resepectively (15). With our method, the limits of quantitation for 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU were 0.056, 0.098, 0.14 and 80 fmol, respectively (Table B.2). Along this line, with the use of external standards, 5-CadC was not detectable in DNA from HEK293T cells or most mouse organs (10), and 5-HmdU could not be detected in DNA isolated from HEK293T cells (15). We, however, were able to quantify these two modified nucleosides in all cell lines and mammalian tissues tested. Moreover, while reasonably good reproducibility was obtained for the levels of the abundant 5-HmdC in cellular or tissue DNA with the use of the previous method, relatively large variations were observed for the levels of 5-FodC between the two replicate measurements of various mouse tissue DNA samples (10). In contrast, excellent reproducibilities were achieved with our isotope-dilution method (Figure 3.3 and Table B.3).

Both endogenous ROS and Tet-mediated oxidation may contribute to the

generation of the 5-mdC oxidation products *in vivo*. Whereas no *in-vitro* experiments have been performed to assess the frequencies of ROS-induced formation of 5-CadC, a previous study showed that the Fenton-type reagent, Cu(II)/H_2O_2/ascorbate, induced the generation of 5-HmdC at a rate that is approximately 1.5 times higher than that of 5-FodC (31). Here we observed that overexpression of the catalytic domain of Tet1 in HEK293T cells led to drastic increases in the levels of 5-HmdC, 5-FodC, and 5-CadC, which is accompanied with a modest increase of 5-HmdU. In addition, we found that 5-HmdC is present at a level that is about 2-3 orders of magnitude greater than 5-FodC, and 3-4 orders of magnitude greater than 5-CadC, in brain tissues of humans and mice, and skin tissues of mice (Figure 3.3). We also observed that 5-HmdC exists at a level that is ~35-70 times greater than 5-HmdU in skin and brain tissues of mice, and ~400 times greater in human brain tissues. To our knowledge, this is the first direct quantitative comparison between the levels of multiple 5-mdC derivatives in mammalian tissues. The difference in the levels of 5-HmdU and 5-HmdC in DNA is remarkable viewing that mouse and human DNA carries at least 20 times more dT than mdC. While relatively high level of 5-HmdC in human brain tissues has been documented (32), this is the first demonstration that this modified nucleoside is also present in mouse skin at levels that are only a few times lower than that in mouse brain. Previous biochemical studies support that human TDG exhibits very high cleavage

rates for 5-FodC, 5-HmdU, 5-CadC, but extremely low cleavage rate for 5-HmdC at CpG dinucleotide site (11,14,33). Therefore, the much higher endogenous levels of 5-HmdC than the other three oxidized pyrimidine nucleosides could be attributed, at least in part, to the poor removal of 5-HmdC from the genome. Considering the recent notion that 5-HmdC, apart from being an intermediate for active cytosine demethylation, may itself serve as an epigenetic mark (34), accumulation of 5-HmdC in skin tissues may also alter the epigenetic regulation of gene expression.

References

1. Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33: 245-254.

2. Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293: 1074-1080.

3. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99: 247-257.

4. Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69: 915-926.

5. Ooi SK, Bestor TH (2008) The colorful history of active DNA demethylation. *Cell* 133: 1145-1148.

6. Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 11: 607-620.

7. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930-935.

8. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466: 1129-1133.

9. He YF, Li BZ, Li Z, Liu P, Wang Y, et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333: 1303-1307.

10. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333: 1300-1303.

11. Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286: 35334-35338.

12. Guo JU, Su Y, Zhong C, Ming GL, Song H (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145: 423-434.

13. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, et al. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146: 67-79.

14. Bennett MT, Rodgers MT, Hebert AS, Ruslander LE, Eisele L, et al. (2006) Specificity of human thymine DNA glycosylase depends on *N*-glycosidic bond stability. *J Am Chem Soc* 128: 12510-12519.

15. Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, et al. (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol* 8: 751-758.

16. Lund AH, van Lohuizen M (2004) Epigenetics and cancer. *Genes Dev* 18: 2315-2335.

17. Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, et al. (2002) LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res* 62: 4075-4080.

18. Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, et al. (2003) TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 17: 637-641.

19. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, et al. (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468: 839-843.

20. Yang H, Liu Y, Bai F, Zhang JY, Ma SH, et al. (2012) Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene*: DOI: 10.1038/onc.2012.1067.

21. Jin SG, Jiang Y, Qiu R, Rauch TA, Wang Y, et al. (2011) 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* 71: 7360-7365.

22. Hong H, Cao H, Wang Y, Wang Y (2006) Identification and quantification of a guanine-thymine intrastrand cross-link lesion induced by Cu(II)/H_2O_2/ascorbate. *Chem Res Toxicol* 19: 614-621.

23. Cao H, Wang Y (2006) Collisionally activated dissociation of protonated

2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives. *J Am Soc Mass Spectrom* 17: 1335-1341.

24. Dai Q, He C (2011) Syntheses of 5-formyl- and 5-carboxyl-dC containing DNA oligos as potential oxidation products of 5-hydroxymethylcytosine in DNA. *Org Lett* 13: 3446-3449.

25. Wang J, Clauson CL, Robbins PD, Niedernhofer LJ, Wang Y (2012) The oxidative DNA lesions 8,5'-cyclopurines accumulate with aging in a tissue-specific manner. *Aging Cell* 11: 714-716.

26. Wang J, Cao H, You C, Yuan B, Bahde R, et al. (2012) Endogenous formation and repair of oxidatively induced G[8-5m]T intrastrand cross-link lesion. *Nucleic Acids Res* 40: 7368-7374.

27. Mitra D, Luo X, Morgan A, Wang J, Hoang MP, et al. (2012) An ultraviolet-radiation-independent pathway to melanoma carcinogenesis in the red hair/fair skin background. *Nature* 491: 449-453.

28. Yuan B, Zhang J, Wang H, Xiong L, Cai Q, et al. (2011) 6-Thioguanine reactivates epigenetically silenced genes in acute lymphoblastic leukemia cells by facilitating proteasome-mediated degradation of DNMT1. *Cancer Res* 71: 1904-1911.

29. Zhang J, Yuan B, Zhang F, Xiong L, Wu J, et al. (2011) Cyclophosphamide perturbs cytosine methylation in Jurkat-T cells through LSD1-mediated stabilization of DNMT1 protein. *Chem Res Toxicol* 24: 2040-2043.

30. Wang J, Yuan B, Guerrero C, Bahde R, Gupta S, et al. (2011) Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal Chem* 83: 2201-2209.

31. Cao H, Wang Y (2007) Quantification of oxidative single-base and intrastrand cross-link lesions in unmethylated and CpG-methylated DNA induced by Fenton-type reagents. *Nucleic Acids Res* 35: 4833-4844.

32. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324: 929-930.

33. Zhang L, Lu X, Lu J, Liang H, Dai Q, et al. (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol* 8: 328-330.

34. Yildirim O, Li R, Hung JH, Chen PB, Dong X, et al. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147: 1498-1510.

Chapter 4: Detection of Oxidation Products of 5-Methyl-2'-deoxycytidine in Arabidopsis DNA

Introduction

DNA methylation at the C5 position of cytosine is a conserved epigenetic mark for transcriptional gene silencing in diverse organisms (1). While the levels of 5-methylcytosine (5-mC) are relatively low in human genomes (~4% of total cytosine), 5-mC is abundantly present in plant genomes (5-25% depending on the species) (2). In addition to the primary methylation at CG sites, cytosine in plants can also be methylated in CHG and, less frequently, in CHH sequences ('H' represents A, C or T) (3). The plant methylation patterns are established by different methyltransferase activities. *De novo* domains rearranged methyltransferases (DRMs) transfer methyl groups to completely unmethylated duplex DNA in all sequence contexts, and chromomethylase 3 (CMT3) can convert cytosine to 5-mC at non-CG sites (4). The CG methylation is propagated during mitotic cell divisions by a group of maintenance methyltransferases (MET1 in Arabidopsis) (5).

Another mechanism for the dynamic regulation of the methylation status of genes is to passively and/or actively remove 5-mC from DNA. Passive demethylation occurs when cells fail to maintain the methylation during DNA replication. In plants, a

subfamily of helix-hairpin-helix-Gly/Pro/Asp (HhH-GPD) DNA glycosylases have been identified as demethylases involved in active cytosine DNA demethylation (6). These bifunctional glycosylases remove the 5-mC base and then cleave the DNA backbone at the resulting abasic site. Subsequent action by base excision repair (BER) machinery results in the replacement of 5-mC with an unmethylated cytosine (7). Previous studies demonstrated the biological role of the glycosylase-mediated demethylation of DNA in Arabidopsis. Loss-of-function mutations in *Demeter* (*DME*), a 5-mC DNA glycosylase gene in Arabidopsis, lead to impaired endosperm and embryo development, and eventually in seed abortion (8). Hypermethylation of cytosine occurred in genomes of plants that lack members of the DNA glycosylase demethylase family, e.g. repressor of silencing 1 (ROS1), DME-like 2 (DML2), and DME-like 3 (DML3) (9).

Mammals appear to lack the activity of glycosylases that can excise 5-mC specifically. However, active DNA demethylation may also be achieved in mammals through a BER pathway by DNA glycosylases, though it requires oxidation of 5-mC as the first step (10). In this vein, 5-mC is converted into 5-hydroxymethylcytosine (5-HmC), 5-formylcytosine (5-FoC), and 5-carboxylcytosine (5-CaC) by the ten-eleven translocation (Tet) family of DNA dioxygenases through iterative oxidation (11,12). 5-FoC and 5-CaC, but not 5-HmC, at CG site can be readily removed by thymine DNA

glycosylase and replaced with unmethylated cytosine by BER proteins (13,14). This differential reactivity toward thymine DNA glycosylase may account for the recent observations that 5-FoC and 5-CaC are much less abundant than 5-HmC in mammalian genomes (12,15).

The discovery of Tet-induced oxidation of 5-mC in mammals raised questions about the possible presence of consecutive oxidations of 5-mC in plants. The mammalian TET proteins responsible for these oxidative modifications contain a catalytic domain that is typical of Fe(II)- and 2-oxoglutarate (2OG)-dependent dioxygenases, members of the cupin superfamily (16). By using computational analysis, Iyer et al. (17) reported several distinct families of Fe(II)- and 2OG-dependent dioxygenases that are likely to be involved in oxidation of 5-mC in mammals and other early branching eukaryotes such as fungi and algae. In contrast, enzymes of this family have not been identified in multicellular plants (17,18). The prediction of lack of this cytosine modifying enzymatic activity in plants was strengthened by the fact that, despite some preliminary results reporting the presence of 5-HmC (19,20), there is yet no definitive evidence supporting the presence of oxidation products of 5-mC in plants.

As it was considered that this conclusion was based on the absence of evidence rather than the evidence of absence, we set out to assess the levels of these modified bases in Arabidopsis DNA using a liquid chromatography coupled with tandem mass

spectrometry (LC-MS/MS/MS) method. In this context, we recently reported the application of LC-MS/MS/MS, along with the isotope-dilution technique, for sensitive and accurate quantification of 5-hydroxymethyl-2'-deoxycytidine (5-HmdC), 5-formyl-2'-deoxycytidine (5-FodC), 5-carboxyl-2'-deoxycytidine (5-CadC), and 5-hydroxymethyl-2'-deoxyuridine (5-HmdU) in mammalian cells and tissues (15). Herein, we measured these four modified nucleosides along with the (5'*S*) diastereomer of 8,5'-cyclo-2'-deoxyguanosine (*S*-cdG), a reliable biomarker for endogenously induced oxidative DNA damage (21), in Arabidopsis DNA. Our results demonstrated the presence of these modified nucleosides in Arabidopsis. Their relatively low levels in genomic DNA, however, suggest that, in contrast to the observations made for mammals, they are not likely to be formed from enzyme-mediated oxidation reactions.

Materials and Methods

Materials

Genomic DNA was isolated from leaves of four individual plants of *Arabidopsis thaliana* Columbia (Col-0) using a DNeasy Plant Maxi Kit (Qiagen, UK) according to the manufacturer's instructions. Nuclease P1 and phosphodiesterases 1 and 2 were from Sigma-Aldrich (St. Louis, MO). Alkaline phosphatase was obtained from New England Biolabs (Ipswich, MA), and erythro-9-(2-hydroxy-3-nonyl)adenine (EHNA) hydrochloride was from Tocris Bioscience (Ellisville, MO). [1,3-$^{15}N_2$, 2'-D]-5-HmdC,

[1,3-$^{15}N_2$-2'-D]-5-FodC, [4-*amino*-1,3-$^{15}N_3$]-5-CadC, [1,3-$^{15}N_2$-2'-D]-5-HmdU, and [1,3,7,9-$^{15}N_4$-(2-*amino*-^{15}N)]-*S*-cdG were synthesized previously (15,22-24). The chemical structures of these modified nucleosides are shown in Figure 4.1.

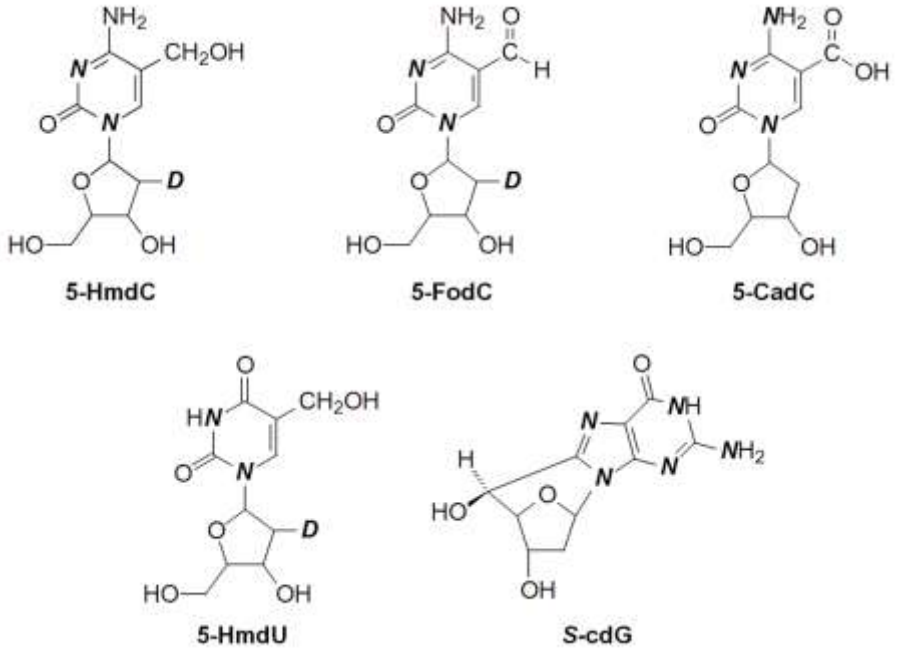


Figure 4.1. Chemical structures of the modified nucleosides measured in this study. The "N" in bold italic font represents ^{15}N, and the site of deuterium atom incorporation is indicated with a "D".

Enzymatic Digestion

Plant DNA (30 μg) was first digested by nuclease P1 (0.1 U/μg DNA) and phosphodiesterase 2 (0.00025 U/μg DNA) in a solution containing 30 mM sodium acetate (pH 5.6), 1.0 mM zinc acetate, and 1 mM EHNA, which is an inhibitor for adenine deaminase. After incubation at 37 ℃ for 48 h, Tris-HCl (pH 8.9) was added to the solution until its final concentration reached 50 mM and the samples were further

treated with alkaline phosphatase (0.05 U/μg DNA) and phosphodiesterase 1 (0.0005 U/μg DNA) at 37°C for 2 h. To the mixture were then added 90 fmol of [1,3-$^{15}N_2$, 2'-D]-5-HmdC, 30 fmol of [1,3-$^{15}N_2$-2'-D]-5-FodC, 75 fmol of [4-*amino*-1,3-$^{15}N_3$]-5-CadC, 2 pmol of [1,3-$^{15}N_2$-2'-D]-5-HmdU, and 128 fmol of [1,3,7,9-$^{15}N_4$-(2-*amino*-^{15}N)]-*S*-cdG. The enzymes were subsequently removed from the digestion mixture by chloroform extraction. The aqueous layer was subjected to off-line HPLC for the enrichment of these five target nucleosides.

HPLC Enrichment

The enrichment was carried out on a Beckman HPLC system with pump module 125 and a UV detector (module 126). A 4.6 × 250 mm Aeris Widepore C18 column (3.6 μm in particle size, Phenomenex, Torrance, CA) was used. An isocratic elution at a flow rate of 0.8 mL/min was employed with a solution of 10 mM ammonium formate (pH 8.5) as mobile phase. The HPLC fractions for 5-HmdC, 5-FodC, 5-CadC, 5-HmdU and *S*-cdG were pooled, dried, redissolved in water, and then injected for LC-MS/MS/MS analysis.

LC-MS/MS/MS Analysis

The LC-MS/MS/MS experiments were conducted using a 0.5 × 250 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies, Santa Clara, CA) and an Agilent 1200 capillary HPLC pump. A solution of 0.1% (v/v) formic acid in water

(solution A) and a solution of 0.1% (v/v) formic acid in methanol (solution B) were used as mobile phases. A gradient of 5 min 0-20% B and 25 min 20-70% B was employed for the separation of the modified nucleosides. The flow rate was 6.0 μL/min. The effluent from the LC column was directed to an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The temperatures of the ion transport tube were maintained 275 and 300 ℃ in the positive- and negative-ion modes, respectively. The electrospray, capillary and tube lens voltages were 5 kV, 4 V and 25 V, respectively, in the positive-ion mode, and 4.5 kV, −12 V and −92 V, respectively, in the negative-ion mode. The sheath gas flow rate was 15 arbitrary units, and no auxiliary gas was used. The mass spectrometer was set up to acquire the MS^3 spectra for fragmentations of the $[M+H]^+$ ions of 5-HmdC, 5-FodC, 5-CadC, and *S*-cdG, and the $[M-H]^-$ ion of 5-HmdU following previously described methods (15,24).

Results

As a first step toward exploring whether the oxidation of 5-mC plays a role in epigenetic regulation in plants, we employed our recently developed LC-MS/MS/MS coupled with the stable isotope-dilution technique (15,24) to measure the levels of 5-HmdC, 5-FodC, and 5-CadC in genomic DNA from Arabidopsis leaves. For comparison, we also quantified 5-HmdU and *S*-cdG in the same DNA samples. In this

regard, we utilized HPLC to enrich these modified nucleosides prior to LC-MS^3 analysis, as described previously (15). As shown in Figure 4.2, the targeted analytes were well resolved from each other and from the canonical ribonucleosides and 2'-deoxyribonucleosides. The identities and quantities of the aforementioned nucleosides were established from LC-MS/MS/MS measurements. The selected-ion chromatograms (SICs) and MS^3 spectra for unlabeled and labeled analytes are displayed in Figure 4.3. The fragmentation behaviors of the protonated ions of 5-HmdC, 5-FodC, 5-CadC, and *S*-cdG and the deprotonated ion of 5-HmdU as well as the calibration curves for each analyte were previously described (15,24). Briefly, collisional activation of protonated ions of the three modified 5-mdC nucleosides gave rise to cleavages of the N-glycosidic linkage and facile elimination of a 2-deoxyribose moiety. Further fragmentation of protonated ions of the nucleobase portion yielded one major fragment ion (*m/z* 142, 140, and 138 for 5-HmdC, 5-FodC, and 5-CadC, respectively) emanating from the neutral losses of H_2O for 5-HmdC and 5-CadC, and HNCO for 5-FodC. On the other hand, collisional activation of the $[M+H]^+$ ion of *S*-cdG leads to the facile cleavages of the N-glycosidic linkage and the bond between C4' and C5' of 2-deoxyribose to give the ion of *m/z* 180 in MS/MS. Further collisional activation of the *m/z*-180 ion results in the losses of NH_3, CO, and both to generate the ions of *m/z* 163, 152, and 135 in MS/MS/MS. As reported previously, 5-HmdU could

be detected with better sensitivity in the negative- than positive-ion mode (25). Upon collisional activation, the $[M-H]^-$ ion of 5-HmdU loses an HNCO from the modified nucleobase moiety to produce the most abundant ion of *m/z* 214 in MS/MS, further activation of which leads to elimination of part of the 2-deoxyribose component to give the ion of *m/z* 124 in MS^3 (15). Corresponding fragments were observed in the MS^3 of the stable isotope-labeled standards (Figure 4.3). The identical elution times in SICs and similar MS^3 spectra for the analytes and their stable isotope-labeled standards confirmed the identities of the modified nucleosides and allowed for their reliable quantification.

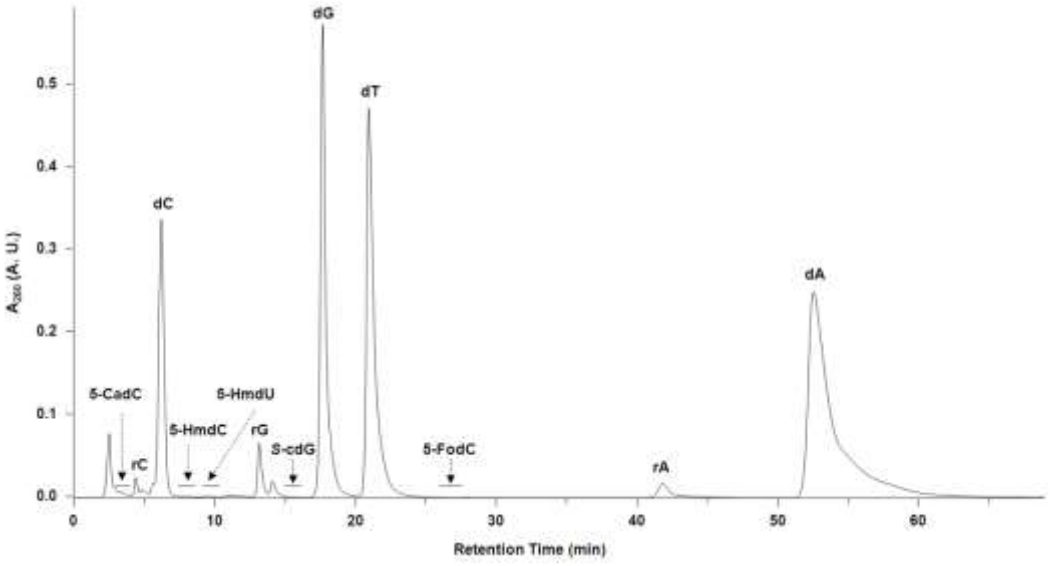


Figure 4.2. Representative HPLC trace for the enrichment of 5-HmdC, 5-FodC, 5-CadC, 5-HmdU, and *S*-cdG from the enzymatic digestion mixture of genomic DNA isolated from Arabidopsis.

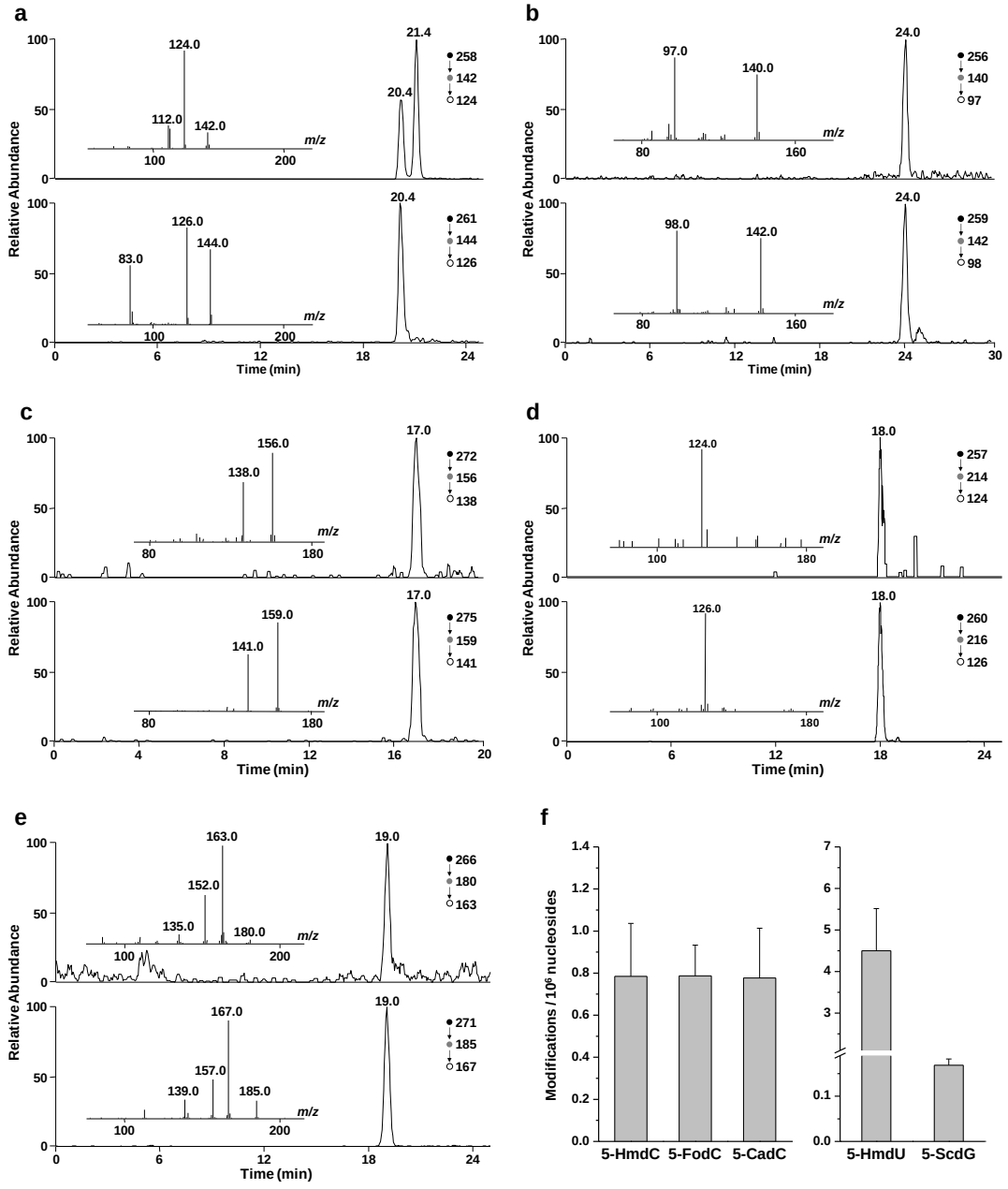


Figure 4.3. Representative LC-MS/MS/MS results for the quantification of 5-HmdC (a), 5-FodC (b), 5-CadC (c), 5-HmdU (d), *S*-cdG (e), and quantification results (f) in Arabidopsis DNA. Shown in panels A-E are the selected-ion chromatograms for monitoring the indicated transitions for the analytes (top trace) and the isotope-labeled standards (bottom trace), and the insets give the MS/MS/MS results for the analytes and internal standards.

With the use of this method, we successfully detected 5-HmdC, 5-FodC, 5-CadC, 5-HmdU, and *S*-cdG in four independent genomic DNA preparations from Arabidopsis leaves (Table 4.1). Our results revealed that 5-HmdC, 5-FodC, and 5-CadC were present at very similar levels in Arabidopsis genomic DNA (0.79, 0.79, and 0.78 per 10^6 nucleosides, respectively, Figure 4.3). In addition, we found that the levels of 5-HmdU and *S*-cdG were approximately 4.5 and 0.17 modifications per 10^6 nucleosides, respectively (Figure 4.3).

Table 4.1. Frequencies of 5-HmdC, 5-FodC, 5-CadC, 5-HmdU and *S*-cdG (in the unit of modifications per 10^6 nucleosides) in Arabidopsis genomic DNA

Nucleosides	**Numbers[*] (n=4)**
5-mdC / 100 dG	9.63 ± 0.92
dT / 100 dN	19.2 ± 0.18
5-HmdC / 10^6 dN	0.79 ± 0.25
5-FodC / 10^6 dN	0.79 ± 0.15
5-CadC / 10^6 dN	0.78 ± 0.24
5-HmdU / 10^6 dN	4.50 ± 1.01
S-cdG / 10^6 dN	0.17 ± 0.01

[*]The data represent mean ± standard deviation

Discussion

The combination of LC-MS^3 with the stable-isotope dilution technique provides a sensitive and accurate method for the measurement of oxidation products of 5-mdC, together with 5-HmdU and *S*-cdG, in Arabidopsis tissues. To our knowledge, this is the

first rigorous quantification of all three 5-mdC modification products in a flowering plant. Since the recent discoveries of Tet-mediated oxidation of 5-mdC to 5-HmdC, 5-FodC, and 5-CadC, many biophysical and biochemical techniques have been employed for their detection, including LC-MS analysis (26-29), thin layer chromatography (30), chemical derivatization followed by sequencing analysis (31,32), single molecule detection (33), antibody-based dot-blot analysis (11,20,34), etc. The development of these methods has provided valuable insights regarding the roles of these 5-mdC oxidation products in processes such as active DNA demethylation in mammals. By using a dot-blot assay, Yao et al. (20) first reported the observation of low levels of 5-HmC in Arabidopsis leaves and flowers. However, a dot-blot assay does not offer an accurate quantification of the modified nucleobase. The relatively low levels of oxidized derivatives of 5-mC in Arabidopsis require more sensitive methods for their reliable detection. With the use of our recently developed LC-MS^3 coupled with isotope-dilution method, we were able to detect these modified bases in Arabidopsis DNA. In contrast to the high levels of 5-HmdC found in mammals (~200-1600 modifications per 10^6 nucleosides), which were about 100 to 1000-fold higher than those of 5-FodC and 5-CadC, our results revealed that the Arabidopsis genome contains similar levels (approximately 0.8 modifications per 10^6 nucleosides, Figure 4.3) of 5-HmdC, 5-FodC, and 5-CadC.

While 5-HmC, 5-FoC, and 5-CaC are considered as products from Fe(II)- and 2OG-dependent dioxygenase-mediated oxidation, they could also arise from endogenous reactive oxygen specie (ROS). Thus, we also measured 5-HmdU and *S*-cdG to estimate the contribution of endogenous ROS to the formation of 5-HmdC, 5-FodC, and 5-CadC. The results showed that the levels of the three 5-mdC modification products are ~0.8 modifications per 10^6 nucleosides, whereas the level of 5-HmdU is approximately 5.7-fold higher than that of 5-HmdC. Considering that there are approximately 6.5 times more thymidine than 5-mdC nucleosides in Arabidopsis, the frequency of occurrence of 5-HmdC (on per mdC basis) is only 1.1 time as high as that of 5-HmdU (on per dT basis). Thus, at least part of 5-mdC modification products arises from ROS present in cells or from artificial oxidation of 5-mdC in DNA during various steps of sample preparation. The latter, however, may make a smaller contribution according to the results for *S*-cdG measurements. The 8,5'-cyclopurine-2'-deoxynucleosides are considered as robust biomarkers for oxidative stress. We previously measured the levels of *S*-cdG and 5-HmdC in mouse skin tissues, which were at the levels of 0.35 and 200 modifications per 10^6 nucleosides, respectively (15,35). In comparison with the results obtained in this study, the difference is huge in the level of 5-HmdC but not that of *S*-cdG between mammalian and Arabidopsis tissues. Given that the level of 5-HmdC in a plant genome is ~2–3 orders of magnitude lower

than that in mammalian tissues while 5-FodC and 5-CadC are present at similar levels (12,15,29), it is unlikely that in plants these oxidized bases serve as intermediates in active DNA demethylation. This finding is in keeping with the fact that to date no putative homologues of Fe(II)- and 2OG-dependent dioxygenase enzymes responsible for 5-mdC oxidation have been unambiguously identified in plants (17).

Reference

1. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, et al. (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* 107: 8689-8694.

2. Rangwala SH, Richards EJ (2004) The value-added genome: building and maintaining genomic cytosine methylation landscapes. *Curr Opin Genet Dev* 14: 686-691.

3. Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. *Nature* 292: 860-862.

4. Gehring M, Henikoff S (2007) DNA methylation dynamics in plant genomes. *Biochim Biophys Acta* 1769: 276-286.

5. Saze H, Mittelsten Scheid O, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* 34: 65-69.

6. Gehring M, Henikoff S (2008) DNA methylation and demethylation in Arabidopsis. *Arabidopsis Book* 6: e0102.

7. Ooi SK, Bestor TH (2008) The colorful history of active DNA demethylation. *Cell* 133: 1145-1148.

8. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, et al. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell* 110: 33-42.

9. Zhu JK (2009) Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet* 43: 143-166.

10. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, et al. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and Gadd45. *Cell* 135: 1201-1212.

11. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466: 1129-1133.

12. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333: 1300-1303.

13. He YF, Li BZ, Li Z, Liu P, Wang Y, et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333: 1303-1307.

14. Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286: 35334-35338.

15. Liu S, Wang J, Su Y, Guerrero C, Zeng Y, et al. (2013) Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res* 41: 6421-6429.

16. Dunwell JM, Purvis A, Khuri S (2004) Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 65: 7-17.

17. Iyer LM, Tahiliani M, Rao A, Aravind L (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8: 1698-1710.

18. Aravind L, Abhiman S, Iyer LM (2011) Natural history of the eukaryotic chromatin protein methylation system. *Prog Mol Biol Transl Sci* 101: 105-176.

19. Krais AM, Park YJ, Plass C, Schmeiser HH (2011) Determination of genomic 5-hydroxymethyl-2'-deoxycytidine in human DNA by capillary electrophoresis with laser induced fluorescence. *Epigenetics* 6: 560-565.

20. Yao Q, Song CX, He C, Kumaran D, Dunn JJ (2012) Heterologous expression and purification of Arabidopsis thaliana VIM1 protein: in vitro evidence for its inability to recognize hydroxymethylcytosine, a rare base in Arabidopsis DNA. *Protein Expr Purif* 83: 104-111.

21. Wang J, Clauson CL, Robbins PD, Niedernhofer LJ, Wang Y (2012) The oxidative DNA lesions 8,5'-cyclopurines accumulate with aging in a tissue-specific manner. *Aging Cell*.

22. Cao H, Wang Y (2006) Collisionally activated dissociation of protonated 2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives. *J Am Soc Mass Spectrom* 17: 1335-1341.

23. Hong H, Cao H, Wang Y (2006) Identification and quantification of a guanine-thymine intrastrand cross-link lesion induced by Cu(II)/H2O2/ascorbate.

Chem Res Toxicol 19: 614-621.

24. Wang J, Yuan B, Guerrero C, Bahde R, Gupta S, et al. (2011) Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal Chem* 83: 2201-2209.

25. Frelon S, Douki T, Ravanat JL, Pouget JP, Tornabene C, et al. (2000) High-performance liquid chromatography--tandem mass spectrometry measurement of radiation-induced base damage to isolated and cellular DNA. *Chem Res Toxicol* 13: 1002-1010.

26. Yin R, Mao SQ, Zhao B, Chong Z, Yang Y, et al. (2013) Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J Am Chem Soc* 135: 10396-10403.

27. Jin SG, Jiang Y, Qiu R, Rauch TA, Wang Y, et al. (2011) 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* 71: 7360-7365.

28. Chen ML, Shen F, Huang W, Qi JH, Wang Y, et al. (2013) Quantification of 5-methylcytosine and 5-hydroxymethylcytosine in genomic DNA from hepatocellular carcinoma tissues by capillary hydrophilic-interaction liquid chromatography/quadrupole TOF mass spectrometry. *Clin Chem* 59: 824-832.

29. Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, et al. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* 50: 7008-7012.

30. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930-935.

31. Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, et al. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* 29: 68-72.

32. Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, et al. (2013) Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153: 678-691.

33. Song CX, Clark TA, Lu XY, Kislyuk A, Dai Q, et al. (2012) Sensitive and specific

single-molecule sequencing of 5-hydroxymethylcytosine. *Nat Methods* 9: 75-77.

34. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, et al. (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468: 839-843.

35. Mitra D, Luo X, Morgan A, Wang J, Hoang MP, et al. (2012) An ultraviolet-radiation-independent pathway to melanoma carcinogenesis in the red hair/fair skin background. *Nature* 491: 449-453.

Chapter 5: Arsenite Targets the Zinc Finger Domains of Tet Proteins and Inhibits Tet-mediated Oxidation of 5-Methylcytosine

Introduction

As a naturally occurring metalloid, arsenic constitutes an important and ubiquitous threat to public health due to the prevalence of its contamination in ground water and the substance's toxic potential (1). Whilst the World Health Organization (2011) recommends a threshold of 10 μg/L for inorganic arsenic concentration in drinking water, over 100 million individuals in nearly 70 nations are exposed to excessive amounts of arsenic through drinking water and diet (2). There is a strong body of evidence linking arsenic exposure with diverse pathophysiological end points in humans, from acute toxicity to chronic diseases including cancer (3). Apart from arsenic-induced oxidative damage in DNA and perturbations of DNA repair (4), it becomes increasingly clear that arsenic accumulation also leads to epigenetic dysregulation, including perturbations of DNA methylation in the global genome and at some specific genomic loci (5,6). However, it remains unexploited how arsenic exposure influences the enzymatic removal of this pivotal epigenetic mark.

Although 5-methylcytosine (5-mC) is a stable epigenetic mark that mediates long-term gene silencing, mounting evidence indicates that DNA methylation is

reversible (7). In this vein, the recent discovery of Ten-eleven translocation (Tet) family of Fe^{2+}- and α-ketoglutarate-dependent dioxygenases shed new lights on the replication-independent active DNA demethylation in mammals. Tet enzymes can induce the iterative oxidation of 5-mC to yield the 5-hydroxymethyl, 5-formyl and 5-carboxyl derivatives of cytosine (5-hmC, 5-foC, and 5-caC) (8,9), and the resultant 5-foC and 5-caC as well as the deamination product of 5-hmC (i.e. 5-hydroxymethyluracil) can then be removed from DNA by thymine DNA glycosylase (10,11). Subsequent repair of the resulting abasic site by the base excision repair machinery restores unmethylated cytosine.

A recent structural analysis of a catalytically active form of human Tet2 revealed three zinc cations in C_3H coordination that are important for the structural integrity, substrate recognition and catalytic activity of Tet2 (12). In addition, the amino acid residues involved in zinc coordination are highly conserved in Tet proteins, and mutations in these residues are found in a number of human cancers (12). Furthermore, As(III) was found to be capable of binding to proteins harboring a C_3H- or C_4-type of zinc fingers, e.g. the promyelocytic leukemia protein (13), estrogen receptor-α (14), poly(ADP-ribose) polymerase-1 (15), DNA repair proteins (16), and some RING-finger histone E3 ubiquitin ligases (17). Hence, we reason that arsenite may interact with the zinc fingers of Tet proteins, thereby altering their conformation and perturbing

Tet-mediated oxidation of 5-mC in mammalian systems.

To test the hypothesis, we examined the interaction between As(III) and zinc finger domains of human Tet proteins *in vitro* and in mammalian cells with the applications of LC-MS, ultraviolet (UV) absorption spectrophotometry, and streptavidin agarose affinity assay. We also assessed how arsenite exposure affects the Tet-mediated oxidation of 5-mC *in vitro* and in cultured mammalian cells by determining the catalytic efficiency of Tet1 in sequentially oxidizing 5-mC in a synthetic duplex DNA and by analyzing the levels of 5-hmC in genomic DNA isolated from arsenite-treated cells. The results uncovered a novel molecular mechanism underlying arsenic-induced perturbation in epigenetic signaling, which deepened our understanding on the epigenotoxicity of arsenic.

Materials and Methods

In Vitro Arsenite Binding Assay

The peptide FPSCRCVERAGHTCEAA, derived from the first zinc finger of human Tet2 with amino acid residues 1130-1137 and 1215-1224, was obtained from Genemed Synthesis Inc. (San Antonio, TX). For mass spectrometric analysis, the lyophilized apopeptide was prepared at a concentration of 1 mM in 100 mM ammonium acetate (pH 7.2) containing 10 mM tris(2-carboxyethyl)phosphine (TCEP).

Aliquots of 100 μM peptide solution were incubated with 100 μM $NaAsO_2$ at room temperature for 1 hr. The resultant solution was diluted by 10 fold in 50% methanol and 0.1% formic acid before injection. Electrospray ionization-mass spectrometry (ESI-MS) experiments were performed on an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Samples were introduced into the inlet at 10 μL/min in 50% methanol. The voltages for the ESI source, capillary, and tube lens were maintained at 5 kV, 8.6 V, and 41.5 V, respectively. The maximum ion injection time was 50 ms and the temperature for the ion transport tube was held constant at 275 ℃ throughout the experiment.

UV Spectrophotometric Analysis of As(III)-Peptide Interaction

Aliquots of 200 μM of the aforementioned apopeptide in a solution containing 20 mM ammonium acetate (pH 7.2) and 1.0 mM TCEP were titrated with $NaAsO_2$ in increments of 20 μM over a range of 20-220 μM at room temperature. The UV absorption spectra of the resultant solutions were collected in the wavelength range of 240-370 nm on a Varian Cary 50 UV-visible spectrophotometer (Palo Alto, CA).

Streptavidin Agarose Affinity Assay and Western Blot

The streptavidin agarose affinity assay was conducted with the use of biotin-As as previously described (17). Briefly, HEK293T cells were transfected with plasmid allowing for expressing the HA-tagged catalytic domain of TET1 (HA-Tet1CD) for 24

hrs, followed by treating with 5 μM biotin-As for 2 hrs. The cells were then lysed in CelLytic^{TM} M lysis buffer supplemented with a protease inhibitor cocktail (Sigma-Aldrich). Afterwards, the cell lysates were incubated with streptavidin agarose beads at room temperature for 3 hrs. Streptavidin agarose beads were subsequently washed with 1×PBS containing 1% IGEPAL CA-630 (Sigma-Aldrich) for 3 times and resuspended in SDS-PAGE loading buffer.

The proteins on the beads were resolved by SDS-PAGE and subsequently transferred to a nitrocellulose membrane using a solution containing 10 mM $NaHCO_3$, 3 mM Na_2CO_3, and 20% methanol. The membranes were blocked with 5% non-fat milk at room temperature for 2 hrs and then incubated with primary anti-HA antibody (1:10000 dilution, Sigma-Aldrich) in PBS-T at 4℃ overnight, washed for 5 times in PBS-T (10 min each), and incubated with secondary anti-rabbit antibody (1:10000) in PBS-T for 1 hr, which was followed with five washes in PBS-T. The secondary antibody was detected by using ECL Advance Western Blotting Detection Kit (GE Healthcare) and visualized with Hyblot CL autoradiography film (Denville Scientific, Inc., Metuchen, NJ). Similar Western blot analysis was carried out for HA-Tet1CD-overexpressing HEK293T cells following a 24-hr treatment with 0-5 μM $NaAsO_2$.

Biochemical Assay of Tet1-Mediated Oxidation of 5-mC in DNA

The 5-mC oxidation assay was performed in triplicate by using a 5-mC-containing

duplex DNA d(AGCTC^{m}CGGTCA)/d(GTGACCGGAGCTG) and purified mouse Tet1 protein included in a 5mC Tet1 Oxidation Kit (Wisegene, IL, USA). The reaction was initiated by incubating 0.15 μL of the Tet1 enzyme with 10 pmol of the duplex DNA at 37 ℃ in a buffer that contained 50 mM HEPES (pH 8.0), 10 mM NaCl, 2 mM ascorbic acid, 1 mM 2-oxoglutarate, 50 μM $(NH_4)_2Fe(SO_4)_2$, 1 mM ATP, and different concentrations of $NaAsO_2$ (i.e. 0, 2, and 5 μM). After reaction for 10 and 30 min, the mixtures were immediately frozen on dry ice.

Prior to LC-MS analysis, the enzyme in the reaction mixture was removed by chloroform extraction. The aqueous layer was subjected to LC-MS analyses on an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) operated in the higher resolution "ultra-zoom-scan" mode, and an Agilent 1200 capillary HPLC (Agilent Technologies, Santa Clara, CA) was used. The separation was carried out on a 0.5 × 250 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies, Santa Clara, CA). The flow rate was 8.0 μL/min, and a gradient of 5 min of 0-20% methanol followed by 35 min of 20-40% methanol in 400 mM 1,1,1,3,3,3-hexafluoro-2-propanol (pH adjusted to 7.0 with triethylamine) was employed. The mass spectrometer was operated in the negative-ion mode to monitor the $[M - 3H]^{3-}$ ions of the initially methylated 11-mer DNA as well as the corresponding products with the 5-mC being oxidized to 5-hmC, 5-foC, and 5-caC. The

ESI source, capillary and tube lens voltages were 4.0 kV, -50 V and -75 V, respectively, and the temperature for the ion transport tube was 300 ℃. Quantification of the oxidation products was conducted based on the intensities of the monoisotopic peak and the M+1 isotopic peak representing the $[M - 3H]^{3-}$ ions of 5-mC-, 5-hmC-, 5-foC-, and 5-caC-containing 11mer DNA. Because the isotopic peaks of the 5-hmC- and 5-foC-bearing DNA partly overlap, the total intensities of the M+2 and M+3 isotopic peaks of the 5-foC-harboring DNA were calculated following its isotopic abundance ratios of M, M+1, M+2, and M+3 (i.e. 73.9:100:76.8:42.9 for an elemental composition of $C_{107}H_{135}N_{41}O_{65}P_{10}$) and subtracted from the intensities of the monoisotopic and M+1 isotopic peaks of the 5-hmC-carrying DNA. The amounts of 5-hmC, 5-foC and 5-caC produced and 5-mC remained in the reaction mixture were expressed as the percentages of the peak intensity for each cytosine modification-harboring 11mer DNA over the total peak intensities of DNA containing all four modifications.

Synthesis of [$^{13}C_5$]-5-methyl-2′-deoxycytidine ([$^{13}C_5$]-5-mdC)

5-methyl-N^4-benzoylcytosine was synthesized following the established procedures for the preparation of N^4-benzoylcytosine (18). Additionally, 3′,5′-di-*O*-acetyl-[$^{13}C_{10}$-$^{15}N_2$]-thymidine was synthesized following our previously established procedures (19). 3′,5′-di-*O*-acetyl-[$^{13}C_{10}$-$^{15}N_2$]-thymidine (5.0 mg, 0.015 mmol, Cambridge Isotope Laboratories) and 5-methyl-N^4-benzoylcytosine (13.3 mg, 0.040

mmol) were added to a 15-mL flask and purged with N_2 gas. Anhydrous acetonitrile (1 mL) was added, and after 10 min of stirring, bis(trimethylsilyl)acetamide (20 μL, 0.070 mmol, Sigma-Aldrich) was added to the reaction mixture. The solution was then heated to 70°C. After 15 min, trimethylsilyl triflate (4 μL, 0.020 mmol, Sigma-Aldrich) was subsequently added and the reaction was maintained at 70°C for 4 hr. The reaction mixture was cooled, removed of solvent, and dried under vacuum. The crude [$^{13}C_5$]-3′,5′-di-*O*-acetyl-5-methyl-2′-deoxycytidine was purged with N_2 gas, and 2 M ammonia in methanol (4 mL, Sigma-Aldrich) was subsequently added. The reaction was heated to 40°C and maintained at that temperature overnight. Ammonium hydroxide solution (30%, 4 mL) was added to the reaction mixture, and the solution was stirred at 40°C for 72 hr. The solvent was removed and [$^{13}C_5$]-5-methyl-2′-deoxycytidine was purified by HPLC. The identity of [$^{13}C_5$]-5-methyl-2′-deoxycytidine was confirmed by LC-MS/MS analysis (Figure C.2).

Cell Culture, Transfection and $NaAsO_2$ treatment

The HEK293T human embryonic kidney cells were cultured in Dulbecco's Modified Eagle Medium (ATCC) supplemented with 10% fetal bovine serum and 100 IU/mL penicillin. The mouse embryonic stem cells were cultured in High Glucose Dulbecco's Modified Eagle Medium (Life Technologies), supplemented with 10% ES-qualified fetal bovine serum (Atlanta Biologicals), 100 IU/mL

penicillin-streptomycin, MEM Non-Essential Amino Acids (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma), and 60,000 units of Leukemia Inhibitory Factor (Stemgent). Culture vessels were pre-treated with 0.1% gelatin (STEMCELL Technologies) for 5-10 min prior to its removal and cell plating. All cells were maintained in a humidified atmosphere with 5% CO_2 at 37 ℃ and media renewed 2-3 times a week according to cell densities. For plasmid transfection, cells were transferred to the same media as described above without the addition of penicillin.

HEK293T cells, seeded in 6-well plates at a density of ~3×10^5 cells per well, were transfected individually with 1.5 μg expression vectors carrying the catalytic domains of human Tet1 (amino acids 1418-2136), human Tet2 (amino acids 1129-2002), or mouse Tet3 (amino acids 697-1669) using Lipofectamine 2000 (Invitrogen), as described previously (20). The media were discarded at 6-hr following transfection, and fresh media containing $NaAsO_2$ at final concentrations of 1, 2, and 5 μM were subsequently added. The cells were cultured for 24 hrs and harvested for DNA extraction. Following the similar procedures, mESCs were seeded in 6-well plates, treated with 0, 1, 2, and 5 μM $NaAsO_2$ for 24 hrs, and harvested for DNA extraction.

DNA Extraction and Enzymatic Digestion

Total genomic DNA was isolated from treated cells using a previously published high-salt method (21). The resulting DNA sample (1 μg) was treated with nuclease P1

(0.5 U) and phosphodiesterase 2 (0.00025 U) at 37 ℃ for 12 hrs in a digestion buffer (30 mM sodium acetate, 1 mM zinc acetate, 1 mM EHNA, pH 5.6), followed by the addition of alkaline phosphatase (0.05 U) and phosphodiesterase 1 (0.0005 U) in 50 mM Tris-HCl (pH 8.9) for 2 hrs. After being neutralized by formic acid, the enzymatic digestion mixture of 50-300 ng genomic DNA was mixed with 48 fmol of [1,3-$^{15}N_2$-2′-D]-5-hmdC, 600 fmol of [$^{13}C_5$]-5-mdC, and 16.2 pmol of [$^{15}N_5$]-dG. The samples were subsequently extracted with chloroform and the aqueous layer was subjected to LC-tandem MS analysis.

LC-Tandem MS Quantification of 5-mdC, dG and 5-hmdC

Measurements of 5-mdC, dG, and 5-hmdC were conducted on an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an Agilent 1200 capillary HPLC (Agilent Technologies, Santa Clara, CA). The mass spectrometer was operated in the positive-ion mode, where the MS/MS for the $[M+H]^+$ ions of 5-mdC (*m/z* 242) and [$^{13}C_5$]-5-mdC (*m/z* 247) were acquired, and the MS/MS/MS arising from the further cleavages of the $[M+H]^+$ ions of the nucleobase portions of dG (*m/z* 152), [$^{15}N_5$]-dG (*m/z* 157), 5-hmdC (*m/z* 142) and [1,3-$^{15}N_2$, 2′-D]-5-hmdC (*m/z* 144) were recorded (Figures C.3-5). The HPLC separation was carried out on a 0.5 × 250 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies, Santa Clara, CA) at a flow rate of 8.0 μL/min. A solution of 2 mM

ammonium bicarbonate (pH 7.0, solution A) and methanol (solution B) were used as mobile phases, and a gradient of 5 min 0-20% B followed by 25 min 20-70% B was employed for the separation. MS settings were as follows: electrospray voltage, 5 kV; capillary temperature, 275 ℃; capillary voltage, 38 V; tube lens voltage, 60 V; sheath gas flow rate, 15 arbitrary units. The calibration curve for 5-hmdC was constructed previously (22), and the calibration curves for 5-mdC and dG were generated by spiking 600 fmol and 16.2 pmol of [$^{13}C_5$]-5mdC and [$^{15}N_5$]-dG with different amounts of 5-mdC and dG, respectively (Figure C.6). The numbers of moles of 5-mdC, dG, and 5-hmdC in each sample were calculated from the peak area ratios versus the molar ratios of the unlabeled analyte to the corresponding stable isotope-labeled standard. The levels of 5-hmdC and 5-mdC were expressed as the percentages of dG.

Results

Arsenite Binds to the Zinc Finger Domains of Tet Proteins In Vitro and In Cells

To test whether $NaAsO_2$ can bind directly to the zinc finger domains of Tet proteins, we first performed a mass spectrometry (MS)-based *in vitro* binding assay using a synthetic peptide derived from the first zinc finger motif of human Tet2 protein, which carries a.a. 1130-1137 and a.a. 1216-1224. Positive-ion electrospray ionization-mass spectrometry (ESI-MS) revealed the $[M + 2H]^{2+}$ ion of the apopeptide

at *m/z* 919 (Figure 5.1a). Incubation of the apopeptide with As(III) leads to a +36 *m/z* shift (Figure 5.1a), which corresponds to the binding with one As(III) after the release of three protons from the Cys residues in the Tet2 peptide. In accordance with the mass increase for coordination of As(III) with other C_3H-type zinc finger proteins (23,24), the MS result revealed that arsenite can occupy the peptide derived from the first zinc binding site of Tet2. The coordination of As(III) with thiol was also confirmed by UV spectrophotometric analysis. In the UV absorption spectra, titrations of arsenite into Tet2 peptide-containing solution led to increased absorbance in the wavelength range of 240-370 nm (Figure 5.1b), indicating the new charge-transfer electronic transitions occurring in the near UV range upon As-S bond formation (25).

To further exploit the potential binding of arsenite to Tet proteins in mammalian cells, we treated HEK293T cells with a *p*-aminophenylarsenoxide-conjugated biotin probe (17,26), and assessed its interaction with ectopically expressed Tet1 by streptavidin agarose affinity assay. The Western blot result showed that biotin-As probe pulled down Tet1 protein that was overexpressed in mammalian cells (Figure 5.1c), confirming Tet1 as an As(III)-binding protein. In a control experiment, we failed to pull down Tet1 without the addition of the biotin-As probe (Figure 5.1c).

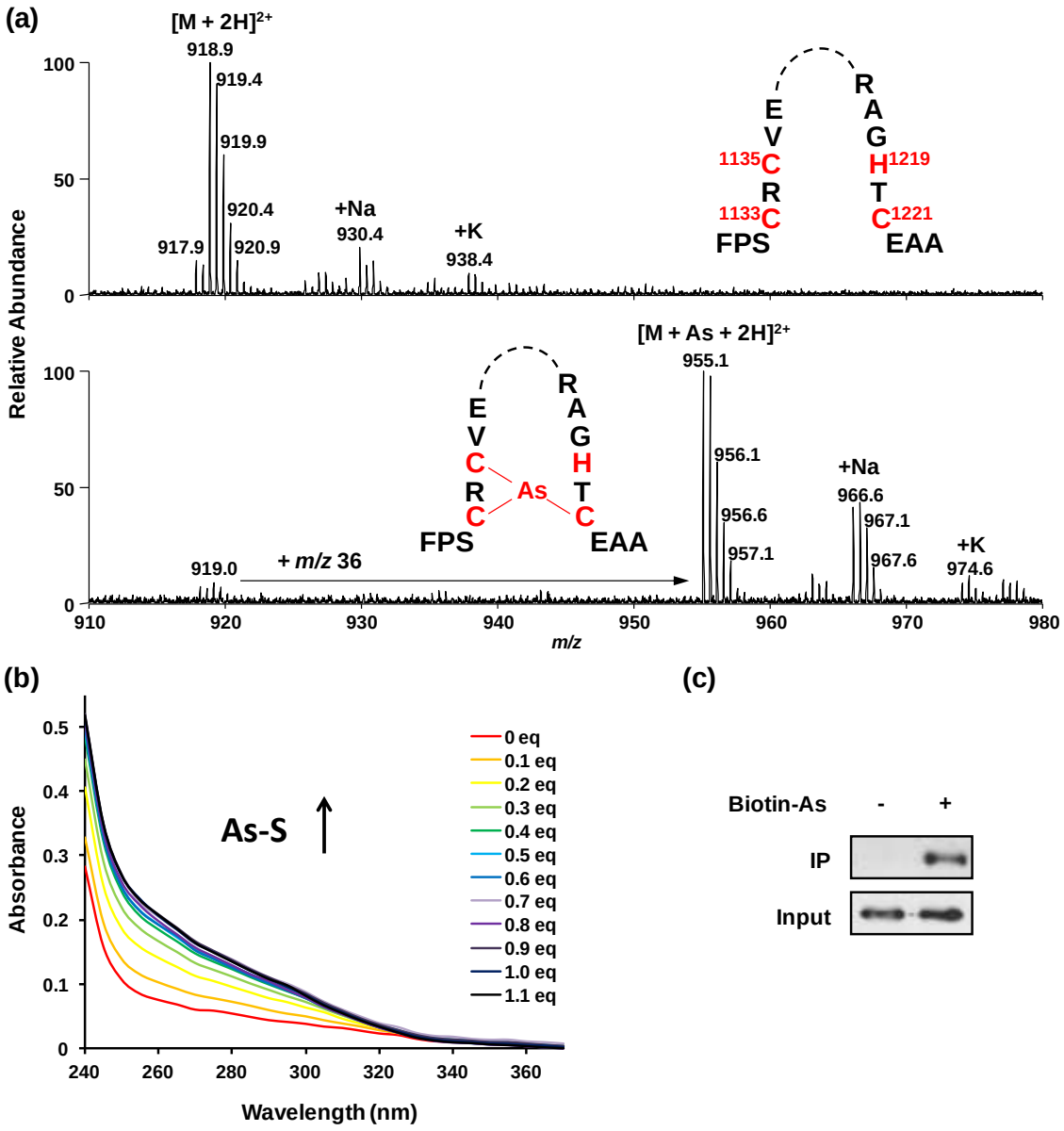


Figure 5.1. The binding behavior between $NaAsO_2$ and human Tet proteins. (a) Higher-resolution "ultra-zoom-scan" ESI-MS showing the $[M + 2H]^{2+}$ ions of a synthetic peptide derived from human Tet2 in the absence or presence of As(III). (b) UV absorption spectra of the synthetic Tet2 peptide titrated with increasing concentrations of arsenite. (c) Streptavidin agarose affinity pull-down assay using biotin-As as a probe revealed the binding between As(III) with Tet1 in cells.

Arsenite Suppresses the Sequential Oxidation Of 5-mC in a Synthetic Duplex DNA

Given that mammalian Tet proteins could induce sequential oxidation of 5-mC to yield 5-hmC, 5-foC, and 5-caC (8-10), we next assessed how binding to arsenite affects the catalytic efficiency of Tet1 in oxidizing 5-mC in duplex DNA. To this end, we conducted an *in vitro* Tet reaction in the presence of different concentrations of arsenite, where a duplex DNA d(AGCTCXGGTCA)/d(GTGACCGGAGCTG) (X=5-mC) was employed as the substrate. The compositions of the resulting reaction mixture were characterized and quantified by LC-ESI-MS analysis (Figures 5.2 and C.1). In the absence of arsenite, the MS results revealed a rapid loss of approximately 83.2% 5-mC in 10 min and an additional decrease of 8.6% in 30 min (Figure 5.3), suggesting the high activity of Tet1 towards oxidizing 5-mC in duplex DNA. Consistently, we observed rapid formation of the three oxidized 5-mC derivatives. At 10 min, the levels of duplex DNA carrying 5-hmC, 5-foC, and 5-caC were increased by 42%, 26%, and 15%, respectively; and at the later time point (30 min), an accumulation of 57% 5-caC was observed, which was accompanied with drops in the levels of 5-hmC and 5-foC by 2.7- and 1.4-fold, respectively. In contrast to the control reactions, the addition of 2 μM arsenite led to an inhibition of Tet1 activity, as reflected by a nearly 50% reduction in the production of 5-caC at 30-min. Such a retardation in Tet1-mediated oxidation of 5-mC to 5-caC was exacerbated when a higher concentration of arsenite (5 μM) was

present in the reaction mixture, where 5-caC was produced at a level that was about one third of its yield in the control reaction. Together, the MS data indicated that $NaAsO_2$ exerted significant inhibitory effects on the oxidation of 5-mC in duplex DNA by Tet enzymes.

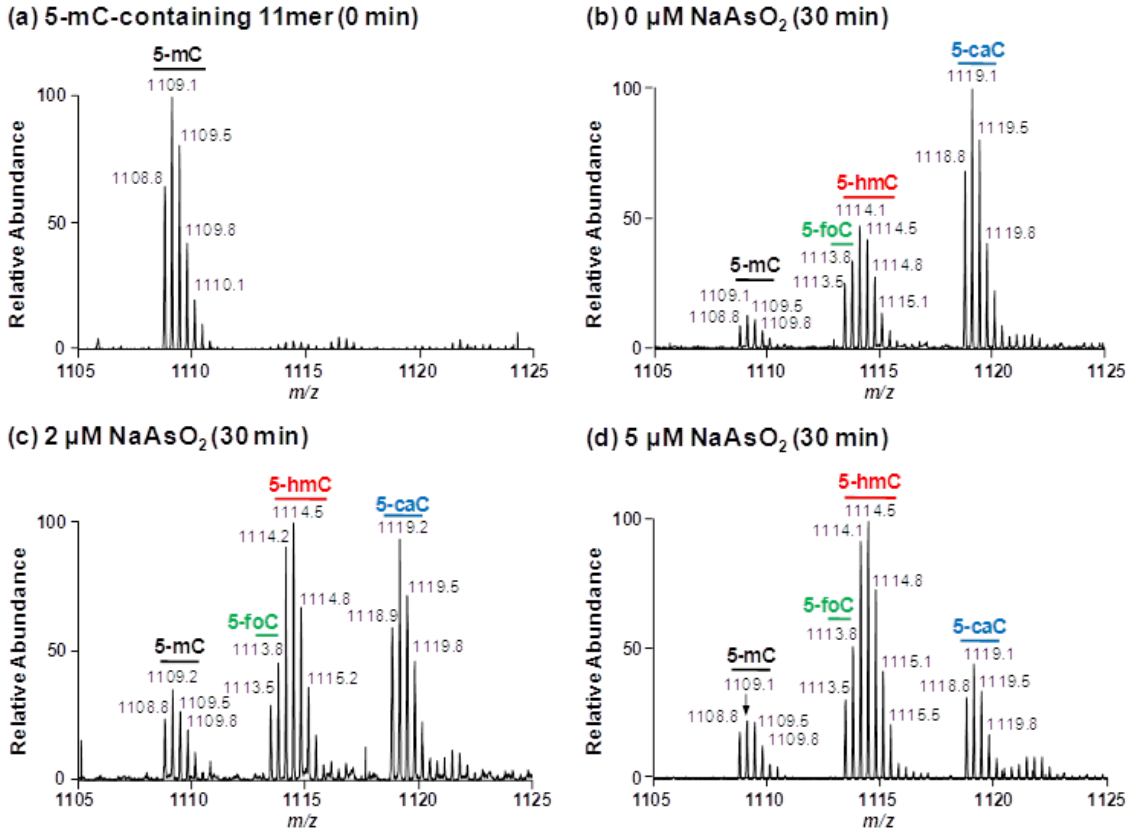


Figure 5.2. LC-ESI-MS for monitoring the Tet1-mediated oxidation of 5-mC in a duplex DNA, d(AGCTCmCGGTCA)/d(GTGACCGGAGCTG), at 0 and 30 min. Shown are the higher resolution "ultra-zoom-scan" ESI-MS results for monitoring the $[M - 3H]^{3-}$ ions of the initial 5-mC-bearing 11mer DNA (a) and its corresponding products with the 5-mC being oxidized to 5-hmC (b), 5-foC (c), and 5-caC (d).

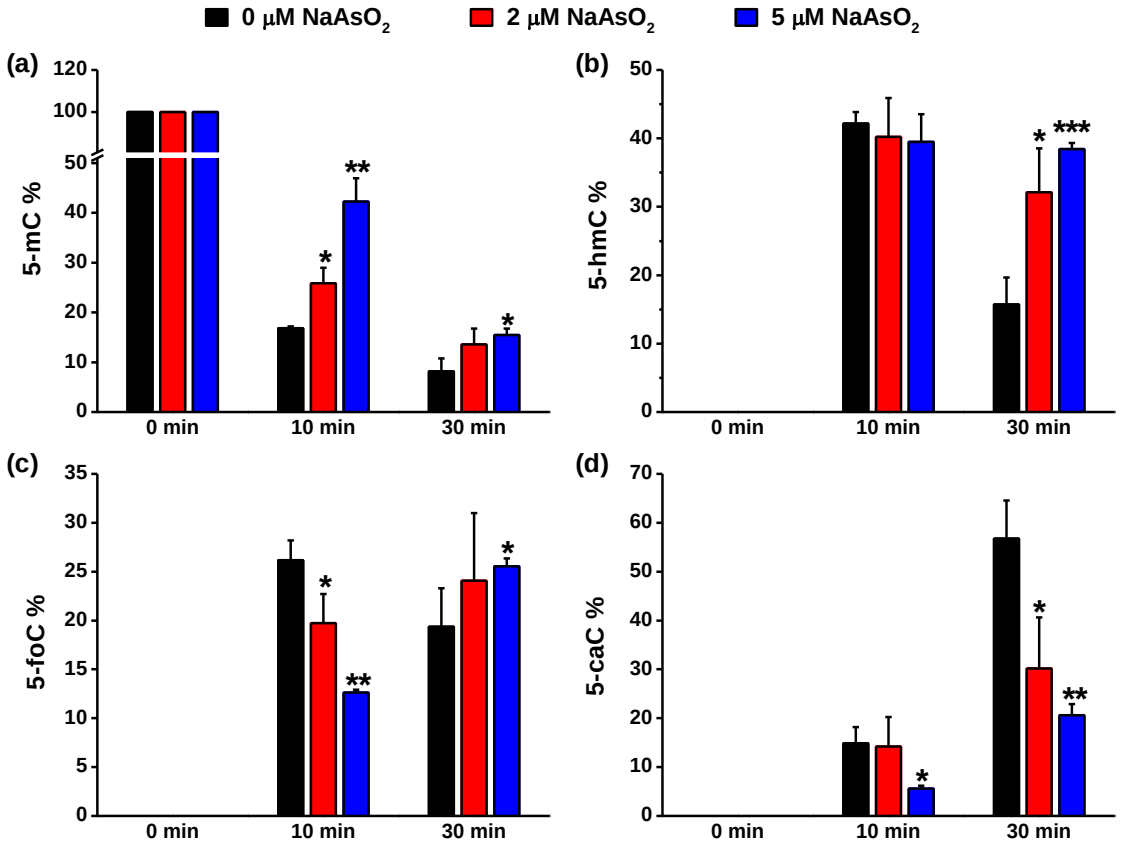


Figure 5.3. Time- and dose-dependent alterations of 5-mC (a) and its oxidation products, i.e. 5-hmC (b), 5-foC (c), and 5-caC (d), in duplex DNA, d(AGCTC^{m}CGGTCA)/d(GTGACCGGAGCTG). The peak intensity for each modification-carrying DNA was quantified by LC-MS (Figures 5.2 and C.1) and expressed as a percentage of the total peak intensities for DNA harboring all the four types of modifications (n=3). The *p* values were calculated using unpaired two-tailed Student's *t*-test, and the asterisks designate significant differences between arsenite treatments and untreated controls at the same time point (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

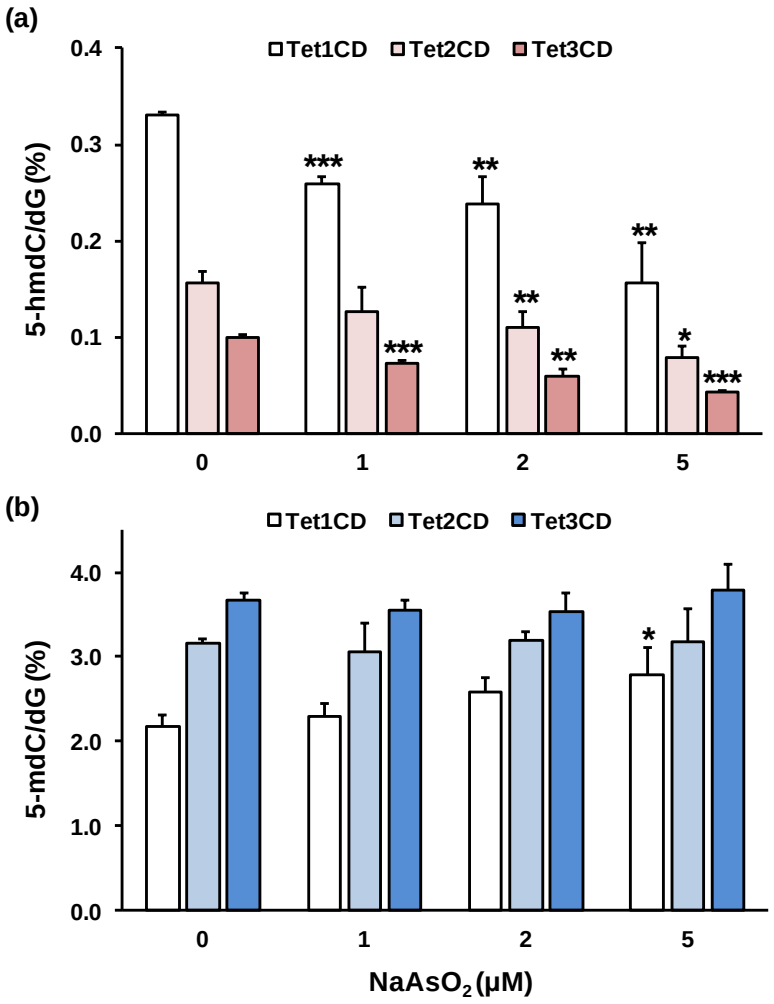


Figure 5.4. $NaAsO_2$ inhibits the formation of 5-hmdC in HEK293T cells individually overexpressing the catalytic domains of Tet1, Tet2, and Tet3 (Tet1CD, Tet2CD, and Tet3CD). Shown are the levels of 5-hmdC (a) and 5-mdC (b) in cells treated with 0, 1, 2, and 5 μM of $NaAsO_2$ (n=3). The *p* values were calculated using unpaired two-tailed Student's *t*-test, and the asterisks designate significant differences between arsenite treatments and untreated controls (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Arsenite Inhibits Tet-mediated Hydroxylation of 5-mdC in Mammalian Cells

Having demonstrated that arsenite suppresses the enzymatic activity of Tet proteins toward 5-mdC oxidation *in vitro*, we next examined how $NaAsO_2$ exposure perturbs the levels of 5-mdC and its oxidation product 5-hmdC in HEK293T cells individually overexpressing the catalytic domains of the three Tet proteins, namely Tet1CD, Tet2CD, and Tet3CD. In this vein, we did not assess if arsenite exposure alters the levels of 5-fodC and 5-cadC in cellular DNA because of the facile cleavage of these two modified nucleosides from DNA by thymine DNA glycosylase (10,27). After arsenic treatments, we isolated genomic DNA from the cells, digested the DNA with enzymes, and subjected the resulting nucleoside mixture to LC-MS/MS for the simultaneous quantification of 5-methyl-2′-deoxycytidine (5-mdC), 2'-deoxyguanosine (dG), and 5-hydroxymethyl-2'-deoxycytidine (5-hmdC). The levels of 5-mdC and 5-hmdC were expressed as the percentages of dG. The identical elution time and similar tandem mass spectra of the analytes and their respective stable isotope-labeled standards facilitated the unambiguous identification and reliable quantification of these nucleosides in the digestion mixture of cellular DNA (Figures C.3-5). Our quantification results revealed that HEK293T cells with ectopic expression of Tet1CD displayed a high level of 5-hmdC (~0.33% dG), whereas 24-hrtreatments with 1, 2, and 5 μM $NaAsO_2$ diminished the levels of 5-hmdC by 1.3-, 1.4-, and 2.1-fold, respectively

(Figure 5.4a). Moreover, we found that the arsenite-induced dose-dependent decrease in the yield of 5-hmdC was accompanied with an elevated level of 5-mdC in genomic DNA (up to a 28% increase at the highest dose) in HEK293T cells ectopically expressing Tet1CD (Figure 5.4b), suggesting that arsenite inhibited the conversion of 5-mdC to 5-hmdC in cells. Likewise, exposure to arsenite resulted in dose-dependent reductions in the levels of 5-hmdC, but not 5-mdC, in HEK293T cells ectopically expressing the catalytic domains of Tet2 and Tet3 (Figure 5.4). Additionally, Western blot result showed that arsenite treatments did not alter significantly the expression level of Tet1CD in HEK293T cells (Figure C.7). Hence, we inferred that the reduced 5-hmdC formation is attributed to impaired catalytic function of Tet enzymes, rather than decreased expression at the protein levels.

We next asked whether similar findings can be extended to mammalian cells without ectopic expression of Tet proteins. It turned out that arsenite exposure did not alter substantially the oxidation of 5-mC to 5-hmC in HEK293T cells without the overexpression of any Tet proteins (Figure C.8). We reason that mouse embryonic stem cells (mESCs) with relatively high levels of Tet-induced oxidation products (28) could be a unique system for studying the impact of arsenic on Tet-mediated oxidation of 5-mdC. Indeed the MS-based quantification data revealed substantial decreases in the levels of 5-hmdC in mESCs treated with $NaAsO_2$, where the levels of 5-hmdC were

0.092, 0.096, 0.070, and 0.063 modifications per 100 dG in cells treated with $NaAsO_2$ at concentrations of 0, 1, 2, and 5 μM, respectively (Figure 5.5a). The drop in the levels of 5-hmdC, particularly in mESCs exposed to 2-5 μM arsenite, was in line with the inhibitory effect of arsenite on 5-hmdC generation in cells ectopically expressing Tet proteins (Figure 5.4a). We also found that the same treatments did not result in statistically significant decreases in the level of 5-mdC (Figure 5.5b), indicating that 5-hmdC diminution was induced by the suppressed enzymatic activity of Tet proteins, but not the reduced level of 5-mC.

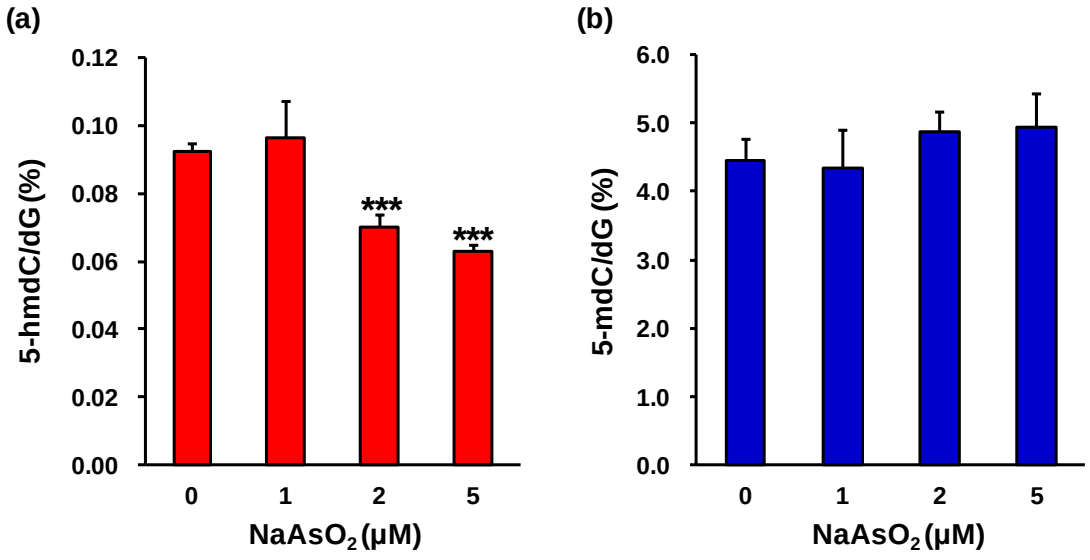


Figure 5.5. $NaAsO_2$ suppresses the formation of 5-hmdC in mouse embryonic stem cells. Displayed are the levels of 5-hmdC (a) and 5-mdC (b) in cells treated with 0, 1, 2, and 5 μM of $NaAsO_2$ (n=3). The *p* values were calculated using unpaired two-tailed *t*-test, and the asterisks designate significant differences between arsenite treatments and untreated controls (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Discussion

Role of Zinc Fingers in the Toxicity of Arsenic and Other Metal Ions

Since As(III) can selectively bind to C_3H- and C_4-type zinc fingers with high affinity (29-31), the interaction between arsenic and zinc finger proteins plays a crucial role in the toxicity and carcinogenesis of the metalloid. Binding of As(III) to the zinc finger domain of poly (ADP-ribose) polymerase 1 (PARP-1) and xeroderma pigmentosum complementation group A (XPA) proteins is known to interfere with DNA repair (32,33). In addition, substitution of Zn(II) with As(III) in the RING finger motif of RNF20-RNF40 heterodimer, a RING finger E3 ubiquitin ligase, was found to impair the ubiquitination of lysine 120 in histone H2B, which gives rise to a chromatin environment that is less biochemically accessible and inhibits DNA double strand break repair (17). Our present study demonstrated that arsenite can bind to the zinc finger motifs of Tet proteins. In this vein, zinc fingers of Tet2 are highly conserved and are essential for the structural integrity, DNA interaction and catalytic activity of the three Tet proteins (12). Therefore, our discovery of diminished Tet activities arising from arsenic binding is closely associated with the disruption of the coordination sphere in the zinc finger environment and consequent perturbation of the zinc finger function. The results also supported that the zinc fingers of Tet proteins serve an essential and critical function in oxidizing 5-mC. Apart from arsenic, other metals and metalloids

with high affinities to thiolates can also compromise zinc fingers by displacing zinc and disrupting the conformations of the zinc finger proteins. For instance, cadmium, cobalt, nickel, and antimony ions were found to cause zinc ion release from the zinc finger motif of XPA protein (34-36). Similarly, the displacement of zinc from zinc finger structures by cadmium and mercury also resulted in the inactivation of bacterial formamidopyrimidine-DNA glycosylase (Fpg), a DNA repair enzyme involved in base excision repair (37). In this vein, owing to the metal binding properties of zinc fingers (16), it can be envisaged that Tet proteins may also be susceptible to inhibition by a variety of carcinogenic metal ions.

Disturbance to DNA Methylation Fidelity by Arsenic

As a milestone in epigenetic research, the recent discovery of Tet family proteins revealed a plausible DNA demethylation process which can be initiated through Tet-mediated successive oxidation of 5-mC to 5-hmC, 5-foC, and 5-caC (8,9). The functions of Tet proteins necessitate their binding with metal ions [i.e. Zn(II) and Fe(II)] and co-substrate (i.e. α-ketoglutarate). It has been demonstrated that some cancer-associated mutations in isocitrate dehydrogenases 1 and 2 (IDH1/2) and the resulting aberrant accumulation of (*R*)-2-hydroxyglutarate could lead to inhibition of Tet enzymes by occupying the α-ketoglutarate binding site, which contributes to diminished levels of 5-hmC and global genome hypermethylation in some cancer cells

(38,39). It was also observed that the catalytic functions of Tet proteins could be stimulated by vitamin C (40-42). Here we found that arsenite could bind to the zinc finger domain of Tet proteins and inhibit the formation of the three oxidized derivatives of 5-mC *in vitro*, and consistently, it compromised the catalytic efficiency of Tet enzymes in cells and led to a substantial loss of 5-hmC in the global genome of cultured mammalian cells. 5-hmC has now been widely accepted as the sixth base owing to its ubiquitous presence in mammalian DNA (43,44). Aside from being an intermediate in active DNA demethylation, 5-hmC can be recognized by specific cellular proteins (45,46), and thus acts as an epigenetic mark on its own right and regulates gene transcription (47). Additionally, changes in the levels of 5-hmC are often accompanied with human diseases, especially cancer (47,48). Therefore, the arsenite-induced alteration of this modification provides a novel mechanism pertaining to carcinogenic effects of arsenic exposure. It is also noteworthy that, failure in efficiently oxidizing 5-mC to 5-hmC may lead to an enrichment of 5-mC in the genome. In this regard, DNA hypermethylation has been observed in the global genome in various human cell lines and human peripheral blood leukocytes after arsenic exposure, and targeted hypermethylation was also found in the promoter regions of some tumor suppressor genes (49). Hence, arsenite-induced impairment of the functions of Tet proteins could confer multiple risks on the control of DNA methylation fidelity and

exacerbate its perturbation of the epigenome.

Implications in Environmental Exposure to Arsenic

While the currently permitted concentration of arsenic in water is 10 μg/L, millions of people worldwide are exposed to arsenic at concentrations that are higher than this limit. Arsenic concentrations of up to 800-7550 μg/L have been detected in drinking water in many countries including Argentina, Bangladesh, Chile, China, India, Mexico, and the United States (3). Even when the levels of arsenic are low and considered safe by the EPA, the impact of long-term cumulative arsenic exposure on public health is substantial (3). For instance, Zierold et al. (50) reported that individuals who drink well water with low concentrations of arsenic (2-10 μg/L) for more than 20 years were significantly more likely to exhibit depressive symptoms than those drinking water with less than 2 μg/L arsenic. On the basis of these studies, although some low-concentration treatments in our work did not result in statistically significant changes in the level of 5-hmdC within 24 hrs, it cannot be ruled out that chronic and/or excessive arsenic exposure may have pronounced effect in inhibiting the activities of mammalian Tet proteins.

From our cellular experiment, the notably dose-dependent accumulation of 5-mC observed in Tet1CD-, but not Tet2CD- and Tet3CD-overexpressing cells can be explained by the expression of a relatively high level of Tet1CD. This result indicates a

potentially high correlation between Tet expression and arsenic toxicity. Viewing that 5-hmC is relatively abundant in the brain and that its presence is Tet protein-dependent, arsenic may have a greater impact on epigenetic regulation in this organ. While exposure to arsenic has been shown to induce neurotoxicity in children and adults (51), our study suggests a link between arsenite-induced aberrant cytosine methylation/ hydroxymethylation and cognitive/mental disorders. In addition, because Tet enzymes and 5-hmC are present in various embryonic and adult cell types, and play crucial roles in pluripotency as well as embryonic and adult development (52), the present work sets the stage for future studies on the roles of arsenic interaction with Tet proteins in perturbing the methylation patterns during development.

References

1. Hunt KM, Srivastava RK, Elmets CA, Athar M (2014) The mechanistic basis of arsenicosis: pathogenesis of skin cancer. *Cancer Lett* 354: 211-219.

2. Martinez VD, Vucic EA, Becker-Santos DD, Gil L, Lam WL (2011) Arsenic exposure and the induction of human cancers. *J Toxicol* 2011: 431287.

3. Naujokas MF, Anderson B, Ahsan H, Aposhian HV, Graziano JH, et al. (2013) The broad scope of health effects from chronic arsenic exposure: update on a worldwide public health problem. *Environ Health Perspect* 121: 295-302.

4. Watanabe T, Hirano S (2013) Metabolism of arsenic and its toxicological relevance. *Arch Toxicol* 87: 969-979.

5. Ren X, McHale CM, Skibola CF, Smith AH, Smith MT, et al. (2011) An emerging role for epigenetic dysregulation in arsenic toxicity and carcinogenesis. *Environ Health Perspect* 119: 11-19.

6. Bailey KA, Wu MC, Ward WO, Smeester L, Rager JE, et al. (2013) Arsenic and the epigenome: interindividual differences in arsenic metabolism related to distinct patterns of DNA methylation. *J Biochem Mol Toxicol* 27: 106-115.

7. Shen L, Song CX, He C, Zhang Y (2014) Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu Rev Biochem* 83: 585-614.

8. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333: 1300-1303.

9. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466: 1129-1133.

10. He YF, Li BZ, Li Z, Liu P, Wang Y, et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333: 1303-1307.

11. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, et al. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146: 67-79.

12. Hu L, Li Z, Cheng J, Rao Q, Gong W, et al. (2013) Crystal structure of TET2-DNA

complex: insight into TET-mediated 5mC oxidation. *Cell* 155: 1545-1555.

13. Zhang XW, Yan XJ, Zhou ZR, Yang FF, Wu ZY, et al. (2010) Arsenic trioxide controls the fate of the PML-RARalpha oncoprotein by directly binding PML. *Science* 328: 240-243.

14. Kitchin KT, Wallace K (2005) Arsenite binding to synthetic peptides based on the Zn finger region and the estrogen binding region of the human estrogen receptor-alpha. *Toxicol Appl Pharmacol* 206: 66-72.

15. Sun X, Zhou X, Du L, Liu W, Liu Y, et al. (2014) Arsenite binding-induced zinc loss from PARP-1 is equivalent to zinc deficiency in reducing PARP-1 activity, leading to inhibition of DNA repair. *Toxicol Appl Pharmacol* 274: 313-318.

16. Witkiewicz-Kucharczyk A, Bal W (2006) Damage of zinc fingers in DNA repair proteins, a novel molecular mechanism in carcinogenesis. *Toxicol Lett* 162: 29-42.

17. Zhang F, Paramasivam M, Cai Q, Dai X, Wang P, et al. (2014) Arsenite binds to the RING finger domains of RNF20-RNF40 histone E3 ubiquitin ligase and inhibits DNA double-strand break repair. *J Am Chem Soc* 136: 12884-12887.

18. Lagoja IM, Pochet S, Boudou V, Little R, Lescrinier E, et al. (2003) A short path synthesis of [13C/15N] multilabeled pyrimidine nucleosides starting from glucopyranose nucleosides. *J Org Chem* 68: 1867-1871.

19. Amato NJ, Zhai Q, Navarro DC, Niedernhofer LJ, Wang Y (2015) In vivo detection and replication studies of alpha-anomeric lesions of 2'-deoxyribonucleosides. *Nucleic Acids Res*: pii: gkv725.

20. Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, et al. (2014) Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J Am Chem Soc* 136: 11582-11585.

21. Wang J, Yuan B, Guerrero C, Bahde R, Gupta S, et al. (2011) Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal Chem* 83: 2201-2209.

22. Liu S, Wang J, Su Y, Guerrero C, Zeng Y, et al. (2013) Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res* 41: 6421-6429.

23. Zhou X, Sun X, Cooper KL, Wang F, Liu KJ, et al. (2011) Arsenite interacts selectively with zinc finger proteins containing C3H1 or C4 motifs. *J Biol Chem* 286: 22855-22863.

24. Demicheli C, Frezard F, Pereira FA, Santos DM, Mangrum JB, et al. (2011) Interaction of arsenite with a zinc finger CCHC peptide: evidence for formation of an As-Zn-peptide mixed complex. *J Inorg Biochem* 105: 1753-1758.

25. Spuches AM, Kruszyna HG, Rich AM, Wilcox DE (2005) Thermodynamics of the As(III)-thiol interaction: arsenite and monomethylarsenite complexes with glutathione, dihydrolipoic acid, and other thiol ligands. *Inorg Chem* 44: 2964-2972.

26. Zhang X, Yang F, Shim JY, Kirk KL, Anderson DE, et al. (2007) Identification of arsenic-binding proteins in human breast cancer cells. *Cancer Lett* 255: 95-106.

27. Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286: 35334-35338.

28. Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 11: 607-620.

29. Zhou X, Sun X, Mobarak C, Gandolfi AJ, Burchiel SW, et al. (2014) Differential binding of monomethylarsonous acid compared to arsenite and arsenic trioxide with zinc finger peptides and proteins. *Chem Res Toxicol* 27: 690-698.

30. Zhao L, Chen S, Jia L, Shu S, Zhu P, et al. (2012) Selectivity of arsenite interaction with zinc finger proteins. *Metallomics* 4: 988-994.

31. Shen S, Li XF, Cullen WR, Weinfeld M, Le XC (2013) Arsenic binding to proteins. *Chem Rev* 113: 7769-7792.

32. Ding W, Liu W, Cooper KL, Qin XJ, de Souza Bergo PL, et al. (2009) Inhibition of poly(ADP-ribose) polymerase-1 by arsenite interferes with repair of oxidative DNA damage. *J Biol Chem* 284: 6809-6817.

33. Schwerdtle T, Walter I, Hartwig A (2003) Arsenite and its biomethylated metabolites interfere with the formation and repair of stable BPDE-induced DNA adducts in human cells and impair XPAzf and Fpg. *DNA Repair (Amst)* 2: 1449-1463.

34. Grosskopf C, Schwerdtle T, Mullenders LH, Hartwig A (2010) Antimony impairs

nucleotide excision repair: XPA and XPE as potential molecular targets. *Chem Res Toxicol* 23: 1175-1183.

35. Kopera E, Schwerdtle T, Hartwig A, Bal W (2004) Co(II) and Cd(II) substitute for Zn(II) in the zinc finger derived from the DNA repair protein XPA, demonstrating a variety of potential mechanisms of toxicity. *Chem Res Toxicol* 17: 1452-1458.

36. Bal W, Schwerdtle T, Hartwig A (2003) Mechanism of nickel assault on the zinc finger of DNA repair protein XPA. *Chem Res Toxicol* 16: 242-248.

37. Asmuss M, Mullenders LH, Eker A, Hartwig A (2000) Differential effects of toxic metal compounds on the activities of Fpg and XPA, two zinc finger proteins involved in DNA repair. *Carcinogenesis* 21: 2097-2104.

38. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, et al. (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18: 553-567.

39. Xu W, Yang H, Liu Y, Yang Y, Wang P, et al. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* 19: 17-30.

40. Yin R, Mao SQ, Zhao B, Chong Z, Yang Y, et al. (2013) Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J Am Chem Soc* 135: 10396-10403.

41. Chen J, Guo L, Zhang L, Wu H, Yang J, et al. (2013) Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat Genet* 45: 1504-1509.

42. Blaschke K, Ebata KT, Karimi MM, Zepeda-Martinez JA, Goyal P, et al. (2013) Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 500: 222-226.

43. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930-935.

44. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324: 929-930.

45. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N (2012) MeCP2 binds to

5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151: 1417-1430.

46. Yildirim O, Li R, Hung JH, Chen PB, Dong X, et al. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147: 1498-1510.

47. Wang J, Tang J, Lai M, Zhang H (2014) 5-Hydroxymethylcytosine and disease. *Mutat Res Rev Mutat Res* 762: 167-175.

48. Ye C, Li L (2014) 5-hydroxymethylcytosine: a new insight into epigenetics in cancer. *Cancer Biol Ther* 15: 10-15.

49. Ray PD, Yosim A, Fry RC (2014) Incorporating epigenetic data into the risk assessment process for the toxic metals arsenic, cadmium, chromium, lead, and mercury: strategies and challenges. *Front Genet* 5: 201.

50. Zierold KM, Knobeloch L, Anderson H (2004) Prevalence of chronic diseases in adults exposed to arsenic-contaminated drinking water. *Am J Public Health* 94: 1936-1937.

51. Tyler CR, Allan AM (2014) The Effects of Arsenic Exposure on Neurological and Cognitive Dysfunction in Human and Rodent Studies: A Review. *Curr Environ Health Rep* 1: 132-147.

52. Dawlaty MM, Breiling A, Le T, Barrasa MI, Raddatz G, et al. (2014) Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell* 29: 102-111.

Chapter 6: Quantitative Mass Spectrometry-Based Analysis of β-D-Glucosyl-5-hydroxymethyluracil in Genomic DNA of *Trypanosoma brucei*

Introduction

β-D-glucosyl-5-hydroxymethyluracil (base J) is the first hyper-modified base discovered in eukaryotic DNA (1). Over the past two decades, this unique modified base has been found within members of unicellular kinetoplastids family, such as *Trypanosoma* and *Leishmania* species (2), and in the related unicellular flagellate protist *Euglena gracilis* (3), where it replaces a fraction of thymine in the genome. In contrast to its discovery in unicellular protozoa, base J was not detectable in animals, plants, or fungi tested, nor in a range of other simple eukaryotes (2). In all kinetoplastid flagellates analyzed, base J is localized primarily in telomeric repeat regions (telomeric J) (4-6) and with a small portion present in other repetitive DNA sequences (7) and in sequences between transcription units (internal J) (8,9). In the parasite *Trypanosoma brucei*, it also appears in the sub-telomeric variant surface glycoprotein (VSG) gene expression sites that are involved in producing a VSG coat on cell surface to evade host immune response (10). The restricted presence of J in the genome of kinetoplastids but not other higher eukaryotes renders base J, and perhaps J biosynthesis, a potential target for parasite-specific chemotherapy (11).

Indirect evidence indicates that biosynthesis of J involves two steps, which starts with *de novo* hydroxylation of a specific thymidine residue to 5-hydroxymethyl-2'-deoxyuridine (5-HmdU) (11). The intermediate 5-HmdU is then converted to β-D-glucosyl-5-hydroxymethyl-2'-deoxyuridine (dJ) by a yet unidentified glucosyl transferase (GT) (11). Two enzymes involved in catalyzing the oxidation of thymidine were identified in trypanosomes, namely, J-binding proteins 1 and 2 (JBP1 and JBP2) (9). Both of them contain a N-terminal thymine hydroxylase (TH) domain (12), but only JBP1 can bind to J in DNA via a particular J-binding domain in its C-terminal half (13). On the other hand, JBP2 contains a SWI2/SNF2 domain homologous to ATPase/DNA helicases, which has been thought to be important for its activity (14). Recently, the JBP proteins have been grouped together with mammalian TET proteins into the new TET/JBP subfamily of Fe(II)- and 2-oxoglutarate (2OG)-dependent dioxygenases (15). Being able to oxidize 5-methylcytosine (5-mC) to 5-hydroxy-methylcytosine (5-HmC) in mammalian DNA (16), TET enzymes are identified as the only homologue of JBP1/2 TH domain in eukaryotes and may play a very important role in active cytosine demethylation in mammals (17).

While much is known about J biosynthesis, the function of base J has long been elusive. However, ice was broken when recent studies showed that J is an epigenetic factor regulating transcription in kinetoplastids. Unusual for eukaryotes, the protein-

coding genes in kinetoplastids are arranged in polycistronic gene clusters transcribed by RNA polymerase II (RNAP II) (18). Genome-wide analysis in trypanosomes revealed an enrichment of internal J at chromosomal regions flanking polycistronic transcription (8). In JBP1/2-knockout *Trypanosoma cruzi*, the loss of J leads to a major increase in transcription initiation and a loose chromatin structure at divergent transcription start regions (19). Differently, loss of internal J, following the deletion of JBP proteins in *Leishmania*, resulted in massive readthrough at RNAP II transcriptional termination sites (20). Although J may assume different functions in different species, the underlying mechanisms could be similar as more repressive chromatin structures are created with the synthesis of J. Furthermore, given the predominant presence of J at telomeric repeats among kinetoplastids analyzed, questions regarding whether J has a telomeric function were addressed, but remained unresolved thus far (11).

To help further dissect the functions of J and verify its presence in other organisms, highly sensitive and accurate detection methods are required. So far several analytical techniques have been employed for detecting base J, including ^{32}P-postlabeling (21), two-dimensional thin-layer chromatography (3), immunoprecipitation experiments (10), immunoblots and Southern blots (6). These methods, however, are limited by their relatively low sensitivity or accuracy and their failure to provide any structural information about J. van Leeuwen *et al.* (21) studied the efficiency of postlabeling, and

found that it could only recover 50% of the total J owing to the poor digestion of this bulky DNA modification during analysis. The following production of rabbit polyclonal antisera against protein-coupled J-deoxyribosemonophosphate (dJMP) raised the sensitivity of J detection (2), but the yield was limited and it was difficult to generate more antisera (11). Since the first application of mass spectrometry techniques in nucleic acid research, HPLC coupled with tandem mass spectrometry (LC-MS/MS) has become one of the standard methods for quantifying DNA modifications (22). For instance, an LC-MS/MS coupled with the stable isotope-dilution method allowed for the examination of the roles of repair proteins in removing bulky DNA lesions from mammalian genome (23). It has also been extensively used for the analysis of 5-mC and its oxidation products, which may be involved in epigenetic regulation of a broad range of biological processes and diseases (24). LC-MS/MS, however, has not been employed for base J detection in any organisms. Herein, we sought to develop a bio-analytical method by using a surrogate internal standard (β-D-glucosyl-5-hydroxymethyl-2'-deoxycytidine, 5-gHmdC) and LC-MS/MS for the accurate measurement of dJ in wild-type *T. brucei* and the isogenic strain that is JBP-null. For comparison, we quantified the levels of 5-HmdU in these strains simultaneously.

Materials and Methods

Materials

All chemicals and enzymes, unless otherwise noted, were purchased from Sigma-Aldrich (St. Louis, MO) and New England Biolabs (Ipswich, MA). Erythro-9-(2-hydroxy-3-nonyl)adenine (EHNA) hydrochloride was from Tocris Bioscience (Ellisville, MO). [1,3-$^{15}N_2$-2'-D]-5-HmdU were previously synthesized (25,26).

Preparation of Standard

Single-stranded oligodeoxyribonucleotide (ODN, 5'-ATGGCGXGCTAT-3') housing a 5-hydroxymethyl-2'-deoxycytidine (5-HmdC) at the X position was incubated with uridine diphosphate (UDP)-glucose and T4 phage β-glucosyltransferase (β-GT) to generate the corresponding 5-gHmdC-containing ODN. The desired ODNs were purified on HPLC and its identity verified by MS and MS/MS. The J-containing ODN was prepared and characterized in a similar fashion.

DNA Extraction and Enzymatic Digestion

Blood stream form *T. brucei* cell line 221a of strain 427 (Wild-type and JBP-null) were cultured as previously described (27). Total genomic DNA was isolated as described (28). To the resulting DNA sample (1 μg) were added 100 fmol of the aforementioned 5-gHmdC-containing ODN, which was the surrogate internal standard

for LC-MS/MS analysis, and a cocktail of enzymes including nuclease P1, phosphodiesterases 1 and 2, and alkaline phosphatase, as described previously (29). To inhibit adenine deaminase during digestion, EHNA was also added to the digestion mixture until its final concentration reached 2.5 mM. To the resultant nucleoside mixture was then added 2 pmol [1,3-$^{15}N_2$-2'-D]-5-HmdU, and the enzymes in the digestion mixture were removed by chloroform extraction. The target nucleosides, dJ, 5-gHmdC, and 5-HmdU, were enriched via off-line HPLC.

HPLC Enrichment

The enrichment was performed on a Beckman HPLC system with pump module 125 and a UV detector (module 126). An Aeris WIDEPORE C18 column (4.6 × 250 mm, 3.6 μm beads, 200 Å pore size, Phenomenex, Torrance, CA) was used. An isocratic elution of 10 mM ammonium formate (pH 8.5) was employed at a flow rate of 0.8 mL/min. The HPLC fractions for dJ and 5-gHmdC were collected together, whereas the fraction for 5-HmdU was collected separately for subsequent LC-MS/MS analysis.

Mass Spectrometric Analysis

LC-MS/MS measurement of dJ was conducted on an LTQ XL linear ion trap mass spectrometer equipped with a nanoelectrospray ionization source coupled to an EASY-nLC II system (Thermo Fisher Scientific, San Jose, CA). Samples were automatically loaded onto a home-made trapping column (150 μm × 40 mm) packed

with Magic AQ reversed-phase C18 material (5 μm beads, 120 Å pore size, Michrom BioResources, Auburn, CA) at a flow rate of 3 μL/min and eluted to an in-house packed Magic C18 AQ column (75 μm × 200 mm, 5 μm beads, 300 Å pore size). A solution of 0.1% (v/v) formic acid in water (solution A) and a solution of 0.1% (v/v) formic acid in acetonitrile (solution B) were used as mobile phases. The modified nucleosides were separated with a gradient of 40 min 0-10% B and at a flow rate of 300 nL/min. The temperature for the ion transport tube was maintained at 275 ℃, the activation Q was 0.25, and activation time was 30 ms. In the positive-ion mode, the spray, capillary and tube lens voltages were 2.0 kV, 12 V and 100 V, respectively.

The enriched fractions containing 5-HmdU were separated on an Agilent 1200 capillary HPLC using a 0.5 × 250 mm Zorbax SB-C18 column (5 μm beads, 80 Å pore size, Agilent Technologies, Santa Clara, CA). The eluent was directed to an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) following the previously reported procedures (30).

Results and Discussion

5-gHmdC-containing ODN as Surrogate Internal Standard for Base J Quantification

LC-MS/MS coupled with the isotope-dilution technique constitutes the most

reliable analytical method for the quantification of modified nucleosides in cellular and tissue DNA (31). Although synthetic methods were previously developed for the synthesis of dJ and for its incorporation into ODNs (32-34), the methods involve a multi-step synthesis and are not readily amenable for the preparation of stable isotope-labeled dJ. Thus, we decided to take an alternative approach for the LC-MS/MS quantification of dJ in cellular DNA, where we employed a 5-gHmdC-bearing single-stranded ODN as a surrogate internal standard. In this context, it is important to note that 5-gHmdC is absent in genomic DNA of *T. brucei*, as revealed by LC-MS/MS analysis. We synthesized the 5-gHmdC-containing ODN using T4 phage *β*-GT following previously published procedures (35). In this vein, the T4 *β*-GT-catalyzed glucosylation of 5-HmdC has been utilized for the measurement of global 5-HmdC levels *in vivo* (35-37). In our method, the 5-gHmdC-carrying standard ODN was digested together with DNA isolated from *T. brucei* prior to HPLC enrichment, which corrects for incomplete release of dJ from DNA during enzymatic hydrolysis and potential loss of the analyte during other stages of sample preparation process, thereby providing more accurate quantification of dJ.

The identity of the standard ODN was verified by ESI-MS and MS/MS analyses (Figure D.1a and D.1b). The MS/MS of the $[M-3H]^{3-}$ ion (*m/z* 1283.5) yielded typical [a_n-base] and w_n series ions (38). In particular, we observed a 481.5-Da difference

between the w_5 (*m/z* 1556.1) and w_6^{2-} (*m/z* 1018.3) ions. This mass difference is consistent with the residue mass of a 5-gHmdC monophosphate, thereby supporting the presence of this modified nucleotide at the 7th position counting from the 5' terminus of the ODN. The similar conclusion can be drawn from the mass difference between the $[a_7\text{-X}]^{2-}$ (*m/z* 994.8) and $[a_8\text{-G}]^{2-}$ (*m/z* 1235.8) ions.

Analytical Strategy for Simultaneous Quantification of dJ and 5-HmdU

To improve the detection sensitivity of dJ and 5-HmdU in isolated cellular DNA, we employed off-line HPLC enrichment to remove the unmodified nucleosides and buffer salts employed in the enzymatic digestion. An example HPLC trace is shown in Figure 6.1, where the three modified nucleosides, 5-HmdU, 5-gHmdC and dJ, eluting at 11 min, 28 min, and 30 min, respectively, were well resolved from each other and from the canonical 2'-deoxyribonucleosides and residual ribonucleosides. In this vein, the amounts of RNA contamination in the DNA samples were corrected based on the peak areas of the 2'-deoxyribonucleosides and ribonucleosides found in the chromatograms for the HPLC enrichment, as described previously (30).

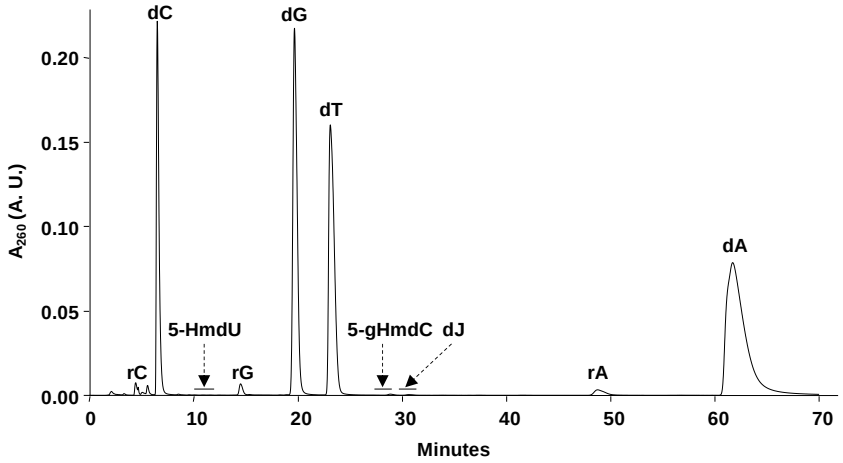


Figure 6.1. A representative HPLC trace for the off-line enrichment of 5-HmdU, 5-gHmdC, and dJ from the enzymatic digestion mixture of genomic DNA isolated from wild-type *T. brucei*.

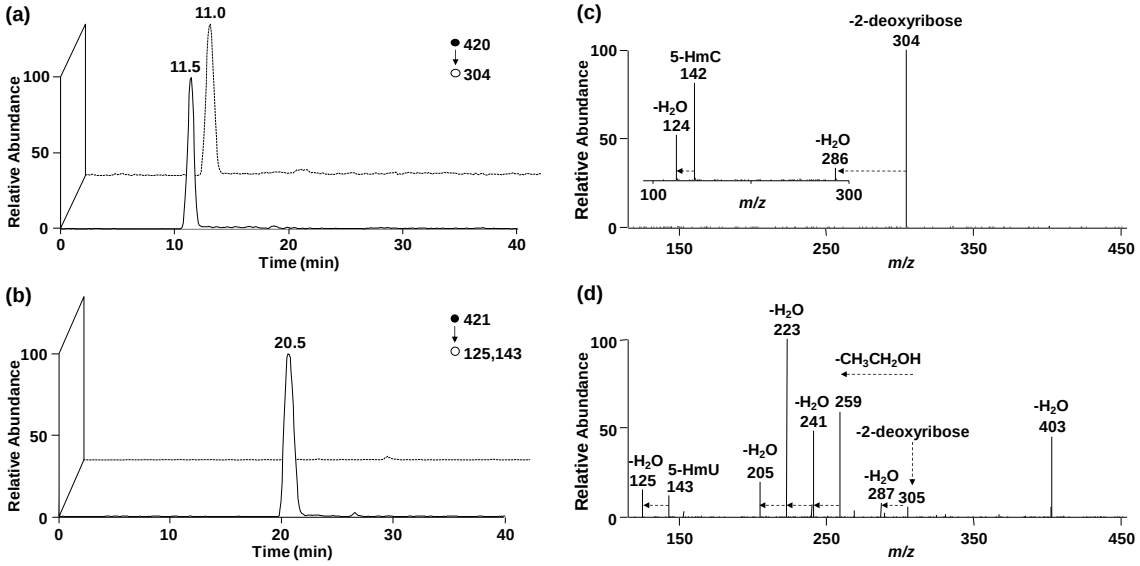
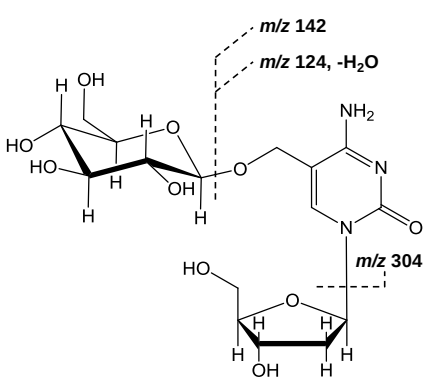
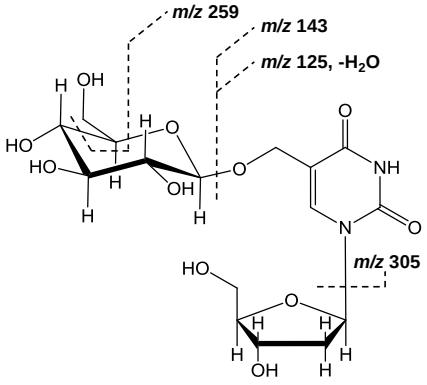


Figure 6.2. Representative LC-MS/MS results for the quantification of dJ. Shown in left panels are the selected-ion chromatograms for monitoring the indicated transitions for the surrogate internal standard (5-gHmdC, a) and the analyte (dJ, b), and solid and dashed lines represent samples from wild-type and JBP-null *T. brucei*, respectively. Right panels give the MS/MS for 5-gHmdC (c) and dJ (d), and the inset (c) gives the MS/MS/MS for 5-gHmdC.

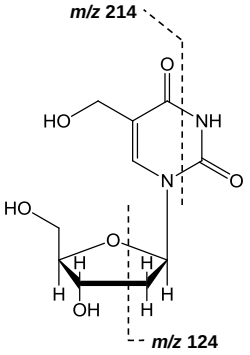
Scheme 6.1. Chemical structures of the modified nucleosides measured in this study and the cleavages for the formation of major fragment ions observed in the MS^2 of the $[M+H]^+$ ion of 5-gHmdC and dJ as well as in the MS^3 of the $[M-H]^-$ ion of 5-HmdU. The sites of ^{15}N and deuterium (D) are indicated in the structure of the isotope-labeled 5-HmdU.



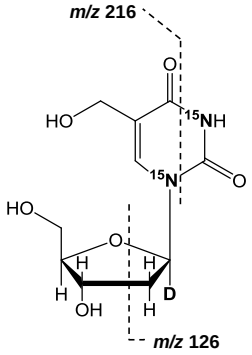
5-gHmdC



dJ



5-HmdU



[1,3-$^{15}N_2$-2'-D]-5-HmdU

To achieve unambiguous identification and quantification of dJ, we subjected the dJ- and 5-gHmdC-containing fractions to nano-scale LC (nanoLC)-MS^2 analysis on an LTQ linear ion-trap mass spectrometer. As displayed in Figure 6.2a and 6.2b, a peak was observed in the selected-ion chromatogram (SIC) for monitoring the *m/z* 420 → 304 transition for 5-gHmdC or *m/z* 421 → 125, 143 transition for dJ. Relative to offline-HPLC enrichment, a much slower gradient was employed to online-HPLC separation in order to avoid the interference between surrogate standard and analyte, the molecular weights of which differ by only 1-Da. The identities of the components eluting at 11 min and 20 min were further confirmed to be 5-gHmdC and dJ from MS/MS measurements. The chemical structures and the proposed major fragmentation pathways for the $[M+H]^+$ ions of 5-gHmdC and dJ found in MS/MS are depicted in Scheme 6.1. Upon collisional activation, the *N*-glycosidic linkage in the $[M+H]^+$ ion of 5-gHmdC (*m/z* 420) was cleaved readily to produce the most abundant product ion of *m/z* 304 in MS/MS (Figure 6.2c). Further collisional activation of the ion of *m/z* 304 led to the generation of three major fragment ions (*m/z* 124, 142, and 286) in MS^3, as a consequence from neutral losses of glucose, part of glucose [$C_6H_{10}O_5$], and H_2O, respectively (see inset of Figure 6.2c). Likewise, collisional activation of the $[M+H]^+$ ions of dJ (*m/z* 421) yielded fragment ions of *m/z* 305, emanating from the elimination of a 2-deoxyribose moiety, and *m/z* 125, 143, and 287, which are attributed to the

subsequent losses of glucose, $C_6H_{10}O_5$, and H_2O from base J portion (Figure 6.2d). Apart from these fragment ions, the MS/MS of the $[M+H]^+$ ion of dJ gave additional fragment ions that are ascribed to the partial losses of the glucosyl unit on J, i.e., the ions of *m/z* 205, 223, 241, and 259 (Figure 6.2d). Because the fragmentation behaviors differed between 5-gHmdC and dJ, we selected unique transitions for monitoring the surrogate internal standard (5-gHmdC, *m/z* 420 → 304) and the target analyte (dJ, *m/z* 421 → 125, 143), which allows for unequivocal identification and reliable quantification of base J in cellular DNA.

A calibration curve was constructed for the quantification of dJ under identical conditions as described for sample analysis (See Experimental Section). Along this line, known amounts of the dJ-containing ODNs and 100 fmol of 5-gHmdC-harboring ODN were digested together with calf thymus DNA at the same amount as the genomic DNA, and then enriched via offline-HPLC. The addition of calf thymus DNA to the calibration mixture serves to mimic the digestion conditions for genomic DNA. This, together with the analysis of 5-gHmdC- and dJ-containing HPLC fractions under the same LC-MS/MS conditions, facilitated the accurate quantification of dJ. It is worth noting that, from the calibration curve, over 100-fold difference was found for the efficiencies in the formation of the ion of *m/z* 304 for 5-gHmdC and those for the formation of the ions of *m/z* 125 and 143 for dJ. At least two factors may account for

this difference. First, the ion of *m/z* 304 is the only product ion found in the MS/MS of the $[M+H]^+$ ion of 5-gHmdC, whereas a large number of fragment ions are observed in the MS/MS of the $[M+H]^+$ ion of dJ (Figure 6.2c and 6.2d). Second, a previous study by Liguori and coworkers (39) suggested that the proton affinity of 2'-deoxyuridine is significantly lower than that of 2'-deoxycytidine. Therefore, the glucosylated 5-HmdC may possess a higher proton affinity than glucosylated 5-HmdU, which may contribute to the more facile ionization of the former nucleoside. In spite of this difference, according to the data gained on three separate days, we obtained excellent linearity for the calibration curve ($R^2 = 0.998$).

Because the sensitivity for the detection of 5-HmdU is better in the negative- than the positive-ion mode, we quantified 5-HmdU by using LC-MS^3 in the negative-ion mode, as previously reported (30). The co-elution of 5-HmdU analyte and standard (Figure 6.3a and 6.3b), together with their identical fragmentation patterns (Figure 6.3c and 6.3d), provides solid evidence for the identification of the component and leads to accurate quantification of 5-HmdU.

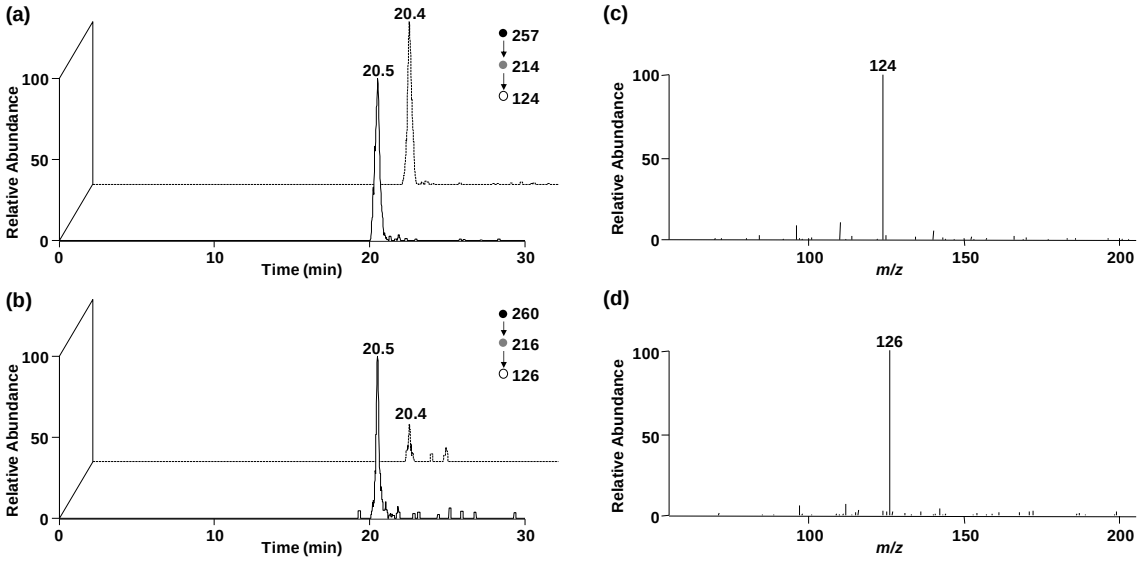


Figure 6.3. Representative LC-MS/MS/MS results for the quantification of 5-HmdU. Shown in left panels are the selected-ion chromatograms for monitoring the indicated transitions for unlabeled 5-HmdU (a) and isotope-labeled 5-HmdU (b), and solid and dashed lines represent samples from wild-type and JBP-null *T. brucei*, respectively. Right panels give the MS/MS/MS for 5-HmdU (c) and labeled 5-HmdU (d).

Quantification of dJ and 5-HmdU in T. brucei

With the use of this method, we measured simultaneously the levels of dJ and 5-HmdU in *T. brucei*. To our knowledge, this is the first rigorous quantification of base J and its intermediate product 5-HmU in the genome of kinetoplastids. It is worth noting that the employment of nanoLC coupled to mass spectrometer decreases the amount of DNA required for J detection (~200 ng for wild-type cells), while in the meantime the detection sensitivity is improved by at least 10 fold relative to the use of regular ESI and a capillary HPLC operating at a flow rate of 8 μL/min. The limit of

quantitation (LOQ) and limit of detection (LOD), which referred to the amounts of dJ giving rise to signal-to-noise ratios of 10 and 3, were 46.8 fmol and 15.6 fmol, respectively. This level of sensitivity renders the method highly powerful for the detection of J in cellular DNA, which will be very useful for dissecting the roles of various protein factors in J biosynthesis and for interrogating the biological functions of base J.

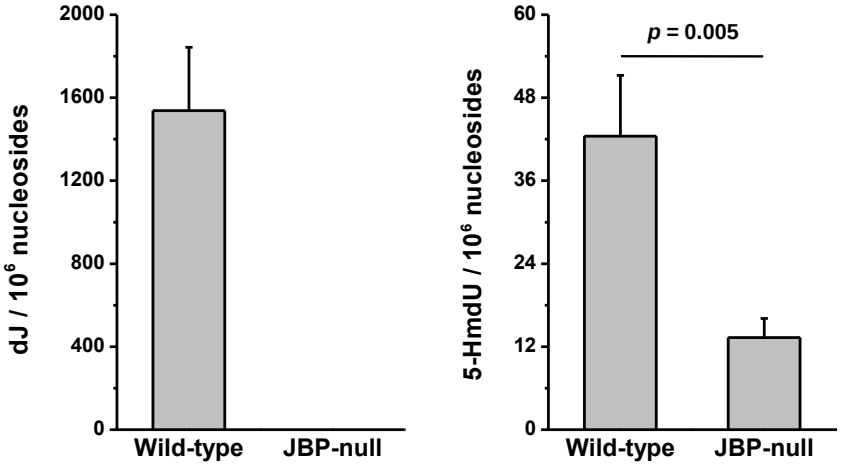


Figure 6.4. Frequencies of dJ and 5-HmdU in *T. brucei* DNA (n=3). The data represent the mean and standard deviation of the measurement results. The *p* values were calculated using unpaired two-tailed *t* test.

Our quantitative result revealed that the level of dJ in *T. brucei* genome was ~1500 modifications per 10^6 nucleosides (Figures 6.2 and 6.4a). Based on the peak areas of the four natural 2'-deoxyribonucleosides found in the HPLC trace with consideration of their extinction coefficients at 260 nm, the percentage of thymidine was quantified to

be approximately 29.1%. Thus, the level of dJ represented 0.53% of dT in the wild-type strain, which is consistent with the levels reported previously (3,4). Conversely, J was not detectable in the JBP-null strain (Figure 6.4a). To confirm that this observation was based on the absence of J but not due to the limit of detection, we increased the amount of DNA used for the analysis by 40 fold. However, there was still no signal exhibited in SICs for the analyte (data not shown). A remarkable difference in the SICs for monitoring dJ in wild-type and mutant *T. brucei* was displayed in Figure 6.2b. Similarly, we found that the level of 5-HmdU in the wild-type strain was 42.5 modifications per 10^6 nucleosides, while it was present at a level that is 3.2 times lower in the JBP-null strain (13.3 modifications per 10^6 nucleosides, Figures 6.3 and 6.4b).

The significant decreases in the levels of dJ and 5-HmdU in JBP-null *T. brucei* are in keeping with the important roles of JBP proteins in the biosynthesis of 5-HmU and J. In this study, we found that the level of 5-HmdU in JBP-null cells was approximately one third of that in wild-type cells. However, considering the LOD and the amount of DNA used for detection, base J, if exists, would be present in the mutant strain at a level that is at least 12,500 times lower than that in the wild-type strain. This is in line with the previous observation that deletion of both JBP1 and JBP2 results in lack of base J in cells (9). The absence of base J in JBP-null background is primarily attributed to the lack of JBP-mediated biosynthesis of the precursor of base J (i.e., 5-HmU). In

this vein, all available evidence now supports that base J synthesis is initiated by JBP1- and JBP2-mediated *de novo* oxidation of dT to 5-HmdU (40-43). Additionally, it was hypothesized that JBP1 can bind to the initially synthesized J via its specific J-binding domain and then oxidize adjacent thymidines followed by GT-catalyzed glucosylation (11). Therefore, the lack of local amplification of J synthesis may also contribute to the absence of J in the trypanosome.

In the presence of 5-HmU, the biosynthesis of J may potentially enter into the second stage where a glucose moiety is transferred onto 5-HmU by the putative GT (11). Thus, one may expect that some of the 5-HmU may be converted to base J in the JBP-null background. Several factors may account for our observation of the absence of base J, but the presence of a significant level of 5-HmdU in the JBP-null *T. brucei*. First, JBPs and putative GT may function together as a protein complex, the loss of function of one (JBPs) may compromise the expression and/or the function of the other (GT). Second, there could be other enzymes involved in the oxidation of thymine to 5-HmU in *T. brucei* and the resulting 5-HmU may not be converted to base J. Moreover, some of the 5-HmdU detected in wild-type and JBP-null *T. brucei* may arise from oxidatively induced DNA damage. Further studies are necessary for a better understanding of this observation.

References

1. Gommers-Ampt JH, Van Leeuwen F, de Beer AL, Vliegenthart JF, Dizdaroglu M, et al. (1993) beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. *Cell* 75: 1129-1136.

2. van Leeuwen F, Taylor MC, Mondragon A, Moreau H, Gibson W, et al. (1998) beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci U S A* 95: 2366-2371.

3. Dooijes D, Chaves I, Kieft R, Dirks-Mulder A, Martin W, et al. (2000) Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of Euglena gracilis. *Nucleic Acids Res* 28: 3017-3021.

4. van Leeuwen F, Wijsman ER, Kuyl-Yeheskiely E, van der Marel GA, van Boom JH, et al. (1996) The telomeric GGGTTA repeats of Trypanosoma brucei contain the hypermodified base J in both strands. *Nucleic Acids Res* 24: 2476-2482.

5. Genest PA, Ter Riet B, Cijsouw T, van Luenen HG, Borst P (2007) Telomeric localization of the modified DNA base J in the genome of the protozoan parasite Leishmania. *Nucleic Acids Res* 35: 2116-2124.

6. Ekanayake DK, Cipriano MJ, Sabatini R (2007) Telomeric co-localization of the modified base J and contingency genes in the protozoan parasite Trypanosoma cruzi. *Nucleic Acids Res* 35: 6367-6377.

7. van Leeuwen F, Kieft R, Cross M, Borst P (2000) Tandemly repeated DNA is a target for the partial replacement of thymine by beta-D-glucosyl-hydroxymethyluracil in Trypanosoma brucei. *Mol Biochem Parasitol* 109: 133-145.

8. Cliffe LJ, Siegel TN, Marshall M, Cross GA, Sabatini R (2010) Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of Trypanosoma brucei. *Nucleic Acids Res* 38: 3923-3935.

9. Cliffe LJ, Kieft R, Southern T, Birkeland SR, Marshall M, et al. (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res* 37: 1452-1462.

10. van Leeuwen F, Wijsman ER, Kieft R, van der Marel GA, van Boom JH, et al.

(1997) Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei. *Genes Dev* 11: 3232-3241.

11. Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* 62: 235-251.

12. Yu Z, Genest PA, ter Riet B, Sweeney K, DiPaolo C, et al. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res* 35: 2107-2115.

13. Cross M, Kieft R, Sabatini R, Wilm M, de Kort M, et al. (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J* 18: 6573-6581.

14. Kieft R, Brand V, Ekanayake DK, Sweeney K, DiPaolo C, et al. (2007) JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form Trypanosoma brucei. *Mol Biochem Parasitol* 156: 24-31.

15. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930-935.

16. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466: 1129-1133.

17. Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 11: 607-620.

18. Ekanayake D, Sabatini R (2011) Epigenetic regulation of polymerase II transcription initiation in Trypanosoma cruzi: modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. *Eukaryot Cell* 10: 1465-1472.

19. Ekanayake DK, Minning T, Weatherly B, Gunasekera K, Nilsson D, et al. (2011) Epigenetic regulation of transcription and virulence in Trypanosoma cruzi by O-linked thymine glucosylation of DNA. *Mol Cell Biol* 31: 1690-1700.

20. van Luenen HG, Farris C, Jan S, Genest PA, Tripathi P, et al. (2012) Glucosylated hydroxymethyluracil, DNA base j, prevents transcriptional readthrough in leishmania. *Cell* 150: 909-921.

21. van Leeuwen F, de Kort M, van der Marel GA, van Boom JH, Borst P (1998) The modified DNA base beta-D-glucosylhydroxymethyluracil confers resistance to micrococcal nuclease and is incompletely recovered by 32P-postlabeling. *Anal Biochem* 258: 223-229.

22. Dudley E, Bond L (2013) Mass spectrometry analysis of nucleosides and nucleotides. *Mass Spectrom Rev*

23. Liu S, Wang Y (2013) A quantitative mass spectrometry-based approach for assessing the repair of 8-methoxypsoralen-induced DNA interstrand cross-links and monoadducts in mammalian cells. *Anal Chem* 85: 6732-6739.

24. Yuan B-F, Feng Y-Q (2014) Recent advances in the analysis of 5-methylcytosine and its oxidation products. *TrAC, Trends Anal Chem* 54: 24-35.

25. Cao H, Wang Y (2006) Collisionally activated dissociation of protonated 2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives. *J Am Soc Mass Spectrom* 17: 1335-1341.

26. Hong H, Cao H, Wang Y (2006) Identification and quantification of a guanine-thymine intrastrand cross-link lesion induced by Cu(II)/H2O2/ascorbate. *Chem Res Toxicol* 19: 614-621.

27. DiPaolo C, Kieft R, Cross M, Sabatini R (2005) Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. *Mol Cell* 17: 441-451.

28. Bernards A, Van der Ploeg LH, Frasch AC, Borst P, Boothroyd JC, et al. (1981) Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3' end. *Cell* 27: 497-505.

29. Liu S, Dunwell TL, Pfeifer GP, Dunwell JM, Ullah I, et al. (2013) Detection of Oxidation Products of 5-Methyl-2'-Deoxycytidine in Arabidopsis DNA. *PLoS One* 8.

30. Liu S, Wang J, Su Y, Guerrero C, Zeng Y, et al. (2013) Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res* 41: 6421-6429.

31. Tretyakova N, Goggin M, Sangaraju D, Janis G (2012) Quantitation of DNA adducts by stable isotope dilution mass spectrometry. *Chem Res Toxicol* 25: 2007-2035.

32. de Kort M, Ebrahimi E, Wijsman ER, van der Marel GA, van Boom JH (1999)

Synthesis of Oligodeoxynucleotides Containing 5-(β-D-Glucopyranosyloxymethyl)-2′-deoxyuridine, a Modified Nucleoside in the DNA of Trypanosoma Brucei. *Eur J Org Chem* 1999: 2337-2344.

33. Turner John J, Meeuwenoord Nico J, Rood A, Borst P, van der Marel Gijs A, et al. (2003) Reinvestigation into the Synthesis of Oligonucleotides Containing 5-(β-D-Glucopyranosyloxymethyl)-2′-deoxyuridine. *Eur J Org Chem* 2003: 3832-3839.

34. Grover RK, Pond SJ, Cui Q, Subramaniam P, Case DA, et al. (2007) O-glycoside orientation is an essential aspect of base J recognition by the kinetoplastid DNA-binding protein JBP1. *Angew Chem Int Ed Engl* 46: 2839-2843.

35. Terragni J, Bitinaite J, Zheng Y, Pradhan S (2012) Biochemical characterization of recombinant beta-glucosyltransferase and analysis of global 5-hydroxymethylcytosine in unique genomes. *Biochemistry* 51: 1009-1019.

36. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* 38: e181.

37. Chen ML, Shen F, Huang W, Qi JH, Wang Y, et al. (2013) Quantification of 5-methylcytosine and 5-hydroxymethylcytosine in genomic DNA from hepatocellular carcinoma tissues by capillary hydrophilic-interaction liquid chromatography/quadrupole TOF mass spectrometry. *Clin Chem* 59: 824-832.

38. McLuckey SA, Van Berkel GJ, Glish GL (1992) Tandem mass spectrometry of small, multiply charged oligonucleotides. *J Am Soc Mass Spectrom* 3: 60-70.

39. Liguori A, Napoli A, Sindona G (1994) Determination of substituent effects on the proton affinities of natural nucleosides by the kinetic method. *Rapid Commun Mass Spectrom* 8: 89-93.

40. Heidebrecht T, Fish A, von Castelmur E, Johnson KA, Zaccai G, et al. (2012) Binding of the J-binding protein to DNA containing glucosylated hmU (base J) or 5-hmC: evidence for a rapid conformational change upon DNA binding. *J Am Chem Soc* 134: 13357-13365.

41. Heidebrecht T, Christodoulou E, Chalmers MJ, Jan S, Ter Riet B, et al. (2011) The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. *Nucleic Acids Res* 39: 5715-5728.

42. Vainio S, Genest PA, ter Riet B, van Luenen H, Borst P (2009) Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. *Mol Biochem Parasitol* 164: 157-161.

43. Cliffe LJ, Hirsch G, Wang J, Ekanayake D, Bullard W, et al. (2012) JBP1 and JBP2 Proteins Are Fe2+/2-Oxoglutarate-dependent Dioxygenases Regulating Hydroxylation of Thymidine Residues in Trypanosome DNA. *J Biol Chem* 287: 19886-19895.

Chapter 7: Concluding Remarks and Future Directions

The goal of work presented in Chapter 2 was to develop an LC-MS/MS method for assessing the repair of DNA damage induced by psoralen and UVA light in mammalian cells. Our results demonstrated that the unequivocal chemical specificity, accuracy, sensitivity, and reproducibility of the LC-MS/MS coupled with isotope dilution technique allowed for monitoring the repair of 8-MOP-induced ICL and MAs in mammalian cells. With the use of this method, we investigated the impact of defective NER pathway on the cellular removal of 8-MOP/UVA-induced DNA damage by comparing human fibroblasts (HFs) derived from normal and XPA individuals, and Chinese hamster ovary (CHO) cells that are repair-competent or deficient in ERCC1. The results revealed a much lower removal rate of ICL and MAs in repair-deficient mammalian cells relative to their repair-competent counterparts, thereby providing direct evidence to unambiguously support the role of NER in repairing these DNA lesions. It can be envisaged that the method can be generally applicable for examining the roles of other DNA repair enzymes and pathways in removing these lesions in mammalian cells.

Our LC-MS/MS method also unveiled variations in repair efficiencies among the three types of mammalian cells (i.e., HFs, CHO cells and human embryonic kidney

epithelial cells), suggesting that the method may also be useful for examining the repair capacities of different individuals toward these DNA lesions. Such a possibility is also manifested by the fact that pharmacologically relevant doses of 8-MOP and UVA light were used in the present study. Thus, our LC-MS/MS-based approach may facilitate the future use of ICL and MAs as biomarkers for defining the optimal dose of psoralen plus UVA treatment in individual patients.

In Chapters 3 and 4, we developed a sensitive LC-MS/MS coupled with stable isotope-dilution method for the simultaneous quantification of 5-hmdC, 5-fodC, 5-cadC, and 5-hmdU in mammalian cellular and tissue DNA and in plant DNA. Our results provided the first direct comparison of the levels of these four modified nucleosides in same mammalian tissue samples. It can be envisaged that the robust analytical method, when combined with manipulations of genes involved in Tet-mediated active cytosine demethylation pathway, will provide important insights into the roles of various protein factors in active DNA cytosine demethylation. Additionally, the method developed here should be generally applicable for the detection of these 5-mdC modification products in other organisms, which will provide important knowledge for understanding the involvement of Tet family enzymes and their homologs in epigenetic regulation of gene expression in these organisms. Moreover, considering that aberrant levels of 5-hmdC has been found in a number of different cancers, the analytical method reported here

should also be useful for examining whether 5-HmdC and other 5-mdC modifications discussed in this paper can be used as biomarkers for cancer and perhaps other human diseases.

In addition, we found that Arabidopsis genomic DNA contains detectable levels of oxidation products of 5-mC. Our quantification results suggest that the modified bases are most likely induced by ROS, though we cannot exclude the possibility that intermediates of sequential oxidation of 5-mdC are present at a small number of specific genomic loci. Further studies would be needed to determine whether 5-HmC, 5-FodC, and 5-CadC are located at specific loci in Arabidopsis DNA, but this approach would be challenging given the low levels of these modified bases that we measured.

In Chapter 5, we examined the interaction between As(III) and zinc finger domains of human Tet proteins *in vitro* and in mammalian cells with the applications of LC-MS, UV absorption spectrophotometry, and streptavidin agarose affinity assay. We also assessed how arsenite exposure affects the Tet-mediated oxidation of 5-mC in vitro and in cultured mammalian cells by determining the catalytic efficiency of Tet1 in sequentially oxidizing 5-mC in a synthetic duplex DNA and by analyzing the levels of 5-hmC in genomic DNA isolated from arsenite-treated cells. We found that arsenite could bind directly to the zinc fingers of Tet proteins *in vitro* and in cells, and this interaction substantially impaired the catalytic efficiency of Tet proteins in oxidizing

5-mC both *in vitro* and in mammalian cells. The results uncovered a novel molecular mechanism underlying the arsenic-induced perturbation in epigenetic signaling, which deepened our understanding on the epigenotoxicity of arsenic. Additionally, because Tet enzymes and 5-hmC are present in various embryonic and adult cell types, and play crucial roles in pluripotency as well as embryonic and adult development, the present work sets the stage for future studies on the roles of arsenic interaction with Tet proteins in perturbing the methylation patterns during development. Owing to the metal binding properties of zinc fingers, it can be envisaged that Tet proteins may also be susceptible to inhibition by a variety of carcinogenic metal ions.

In Chapter 6, the mass spectrometry-based analytical method developed in this study provides the first accurate and simultaneous measurements of base J and its intermediate product 5-hmdU in kinetoplastid flagellates. Our results are consistent with the important roles of JBP proteins in J synthesis. It can be envisaged that the analytical method will be generally applicable for examining the roles of other proteins involved in J synthesis. Additionally, being the only hyper-modified base discovered in eukaryotic genome, J has been found in unicellular kinetoplastids but not in higher animals. Given that TET enzymes are the only homologue of JBP1/2 in mammals, it is still an open question whether hyper-modifications, such as J derivatives, are synthesized in higher organisms. In this context, the sensitive method reported here

could be adapted readily for exploring the potential presence of other hyper-modifications and their biological functions in other eukaryotes.

Appendix A: Supporting information for Chapter 2

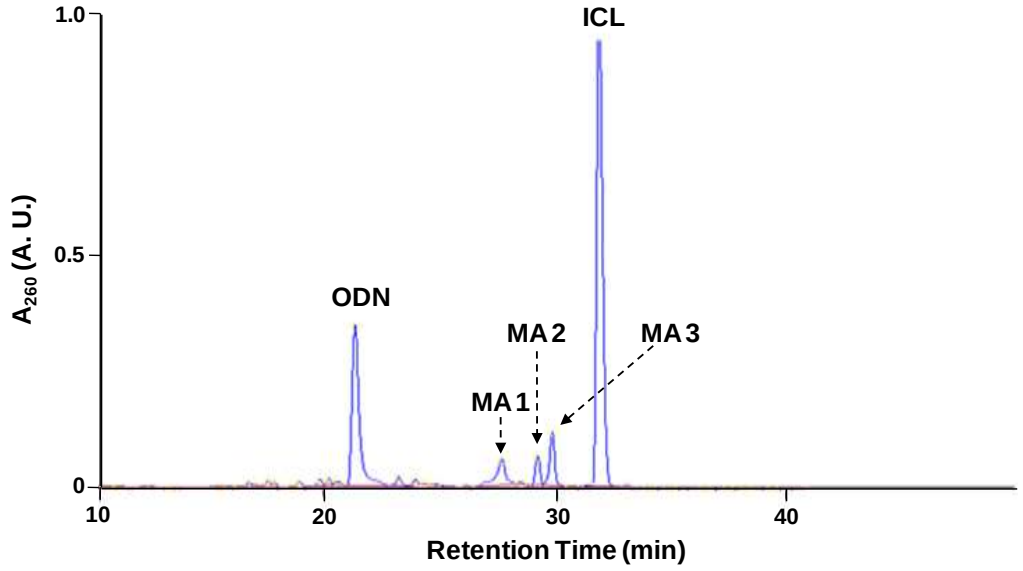


Figure A.1. Representative HPLC trace for the purification of 8-MOP-ICL or 8-MOP-MA bearing self-complementary duplex ODN, d(CGCGCTAGCGCG), from the reaction mixture.

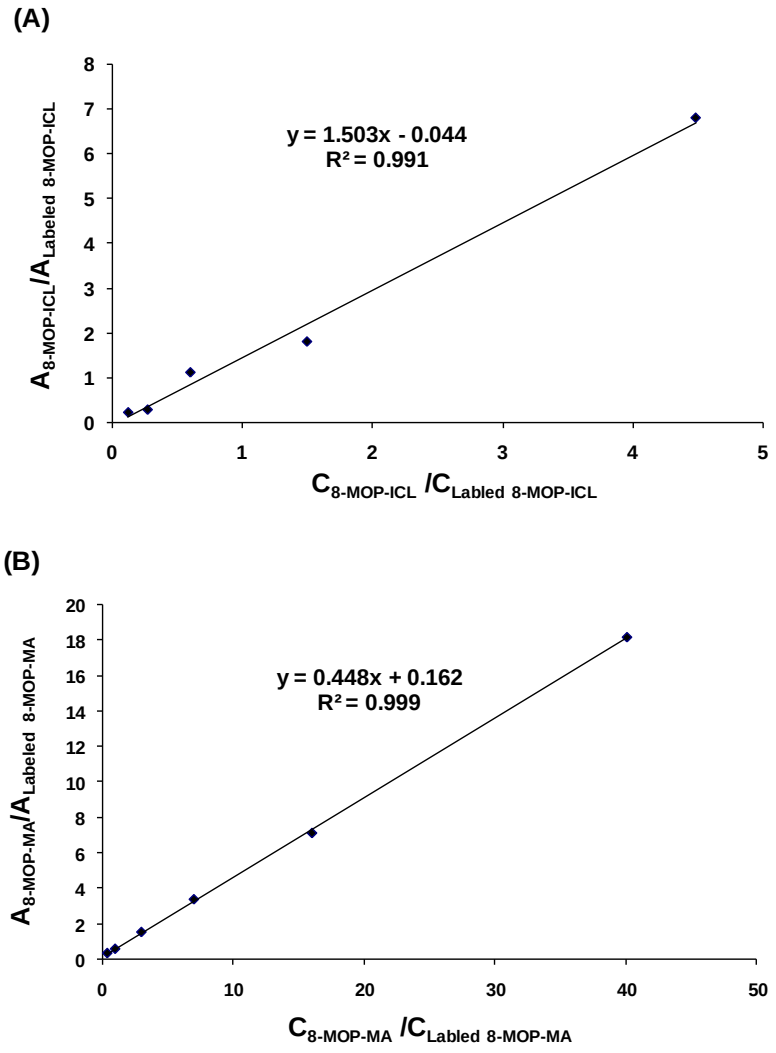


Figure A.2. Calibration curves for the quantification of 8-MOP-ICL (A) and 8-MOP-MA (B) by LC-MS/MS with the stable isotope dilution method. Plotted are the ratios of areas of peaks found in the selected-ion chromatograms for monitoring the transitions as shown in Figures 2.1 and 2.2 versus the molar ratios of the lesion-containing unlabeled 8-MOP over that of the corresponding stable isotope-labeled standard (0.5 pmol for 8-MOP-ICL and 0.2 pmol for 8-MOP-MA).

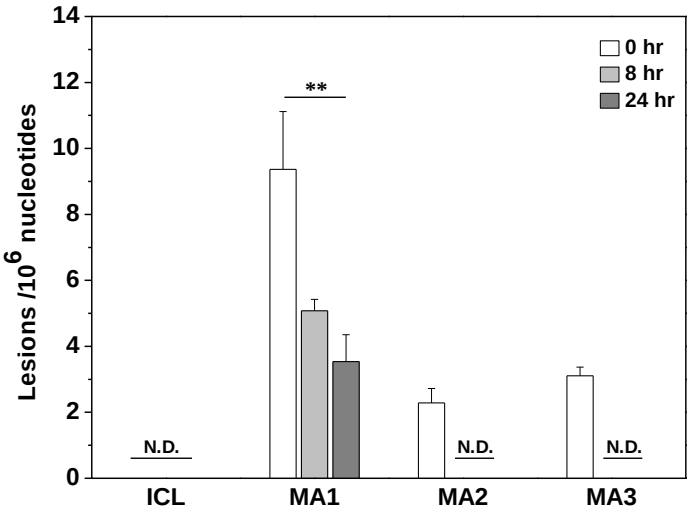


Figure A.3. LC-MS/MS quantification results for 8-MOP photo-activated ICL and MAs in DNA extracted from HEK293T cells. Data represent the means and standard deviations of results obtained from three independent experiments. **, $p < 0.01$; N.D., not detectable. The *p*-values were calculated by using unpaired two-tailed *t*-test.

Appendix B: Supporting information for Chapter 3

Table B.1. Information of 4 patients from whom the brain tissues were obtained.

Patient #	Age	Sex	Race
1	44 years 60 days	Female	Caucasian
2	46 years 163 days	Female	Caucasian
3	35 years 345 days	Female	Caucasian
4	23 years 228 days	Female	Asian

Table B.2. Optimized instrument conditions used for the LC-ESI-MS3 analysis of the 5-HmdC, 5-FodC, 5-CadC and 5-HmdU and detection limits of these modifications. The LOQ (limit of quantitation, which is defined as the amount of analyte that gives rise to a signal-to-noise ratio of 10) represents the means and standard deviations of the results from three measurements of the unlabeled standards. A constant activation time of 30 ms was employed for this experiment.

Analyte	**SIM (MS^3) of lesions**		**Optimized LTQ parameters**		**LOQ, fmol**
	Unlabeled	**Isotope-labeled**	**Normalized collision energy**	**Activation Q**	
5-HmdU	257 → 214	260 → 216	30	0.25	80 ± 15
	214 → 124	216 → 126	33	0.25	
5-HmdC	258 → 142	261 → 145	48	0.25	0.056 ± 0.011
	142 → 124	145 → 126	27	0.25	
5-FodC	256 → 140	259 → 142	33	0.27	0.098 ± 0.007
	140 → 97	142 → 98	36	0.43	
5-CadC	272 → 156	275 → 159	15	0.35	0.14 ± 0.03
	156 → 138	159 → 141	21	0.37	

Table B.3. Levels of 5-mdC, 5-HmdC, 5-FodC, 5-CadC, and 5-HmdU in genomic DNA of HEK293T cells with the overexpression of Tet1 catalytic domain (HEK293T-Tet1) and its corresponding mutant (HEK293T-Tet1m), HeLa cells, WM-266-4 cells, mouse ES cells, human brain, mouse brain and mouse skin tissues. The data represent mean ± standard deviation of results from at least three independent measurements. The "n" for brain and skin tissue DNA designates the number of different animal/human tissues employed for the study.

	5-mdC	5-HmdC		5-FodC	
	% of dG	/10^6 bases	/10^3 5-mdC	/10^6 bases	/10^3 5-mdC
HEK293T-Tet1 (n=3)	2.7 ± 0.3	960 ± 140	161 ± 35	76 ± 3	12.8 ± 1.6
HEK293T-Tet1m (n=3)	3.6 ± 0.3	32.5 ± 2.1	4.0 ± 0.2	0.23 ± 0.01	0.028 ± 0.002
HeLa (n=3)	2.8 ± 0.2	31.2 ± 3.7	5.3 ± 0.6	0.67 ± 0.03	0.114 ± 0.012
WM-266-4 (n=3)	2.8 ± 0.1	12.2 ± 2.4	2.1 ± 0.4	0.69 ± 0.18	0.116 ± 0.031
Mouse ES*	-	163 ± 11	-	3.5 ± 0.7	-
Human Brain (n=4)	5.3 ± 1.0	1550 ± 250	133 ± 18	1.7 ± 0.4	0.146 ± 0.028
Mouse Brain (n=3)	4.6 ± 1.1	560 ± 70	59 ± 17	1.4 ± 0.2	0.150 ± 0.041
Mouse Skin (Red head, n=3)	3.4 ± 0.2	277 ± 36	35 ± 5	1.2 ± 0.1	0.153 ± 0.013
Mouse Skin (Albino, n=3)	3.4 ± 0.2	217 ± 3	27 ± 1	0.7 ± 0.1	0.085 ± 0.017

	5-CadC		5-HmdU
	/10^6 bases	/10^3 5-mdC	% of dG
HEK293T-Tet1 (n=3)	102 ± 13	17.2 ± 3.0	2.7 ± 0.3
HEK293T-Tet1m (n=3)	0.18 ± 0.02	0.022 ± 0.004	3.6 ± 0.3
HeLa (n=3)	0.27 ± 0.04	0.029 ± 0.006	2.8 ± 0.2
WM-266-4 (n=3)	0.29 ± 0.03	0.049 ± 0.004	2.8 ± 0.1
Mouse ES*	0.83 ± 0.08	-	-
Human Brain (n=4)	0.15 ± 0.02	0.013 ± 0.001	5.3 ± 1.0
Mouse Brain (n=3)	0.12 ± 0.01	0.012 ± 0.002	4.6 ± 1.1
Mouse Skin (Red head, n=3)	0.21 ± 0.03	0.026 ± 0.004	3.4 ± 0.2
Mouse Skin (Albino, n=3)	0.19 ± 0.01	0.023 ± 0.002	3.4 ± 0.2

*The data for mouse ES cells were obtained from triplicate LC-MS/MS/MS measurements of samples from a single off-line HPLC enrichment. The amount of DNA was inadequate for 5-HmdU measurement.

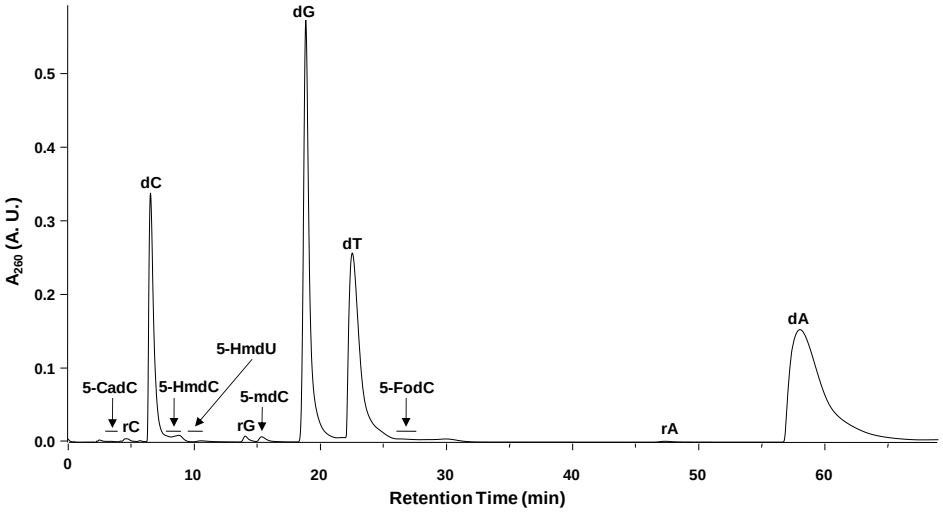


Figure B.1. Representative HPLC trace for the enrichment of 5-FodC, 5-CadC, and 5-HmdU from the enzymatic digestion mixture of genomic DNA isolated from cellular or tissue DNA. Shown is the trace for a DNA sample isolated from skin tissue of an albino mouse.

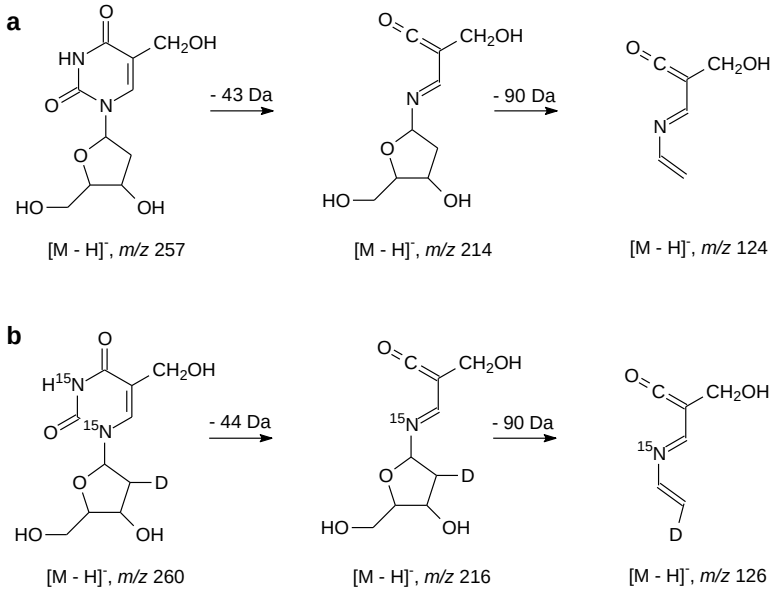


Figure B.2. Proposed fragmentation pathways for the [M – H]– ions of unlabeled (a) and stable isotope-labeled (b) 5-HmdU found in MS/MS and MS/MS/MS.

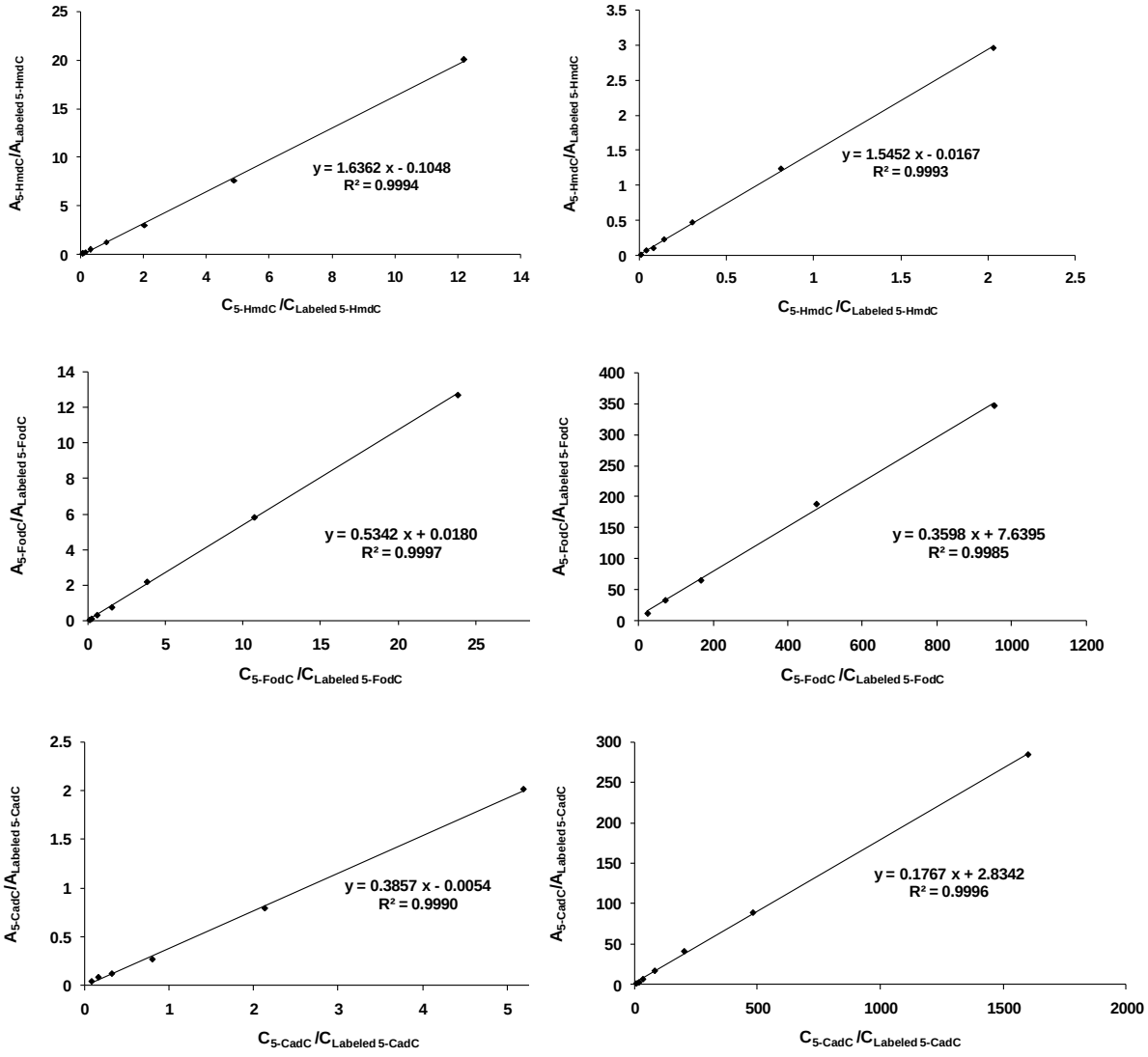


Figure B.3. Calibration curves for the quantification of 5-HmdC, 5-FodC, and 5-CadC by LC-MS/MS/MS with the stable isotope dilution method. Plotted are the ratios of areas of peaks found in the selected-ion chromatograms for monitoring the transitions as shown in Figure 3.2 versus the molar ratios of the unlabeled nucleoside over that of the corresponding stable isotope-labeled standard (30 fmol each for 5-HmdC and 5-FodC, 25 fmol for 5-CadC). Two calibration curves were constructed for different concentration ranges so that the modified nucleosides in a wide concentration range can be accurately quantified.

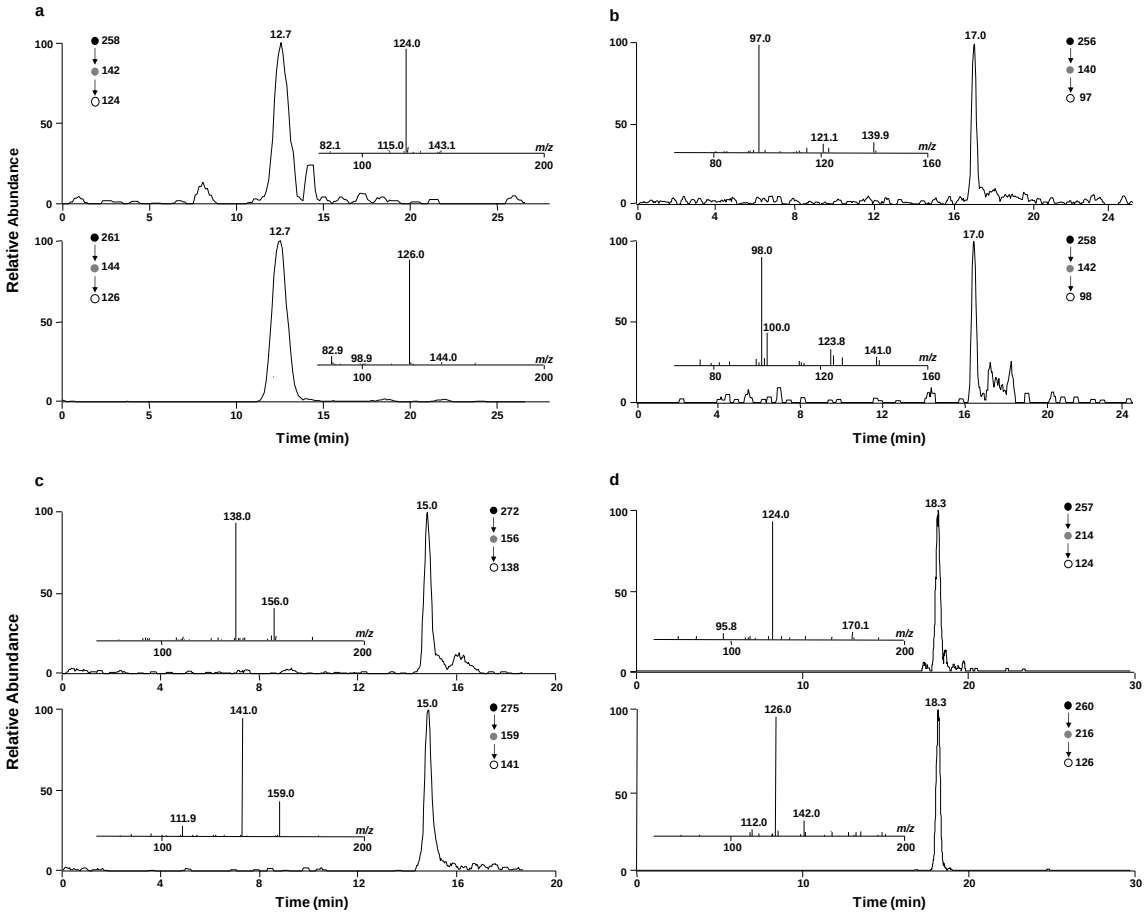


Figure B.4. Representative LC-MS/MS/MS results for the quantification of 5-HmdC (a), 5-FodC (b), 5-CadC (c), and 5-HmdU (d) in skin tissue of an albino mouse. Shown are the selected-ion chromatograms for monitoring the indicated transitions, and the insets give the MS/MS/MS for the analytes and internal standards.

Appendix C: Supporting information for Chapter 5

Table C.1. The relative abundances of 5-mC and its oxidation products, 5-hmC, 5-foC, and 5-caC, in duplex DNA d(AGCTC^{m}CGGTCA)/d(GTGACCGGAGCTG) after incubation with mouse Tet1 and different concentrations of $NaAsO_2$ (i.e. 0, 2, and 5 μM). The data, expressed as average ± standard deviation, represent the results from three independent reactions stopped at 0, 10, and 30 min.

	Time (min)	0 μM $NaAsO_2$	2 μM $NaAsO_2$	5 μM $NaAsO_2$
5-mC	0	100 ± 0	100 ± 0	100 ± 0
	10	16.8 ± 0.3	26 ± 3	42 ± 5
	30	8 ± 3	14 ± 3	15 ± 1
5-hmC	0	-	-	-
	10	42 ± 2	40 ± 6	40 ± 4
	30	16 ± 4	32 ± 6	38.4 ± 0.9
5-foC	0	-	-	-
	10	26 ± 2	20 ± 3	12.6 ± 0.3
	30	19 ± 4	24 ± 7	25.5 ± 0.8
5-caC	0	-	-	-
	10	15 ± 3	14 ± 6	5.6 ± 0.6
	30	57 ± 8	30 ± 10	21 ± 2

Table C.2. The levels of 5-mdC and 5-hmdC in DNA samples isolated from HEK293T cells individually overexpressing the catalytic domains of Tet proteins with exposure to different concentrations of $NaAsO_2$ (i.e. 0, 1, 2, and 5 μM). The data, expressed as average ± standard deviation, represent the results from three independent transfections, $NaAsO_2$ treatments, and LC-tandem MS measurements.

	$NaAsO_2$ (μM)	5-mdC/dG (%)	5-hmdC/dG (%)
Tet1CD	**0**	2.2 ± 0.1	0.331 ± 0.003
	1	2.3 ± 0.2	0.259 ± 0.008
	2	2.6 ± 0.2	0.24 ± 0.03
	5	2.8 ± 0.3	0.16 ± 0.04
Tet2CD	**0**	3.2 ± 0.1	0.16 ± 0.01
	1	3.1 ± 0.3	0.13 ± 0.03
	2	3.2 ± 0.1	0.11 ± 0.02
	5	3.2 ± 0.4	0.08 ± 0.01
Tet3CD	**0**	3.7 ± 0.1	0.099 ± 0.003
	1	3.6 ± 0.1	0.073 ± 0.004
	2	3.5 ± 0.2	0.060 ± 0.007
	5	3.8 ± 0.3	0.043 ± 0.002

Table C.3. The levels of 5-mdC and 5-hmdC in DNA samples isolated from mouse embryonic stem cells with exposure to different concentrations of $NaAsO_2$ (i.e. 0, 1, 2, and 5 μM). The data, expressed as average ± standard deviation, represent the results from three independent transfections, $NaAsO_2$ treatments, and LC-tandem MS measurements.

$NaAsO_2$ (μM)	5-mdC/dG (%)	5-hmdC/dG (%)
0	4.5 ± 0.3	0.092 ± 0.002
1	4.4 ± 0.6	0.096 ± 0.011
2	4.9 ± 0.3	0.070 ± 0.004
5	4.9 ± 0.5	0.063 ± 0.001

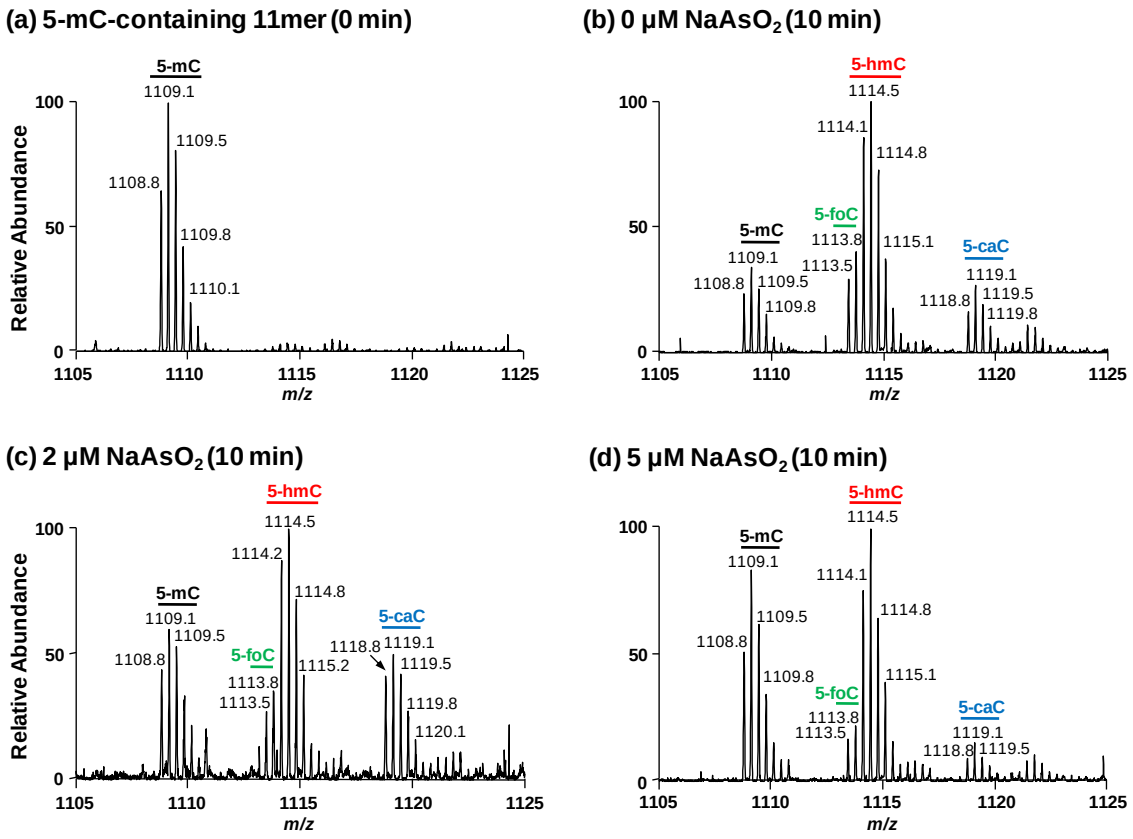


Figure C.1. LC-ESI-MS for monitoring the Tet1-mediated oxidation of 5-mC in a duplex DNA d(AGCTCmCGGTCA)/d(GTGACCGGAGCTG) at 0- and 10-min. Shown are the higher resolution "ultra-zoom-scan" ESI-MS results for monitoring the $[M - 3H]^{3-}$ ions of the initial 5-mC-bearing 11mer DNA (a) and its corresponding products with the 5-mC being oxidized to 5-hmC (b), 5-foC (c), and 5-caC (d).

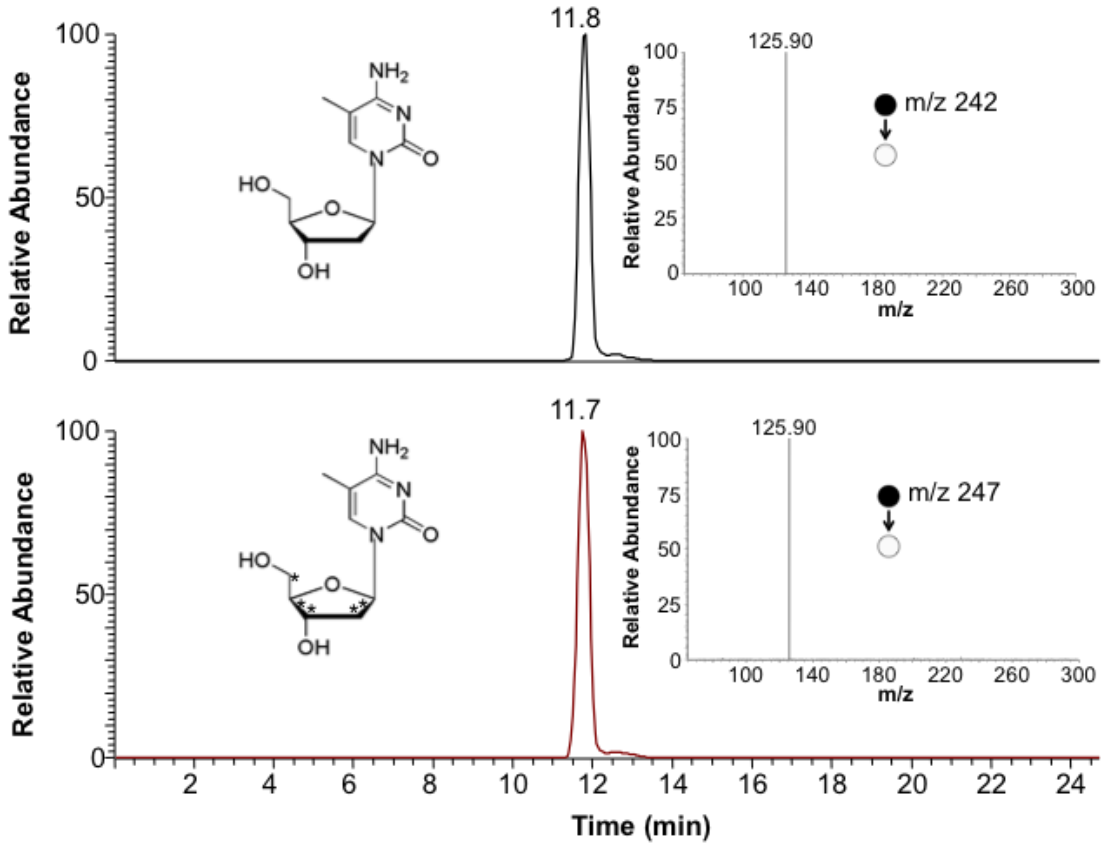


Figure C.2. LC-MS/MS analysis of purified [$^{13}C_5$]-5-mdC in the presence of unlabeled 5-mdC. The top panel is the selected-ion chromatogram for the *m/z* 242→126 transition, monitoring the loss of the 2-deoxyribose moiety from the $[M+H]^+$ ion of unlabeled 5-mdC. The bottom panel is the selected-ion chromatogram for the *m/z* 247→126 transition, monitoring the loss of [$^{13}C_5$]-2-deoxyribose moiety from the $[M+H]^+$ ion of unlabeled 5-mdC. The MS/MS for the $[M+H]^+$ ions of 5-mdC and [$^{13}C_5$]-5-mdC are shown in the insets.

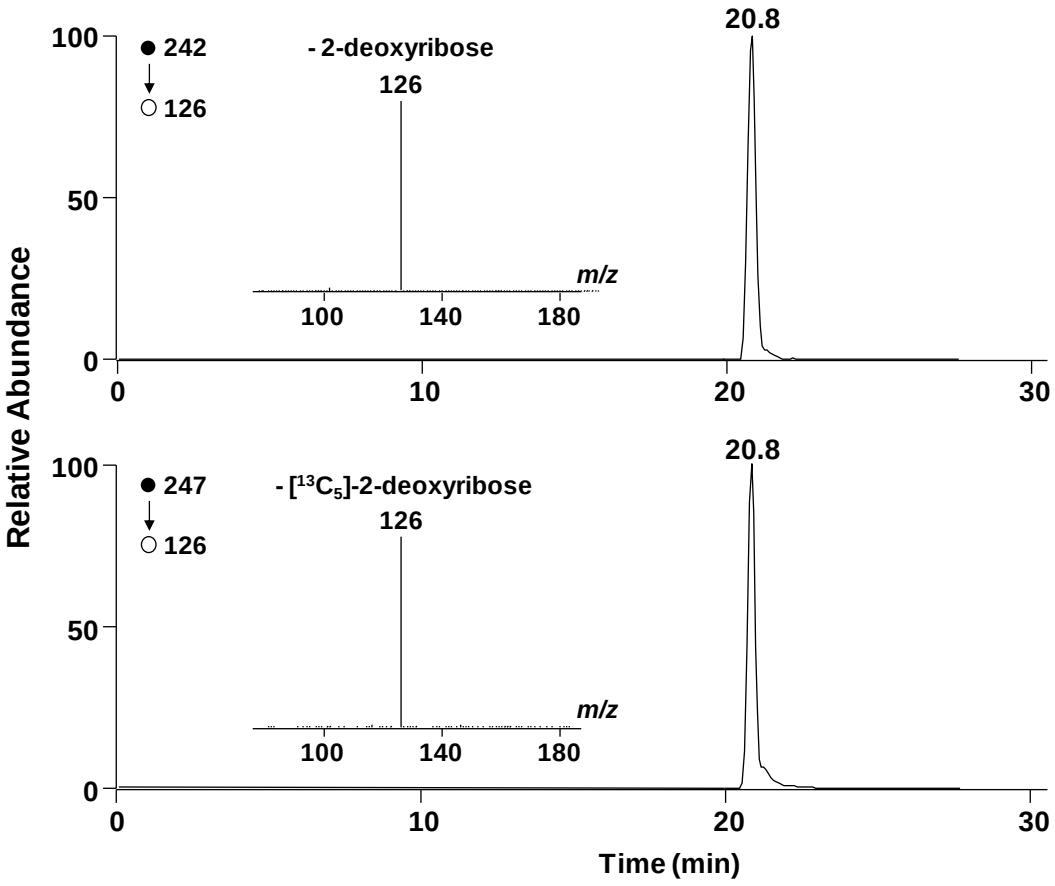


Figure C.3. Representative LC-MS/MS results for the quantification of 5-mdC in mouse embryonic stem cells. Shown are the selected-ion chromatograms for monitoring the indicated transitions for 5mdC (top trace) and [$^{13}C_5$]-5-mdC (bottom trace). Shown in the insets are the corresponding MS^2 for the analyte (top) and the internal standard (bottom).

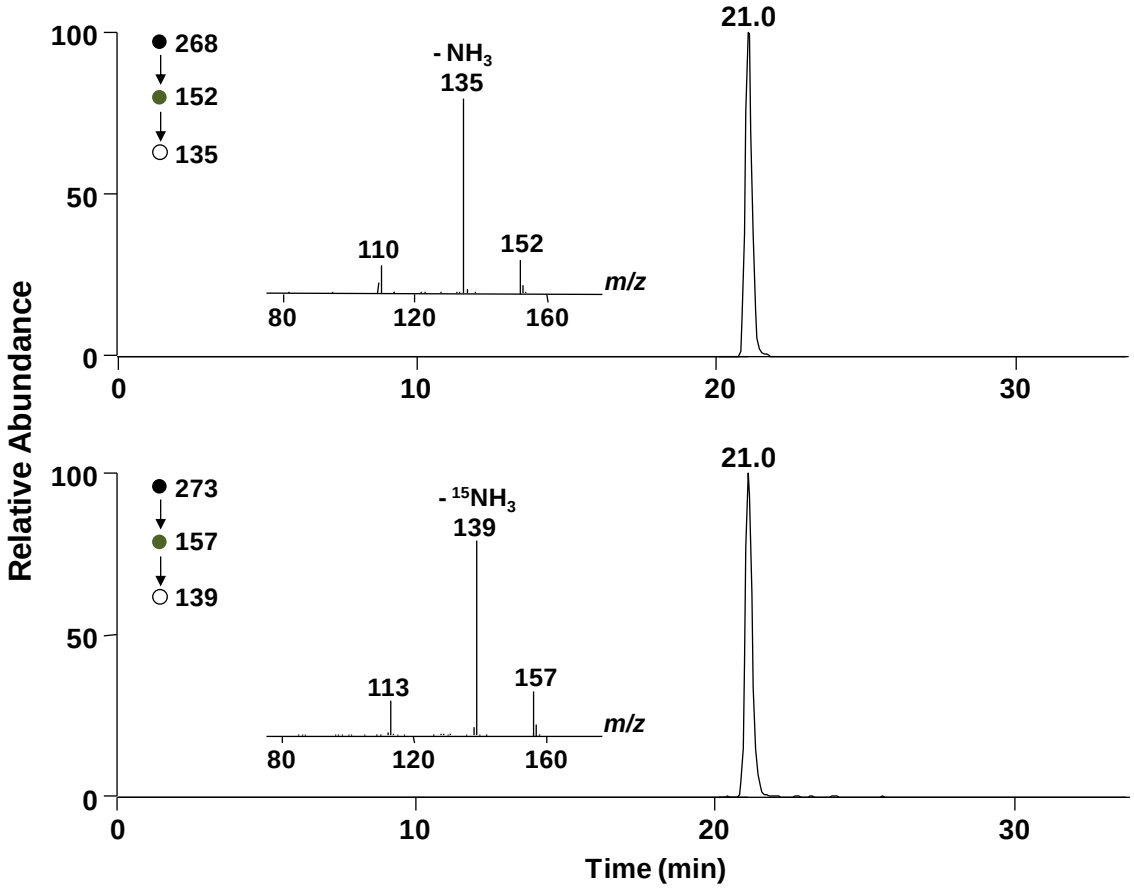


Figure C.4. Representative LC-MS/MS/MS results for the quantification of dG in mouse embryonic stem cells. Shown are the selected-ion chromatograms for monitoring the indicated transitions for dG (top trace) and [$^{15}N_5$]-dG (bottom trace), and the insets give the corresponding MS^3 for the analyte (top) and the internal standard (bottom).

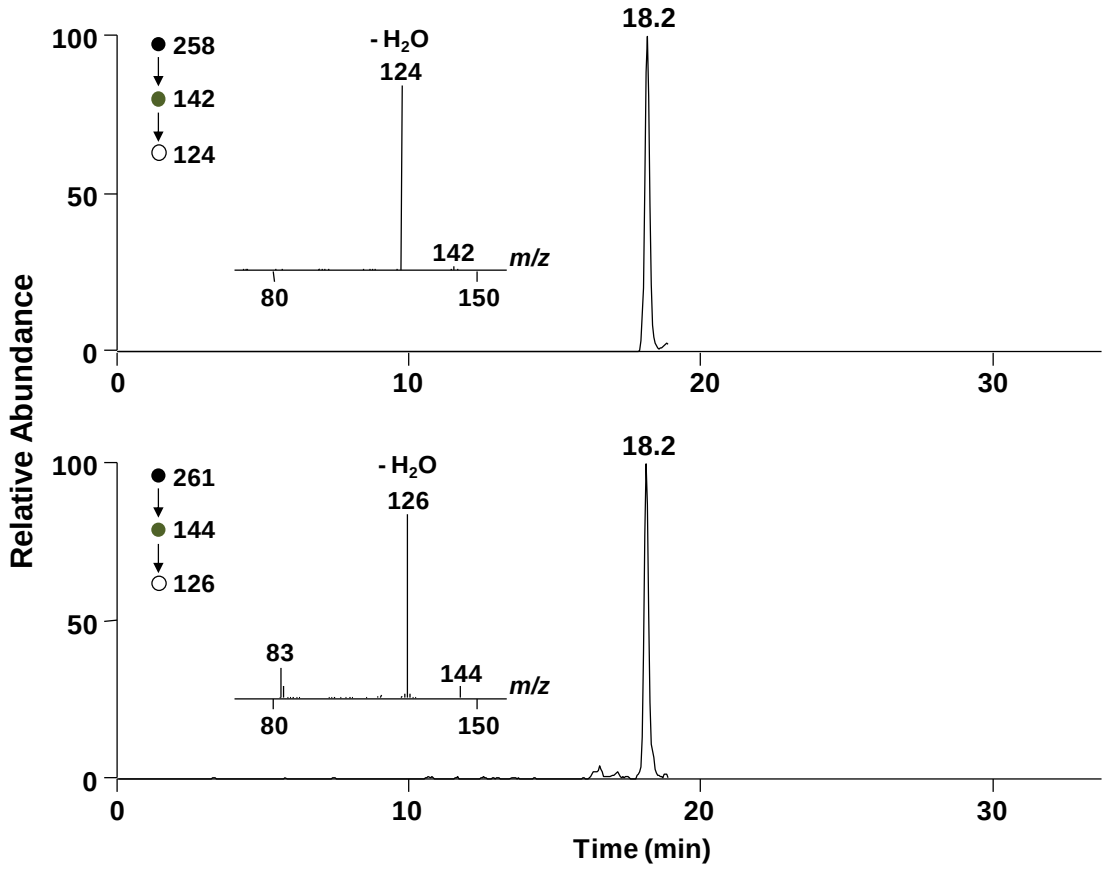
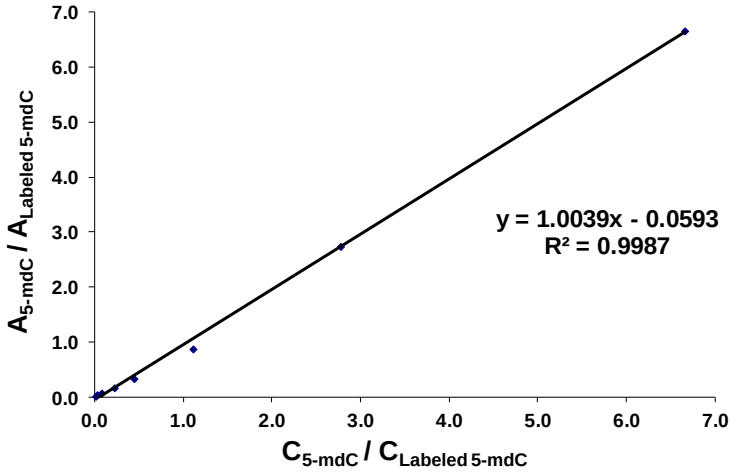


Figure C.5. Representative LC-MS/MS/MS results for the quantification of 5-hmdC in mouse embryonic stem cells. Shown are the selected-ion chromatograms for monitoring the indicated transitions for 5-hmdC (top trace) and [1,3-$^{15}N_2$-2′-D]-5-hmdC (bottom trace), and displayed in the insets are the corresponding MS^3 for the analyte (top) and the internal standard (bottom).

(a)



(b)

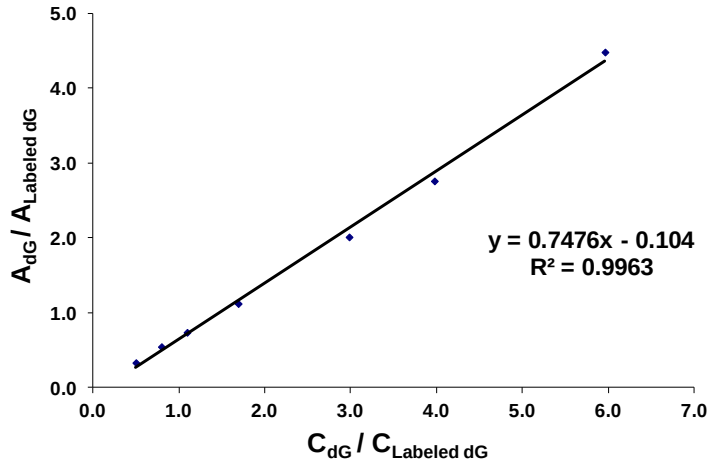


Figure C.6. Calibration curves for the quantifications of 5-mdC (a) and dG (b) by LC-tandem MS with the stable isotope dilution method. Plotted are the ratios of areas of peaks found in the selected-ion chromatograms for monitoring the transitions as shown in Figure C.3 and C.4 versus the molar ratios of the unlabeled nucleoside over that of the corresponding stable isotope-labeled standard (600 fmol for 5-mdC and 16.2 pmol for dG).

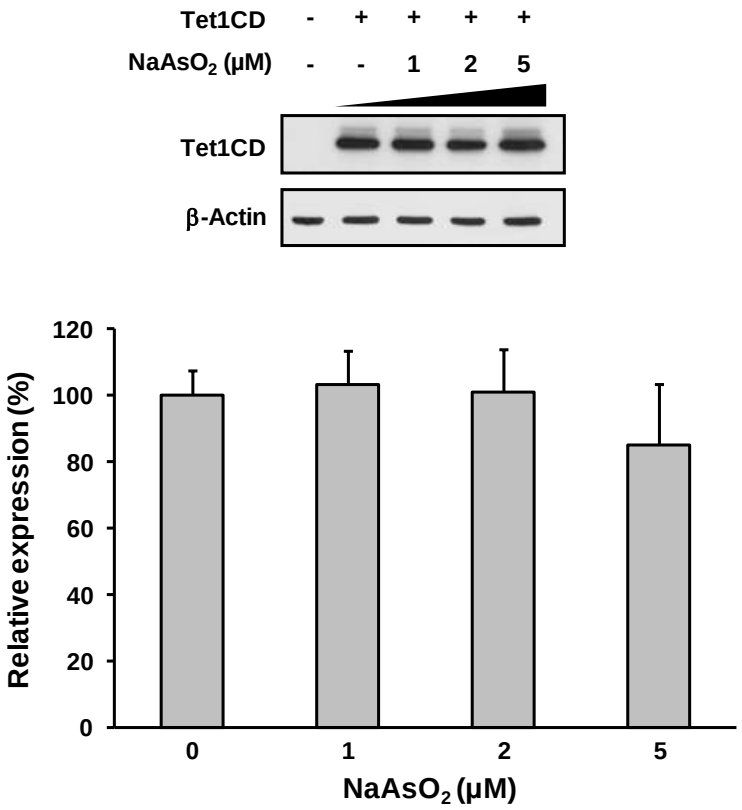


Figure C.7. $NaAsO_2$ treatments did not significantly alter the expression level of Tet1CD in HEK293T cells that were pre-transfected with Tet1CD plasmid. Shown are the expression levels of Tet1CD protein measured by Western blot with the use of β-actin as the loading control. The quantification data represent the mean and standard deviation of results from six biological replicates (calculated based on the ratios of band intensities of Tet1CD over β-actin and normalized to what was found for control cells without $NaAsO_2$ treatment).

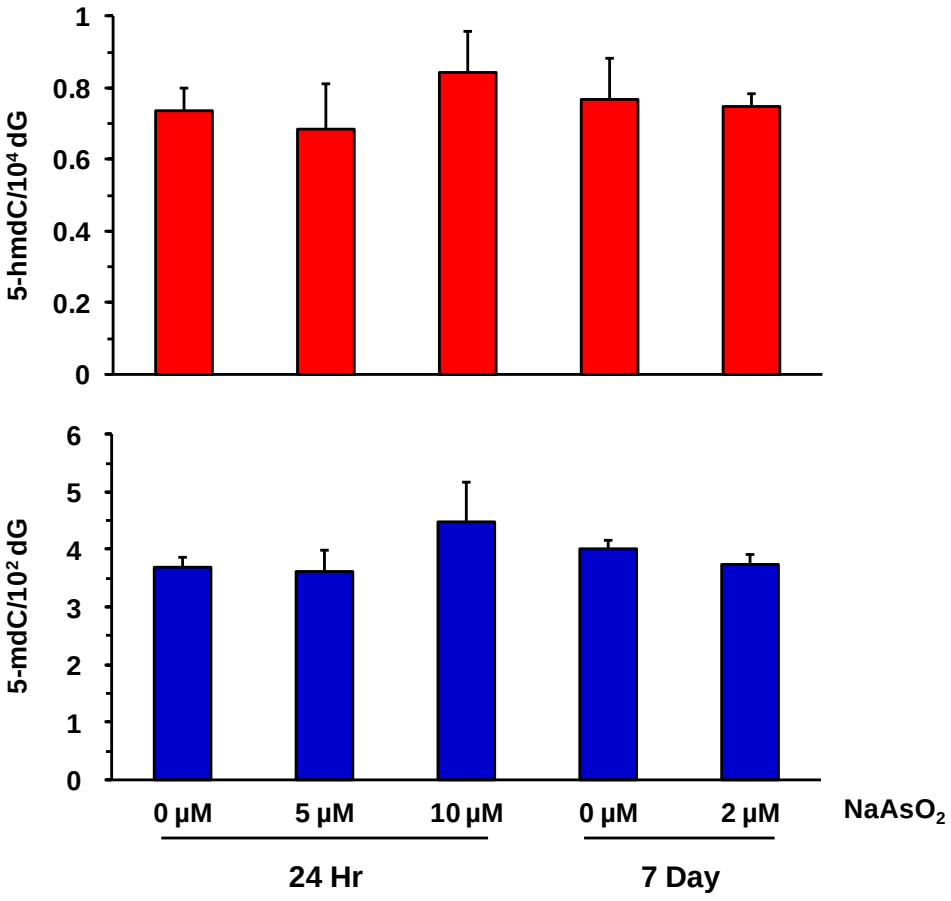


Figure C.8. High concentration exposure or prolonged treatment of $NaAsO_2$ did not lead to a significant alteration in the levels of 5-hmdC and 5-mdC in HEK293T cells. Shown are the levels of 5-hmdC and 5-mdC in DNA isolated from cells following a 24-hr treatment with 5 or 10 μM $NaAsO_2$ or following a 7-day exposure to 2 μM $NaAsO_2$.

Appendix D: Supporting information for Chapter 6

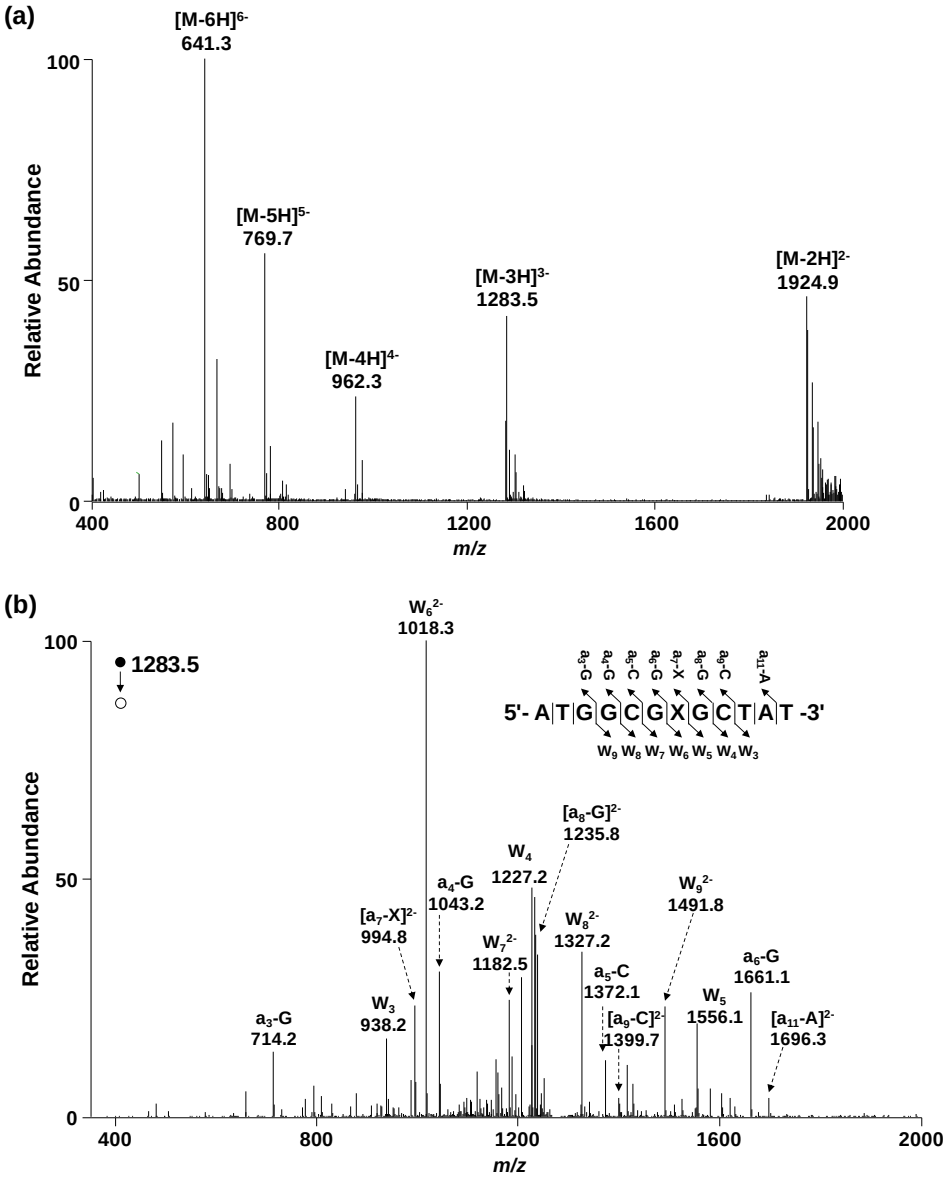


Figure D.1. Negative-ion ESI-MS of the standard ODN harboring a 5-gHmdC (a) and the product-ion spectrum of the $[M-3H]^{3-}$ ion (b).

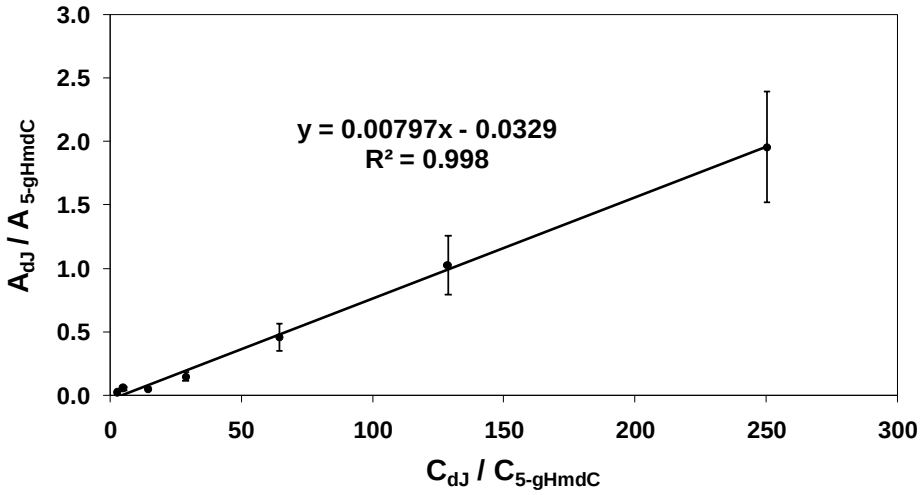


Figure D.2. The calibration curve for the quantification of dJ by LC-MS/MS with the use of a surrogate internal standard method. Plotted are the ratios of areas of peaks found in the selected-ion chromatograms for monitoring the transitions as shown in Figures 6.2b and 6.2a versus the molar ratios of dJ over that of 5-gHmdC standard (100 fmol for 5-gHmdC-containing ODN was used). The data represent the mean and standard deviation of the measurement results obtained on three separated days.