

A Biochemical Characterization of the DNA Glycosylase DEMETER

By

Sonja Claudia Brooks

Dissertation

Submitted to the Faculty of the  
Graduate School of Vanderbilt University  
in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical and Physical Biology

August, 2014

Nashville, Tennessee

Approved:

Walter Chazin

Carmelo Rizzo

Martin Egli

Neil Osheroff

Brandt Eichman

For Braden,  
Thank you for all your love and support

## ACKNOWLEDGMENTS

This work was supported through a National Science Foundation Graduate Research Fellowship grant, a position on the National Institutes of Health Molecular Biophysics Training Program, and funding from the National Institutes of Health.

I am grateful for the continuous support and guidance from my advisor, Brandt Eichman. His excitement and dedication to his research projects has always been inspiring. I would also like to thank my committee members for their direction throughout my dissertation. In addition, the support of my lab members over the last few years has been invaluable. I especially thank Emily Rubinson and Suraj Adhikary, who taught me the ins and outs of the lab and structural biology, while we were becoming friends. I am forever grateful for current lab members and friends Briana Greer and Claire Cato, for always offering their help when needed and having plenty of fun in the process.

Finally, I thank my family and friends for their continued support. My parents have always encouraged me to pursue my dreams, and for that I am forever grateful. My friends in Nashville have made this experience fun, and I will fondly remember our adventures here. I would not be here without the love and support of my fiancé Braden, who encouraged me from qualifying exam to defense.

# TABLE OF CONTENTS

	Page
DEDICATION .....	ii
ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
LIST OF ABBREVIATIONS.....	ix
 Chapter	
I. Introduction.....	1
5mC is an Epigenetic Mark.....	1
Mechanism of Transcriptional Repression .....	5
DNA Methyltransferases .....	6
Dysregulation of Methylation Leads to Disease.....	11
Environmental Epigenetics .....	13
Base Excision Repair .....	14
Passive DNA Demethylation .....	17
Active DNA Demethylation .....	17
5mC Oxidation Derivatives .....	18
Active DNA Demethylation in Plants.....	24
Scope of this Work.....	27
II. Structural Mechanisms of DNA Glycosylases .....	29
Introduction.....	29
Oxidative Damage .....	33
Alkylation Damage .....	35
AAG.....	37
Helix Hairpin Helix Superfamily.....	42
Uracil/Thymine/5mC .....	51
A Possible Role of BER in DNA demethylation .....	53
Structural Insight into TDG Function.....	54
MBD4 .....	62
DME/ROS1.....	63

III.	5-methylcytosine Recognition by <i>Arabidopsis thaliana</i> DNA Glycosylases	
	DEMETER and DML3 .....	66
	Introduction.....	66
	Materials and Methods.....	70
	Protein Expression and Purification.....	70
	Glycosylase Activity Assay .....	71
	DNA Binding .....	72
	DME Homology Model .....	73
	Results.....	73
	A Model for DNA Binding by DME .....	73
	Specificity of DME and DML3 for oxidized 5mC .....	79
	Active Site Residues Confer Specificity for Substrate DNA.....	83
	Discussion .....	86
IV.	Discussion and Future Directions .....	92
	DNA Glycosylases Regulate DNA Methylation .....	92
	DME is a Multi-Domain Glycosylase: the Role of Domain B .....	94
	DME Paralogs Have Distinct Functions .....	95
	DME Unstructured Regions of Unknown Function .....	97
	Role of the Iron-Sulfur Cluster .....	99
	Epigenetics and Human Health.....	101
	State of the Field and Future Directions .....	101
	Appendix I .....	104
	Appendix II .....	110
	REFERENCES .....	113

## LIST OF TABLES

Table	Page
1. DNA binding and 5mC excision activities for DME point mutants.....	77
2. Rates of excision of oxidized 5mC derivatives by DME and DML3 .....	79
3. Substrate specificity of DME.....	81
4. Effects of predicted DME active site residues on nucleobase excision.....	86
5. Thermal denaturation of 5mC oxidation derivatives .....	106
A1. DME, DML3, and DMLa Constructs .....	110

## LIST OF FIGURES

Figure	Page
1. DNA methyltransferase mechanism .....	8
2. Base Excision Repair (BER) Pathway.....	16
3. 5mC derivatives .....	20
4. Common DNA lesions.....	30
5. Chemical reaction catalyzed by DNA glycosylases .....	31
6. DNA glycosylase structural superfamilies.....	33
7. Alkylpurine DNA glycosylases .....	38
8. Binding of $\epsilon$ A and $\epsilon$ C to AAG.....	41
9. Yeast 3-methyladenine DNA glycosylases MAG and Mag1 .....	47
10. Protein-DNA contacts within the TDG active site .....	56
11. Comparison of TDG and MBD4 contacts to the strand opposite the lesion.....	59
12. SUMO1 modified TDG creates steric clash with DNA.....	61
13. DME paralogs and constructs .....	69
14. 5mC excision kinetics and DNA binding of DME .....	74
15. DME domain structure and distribution of critical residues .....	76
16. Identification of DNA intercalation residues in DME and DML3 .....	78
17. Excision of oxidized 5mC nucleobases by DME and DML3.....	80
18. Inhibition of 5mC and T/G excision by reaction intermediates.....	82
19. DME binding to 5mC- and T/G-DNA.....	83
20. DME active site mutants affect base excision .....	84
21. 5hmC, 5caC, and T/G excision kinetics.....	85

22. Representative thermal melt of 5fC-DDD .....106

## LIST OF ABBREVIATIONS

$\mu\text{M}$	Micromolar
1mA	<i>N1</i> -methyladenine
3mA	<i>N3</i> -methyladenine
3mC	<i>N3</i> -methylcytosine
3mG	<i>N3</i> -methylguanine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
5-OHC	5-hydroxycytosine
5-OHU	5-hydroxyuracil
7mG	<i>N7</i> -methylguanine
8-OHG	7,8-dihydro-8-hydroxyguanine
8-OxoG	7,8-dihydro-8-oxoguanine
A (Ade)	Adenine
AAG	Alkyladenine DNA glycosylase
AID	Activation-induced deaminase
Airn	Antisense Igf2r RNA noncoding
AlkA	3mA DNA glycosylase II
AP	Apurinic/apyrimidinc
APE	AP endonuclease

APOBEC	Apolipoprotein B RNA-editing catalytic component
Base J	Beta-D-glycosyl hydroxymethyluracil
BER	Base excision repair
C (Cyt)	Cytosine
CBP	CREB binding protein
CMT	Chromomethylase
CpG	Cytosine (phosphate) Guanine dinucleotide
CpN	Cytosine (phosphate) Nucleotide dinucleotide
DDD	Dickerson dodecamer
DHT	Dihydrothymine
DHU	Dihyrouracil
DME	DEMETER
DML	DME-like
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DRM	Domains rearranged methyltransferase
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraceticacid
EndoIII	Endonuclease III
ER $\alpha$	Estrogen receptor alpha
ES	Embryonic stem cells

FAM	6-carboxyfluorescein
FapyG	2,6-diamino-4-hydroxy-5-formamidopyrimidine
FIE	Fertilization independent endosperm
FIS2	Fertilization independent seed 2
G (Gua)	Guanine
GADD45 $\alpha$	Growth arrest and DNA damage-inducible protein 45 $\alpha$
GFP	Green fluorescent protein
Gh	Guanidinohydantoin
GUS	$\beta$ -glucuronidase
H2TH	Helix-two turn-helix
H3K4	Histone 3 Lysine 4
H3K9	Histone 3 Lysine 9
HDAC1	Histone deacetylase 1
HhH	Helix hairpin helix
HOMe-U	Hydroxymethyldeoxyuridine
Hx	Hypoxanthine
Ia	Iminoallantoin
Igf2	Insulin-like growth factor 2
Igf2r	Insulin-like growth factor 2 receptor
IPTG	Isopropyl thio- $\beta$ -D-galactopyranoside
JBP	J binding protein
$K_{1/2}$	Apparent $K_M$
kcal	Kilocalorie

$K_d$	Equilibrium dissociation constant
kDa	Kilodalton
$K_i$	Inhibition constant
$K_M$	Michaelis-Menton equilibrium constant
$k_{obs}$	Observed rate constant
K <sub>st</sub>	Single-turnover rate constant
LB	Luria broth
LIG1	DNA ligase I
LIG3	DNA ligase III
LRC	Lesion recognition complex
MAG	Methyladenine DNA glycosylase
MALDI	Matrix assisted laser desorption ionization
MBD	Methyl-CpG-binding domain
MBD4	Methyl-CpG-binding domain protein 4
MEA	MEDEA
MeDIP	Methyl-DNA immunoprecipitation
mFapyG	<i>N</i> 7-methylFapyG
MIG	Mismatch specific glycosylase
mL	Milliliter
mM	Millimolar
mol	Mole
Mpg	Methylpurine DNA glycosylase
MPOA	Medial preoptic area

MS	Mass spectrometry
MUG	Mispaired uracil glycosylase
MutY	A/G specific adenine glycosylase
NHE	Normal Hydrogen electrode
Ni-NTA	Nickel-nitrolotriacetic acid
NMR	Nuclear magnetic resonance
Nth	Endonuclease III
OGG	8-oxoguanine DNA glycosylase
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PGCs	Primordial germ cells
POL II	DNA Polymerase II
POL $\beta$	DNA polymerase $\beta$
RAR	Retinoic acid receptor
RD	Regulatory domain
RdDM	RNA directed DNA methylation
RNA	Ribonucleic acid
ROS	Reactive oxygen species
ROS1	Repressor of silencing 1
RXR	Retinoid X receptor
SAM	S-adenosylmethionine

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIM	SUMO-interacting site
SMUG	Single-stranded monofunctional uracil glycosylase
Sp	Spiroiminodihydantoin
ssDNA	Single-stranded DNA
SUMO	Small ubiquitin-like modifier protein
T (Thy)	Thymine
TAG	3mA DNA glycosylase I
TBE	Tris Boric Acid EDTA Buffer
TDG	Thymine DNA Glycosylase
TEs	Transposable elements
TET	Ten-eleven translocation protein
Tg	Thymine glycol
THF	Tetrahydrofuran
TOF	Time-of-flight
TRIS	Tris-(hydroxymethyl)-aminomethane
TTF1	Thyroid transcription factor 1
U	Uracil
UDC	Undamaged DNA complex
UDG	Uracil DNA glycosylase
UF	2'-deoxy-2'-fluroarabinouridine
VSG	Variant surface glycoprotein
WT	Wild-type

$\epsilon\text{A}$	1, $N^6$ -ethenoadenine
$\epsilon\text{C}$	3, $N^4$ -ethenocytosine

## **Chapter I**

### **INTRODUCTION**

Genetic information is transmitted through the use of a primary DNA structure composed of four canonical bases: thymine, cytosine, guanine, and adenine. A chemical modification to cytosine to create 5-methylcytosine (5mC) is an example of a more complex transmission of genetic information called epigenetics. Epigenetics refers to heritable changes to DNA and histones that affect gene expression but do not modify the DNA sequence. Cytosine methylation is essential for gene expression, cell differentiation, development, genomic imprinting, X chromosome inactivation, and regulation of disease states. Although 5mC is referred to as the fifth base, it is not synthesized into DNA like the other four bases, which are made into a polymer by DNA polymerases. Instead, other enzymes, such as DNA methyltransferases, are responsible for establishing the methylation patterns that lead to epigenetic control of genes. Although the process of DNA methylation has been well studied, the enzymes responsible for DNA demethylation are less well understood.

#### **5mC is an Epigenetic Mark**

In the 1970s, it was proposed that DNA methylation is responsible for maintaining gene expression patterns through mitotic cell division (Holliday and Pugh 1975; Riggs 1975). Today it is well accepted that DNA methylation patterns contribute to gene expression states by silencing genes through interaction with proteins that modify

nucleosomes (Wolffe and Matzke 1999; Urnov and Wolffe 2001). DNA methylation is found in most plant, animal, and fungal models and has a great impact on genome stability, gene expression, and development (Jaenisch and Bird 2003; Law and Jacobsen 2010; Smith and Meissner 2013).

Although 5mC constitutes only one percent of human DNA, seventy to eighty percent of CpG dinucleotides are methylated, while the majority of CpG islands are not methylated (Ehrlich and Wang 1981; Ehrlich, Gama-Sosa et al. 1982; Weber, Davies et al. 2005). However, CpG islands are often hypermethylated during tumorigenesis or ageing (See *Dysregulation of Methylation Leads to Disease* below) (Meissner, Mikkelsen et al. 2008). In addition, there exists a small amount of non-CG methylation in embryonic stem (ES) cells (Ramsahoye, Biniszkiwicz et al. 2000; Bird 2002; Lister, Pelizzola et al. 2009). The spectrum of DNA methylation in other animals is very broad: the nematode worm *Caenorhabditis elegans* has no detectable 5mC, while vertebrates have very high levels of 5mC (Colot and Rossignol 1999; Hung, Karthikeyan et al. 1999; Tweedie, Ng et al. 1999; Bird 2002). In plants, methylation occurs in all cytosine sequence contexts (Henderson and Jacobsen 2007). Specifically in *Arabidopsis thaliana*, twenty-four percent of CpG sites, seven percent of CpNpG, and two percent of CpNpN sites are methylated, where N is any nucleotide (Cokus, Feng et al. 2008). DNA methylation predominantly occurs on transposons and other repetitive elements (Zhang, Yazaki et al. 2006). It has been suggested that DNA methylation may have evolved as a defense against transposable elements (TEs) (Goll and Bestor 2005). Because transposons may cause mutagenesis and threaten genomic integrity by potentially causing antisense transcripts that activate RNAi pathways or by promoting improper gene expression, most

genomic methylation is on transposons, in order to regulate and inactivate them (Chang-Yeh, Mold et al. 1991; Takahara, Ohsumi et al. 1996; Yoder, Walsh et al. 1997). Evidence for this hypothesis exists in plants, as in cases of genome demethylation, the rate of transposon insertion increases (Hirochika, Okamoto et al. 2000; Miura, Yonebayashi et al. 2001; Singer, Yordan et al. 2001; Kato, Miura et al. 2003). Transposons can be permanently inactivated when the methylated cytosines are deaminated, leading to C → T transition mutations (Schorderet and Gartler 1992).

Epigenetic modifications can be inherited during development of an organism (during mitosis) or across generations (reviewed in Danchin, Charmantier et al. 2011). The transmission of epigenetic marks, such as cytosine methylation (see *DNA Methyltransferases* below) results in transgenerational epigenetic inheritance or mitotic epigenetic inheritance, which allows for cell differentiation. Transgenerational epigenetic inheritance can occur through germline epigenetic inheritance or experience-dependent epigenetic inheritance. In germline epigenetic inheritance, the epigenetic state of the DNA is present in germline cells and is transmitted over many generations. One of the first examples of this inheritance due to environmental changes was observed through the prenatal exposure of the pesticide vinclozolin. Resulting changes in DNA methylation were inherited to the F<sub>4</sub> generation and beyond (Anway, Cupp et al. 2005). On the other hand, experience-dependent epigenetic inheritance occurs when an epigenetic state in a parent induces behavior that generates the same epigenetic state in the offspring. For example, low levels of licking/grooming due to early weaning of female mouse pups results in methylation of the estrogen receptor alpha (ER $\alpha$ ) promoter, thereby reducing the amount of ER $\alpha$  expressed in the medial preoptic area (MPOA) of

the brain. The MPOA has been found to be necessary for post-partum maternal interactions with offspring; thus, the pups that are weaned early exhibit reduced maternal care as adults (Gonzalez, Lovic et al. 2001; Lovic, Gonzalez et al. 2001; Kikusui, Isaka et al. 2005; reviewed in Champagne and Meaney 2007).

Transgenerational epigenetic inheritance includes parent-of-origin expression patterns established through gene imprinting. Imprinted genes are expressed from the allele of one parent, while the other is silenced. This control of gene expression occurs in insects, mammals, and flowering plants (Jurkowska and Jeltsch 2013). In humans, approximately one hundred genes are effected by gene imprinting, which allow for proper growth and development (Jurkowska and Jeltsch 2013). The first evidence for gene imprinting was observed in 1984, when it was determined that a copy of both the maternal and paternal genomes were necessary for embryogenesis in mice. A bimaternal zygote, in which the paternal pronucleus was replaced with a maternal pronucleus, was not able to complete embryogenesis; a bipaternal zygote suffered the same problem (Barton, Surani et al. 1984; McGrath and Solter 1984; Surani, Barton et al. 1984). Examples of specific imprinted genes followed a few years later, when *insulin-like growth factor 2 (Igf2)*, which expresses from the paternal allele, and *insulin-like growth factor type-2 receptor (Igf2r)* and H19 noncoding RNA, which express from the maternal allele, were identified in mice (Barlow, Stoger et al. 1991; Bartolomei, Zemel et al. 1991; DeChiara, Robertson et al. 1991; Ferguson-Smith, Cattanaach et al. 1991). It is now known that DNA methylation, combined with additional histone modifications, is responsible for the differential expression of maternal or paternal alleles (Bartolomei, Webber et al. 1993; Ferguson-Smith, Sasaki et al. 1993; Stoger, Kubicka et al. 1993).

Gene imprinting affects growth regulation in the embryo and affects metabolism and behavior (Jurkowska and Jeltsch 2013).

### **Mechanism of Transcriptional Repression**

DNA methylation may function to prevent transcription by disrupting the binding of a protein to the DNA sequence (Bird 2002). Some factors that are known to bind CpG sequences are not able to bind the DNA when the CpG is methylated. For example, the transcription factor CTCF, which is involved in imprinting the *H19/Igf2* locus in mice, is not able to bind the promoter of the paternal allele due to the methylation at CpG sites. This allows for the downstream enhancer to activate *Igf2* expression of the paternal allele, while the maternal allele is silenced as a result of CTCF binding to the maternal promoter and downstream enhancer (Bell, West et al. 1999; Bell and Felsenfeld 2000; Hark, Schoenherr et al. 2000; Szabo, Tang et al. 2000; Holmgren, Kanduri et al. 2001). Alternatively, the sites of methylation may recruit proteins that preferentially bind 5mC over C, functioning to repress transcription (Bird 2002). Five methyl-CpG-binding proteins have been characterized that are related to the methyl-CpG-binding domain (MBD) of methyl CpG binding protein 2 (MeCP2) (Nan, Meehan et al. 1993; Cross, Meehan et al. 1997; Nan, Campoy et al. 1997; Hendrich and Bird 1998). MBD1, MBD2, MBD3, and MeCP2 have been shown to repress transcription in a methylation-dependent manner (reviewed in Bird and Wolffe 1999).

The mechanism of gene imprinting functions through a cluster of genes that contains some imprinted genes and other non-imprinted genes. Imprinting control regions (ICRs) contain a region with parent-specific methylation patterns, called a

differentially methylated region (DMR). Usually each cluster is associated with one ICR that functions to regulate expression (Barlow 2011). For example, the *Igf2r* gene, one of the first genes implicated as an imprinted gene, is regulated by a large noncoding RNA called Airn (antisense *Igf2r* RNA noncoding), which is expressed from the paternal allele. Because Airn is transcribed antisense to *Igf2r*, the expression of *Airn* prevents the expression of *Igf2r* from the paternal chromosome (Sleutels, Zwart et al. 2002).

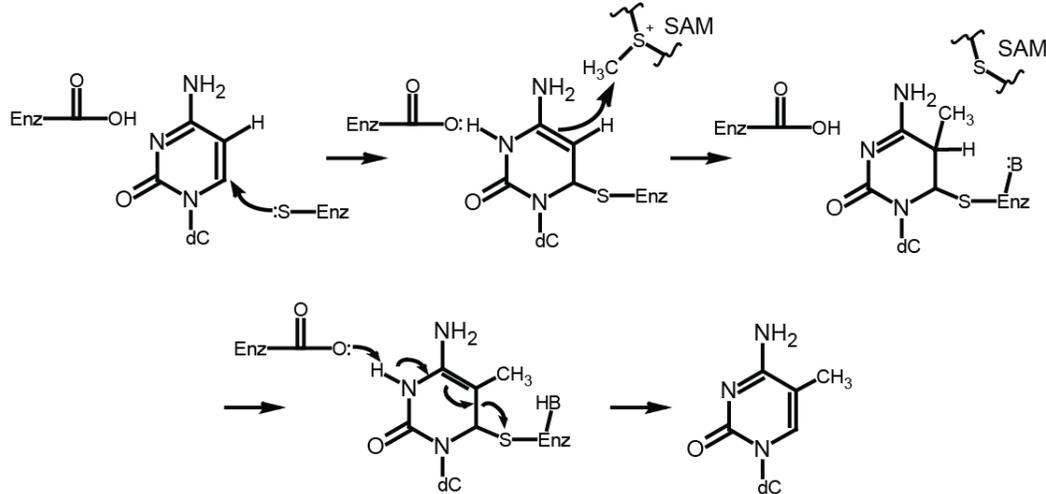
### DNA Methyltransferases

Nearly 40 years ago, two classes of DNA methyltransferases were predicted to be responsible for the stable inheritance of DNA methylation across cell cycles and generations (Holliday and Pugh 1975; Riggs 1975) (Bestor, Laudano et al. 1988). Since then, the mechanism and functions for the essential DNA methyltransferases have been characterized. DNA methyltransferases, including DNA methyltransferase 1 (DNMT1) and DNA methyltransferase 3A and 3B (DNMT3A and DNMT3B) allow for DNA methylation patterns to be maintained throughout many cell cycles (Wigler, Levy et al. 1981; Schubeler, Lorincz et al. 2000). The function of these enzymes is critical for development. Indeed, null mutations of DNMT1 or DNMT3A/B are embryonic lethal (Li, Bestor et al. 1992; Okano, Bell et al. 1999; Chan, Henderson et al. 2005; Goll and Bestor 2005).

DNA methyltransferases catalyze the transfer of a methyl group to the 5 position of cytosine residues (**Figure 1**). Although the methylation of the base at this location is predicted to be chemically improbable, methyltransferases utilize a covalent catalysis mechanism to overcome this (Santi, Garrett et al. 1983; Chen, MacMillan et al. 1991). A

cysteine residue that is conserved and essential in eukaryotic methyltransferases forms a covalent bond to C6 of cytosine (Bestor and Verdine 1994; Hsieh 1999). Ten conserved motifs are utilized by both eukaryotic and bacterial methyltransferases (Lauster, Trautner et al. 1989; Posfai, Bhagwat et al. 1989; Kumar, Cheng et al. 1994). The function of each motif has been elucidated through crystallographic studies of the enzymes in complex with intermediates of the reaction (Trautner, Balganesch et al. 1988; Wilke, Rauhut et al. 1988; Reinisch, Chen et al. 1995; Dong, Yoder et al. 2001). Sequence specificity for the bacterial methyltransferases is determined by a region between motif VIII and IX; however, eukaryotic methyltransferases do not have innate sequence specificity (Trautner, Balganesch et al. 1988; Wilke, Rauhut et al. 1988). Like DNA glycosylases, DNA methyltransferases utilize a base flipping mechanism to access C5 and C6 of the base (See Chapter II for DNA glycosylase mechanism) (Horton, Ratner et al. 2004). DNA methyltransferases have a well conserved “large domain,” which includes the active site motifs and the binding site for SAM (AdoMet).

DNA methyltransferases are now grouped into four families based on sequence homology within the C-terminal catalytic domain (Goll and Bestor 2005). The families include DNA methyltransferase 1, 2, and 3 (DNMT1, DNMT2, and DNMT3, respectively), and the chromomethylase family, found only in plants. DNMT2 is present in all organisms that contain DNMT1 and DNMT3 homologs, but in some organisms, a DNMT2 homolog is the only DNA methyltransferase. *Caenorhabditis elegans* and *Saccharomyces cerevisiae* do not contain any methyltransferase homologs (Goll and Bestor 2005).



**Figure 1.** DNA methyltransferase mechanism. DNA methyltransferases utilize a covalent catalysis mechanism in order to transfer a methyl group from *S*-adenosylmethionine (SAM) to position 5 on a cytosine base, creating 5-methylcytosine (5mC). [Adapted from (Goll and Bestor 2005).]

DNMT1, the first DNA methyltransferase to be identified on the basis of *de novo* methylase activity, is responsible for maintaining symmetrically methylated CpGs after DNA replication during mitosis. DNMT1 methylates hemimethylated substrates better than unmethylated DNA (Bestor, Laudano et al. 1988; Yoder, Soman et al. 1997; Song, Rechkooblit et al. 2011). DNMT1 is expressed in dividing cells and is recruited to DNA replication sites through interactions with proliferating cell nuclear antigen (PCNA) and ubiquitin-like with PHD and RING finger domains 1 (UHRF1), which specifically directs DNMT1 to hemimethylated sites (Chuang, Ian et al. 1997; Kishikawa, Murata et al. 2003; Bostick, Kim et al. 2007; Sharif, Muto et al. 2007). The activity of DNMT1 is dependent on a larger complex of proteins that regulate its activity through post-translational modifications, which oppose DNMT1 activity. The amino-terminal domain of DNMT1 and the E3 ubiquitin ligase RING domain of UHRF1 are targets for the histone acetyltransferase TIP60 and methyltransferase SET7 (Esteve, Chin et al. 2009;

Du, Song et al. 2010). However, when active DNA-bound DNMT1 is in complex with histone deacetylase 1 (HDAC1), these modifications are opposed (Fuks, Burgers et al. 2000; Robertson, Ait-Si-Ali et al. 2000; Wang, Hevi et al. 2009; Du, Song et al. 2010). The first plant DNA methyltransferase to be identified was MET1, which belongs to the DNMT1 family (Finnegan and Dennis 1993). MET1 is considered a maintenance methyltransferase, functioning to maintain methylation patterns after DNA replication, although some data indicate that it has *de novo* methylation activity as well (Aufsatz, Mette et al. 2004).

DNMT3A and DNMT3B function as *de novo* methyltransferases in mammals. A third homolog, DNMT3L does not have methyltransferase activity but functions as a regulatory factor in germ cells (Bourc'his, Xu et al. 2001; Goll and Bestor 2005). DNMT3A and DNMT3B are able to methylate both hemi- and unmethylated substrates at equal rates (Okano, Xie et al. 1998). These enzymes are expressed in adult tissues, but not as robustly as DNMT1 (Goll and Bestor 2005). DNMT3L is required for maternal genomic imprinting, and *DNMT3L* homozygous null mice are sterile, but viable (Bourc'his, Xu et al. 2001). Evidence suggests that DNMT3A and DNMT3L likely interact to establish maternal genomic imprinting (Chedin, Lieber et al. 2002; Kaneda, Okano et al. 2004). Repressed CpG island promoters, although usually unmethylated, may be silenced through both histone H3 lysine 9 (H3K9) methylation and DNA methylation initiated by DNMT3A and DNMT3B (Ayyanathan, Lechner et al. 2003). In plants, the DNMT3 family is referred to as the domains rearranged methyltransferase (DRM) family. *A. thaliana* encodes DRM1 and DRM2, which establish CpN, CpG, and CpNpG methylation through RNA directed DNA methylation (RdDM) (Cao, Springer et

al. 2000; Cao and Jacobsen 2002; Cao and Jacobsen 2002). The generation of short RNAs signals the methylation of homologous DNA sequences and is a necessary control of DNA methylation (Cao and Jacobsen 2002). Evidence for RdDM has only been found in the plant kingdom, and proteins that are necessary for RdDM in *A. thaliana*, such as an RNA-dependent RNA polymerase or SNF2 chromatin-remodeling protein DRD1, are not present in mammals (Chan, Zilberman et al. 2004; Freitag, Hickey et al. 2004; Kanno, Mette et al. 2004).

The DNMT2 family is the most well conserved and ubiquitous DNA methyltransferase (Yoder and Bestor 1998; Goll and Bestor 2005). However, no DNA methyltransferase activity has been detected for DNMT2, although it contains all ten conserved motifs in the expected order, and the structure of the large domain is very similar to the bacterial methyltransferase M.HhaI (Okano, Xie et al. 1998; Yoder and Bestor 1998; Dong, Yoder et al. 2001; Hermann, Schmitt et al. 2003; Tang, Reddy et al. 2003). Interestingly, the crystallographic position of an invariant tyrosine residue in the DNMT2 family is predicted to make a steric clash with DNA modeled from M.HhaI. This residue is a glycine in M.HhaI, and no other residue can be tolerated at this site (Lee, Tawfik et al. 2002). Furthermore, deletion of DNMT2 does not affect methylation, nor does it cause any phenotypic abnormalities in ES cells (Okano, Xie et al. 1998). Therefore, it is unlikely that DNMT2 methylates duplex DNA.

The chromomethylase family in plants methylates the symmetrical CpNpG, which is rarely methylated in vertebrates (Henikoff and Comai 1998). These enzymes have a chromodomain between the conserved methyltransferase motifs II and IV. Chromodomains, which were first identified in the Polycomb group proteins, function to

target proteins to heterochromatin (Jones, Cowell et al. 2000). *A. thaliana* contains three chromomethylase homologs (CMT1, CMT2, and CMT3), although only CMT3 is necessary for CpNpG methylation (Henikoff and Comai 1998; Bartee, Malagnac et al. 2001; Lindroth, Cao et al. 2001). However, the loss of CpNpG methylation due to mutations in *CMT3* does not cause any severe morphological phenotypes (Bartee, Malagnac et al. 2001; Lindroth, Cao et al. 2001). CpNpG methylation may have evolved as a defense against transposons that evolved CpG-free promoters that were not silenced by CpG methylation (Goll and Bestor 2005).

### **Dysregulation of Methylation Leads to Disease**

The proper establishment and maintenance of DNA methylation is critical for development and continued health of an organism. Dysregulation of cytosine methylation may lead to the pathogenesis of disease. In cancer cells, CpG islands that are normally unmethylated become hypermethylated, thereby silencing normally expressed genes (Gal-Yam, Saito et al. 2008). In addition, repeat sequences that are usually methylated become hypomethylated. This change in DNA methylation also is associated with changes to histone modifications, including reduced H3K4 methylation, increased H3K9/K27 methylation, and decreased acetylation (Riggs and Jones 1983; Feinberg and Tycko 2004; Liang, Lin et al. 2004; Heintzman, Stuart et al. 2007; Kouzarides 2007). These changes are epigenetic patterns common to cancer cells. Hypomethylation in cancer cells may lead to overexpression of oncogenic proteins. For example, the activation of the normally silent maternally imprinted gene *IGF2*, described above, is associated with increased risk of colon cancer (Kaneda and Feinberg 2005). On the other

hand, hypermethylation of CpG islands leads to the silencing of tumor suppressors, such as the *Retinoblastoma* gene, which is hypermethylated in retinoblastoma (Greger, Passarge et al. 1989). Additional examples of this phenomenon have been found in renal cancer (hypermethylation of *von Hippel-Lindau*), bladder cancer (hypermethylation of the cell cycle regulator *DKN2 A/p16*), and colon cancer (hypermethylation of the mismatch repair gene *hMLH1*) (Herman, Latif et al. 1994; Gonzalez-Zulueta, Bender et al. 1995; Kane, Loda et al. 1997). A better understanding of DNA methylation in various cancers may help improve diagnosis and prognosis of disease states.

Aberrant DNA methylation is associated with a number of diseases in addition to various cancers. Hypermethylation of certain gene promoters has been reported in schizophrenia, while hypomethylation of other promoters is associated with bipolar disorder (Kato and Iwamoto, 2014, Wockner, 2014). Aberrant DNA methylation also has been associated with other neurologic disorders, such as Rett syndrome, alpha-thalassemia X-linked retardation (Jakovcevski and Akbarian 2012), Prader-Willi syndrome, Angelman syndrome, Beckwith-Weidemann syndrome, congenital malformation, aging (Ballestar and Esteller 2005), and neoplasia (Baylin, Esteller et al. 2001; Esteller and Herman 2002; Esteller 2003; Robertson 2005; Hansen, Timp et al. 2011). Environmental factors also contribute to the development of psychiatric disorders (Schmitt, Malchow et al. 2014) (See *Environmental Epigenetics* below). Furthermore, differential methylation patterns have been associated with cardiac dysfunction and regeneration (Chaturvedi and Tyagi 2014). As seen in cancer cells, disease states of cardiovascular disorders often contain increased methylation at CpG islands in the promoters of genes (Baccarelli, Rienstra et al. 2010). In addition, the hypomethylation of

DNMTs and methylenetetrahydrofolate reductase (MTHFR) has been implicated in atherosclerosis and cardiovascular disease (Chen, Karaplis et al. 2001; Lund, Andersson et al. 2004). Irregular DNA methylation and gene imprinting result in many more disorders not described here (Paulsen and Ferguson-Smith 2001; Smith and Meissner 2013).

### **Environmental Epigenetics**

Genetic inheritance is often thought of the transmission of DNA sequence from one generation to the next, and species evolution is considered a result of variation in DNA sequence yielding phenotypic variations that are subject to natural selection (Ellegren and Sheldon 2008). However, changes to phenotype also result from differences in gene expression, which is controlled through epigenetic mechanisms (Cubas, Vincent et al. 1999; Richards 2006; Bossdorf, Richards et al. 2008). Although environmental changes to DNA are typically associated with various types of DNA damage leading to mutations or cell death (see *Base Excision Repair* section below), an organism's surroundings also can have an effect on the epigenetic level (Danchin, Charmantier et al. 2011; Varriale 2014). Factors such as xenobiotics, social behavior, metabolism, and nutritional deficiencies can affect epigenetic changes (Anway, Cupp et al. 2005; Petronis 2010; Danchin, Charmantier et al. 2011). Several authors speculate that epigenetic changes may arise due to environmental changes, causing the expression of normally silenced genes or inducing transposon activity (Jaenisch and Bird 2003; Jablonka and Lamb 2007; Bossdorf, Richards et al. 2008; Rebollo, Horard et al. 2010). For example, studies in humans have revealed that exposure to certain air pollutants, such as benzene or particulate matter with aerodynamic diameter less than 10  $\mu\text{m}$ , is

associated with reduced methylation of repetitive elements and gene promoters, which may be associated with resulting disease states including cardiorespiratory disease and lung cancer (Samet, Dominici et al. 2000; Snyder 2002; Brook, Franklin et al. 2004; Peters 2005; Vineis and Husgafvel-Pursiainen 2005; Bollati, Baccarelli et al. 2007; Baccarelli, Martinelli et al. 2008; Tarantini, Bonzini et al. 2009).

### **Base Excision Repair**

Because DNA is a dynamic, reactive molecule, it participates in chemical reactions with endogenous and exogenous reagents. When the DNA molecule is altered, the information encoded is changed and may lead to modification to nucleobases that can cause disease or may disrupt DNA replication causing cell death. Modifications to a nucleobase may disrupt base pairing and could cause the misincorporation of the wrong nucleotide. Hence, these modifications are called DNA damage. The base excision repair (BER) pathway is one of several pathways to repair damage to DNA. BER is predominantly used to repair small lesions to DNA bases, caused by reactive oxygen species (ROS), alkylating agents, or deamination events that produce aberrant nucleobases (See Chapter II). BER functions through either short- or long-patch pathways, in which a single nucleotide or at least two nucleotides, respectively, are removed. This pathway and the enzymes involved are conserved from bacteria to mammals (Lindahl 1993; Friedberg, Aguilera et al. 2006).

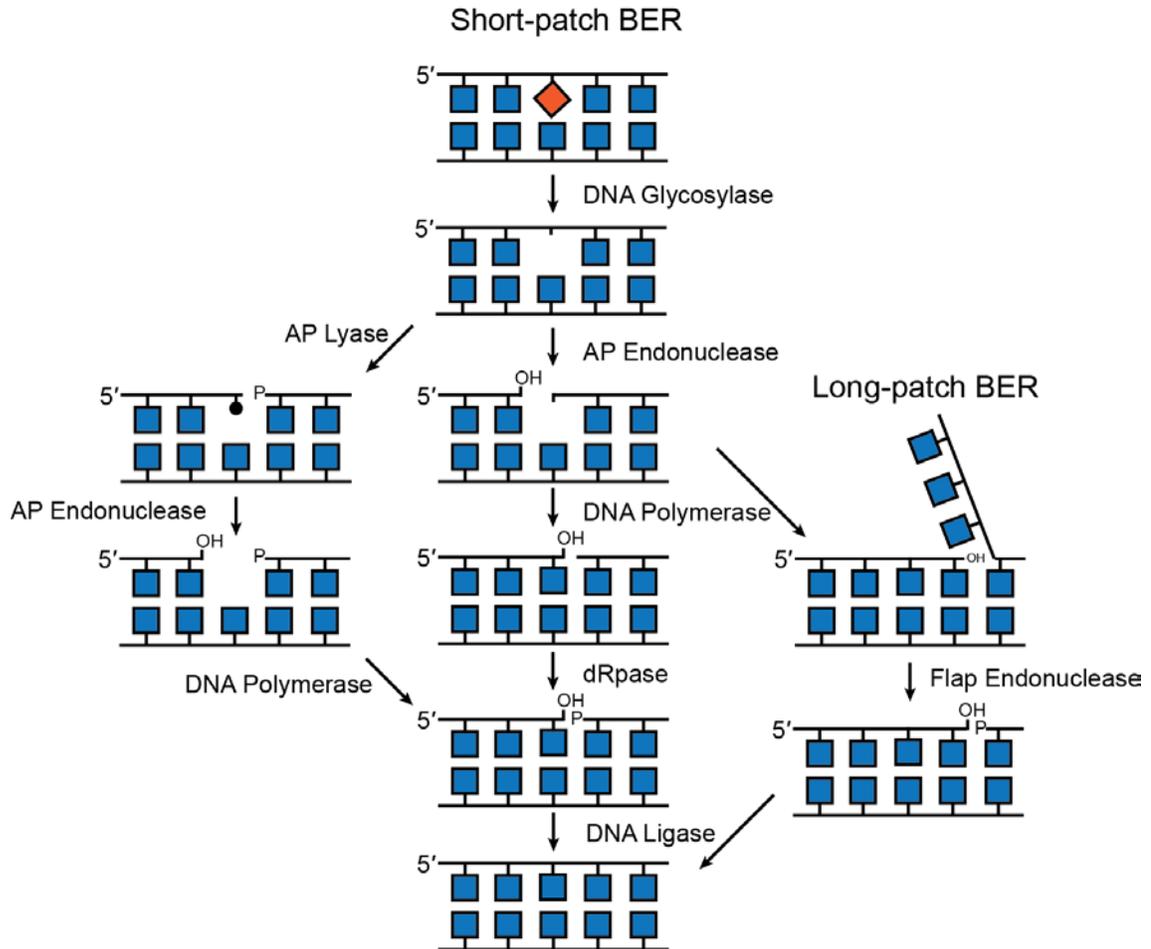
The simple short-patch BER pathway utilizes five enzymes: a DNA glycosylase, an AP endonuclease or AP DNA lyase, a DNA polymerase, a dRpase, and a DNA ligase (**Figure 2**). The DNA glycosylase recognizes the damaged base and cleaves the *N*-

glycosidic bond, releasing the base and creating an apurinic or apyrimidinic (AP) site. Next, the DNA backbone is cleaved through the use of an AP endonuclease or AP lyase. Some glycosylases contain lyase activity. The AP endonuclease cleaves the backbone, resulting in a nick 5' to the AP site, while AP lyase activity creates a nick 3' to the AP site. The AP endonuclease processes the nick to create a single nucleotide gap, with 3' hydroxyl and a 5' phosphate, creating a substrate for a DNA polymerase. The polymerase fills the gap with the correct nucleotide, and a DNA ligase seals the nick in the DNA to complete the repair process (Lindahl 2000).

The enzymes responsible for BER in humans include a number of glycosylases specific for various types of lesions (reviewed in Stivers and Jiang 2003). APEX nuclease 1 (APEX1), DNA polymerase  $\beta$  (POL $\beta$ ), and DNA ligase III (LIG3) have been identified as components of short-patch BER (Laval 1977; Matsumoto and Kim 1995; Kubota, Nash et al. 1996; Sobol, Horton et al. 1996). Long-patch BER requires the use of either POL $\beta$  or DNA polymerase  $\delta$  (POL $\delta$ ), PCNA, and flap structure-specific endonuclease 1 (FEN1) to remove the flap of DNA that is displaced during POL $\beta$  synthesis of a new strand of DNA (Klungland and Lindahl 1997; Kim, Biade et al. 1998).

Studies of BER enzymes in plants are less numerous than those in animals. BER enzymes in plants again include various DNA glycosylases specific for different base modifications (Krishnamurthy, Zhao et al. 2008; Kathe, Barrantes-Reynolds et al. 2009; Duclos, Aller et al. 2012). DNA LIGASE I (LIGI) may participate in the BER pathway in plants, but the DNA polymerase necessary for repair in plants is unknown (Andreuzza, Li et al. 2010). Plants may use either short- or long-patch BER to fill gaps generated by AP endonucleases and AP lyases (Cordoba-Canero, Morales-Ruiz et al. 2009). A

detailed discussion on DNA glycosylase structures and mechanisms can be found in Chapter II.



**Figure 2.** Base Excision Repair (BER) Pathway. A repair pathway for detrimental nucleobase modifications is initiated by damage-specific DNA glycosylases, which are either mono-functional (center pathway) or bi-functional (left pathway). An AP endonuclease processes the abasic site product, and a DNA polymerase functions to synthesize DNA. The DNA backbone is restored by a DNA ligase. In long-patch BER (right pathway), a 5' flap of DNA created by strand displacement synthesis is cleaved by a FLAP endonuclease before the final step is performed by a DNA ligase. [Adapted from (Scharer and Jiricny 2001)].

### **Passive DNA Demethylation**

As discussed above, the DNA methyltransferases DNMT3a and 3b are responsible for *de novo* methylation of CpG dinucleotides in vertebrates. DNMT1 functions to maintain methylation at newly synthesized daughter strands during semiconservative DNA replication. The interaction of DNMT1 and PCNA targets DNMT1 to sites of demethylation. If the activity of DNMT1 is absent, due to inhibition, depletion, or nuclear exhaustion of DNMT1, passive demethylation will occur. Thus, up to 50% of DNA methylation will be lost in each round of DNA replication. This process does not occur in terminally differentiated cells (Niehrs 2009).

### **Active DNA Demethylation**

The enzymatic removal of DNA methylation is critical for developmental processes in plants and animals (Zhang and Zhu 2012). Two genome-wide demethylation events occur in developing humans. First, the paternal pronuclei in the zygote undergoes demethylation, while the maternal genome maintains its methylation patterns. However, some portions of the paternal genome, including centromeric and pericentromeric heterochromatin, some transposons, and imprinted regions, are not demethylated (Olek and Walter 1997; Rougier, Bourc'his et al. 1998; Mayer, Niveleau et al. 2000; Oswald, Engemann et al. 2000; Santos, Hendrich et al. 2002; Lane, Dean et al. 2003). This first demethylation event is followed by passive DNA demethylation in the maternal genome as a result of replication. Then, *de novo* DNA methyltransferases reestablish genomic DNA methylation (Wu and Zhang 2010). The second demethylation event follows in the primordial germ cells (PGCs) and functions to erase much of the

newly created DNA methylation (Morgan, Santos et al. 2005; Hajkova, Ancelin et al. 2008). These two global demethylation events, combined with subsequent chromatin rearrangement, allow for totipotency (Hajkova, Ancelin et al. 2008). Additional evidence for active DNA demethylation was observed in somatic differentiation of the developing embryo, allowing for the establishment of tissue-specific expression patterns (Kress, Thomassin et al. 2006; Niehrs 2009). Furthermore, active DNA demethylation has been detected during gene activation in adult kidney and brain (Kim, Kondo et al. 2009; Ma, Jang et al. 2009).

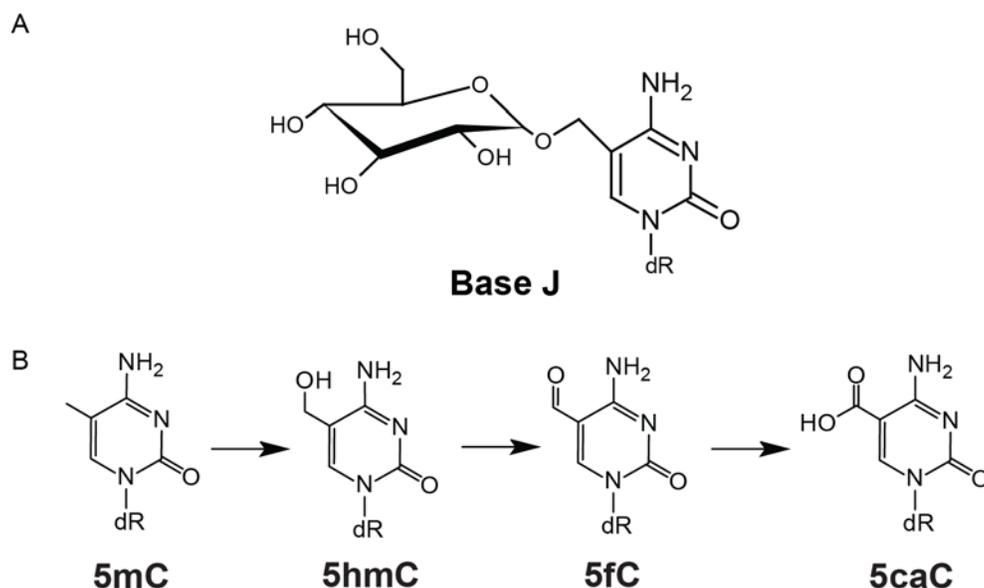
### **5mC Oxidation Derivatives**

The existence and maintenance of DNA methylation has been known for decades, as described above. However, only recently has the mechanism of active DNA demethylation begun to be understood. The proteins responsible for recognition of and enzymatic removal of 5mC have been identified in plants over the last 15 years (see *Active DNA Demethylation in Arabidopsis* below). However, the mechanism of active DNA demethylation in animals is not as clearly defined. In animals, there is mounting evidence for methylation removal by the coupling of DNA repair pathways and derivatives of 5mC. Although thymine DNA glycosylase (TDG) and methyl binding domain 4 (MBD4) are able to remove 5mC, their activities for this substrate are much lower than for T/G mismatches (Kress, Thomassin et al. 2001). Thus, it is likely the activity for 5mC is a minor side reaction and not physiologically relevant. To date, there is no known mammalian glycosylase with 5mC as the preferred substrate. However,

many reports now indicate that DNA glycosylases may participate in DNA demethylation by excising derivatives of 5mC.

The oxidized derivatives of 5mC were first discovered by searching the human genome for proteins similar to trypanosome enzymes JBP1 and JBP2 (Tahiliani, Koh et al. 2009). The enzymes are predicted to catalyze the first step of the biosynthesis of base J (beta-D-glycosyl hydroxymethyluracil, **Figure 3A**) (Yu, Genest et al. 2007; Cliffe, Kieft et al. 2009). Base J is a hypermodified thymine that is present in kinetoplastids, *Diplonema* (van Leeuwen, Taylor et al. 1998), and the green algae *Euglena* (Dooijes, Chaves et al. 2000), but has not been found in other protozoa, fungi, or vertebrates (van Leeuwen, Taylor et al. 1998). Because the synthesis of base J proceeds through the intermediate hydroxymethyldeoxyuridine (HOME-U), a base that most organisms remove using DNA glycosylases, it is unlikely that base J exists in other organisms. Like 5mC, Base J has been associated with gene silencing. The modification exists at expression sites for the variant surface glycoprotein (VSG) gene when it is not expressed; base J is absent at these sites when VSG is expressed (van Leeuwen, Wijsman et al. 1997). Base J also has been found at telomeric repeats in kinetoplastid flagellates, with at least 13% of thymines containing this modification at telomeres, compared to 0.8% in total DNA (van Leeuwen, de Kort et al. 1998). When base J is reduced at transcription start sites, nucleosome abundance is decreased, histone acetylation is increased, and Pol II is recruited to promoter regions (Ekanayake and Sabatini 2011; Ekanayake, Minning et al. 2011). These changes are associated with increased transcription initiation (Ekanayake and Sabatini 2011).

Base J is synthesized through two steps: first, a thymidine hydroxylase oxidizes the 5-methyl group of a thymidine residue to create HOMe-U; second, a glucosyl transferase adds a glucose group to the hydroxymethyl modification (Gommers-Ampt and Borst 1995; van Leeuwen, Kieft et al. 1998; Ulbert, Cross et al. 2002; Borst and Sabatini 2008). J binding protein 1 and 2, JBP1 and JBP2, are members of the 2-oxoglutarate- and Fe(II)-dependent [2OG-Fe(II)] oxygenase superfamily, and they have been implicated *in vitro* and *in vivo* as the thymidine hydroxylases responsible for the first step in base J synthesis (Cross, Kieft et al. 1999; Hausinger 2004; DiPaolo, Kieft et al. 2005; Yu, Genest et al. 2007; Cliffe, Kieft et al. 2009; Cliffe, Hirsch et al. 2012).



**Figure 3.** 5mC derivatives. (A) Base J is a cytosine modification present in trypanosomes. (B) Tet enzymes successively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC).

Homologs of the oxygenase domains of JBP1 and JBP2 were found by iterative sequence profile searches. Three paralogous human proteins called ten eleven

translocation proteins, Tet1, Tet2, and Tet3 were identified as 2OG-Fe(II) oxygenases, in addition to homologous domains in metazo, fungi, and algae (Tahiliani, Koh et al. 2009). In the same study, it was determined that TET1 produced 5-hydroxymethylcytosine (5hmC) in mouse embryonic stem cells (Tahiliani, Koh et al. 2009) (**Figure 3B**). A subsequent study confirmed that mouse Tet1 and Tet2 reduce 5mC levels and result in the generation of 5hmC *in vivo*. Although Tet3 expression did not reduce 5mC levels *in vivo*, 5hmC was generated (Ito, D'Alessio et al. 2010). Recombinant purified Tet1, Tet2, and Tet3 generated 5hmC *in vitro* as well, while catalytic mutants did not. Tet1 and Tet2 are expressed in mouse embryonic stem cells, but Tet3 is not. The knockdown of Tet1 resulted in abnormal mouse morphology and reduced cell growth rate, but the knockdown of Tet2 and Tet3 had no effect (Ito, D'Alessio et al. 2010). The Tet1 knockdown resulted in reduced expression of key stem cell factors, such as Nanog, Oct4, and Sox2. Therefore, Tet1 likely functions in embryonic stem cell self-renewal and maintenance (Ito, D'Alessio et al. 2010). Furthermore, chromatin immunoprecipitation (ChIP) and methyl-DNA immunoprecipitation (MeDIP) experiments indicated Tet1 functions to maintain the *Nanog* promoter in a hypomethylated state. Combined with knockdown experiments in cells lacking DNA methyltransferases, which showed Tet1 control of *Nanog* is dependent on DNA methylation, these results indicate that Tet1 regulates Nanog expression (Ito, D'Alessio et al. 2010).

The extent to which 5mC is modified in animals was found to include two additional oxidized derivatives – 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) – in 2011 (Ito, Shen et al. 2011) (**Figure 3B**). Through the use of a more sensitive assay, it was determined that Tet proteins generate these derivatives, in addition to 5hmC. In a

follow-up study to the 2010 report of Tet activity, it was found that the restriction enzyme used to digest the oligonucleotides (MspI) was not able to digest DNA containing 5fC and 5caC (Ito, Shen et al. 2011). Using a different enzyme (TaqI) to digest the DNA in conjunction with two-dimensional thin-layer chromatography (2D-TLC), the authors were able to detect 5fC and 5caC reaction products from Tet reactions with 5mC DNA. Furthermore, when the Tet proteins are overexpressed, the genomic content of these derivatives increases. The authors further quantified the activity of Tet2 by incubating the enzyme with 5mC-, 5hmC-, or 5fC-containing DNA and quantifying the products using LC-MS. Tet2 was able to convert over 95% of 5mC to 5hmC (60%), 5fC (30%), and 5caC (5%) in 10 minutes, but only converted 40% of 5hmC-DNA substrate and 25% of 5fC-DNA substrate (Ito, Shen et al. 2011).

The oxidized derivatives 5hmC, 5fC, and 5caC have all been detected in the genomic DNA of mouse embryonic stem cells (Penn, Suwalski et al. 1972; Ito, Shen et al. 2011; Pfaffeneder, Hackner et al. 2011). It was determined that about  $1.3 \times 10^3$  5hmC, 20 5fC, and 3 5caC are present in every  $10^6$  cytosine bases in mouse embryonic stem cells. 5fC was also detected in genomic DNA of major mouse organs, but 5caC was not (Ito, Shen et al. 2011). These derivatives have not been detected in plants (Jang, Shin et al. 2014). These derivatives each have unique steric and electronic properties, including alternative tautomers and differential *N*-glycosidic bond stabilities (Bennett, Rodgers et al. 2006; Williams and Wang 2012; Hashimoto, Zhang et al. 2013; Maiti, Michelson et al. 2013).

The role of TDG in DNA demethylation has long been hypothesized, due to its documented involvement in transcriptional regulation and demethylation. TDG interacts

with transcription factors, including retinoic acid receptor (RAR), retinoid X receptor (RXR), estrogen receptor  $\alpha$  (ER $\alpha$ ), thyroid transcription factor 1 (TTF1), and histone acetyl-transferases p300 and CBP (Um, Harbers et al. 1998; Missero, Pirro et al. 2001; Tini, Benecke et al. 2002; Chen, Lucey et al. 2003; Leger, Smet-Nocca et al. 2014). Furthermore, TDG is required for embryonic development (Cortazar, Kunz et al. 2011; Cortellino, Xu et al. 2011). Although it was proposed that TDG may be able to remove 5mC directly, additional studies have shown that activity for 5mC is very weak, while TDG is much more active for excision of mismatched T (Zhu, Zheng et al. 2000; Cortazar, Kunz et al. 2007; Cortellino, Xu et al. 2011). However, recent studies have illuminated the role of TDG in DNA demethylation: evidence for TDG excision of 5fC and 5caC, two of the oxidation derivatives of 5mC generated by Tet proteins, was found in 2011 (He, Li et al. 2011; Maiti and Drohat 2011). Further structural and biochemical studies indicate that TDG has a well-organized binding pocket for 5caC that excludes binding of 5mC and 5hmC (Zhang, Lu et al. 2012). Additional discussion on TDG structures and its substrate specificity can be found in Chapter II.

In addition to the oxidation of 5mC by Tet proteins, 5mC also may be removed by DNA glycosylases as a result of deamination of the base to thymine. Demethylation events have been linked to the genome stability protein GADD45 $\alpha$  (growth arrest and DNA damage-inducible protein 45  $\alpha$ ) in *Xenopus* (Barreto 2007). However, the role of GADD45 $\alpha$  in mammals has been disputed (Barreto 2007, Schmitz, 2009, Jin, 2008). In zebrafish embryos, it was found that 5mC demethylation to thymine may be catalyzed by activation-induced deaminase (AID) or apolipoprotein B RNA-editing catalytic component 2b and 2a (APOBEC2B, 2A) (Rai, Huggins et al. 2008). The removal of the

resulting mismatched thymine by the thymine DNA glycosylase MBD4 was suspected due to increased methylation as a result of MBD4 knockdown. Furthermore, the non-enzymatic GADD45a was found to promote interactions between AID and MBD4 (Rai, Huggins et al. 2008). However, some reports have challenged the role of GADD45a in DNA demethylation (Jin, Guo et al. 2008). In addition, there are no developmental effects due to AID/APOBEC deficiency (Nabel, Manning et al. 2012). Therefore, it is unlikely that the removal of 5mC after deamination and subsequent repair by MBD4 or TDG is the mechanism of active DNA demethylation in mammals.

### **Active DNA Demethylation in Plants**

One of the first indications that the base excision repair pathway is used to remove cytosine methylation was found in *Arabidopsis thaliana*. Control of gene transcription is necessary for the female gametophyte to develop properly into embryo and endosperm once fertilized. The endosperm is formed from fertilization of the central cell, which exists as a diploid nucleus resulting from the fusion of two haploid nuclei within the ovule. A separate fertilization of the adjacent egg cell forms the embryo, thereby forming a seed containing a triploid endosperm and diploid embryo. While the embryo develops organs and tissues, the endosperm functions to transfer nutrients from maternal tissues to the developing embryo, similar to the placenta in mammals (Brown, Lemmon et al. 1999). Some maternal alleles have been discovered to be necessary for proper seed development. In particular, the Polycomb group proteins, such as Medea (MEA), Fertilization Independent Endosperm (FIE), and Fertilization Independent Seed2 (FIS2), are SET-domain, WD domain, and zinc-finger containing proteins that function to

repress gene transcription (Grossniklaus, Vielle-Calzada et al. 1998; Kiyosue, Ohad et al. 1999; Luo, Bilodeau et al. 1999; Ohad, Yadegari et al. 1999; Birve, Sengupta et al. 2001). These proteins are necessary to control central cell proliferation and endosperm development. In particular, the *MEDEA* gene is imprinted, and only the maternal allele is expressed during endosperm development, while the paternal allele is silenced (Kiyosue, Ohad et al. 1999; Vinkenoog, Spielman et al. 2000).

In 2002, a mutation in the maternal *DEMETER* allele was found to cause abnormal seed development in *A. thaliana* (Choi, Gehring et al. 2002). The maternal *DME* allele is necessary for maternal *MEA* expression in the central cell and endosperm, and only the maternal *DME* allele is necessary for seed viability. The Fisher lab found that *DME* prevents developmental abnormalities, such as reduced or increased petal and sepal number that appeared in *dme-1* plants. Additional abnormalities in leaf and stem morphology also were observed. Therefore, *DME* is necessary for proper floral and vegetative development, in addition to seed viability in *A. thaliana* (Choi, Gehring et al. 2002). Using semiquantitative reverse transcriptase-polymerase chain reaction (RT-PCR), it was determined that *DME* is expressed in maternal reproductive structures, including immature flower buds and ovules, prior to fertilization events. The expression of *DME* during female gametophyte development was confirmed through the use of the reporter gene  $\beta$ -glucuronidase (GUS) and Green Fluorescent Protein (GFP) labeling, which were detected in the central cell until fertilization. It was further found that *DME* is necessary for *MEA* and *FWA* gene expression in the central cell and in the endosperm after fertilization, thereby controlling *A. thaliana* development (Choi, Gehring et al. 2002) (Kinoshita, Miura et al. 2004).

In the same report, it was determined that *DME* encodes a DNA glycosylase of 1729 amino acids, including a 200 residue domain that is related to the helix-hairpin-helix (HhH) structural superfamily of DNA glycosylases. (See Chapter II, *HhH Superfamily* for more details). The glycosylase domain of DME also contains four conserved cysteine residues that comprise a [4Fe-4S] cluster, which is found in other glycosylases as well (Choi, Gehring et al. 2002). The role of the iron sulfur cluster has not been determined, though some speculate it is necessary for DNA binding and/or lesion recognition (see Chapter IV for discussion). These results indicated that DME functions to excise 5mC (Choi, Gehring et al. 2002). DME contains an aspartic acid residue (Asp1304) that aligns with conserved HhH catalytic aspartate residues necessary for glycosylase function. When this residue is mutated to asparagine (Asp1304Asn), DME is no longer able to rescue seed abortion due to mutation in the endogenous *DME* gene (Choi, Harada et al. 2004). The Fisher lab demonstrated that DME causes hypomethylation in the maternal endosperm compared to the paternal endosperm, while the *dme* mutant does not show this difference (Gehring, Huh et al. 2006). Although initially thought to be a monofunctional glycosylase, it was later determined that DME utilizes a bifunctional mechanism to excise 5mC (Gehring, Huh et al. 2006). Repressor of Silencing 1 (ROS1), along with two DME-like paralogs, DML2 and DML3, are also expressed in *A. thaliana*. ROS1, DML2, and DML3 function as genome-wide demethylases to counteract DNA methylation established by RdDM (Gong, Morales-Ruiz et al. 2002; He, Hsu et al. 2009; Hsieh, Ibarra et al. 2009; Zhu 2009; Gao, Liu et al. 2010). Additional discussion on the biochemical mechanisms of DME/ROS1/DML family continues in Chapter II, and an in depth analysis of predicted DME structure,

essential residues, and excision of 5mC and its oxidation derivatives can be found in Chapter III.

Recent reports of DME activities in many plant systems indicate that the control of DNA methylation can be utilized agriculturally to benefit crop production. For example, a study of a DME homolog from barley (HvDME) indicates that DME is upregulated during drought and confirms DME function in endosperm and seed development (Kapazoglou, Drosou et al. 2013). Results from studies of important cereal crops can be extended to develop therapies for patients suffering from celiac disease (Osorio, Wen et al. 2012). The expression levels of gliadins and low-molecular-weight glutenins (LMWGs), two proteins that are the source of immunogenic epitopes in celiac disease, are maintained by the methylation status of their promoters (Qi, Wei et al. 2009; Mitea, Salentijn et al. 2010). Wen, et al. determined that *DME* can be suppressed using RNA interference, which prevents DME from functioning to activate expression of gliadins and glutenins, thereby reducing the amount of these proteins in the endosperm of wheat (Wen, Wen et al. 2012).

### **Scope of this Work**

The central topic of this work is to determine the role of a DNA glycosylase, DEMETER, in the regulation of gene transcription via excision of 5mC. This connection between DNA repair and gene regulation is unique, since the usual role of DNA glycosylases is to recognize and initiate the repair of damage. The study of DNA glycosylases is intriguing given their ability to locate a single damaged base among a large excess of unmodified DNA. Chapter II begins a discussion on the structural

mechanisms of DNA glycosylases, as the 3-dimensional structures of these enzymes have yielded significant insight into substrate specificity and function within the BER pathway. The biochemistry and predicted structure of DME is investigated in Chapter III. The ability of DME and its paralog DML3 to remove 5mC and its oxidation derivatives yields valuable information on the active site of these enzymes. Finally, in Chapter IV, the implications of this work are discussed, in addition to plans for future experiments.

## Chapter II

### STRUCTURAL MECHANISMS OF DNA GLYCOSYLASES\*

#### Introduction

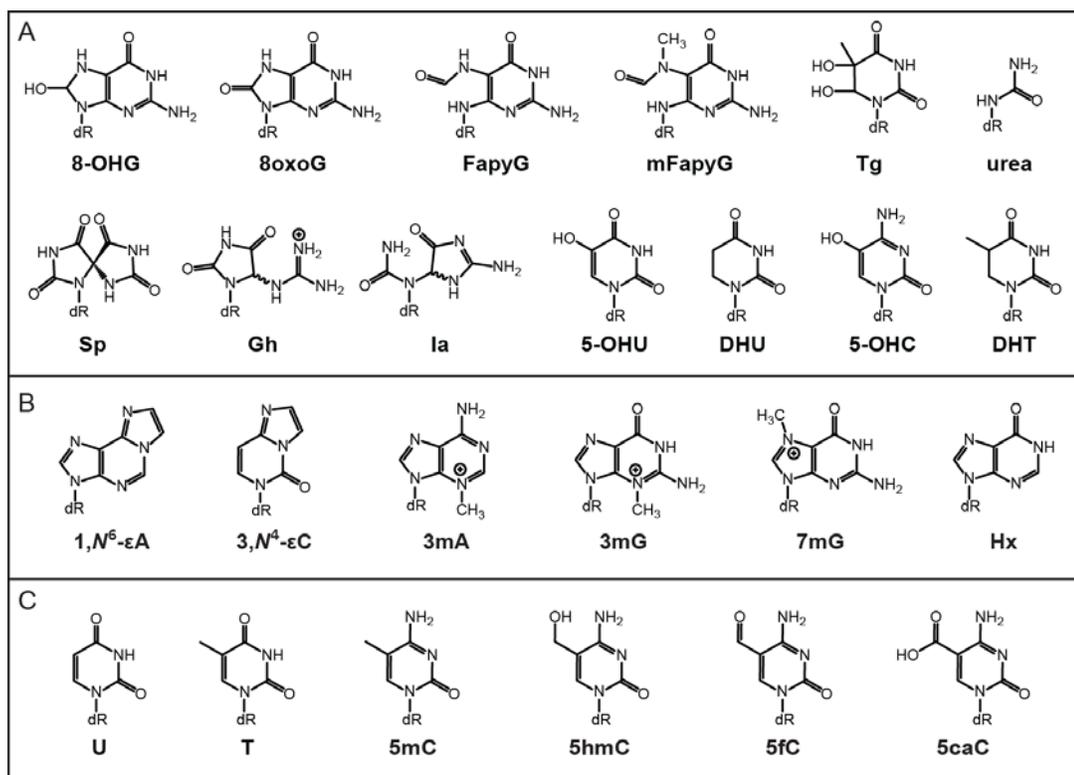
The integrity of the chemical structure of DNA and its interactions with replication and transcription machinery is important for the faithful transmission and interpretation of genetic information. Oxidation, alkylation, and deamination of the nucleobases by a number of endogenous and exogenous agents create aberrant nucleobases (**Figure 4**) that alter normal cell progression, cause mutations and genomic instability, and can lead to a number of diseases including cancer (reviewed in Friedberg, Aguilera et al. 2006). Many of these lesions are removed by the base excision repair (BER) pathway (Lindahl 2000), which is initiated by a DNA glycosylase specialized for a particular type of chemical damage. Upon locating its particular lesion within the DNA, glycosylases catalyze the excision of the nucleobase from the phosphoribose backbone by cleaving the *N*-glycosidic bond, generating an apurinic/apyrimidinic (AP) site (**Figure 5**). Monofunctional glycosylases catalyze only base excision, whereas bifunctional glycosylases also contain a lyase activity that cleaves the backbone immediately 3' to the

---

\*Part of the work presented in this chapter was published in Brooks, S. B., Adhikary, S., Rubinson, E. H., and Eichman, B.F. (2013) Recent advances in structural mechanisms of DNA glycosylases. *Biochimica et Biophysica Acta* **1834**, 247–271.

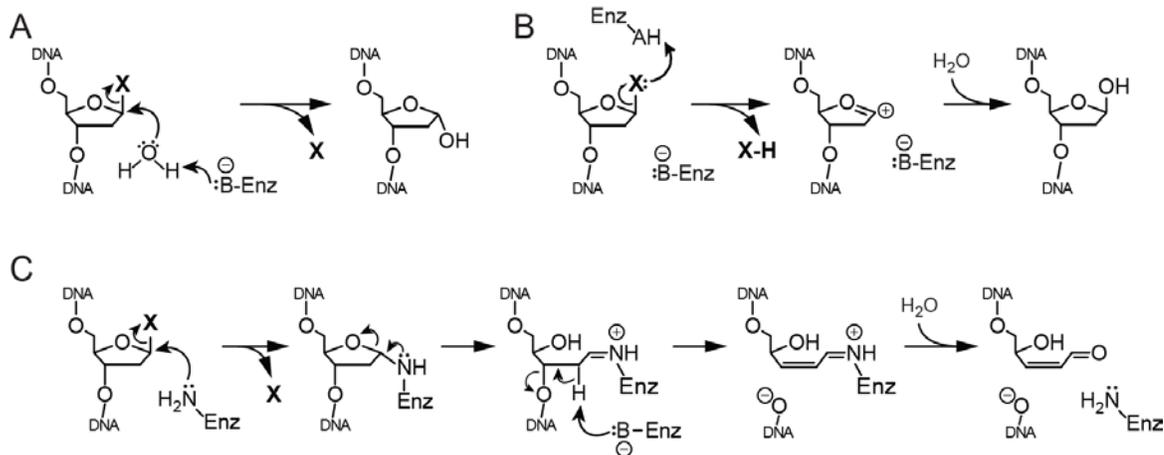
AP site. The resulting single-stranded and nicked AP sites are processed by AP endonuclease 1 (APE1), which hydrolyzes the phosphodiester bond 5' to the AP site. This generates a 3' hydroxyl substrate for replacement synthesis by DNA polymerase  $\beta$ , followed by sealing of the resulting nick by DNA ligase.

Since the glycosylases are the first line of defense against a vast array of DNA damage, they have been the subject of a large body of work to understand their



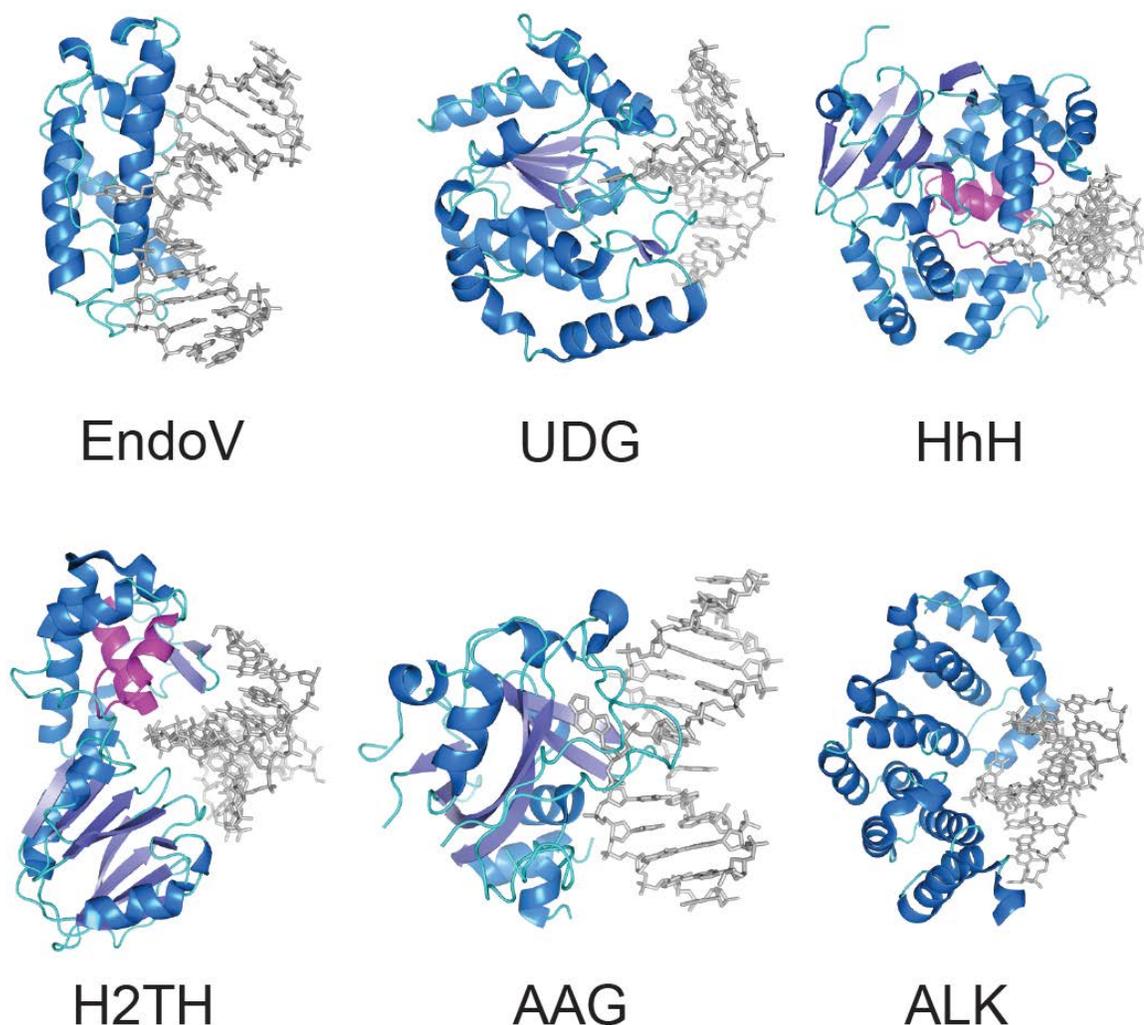
**Figure 4.** Common DNA lesions. (A) Oxidized nucleobases. 8-OHG, 7,8-dihydro-8-hydroxyguanine; 8-OxoG, 7,8-dihydro-8-oxoguanine; FapyG, 2,6-diamino-4-hydroxy-5-formamidopyrimidine; mFapyG, N7-methylFapyG; Tg, thymine glycol; Sp, spiroiminodihydantoin; Gh, guanidinohydantoin; Ia, iminoallantoin; 5-OHU, 5-hydroxyuracil; DHU, dihydrouracil; 5-OHC, 5-hydroxycytosine; DHT, dihydrothymine. (B) Alkylated nucleobases. εA, 1,N<sup>6</sup>-ethenoadenine; εC, 3,N<sup>4</sup>-ethenocytosine; 3mA, N3-methyladenine; 3mG, N3-methylguanine; 7mG, N7-methylguanine; Hx, hypoxanthine. (C) Nucleobases repaired by the UDG/TDG family of DNA glycosylases. U, uracil; T, thymine; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine.

mechanism of action (Scharer and Jiricny 2001; Stivers and Jiang 2003; Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; David, O'Shea et al. 2007; Dalhus, Laerdahl et al. 2009; Friedman and Stivers 2010; Li 2010; Zharkov, Mechetin et al. 2010; Rubinson and Eichman 2012). The first crystal structures of DNA glycosylases were reported in 1992 for bacteriophage T4 Endonuclease V (EndoV) and *Escherichia coli* (*E. coli*) Endonuclease III (EndoIII), which remove pyrimidine dimers and oxidized pyrimidines, respectively (Kuo, McRee et al. 1992; Morikawa, Matsumoto et al. 1992). Soon thereafter, DNA or inhibitor-bound structures of EndoV and uracil DNA glycosylase (UDG) established that these enzymes use a base-flipping mechanism to gain access to modified nucleobases in DNA (Mol, Kuo et al. 1995; Savva, McAuley-Hecht et al. 1995; Vassylyev, Kashiwagi et al. 1995; Slupphaug, Mol et al. 1996).



**Figure 5.** Chemical reaction catalyzed by DNA glycosylases. (A,B) Monofunctional glycosylases cleave the *N*-glycosidic bond to liberate free nucleobase (N) from the phosphoribose backbone through either associative (A) or dissociative (B) mechanisms. (C) Bifunctional mechanism, in which both the *N*-glycosidic bond and the DNA backbone are cleaved.

Subsequent studies established that glycosylases fall into one of six structural superfamilies (**Figure 6**). Despite their divergent architectures, these proteins, with the exception of the ALK family (Rubinson and Eichman 2012), have evolved the base flipping strategy to correctly identify and orient their substrates for catalysis. Recognition of the target modification likely proceeds in several steps, in which the protein probes the stability of the base pairs through processive interrogation of the DNA duplex, followed by extrusion of the aberrant nucleobase into a specific active site pocket on the enzyme (Stivers and Jiang 2003; Banerjee, Santos et al. 2006). The enzyme-substrate complex is stabilized by nucleobase contacts within the active site and a pair of side chains that plug the gap in the DNA left by the extrahelical nucleotide and wedge into the DNA base stack on the opposite strand (Scharer and Jiricny 2001; Stivers and Jiang 2003; Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; David, O'Shea et al. 2007; Dalhus, Laerdahl et al. 2009; Friedman and Stivers 2010; Li 2010; Zharkov, Mechetin et al. 2010; Rubinson and Eichman 2012).



**Figure 6.** DNA glycosylase structural superfamilies. Representative crystal structures from each class shown are: EndoV, T4 pyrimidine dimer DNA glycosylase EndoV (PDB ID 1VAS); UDG, human uracil-DNA glycosylase UDG (1EMH); Helix-hairpin-Helix (HhH), human 8-oxoguanine DNA glycosylase OGG1 (1YQK); Helix-two turn-helix (H2TH), *Bacillus stearothermophilus* 8-oxoguanine DNA glycosylase MutM (1L1T); AAG, human alkyladenine DNA glycosylase AAG/MPG (1EWN); ALK, *Bacillus cereus* alkylpurine DNA glycosylase AlkD (3JXZ). Proteins are colored according to secondary structure with the HhH and H2TH domains magenta. DNA is shown as gray sticks.

### Oxidative Damage

DNA bases undergo oxidative damage from chemical oxidants, free radicals and reactive oxygen species (ROS) produced from cellular respiration, inflammatory responses, and ionizing radiation (Klaunig and Kamendulis 2004; Valko, Rhodes et al.

2006; van Loon, Markkanen et al. 2010). Oxidized bases are often used as biomarkers for oxidative stress and cancer (Klaunig and Kamendulis 2004; Kryston, Georgiev et al. 2011). Guanines are especially susceptible to oxidation, leading to a number of lesions that are substrates for BER (**Figure 4A**) (Neeley and Essigmann 2006). Attack of a hydroxyl radical at the C8 position of guanine produces a guanyl radical, which can be reduced to 7,8-dihydro-8-hydroxyguanine (8-OHG). Ring-opening of 8-OHG results in 2,6-diamino-5-formamido-4-hydroxy-pyrimidine (FapyG). Alternatively, the guanyl radical may be oxidized to 8-oxo-7,8-dihydroguanine (8-OxoG). 8-OxoG and FapyG are two of the most abundant oxidative DNA adducts (Burrows and Muller 1998; Evans, Dizdaroglu et al. 2004). 8-OxoG is a particularly insidious lesion because of its dual coding potential by replicative polymerases, leading to G→T transversion mutations likely as a result of its ability to form both 8-OxoG(syn)•A(anti) and 8-OxoG(anti)•C(anti) base pairs (Cheng, Cahill et al. 1992; Brieba, Eichman et al. 2004; Hsu, Ober et al. 2004; Klaunig and Kamendulis 2004; van Loon, Markkanen et al. 2010). Oxidation of guanine and 8-OxoG also produces a variety of ring-opened purines in addition to FapyG, including hydantoin lesions, spiroiminodihydantoin (Sp), guanidinohydantoin (Gh), and its isomer iminoallantoin (Ia) (**Figure 4A**) (Luo, Muller et al. 2001; Burrows, Muller et al. 2002; Henderson, Delaney et al. 2003). Fapy lesions inhibit DNA polymerases and are potentially mutagenic (Tudek 2003). Hydantoin lesions have been suggested to lead to an increase in G→T and G→C transversions and stall the replication machinery (Leipold, Muller et al. 2000; Burrows, Muller et al. 2002; Henderson, Delaney et al. 2003; Delaney, Neeley et al. 2007). In addition to purines, reaction of hydroxyl radicals at positions 5 or 6 of thymine produces 5,6-dihydroxy-5,6-

dihydrothymine (thymine glycol, Tg), a cytotoxic lesion that distorts the DNA duplex and can inhibit replication (Evans, Dizdaroglu et al. 2004; Aller, Rould et al. 2007). Other potentially harmful pyrimidines include dihydrothymine (DHT), dihydrouracil (DHU), 5-hydroxyuracil (5-OHU), 5-hydroxycytosine (5-OHC), 5-hydroxymethyluracil (5hmU), and 5-formyluracil (5fU) (Boorstein and Teebor 1988; Purmal, Kow et al. 1994; Yoshida, Makino et al. 1997; Kreutzer and Essigmann 1998; Liu and Doetsch 1998; Shikazono, Pearson et al. 2006).

### Alkylation Damage

A diverse array of alkylated DNA adducts are produced by environmental mutagens, cellular metabolites, and chemotherapeutic agents (**Figure 4B**) (Lawley and Brookes 1963; Sedgwick 2004; Friedberg, Aguilera et al. 2006; Gates 2009). The major and minor groove-exposed N7 and N3 positions of purines make them susceptible to reaction with electrophiles, with guanine N7 being the most nucleophilic (Gates, Nooner et al. 2004). Whereas *N*7-methylguanine (7mG) is relatively innocuous compared to larger *N*7-alkyl substituents, the positive charge generated from *N*7-substitution destabilizes the base and leads to spontaneous depurination and ring decomposition to produce, for example, 5-*N*-Methyl-2,6-diamino-4-hydroxyformamidopyrimidine (mFapyG). The glycosidic linkage of *N*3-methyladenine (3mA) is especially unstable, with a half-life for 3mA depurination as short as 24 h at 37°C (Margison and O'Connor 1973). Reactive aldehydes and epoxides generated from lipid peroxidation produce a number of ethenoadducts with A, G, and C, including 1,*N*<sup>6</sup>-ethenoadenine ( $\epsilon$ A), 1,*N*<sup>2</sup>- and *N*<sup>2</sup>,3-ethenoguanine, and 3,*N*<sup>4</sup>-ethenocytosine ( $\epsilon$ C) (Nair, Barbin et al. 1999; Barbin 2000;

Friedberg, Aguilera et al. 2006). In general, these lesions cause genomic instability through mutations and strand breaks (Friedberg, Aguilera et al. 2006). 3mA is cytotoxic, likely as a result of inhibition of DNA synthesis by disruption of DNA polymerase contacts to the N3 position in the minor groove (Boiteux, Huisman et al. 1984; Larson, Sahm et al. 1985; Fronza and Gold 2004; Plosky, Frank et al. 2008).

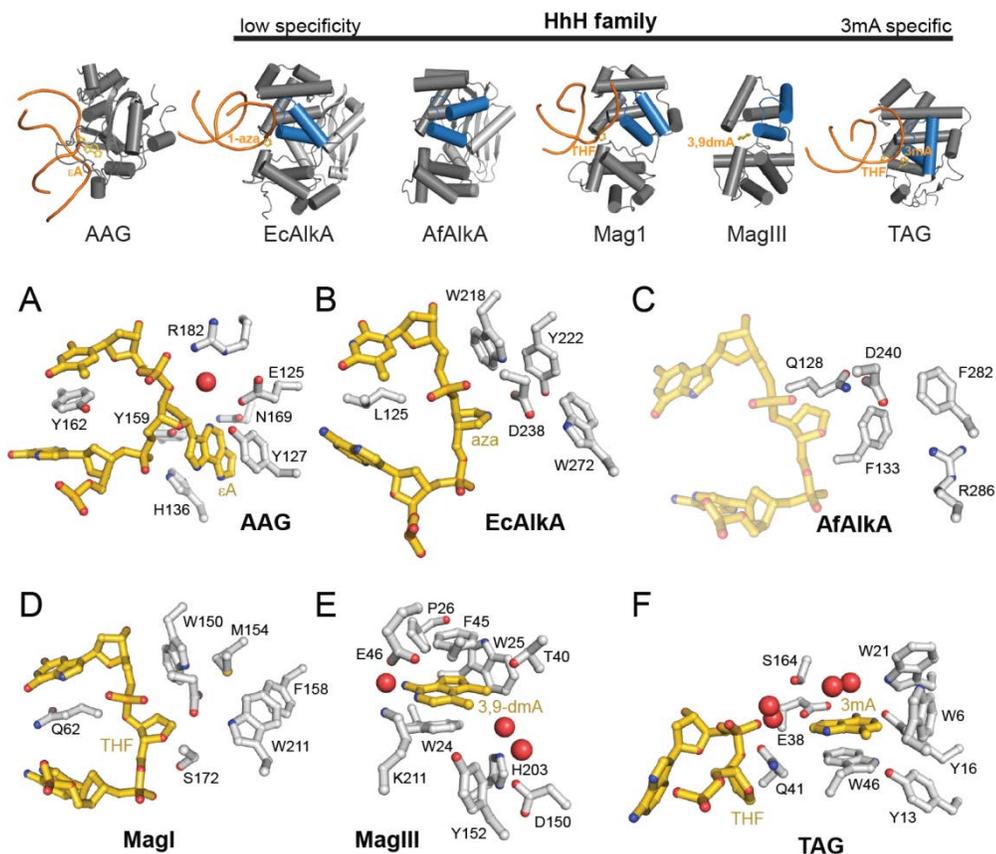
DNA glycosylases specific for alkylation damage have been characterized from eukaryotes, archaea, and bacteria. These include human AAG/MPG/ANPG (Brent 1979; Karran and Lindahl 1980), *Saccharomyces cerevisiae* MAG and *Schizosaccharomyces pombe* Mag1 (Berdal, Bjørås et al. 1990; Chen, Derfler et al. 1990; Memisoglu and Samson 1996), *E. coli* 3mA DNA glycosylase I (TAG) and II (AlkA) (Riazuddin and Lindahl 1978; Thomas, Yang et al. 1982), *Archaeoglobus fulgidus* AlkA (AfAlkA) (Mansfield, Kerins et al. 2003; Leiros, Nabong et al. 2007), *Deinococcus radiodurans* AlkA (DrAlkA) (Moe, Hall et al. 2012), *Thermotoga maritima* MpgII (Begley, Haas et al. 1999), *Helicobacter pylori* MagIII (O'Rourke, Chevalier et al. 2000), and *Bacillus cereus* AlkC and AlkD (Alseth, Rognes et al. 2006). AAG, AlkA, and MAG/Mag1 excise a broad range of alkylated and deaminated bases (McCarthy, Karran et al. 1984; Bjelland, Birkeland et al. 1994; Saparbaev, Kleibl et al. 1995; Birkeland, Anensen et al. 2002; Mansfield, Kerins et al. 2003; Alseth, Osman et al. 2005; Lingaraju, Kartalou et al. 2008; Adhikary and Eichman 2011), TAG is specific for N3-methylpurines 3mA and 3mG (Bjelland, Bjørås et al. 1993), and MagIII, MpgII and AlkC/D have an intermediate selectivity for positively charged lesions 3mA and 7mG (Begley, Haas et al. 1999; O'Rourke, Chevalier et al. 2000; Alseth, Rognes et al. 2006). Interestingly, AfAlkA has robust activity toward N1-methyladenine (1mA) and N3-methylcytosine (3mC), which

are normally repaired by oxidative demethylation (Birkeland, Anensen et al. 2002; Mansfield, Kerins et al. 2003; Leiros, Nabong et al. 2007). The alkylpurine DNA glycosylases can be grouped into three structural classes: 1) AAG, defined by the human enzyme, 2) ALK, including AlkC and AlkD, and 3) HhH, comprising all others (**Figure 6**). AAG and HhH subfamilies have a collection of aromatic, electron-rich active sites specialized to accommodate electron-poor alkylpurines despite the divergence in architectures (**Figure 7**) (Labahn, Scharer et al. 1996; Lau, Scharer et al. 1998; Hollis, Ichikawa et al. 2000; Rubinson, Gowda et al. 2010). The details of the alkylpurine glycosylase structures are the subject of a recent review (Rubinson, Adhikary et al. 2010).

### *AAG*

Human AAG, also known as MPG and ANPG, excises a variety of alkylated purines, including 3mA, 7mG, and  $\epsilon$ A, as well as hypoxanthine (Hx), the oxidative deamination product of adenine (**Figure 4B**) (McCullough, Dodson et al. 1999; Wyatt, Allan et al. 1999). The exceptional rate enhancement of Hx excision relative to alkylated substrates suggests that Hx is the predominant biological substrate (O'Brien and Ellenberger 2004). AAG has also been shown to excise 1-methylguanine and 1, $N^2$ - $\epsilon$ G (Saparbaev, Langouet et al. 2002; Lee, Delaney et al. 2009). Crystal structures of a catalytic fragment of AAG bound to oligonucleotides containing a pyrrolidine transition-state analog and an  $\epsilon$ A base showed that AAG is a single domain protein with a mixed  $\alpha/\beta$  structure and a positively charged DNA binding surface unique to DNA glycosylases (Lau, Scharer et al. 1998; Lau, Wyatt et al. 2000). Tyr162 on the tip of a  $\beta$ -hairpin plugs the gap in the DNA left by the flipped nucleotide, and the flipped  $\epsilon$ A base is stacked between two tyrosine residues (Tyr127 and Tyr159) and His136 inside the active site

cavity (**Figure 7A**). These structures provided a framework for a number of recent kinetic and thermodynamic studies aimed at dissecting the mechanism and substrate specificity of AAG and its collaboration with other BER enzymes.



**Figure 7.** Alkylpurine DNA glycosylases. Overall architectures are shown on the top row, with HhH enzymes arranged in order of increasing specificity for 3mA. (**A-F**) Active sites. Protein and nucleic acid atoms are shaded grey and gold, respectively, and waters are shown as red spheres. (**A**) Human AAG/ $\epsilon$ A-DNA substrate complex (1EWN). (**B**) *E. coli* AlkA bound to 1-azaribose-DNA (1DIZ). (**C**) *A. fulgidus* AlkA (2JHJ) with THF-DNA modeled from the *S. pombe* MagI/DNA complex (3S6I). (**D**) *S. pombe* MagI/THF-DNA (3S6I). (**E**) *H. pylori* MagIII/3,9-dimethyladenine (1PU7). (**F**) *E. coli* TAG/THF-DNA/3mA product complex (2OFI).

### *1. Mechanism of base flipping and substrate discrimination*

A series of careful biochemical examinations of substrate binding, flipping, and excision by AAG has recently been reported. As is typically true for other glycosylases, substrates that decrease the stability of the DNA increase the efficiency of excision by AAG, with bulged nucleotides excised more efficiently than mismatched base pairs (O'Brien and Ellenberger 2004; Lyons and O'Brien 2009). Interestingly, the strength of AAG binding to bulges correlates with increased spontaneous frameshift mutations upon overexpression of the enzyme, which may be a result of AAG shielding bulged bases from mismatch repair (Klapacz, Lingaraju et al. 2010). Discrimination of nucleobases on the basis of their stability within the DNA duplex can be rationalized by the barrier to base flipping. Kinetic analysis using intrinsic  $\epsilon$ A fluorescence revealed that  $\epsilon$ A flipping by AAG is highly favorable, which helps to explain discrimination of this lesion from undamaged bases (Wolfe and O'Brien 2009). These experiments also generated a two-step binding regime in which distinct DNA-bound and base flipped complexes form on the millisecond to second time scale, whereas *N*-glycosidic bond cleavage takes place on the minute time scale (Wolfe and O'Brien 2009; Hendershot, Wolfe et al. 2011). Thus, destabilized base pairing allows AAG to selectively excise DNA lesions. More stringent selection takes place inside the active site, in which side chains create steric clashes with unmodified A and G bases (Lau, Wyatt et al. 2000; Connor and Wyatt 2002; O'Brien and Ellenberger 2004).

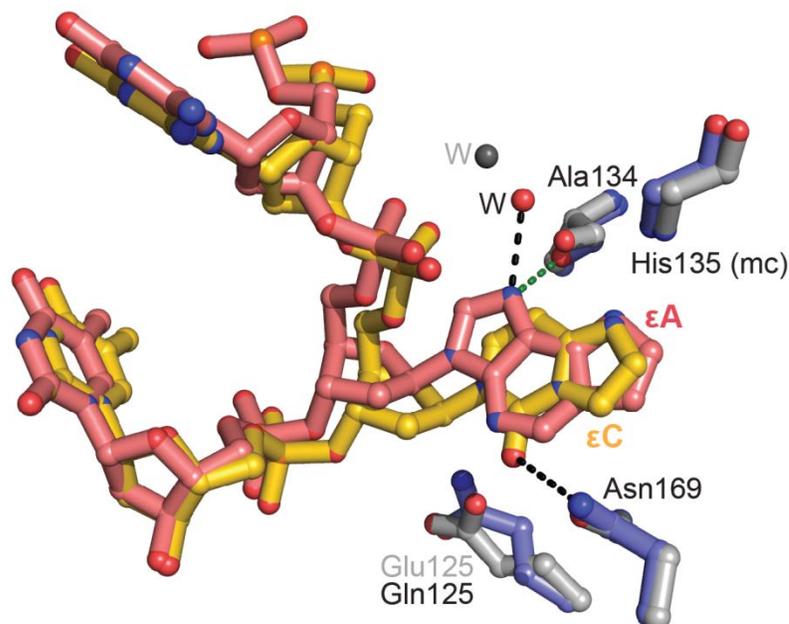
Excision of neutral substrates by AAG has been shown by pH-activity profiles to employ both a general acid and general base (O'Brien and Ellenberger 2003). The general acid acts to protonate the nucleobase, facilitating its dissociation, and the general base

would deprotonate a catalytic water molecule to attack C1' (O'Brien and Ellenberger 2003). Although the identity of the general acid has not been determined, the necessity to protonate the base explains the specificity of AAG for purines versus pyrimidines, and excision of positively charged lesions (e.g., 7mG) does not require the general acid (O'Brien and Ellenberger 2003). Quantum mechanical modeling studies indicate that base excision by AAG is facilitated by  $\pi$ - $\pi$  interactions between the enzyme and its substrate DNA, and suggest that the nucleobase is not fully protonated but rather hydrogen bond donation by a protein-bound water molecule lowers the catalytic barrier (Rutledge and Wetmore 2011).

## 2. Structural Basis of AAG Inhibition by $\epsilon$ C

In addition to  $\epsilon$ A, AAG has a modest activity toward 1, $N^2$ - $\epsilon$ G (Saparbaev, Langouet et al. 2002). Although AAG binds  $\epsilon$ C with a 2-fold greater affinity than  $\epsilon$ A (Lingaraju, Davis et al. 2011), AAG is incapable of excising  $\epsilon$ C, which is normally removed by the uracil/thymine DNA glycosylase family of enzymes (Saparbaev and Laval 1998; Gros, Ishchenko et al. 2003; Gros, Maksimenko et al. 2004). A recent structure of AAG in complex with  $\epsilon$ C-DNA showed  $\epsilon$ C to reside in the active site in a virtually identical position as  $\epsilon$ A (Lingaraju, Davis et al. 2011) (**Figure 8**). The hydrogen bond between His136 and  $\epsilon$ A ( $N^6$ ) is preserved to  $\epsilon$ C ( $N^4$ ), and as a consequence the  $\epsilon$ C nucleotide is pulled slightly farther into the binding pocket. The enhanced binding to  $\epsilon$ C may be explained by one additional hydrogen bond between the protein (Asn169) and  $O^2$  of  $\epsilon$ C, which is not present in  $\epsilon$ A. Regarding inhibition, protonation of substrate purines likely occurs at the N7 nitrogen (O'Brien and Ellenberger 2003), and crystal structures

suggest that a protonated N7 would be stabilized by a hydrogen bond to the backbone oxygen of Ala134 (**Figure 8**). The AAG/ $\epsilon$ C-DNA structure proposes that inhibition by  $\epsilon$ C is due to the inability of AAG to protonate  $\epsilon$ C, which lacks a nitrogen at the position corresponding to N7 of  $\epsilon$ A. In addition, this structure also showed an octahedral coordinate  $Mn^{2+}$  ion bound to the guanine opposite the  $\epsilon$ C that perturbed the guanine sugar pucker. This was the first observation of bound divalent ion to AAG and suggested that inhibition of the enzyme by divalent ions might be a consequence of impaired base flipping or duplex opening to expose the substrate base (Lingaraju, Davis et al. 2011). AAG has also been trapped onto  $\epsilon$ C-DNA in a non-specific orientation, providing a structural basis for the ability of the enzyme to bind single-base bulges (Klapacz, Lingaraju et al. 2010; Setser, Lingaraju et al. 2011).



**Figure 8.** Binding of  $\epsilon$ A and  $\epsilon$ C to AAG. The  $\epsilon$ A complex (PDB ID 1EWN) is colored blue (protein) and salmon (DNA), and the  $\epsilon$ C complex (PDB ID 3QI5) is silver and gold. Hydrogen bonds are depicted as dashed lines. A water molecule (red sphere) is in position to protonate N7 of  $\epsilon$ A, and protonated N7 would donate a hydrogen bond to Ala134 (green dashed line).  $\epsilon$ C does not have an ionizable group at this position.

### *3. Product release and diffusion along DNA*

Single- and multiple-turnover kinetic experiments have shown that the rate-limiting step of hypoxanthine hydrolysis by AAG is the release of the abasic DNA product (Baldwin and O'Brien 2010). In fact, the tight binding of AAG to product DNA enables AAG to catalyze the reverse reaction and re-form the *N*-glycosidic bond (Admiraal and O'Brien 2010). Product release is promoted by APE1, the next enzyme in the BER pathway (Baldwin and O'Brien 2009). Displacement of the glycosylase by APE1 has also been observed for TDG and OGG1 (Waters, Gallinari et al. 1999; Sidorenko, Nevinsky et al. 2007; Fitzgerald and Drohat 2008). The nonspecific binding of both AAG and APE1 to DNA suggests that these enzymes may bind DNA simultaneously and facilitate a handoff of the abasic site from AAG to APE1. Baldwin & O'Brien propose that APE1 displaces AAG from the AP site without a direct protein-protein interaction, and that AAG remains bound to the DNA upon AP dissociation (Baldwin and O'Brien 2010). The processivity of AAG along DNA is dependent on ionic strength, indicating a reliance on electrostatic interactions with the DNA backbone. Furthermore, the amino terminal 80 amino acids, which are not necessary for catalysis by AAG, contribute to the enzyme's ability to diffuse along DNA (Hedglin and O'Brien 2008).

### *Helix hairpin Helix Superfamily*

The majority of yeast, archaeal, and bacterial glycosylases adopt the HhH protein fold, with the exception of AlkC/AlkD and bacterial orthologs of human AAG (Aamodt, Falnes et al. 2004; Rubinson, Metz et al. 2008). The HhH glycosylases contain two  $\alpha$ -

helical domains with the active site cleft located at their interface. The domain containing the HhH motif and DNA intercalating residues is formed from an internal region of the primary structure and has a relatively conserved tertiary structure. The HhH anchors the protein to the DNA through a series of hydrogen bonds between main-chain atoms of the hairpin and the phosphoribose backbone downstream of the lesion. At the damage site, bulky side chains from neighboring loops fill the void left by the extrahelical nucleobase target and wedge into the base stack opposite the flipped out nucleotide. Both plug and wedge residues are important for stabilizing the bent conformation of the DNA and have been implicated in probing the DNA helix during the search process (Bowman, Lee et al. 2010). The second domain, formed from the N- and C-termini, is more structurally divergent and often contains additional structural elements, such as a zinc ion (TAG), iron-sulfur cluster (MpgII), or carbamylated lysine (MagIII) (Rubinson, Gowda et al. 2010).

Comparative analysis of the HhH alkylpurine glycosylases has been instrumental in deciphering the physical and chemical determinants of substrate recognition (Rubinson, Adhikary et al. 2010). On one hand, we have learned that the HhH scaffold accommodates a diverse array of nucleobase binding pockets that discriminate between lesions on the basis of shape complementarity. For example, the nucleobase binding surface of AlkA is a shallow cleft that can accommodate a variety of alkylpurines, whereas the active sites of TAG and MagIII are constrained and perfectly shaped for 3mA. On the other hand, this steric selection is not the only determinant of specificity since some active sites can accommodate nucleobases for which they do not excise (e.g., Mag1) (Adhikary and Eichman 2011). In addition, the catalytic requirements for excision

of cationic lesions 3mA and 7mG differ from the uncharged alkylpurines (e.g., εA) by virtue of their weaker *N*-glycosidic bonds (Stivers and Jiang 2003). Hence, the inherent instability of these lesions render their excision highly dissociative, and recent reports suggest that cationic lesions may be removed and even detected within DNA differently than neutral lesions (Metz, Hollis et al. 2007; Rubinson, Gowda et al. 2010; Adhikary and Eichman 2011).

### *1. E. coli AlkA*

Crystal structures of unliganded AlkA identified the enzymes as a member of the HhH superfamily and revealed a shallow nucleobase binding surface that can accommodate a variety of alkylpurines, a feature that helped to explain its broad specificity (Labahn, Scharer et al. 1996; Yamagata, Kato et al. 1996) (**Figure 7B**). In addition to the two-domain HhH architecture, AlkA contains an amino-terminal β-sheet domain of unknown function that is also present in OGG1 (**Figure 7**). A structure of AlkA bound to DNA containing 1-azaribose, which mimics the oxocarbenium reaction intermediate, has contributed greatly to our understanding of these enzymes (Hollis, Ichikawa et al. 2000; Hollis, Lau et al. 2000). The HhH anchors the protein to the DNA and does not directly participate in lesion recognition. The DNA is kinked by ~60° around the 1-azaribose, is rotated 180° around the phosphoribose backbone, and is pointed into the active site cleft (**Figure 7B**). Leu125 plugs the gap in the DNA. Rotation of the 1-azaribose into the active site places the N1' nitrogen directly adjacent to the carboxylate group of the catalytic Asp238, which likely stabilizes the oxocarbenium intermediate (Hollis, Ichikawa et al. 2000). In addition to this lesion-specific binding

mode, AlkA has the ability to bind to DNA ends (Zhao and O'Brien 2011), which may explain why a structure of AlkA bound to a substrate DNA has not been determined. Nonetheless, this feature was exploited to develop a host-guest crystallization strategy to determine structures of various lesions in DNA (Bowman, Lee et al. 2008).

High resolution structures of AlkA cross-linked to undamaged DNA bases provided insight into how the enzyme detects damage within the context of unmodified DNA (Bowman, Lee et al. 2010). Not surprisingly, the most notable differences between these undamaged DNA complexes (UDCs) and the 1-azaribose lesion recognition complex (LRC) are centered around the lesion. The UDCs do not exhibit the kink present in the LRC DNA. The domain containing most of the catalytically important residues, including Asp238, is shifted 2.4 Å toward the lesion strand in the LRC compared to the UDCs. This movement, combined with a modest 1 Å shift of the Leu125 plug residue toward the lesion strand, clamps the lesion between the two domains and creates additional protein contacts that stabilize the LRC. In contrast, the HhH motif makes the same DNA contacts in LRC and UDC structures, providing additional evidence that the HhH motif is a non-specific DNA binding motif and is not involved in distorting the DNA for catalysis. Leu125 in the UDCs does not interact with the DNA, although it is still present in the minor groove. The phosphate backbone in the LRC is significantly (~9 Å) closer to the protein, which allows the Leu125 side-chain to intercalate into the DNA base stack in that structure. A 3mA base modeled in place of a centrally located cytosine indicates that Leu125 likely makes van der Waals contacts to the *N*3-methyl group of the 3mA (Bowman, Lee et al. 2010). These observations suggest that AlkA employs a

passive scanning mechanism along the minor groove and uses the Leu125 side chain to detect abnormal bases and flip them into the active site.

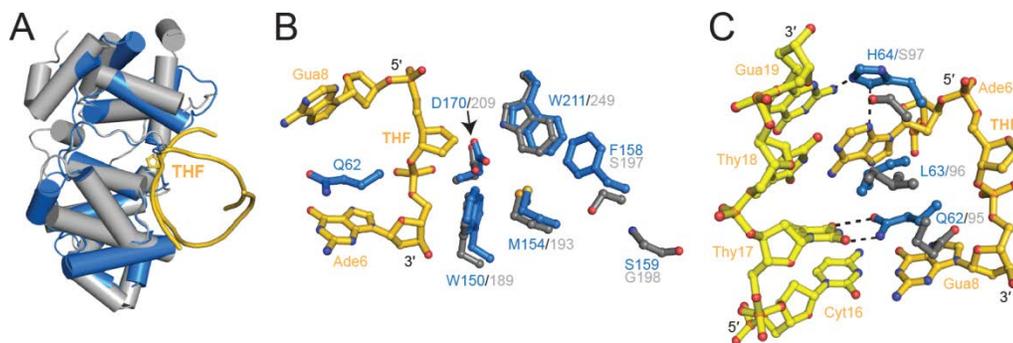
## 2. *Archaeal AlkA*

An AlkA ortholog from the archaeon *Archaeoglobus fulgidus* (AfAlkA), has been shown to excise 1mA and 3mC in addition to 3mA, 7mG,  $\epsilon$ A and Hx from DNA (Birkeland, Anensen et al. 2002; Mansfield, Kerins et al. 2003; Leiros, Nabong et al. 2007). The crystal structure of this ortholog shows that the nucleobase binding pockets of AfAlkA and *E. coli* AlkA are strikingly different despite the similarity in their overall architectures (Leiros, Nabong et al. 2007) (**Figure 7B,C**). The substrate nucleobase is predicted to stack between Phe133 and Phe282, similar to stabilization of 3mA by MagIII (**Figure 7E**). In support of this, substitution of Phe133 or Phe282 with alanine diminishes  $\epsilon$ A and 1mA base excision, and the double mutant abrogates activity. Arg286 is predicted to orient  $\epsilon$ A in the active site through hydrogen bonding, but would potentially repel the protonated amine groups of 1mA and 3mC (Leiros, Nabong et al. 2007). Mutation of the catalytic Asp240 (Asp238 in EcAlkA) completely eliminates base excision activity in AfAlkA.

## 3. *Yeast MAG/Mag1*

*S. cerevesiae* MAG and *S. pombe* Mag1 are 42% and 47% similar in sequence to *E. coli* AlkA, respectively, but have a more restricted substrate specificity (Rubinson, Adhikary et al. 2010). MAG excises 3mA, 7mG,  $\epsilon$ A, Hx, and guanine, but not oxidized substrates (e.g., O<sup>2</sup>-methylthymine) from DNA, while Mag1 is more restricted to 3mA,

3mG, and 7mG and has only a modest activity toward  $\epsilon$ A (Saparbaev and Laval 1994; Bjørås, Klungland et al. 1995; Saparbaev, Kleibl et al. 1995; Berdal, Johansen et al. 1998; Alseth, Osman et al. 2005; Lingaraju, Kartalou et al. 2008; Adhikary and Eichman 2011). Differences in substrate preferences are reflected in cells. For example, MAG deletion strains are more sensitive to alkylation agents than are *S. pombe mag1*. In addition, MAG expression is induced to higher levels than Mag1 upon exposure to alkylation agents (Chen and Samson 1991; Memisoglu and Samson 2000). These phenotypic differences suggest that these proteins have different roles in protecting cells against alkylation damage (Memisoglu and Samson 2000; Memisoglu and Samson 2000).



**Figure 9.** Yeast 3-methyladenine DNA glycosylases MAG and Mag1. In all panels, the unbound *Saccharomyces cerevisiae* MAG (grey) free enzyme is superimposed onto the *Schizosaccharomyces pombe* Mag1/THF-DNA complex (blue/gold, PDB ID 3S6I). (A) Overall structures. (B) Active sites. Mag1 residues Phe158 and Ser159 are the only two active site residues that differ between the two enzymes. (C) Close-up of Mag1-DNA contacts at the lesion. Swapping Mag1 His64 and MAG Ser97 between the two enzymes effectively swaps their respective abilities to remove  $\epsilon$ A (see text for details).

Our laboratory recently determined crystal structures of Mag1 bound to DNA containing a THF abasic analog (Adhikary and Eichman 2011) and of free MAG (unpublished results) (**Figure 9**). Neither MAG nor Mag1 contain the mixed  $\alpha/\beta$  domain present at the N-terminus of the AlkA orthologs (**Figure 9A**). Nevertheless, Mag1

engages the THF-DNA similarly to AlkA, with the DNA bent by  $\sim 60^\circ$  and the THF moiety rotated around the phosphate backbone toward the nucleobase binding pocket. Inside the active site, there are only two notable differences between MAG and Mag1. Mag1 residues Phe158 and Ser159 at the back of the binding cleft are occupied by Ser197 and Gly198 in MAG (**Figures 7D and 9B**). Swapping these residues (Mag1 FS $\rightarrow$ SG and MAG SG $\rightarrow$ FS double mutants) did not affect their relative  $\epsilon$ A activities, providing evidence that the bulky Phe residue in the binding pocket is not responsible for the lower  $\epsilon$ A excision activity of Mag1 (Adhikary and Eichman 2011). Interestingly, substitution of the catalytic aspartate residues had dramatically different effects. MAG Asp209Asn completely abrogated  $\epsilon$ A and 7mG excision activities similar to that observed for AlkA Asp238 (Labahn, Scharer et al. 1996), while Mag1 Asp170Asn had a more modest effect, implying that this residue in MAG plays a more significant role in catalysis, which may explain the broader substrate preference of this enzyme (Adhikary and Eichman 2011).

Outside of the active site, there is a notable difference between MAG and Mag1 at the point of contact with the DNA minor groove flanking the damage site (**Figure 9C**). In addition to the plug and wedge residues, His64 in Mag1 is in position to hydrogen bond with either the N3 of the adenine immediately 5' to the lesion or to the exocyclic N2 of the guanine on the opposite strand. MAG and AlkA orthologs, including those from *Bacillus halodurans* and *Deinococcus radiodurans*, which have broad substrate preferences and for which crystal structures are available, contain a serine residue at this position (**Figure 9C**) (Rubinson, Adhikary et al. 2010; Adhikary and Eichman 2011; Moe, Hall et al. 2012). Surprisingly, swapping histidine and serine between Mag1 and

MAG led to dramatic increase in  $\epsilon$ A excision rate in MagI and a decrease in  $\epsilon$ A excision in MAG, whereas the 7mG excision rates in both enzymes remained the same (Adhikary and Eichman 2011). This is another illustration of how contacts to the minor groove may be important for damage detection and/or stabilizing a specific enzyme-substrate complex for catalysis. These results also suggest that cationic lesions like 7mG may be detected or stabilized differently than neutral alkylpurines.

#### 4. *H. pylori* MagIII and *T. maritima* MpgII

MagIII and MpgII are related alkylpurine glycosylases identified by their sequence similarity to EndoIII (Begley, Haas et al. 1999; O'Rourke, Chevalier et al. 2000). MagIII is highly specific for 3mA but can excise mispaired 7mG, whereas MpgII can excise both 3mA and 7mG (Begley, Haas et al. 1999; O'Rourke, Chevalier et al. 2000). The crystal structure of MagIII showed a unique feature in the N/C-terminal domain, which contains a carbamylated lysine (Lys205) that neutralizes an otherwise highly positively charged region of the protein (Eichman, O'Rourke et al. 2003). MagIII's preference for 3mA can be explained by the snug fit of 3mA inside the active site, which partially excludes *N*7-substituted purines. Structures of MagIII bound to positively charged 3,9-dimethyladenine (3,9-dmA) and neutral  $\epsilon$ A bases showed the nucleobases to stack between Phe45 and Trp24 and to be bound on three sides by Trp25, Pro26, and Lys211 (**Figure 7E**). Other than these van der Waals and  $\pi$ -stacking interactions, there are no specific hydrogen bonding or polar contacts to the adenine ring like those observed in TAG (see section 5). Similar to MagI, mutation of the putative catalytic aspartate Asp150 in MagIII does not completely abrogate base excision activity, again suggesting

that the catalytic power of this residue determines the ability of the HhH enzymes to remove more stable, neutral nucleobases from DNA, and that little catalytic assistance is required for hydrolysis of the labile 3mA glycosidic bond (Eichman, O'Rourke et al. 2003; Stivers and Jiang 2003).

Unlike MagIII, MpgII contains an iron-sulfur cluster and shows robust activity toward 7mG, which is intriguing given the sequence similarity between MagIII and MpgII (Begley, Haas et al. 1999; Rubinson, Adhikary et al. 2010). Although there is no structure for MpgII, sequence comparison predicts that only two residues differ within the active site: MpgII Trp52 and Lys53 are occupied by Phe45 and Glu46 in MagIII, respectively. The MagIII active site is constrained by a salt bridge between Glu46-Lys211. Substitution of Glu46 with the corresponding lysine residue (Lys53) in MpgII should relieve this constraint from electrostatic repulsion. Indeed, a MagIII Glu46Lys mutant resulted in an 8-fold increase in 7mG•T activity, suggesting that steric exclusion of 7mG partially accounts for the low activity of MagIII toward methylguanine bases (Eichman, O'Rourke et al. 2003).

#### *5. E. coli TAG*

TAG substrate preference is strictly limited to *N*3-substituted purines 3mA and 3mG (Bjelland, Bjørås et al. 1993) and does not have the catalytic aspartate residue present in other 3mA DNA glycosylases. NMR studies of *E. coli* TAG showed it to be a structurally divergent member of the HhH family and to contain a zinc ion in the N/C-terminal domain (Drohat, Kwon et al. 2002; Cao, Kwon et al. 2003; Kwon, Cao et al. 2003). Similar to MagIII, TAG's specificity can be partially attributed to the fact that the

3mA binding pocket would sterically exclude all other nucleobases (**Figure 7F**). Binding studies and NMR investigation of 3mA in the active site led to the suggestion that TAG enhances the rate of 3mA depurination by binding tightly to the nucleobase, thereby destabilizing the ground state of the enzyme-substrate complex (Cao, Kwon et al. 2003). This idea was illustrated by crystal structures of a TAG/abasic-DNA/3mA product complex using the *Salmonella typhi* ortholog, which is 82% identical and 92% conserved overall with *E. coli* TAG (Metz, Hollis et al. 2007). In that structure, the bound DNA is more B-form when compared to the highly distorted 1-azaribose DNA bound to AlkA, and there was a large (7 Å) separation between the THF, which is not fully engaged inside the active site, and 3mA, which is buried deep inside the cleft. These observations indicated that the DNA undergoes significant relaxation upon breakage of the *N*-glycosidic bond, and it was suggested that steric strain may contribute to bond cleavage (Metz, Hollis et al. 2007). A recent structure of *Staphylococcus aureus* TAG recapitulates the structural features observed in the *E. coli* and *S. typhi* structures, and the authors suggested that tautomerization of 3mA contributes to its recognition by TAG (Zhu, Yan et al. 2012).

### **Uracil / Thymine / 5mC**

G•U and G•T mismatches arise from deamination of cytosine and 5-methylcytosine (5mC), respectively, and lead to A•T transition mutations (Coulondre, Miller et al. 1978; Duncan and Miller 1980). Uracil is excised in eukaryotes by uracil DNA glycosylase (UDG, also known as UNG), single-stranded monofunctional uracil glycosylase (SMUG), and to a lesser extent by thymine DNA glycosylase (TDG). In

bacteria, uracil is removed by the UDG ortholog, Ung, and mispaired uracil glycosylase (MUG) (Lindahl 1974; Gallinari and Jiricny 1996; Haushalter, Todd Stukenberg et al. 1999; Kavli, Sundheim et al. 2002). Thymine is removed from G•T mismatches by TDG and MBD4 in eukaryotes and by archaeal mismatch specific glycosylase (MIG) (Brown and Jiricny 1987; Begley and Cunningham 1999; Hendrich, Hardeland et al. 1999). With the exception of MBD4 and MIG, which belong to the HhH superfamily, the UDG/TDG glycosylases adopt a highly conserved  $\alpha/\beta$  fold (**Figure 6**) and can be divided into 4 subfamilies on the basis of sequence similarity and substrate specificity (Mol, Arvai et al. 1995; Mol, Arvai et al. 2002; Wu, Qiu et al. 2003). UDG family 1 contains UDG/UNG and is defined by the landmark structures of the human and viral enzymes in various states, which revealed mechanistic details about substrate recognition and catalysis common to the entire superfamily (Mol, Arvai et al. 1995; Mol, Kuo et al. 1995; Savva, McAuley-Hecht et al. 1995; Slupphaug, Mol et al. 1996). Family 2 is composed of thymine-specific TDG and MUG, which are homologous to UDG in structure but not sequence (Barrett, Savva et al. 1998; Barrett, Savva et al. 1998; Barrett, Scharer et al. 1999; Maiti, Morgan et al. 2008). The third family is defined by SMUG, and the fourth by *Thermus thermophilus* TDG. The common  $\alpha/\beta$  fold of the UDG superfamily contains a positively-charged groove approximately the width of a DNA duplex that is ideal for binding double-strand DNA (Mol, Arvai et al. 1995).

UDG has served as a model for understanding the structural and biochemical functions of DNA glycosylases in general, and recent work has focused on the mechanism by which the enzyme locates uracil amidst undamaged DNA. This collective body of work on UDG has been the subject of several recent reviews (Krokan, Drablos et

al. 2002; Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; O'Brien 2006; Stivers 2008; Friedman and Stivers 2010; Zharkov, Mechetin et al. 2010), and thus will not be discussed here. We instead focus on recent structural results for TDG in light of new evidence implicating this enzyme in active 5mC demethylation (Cortellino, Xu et al. 2011; Jacobs and Schar 2011).

#### *A possible role of BER in DNA demethylation*

In addition to repair of thymine mismatches, the biological functions of TDG and MBD4 may extend beyond DNA repair as a defense against mutations to a potential role in regulating gene expression and DNA demethylation (Cortazar, Kunz et al. 2007; Cortellino, Xu et al. 2011). 5mC is an important marker for gene expression, X-chromosome inactivation, and transposon silencing among other developmental processes (Chan, Henderson et al. 2005; Huh, Bauer et al. 2008; Law and Jacobsen 2010). Whereas DNA methylation mechanisms are relatively well understood (Goll and Bestor 2005), the demethylation pathways are not. Demethylation can occur passively after replicative synthesis of unmethylated daughter strands or actively by demethylase enzymes. In plants, active demethylation takes place by the DME/ROS1 family of 5mC DNA glycosylases, but an analogous 5mC glycosylase has not been discovered in mammals. TDG and MBD4 have been implicated in active demethylation on the basis of their abilities to excise mispaired thymine produced from AID- or APOBEC-dependent 5mC deamination, and recent studies have shown TDG to be necessary for maintenance of epigenetic stability (Rai, Huggins et al. 2008; Cortazar, Kunz et al. 2011; Cortellino, Xu et al. 2011). In addition, the recent discoveries that the ten-eleven translocation (TET)

proteins oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (**Figure 4C**), and that 5hmC is found at transcriptional start sites and within actively transcribed genes raises the distinct possibility that these 5mC derivatives and their deamination products are intermediates in a BER-dependent active demethylation pathway (Tahiliani, Koh et al. 2009; Ito, D'Alessio et al. 2010; Ficz, Branco et al. 2011; Guo, Su et al. 2011; Guo, Su et al. 2011; He, Li et al. 2011; Ito, Shen et al. 2011; Pastor, Pape et al. 2011; Veron and Peters 2011; Wu and Zhang 2011). Indeed, TDG is capable of excising 5mC oxidation products 5fC and 5caC (Maiti and Drohat 2011; Zhang, Lu et al. 2012), further implicating the thymine glycosylases in active demethylation.

### *Structural Insight into TDG Function*

Structures of the catalytic domain of TDG (residues 111-308) bound to substrate and product DNA and conjugated by the regulatory protein SUMO have provided a basis to understand the sequence specificity and mechanisms of base excision and product release of TDG (Baba, Maita et al. 2005; Maiti, Morgan et al. 2008; Manvilla, Maiti et al. 2012; Zhang, Lu et al. 2012). In addition, an NMR study of the N-terminal regulatory domain of TDG (residues 1-111) supports models for allosteric control of TDG activity (Steinacher and Schar 2005; Smet-Nocca, Wieruszeski et al. 2008).

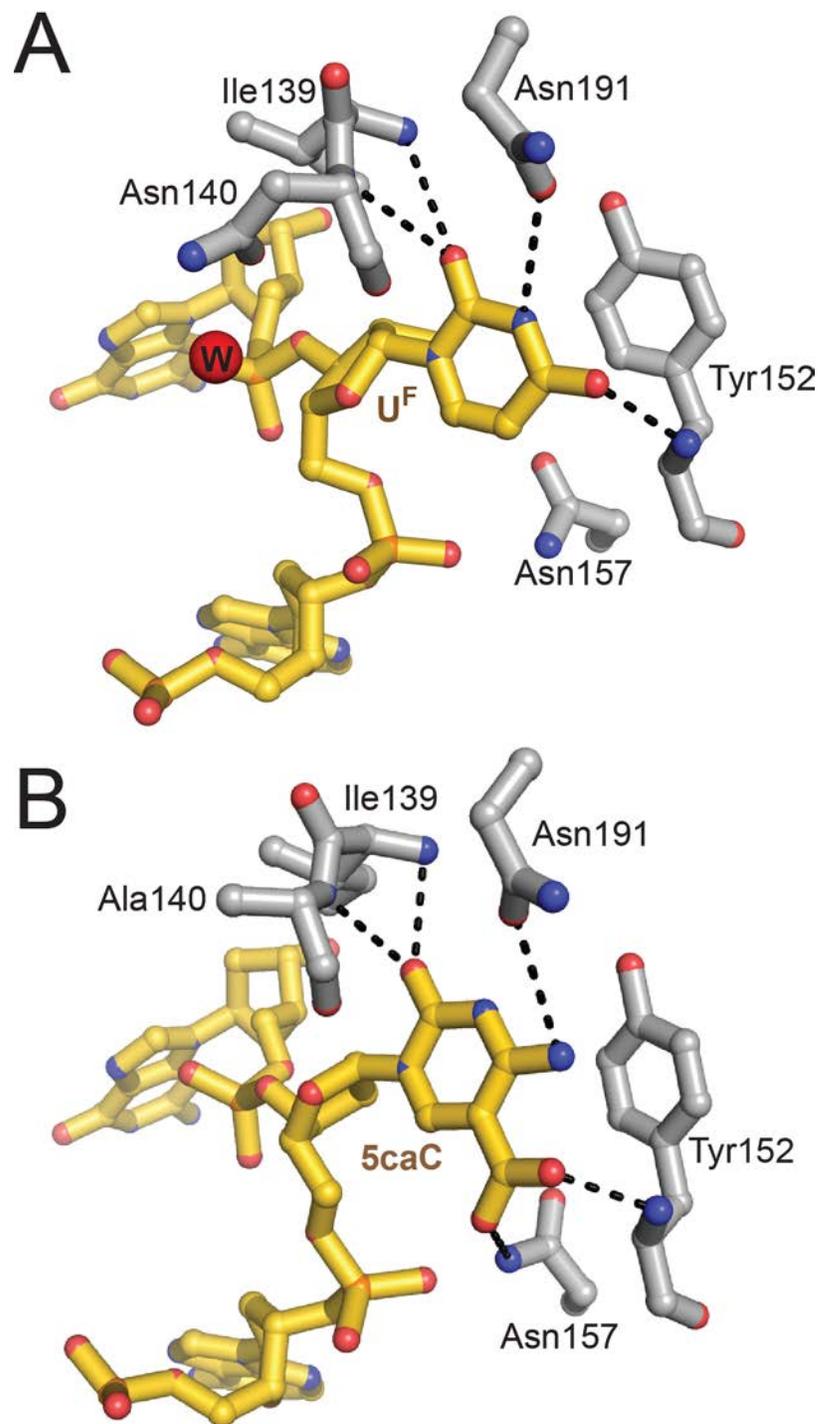
#### *1. TDG-DNA complexes*

Structures of TDG bound to duplex DNA containing a THF product mimic provided the general features of DNA binding and the first glimpse into thymine

recognition (Maiti, Morgan et al. 2008). In this structure, two TDG molecules bound to a single 22-nucleotide DNA duplex, with one protein anchored at the abasic site and the other positioned at an undamaged site, although biochemical analysis indicated that only one protein per lesion is required for catalysis (Maiti, Morgan et al. 2008; Morgan, Maiti et al. 2011).

The TDG complex is very similar to DNA-bound structures of UDG and *E. coli* MUG, with notable exceptions. Both TDG and UDG impose a  $\sim 43^\circ$  bend in the substrate DNA at the abasic site, although MUG does not significantly bend the DNA (Barrett, Savva et al. 1998; Parikh, Mol et al. 1998; Barrett, Scharer et al. 1999; Stivers and Jiang 2003). TDG utilizes an arginine (Arg275) to plug the gap created by the flipped nucleotide, whereas UDG and MUG have leucine plugs. Substitution of Arg275 with either alanine or leucine results in a significant decrease in both the rate of thymine excision and substrate binding (Maiti, Morgan et al. 2009). Another unique aspect of the TDG family are lysine residues Lys246 and Lys248, which make contacts to the DNA backbone of the non-damaged strand, 8 and 9 nucleotides away from the damage site, providing an explanation for TDG's requirement for 9-10 bases 5' to the lesion (Scharer, Kawate et al. 1997; Maiti, Morgan et al. 2008).

Two TDG-substrate-DNA complexes were recently determined that provided a snapshot of an uncleaved nucleobase in the active site (Maiti, Noon et al. 2012; Zhang, Lu et al. 2012). One structure contained the wild-type enzyme bound to DNA containing a non-hydrolyzable dU mimetic (2'-deoxy-2'-fluoroarabinouridine, U<sup>F</sup>) (**Figure 10A**) and the other trapped 5caC in the active site by utilizing a variant TDG containing an Asn140Ala substitution (**Figure 10B**), which had previously been shown to decrease the



**Figure 10.** Protein-DNA contacts within the TDG active site. **(A)** TDG bound to DNA containing 2'-deoxy-2'-fluoroarabinouridine ( $U^F$ ) (PDB ID 3UFJ). Hydrogen bonds are shown as dashed lines and the putative catalytic water molecule is a red sphere. **(B)** TDG in complex with 5-carboxylcytosine (5caC)-DNA.

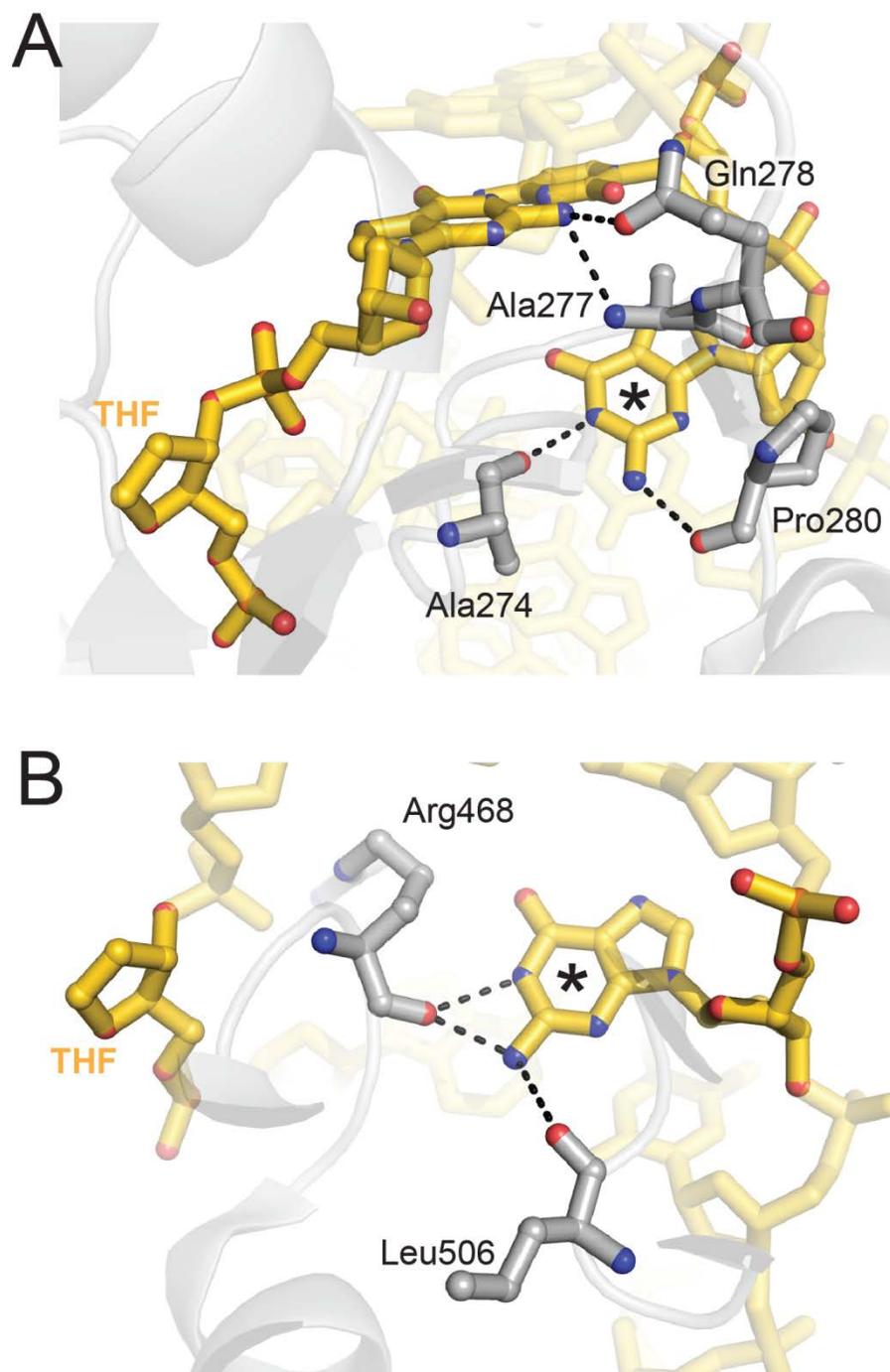
rate of thymine excision while having only marginal effect on substrate binding (Maiti, Morgan et al. 2008; Maiti, Morgan et al. 2009).

The overall structures of U<sup>F</sup> and 5caC substrate complexes are similar to the THF product complex, and the active sites reveal common modes of recognition of the two substrates. A hydrogen bond was observed from Asn191 to N4 in the 5caC structure and to pyrimidine N3 in the U<sup>F</sup> structure, and this contact is conserved in UNG and SMUG1 but not MUG. In UNG and SMUG1, this asparagine side chain forms an additional hydrogen bond to the pyrimidine O4 (Parikh, Mol et al. 1998; Barrett, Scharer et al. 1999; Wibley, Waters et al. 2003). Maiti et al suggest the differential orientation of this residue in UNG and SMUG1 prevents these enzymes from excising cytosine analogs such as 5fC and 5caC (Maiti, Noon et al. 2012). In addition, 5caC and likely thymine participate in van der Waals interactions with Ala145 and both U<sup>F</sup> and 5caC form hydrogen bonds with main chain atoms of Tyr152, at either the O4 of uracil (and thymine) or the carboxyl group of 5caC. The 5caC forms an additional interaction with Asn157. Even though modeling a thymine base into the active site of the U<sup>F</sup> structure shows a steric clash with Ala145, this residue is able to accommodate the 5-carboxyl group in the 5caC structure. Nevertheless, Ala145Gly and His151Ala mutants both increase TDG's thymine excision activity by 13-fold over the wild-type. Maiti et al propose that His151 slows the cleavage reaction by destabilizing the partial negative charge that develops during the reaction (Maiti, Noon et al. 2012). The mutations showed an even greater increase in activity for thymine from normal A•T base pairs, suggesting that these highly conserved residues are needed to limit aberrant action on undamaged DNA (Maiti, Noon et al. 2012). Zhang, et al. found that TDG binds DNA containing

5caC with higher affinity than 5fC, U, and T, which they attribute to the hydrogen bonds to the electronegative 5caC carboxyl group from Asn157, His151, and Tyr152 (Zhang, Lu et al. 2012). The inability of other members of the UDG family, including SMUG1 and UDG, to bind 5caC and 5fC may result from the presence of residues that interfere with these interactions (Zhang, Lu et al. 2012).

Regarding catalysis, stabilization of Asn140 and the  $\beta$ 2- $\alpha$ 4 catalytic loop, which encircles the active site, by Thr197 was found to be important for TDG function, as a Thr197Ala substitution resulted in a 32-fold reduction in base excision activity (Maiti, Noon et al. 2012). Interestingly, the U<sup>F</sup> structure suggested that a putative water nucleophile, which is absent from other TDG structures and the enzyme-substrate complex for related MUG enzymes, is positioned by the side chain and main chain of Asn140 and Thr197, respectively. The low resolution and extensive merohedral twinning of the crystallographic data precludes an unambiguous assignment of such a water molecule. However, the presence of a water at this position in TDG is consistent with those observed in the high-resolution structure of free MUG (Barrett, Savva et al. 1998) and with the putative water nucleophiles located in high-resolution structures of UNG (Parikh, Mol et al. 1998; Xiao, Tordova et al. 1999).

TDG makes several contacts to the guanine opposite the lesion that offer an explanation for the enzyme's specificity for thymine in certain sequence contexts. Specificity for G•T mispairs (Waters and Swann 1998) is dictated by Ala274 and Pro280, which make guanine-specific hydrogen bonds from their backbone oxygen atoms to N1 and N2 of the G opposite the abasic site (**Figure 11A**). The Ala274 contact is conserved by UDG enzymes, whereas the Pro280 interaction is unique to TDG. In addition, TDG



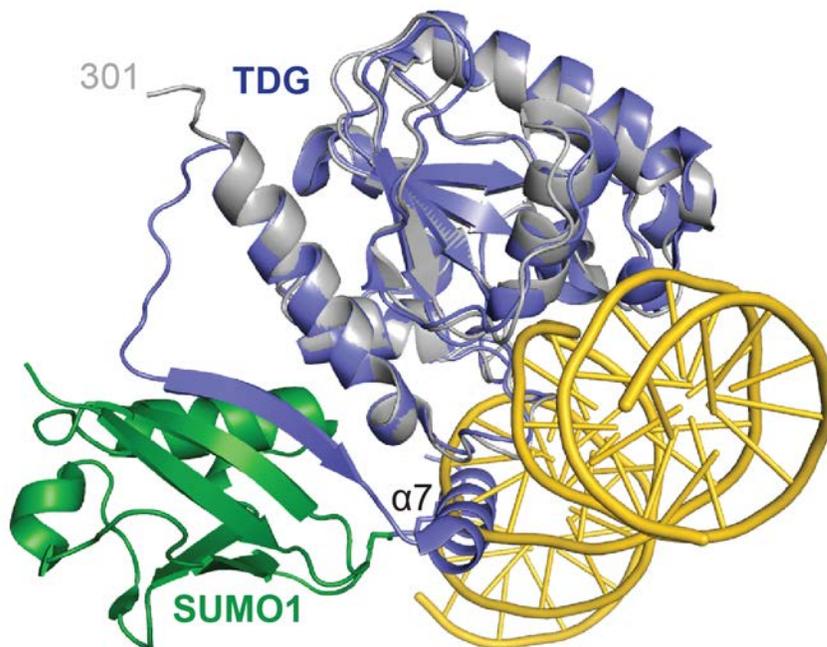
**Figure 11.** Comparison of TDG (A) and MBD4 (B) contacts to the strand opposite the lesion. The guanine base opposite the THF abasic site is marked with an asterisk. (A) TDG/THF-DNA complex (PDB ID 2RBA). (B) MBD4/THF-DNA complex (4DK9).

has the highest affinity for thymine that is immediately 5' to a guanine (i.e., in a TpG/CpG dinucleotide step) (Sibghat, Gallinari et al. 1996; Morgan, Bennett et al. 2007). Interestingly, TDG does not contact the 5' cytosine on the non-lesion strand. The specificity for TpG/CpG is likely explained by contacts to the TpG guanine from the conserved Gln278 side chain and Ala277 main chain, as well as the Ala274/Pro280 contacts to the G•T guanine described above (**Figure 11A**) (Waters and Swann 1998; Morgan, Bennett et al. 2007; Maiti, Morgan et al. 2008).

## 2. TDG-SUMO interaction

Sumoylation of TDG at lysine 330, located at the C-terminal end of the catalytic domain, weakens TDG-DNA binding (Hardeland, Steinacher et al. 2002; Steinacher and Schar 2005). The crystal structure of human TDG conjugated with SUMO1 was determined in 2005 and provided insight into how SUMO1 modulates TDG-DNA binding (Baba, Maita et al. 2005). C-terminal residues of TDG (307-330) form a crossover  $\beta$ -strand that extends a  $\beta$ -sheet with SUMO1. In addition, TDG and SUMO1 interact through a series of main chain hydrogen bonds and side chain hydrophobic and polar interactions at its SUMO-interacting site (SIM). A mutational analysis revealed that these noncovalent bonds are necessary for SUMO1-induced disruption of DNA binding by TDG, confirming an earlier study (Hardeland, Steinacher et al. 2002; Baba, Maita et al. 2005). The covalent tethering of SUMO1 to the C-terminus of TDG places helix  $\alpha 7$  in an outwardly extended conformation from the rest of the protein. Superposition of sumoylated and DNA-bound forms of TDG predict that helix  $\alpha 7$  in this orientation collides with the DNA strand immediately opposite the lesion (**Figure 12**). The nature of

the conformational change in the C-terminal end of the TDG glycosylase domain is uncertain since helix  $\alpha 7$  was not present in the DNA-bound structures. Nevertheless, the SUMO1-TDG structure suggests that sumoylation of TDG locks helix  $\alpha 7$  in an orientation that promotes DNA release. A structure of SUMO3 conjugated to TDG shows very similar binding between the two proteins as with SUMO1 (Baba, Maita et al. 2006).



**Figure 12.** SUMO1 modified TDG creates steric clash with DNA. The SUMO1-modified TDG structure (blue TDG, green SUMO1, PDB ID 1WYW) is superimposed onto the TDG/THF-DNA complex (silver/gold, 2RBA). SUMO1 modification holds helix  $\alpha 7$  in a position that would presumably clash with the DNA.

### 3. TDG regulatory domain

TDG contains at its N-terminus a lysine-rich regulatory domain (RD) that interacts with DNA and a number of proteins involved in genome maintenance. The RD binds to DNA methyltransferase DNMT3a (Li, Zhou et al. 2007) and is a target for

acetylation by transcriptional co-activators CBP (CREB binding protein)/p300, which aids in recruitment of APE1 (Tini, Benecke et al. 2002). In addition, the RD is important for TDG specificity for G•T mismatches (Gallinari and Jiricny 1996). This regulatory domain is highly flexible and contains a non-specific DNA binding function that is modulated by sumoylation of the catalytic domain, thereby affecting its enzymatic activity (Steinacher and Schar 2005). A recent NMR study of this domain revealed residues 1-50 to be unstructured even in the context of the full protein. Residues 51-111, on the other hand, showed a modest degree of structure and an extended conformation that contacts the catalytic domain in the context of the full-length protein (Smet-Nocca, Wieruszeski et al. 2008). The authors propose an electrostatic interaction between the regulatory and catalytic domains that modulates rates of thymine and uracil excision, supporting previous models for allosteric control of G•T specificity (Steinacher and Schar 2005; Smet-Nocca, Wieruszeski et al. 2008).

#### *MBD4*

Mismatch specific thymine glycosylase MBD4 contains a methyl-CpG-binding domain (MBD) and a C-terminal glycosylase domain that preferentially excises T mispaired with G (Hendrich, Hardeland et al. 1999). Crystal structures of the glycosylase domains of mouse and human MBD4 showed that the enzyme belongs to the helix-hairpin-helix structural superfamily (Wu, Qiu et al. 2003; Zhang, Liu et al. 2011). A recent structure of MBD4's glycosylase domain bound to THF-DNA showed the overall DNA binding regime characteristic of HhH glycosylases, including a 57° bend in the DNA, an extrahelical THF moiety, the opposite base (guanine) stacked in the duplex, and

plug (Leu506) and wedge (Arg468) residues to stabilize the flipped nucleotide and distorted duplex (Manvilla, Maiti et al. 2012). Like TDG, MBD4 makes several guanine-specific contacts to the base opposite the thymine, and thus the structure provides a rationale for discrimination of G•T over A•T base pairs (**Figure 11B**). Specifically, Leu506 and Arg468 mainchain oxygens participate in hydrogen bonds or polar interactions with the N1 and exocyclic N2 of guanine, contacts which cannot be made to adenine (Maiti, Morgan et al. 2008; Manvilla, Maiti et al. 2012) (**Figure 11B**). Finally, although MBD4 has been proposed to process G•T mispairs created by active demethylation (Rai, Huggins et al. 2008), MBD4 is not able to excise 5fC or 5caC (He, Li et al. 2011; Manvilla, Maiti et al. 2012). Manvilla et al attribute this lack of activity to incompatibility between the MBD4 active site and the negative charges that develop on 5fC and 5caC upon dissociation (Manvilla, Maiti et al. 2012).

### *DME/ROS1*

Plants contain a family of 5mC glycosylases, represented by the *Arabidopsis* proteins DEMETER (DME), Repressor of Silencing 1 (ROS1), DME like 2 (DML2) and DML3, that are responsible for active demethylation via BER. DME is responsible for endosperm gene imprinting and is necessary for seed viability (Choi, Gehring et al. 2002). The DML enzymes, including ROS1, likely function as genome wide demethylases, particularly at sites 5' and 3' to genes to regulate transcription, protecting plants from erroneous gene silencing (Gong, Morales-Ruiz et al. 2002; Zhang, Yazaki et al. 2006; Zhu, Kapoor et al. 2007; Zilberman, Gehring et al. 2007). DME, ROS1, and DML3 excises 5mC in CpG, CpNpG, and CpNpN contexts, and DML2 was shown to

have 5mC excision activity in a CpG context (Gehring, Huh et al. 2006; Morales-Ruiz, Ortega-Galisteo et al. 2006; Penterman, Zilberman et al. 2007).

The DME/ROS1 family of enzymes utilize a bifunctional glycosylase-lyase mechanism and adopt an iron-sulfur-containing HhH glycosylase domain similar to EndoIII and MutY (Choi, Gehring et al. 2002; Gehring, Huh et al. 2006; Penterman, Zilberman et al. 2007; Ponferrada-Marin, Roldan-Arjona et al. 2009). A homology model of ROS1 using BsEndoIII as a template helped to identify several residues conserved among the HhH superfamily that are important for ROS1 function (Ponferrada-Marin, Parrilla-Doblas et al. 2011). Tyr606 and Asp611 are predicted to reside near the base recognition pocket and are necessary for the excision of both 5mC and thymine from G•T mismatches. A Tyr606Leu mutant slightly reduced the DNA binding activity, whereas Asp611Val mutation increased DNA binding with respect to wild-type ROS1 (Ponferrada-Marin, Parrilla-Doblas et al. 2011). The homology model also revealed that two aromatic residues (Phe589 and Tyr1028) conserved in the DME family are positioned to interact with the nucleobase lesion. Interestingly, Phe589Ala and Tyr1028Ser mutations changed the preference of ROS1 from 5mC to T•G mismatches, indicating that they are important for substrate recognition (Ponferrada-Marin, Parrilla-Doblas et al. 2011). We have constructed a similar homology model for the glycosylase domain of DME, and find analogous residues predicted to be necessary for catalysis (See Chapter III).

Unlike other glycosylases, DME/ROS1 proteins contain two additional domains flanking the glycosylase domain and that are essential for 5mC excision (Morales-Ruiz, Ortega-Galisteo et al. 2006; Mok, Uzawa et al. 2010; Ponferrada-Marin, Martinez-Macias

et al. 2010). The C-terminal domain lacks any identifiable sequence homology to proteins outside of the DME/ROS1 family, and thus additional structural and functional studies will be necessary to ascertain its function. The N-terminal domain, on the other hand, contains a conserved lysine-rich region required for ROS1 binding to non-methylated DNA and enhances ROS1 specificity for 5mC over T. Deletion of this domain caused ROS1 to process long DNA substrates less effectively (Ponferrada-Marin, Martinez-Macias et al. 2010). On its own, the N-terminal region binds DNA with a high affinity. Deletion of the N-terminal domain reduces the ability of ROS1 to bind DNA and excise 5mC, but does not affect the ability of DME to bind methylated and non-methylated DNA (Mok, Uzawa et al. 2010; Ponferrada-Marin, Martinez-Macias et al. 2010). The N-terminal and glycosylase domains are separated by ~240 and 400 residues in ROS1 and DME, respectively. Deletion of this interdomain region (IDR1) in DME does not affect 5mC excision activity (Mok, Uzawa et al. 2010). The ROS1-EndoIII homology model predicted that this linker region is an inserted sequence within the glycosylase domain and that the lesion-intercalating plug residue is N-terminal to the IDR1 insertion (Ponferrada-Marin, Parrilla-Doblas et al. 2011). A study of DME and DML3 and comparison to the ROS1 homology model continues in Chapter III, where the specificity of the DME/ROS1/DML3 enzymes for 5mC is investigated.

## Chapter III

### 5-METHYLCYTOSINE RECOGNITION BY *ARABIDOPSIS THALIANA* DNA GLYCOSYLASES DEMETER AND DML3\*

#### Introduction

5-Methylcytosine (5mC) is an important epigenetic modification that serves as a marker for gene expression, X-chromosome inactivation, and transposon silencing, among other developmental processes (Chan, Henderson et al. 2005; Huh, Bauer et al. 2008; Law and Jacobsen 2010). Abnormalities in the regulation of DNA methylation may lead to neuronal disorders and cancer development (Feinberg, Ohlsson et al. 2006; Dulac 2010). 5mC is generated by a class of DNA methyltransferases, which use S-adenosylmethionine to methylate cytosine at position C5 by a relatively well characterized mechanism (Goll and Bestor 2005). In contrast, the mechanism of 5mC → C demethylation is less well understood. DNA demethylation may occur passively after synthesis of unmethylated daughter strands during replication or actively by enzymes that remove the methyl group through chemical modification. Active demethylation has recently been linked to the base excision repair (BER) pathway, which is normally associated with removal of detrimental nucleobase modifications from the genome (Lindahl 2000; Piccolo and Fisher 2013).

---

\* Part of the work presented in this chapter was published in Brooks, S. B., Fischer, R. L., Huh, J. H., and Eichman, B.F. (2014) 5-methylcytosine recognition by *Arabidopsis thaliana* DNA glycosylases DEMETER and DML3. *Biochemistry*, 53(15): 2525-2532.

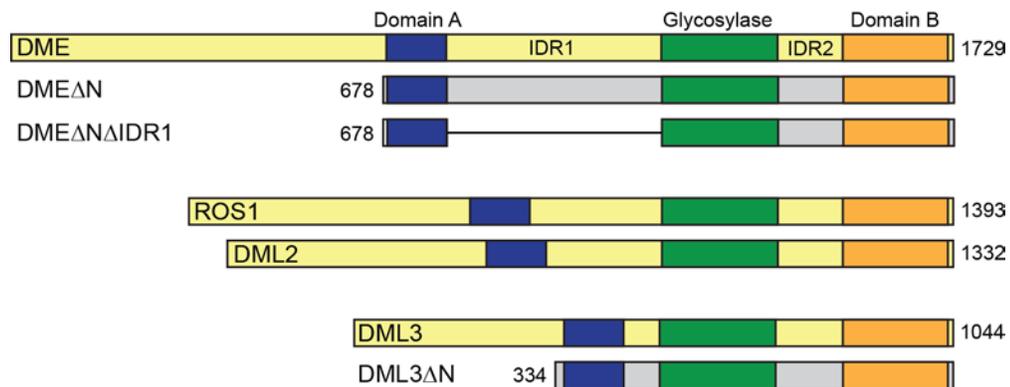
BER is initiated by DNA glycosylases, which recognize and remove a specifically modified nucleobase by cleaving the *N*-glycosidic bond (reviewed in Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; Brooks, Adhikary et al. 2013). In mammals, 5mC may be removed in two ways. In one pathway, activation-induced (DNA-cytosine) deaminase (AID) converts 5mC to thymine, which is excised from the resulting T/G mismatch by thymine DNA glycosylase (TDG) and MBD4 (Rai, Huggins et al. 2008; Cortazar, Kunz et al. 2011; Cortellino, Xu et al. 2011). Alternatively, 5mC is oxidized by ten-eleven translocation (TET) proteins to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Tahiliani, Koh et al. 2009; Ito, Shen et al. 2011; Veron and Peters 2011; Wu and Zhang 2011). Of these, 5fC and 5caC are substrates for TDG (Maiti and Drohat 2011; Zhang, Lu et al. 2012).

In contrast to mammals, plants have evolved specific 5mC DNA glycosylases that remove DNA methylation (Choi, Gehring et al. 2002; Gehring, Huh et al. 2006; Morales-Ruiz, Ortega-Galisteo et al. 2006). *Arabidopsis thaliana* DEMETER (DME) functions during plant gametogenesis before fertilization and is responsible for imprinting specific genes in the endosperm, which is necessary for seed viability (Choi, Gehring et al. 2002; Hsieh, Shin et al. 2011). DME demethylation in the central and vegetative cells is also thought to produce small interfering RNAs (siRNA) that guide methylation at transposons in the egg and sperm cells, respectively (Calarco, Borges et al. 2012; Ibarra, Feng et al. 2012). *Arabidopsis* also contains three DME paralogs—Repressor of Silencing 1 (ROS1), DME-like 2 (DML2), and DML3—which function in adult cells as genome-wide demethylases that remove 5mC marks at sites 3' and 5' to genes (Gong, Morales-Ruiz et al. 2002; Zhang, Yazaki et al. 2006; Penterman, Uzawa et al. 2007; Penterman,

Zilberman et al. 2007; Zhu, Kapoor et al. 2007; Zilberman, Gehring et al. 2007; Ortega-Galisteo, Morales-Ruiz et al. 2008). Recent reports of DME activities indicate that the control of DNA methylation can be utilized agriculturally to benefit crop production and medically to develop therapies against celiac disease by suppressing DME expression to produce wheat varieties lacking gliadins and glutenins that cause immunogenic epitopes (Qi, Wei et al. 2009; Kim, Kwak et al. 2010; Mitea, Salentijn et al. 2010; Osorio, Wen et al. 2012; Wen, Wen et al. 2012; Kapazoglou, Drosou et al. 2013).

The DME/ROS1/DML enzymes contain a conserved DNA glycosylase domain belonging to the helix-hairpin-helix [4Fe-4S] iron-sulfur cluster superfamily (Choi, Gehring et al. 2002). DME and ROS1 utilize a bifunctional glycosylase/lyase mechanism to cleave the glycosidic bond and the phosphodiester backbone through  $\beta,\delta$ -elimination to create a one-nucleotide gap in the DNA (Agius, Kapoor et al. 2006; Gehring, Huh et al. 2006). DME/ROS1/DML enzymes are much larger than other glycosylases, ranging from 1100 to over 1700 residues, and contain two additional domains (A and B) with no known homology to other proteins (**Figure 13**). Like the glycosylase domain, domains A and B are conserved among the DME paralogs and are essential for DME function (Mok, Uzawa et al. 2010). Homology modeling of ROS1 predicted a bipartite discontinuous glycosylase domain separated in sequence by an unstructured interdomain region (IDR) (Ponferrada-Marin, Parrilla-Doblas et al. 2011). All DME paralogs contain an N-terminal lysine-rich region that in ROS1 is necessary for binding long stretches of DNA in a methylation-independent manner (Choi, Gehring et al. 2002; Ponferrada-Marin, Martinez-Macias et al. 2010).

In contrast to ROS1, which has been the subject of several biochemical studies (Ponferrada-Marin, Roldan-Arjona et al. 2009; Ponferrada-Marin, Martinez-Macias et al. 2010; Ponferrada-Marin, Parrilla-Doblas et al. 2011; Parrilla-Doblas, Ponferrada-Marin et al. 2013), little is known about the molecular mechanisms of DME and DML enzymes. Here, we investigated 5mC excision and substrate specificity of DME and DML3 by mutational analysis. Residues that constitute the active site and DNA damage binding face of DME were predicted from Endonuclease III (EndoIII/Nth) homology modeling and the importance of these residues to DNA binding and 5mC excision tested. In addition, we previously showed that DME is able to excise 5hmC (Jang, Shin et al. 2014), and we now extend this activity to DML3 and report a comprehensive kinetic analysis of oxidized 5mC excision by DME and DML3. These results provide new insight into the molecular underpinnings for 5mC specificity and excision by the DME family of enzymes.



**Figure 13.** DME paralogs and constructs. Schematic representation of the four *Arabidopsis thaliana* DME proteins (yellow) and DME and DML3 deletion constructs (grey) used in this study. Conserved domains are indicated by blue (domain A), green (glycosylase), and orange (domain B) boxes. Deletion constructs ( $\Delta$ N) lack the N-terminal region preceding domain A, and the inter-domain region (IDR) 1 has been replaced with a synthetic dodecapeptide linker in DME $\Delta$ N $\Delta$ IDR1. Numbers refer to amino acid positions at the termini.

## Materials and Methods

### *Protein Expression and Purification*

*Arabidopsis thaliana* DME $\Delta$ N (residues 678-1729) and DME $\Delta$ N $\Delta$ IDR1, in which residues 798-1188 were replaced with the dodecapeptide AGSSGNGSSGNG, were overexpressed as N-terminal His<sub>6</sub>-MBP-fusion proteins as described previously (Mok, Uzawa et al. 2010). The region encoding *A. thaliana* DML3 amino acids 334-1044 (DML3 $\Delta$ N) was PCR-amplified from full-length DML3 cDNA template (Penterman, Zilberman et al. 2007) using primers 5'- AAT TGG ATC CGT AAC AAC GAT GAT CAA AGC) and 5'-AAT TGA ATT CCT ATA TAT CAT CAT CAC TCA TAA AC. The PCR amplicon was digested with BamHI and EcoRI and cloned into the pET27 (Novagen) derived expression vector pBG102 (Vanderbilt Center for Structural Biology), which produces a cleavable N-terminal His<sub>6</sub>-SUMO-fusion protein. All constructs were transformed into *E. coli* Rosetta2 (DE3) cells (Novagen) and propagated in LB media to an OD of 0.5. Proteins were overexpressed for 18 h at 16 °C upon addition of 0.1 mM IPTG. Cells were harvested by centrifugation at 6,000 rpm and lysed in 50 mM Tris-HCl pH 8.5, 500 mM NaCl, and 10% glycerol with an Avestin Emulsifier C3 homogenizer operating at ~20,000 psi. The lysates were cleared by centrifugation at 22,000 rpm for 20 minutes at 4 °C. Fusion proteins were purified using Ni-NTA (Qiagen) affinity chromatography, followed by cleavage of the affinity tag by PreScission Protease overnight at 4 °C. The proteins were further purified by heparin-sepharose (GE Healthcare) in 50 mM Tris-HCl pH 8.5, 10% glycerol, 0.1 mM DTT, 0.1 mM EDTA and a NaCl gradient (0.1 – 1 M), followed by gel filtration on a 16/60 Superdex 200 column (GE Healthcare) in 20 mM Tris-HCl pH 8.5, 200 mM NaCl, 10% glycerol, 0.1 mM DTT,

and 0.1 mM EDTA. Protein was concentrated to 100-150  $\mu$ M and stored at -80  $^{\circ}$ C in 20 mM Tris-HCl pH 8.5, 150 mM NaCl, 40% glycerol, 0.1 mM DTT, and 0.1 mM EDTA. Mutant proteins were prepared by site-directed mutagenesis using a Quik-Change Kit (Stratagene) and purified in the same manner as the wild-type proteins.

### *Glycosylase Activity Assay*

Proteins were dialyzed in reaction buffer (20 mM Tris-HCl pH 8.5, 150 mM NaCl, 0.1 mM DTT, 0.1 mM EDTA) prior to use. Oligonucleotides of the sequence d(TGACTACTACATGXTTGCCTACCAT), in which X is 5mC, 5hmC, 5fC, 5caC, or T, were synthesized with 6-carboxyfluorescein (FAM) at the 5' end by Integrated DNA Technologies (5mC and T) and Midland Certified Reagents (5hmC, 5fC, and 5caC) and annealed to the complementary strand containing G opposite base X. Enzyme (at least 10  $\mu$ M for single-turnover conditions) was incubated with 100 nM duplex DNA at 25  $^{\circ}$ C in reaction buffer. For data shown in **Figure 14C**, 2  $\mu$ M protein (non-saturating) was used. Aliquots of 8  $\mu$ L were removed from each reaction and terminated with stop buffer (10 mM EDTA in formaldehyde with xylene cyanol and bromophenol blue). Substrate and product DNA were separated by gel electrophoresis with a 20% acrylamide gel in TBE buffer run at 40 W for 1 hour. Gels were imaged using a Typhoon Trio variable mode imager and quantified using ImageQuant software. Single-turnover rate constants ( $k_{st}$ ) were determined by exponential fit of the fraction product versus time. To establish the apparent  $K_M$  ( $K_{1/2}$ ) for each substrate, rate constants ( $k_{obs}$ ) were determined at varying enzyme concentrations, and plots of  $k_{obs}$  vs [E] were fit to the equation,  $k_{obs} = k_{st}[E] / (K_{1/2}$

+ [E]). Under non-saturating conditions,  $k_{\text{obs}}$  is dependent on enzyme concentration and is thus reported in units of  $\text{M}^{-1}\text{s}^{-1}$ .

For inhibition experiments, the glycosylase assay was carried out with 10  $\mu\text{M}$  (5mC excision) and 2  $\mu\text{M}$  (T/G excision) DME $\Delta\text{N}\Delta\text{IDR1}$  in the presence of varying concentrations of unlabeled competitor DNA containing tetrahydrofuran (THF), a 1-nucleotide gap, or thymine (T/G mismatch) in the central position. Gap-DNA was assembled by annealing the two 12-mer oligonucleotide sequences flanking the central position to the 25-mer opposite strand. To determine the inhibitory effects of free 5mC or Thy base on 5mC excision, 10  $\mu\text{M}$  DME $\Delta\text{N}\Delta\text{IDR1}$  was incubated with varying concentrations of free base (0.01 – 31 mM) for 30 minutes prior to the addition of 100 nM FAM-5mC-DNA substrate. The  $K_i$  was determined by plotting  $k_{\text{obs}}$  as a function of competitor DNA concentration and fitting the data to the equation,  $k_{\text{obs}} = k_{\text{sr}}[\text{E}] / (K_i + [\text{E}])$ .

### *DNA Binding*

Oligonucleotides containing the same sequence as above and a centrally located 5mC were annealed to 6-carboxyfluorescein at the 3'-end. Varying concentrations of enzyme (0-100  $\mu\text{M}$ ) were incubated with 50 nM duplex DNA in reaction buffer for 10 min at 25 °C. Fluorescence anisotropy was measured using a Biotek Synergy H1 hybrid multi-mode microplate reader at excitation and emission wavelengths of 485 and 538 nm, respectively. Dissociation constants ( $K_d$ ) were calculated by fitting the data to a two-state binding model. Reported values are averages from three independent experiments. Because DME reacts slowly, there is no significant product generated within 10 minutes.

Anisotropy values measured 0, 10, and 30 minutes after addition of wild-type enzyme did not change.

#### *DME Homology Model*

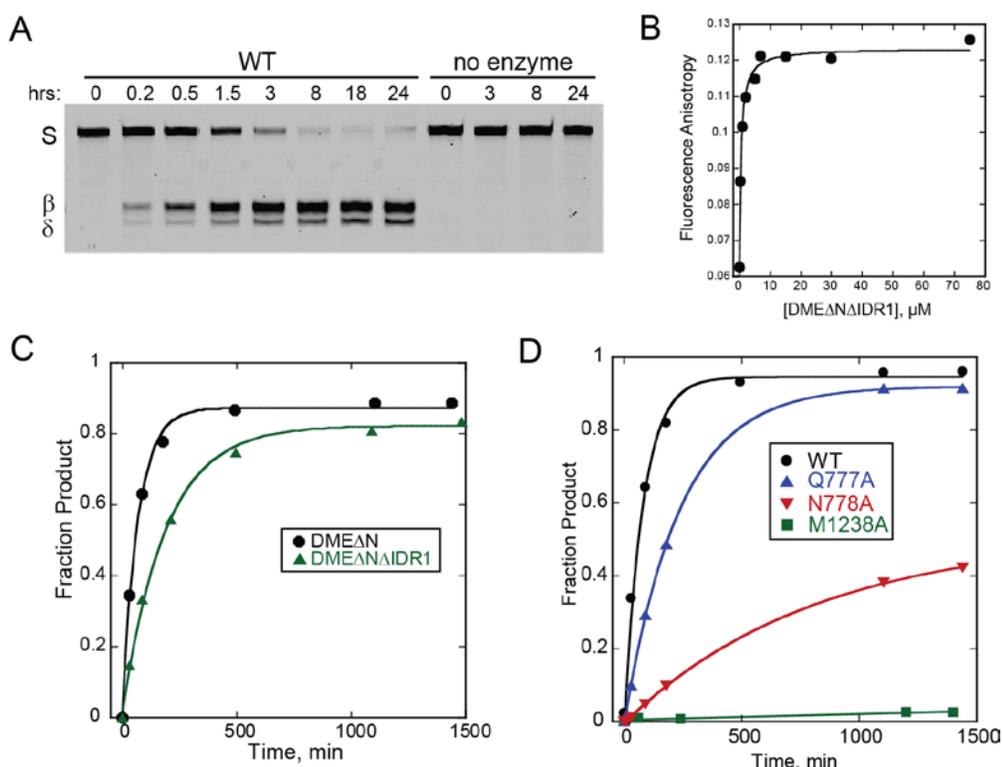
A homology model of DME bound to abasic-DNA was constructed using Swiss-Model (Guex and Peitsch 1997) and the structure of *Bacillus stearothermophilus* EndoIII/Nth (PDB ID 1P59 (Fromme and Verdine 2003)) as a template, similar to that previously described for ROS1 (Ponferrada-Marin, Parrilla-Doblas et al. 2011). The final model incorporated DME residues 737 – 796 of domain A and 1217 –1396 of the glycosylase domain. Protein-DNA contacts were predicted using the coordinates of the tetrahydrofuran (THF)-containing DNA from the Nth structure superimposed on the DME model.

## **Results**

#### *A Model for DNA Binding by DME*

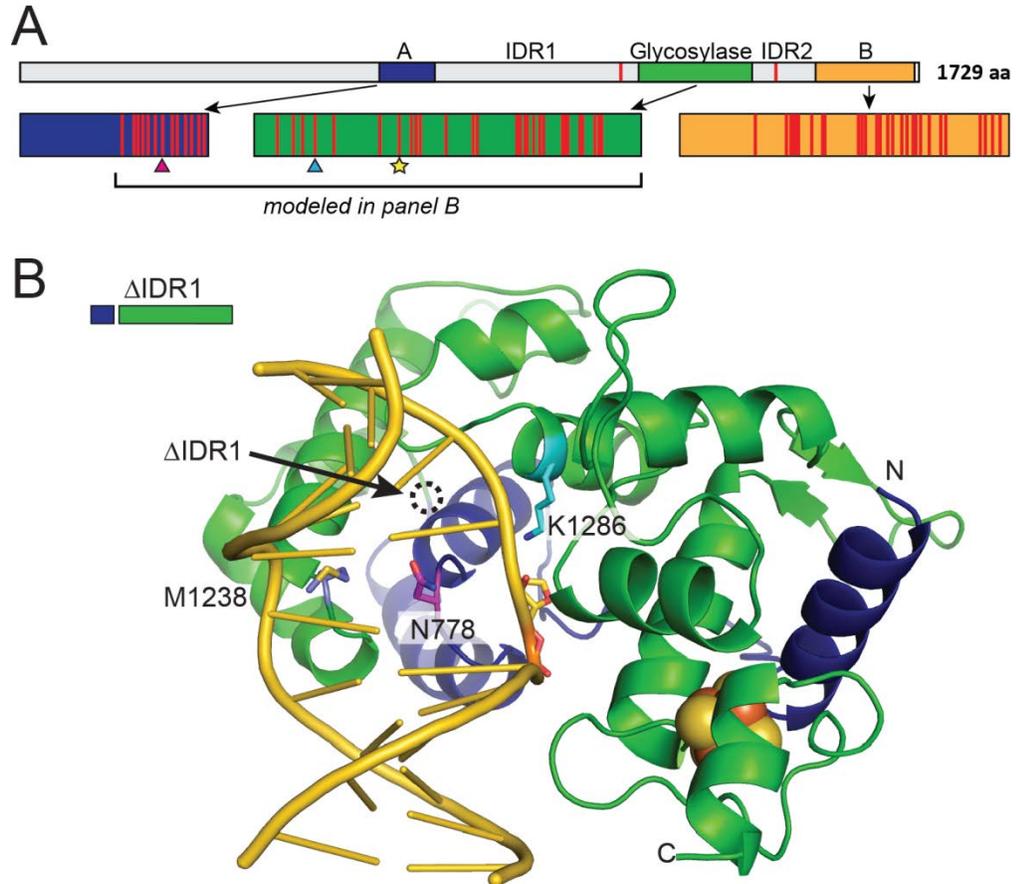
*Arabidopsis* DME, ROS1, DML2, and DML3 each contain three highly conserved domains important for function (**Figure 13**). The N-terminal and inter-domain regions (IDRs) have very low homology and little predicted secondary structure. We previously showed that the three conserved domains are necessary and sufficient for DME 5mC excision activity, as constructs lacking the N-terminal 677 residues (DME $\Delta$ N) and with the IDR1 replaced with a short dodecapeptide linker (DME $\Delta$ N $\Delta$ IDR1) retain activity (Mok, Uzawa et al. 2010). In order to quantify the effect of IDR1 removal on enzymatic activity, we compared the kinetics of 5mC excision of DME $\Delta$ N and DME $\Delta$ N $\Delta$ IDR1 proteins. Under non-saturating conditions, the observed rate constant

( $k_{obs}$ ) of DME $\Delta$ N is  $134 \pm 9 \text{ M}^{-1}\text{s}^{-1}$ , compared to  $43 \pm 1 \text{ M}^{-1}\text{s}^{-1}$  for DME $\Delta$ N $\Delta$ IDR1 (Figure 13, 14). Thus, removal of the 393-residue region between domain A and glycosylase domain has only a 3-fold effect on the rate of 5mC excision, which, under the conditions of this assay, is not significant. However, deletion of IDR1 significantly improves protein solubility, purity, and stability, mainly as a result of proteolytic sensitivity of IDR1, making DME $\Delta$ N $\Delta$ IDR1 more amenable to biochemical analysis than DME or DME $\Delta$ N. Therefore, the DME $\Delta$ N $\Delta$ IDR1 construct was used for the remainder of this study.



**Figure 14.** 5mC excision kinetics and DNA binding of DME. (A) Representative denaturing polyacrylamide gel showing DNA glycosylase activity. Reactions contained FAM-5mC-DNA and either 17  $\mu\text{M}$  or 0  $\mu\text{M}$  DME $\Delta$ N $\Delta$ IDR1. Numbers above the gel represent reaction times in hours. Substrate (S) and  $\beta$  and  $\delta$  elimination products are labeled. (B) Representative fluorescence anisotropy binding isotherm for wild-type DME $\Delta$ N $\Delta$ IDR1 binding to 25mer 5mC-DNA. (C) 5mC excision activities of DME $\Delta$ N (black circles) and DME $\Delta$ N $\Delta$ IDR1 (green triangles). (D) 5mC excision activities of plug and wedge mutants compared to DME $\Delta$ N $\Delta$ IDR1 (WT) used to determine single-turnover rates ( $k_{st}$ ) shown in Table 1.

In order to gain insight into the DME structural features important for 5mC excision, we constructed a homology model of the glycosylase domain using the structure of *Bacillus stearothermophilus* Nth, a closely related [4Fe-4S]-containing glycosylase specific for oxidized purines. As reported previously for ROS1 (Ponferrada-Marin, Parrilla-Doblas et al. 2011), the C-terminal region (residues 737-796) of domain A is predicted to complete the helix-hairpin-helix glycosylase fold based on its significant sequence conservation with the N-terminal 60 residues of Nth (**Figure 15B**). The conserved regions place DME Phe796 (domain A) in close proximity to DME Asp1217 (glycosylase domain) such that IDR1 is excluded entirely and most likely an independent domain. This aspect of the model is validated by the retention of activity in the DME $\Delta$ N $\Delta$ IDR1 mutant and the inclusion of functionally important residues within the region contributed by domain A (Mok, Uzawa et al. 2010). In addition, the model places the catalytic Asp1304 and Lys1286 in close proximity to the flipped abasic site. Substitution of either of these residues (Asp1304Asn and Lys1286Gln) abrogates 5mC excision activity (**Table 1**) (Gehring, Huh et al. 2006).



**Figure 15.** DME domain structure and distribution of critical residues. **(A)** Schematic of *A. thaliana* DEMETER (DME), showing the location of the glycosylase domain (green) and domains A (blue) and B (gold) of unknown function. IDR1, inter-domain region 1; IDR2, inter-domain region 2. Red bars mark the locations of residues identified by random mutagenesis to be critical for DME function. Putative residues important for DNA binding and catalysis and conserved in other HhH enzymes are marked by symbols below the schematic (Asn778 DNA plug, magenta triangle; Met1238 DNA wedge, blue triangle; catalytic Lys1286, yellow star). **(B)** Homology model of the C-terminal half of domain A and the glycosylase domain with DNA superimposed from the structure of *Bacillus stearothermophilus* EndoIII bound to abasic DNA (PDB ID 1P59). The putative catalytic Lys1286 (cyan) Met1238 wedge (slate) residues are contributed by the glycosylase domain (green), the putative Asn778 plug residue (magenta) is from domain A.

Damage recognition and catalysis by DNA glycosylases is facilitated by extrusion of the target nucleobase inside an active site pocket. This extrahelical conformation is stabilized by intercalating side chains that plug the gap left by the flipped nucleobase and

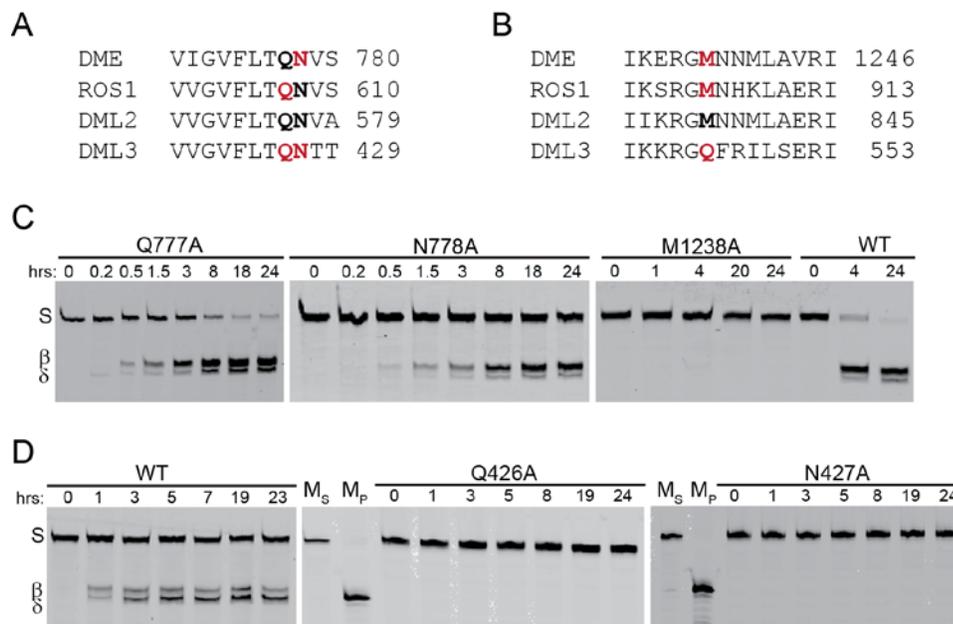
create a wedge in the opposite strand to create a sharp kink in the DNA duplex (Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; David, O'Shea et al. 2007; Dalhus, Laerdahl et al. 2009; Friedman and Stivers 2010; Li 2010; Brooks, Adhikary et al. 2013). Substitutions to these intercalating residues typically have a dramatic effect on glycosylase activity. Our homology model predicts that Asn778 and Met1238 would serve as the plug and wedge residues in DME (**Figure 15B**). Substitution of Met1238 with alanine in DME $\Delta$ N $\Delta$ IDR1 abrogated 5mC excision activity and had no effect on DNA binding (**Figure 14D** and **16, Table 1**), consistent with results from the

<b>Table 1.</b> DNA binding and 5mC excision activities for DME $\Delta$ N $\Delta$ IDR1 point mutants				
	$K_d$ ( $\mu$ M)	Relative affinity	$k_{st}$ ( $\times 10^{-5} \text{ s}^{-1}$ )	Relative to WT
WT	$8.3 \pm 1.2$	1.0	$18.1 \pm 1.2$	1.0
Q777A	$8.1 \pm 1.1$	1.0	$7.6 \pm 0.4$	0.4
N778A	$18.9 \pm 3.4$	0.4	$2.7 \pm 0.6$	0.1
M1238A	$8.5 \pm 1.0$	1.0	$0.02 \pm 0.01$	0.001
K1286Q	$8.2 \pm 1.8$	1.0	$< 0.01^\dagger$	0.0005
D1304N	$0.6 \pm 0.1$	14	$< 0.01^\dagger$	0.0005

Dissociation constants ( $K_d$ ) and single-turnover rate constants for 5mC ( $k_{st}$ ) excision were measured at 25 °C, pH 8.5 and 170 mM ionic strength using a 25mer DNA duplex containing 5mC and end-labeled with fluorescein. Enzyme concentration was saturating (10  $\mu$ M) in activity assays. Values are reported as averages  $\pm$  S.D. from at least three experiments. Relative binding affinities calculated as ( $K_d$  WT)/( $K_d$  Mutant); Relative to WT rates calculated as ( $k_{st}$  Mutant)/( $k_{st}$  WT).

$^\dagger$  Activity is below limit of detection.

corresponding residue (Met905) in ROS1 (**Figure 16**) (Ponferrada-Marin, Parrilla-Doblas et al. 2011; Parrilla-Doblas, Ponferrada-Marin et al. 2013). Similarly, an Asn778Ala substitution had a 10-fold reduction in activity relative to WT (**Figure 14D, Table 1**). Interestingly, the corresponding substitution in ROS1 (Asn608Ala) had no effect (Ponferrada-Marin, Parrilla-Doblas et al. 2011). Rather, the neighboring ROS1 residue (Gln607) was found to be necessary for 5mC removal and duplex interrogation (Ponferrada-Marin, Parrilla-Doblas et al. 2011; Parrilla-Doblas, Ponferrada-Marin et al. 2013). In contrast, substitution of this neighboring residue in DME (Gln777) with alanine had only a 2-fold effect on 5mC excision and no effect on DNA binding relative to wild-type (**Figure 14D and 16, Table 1**).



**Figure 16.** Identification of DNA intercalation residues in DME and DML3. (A,B) Sequence alignments of the regions containing the plug (A) and wedge (B) DNA intercalating residues. Residues predicted and confirmed by mutagenesis to intercalate DNA are boldface and red, respectively. (C,D) Base excision activity of DME $\Delta$ N $\Delta$ IDR1 (C) and DML3 $\Delta$ N (D) point mutants. Numbers above the gels represent the reaction times in hours. MS, molecular weight standard for 25mer substrate; MP, molecular weight standard for 12mer product.

We also tested these two putative plug residues in DML3ΔN by making alanine substitutions corresponding to DME Gln777 and Asn778. In this case, both DML3 Gln426Ala and Asn427Ala abrogated 5mC excision activity (**Figure 16**), suggesting that DME and DML3 utilize subtly different modes of damage recognition. Thus, we have identified DME and DML3 residues from discrete domains as likely candidates for duplex interrogation and/or stabilization of extrahelical 5mC.

<b>Table 2.</b> Rates of excision of oxidized 5mC derivatives by DME and DML3.					
	5mC	5hmC	5fC	5caC <sup>†</sup>	T/G
DMEΔNΔIDR1	18.1 ± 1.2	8.3 ± 1.1	0.02 ± 0.01	8.4 ± 1.6	6.5 ± 0.6
DML3ΔN	9.1 ± 0.3	11.4 ± 5.3	0.04 ± 0.02	35 ± 10	9.2 ± 0.6

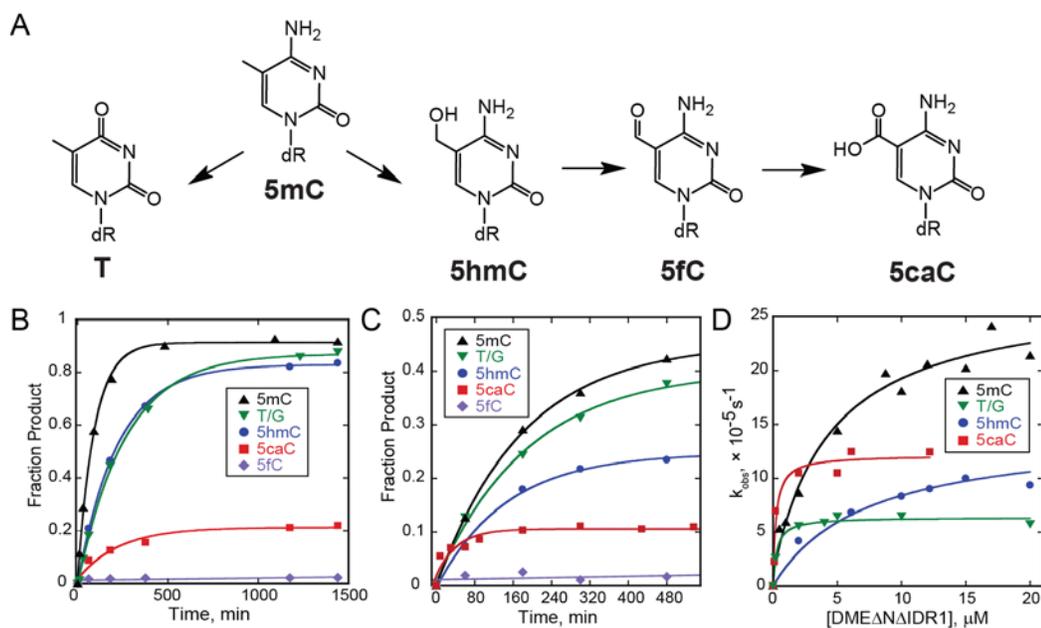
Single-turnover rate constants ( $k_{st}$ , x  $10^{-5}$  s<sup>-1</sup>) for excision of oxidized 5mC derivatives from a 25mer oligonucleotide at 25 °C, pH 8.5 and 170 mM ionic strength, and 10μM enzyme. Values are reported as averages ± S.D. from at least three experiments.

<sup>†</sup> Enzyme removes only 20% 5caC.

#### *Specificity of DME and DML3 for oxidized 5mC*

In mammals, 5mC is oxidized to 5hmC, 5fC, and 5caC by TET proteins (Tahiliani, Koh et al. 2009; Ito, Shen et al. 2011; Veron and Peters 2011; Wu and Zhang 2011) (**Figure 17A**). We recently found that DME and ROS1 have activity toward 5hmC (Jang, Shin et al. 2014). Although we were unable to detect these derivatives in *Arabidopsis*, we compared their rates of excision by DMEΔNΔIDR1 and DML3ΔN against 5mC and thymine from T/G-mismatches as a means to understand the molecular basis for the unique 5mC specificity of the DME family of enzymes. Under single-turnover conditions, DME exhibited the same activity for T/G and 5hmC, with first order rate constants ( $k_{st}$ ) only 2-3-fold less than for 5mC excision (**Figure 17B, Table 2**).

Interestingly, the rate of 5caC excision also was 2-fold less than 5mC (**Table 2**), although the enzyme was only able to process 20% of the substrate (**Figure 17B**), suggesting either a mixture of 5caC species (e.g., amino and imino forms) (Maiti, Michelson et al. 2013 1045) was present in the DNA or a product of 5caC excision inhibited the reaction. Under the same conditions, DME was inactive toward 5fC (**Figure 17B**). DML3 excised 5mC 2-fold slower than DME and showed the same general trend toward excision of oxidized 5mC derivatives (**Figure 17C and Table 2**).



**Figure 17.** Excision of oxidized 5mC nucleobases by DME and DML3. (A) 5mC can be deaminated to form thymine or successively oxidized to 5hmC, 5fC, and 5caC. (B,C) Glycosylase activities of DMEΔNΔIDR1 (B) and DML3ΔN (C) used to determine single-turnover rates ( $k_{st}$ ) shown in Table 2. (D) DMEΔNΔIDR1 concentration dependence on rates of 5mC, 5hmC, 5caC, and T/G excision used to determine  $K_{1/2}$  values shown in Table 3.

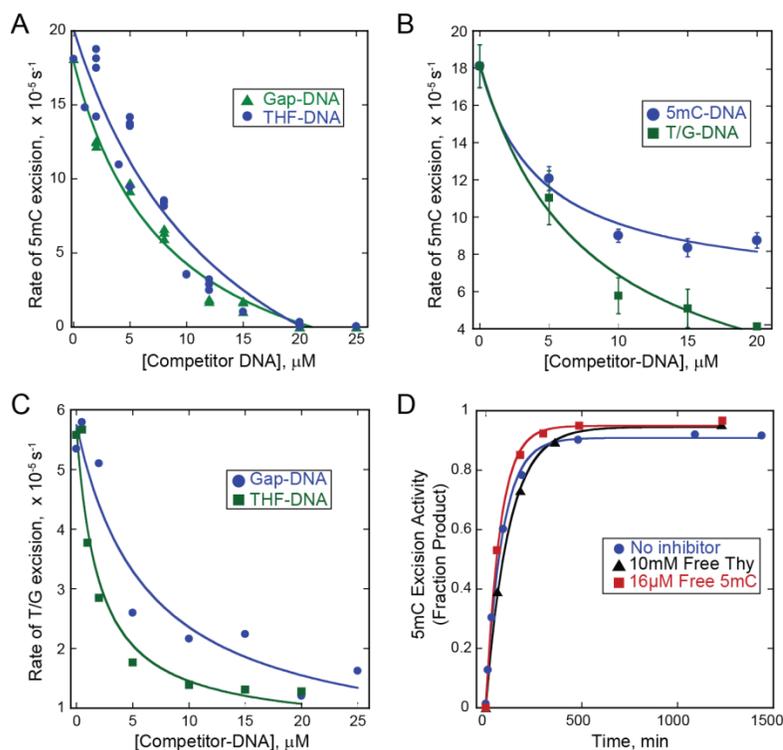
In order to establish the preferred substrate of DME, we determined the apparent  $K_M$  ( $K_{1/2}$ ) using the DMEΔNΔIDR1 construct. Interestingly, the preference of DME for each substrate varied greatly (**Figure 17D**). The  $K_{1/2}$  for 5hmC was comparable to 5mC,

whereas  $K_{1/2}$  values for 5caC and T/G were at least 25-fold smaller, indicating a preference for binding these substrates (**Table 3**). However, the  $k_{st}$  for 5hmC, 5caC, and T/G were all lower than 5mC, suggesting inhibition by an intermediate or product of these reactions (**Figure 17D**).

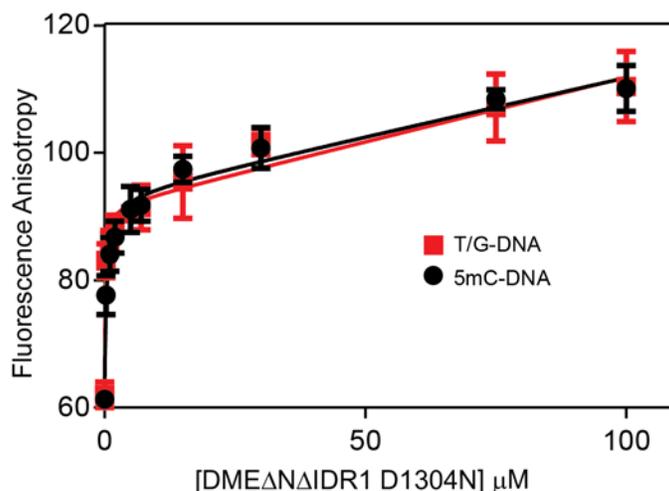
<b>Table 3. Substrate specificity of DME</b>	
Substrate	$K_{1/2}$ , $\mu\text{M}$
5mC	$5.1 \pm 1.7$
5hmC	$7.4 \pm 0.8$
5fC	ND
5caC	$0.2 \pm 0.9$
T/G	$0.12 \pm 0.04$
ND, not determined. Values are reported as averages $\pm$ S.D. from at least three experiments.	

Because DNA glycosylases are often product inhibited, and abasic sites opposite 5mC inhibit DME activity (Gehring, Huh et al.), we tested the inhibitory effects of DNA containing a THF abasic site mimetic or a 1-nucleotide gap, which resemble the reaction intermediate and product, respectively, on 5mC and T/G excision activity by DME (**Figure 18**). The abasic mimic and gapped DNA products inhibited 5mC excision with  $K_i$  values of  $11.3 \pm 4.3 \mu\text{M}$  and  $12.6 \pm 3.3 \mu\text{M}$ , respectively, which were roughly equivalent to the enzyme concentration used ( $10 \mu\text{M}$ ) and to the strength of competition by 5mC-DNA ( $K_i = 4.3 \pm 2.1 \mu\text{M}$ ) and T/G-DNA ( $K_i = 7.7 \pm 3.3 \mu\text{M}$ ) substrates (**Figure 18**). Similarly, the DNAs inhibited T/G excision with  $K_i$  values of  $2.4 \pm 0.6 \mu\text{M}$  (THF-DNA) and  $6.8 \pm 0.8 \mu\text{M}$  (Gap-DNA) (**Figure 18**), similar to the enzyme concentration

used (2  $\mu\text{M}$ ). Therefore, the DNA products inhibited DME excision of 5mC and T/G to the same extent. Consistent with this,  $K_d$  values measured by fluorescence anisotropy for binding of the catalytically dead D1304N mutant to DNA containing 5mC and T/G were  $0.6 \pm 0.1 \mu\text{M}$  and  $0.2 \pm 0.1 \mu\text{M}$ , respectively (**Figure 19**). In contrast to DNA intermediates and products, we were unable to inhibit 5mC excision by free nucleobases, even at high concentrations (**Figure 18D**).



**Figure 18.** Inhibition of 5mC and T/G excision by reaction intermediates. (A,B) Rate constants for reactions containing 10  $\mu\text{M}$  DME $\Delta\text{N}\Delta\text{IDR1}$ , 100 nM radiolabeled 5mC-DNA, and indicated amounts of cold 25-mer competitor DNA duplex containing a 1-nucleotide gap or tetrahydrofuran (THF) abasic site mimetic (A), or containing 5mC- or T/G substrates (B). (C) Rate constants for T/G excision by 2  $\mu\text{M}$  DME $\Delta\text{N}\Delta\text{IDR1}$  incubated with 100 nM radiolabeled T/G-DNA in the presence of the indicated amounts of Gap- and THF-DNA. (D) 5mC excision activity for 10  $\mu\text{M}$  DME $\Delta\text{N}\Delta\text{IDR1}$  incubated with 100 nM radiolabeled 5mC-DNA in the presence of 16  $\mu\text{M}$  5mC or 10 mM thymine.

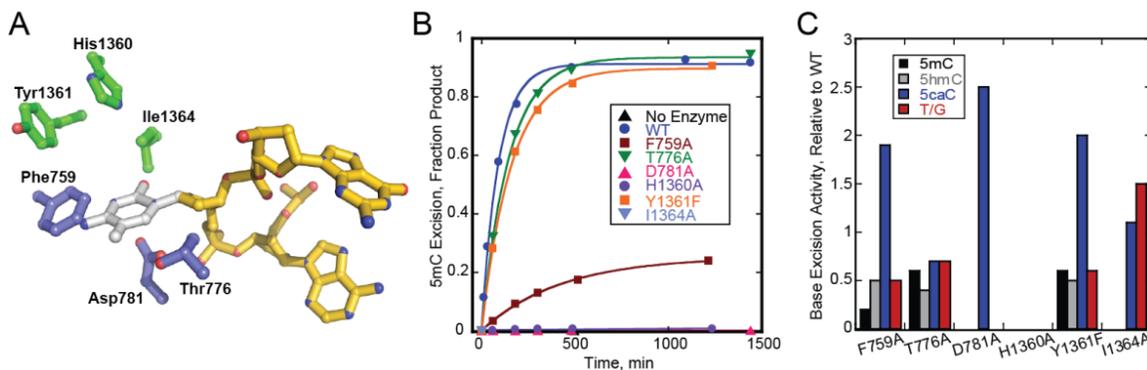


**Figure 19.** DME binding to 5mC- and T/G-DNA. Fluorescence anisotropy binding isotherm for catalytic mutant DME $\Delta$ N $\Delta$ IDR1 D1304N binding to 25mer 5mC-DNA (black circles) and T/G-DNA (red squares). Data was fit using a two-state binding model in GraphPad Prism 6.

*Active site residues confer specificity for substrate DNA*

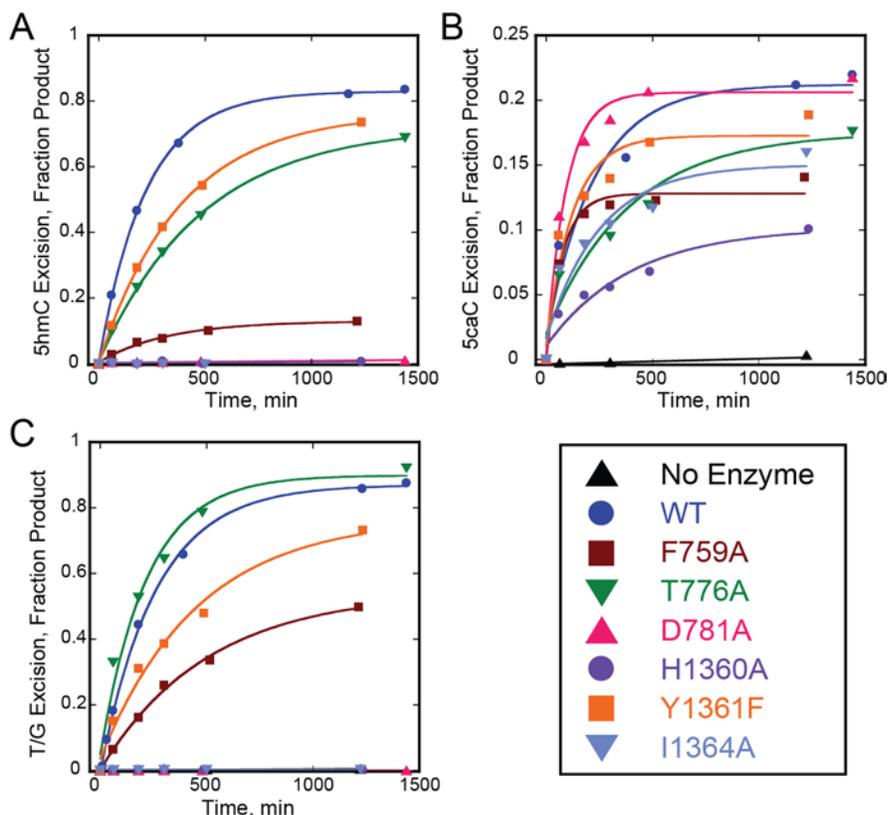
Substrate specificity of the HhH superfamily of DNA glycosylases is largely determined by the shape and chemical complementarity between the nucleobase within the active site pocket (Brooks, Adhikary et al. 2013). The active site of DME predicted by the homology model is formed by residues contributed by both the A and glycosylase domains (**Figure 15B**), similar to that previously shown for ROS1 (Ponferrada-Marin, Parrilla-Doblas et al. 2011). Superimposing an extrahelical 5mC onto our homology model and an examination of other glycosylase-DNA structures, including MBD4 in complex with T/G-DNA (PDB ID 4EVV), led to the identification of six residues lining the DME nucleobase binding pocket: Phe759, Thr776, and Asp781 from Domain A, and His1360, Tyr1361, and Ile1364 from the glycosylase domain (**Figure 20A**). Phe759 is predicted to reside on an unstructured loop and to stack against the N4 nitrogen of 5mC.

Thr776 and Asp781 sit adjacent to the abasic site DNA in the DME homology model and correspond to two MBD4 residues that were observed to make contacts to the flipped thymine substrate in the crystal structures of MBD4 bound to DNA (Manvilla, Maiti et al. 2012; Morera, Grin et al. 2012; Otani, Arita et al. 2013). His1360, Tyr1361, and Ile1364 reside near the predicted location of 5mC in the active site.



**Figure 20.** DME active site mutants affect base excision. (A) Homology model of DME's active site. Protein residues are colored by domain as in **Figure 13/14**. THF-containing DNA (gold) is modeled from the structure of Nth bound to THF-DNA (PDB ID 1P59), and a modeled 5mC is shown in silver. (B) Single-turnover excision rates of 5mC by DME active site mutants compared with wild-type (WT) DME $\Delta$ N $\Delta$ IDR1. (C) Single-turnover rates of 5mC, 5hmC, 5caC, or T/G excision by each DME active site mutant relative to wild-type.

We examined the contribution of each predicted active site residue to the substrate specificity of DME by mutational analysis of 5mC, 5hmC, 5caC, and T/G excision activity using the DME $\Delta$ N $\Delta$ IDR1 construct. All mutants purified well and retained the brilliant yellow color of wild-type DME, indicating proper folding of the proteins. Thr776Ala had only a modest reduction in 5mC excision activity relative to wild-type. In contrast, leucine substitution of the corresponding residue in ROS1 (Thr606) abolished ROS1 activity (Ponferrada-Marin, Parrilla-Doblas et al. 2011). DME activity also differed from ROS1 with respect to Tyr 1361 (Tyr1028 in ROS1). DME Tyr1361Phe



**Figure 21.** 5hmC, 5caC, and T/G excision kinetics. Single-turnover excision rates of 5hmC (A), 5caC (B), and T/G (C) by active site mutants compared with wild-type (WT) DME $\Delta$ N $\Delta$ IDR1.

reduced 5mC and T/G excision activity nearly two-fold, whereas ROS1 Tyr1028Ser reduced 5mC activity two-fold but increased preference for T/G excision (Ponferrada-Marín, Parrilla-Doblas et al. 2011). In contrast to Thr776Ala and Tyr1361Phe, which had only a modest effect on 5mC excision by DME, Phe759Ala significantly reduced 5mC excision and Asp781Ala, His1360Ala, and Ile1364Ala abrogated 5mC excision (**Figure 20B, C**). A similar reduction in 5mC activity was observed for ROS1 Asp611Val, which corresponds to DME Asp781 (Ponferrada-Marín, Parrilla-Doblas et al. 2011). Similarly, DME Phe759Ala showed a higher activity for T/G excision than 5mC excision, consistent with the corresponding residue in ROS1 (Phe589Ala) (Ponferrada-

Marin, Parrilla-Doblas et al. 2011). His1360Ala was the only mutation in DME that abrogated excision of all substrates (**Figure 20C, 21**). Although Ile1364Ala had no activity for 5mC or 5hmC, it retained activity for 5caC and T/G at rates similar to wild-type (**Figure 20C** and **Table 4**). Surprisingly, three mutations (Phe759Ala, Asp781Ala, and Tyr1361Phe) showed a modest increase in 5caC excision rates relative to wild-type, but remained unable to excise more than 20% of the 5caC-DNA substrate (**Figure 20C, 21**). Interestingly, the Asp781Ala substitution showed the highest increase in 5caC activity despite complete loss of activity toward the other nucleobases tested. Taken together, these results validate the homology model, identify His1360, Asp781, and Ile1364 as important for DME activity, and suggest that the various 5mC derivatives may be recognized through architectural changes to the active site pocket.

Mutant	$k_{st}$ 5mC-DNA	Relative to WT	$k_{st}$ 5hmC-DNA	Relative to WT	$k_{st}$ 5caC-DNA	Relative to WT	$k_{st}$ T/G-DNA	Relative to WT
F759A	4.0 ± 0.2	0.2	4.3 ± 2.0	0.5	16 ± 5	1.9	3.2 ± 0.2	0.5
T776A	11.2 ± 0.2	0.6	3.5 ± 0.1	0.4	6.0 ± 1.6	0.7	7.7 ± 0.3	0.7
D781A	ND	0	ND	0	21 ± 4	2.5	ND	0
H1360A	ND	0	ND	0	ND	0	ND	0
Y1361F	10.9 ± 0.4	0.6	4.4 ± 0.3	0.5	17 ± 3	2	3.7 ± 0.2	0.6
I1364A	ND	0	ND	0	7.1 ± 1.7	1.1	10 ± 2	1.5

First-order single-turnover rate constants ( $k_{st}$ ,  $\times 10^{-5} \text{ s}^{-1}$ ) for excision of oxidized 5mC nucleobases from a 25mer oligonucleotide at 25 °C, pH 8.5 and 170 mM ionic strength. Values represent the averages and standard deviations from at least three experiments. ND, no detectable activity.

## Discussion

This work represents the first quantitative structure-function study of DME and DML3, providing a molecular rationale for their similarities and differences to ROS1, which has been extensively characterized (Morales-Ruiz, Ortega-Galisteo et al. 2006; Ponferrada-Marin, Roldan-Arjona et al. 2009; Ponferrada-Marin, Martinez-Macias et al.

2010; Ponferrada-Marin, Parrilla-Doblas et al. 2011; Parrilla-Doblas, Ponferrada-Marin et al. 2013). We previously identified the catalytic core of DME to be composed of domains A and B in addition to the glycosylase domain and found that IDR1 is dispensable for 5mC excision (Mok, Uzawa et al. 2010). Here, using the structure of the Nth-DNA complex as a guide, homology modeling of DME revealed that domain A contributes three  $\alpha$ -helices to the glycosylase domain as previously identified in ROS1 (Ponferrada-Marin, Parrilla-Doblas et al. 2011). The model also predicts that the IDR1 insertion extrudes from the glycosylase domain on the side opposite the DNA binding site, consistent with the fact that deletion of IDR1 had only a modest effect on DME glycosylase activity *in vitro*, explaining why deletion of IDR1 resulted in only a two-fold decrease in DME activity. Moreover, the rate of 5mC excision by DME was comparable to DML3 (**Table 1**) and ROS1 (Ponferrada-Marin, Roldan-Arjona et al. 2009), which contain variable IDR1s in both length and sequence. This large insertion to a glycosylase fold is unique to the DME family, leading us to speculate that these large insertions function to regulate *in vivo* activity, possibly by mediating protein interactions. For example, specific IDR1-protein interactions may facilitate DME activation of imprinted genes at specific loci and the genome wide demethylase activity of ROS1 and DML3 (Gong, Morales-Ruiz et al. 2002; Zhang, Yazaki et al. 2006; Zhu, Kapoor et al. 2007; Zilberman, Gehring et al. 2007).

Our homology model and mutagenesis results indicate that Asn778 and Met1238 are plug and wedge residues, respectively, important for 5mC detection. These two residues were identified to be critical for DME activity from previous random mutagenesis experiments, which were performed *in vivo* in the context of full-length

DME (Mok, Uzawa et al. 2010) (**Figure 15A**). Interestingly, despite the similarity in their predicted structures, DME, ROS1, and DML3 seem to utilize different duplex interrogating residues. All DME family members contain an Asn-Gln moiety predicted to reside on the loop containing the plug residue in other glycosylases (**Figure 16**). Whereas alanine substitution of the Gln had a 10-fold greater effect than the Asn residue on ROS1 activity, we found the opposite to be true for DME. Moreover, alanine substitution of either Asn or Gln abolished DML3 activity. These data suggest that the DNA intercalation loop in each DME paralog engages the DNA differently. Because the residues from this loop have been shown to be important for interrogation of DNA during discrimination of modified versus unmodified nucleobases prior to base flipping (Banerjee, Santos et al. 2006), a different mode of intercalation would likely alter the mechanism by which each enzyme locates its 5mC target. Moreover, this interrogating loop in HhH glycosylases plays a role in recognition of the base opposite the lesion. For example, Nth contacts the opposite base through hydrogen bonds from the backbone carbonyl of Gln42 (Fromme and Verdine 2003), which aligns with DME Asn778. There are no other candidate interactions with the opposing G in our homology model, and thus we do not expect there to be a strong opposite base specificity in DME. Consistent with this, ROS1 excises 5mC more efficiently when mispaired with C, T, or A, which is likely due to the reduced thermodynamic stability of the base pair (Ponferrada-Marin, Roldan-Arjona et al. 2009). Given the location of the plug residues on the A-domain segment of the glycosylase, it is conceivable that the variable IDR1 also influences duplex interrogation by the Asn-Gln loop.

Indeed, despite differences in the identity of the plug residue, both DME and ROS1 utilize a conserved Met residue from the glycosylase domain to wedge between the nucleobases opposite 5mC (Ponferrada-Marin, Parrilla-Doblas et al. 2011). Furthermore, we observed both similarities and differences between the predicted active sites of DME and ROS1. Most notably, the residue corresponding to DME Thr776 is essential to 5mC excision in ROS1 (Ponferrada-Marin, Parrilla-Doblas et al. 2011), but not in DME. Likewise, DME Tyr1361Phe did not cause a change in substrate specificity between 5mC and T/G, although the corresponding mutation in ROS1 did (Ponferrada-Marin, Parrilla-Doblas et al. 2011). Therefore, despite the similarity among the DME family of enzymes, we find subtle differences in substrate recognition by both DNA intercalation and active site residues that may help explain the differences in sites of demethylation in the Arabidopsis genome (Gong, Morales-Ruiz et al. 2002; Zhang, Yazaki et al. 2006; Penterman, Uzawa et al. 2007; Penterman, Zilberman et al. 2007; Zhu, Kapoor et al. 2007; Zilberman, Gehring et al. 2007; Ortega-Galisteo, Morales-Ruiz et al. 2008).

We found the order of substrate preference for DME and DML3 to be 5mC > 5hmC ~ T/G > 5caC > 5fC. However, when comparing the rates of excision, it appeared that DME bound both 5caC and T/G tighter than 5mC and even excised 5caC faster than 5mC at low levels of enzyme. DME was not inhibited by free 5mC or thymine base, and thus the low maximal rate of T/G excision compared to 5mC cannot be explained by product thymine remaining in the active site. In contrast, we found that both 5mC and T/G excision were inhibited in the presence of product-DNA. However, less product-DNA was needed to reduce T/G excision, indicating a higher affinity of DME for 5mC-DNA as compared with T/G-DNA. Therefore, we conclude that the apparent high affinity

of DME for T/G- and 5caC-DNA can be attributed to product DNA inhibition and to the slow turnover of the enzyme for these substrates, as evidenced by their low  $k_{st}$  values (**Figure 17D**). Release of the abasic DNA product is often the rate-limiting step in glycosylase reactions and is promoted by AP-endonuclease, the next enzyme in BER pathway (Waters, Gallinari et al. 1999; Sidorenko, Nevinsky et al. 2007; Fitzgerald and Drohat 2008; Baldwin and O'Brien 2010). Consistent with this interpretation, the Ariza group has shown that ROS1 excises 5mC and binds to the resulting abasic site in a distributive manner (Ponferrada-Marin, Roldan-Arjona et al. 2009).

DNA glycosylases, including those belonging to the helix-hairpin-helix superfamily, are either specific for oxidation or alkylation damage, but not both (Brooks, Adhikary et al. 2013). The ability of the DME family of enzymes to remove both methylated 5mC and oxidized 5hmC makes them unusual. This unique specificity is important in light of the recent progress to understand active demethylation by TDG in mammals (Wu and Zhang 2014). While DME excises 5hmC relatively well compared to 5mC, it does not effectively remove 5fC nor 5caC. In contrast, TDG shows no activity for 5hmC, and has 40% greater activity for 5fC and 25% residual activity for 5caC relative to its preferred T/G substrate. Thus, DME and TDG have reciprocal specificities with respect to the oxidized derivatives. Although there is no evidence to suggest that oxidized 5mC nucleobases are biologically relevant substrates in *Arabidopsis* (Jang, Shin et al. 2014), the ability of DME to remove these bases provides insight into 5mC discrimination inside the binding pocket.

Residues contributed by both domain A (Phe759 and Asp781) and the glycosylase domain (His1360 and Ile1364) are necessary for full 5mC excision activity. These data

are consistent with our previous random mutagenesis screen, in which mutations of Asp781 and His1360 abolished DME activity (Mok, Uzawa et al. 2010). Similarly, the lack of an effect of Thr776Ala or Tyr1361Phe on 5mC excision correlates with the absence of Thr776 or Tyr1361 substitutions from the random screen. Interestingly, Ile1364 abolished 5mC and 5hmC excision, but did not affect activity toward T/G, implicating this residue as an important recognition element of cytosine derivatives. Phe579, located in close proximity to the 5mC methyl substituent, was the only residue to show a difference, albeit modest, between 5mC and 5hmC activity. In contrast, the relatively large distance of His1360 from the flipped base suggests that substitution of this residue affected the folding of the binding pocket. Although these data start to paint a picture of the 5mC active site, it is conceivable that domain B, lying outside of the nucleobase binding pocket, contributes to 5mC recognition, similar to the 8-oxoguanine recognition domain of MutY (Fromme, Banerjee et al. 2004). Indeed, domain B contains residues essential to DME activity (Mok, Uzawa et al. 2010). Further structural analysis will be necessary to elucidate the roles of domains A and B in DME.

## Chapter IV

### DISCUSSION AND FUTURE DIRECTIONS

DNA is a reactive, dynamic molecule that is subject to modification in the form of damage and regulation. The base excision repair (BER) pathway that is necessary for removal of single nucleobase adducts is also necessary for the removal of DNA methylation, thereby linking the regulation of gene expression to the maintenance of genome integrity. While the structures, biochemistry, and mechanisms of the DNA glycosylases that initiate BER have been studied extensively, these enzymes have only recently been implicated in DNA demethylation. Therefore, many questions on the mechanism of DNA demethylation remain unanswered.

#### **DNA Glycosylases Regulate DNA Methylation**

The transfer of a methyl group from *S*-adenosylmethionine (SAM) to position 5 of a cytosine nucleotide is a well characterized reaction. The formation of the C-C bond creates a stable marker for various epigenetic processes. Although the modification is small, its effects are large and numbered. Combined with the modification of histones, 5-methylcytosine (5mC) is responsible for regulating gene expression throughout organism development and cell differentiation to disease progression (see *Epigenetics and Human Health* below). The C-C bond is so thermodynamically stable that it is highly unlikely that there exists a demethylase capable of directly removing the methyl modification. Instead, it has been speculated since at least the early 2000s that a DNA glycosylase is

responsible for the removal of the modified base. Compared to the single enzyme transfer of the methyl group onto the cytosine base, removing the modified base through the BER pathway utilizes at least five separate enzymes to accomplish the 5mC to C conversion. In addition, according to the proposed mammalian DNA demethylation pathway, Tet proteins are necessary to oxidize 5mC before removal by TDG, thereby requiring even more enzymatic steps for the demethylation event. This labor-intensive pathway is indicative of the importance of proper regulation of DNA methylation and its necessity for organism development. Furthermore, the fact that DNA glycosylases have been implicated in the removal of DNA methylation reiterates the significance of 5mC, 5fC, or 5caC recognition by a class of enzymes that are trusted to locate and remove damaged DNA bases amongst the vast amount of undamaged DNA.

DNA glycosylases have been associated with control of gene transcription before the evidence for TDG removal of DNA methylation was found. TDG has been found to bind to hormone receptors, including retinoic acid receptors and estrogen receptor alpha (Um, Harbers et al. 1998; Chen, Lucey et al. 2003), and transcription factors, such as thyroid transcription factor-1 and CBP/p300 acetylase (Missero, Pirro et al. 2001; Tini, Benecke et al. 2002; Chen, Lucey et al. 2003). Interestingly, TDG may function to control CpG sites beyond regulation of DNA methylation. A recent report on the aberrant TDG activity for excising thymine opposite various adenine lesions in a TpG sequence contexts suggests that this activity may introduce TpG, CpA → CpG mutations, thereby stabilizing and extending CG content in CpG-rich promoters (Talhoui, Couve et al. 2014). Therefore, the CpG islands present near the transcription start sites of genes may in fact be a result of aberrant TDG activity. This hypothesis proposes that the DME

family of DNA glycosylases in plants, which also have some mismatch repair activity, allows for increased CG content of plant genomes as well (Talhaoui, Couve et al. 2014). Additional studies are needed to determine if this mechanism is responsible for CpG content in both plants and mammals.

### **DME is a Multi-Domain Glycosylase: the Role of Domain B**

As described in previous work, the three conserved domains of DME are necessary and sufficient for 5mC excision activity. To date, there is little known about the function of domain B, the C-terminal domain of the enzyme. However, many random mutations within domain B cause loss of DME function (Mok, Uzawa et al. 2010). Secondary structure calculations predict a mixture of alpha helices and beta strands in domain B. Unlike the glycosylase and A domain of DME, domain B has no significant sequence similarity to other proteins, preventing the construction of a homology model for this portion of the enzyme. The necessity for domain B, coupled with the lack of sequence similarity to known DNA glycosylases, indicates that this domain may not function during catalysis of the glycosylase reaction. Instead, domain B may be necessary for proper folding and structure of the bipartite glycosylase domain. This hypothesis is supported by an inability to purify recombinant DME lacking domain B.

On the contrary, domain B could be necessary for substrate identification. DME has higher activity for 5mC in a hemimethylated state, as opposed to fully methylated DNA and is inhibited by abasic sites on the strand opposite 5mC (Gehring, Huh et al. 2006). This difference in activity is consistent with other DNA glycosylases, which are inhibited by abasic sites or nicks on the strand opposite a DNA lesion, in order to prevent

detrimental double-strand breaks (Hanai, Yazu et al. 1998; David-Cordonnier, Boiteux et al. 2001; Weinfeld, Rasouli-Nia et al. 2001). Domain B may function to identify the methylation state of the DNA, in order to prevent activity on fully methylated substrates. The effect of abasic site inhibition on DME activity is lost when the site is 7 nucleotides away from the position of 5mC (Gehring, Huh et al. 2006). The homology model of DME predicts that the glycosylase fold is in contact with only the nucleotides flanking the 5mC and opposite G. Thus, domain B may be positioned to interrogate the duplex farther in either the 3' or 5' direction of the substrate 5mC. Domain B is predicted to be nearly 32 kDa, but is not likely to be a separate structural domain, since efforts to purify it independently of domain A and the glycosylase domain have failed (see Appendix II). The overall structure of DME is not predicted to be a modular structure of separately folded and independent domains. Instead, all indications point to a single globular module composed of the three interdependent domains of DME. Further conclusions await structural data.

### **DME Paralogs Have Distinct Functions**

As described previously, there are four paralogs of DME in *A. thaliana*: DME, ROS1, DML2, and DML3. The sequences of domain A, the glycosylase domain, and domain B are very well conserved among the four paralogs. However, the length and sequence content of the N-terminal residues preceding domain A of each of the paralogs are not similar, nor are the sequences of the two inter-domain regions, IDR1 and IDR2. These unconserved regions may be responsible for the variation in activities by these enzymes.

While DME functions to activate transcription of specific imprinted genes, the DML enzymes, including ROS1, function genome wide, at sites 5' and 3' to genes to regulate transcription, protecting plants from erroneous gene silencing (Gong, Morales-Ruiz et al. 2002; Zhang, Yazaki et al. 2006; Zhu, Kapoor et al. 2007; Zilberman, Gehring et al. 2007). An additional distinction among the paralogs is that although DME, ROS1, and DML3 excise 5mC in CpG, CpNpG, and CpNpN contexts, DML2 is only able to excise 5mC in a CpG context (Gehring, Huh et al. 2006; Morales-Ruiz, Ortega-Galisteo et al. 2006; Penterman, Zilberman et al. 2007). Although the majority of the residues identified as critical for DME function in a random mutagenesis screen are contained within the three conserved domains, one residue was found to be necessary for 5mC excision activity in each of the two IDRs (Mok, Uzawa et al. 2010). It is plausible that these regions may facilitate protein-protein interactions that regulate the specific recruitment of each of the DME-like enzymes, which are expressed in adult cells, unlike DME, which is expressed in the central and vegetative cells. The IDRs contain some predicted alpha helices and beta strands and therefore are more likely to have functional interactions with regulatory proteins than the N-terminal residues preceding domain A that have very little predicted secondary structure (Mok, Uzawa et al. 2010).

Although the three domains are well conserved among the DME paralogs, some differences in sequence within these domains and consequent 5mC excision ability contribute to the differentiation of these enzymes. It is of particular importance to note the variable interrogation and active site residues in the DME family of enzymes. The interrogation residues of DNA glycosylases have been implicated as not only necessary for stabilization of the duplex during catalysis, but also as important detectors of DNA

damage and substrate identification (Stivers and Jiang 2003; Banerjee, Santos et al. 2006; Adhikary and Eichman 2011). As described in Chapter III, the wedge residues identified in DME, ROS1, and DML3 are identical, while there are discrepancies among the plug residues. Additional differences are found for the active site residues that discriminate substrate DNA, which confer specificity for 5mC over Thy.

The investigation of the active site residues and interrogation residues of the DME family contribute to the study of the mechanism of DNA glycosylases in general. An outstanding problem in the field is a lack of understanding of the ability of DNA glycosylases to locate and maintain specificity for DNA lesions from the 700,000-fold abundance of unmodified DNA. This problem extends to the glycosylases responsible for initiating DNA demethylation, since 5mC constitutes only one percent of human DNA. This problem is compounded by the fact that not all cytosine methylation needs to be removed, unlike the necessity of removing harmful DNA lesions. Thus, it is likely that cells maintain a complex pattern of DNA methylation through the use of transcriptional repressors that bind methylated DNA when genes are silenced, thereby excluding DNA glycosylases that could function to activate the genes.

### **DME Unstructured Regions of Unknown Function**

The DME family of enzymes are very large proteins: DME exceeds 1700 amino acid residues. However, over half of the protein sequence is not conserved among DME homologs. The N-terminal 600 residues and the two inter-domain regions (IDR1 and IDR2) have no significant sequence similarity among homologs, nor are these regions predicted to have substantial secondary structure. Because no splice variants of DME

have been found, it is likely that the full protein is expressed in plant cells. However, it is possible that these regions are subject to protease degradation *in vivo*, similar to the degradation of IDR1 observed *in vitro*. If this is the case, it is expected that the functional DME protein *in vivo* would resemble the DME $\Delta$ N $\Delta$ IDR1 construct used in biochemical experiments in this thesis.

Alternatively, the regions of low sequence similarity and predicted structural complexity may function to regulate DME activity. These regions may interact with other proteins to signal the need for DNA demethylation. For example, the N-terminal region or the IDR1 of DME may interact with histone modifications to signal when DME is needed to remove methylation. Because IDR1 is predicted to extend from the core glycosylase fold formed from both domain A and the conserved glycosylase domain of DME, it is plausible that these interactions may trigger a conformational change in DME, allowing for enzymatic function. Other protein interactions with these regions may lead to the recruitment of DME to sites of DNA methylation. However, there are no established proteins that recruit other DNA glycosylases to sites of damage in need of repair. Instead, it is theorized that the DNA glycosylases themselves are responsible for locating modified nucleobases. The method of DME localization to sites of DNA methylation may function differently than other DNA glycosylases. This is to be expected, since unlike damaged nucleobases that could lead to disease or cell death, not all DNA methylation needs to be removed. Instead, the processes of DNA methylation and demethylation are precisely timed for proper development and cell differentiation. Therefore, it is quite reasonable to expect that the regions of low sequence similarity in these enzymes may play a role to regulate the differential function of the DME homologs.

## **Role of the Iron-Sulfur Cluster**

Many DNA repair proteins contain iron-sulfur clusters, including EndoIII/Nei, MutY, and DME. The role of the [4Fe-4S] iron-sulfur cluster in EndoIII has long been debated (Cunningham, Asahara et al. 1989). In 1992, it was determined that the iron-sulfur cluster of EndoIII is redox-inert and therefore would not cause damage to DNA (Fu, O'Handley et al. 1992). Evidence from crystal structures of EndoIII and MutY indicate that one loop of the cluster presents positively charged residues to the DNA backbone, however these residues are not well conserved (Thayer, Ahern et al. 1995; Guan, Manuel et al. 1998; Chepanoske, Golinelli et al. 2000; Fromme and Verdine 2003). Today, the role of the iron-sulfur cluster is debated, as many recent reports indicate that it may facilitate lesion searching, as opposed to functioning purely as a structural component of the proteins.

Although the role of iron-sulfur clusters in repair is not fully understood, many suspect that repair enzymes may utilize changes in the electrochemistry of damaged DNA to sense lesions (Fan, Fuss et al. 2008; Netz, Pierik et al. 2012; White and Dillingham 2012). The structure of DNA allows for charge transport due to the overlapping  $\pi$  orbitals of the stacked nitrogenous bases (Wagenknecht 2005; Genereux and Barton 2010; Genereux, Boal et al. 2010). Undamaged, stacked DNA has the same conductivity as graphite (Gup, Gorodetsky et al. 2008), and charge transport can occur over hundreds of base pairs on a picosecond time scale (Wagenknecht 2005). DNA charge transport chemistry distinguishes between well-matched base pairs and those that are mismatched or contain lesions, due to changes in conductivity.

Iron-sulfur clusters are a common redox cofactor for bioinorganic chemistry across the kingdoms of life (Sontz, Muren et al. 2012). The DNA-bound potentials of EndoIII and MutY are ~80 mV versus NHE, which is within the range for a biological redox switch (Boal, Genereux et al. 2009; Genereux, Boal et al. 2010). Binding of DNA by these proteins causes a large negative shift in potential (Boal, Genereux et al. 2009; Genereux, Boal et al. 2010). A proposed model for these proteins to locate sites of DNA damage begins with this shift in potential when the enzyme binds DNA. The protein may be activated toward oxidation when in the vicinity of DNA and would then release an electron that travels along the length of unmodified DNA to another protein with similar DNA-bound potential, which would be reduced by the electron and subsequently dissociated from the DNA. In this way, the proteins are scanning the DNA for lesions, and if the electron transfer is disrupted by an unmatched base pair, the second protein will not dissociate, but will instead remain bound and continue to search along the DNA for a site of damage (Sontz, Muren et al. 2012).

Although not investigated in this work, the role of the iron-sulfur cluster in DME presents an interesting perspective on the hypothesis of disruption of DNA charge transport as a means to lesion detection. Since 5mC does not disrupt the base stacking or base pairing of the G/C base pair, it is unlikely that charge transport is changed along the DNA. Much of the work on iron-sulfur clusters have focused on EndoIII, Nei, and MutY, which repair oxidized pyrimidines that do change the electrochemical environment of the DNA duplex (David and Williams 1998; Barton, Olmon et al. 2011; Romano, Sontz et al. 2011; Muren, Olmon et al. 2012; Sontz, Muren et al. 2012). While the currently accepted model of damage location is 3D diffusion and short-distance

sliding along the DNA, it is possible the iron-sulfur cluster of some repair proteins may utilize charge transport in addition to diffusion and sliding (Zharkov and Grollman 2005; Friedman and Stivers 2010).

### **Epigenetics and Human Health**

The status of DNA methylation on certain gene promoters plays a large part in determining disease progression for developmental abnormalities, cancer, and neurodegenerative diseases (See *Dysregulation of Methylation Leads to Disease* in Chapter I) (Robertson 2005; Jakovcevski and Akbarian 2012). In this way, methylation patterns can be used in personalized medicine to determine treatment options and disease progression. Additional studies on the methylation patterns of various diseases will be needed in order to fully utilize the epigenome of each patient. On the other hand, epigenetics can affect human health beyond disease progression, since people are consumers of plants and animals that also utilize DNA methylation to control gene expression. As described in Chapter I, laboratories are seeking to engineer wheat varieties that lack certain proteins that induce immune responses in patients with celiac disease or other gluten intolerances. This idea may be extended to other plants containing DME family enzymes. If the methylation state of particular genes can be turned on or off, many new modified plants could be selected for specific traits.

### **State of the Field and Future Directions**

An increased understanding of the plant DME glycosylases has implications beyond the possibility of engineering wheat varieties with different protein content.

Several questions remain outstanding in the field of DNA glycosylases. In particular, the question of substrate recognition is particularly interesting with respect to 5mC, since it is structurally very similar to Cyt and Thy. Although the model predicted through sequence homology provided some insight into the active site of DME, a crystal structure of DME in complex with 5mC-DNA would provide details on the specific protein-DNA contacts that distinguish 5mC from other substrates. Domain B may play an important role in stabilizing the structure of the enzyme alone or in complex with its substrate DNA. The question of the role of domain B in substrate recognition would be answered with additional structural information that cannot be elucidated through homology modeling.

The results of this body of work focus on the ability of DME to utilize a DNA repair mechanism to remove a non-toxic and non-mutagenic base. Through base excision repair, DME is able to directly remove 5mC. Current studies indicate that animals require oxidation of 5mC to 5fC and 5caC before the base can be efficiently removed by a DNA glycosylase. Since there is no known mammalian 5mC glycosylase, DME glycosylases are an interesting distinction between plant and animal development. The control of DNA methylation throughout the development and life cycle of a plant is very complex. Yet, these processes in animals require additional levels of controls through the TET proteins that oxidize 5mC. Furthermore, the method of TDG removal of 5fC and 5caC across from Gua may proceed through recognition of a wobble base pair, which is not likely formed between 5mC and Gua. Therefore, the recognition of 5mC by DME is likely due to different contacts between the enzyme and substrate DNA than those formed through TDG recognition of 5fC and 5caC.

The work presented here can be furthered through continuing crystallization attempts of DME in complex with DNA. The purification methods and buffer conditions have been optimized for glycosylase activity and DNA binding throughout the course of this project. A high quality crystal of DME may be more readily obtainable using the improved method of purification as described in Chapter III. In addition, the effects of mutations and deletion constructs on DME activity tested *in vitro* in this work may be extended to *in vivo* studies to explore the hypotheses on the function of the unstructured regions and their effects on activity of the DME homologs. Studies of the binding partners of DME *in vivo* also would be enlightening and may lead to further insight on the function of the unstructured regions and/or Domain B. There remains much to be understood on the function and structure of DME and its role in epigenetics.

## Appendix I

### STABILITY OF 5-METHYLCYTOSINE DERIVATIVES

#### *Introduction*

In order to understand the mechanism of damage recognition and removal by DNA glycosylases, the inherent stability of the DNA duplex is often investigated (Brown, Hunter et al. 1986; Hunter, Kneale et al. 1986; Hunter, Brown et al. 1987). Glycosylases utilize DNA interrogation residues to probe the substrate duplex for changes to the DNA structure. Once a modification or damage is found, the glycosylase flips the DNA base into an active site pocket, where catalysis occurs. Often, there are specific interactions between the enzyme and the base in the active site pocket that allow for and/or exclude modifications, thereby conferring specificity to each DNA glycosylase (Scharer and Jiricny 2001; Stivers and Jiang 2003; Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; David, O'Shea et al. 2007; Dalhus, Laerdahl et al. 2009; Friedman and Stivers 2010; Li 2010; Zharkov, Mechetin et al. 2010; Rubinson and Eichman 2012). Due to the increasing interest in 5mC oxidation derivatives and their role in DNA demethylation pathways, the inherent stability of each of the derivatives in the context of duplex DNA was investigated in this ongoing work. Results of previous, unpublished work by Szulik and Stone on 5mC and 5hmC were extrapolated to 5fC and 5caC derivatives. Calorimetric measurements combined with NMR studies and x-ray crystallography were used to compare the 5mC derivatives with unmodified DNA. Calorimetric measurements revealed similar thermodynamic stabilities for each of the derivatives. However, NMR

measurements indicated that the 5caC modification increased stability over the unmodified duplex, while the other modifications did not affect stability.

### *Methods*

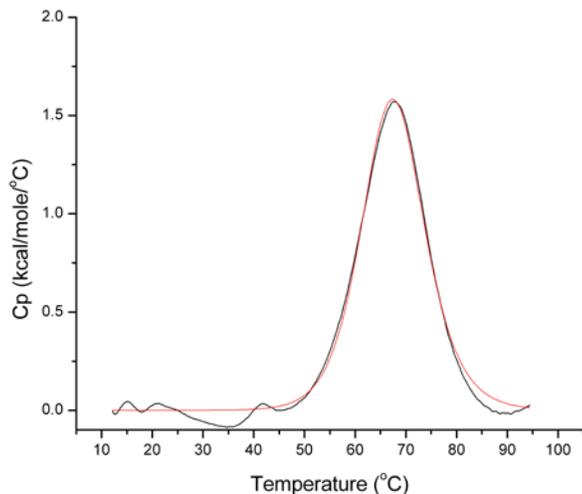
Oligonucleotides were synthesized by Midland Certified Reagents with the self-complementary Dickerson-Drew dodecamer (DDD) sequence d(5' - CGCGAATTXGCG), where X is 5fC or 5caC. The samples were desalted in the Stone lab by gel-permeation chromatography using a Sephadex G-25 column (GE Healthcare) and were lyophilized to dryness. The samples were subsequently characterized by MALDI-TOF-MS. The dry oligomers were annealed by dissolving the single-stranded oligonucleotides in buffer (10 mM sodium phosphate pH 7 and 10 mM or 100 mM NaCl), heating the solution up to 90 °C for 10 min, and allowing it to cool down slowly to room temperature for a final concentration of 100 μM duplex DNA. Calorimetry experiments were performed using a VP-DSC differential scanning calorimeter (DSC, Microcal, Inc., Northampton, MA) from 10 to 95 °C at a rate of 1 °C per minute. The reference cell was filled with buffer containing either 10 mM or 100 mM NaCl, according to the experimental condition.

Crystallization screens were set up using the Nucleic Acid Mini Screen (Hampton Research, Aliso Viejo, CA). The 24 reagents were combined in a sitting drop tray with either 5fC- or 5caC-DDD in a 1:1 ratio, for a final DNA duplex concentration of 0.6 mM. Reservoirs were filled with 500 μL of 10% 2-methyl-2,4-pentanediol (MPD), and crystals were grown at 18 °C for at least two weeks. Crystals were harvested, flash-frozen in

liquid N<sub>2</sub>, and sent to LS-CAT (Argonne National Laboratory, Lemont, IL) for data collection.

### Results

Thermal melts of the Dickerson-Drew dodecamer containing 5fC and 5caC were performed to determine differences in thermodynamic stability between the two modifications and to compare to the stability of 5hmC- and 5mC-containing DNA. The melting temperature of each oligonucleotide as determined by calorimetry was found to be 67 to 68 °C (**Table 5**). A representative thermal melt of 5fC-DDD is shown in **Figure 22**.



**Figure 22.** Representative thermal melt of 5fC-DDD with 10 mM sodium phosphate pH 7 and 100 mM NaCl.

<b>Table 5.</b> Thermal denaturation of 5mC oxidation derivatives in DDD.		
		T <sub>M</sub> °C
A	Cyt	67
	5mC	68
	5hmC	67
	5fC	68
	5caC	67
Melting temperatures determined via DSC thermal melts of 100 μM duplex DNA.		

crystal structure of the Dickerson-dodecamer containing 5hmC was determined by Szulik, et al., at a resolution of 1 Å and  $R_{\text{work}} / R_{\text{free}}$  of 15% / 17% (99.6% / 98.5% completeness) (PDB ID 4I9V). Like 5mC, the 5hmC modification does not alter base stacking or pairing to the opposite G (Renciuk, Blacque et al. 2013). Crystal trials were conducted using the 5fC and 5caC oligonucleotides in the DDD sequence context, in order to determine if any structural variances may elucidate the differences in recognition by DNA glycosylases. Crystals of 5fC-DDD were obtained in conditions including 10% MPD, 40 mM sodium cacodylate trihydrate pH 5.5 – 6.0, 20 mM hexamine cobalt (III) chloride, 40 mM lithium chloride, and 20 mM magnesium chloride hexahydrate. Similar conditions were found for 5caC-DDD crystals, except with 12 mM spermine tetrahydrochloride instead of hexamine cobalt (III) chloride. Although many crystals were obtained from these screens, none diffracted to more than 3.2 Å, thereby precluding molecular replacement solution with unmodified DDD structures.

### *Discussion*

The calorimetric measurements of  $T_M$  for each 5mC oxidation derivative indicate that the modifications do not affect stability of the DNA duplex. The cytosine C5 substituent does not interfere with the G-C hydrogen bonding or base stacking to neighboring bases. These results indicate that the oxidation derivatives do not have significant electron-withdrawing effects that weaken or strengthen the base interactions. Since the oxidation derivatives of 5mC do not change the inherent stability of the duplex, it is unlikely that DNA glycosylases, such as DME and TDG, are able to sense a perturbation in duplex stability in order to locate the DNA modification. Instead, the

DNA glycosylase may require a specific binding pocket interaction to recognize its preferred substrate. Structures of TDG bound to substrate DNA discussed in previous chapters support this hypothesis, since the enzyme makes specific contacts to the extruded base.

Although the calorimetric melting temperatures determined by DSC were very similar for all oxidation derivatives of 5mC, NMR experiments conducted by the Stone lab at varying temperatures indicated that the 5-carboxyl modification *increased* stability of the duplex compared to 5mC, 5hmC, and 5fC. Monitoring the imino protons of the modified C-G base pair revealed that this interaction is lost between 46 to 48 °C for unmodified DDD (48 °C), 5mC- (47 °C), 5hmC- (47 °C) and 5fC-DNA (46 °C). Surprisingly, the melting temperature for 5caC was determined to be 62 °C (unpublished results from M. Szulik and M. Stone). It is unclear what the cause of this discrepancy is, but we speculate that the difference in melting temperatures between the two methods may result from a mixture of hairpin and duplex. Due to the high DNA strand concentration, repeated annealing of the samples during the DSC experiments, and high salt concentration (100 mM NaCl), the DNA forms a bimolecular duplex. However, at the lower strand concentrations used for NMR stability studies, the DNA may form an intramolecular hairpin structure, which may be more affected by modifications to the DNA bases. Because the structure of the DNA was not controlled for in the NMR experiments, this explanation cannot be ruled out.

Crystal structures recently have been determined for the formyl- and carboxyl-cytosine modifications by M. Szulik (unpublished results). These structures indicate that none of the modifications at the C5-position cytosine in the DDD sequence context cause

structural perturbations. Each of the base pairs between the modified nucleobase and the opposite guanine exhibits an unchanged hydrogen bonding interface, and base stacking throughout the duplex is not disrupted. This lack of perturbation of the duplex structure is consistent with the calorimetric results of this study indicating very similar thermodynamic stabilities among the derivatives. Additional experiments that may be performed to elucidate the differences between the oxidized 5mC derivatives include continuing calorimetric studies in the presence of increasing amounts of ethylene glycol to determine the number of waters associated with each modified oligonucleotide. Furthermore, the C5 cytosine modifications may be placed within different sequence contexts, in order to determine any changes to DNA stability due to neighbor base interactions.

## Appendix II

Table A1. DME, DML3, and DMLa Constructs and Solubility					
Construct	Description	pI <sup>†</sup>	MW (kDa) <sup>‡</sup>	Evaluation	
pLM302-DMEΔ677	N-terminal unstructured region truncated before beginning of Domain A	5.53	119	Very little expression, highly insoluble and does not purify completely; Many truncation products due to protease activity	
pBG100-DMEΔ677ΔIDR1	12 amino acid linker replaces IDR1	6.31	76	Test expression revealed expressed protein is insoluble	
pBG102-DMEΔ677ΔIDR1	12 amino acid linker replaces IDR1	6.31	76	SUMO tag increases solubility over 6x-His tag (pBG100); Expression levels low, but can be extracted from lysate to pure protein	
pLM302-DMEΔ677ΔIDR1	12 amino acid linker replaces IDR1	6.31	76	MBP tag greatly increases solubility over SUMO (pBG102) or 6x-His tag (pBG100); Expression levels high and can be extracted from lysate to yield very pure protein	
pBG102-DMEΔ1190ΔIDR1	Missing entire A domain; 12 amino acid linker replaces IDR1	5.89	61	Very poor solubility and tag is largely overexpressed while DME is not.	
pLM302-DMEΔ677Δ(797-1216)	Domain A fused to glycosylase domain without 12 amino acid linker	6.43	72	Did not express well	

pLM302-DMEΔ736ΔIDR1	N-terminus of Domain A truncated; 12 amino acid linker replaces IDR1	6.42	69	Did not express well; No yellow protein (indicating improper folding)
pLM302-DMEΔ736Δ(797-1216)	N-terminus of Domain A truncated; Domain A fused to glycosylase domain without 12 amino acid linker	6.55	65	Did not express well; No yellow protein (indicating improper folding)
pLM302-DMEΔ1512	Isolated Domain B	9.12	25	Insoluble/aggregates when cleaved from MBP tag
pLM302-DMEΔ677ΔIDR1::1401	Domain A and Glycosylase domain with 12 amino acid linker replaces IDR1; No Domain B – due to stop codon at 1401	9.00	39	Did not express well
pLM302-DMEΔ677::1401	Domain A and Glycosylase domain; No Domain B – due to stop codon at 1401	5.79	82	Did not express well
pFB1-DMEΔ677ΔIDR1-TEV6xHis	N-terminal truncation; 12 amino acid linker replaces IDR1	6.31	76	Insect cell expression does not improve expression
pFB HT A-DMEΔ677ΔIDR1	N-terminal truncation; 12 amino acid linker replaces IDR1	6.31	76	Insect cell expression does not improve expression
pFB1-DMEΔ677-TEV6xHis	N-terminal truncation; 12 amino acid linker replaces IDR1	5.53	119	Insect cell expression does not improve expression
pFB HT A-DMEΔ677	N-terminal truncation; 12 amino acid linker replaces IDR1	5.53	119	Insect cell expression does not improve expression
pMAL-DMEΔ677ΔIDR1	N-terminal truncation; 12 amino acid linker replaces IDR1	6.31	76	pMAL vector has lower expression than pLM302; Protein does not purify completely

pBG102-DML3Δ394	N-terminal unstructured region truncated before beginning of Domain A	6.6	82	Expresses and purifies completely; Activity assay reveals protein is not stable at 25 °C for over 3 hours
pLM302-DMLa(673-1207)	Glycosylase domain and domain B	7.07	60	Did not express (overexpression of <i>E. coli</i> proteins)
pLM302-DMLa(927-1207)	Isolated domain B	5.36	32	Did not express
pLM302-DMLa(673-888)	Isolated glycosylase domain	9.41	24.5	Did not express
pLM302-DMLa(407-1207)	N-terminal deletion 37 residues upstream of domain A (3 predicted $\alpha$ helices not homologous to DME)	6.45	91	Expressed, soluble; can be extracted from lysate to yield 80% pure protein
pLM302-DMLa(444-1207)	N-terminal deletion immediately before domain A	6.41	87	Expressed, but insoluble
<p><sup>†</sup>pI, isoelectric point of the construct after cleavage; <sup>‡</sup>MW, molecular weight of the construct in kDa after cleavage; Protein purification evaluated after test expression in various cell lines to determine solubility, *DMLa cloned from <i>Oryza sativa DMLa</i></p> <p>Expression vectors: pLM302: N-terminal 6xHis + Maltose Binding Protein (MBP) tag, 3C protease cleavable, Kan resistant, pET27 derivative with <i>T7lac</i> promoter; pBG102: N-terminal 6xHis + SUMO tag, 3C protease cleavable, Kan resistant, pET27 derivative with <i>T7lac</i> promoter; pBG100: N-terminal 6xHis, 3C protease cleavable, Kan resistant, pET27 derivative with <i>T7lac</i> promoter; pFBI (pFASTBAC1): Transfer vector (to <i>E. coli</i> from expression bacmid) with polyhedrin promoter, adds SV40 polyA site; amp resistance (gentamycin for selection in insect cells); restriction enzyme cloning; pFB HT A: Transfer vector (to <i>E. coli</i> from expression bacmid) with polyhedrin promoter, adds SV40 polyA site; amp resistance (gentamycin for selection in insect cells); restriction enzyme cloning; pMAL: N-terminal MBP tag, Amp resistant</p>				

## REFERENCES

- Aamodt, R. M., P. O. Falnes, et al. (2004). "The *Bacillus subtilis* counterpart of the mammalian 3-methyladenine DNA glycosylase has hypoxanthine and 1,N<sup>6</sup>-ethenoadenine as preferred substrates." *J Biol Chem* **279**(14): 13601-13606.
- Adhikary, S. and B. F. Eichman (2011). "Analysis of substrate specificity of *Schizosaccharomyces pombe* Mag1 alkylpurine DNA glycosylase." *EMBO Rep* **12**(12): 1286-1292.
- Admiraal, S. J. and P. J. O'Brien (2010). "N-glycosyl bond formation catalyzed by human alkyladenine DNA glycosylase." *Biochemistry* **49**(42): 9024-9026.
- Agius, F., A. Kapoor, et al. (2006). "Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation." *Proc Natl Acad Sci U S A* **103**(31): 11796-11801.
- Aller, P., M. A. Rould, et al. (2007). "A structural rationale for stalling of a replicative DNA polymerase at the most common oxidative thymine lesion, thymine glycol." *Proc Natl Acad Sci U S A* **104**(3): 814-818.
- Alseth, I., F. Osman, et al. (2005). "Biochemical characterization and DNA repair pathway interactions of Mag1-mediated base excision repair in *Schizosaccharomyces pombe*." *Nucleic Acids Res* **33**(3): 1123-1131.
- Alseth, I., T. Rognes, et al. (2006). "A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD." *Mol Microbiol* **59**(5): 1602-1609.
- Andreuzza, S., J. Li, et al. (2010). "DNA LIGASE I exerts a maternal effect on seed development in *Arabidopsis thaliana*." *Development* **137**(1): 73-81.
- Anway, M. D., A. S. Cupp, et al. (2005). "Epigenetic transgenerational actions of endocrine disruptors and male fertility." *Science* **308**(5727): 1466-1469.
- Aufsatz, W., M. F. Mette, et al. (2004). "The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides." *Plant Mol Biol* **54**(6): 793-804.
- Ayyanathan, K., M. S. Lechner, et al. (2003). "Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation." *Genes Dev* **17**(15): 1855-1869.
- Baba, D., N. Maita, et al. (2005). "Crystal structure of thymine DNA glycosylase conjugated to SUMO-1." *Nature* **435**(7044): 979-982.
- Baba, D., N. Maita, et al. (2006). "Crystal structure of SUMO-3-modified thymine-DNA glycosylase." *J Mol Biol* **359**(1): 137-147.
- Baccarelli, A., I. Martinelli, et al. (2008). "Exposure to particulate air pollution and risk of deep vein thrombosis." *Arch Intern Med* **168**(9): 920-927.
- Baccarelli, A., M. Rienstra, et al. (2010). "Cardiovascular epigenetics: basic concepts and results from animal and human studies." *Circ Cardiovasc Genet* **3**(6): 567-573.
- Baldwin, M. R. and P. J. O'Brien (2009). "Human AP endonuclease 1 stimulates multiple-turnover base excision by alkyladenine DNA glycosylase." *Biochemistry* **48**(25): 6022-6033.

- Baldwin, M. R. and P. J. O'Brien (2010). "Nonspecific DNA binding and coordination of the first two steps of base excision repair." Biochemistry **49**(36): 7879-7891.
- Ballestar, E. and M. Esteller (2005). "Methyl-CpG-binding proteins in cancer: blaming the DNA methylation messenger." Biochem Cell Biol **83**(3): 374-384.
- Banerjee, A., W. L. Santos, et al. (2006). "Structure of a DNA glycosylase searching for lesions." Science **311**(5764): 1153-1157.
- Barbin, A. (2000). "Etheno-adduct-forming chemicals: from mutagenicity testing to tumor mutation spectra." Mutat Res **462**(2-3): 55-69.
- Barlow, D. P. (2011). "Genomic imprinting: a mammalian epigenetic discovery model." Annual Rev Genet **45**: 379-403.
- Barlow, D. P., R. Stoger, et al. (1991). "The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus." Nature **349**(6304): 84-87.
- Barrett, T. E., R. Savva, et al. (1998). "Structure of a DNA base-excision product resembling a cisplatin inter-strand adduct." Nat Struct Biol **5**(8): 697-701.
- Barrett, T. E., R. Savva, et al. (1998). "Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions." Cell **92**(1): 117-129.
- Barrett, T. E., O. D. Scharer, et al. (1999). "Crystal structure of a thwarted mismatch glycosylase DNA repair complex." EMBO J **18**(23): 6599-6609.
- Bartee, L., F. Malagnac, et al. (2001). "Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene." Genes Dev **15**(14): 1753-1758.
- Bartolomei, M. S., A. L. Webber, et al. (1993). "Epigenetic mechanisms underlying the imprinting of the mouse H19 gene." Genes Dev **7**(9): 1663-1673.
- Bartolomei, M. S., S. Zemel, et al. (1991). "Parental imprinting of the mouse H19 gene." Nature **351**(6322): 153-155.
- Barton, J. K., E. D. Olmon, et al. (2011). "Metal Complexes for DNA-Mediated Charge Transport." Coord Chem Rev **255**(7-8): 619-634.
- Barton, S. C., M. A. Surani, et al. (1984). "Role of paternal and maternal genomes in mouse development." Nature **311**(5984): 374-376.
- Baylin, S. B., M. Esteller, et al. (2001). "Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer." Hum Mol Genet **10**(7): 687-692.
- Begley, T. J. and R. P. Cunningham (1999). "Methanobacterium thermoformicicum thymine DNA mismatch glycosylase: conversion of an N-glycosylase to an AP lyase." Protein Eng **12**(4): 333-340.
- Begley, T. J., B. J. Haas, et al. (1999). "A new member of the endonuclease III family of DNA repair enzymes that removes methylated purines from DNA." Curr Biol **9**(12): 653-656.
- Bell, A. C. and G. Felsenfeld (2000). "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene." Nature **405**(6785): 482-485.
- Bell, A. C., A. G. West, et al. (1999). "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators." Cell **98**(3): 387-396.

- Bennett, M. T., M. T. Rodgers, et al. (2006). "Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability." J Am Chem Soc **128**(38): 12510-12519.
- Berdal, K. G., M. Bjørås, et al. (1990). "Cloning and expression in *Escherichia coli* of a gene for an alkylbase DNA glycosylase from *Saccharomyces cerevisiae*; a homologue to the bacterial alkA gene." EMBO J **9**(13): 4563-4568.
- Berdal, K. G., R. F. Johansen, et al. (1998). "Release of normal bases from intact DNA by a native DNA repair enzyme." EMBO J **17**(2): 363-367.
- Bestor, T., A. Laudano, et al. (1988). "Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases." J Mol Biol **203**(4): 971-983.
- Bestor, T. H. and G. L. Verdine (1994). "DNA methyltransferases." Curr Opin Cell Biol **6**(3): 380-389.
- Bird, A. (2002). "DNA methylation patterns and epigenetic memory." Genes Dev **16**(1): 6-21.
- Bird, A. P. and A. P. Wolffe (1999). "Methylation-induced repression--belts, braces, and chromatin." Cell **99**(5): 451-454.
- Birkeland, N. K., H. Anensen, et al. (2002). "Methylpurine DNA glycosylase of the hyperthermophilic archaeon *Archaeoglobus fulgidus*." Biochemistry **41**(42): 12697-12705.
- Birve, A., A. K. Sengupta, et al. (2001). "Su(z)12, a novel *Drosophila* Polycomb group gene that is conserved in vertebrates and plants." Development **128**(17): 3371-3379.
- Bjelland, S., N. K. Birkeland, et al. (1994). "DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the AlkA enzyme in *Escherichia coli*." J Biol Chem **269**(48): 30489-30495.
- Bjelland, S., M. Bjørås, et al. (1993). "Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of *Escherichia coli*." Nucleic Acids Res **21**(9): 2045-2049.
- Bjørås, M., A. Klungland, et al. (1995). "Purification and properties of the alkylation repair DNA glycosylase encoded the MAG gene from *Saccharomyces cerevisiae*." Biochemistry **34**(14): 4577-4582.
- Boal, A. K., J. C. Genereux, et al. (2009). "Redox signaling between DNA repair proteins for efficient lesion detection." Proc Natl Acad Sci U S A **106**(36): 15237-15242.
- Boiteux, S., O. Huisman, et al. (1984). "3-Methyladenine residues in DNA induce the SOS function *sfiA* in *Escherichia coli*." EMBO J **3**(11): 2569-2573.
- Bollati, V., A. Baccarelli, et al. (2007). "Changes in DNA methylation patterns in subjects exposed to low-dose benzene." Cancer Res **67**(3): 876-880.
- Boorstein, R. J. and G. W. Teebor (1988). "Mutagenicity of 5-hydroxymethyl-2'-deoxyuridine to Chinese hamster cells." Cancer Res **48**(19): 5466-5470.
- Borst, P. and R. Sabatini (2008). "Base J: discovery, biosynthesis, and possible functions." Annu Rev Microbiol **62**: 235-251.

- Bossdorf, O., C. L. Richards, et al. (2008). "Epigenetics for ecologists." Ecol Lett **11**(2): 106-115.
- Bostick, M., J. K. Kim, et al. (2007). "UHRF1 plays a role in maintaining DNA methylation in mammalian cells." Science **317**(5845): 1760-1764.
- Bourc'his, D., G. L. Xu, et al. (2001). "Dnmt3L and the establishment of maternal genomic imprints." Science **294**(5551): 2536-2539.
- Bowman, B. R., S. Lee, et al. (2008). "Structure of the E. coli DNA glycosylase AlkA bound to the ends of duplex DNA: a system for the structure determination of lesion-containing DNA." Structure **16**(8): 1166-1174.
- Bowman, B. R., S. Lee, et al. (2010). "Structure of Escherichia coli AlkA in complex with undamaged DNA." J Biol Chem **285**(46): 35783-35791.
- Brent, T. P. (1979). "Partial purification and characterization of a human 3-methyladenine-DNA glycosylase." Biochemistry **18**(5): 911-916.
- Briebe, L. G., B. F. Eichman, et al. (2004). "Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase." EMBO J **23**(17): 3452-3461.
- Brook, R. D., B. Franklin, et al. (2004). "Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association." Circulation **109**(21): 2655-2671.
- Brooks, S. C., S. Adhikary, et al. (2013). "Recent advances in the structural mechanisms of DNA glycosylases." Biochim Biophys Acta **1834**(1): 247-271.
- Brown, R. C., B. E. Lemmon, et al. (1999). "Development of endosperm in *Arabidopsis thaliana*." Sex Plant Reproduction **12**: 32-42.
- Brown, T., W. N. Hunter, et al. (1986). "Molecular structure of the G.A base pair in DNA and its implications for the mechanism of transversion mutations." Proc Natl Acad Sci U S A **83**(8): 2402-2406.
- Brown, T. C. and J. Jiricny (1987). "A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine." Cell **50**(6): 945-950.
- Burrows, C. J. and J. G. Muller (1998). "Oxidative Nucleobase Modifications Leading to Strand Scission." Chem Rev **98**(3): 1109-1152.
- Burrows, C. J., J. G. Muller, et al. (2002). "Structure and potential mutagenicity of new hydantoin products from guanosine and 8-oxo-7,8-dihydroguanine oxidation by transition metals." Environ Health Perspect **110 Suppl 5**: 713-717.
- Calarco, J. P., F. Borges, et al. (2012). "Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA." Cell **151**(1): 194-205.
- Cao, C., K. Kwon, et al. (2003). "Solution structure and base perturbation studies reveal a novel mode of alkylated base recognition by 3-methyladenine DNA glycosylase I." J Biol Chem **278**(48): 48012-48020.
- Cao, X. and S. E. Jacobsen (2002). "Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes." Proc Natl Acad Sci U S A **99 Suppl 4**: 16491-16498.

- Cao, X. and S. E. Jacobsen (2002). "Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing." Curr Biol **12**(13): 1138-1144.
- Cao, X., N. M. Springer, et al. (2000). "Conserved plant genes with similarity to mammalian de novo DNA methyltransferases." Proc Natl Acad Sci U S A **97**(9): 4979-4984.
- Champagne, F. A. and M. J. Meaney (2007). "Transgenerational effects of social environment on variations in maternal care and behavioral response to novelty." Behav Neurosci **121**(6): 1353-1363.
- Chan, S. W., I. R. Henderson, et al. (2005). "Gardening the genome: DNA methylation in *Arabidopsis thaliana*." Nat Rev Genet **6**(5): 351-360.
- Chan, S. W., D. Zilberman, et al. (2004). "RNA silencing genes control de novo DNA methylation." Science **303**(5662): 1336.
- Chang-Yeh, A., D. E. Mold, et al. (1991). "Identification of a novel murine IAP-promoted placenta-expressed gene." Nucleic Acids Res **19**(13): 3667-3672.
- Chaturvedi, P. and S. C. Tyagi (2014). "Epigenetic mechanisms underlying cardiac degeneration and regeneration." Int J Cardiol **173**(1): 1-11.
- Chedin, F., M. R. Lieber, et al. (2002). "The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a." Proc Natl Acad Sci U S A **99**(26): 16916-16921.
- Chen, D., M. J. Lucey, et al. (2003). "T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha." J Biol Chem **278**(40): 38586-38592.
- Chen, J., B. Derfler, et al. (1990). "*Saccharomyces cerevisiae* 3-methyladenine DNA glycosylase has homology to the AlkA glycosylase of *E. coli* and is induced in response to DNA alkylation damage." EMBO J **9**(13): 4569-4575.
- Chen, J. and L. Samson (1991). "Induction of *S.cerevisiae* MAG 3-methyladenine DNA glycosylase transcript levels in response to DNA damage." Nucleic Acids Res **19**(23): 6427-6432.
- Chen, L., A. M. MacMillan, et al. (1991). "Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase." Biochemistry **30**(46): 11018-11025.
- Chen, Z., A. C. Karaplis, et al. (2001). "Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition." Hum Mol Genet **10**(5): 433-443.
- Cheng, K. C., D. S. Cahill, et al. (1992). "8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G---T and A---C substitutions." J Biol Chem **267**(1): 166-172.
- Chepanoske, C. L., M. P. Golinelli, et al. (2000). "Positively charged residues within the iron-sulfur cluster loop of *E. coli* MutY participate in damage recognition and removal." Arch Biochem Biophys **380**(1): 11-19.
- Choi, Y., M. Gehring, et al. (2002). "DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis." Cell **110**(1): 33-42.

- Choi, Y., J. J. Harada, et al. (2004). "An invariant aspartic acid in the DNA glycosylase domain of DEMETER is necessary for transcriptional activation of the imprinted MEDEA gene." Proc Natl Acad Sci U S A **101**(19): 7481-7486.
- Chuang, L. S., H. I. Ian, et al. (1997). "Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1." Science **277**(5334): 1996-2000.
- Cliffe, L. J., G. Hirsch, et al. (2012). "JBP1 and JBP2 proteins are Fe<sup>2+</sup>/2-oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA." J Biol Chem **287**(24): 19886-19895.
- Cliffe, L. J., R. Kieft, 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**(5): 1452-1462.
- Cokus, S. J., S. Feng, et al. (2008). "Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning." Nature **452**(7184): 215-219.
- Colot, V. and J. L. Rossignol (1999). "Eukaryotic DNA methylation as an evolutionary device." BioEssays **21**(5): 402-411.
- Connor, E. E. and M. D. Wyatt (2002). "Active-site clashes prevent the human 3-methyladenine DNA glycosylase from improperly removing bases." Chem Biol **9**(9): 1033-1041.
- Cordoba-Canero, D., T. Morales-Ruiz, et al. (2009). "Single-nucleotide and long-patch base excision repair of DNA damage in plants." Plant J **60**(4): 716-728.
- Cortazar, D., C. Kunz, et al. (2007). "The enigmatic thymine DNA glycosylase." DNA Repair **6**(4): 489-504.
- Cortazar, D., C. Kunz, et al. (2011). "Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability." Nature **470**(7334): 419-423.
- Cortellino, S., J. Xu, et al. (2011). "Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair." Cell **146**(1): 67-79.
- Coulondre, C., J. H. Miller, et al. (1978). "Molecular basis of base substitution hotspots in Escherichia coli." Nature **274**(5673): 775-780.
- Cross, M., R. Kieft, et al. (1999). "The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans." EMBO J **18**(22): 6573-6581.
- Cross, S. H., R. R. Meehan, et al. (1997). "A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins." Nat Genet **16**(3): 256-259.
- Cubas, P., C. Vincent, et al. (1999). "An epigenetic mutation responsible for natural variation in floral symmetry." Nature **401**(6749): 157-161.
- Cunningham, R. P., H. Asahara, et al. (1989). "Endonuclease III is an iron-sulfur protein." Biochemistry **28**(10): 4450-4455.
- Dalhous, B., J. K. Laerdahl, et al. (2009). "DNA base repair--recognition and initiation of catalysis." FEMS Microbiol Rev **33**(6): 1044-1078.
- Danchin, E., A. Charmantier, et al. (2011). "Beyond DNA: integrating inclusive inheritance into an extended theory of evolution." Nat Rev Genet **12**(7): 475-486.

- David-Cordonnier, M. H., S. Boiteux, et al. (2001). "Efficiency of excision of 8-oxo-guanine within DNA clustered damage by XRS5 nuclear extracts and purified human OGG1 protein." Biochemistry **40**(39): 11811-11818.
- David, S. S., V. L. O'Shea, et al. (2007). "Base-excision repair of oxidative DNA damage." Nature **447**(7147): 941-950.
- David, S. S. and S. D. Williams (1998). "Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair." Chem Rev **98**(3): 1221-1262.
- DeChiara, T. M., E. J. Robertson, et al. (1991). "Parental imprinting of the mouse insulin-like growth factor II gene." Cell **64**(4): 849-859.
- Delaney, S., W. L. Neeley, et al. (2007). "The substrate specificity of MutY for hyperoxidized guanine lesions in vivo." Biochemistry **46**(5): 1448-1455.
- DiPaolo, C., R. Kieft, et al. (2005). "Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein." Mol Cell **17**(3): 441-451.
- Dong, A., J. A. Yoder, et al. (2001). "Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA." Nucleic Acids Res **29**(2): 439-448.
- Dooijes, D., I. Chaves, et al. (2000). "Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*." Nucleic Acids Res **28**(16): 3017-3021.
- Drohat, A. C., K. Kwon, et al. (2002). "3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member." Nat Struct Biol **9**(9): 659-664.
- Du, Z., J. Song, et al. (2010). "DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination." Sci Signal **3**(146): ra80.
- Duclos, S., P. Aller, et al. (2012). "Structural and biochemical studies of a plant formamidopyrimidine-DNA glycosylase reveal why eukaryotic Fpg glycosylases do not excise 8-oxoguanine." DNA Repair **11**(9): 714-725.
- Dulac, C. (2010). "Brain function and chromatin plasticity." Nature **465**(7299): 728-735.
- Duncan, B. K. and J. H. Miller (1980). "Mutagenic deamination of cytosine residues in DNA." Nature **287**(5782): 560-561.
- Ehrlich, M., M. A. Gama-Sosa, et al. (1982). "Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells." Nucleic Acids Res **10**(8): 2709-2721.
- Ehrlich, M. and R. Y. Wang (1981). "5-Methylcytosine in eukaryotic DNA." Science **212**(4501): 1350-1357.
- Eichman, B. F., E. J. O'Rourke, et al. (2003). "Crystal structures of 3-methyladenine DNA glycosylase MagIII and the recognition of alkylated bases." EMBO J **22**(19): 4898-4909.
- Ekanayake, D. and R. Sabatini (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**(11): 1465-1472.

- Ekanayake, D. K., T. Minning, et al. (2011). "Epigenetic regulation of transcription and virulence in *Trypanosoma cruzi* by O-linked thymine glucosylation of DNA." Mol Cell Biol **31**(8): 1690-1700.
- Ellegren, H. and B. C. Sheldon (2008). "Genetic basis of fitness differences in natural populations." Nature **452**(7184): 169-175.
- Esteller, M. (2003). "Cancer epigenetics: DNA methylation and chromatin alterations in human cancer." Adv Exp Med Biol **532**: 39-49.
- Esteller, M. and J. G. Herman (2002). "Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours." J Pathol **196**(1): 1-7.
- Esteve, P. O., H. G. Chin, et al. (2009). "Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells." Proc Natl Acad Sci U S A **106**(13): 5076-5081.
- Evans, M. D., M. Dizdaroglu, et al. (2004). "Oxidative DNA damage and disease: induction, repair and significance." Mutat Res **567**(1): 1-61.
- Fan, L., J. O. Fuss, et al. (2008). "XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations." Cell **133**(5): 789-800.
- Feinberg, A. P., R. Ohlsson, et al. (2006). "The epigenetic progenitor origin of human cancer." Nat Rev Genet **7**(1): 21-33.
- Feinberg, A. P. and B. Tycko (2004). "The history of cancer epigenetics." Nat Rev Cancer **4**(2): 143-153.
- Ferguson-Smith, A. C., B. M. Cattanach, et al. (1991). "Embryological and molecular investigations of parental imprinting on mouse chromosome 7." Nature **351**(6328): 667-670.
- Ferguson-Smith, A. C., H. Sasaki, et al. (1993). "Parental-origin-specific epigenetic modification of the mouse H19 gene." Nature **362**(6422): 751-755.
- Ficz, G., M. R. Branco, et al. (2011). "Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation." Nature **473**(7347): 398-402.
- Finnegan, E. J. and E. S. Dennis (1993). "Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*." Nucleic Acids Res **21**(10): 2383-2388.
- Fitzgerald, M. E. and A. C. Drohat (2008). "Coordinating the initial steps of base excision repair. Apurinic/aprimidinic endonuclease 1 actively stimulates thymine DNA glycosylase by disrupting the product complex." J Biol Chem **283**(47): 32680-32690.
- Freitag, M., P. C. Hickey, et al. (2004). "HP1 is essential for DNA methylation in *Neurospora*." Mol Cell **13**(3): 427-434.
- Friedberg, E. C., A. Aguilera, et al. (2006). "DNA repair: from molecular mechanism to human disease." DNA Repair **5**(8): 986-996.
- Friedman, J. I. and J. T. Stivers (2010). "Detection of damaged DNA bases by DNA glycosylase enzymes." Biochemistry **49**(24): 4957-4967.
- Fromme, J. C., A. Banerjee, et al. (2004). "Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase." Nature **427**(6975): 652-656.

- Fromme, J. C., A. Banerjee, et al. (2004). "DNA glycosylase recognition and catalysis." Curr Opin Struct Biol **14**(1): 43-49.
- Fromme, J. C. and G. L. Verdine (2003). "Structure of a trapped endonuclease III-DNA covalent intermediate." EMBO J **22**(13): 3461-3471.
- Fronza, G. and B. Gold (2004). "The biological effects of N3-methyladenine." J Cell Biochem **91**(2): 250-257.
- Fu, W., S. O'Handley, et al. (1992). "The role of the iron-sulfur cluster in Escherichia coli endonuclease III. A resonance Raman study." J Biol Chem **267**(23): 16135-16137.
- Fuks, F., W. A. Burgers, et al. (2000). "DNA methyltransferase Dnmt1 associates with histone deacetylase activity." Nat Genet **24**(1): 88-91.
- Gal-Yam, E. N., Y. Saito, et al. (2008). "Cancer epigenetics: modifications, screening, and therapy." Annu Rev Med **59**: 267-280.
- Gallinari, P. and J. Jiricny (1996). "A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase." Nature **383**(6602): 735-738.
- Gao, Z., H. L. Liu, et al. (2010). "An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation." Nature **465**(7294): 106-109.
- Gates, K. S. (2009). "An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals." Chem Res Toxicol **22**(11): 1747-1760.
- Gates, K. S., T. Nooner, et al. (2004). "Biologically relevant chemical reactions of N7-alkylguanine residues in DNA." Chem Res Toxicol **17**(7): 839-856.
- Gehring, M., J. H. Huh, et al. (2006). "DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation." Cell **124**(3): 495-506.
- Genereux, J. C. and J. K. Barton (2010). "Mechanisms for DNA charge transport." Chem Rev **110**(3): 1642-1662.
- Genereux, J. C., A. K. Boal, et al. (2010). "DNA-mediated charge transport in redox sensing and signaling." J Am Chem Soc **132**(3): 891-905.
- Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu Rev Biochem **74**: 481-514.
- Gommers-Ampt, J. H. and P. Borst (1995). "Hypermodified bases in DNA." FASEB J **9**(11): 1034-1042.
- Gong, Z., T. Morales-Ruiz, et al. (2002). "ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase." Cell **111**(6): 803-814.
- Gonzalez-Zulueta, M., C. M. Bender, et al. (1995). "Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing." Cancer Res **55**(20): 4531-4535.
- Gonzalez, A., V. Lovic, et al. (2001). "Intergenerational effects of complete maternal deprivation and replacement stimulation on maternal behavior and emotionality in female rats." Dev Psychobiol **38**(1): 11-32.

- Greger, V., E. Passarge, et al. (1989). "Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma." Hum Genet **83**(2): 155-158.
- Gros, L., A. A. Ishchenko, et al. (2003). "Enzymology of repair of etheno-adducts." Mutat Res **531**(1-2): 219-229.
- Gros, L., A. V. Maksimenko, et al. (2004). "Hijacking of the human alkyl-N-purine-DNA glycosylase by 3,N4-ethenocytosine, a lipid peroxidation-induced DNA adduct." J Biol Chem **279**(17): 17723-17730.
- Grossniklaus, U., J. P. Vielle-Calzada, et al. (1998). "Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis." Science **280**(5362): 446-450.
- Guan, Y., R. C. Manuel, et al. (1998). "MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily." Nat Struct Biol **5**(12): 1058-1064.
- Guex, N. and M. C. Peitsch (1997). "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling." Electrophoresis **18**(15): 2714-2723.
- Guo, J. U., Y. Su, et al. (2011). "Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond." Cell Cycle **10**(16): 2662-2668.
- Guo, J. U., Y. Su, et al. (2011). "Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain." Cell **145**(3): 423-434.
- Gup, X., A. A. Gorodetsky, et al. (2008). "Conductivity of a single DNA duplex bridging a carbon nanotube gap." Nat Biotechnol **18**: 1096-1100.
- Hajkova, P., K. Ancelin, et al. (2008). "Chromatin dynamics during epigenetic reprogramming in the mouse germ line." Nature **452**(7189): 877-881.
- Hanai, R., M. Yazu, et al. (1998). "On the experimental distinction between ssbs and dsbs in circular DNA." Int J Radiat Biol **73**(5): 475-479.
- Hansen, K. D., W. Timp, et al. (2011). "Increased methylation variation in epigenetic domains across cancer types." Nat Genet **43**(8): 768-775.
- Hardeland, U., R. Steinacher, et al. (2002). "Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover." EMBO J **21**(6): 1456-1464.
- Hark, A. T., C. J. Schoenherr, et al. (2000). "CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus." Nature **405**(6785): 486-489.
- Hashimoto, H., X. Zhang, et al. (2013). "Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant." J Mol Biol.
- Haushalter, K. A., M. W. Todd Stukenberg, et al. (1999). "Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors." Curr Biol **9**(4): 174-185.
- Hausinger, R. P. (2004). "FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes." Crit Rev Biochem Mol Biol **39**(1): 21-68.
- He, X. J., Y. F. Hsu, et al. (2009). "An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein." Cell **137**(3): 498-508.

- He, Y. F., B. Z. Li, et al. (2011). "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." Science **333**(6047): 1303-1307.
- Hedglin, M. and P. J. O'Brien (2008). "Human alkyladenine DNA glycosylase employs a processive search for DNA damage." Biochemistry **47**(44): 11434-11445.
- Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-318.
- Hendershot, J. M., A. E. Wolfe, et al. (2011). "Substitution of active site tyrosines with tryptophan alters the free energy for nucleotide flipping by human alkyladenine DNA glycosylase." Biochemistry **50**(11): 1864-1874.
- Henderson, I. R. and S. E. Jacobsen (2007). "Epigenetic inheritance in plants." Nature **447**(7143): 418-424.
- Henderson, P. T., J. C. Delaney, et al. (2003). "The hydantoin lesions formed from oxidation of 7,8-dihydro-8-oxoguanine are potent sources of replication errors in vivo." Biochemistry **42**(31): 9257-9262.
- Hendrich, B. and A. Bird (1998). "Identification and characterization of a family of mammalian methyl-CpG binding proteins." Mol Cell Biol **18**(11): 6538-6547.
- Hendrich, B., U. Hardeland, et al. (1999). "The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites." Nature **401**(6750): 301-304.
- Henikoff, S. and L. Comai (1998). "A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis." Genetics **149**(1): 307-318.
- Herman, J. G., F. Latif, et al. (1994). "Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma." Proc Natl Acad Sci U S A **91**(21): 9700-9704.
- Hermann, A., S. Schmitt, et al. (2003). "The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity." J Biol Chem **278**(34): 31717-31721.
- Hirochika, H., H. Okamoto, et al. (2000). "Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation." Plant Cell **12**(3): 357-369.
- Holliday, R. and J. E. Pugh (1975). "DNA modification mechanisms and gene activity during development." Science **187**(4173): 226-232.
- Hollis, T., Y. Ichikawa, et al. (2000). "DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, Escherichia coli AlkA." EMBO J **19**(4): 758-766.
- Hollis, T., A. Lau, et al. (2000). "Structural studies of human alkyladenine glycosylase and E. coli 3-methyladenine glycosylase." Mutat Res **460**(3-4): 201-210.
- Holmgren, C., C. Kanduri, et al. (2001). "CpG methylation regulates the Igf2/H19 insulator." Curr Biol **11**(14): 1128-1130.
- Horton, J. R., G. Ratner, et al. (2004). "Caught in the act: visualization of an intermediate in the DNA base-flipping pathway induced by HhaI methyltransferase." Nucleic Acids Res **32**(13): 3877-3886.

- Hsieh, C. L. (1999). "In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b." Mol Cell Biol **19**(12): 8211-8218.
- Hsieh, T. F., C. A. Ibarra, et al. (2009). "Genome-wide demethylation of Arabidopsis endosperm." Science **324**(5933): 1451-1454.
- Hsieh, T. F., J. Shin, et al. (2011). "Regulation of imprinted gene expression in Arabidopsis endosperm." Proc Natl Acad Sci U S A **108**(5): 1755-1762.
- Hsu, G. W., M. Ober, et al. (2004). "Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase." Nature **431**(7005): 217-221.
- Huffman, J. L., O. Sundheim, et al. (2005). "DNA base damage recognition and removal: new twists and grooves." Mutat Res **577**(1-2): 55-76.
- Huh, J. H., M. J. Bauer, et al. (2008). "Cellular programming of plant gene imprinting." Cell **132**(5): 735-744.
- Hung, M. S., N. Karthikeyan, et al. (1999). "Drosophila proteins related to vertebrate DNA (5-cytosine) methyltransferases." Proc Natl Acad Sci U S A **96**(21): 11940-11945.
- Hunter, W. N., T. Brown, et al. (1987). "The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution." J Biol Chem **262**(21): 9962-9970.
- Hunter, W. N., G. Kneale, et al. (1986). "Refined crystal structure of an octanucleotide duplex with G . T mismatched base-pairs." J Mol Biol **190**(4): 605-618.
- Ibarra, C. A., X. Feng, et al. (2012). "Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes." Science **337**(6100): 1360-1364.
- Ito, S., A. C. D'Alessio, et al. (2010). "Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification." Nature **466**(7310): 1129-1133.
- Ito, S., L. Shen, et al. (2011). "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." Science **333**(6047): 1300-1303.
- Jablonka, E. and M. J. Lamb (2007). "Precis of Evolution in Four Dimensions." Behav Brain Sci **30**(4): 353-365; discussion 365-389.
- Jacobs, A. L. and P. Schar (2011). "DNA glycosylases: in DNA repair and beyond." Chromosoma.
- Jaenisch, R. and A. Bird (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nat Genet **33** Suppl: 245-254.
- Jakovcevski, M. and S. Akbarian (2012). "Epigenetic mechanisms in neurological disease." Nat Med **18**(8): 1194-1204.
- Jang, H., H. Shin, et al. (2014). "Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases." Biochem Biophys Res Commun.
- Jin, S. G., C. Guo, et al. (2008). "GADD45A does not promote DNA demethylation." PLoS Genet **4**(3): e1000013.
- Jones, D. O., I. G. Cowell, et al. (2000). "Mammalian chromodomain proteins: their role in genome organisation and expression." BioEssays **22**(2): 124-137.
- Jurkowska, R. Z. and A. Jeltsch (2013). "Genomic imprinting--the struggle of the genders at the molecular level." Angew Chem Int Ed Engl **52**(51): 13524-13536.

- Kane, M. F., M. Loda, et al. (1997). "Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines." *Cancer Res* **57**(5): 808-811.
- Kaneda, A. and A. P. Feinberg (2005). "Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk." *Cancer Res* **65**(24): 11236-11240.
- Kaneda, M., M. Okano, et al. (2004). "Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting." *Nature* **429**(6994): 900-903.
- Kanno, T., M. F. Mette, et al. (2004). "Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation." *Curr Biol* **14**(9): 801-805.
- Kapazoglou, A., V. Drosou, et al. (2013). "The study of a barley epigenetic regulator, HvDME, in seed development and under drought." *BMC Plant Biol* **13**(1): 172.
- Karran, P. and T. Lindahl (1980). "Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus." *Biochemistry* **19**(26): 6005-6011.
- Kathe, S. D., R. Barrantes-Reynolds, et al. (2009). "Plant and fungal Fpg homologs are formamidopyrimidine DNA glycosylases but not 8-oxoguanine DNA glycosylases." *DNA Repair* **8**(5): 643-653.
- Kato, M., A. Miura, et al. (2003). "Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis." *Curr Biol* **13**(5): 421-426.
- Kavli, B., O. Sundheim, et al. (2002). "hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup." *J Biol Chem* **277**(42): 39926-39936.
- Kikusui, T., Y. Isaka, et al. (2005). "Early weaning deprives mouse pups of maternal care and decreases their maternal behavior in adulthood." *Behav Brain Res* **162**(2): 200-206.
- Kim, J. Y., K. J. Kwak, et al. (2010). "MicroRNA402 affects seed germination of Arabidopsis thaliana under stress conditions via targeting DEMETER-LIKE Protein3 mRNA." *Plant Cell Physiol* **51**(6): 1079-1083.
- Kim, K., S. Biade, et al. (1998). "Involvement of flap endonuclease 1 in base excision DNA repair." *J Biol Chem* **273**(15): 8842-8848.
- Kim, M. S., T. Kondo, et al. (2009). "DNA demethylation in hormone-induced transcriptional derepression." *Nature* **461**(7266): 1007-1012.
- Kinoshita, T., A. Miura, et al. (2004). "One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation." *Science* **303**(5657): 521-523.
- Kishikawa, S., T. Murata, et al. (2003). "Control elements of Dnmt1 gene are regulated in cell-cycle dependent manner." *Nucleic Acids Res Suppl*(3): 307-308.
- Kiyosue, T., N. Ohad, et al. (1999). "Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis." *Proc Natl Acad Sci U S A* **96**(7): 4186-4191.
- Klapacz, J., G. M. Lingaraju, et al. (2010). "Frameshift mutagenesis and microsatellite instability induced by human alkyladenine DNA glycosylase." *Mol Cell* **37**(6): 843-853.

- Klaunig, J. E. and L. M. Kamendulis (2004). "The role of oxidative stress in carcinogenesis." Ann Rev Pharmacol Toxicol **44**: 239-267.
- Klungland, A. and T. Lindahl (1997). "Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1)." EMBO J **16**(11): 3341-3348.
- Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.
- Kress, C., H. Thomassin, et al. (2001). "Local DNA demethylation in vertebrates: how could it be performed and targeted?" FEBS Lett **494**(3): 135-140.
- Kress, C., H. Thomassin, et al. (2006). "Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks." Proc Natl Acad Sci U S A **103**(30): 11112-11117.
- Kreutzer, D. A. and J. M. Essigmann (1998). "Oxidized, deaminated cytosines are a source of C -> T transitions in vivo." Proc Natl Acad Sci U S A **95**(7): 3578-3582.
- Krishnamurthy, N., X. Zhao, et al. (2008). "Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1." Biochemistry **47**(27): 7137-7146.
- Krokan, H. E., F. Drablos, et al. (2002). "Uracil in DNA--occurrence, consequences and repair." Oncogene **21**(58): 8935-8948.
- Kryston, T. B., A. B. Georgiev, et al. (2011). "Role of oxidative stress and DNA damage in human carcinogenesis." Mutat Res **711**(1-2): 193-201.
- Kubota, Y., R. A. Nash, et al. (1996). "Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein." EMBO J **15**(23): 6662-6670.
- Kumar, S., X. Cheng, et al. (1994). "The DNA (cytosine-5) methyltransferases." Nucleic Acids Res **22**(1): 1-10.
- Kuo, C. F., D. E. McRee, et al. (1992). "Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III." Science **258**(5081): 434-440.
- Kwon, K., C. Cao, et al. (2003). "A novel zinc snap motif conveys structural stability to 3-methyladenine DNA glycosylase I." J Biol Chem **278**(21): 19442-19446.
- Labahn, J., O. D. Scharer, et al. (1996). "Structural basis for the excision repair of alkylation-damaged DNA." Cell **86**(2): 321-329.
- Lane, N., W. Dean, et al. (2003). "Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse." Genesis **35**(2): 88-93.
- Larson, K., J. Sahn, et al. (1985). "Methylation-induced blocks to in vitro DNA replication." Mutat Res **150**(1-2): 77-84.
- Lau, A. Y., O. D. Scharer, et al. (1998). "Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision." Cell **95**(2): 249-258.
- Lau, A. Y., M. D. Wyatt, et al. (2000). "Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG." Proc Natl Acad Sci U S A **97**(25): 13573-13578.

- Lauster, R., T. A. Trautner, et al. (1989). "Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains." J Mol Biol **206**(2): 305-312.
- Laval, J. (1977). "Two enzymes are required from strand incision in repair of alkylated DNA." Nature **269**(5631): 829-832.
- Law, J. A. and S. E. Jacobsen (2010). "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." Nat Rev Genet **11**(3): 204-220.
- Lawley, P. D. and P. Brookes (1963). "The Action of Alkylating Agents on Deoxyribonucleic Acid in Relation to Biological Effects of the Alkylating Agents." Exp Cell Res **24**: SUPPL9:512-520.
- Lee, C. Y., J. C. Delaney, et al. (2009). "Recognition and processing of a new repertoire of DNA substrates by human 3-methyladenine DNA glycosylase (AAG)." Biochemistry **48**(9): 1850-1861.
- Lee, Y. F., D. S. Tawfik, et al. (2002). "Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation." Nucleic Acids Res **30**(22): 4937-4944.
- Leger, H., C. Smet-Nocca, et al. (2014). "A TDG/CBP/RARalpha Ternary Complex Mediates the Retinoic Acid-dependent Expression of DNA Methylation-sensitive Genes." Genomics Proteomics Bioinformatics **12**(1): 8-18.
- Leipold, M. D., J. G. Muller, et al. (2000). "Removal of hydantoin products of 8-oxoguanine oxidation by the Escherichia coli DNA repair enzyme, FPG." Biochemistry **39**(48): 14984-14992.
- Leiros, I., M. P. Nabong, et al. (2007). "Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA." EMBO J **26**(8): 2206-2217.
- Li, E., T. H. Bestor, et al. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-926.
- Li, G. M. (2010). "Novel molecular insights into the mechanism of GO removal by MutM." Cell Res **20**(2): 116-118.
- Li, Y. Q., P. Z. Zhou, et al. (2007). "Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair." Nucleic Acids Res **35**(2): 390-400.
- Liang, G., J. C. Lin, et al. (2004). "Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome." Proc Natl Acad Sci U S A **101**(19): 7357-7362.
- Lindahl, T. (1974). "An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues." Proc Natl Acad Sci U S A **71**(9): 3649-3653.
- Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." Nature **362**(6422): 709-715.
- Lindahl, T. (2000). "Suppression of spontaneous mutagenesis in human cells by DNA base excision-repair." Mutat Res **462**(2-3): 129-135.

- Lindroth, A. M., X. Cao, et al. (2001). "Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation." Science **292**(5524): 2077-2080.
- Lingaraju, G. M., C. A. Davis, et al. (2011). "Structural basis for the inhibition of human alkyladenine DNA glycosylase (AAG) by 3,N4-ethenocytosine-containing DNA." J Biol Chem **286**(15): 13205-13213.
- Lingaraju, G. M., M. Kartalou, et al. (2008). "Substrate specificity and sequence-dependent activity of the *Saccharomyces cerevisiae* 3-methyladenine DNA glycosylase (Mag)." DNA Repair **7**(6): 970-982.
- Lister, R., M. Pelizzola, et al. (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature **462**(7271): 315-322.
- Liu, J. and P. W. Doetsch (1998). "Escherichia coli RNA and DNA polymerase bypass of dihydrouracil: mutagenic potential via transcription and replication." Nucleic Acids Res **26**(7): 1707-1712.
- Lovic, V., A. Gonzalez, et al. (2001). "Maternally separated rats show deficits in maternal care in adulthood." Dev Psychobiol **39**(1): 19-33.
- Lund, G., L. Andersson, et al. (2004). "DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E." J Biol Chem **279**(28): 29147-29154.
- Luo, M., P. Bilodeau, et al. (1999). "Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*." Proc Natl Acad Sci U S A **96**(1): 296-301.
- Luo, W., J. G. Muller, et al. (2001). "Characterization of hydantoin products from one-electron oxidation of 8-oxo-7,8-dihydroguanosine in a nucleoside model." Chem Res Toxicol **14**(7): 927-938.
- Lyons, D. M. and P. J. O'Brien (2009). "Efficient recognition of an unpaired lesion by a DNA repair glycosylase." J Am Chem Soc **131**(49): 17742-17743.
- Ma, D. K., M. H. Jang, et al. (2009). "Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis." Science **323**(5917): 1074-1077.
- Maiti, A. and A. C. Drohat (2011). "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites." J Biol Chem **286**(41): 35334-35338.
- Maiti, A., A. Z. Michelson, et al. (2013). "Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA." J Am Chem Soc **135**(42): 15813-15822.
- Maiti, A., M. T. Morgan, et al. (2009). "Role of two strictly conserved residues in nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase." J Biol Chem **284**(52): 36680-36688.
- Maiti, A., M. T. Morgan, et al. (2008). "Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition." Proc Natl Acad Sci U S A **105**(26): 8890-8895.

- Maiti, A., M. S. Noon, et al. (2012). "Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA." Proc Natl Acad Sci U S A.
- Mansfield, C., S. M. Kerins, et al. (2003). "Characterisation of *Archaeoglobus fulgidus* AlkA hypoxanthine DNA glycosylase activity." FEBS Lett **540**(1-3): 171-175.
- Manvilla, B. A., A. Maiti, et al. (2012). "Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA." J Mol Biol **420**(3): 164-175.
- Margison, G. P. and P. J. O'Connor (1973). "Biological implications of the instability of the N-glycosidic bond of 3-methyldeoxyadenosine in DNA." Biochim Biophys Acta **331**(3): 349-356.
- Matsumoto, Y. and K. Kim (1995). "Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair." Science **269**(5224): 699-702.
- Mayer, W., A. Niveleau, et al. (2000). "Demethylation of the zygotic paternal genome." Nature **403**(6769): 501-502.
- McCarthy, T. V., P. Karran, et al. (1984). "Inducible repair of O-alkylated DNA pyrimidines in *Escherichia coli*." EMBO J **3**(3): 545-550.
- McCullough, A. K., M. L. Dodson, et al. (1999). "Initiation of base excision repair: glycosylase mechanisms and structures." Annu Rev Biochem **68**: 255-285.
- McGrath, J. and D. Solter (1984). "Completion of mouse embryogenesis requires both the maternal and paternal genomes." Cell **37**(1): 179-183.
- Meissner, A., T. S. Mikkelsen, et al. (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells." Nature **454**(7205): 766-770.
- Memisoglu, A. and L. Samson (1996). "Cloning and characterization of a cDNA encoding a 3-methyladenine DNA glycosylase from the fission yeast *Schizosaccharomyces pombe*." Gene **177**(1-2): 229-235.
- Memisoglu, A. and L. Samson (2000). "Base excision repair in yeast and mammals." Mutat Res **451**(1-2): 39-51.
- Memisoglu, A. and L. Samson (2000). "Contribution of base excision repair, nucleotide excision repair, and DNA recombination to alkylation resistance of the fission yeast *Schizosaccharomyces pombe*." J Bacteriol **182**(8): 2104-2112.
- Metz, A. H., T. Hollis, et al. (2007). "DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG)." EMBO J **26**(9): 2411-2420.
- Missero, C., M. T. Pirro, et al. (2001). "The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription." J Biol Chem **276**(36): 33569-33575.
- Mitea, C., E. M. Salentijn, et al. (2010). "A universal approach to eliminate antigenic properties of alpha-gliadin peptides in celiac disease." PloS One **5**(12): e15637.
- Miura, A., S. Yonebayashi, et al. (2001). "Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*." Nature **411**(6834): 212-214.

- Moe, E., D. R. Hall, et al. (2012). "Structure-function studies of an unusual 3-methyladenine DNA glycosylase II (AlkA) from *Deinococcus radiodurans*." Acta Crystallogr D Biol Crystallogr **68**(Pt 6): 703-712.
- Mok, Y. G., R. Uzawa, et al. (2010). "Domain structure of the DEMETER 5-methylcytosine DNA glycosylase." Proc Natl Acad Sci U S A **107**(45): 19225-19230.
- Mol, C. D., A. S. Arvai, et al. (2002). "Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases." J Mol Biol **315**(3): 373-384.
- Mol, C. D., A. S. Arvai, et al. (1995). "Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis." Cell **80**(6): 869-878.
- Mol, C. D., C. F. Kuo, et al. (1995). "Structure and function of the multifunctional DNA-repair enzyme exonuclease III." Nature **374**(6520): 381-386.
- Morales-Ruiz, T., A. P. Ortega-Galisteo, et al. (2006). "DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases." Proc Natl Acad Sci U S A **103**(18): 6853-6858.
- Morera, S., I. Grin, et al. (2012). "Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA." Nucleic Acids Res **40**(19): 9917-9926.
- Morgan, H. D., F. Santos, et al. (2005). "Epigenetic reprogramming in mammals." Hum Mol Genet **14 Spec No 1**: R47-58.
- Morgan, M. T., M. T. Bennett, et al. (2007). "Excision of 5-halogenated uracils by human thymine DNA glycosylase. Robust activity for DNA contexts other than CpG." J Biol Chem **282**(38): 27578-27586.
- Morgan, M. T., A. Maiti, et al. (2011). "Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA." Nucleic Acids Res **39**(6): 2319-2329.
- Morikawa, K., O. Matsumoto, et al. (1992). "X-ray structure of T4 endonuclease V: an excision repair enzyme specific for a pyrimidine dimer." Science **256**(5056): 523-526.
- Muren, N. B., E. D. Olmon, et al. (2012). "Solution, surface, and single molecule platforms for the study of DNA-mediated charge transport." Phys Chem Chem Phys **14**(40): 13754-13771.
- Nabel, C. S., S. A. Manning, et al. (2012). "The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential." ACS Chem Biol **7**(1): 20-30.
- Nair, J., A. Barbin, et al. (1999). "Etheno DNA-base adducts from endogenous reactive species." Mutat Res **424**(1-2): 59-69.
- Nan, X., F. J. Campoy, et al. (1997). "MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin." Cell **88**(4): 471-481.
- Nan, X., R. R. Meehan, et al. (1993). "Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2." Nucleic Acids Res **21**(21): 4886-4892.

- Neeley, W. L. and J. M. Essigmann (2006). "Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products." Chem Res Toxicol **19**(4): 491-505.
- Netz, D. J., A. J. Pierik, et al. (2012). "A bridging [4Fe-4S] cluster and nucleotide binding are essential for function of the Cfd1-Nbp35 complex as a scaffold in iron-sulfur protein maturation." J Biol Chem **287**(15): 12365-12378.
- Niehrs, C. (2009). "Active DNA demethylation and DNA repair." Differentiation **77**(1): 1-11.
- O'Brien, P. J. (2006). "Catalytic promiscuity and the divergent evolution of DNA repair enzymes." Chem Rev **106**(2): 720-752.
- O'Brien, P. J. and T. Ellenberger (2003). "Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines." Biochemistry **42**(42): 12418-12429.
- O'Brien, P. J. and T. Ellenberger (2004). "Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase." J Biol Chem **279**(11): 9750-9757.
- O'Rourke, E. J., C. Chevalier, et al. (2000). "A novel 3-methyladenine DNA glycosylase from helicobacter pylori defines a new class within the endonuclease III family of base excision repair glycosylases." J Biol Chem **275**(26): 20077-20083.
- Ohad, N., R. Yadegari, et al. (1999). "Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization." Plant Cell **11**(3): 407-416.
- Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.
- Okano, M., S. Xie, et al. (1998). "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." Nat Genet **19**(3): 219-220.
- Okano, M., S. Xie, et al. (1998). "Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells." Nucleic Acids Res **26**(11): 2536-2540.
- Olek, A. and J. Walter (1997). "The pre-implantation ontogeny of the H19 methylation imprint." Nat Genet **17**(3): 275-276.
- Ortega-Galisteo, A. P., T. Morales-Ruiz, et al. (2008). "Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks." Plant Mol Biol **67**(6): 671-681.
- Osorio, C., N. Wen, et al. (2012). "Targeted modification of wheat grain protein to reduce the content of celiac causing epitopes." Funct Integr Genomics **12**(3): 417-438.
- Oswald, J., S. Engemann, et al. (2000). "Active demethylation of the paternal genome in the mouse zygote." Curr Biol **10**(8): 475-478.
- Otani, J., K. Arita, et al. (2013). "Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4." J Biol Chem **288**(9): 6351-6362.
- Parikh, S. S., C. D. Mol, et al. (1998). "Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA." EMBO J **17**(17): 5214-5226.

- Parrilla-Doblas, J. T., M. I. Ponferrada-Marin, et al. (2013). "Early steps of active DNA demethylation initiated by ROS1 glycosylase require three putative helix-invading residues." Nucleic Acids Res **41**(18): 8654-8664.
- Pastor, W. A., U. J. Pape, et al. (2011). "Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells." Nature **473**(7347): 394-397.
- Paulsen, M. and A. C. Ferguson-Smith (2001). "DNA methylation in genomic imprinting, development, and disease." J Pathol **195**(1): 97-110.
- Penn, N. W., R. Suwalski, et al. (1972). "The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid." Biochem J **126**(4): 781-790.
- Penterman, J., R. Uzawa, et al. (2007). "Genetic interactions between DNA demethylation and methylation in Arabidopsis." Plant Physiol **145**(4): 1549-1557.
- Penterman, J., D. Zilberman, et al. (2007). "DNA demethylation in the Arabidopsis genome." Proc Natl Acad Sci U S A **104**(16): 6752-6757.
- Peters, A. (2005). "Particulate matter and heart disease: evidence from epidemiological studies." Toxicol Appl Pharmacol **207**(2 Suppl): 477-482.
- Petronis, A. (2010). "Epigenetics as a unifying principle in the aetiology of complex traits and diseases." Nature **465**(7299): 721-727.
- Pfaffeneder, T., B. Hackner, et al. (2011). "The discovery of 5-formylcytosine in embryonic stem cell DNA." Angew Chem Int Ed Engl **50**(31): 7008-7012.
- Piccolo, F. M. and A. G. Fisher (2013). "Getting rid of DNA methylation." Trends Cell Biol.
- Plosky, B. S., E. G. Frank, et al. (2008). "Eukaryotic Y-family polymerases bypass a 3-methyl-2'-deoxyadenosine analog in vitro and methyl methanesulfonate-induced DNA damage in vivo." Nucleic Acids Res **36**(7): 2152-2162.
- Ponferrada-Marin, M. I., M. I. Martinez-Macias, et al. (2010). "Methylation-independent DNA binding modulates specificity of Repressor of Silencing 1 (ROS1) and facilitates demethylation in long substrates." J Biol Chem **285**(30): 23032-23039.
- Ponferrada-Marin, M. I., J. T. Parrilla-Doblas, et al. (2011). "A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine." Nucleic Acids Res **39**(4): 1473-1484.
- Ponferrada-Marin, M. I., T. Roldan-Arjona, et al. (2009). "ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion." Nucleic Acids Res **37**(13): 4264-4274.
- Posfai, J., A. S. Bhagwat, et al. (1989). "Predictive motifs derived from cytosine methyltransferases." Nucleic Acids Res **17**(7): 2421-2435.
- Purmal, A. A., Y. W. Kow, et al. (1994). "Major oxidative products of cytosine, 5-hydroxycytosine and 5-hydroxyuracil, exhibit sequence context-dependent mispairing in vitro." Nucleic Acids Res **22**(1): 72-78.
- Qi, P. F., Y. M. Wei, et al. (2009). "The gamma-gliadin multigene family in common wheat (*Triticum aestivum*) and its closely related species." BMC Genomics **10**: 168.

- Rai, K., I. J. Huggins, et al. (2008). "DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45." Cell **135**(7): 1201-1212.
- Ramsahoye, B. H., D. Binizskiewicz, et al. (2000). "Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a." Proc Natl Acad Sci U S A **97**(10): 5237-5242.
- Rebollo, R., B. Horard, et al. (2010). "Jumping genes and epigenetics: Towards new species." Gene **454**(1-2): 1-7.
- Reinisch, K. M., L. Chen, et al. (1995). "The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing." Cell **82**(1): 143-153.
- Renciuk, D., O. Blacque, et al. (2013). "Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine." Nucleic Acids Res **41**(21): 9891-9900.
- Riazuddin, S. and T. Lindahl (1978). "Properties of 3-methyladenine-DNA glycosylase from *Escherichia coli*." Biochemistry **17**(11): 2110-2118.
- Richards, E. J. (2006). "Inherited epigenetic variation--revisiting soft inheritance." Nat Rev Genet **7**(5): 395-401.
- Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation." Cytogenet Cell Genet **14**(1): 9-25.
- Riggs, A. D. and P. A. Jones (1983). "5-methylcytosine, gene regulation, and cancer." Adv Cancer Res **40**: 1-30.
- Robertson, K. D. (2005). "DNA methylation and human disease." Nat Rev Genet **6**(8): 597-610.
- Robertson, K. D., S. Ait-Si-Ali, et al. (2000). "DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters." Nat Genet **25**(3): 338-342.
- Romano, C. A., P. A. Sontz, et al. (2011). "Mutants of the base excision repair glycosylase, endonuclease III: DNA charge transport as a first step in lesion detection." Biochemistry **50**(27): 6133-6145.
- Rougier, N., D. Bourc'his, et al. (1998). "Chromosome methylation patterns during mammalian preimplantation development." Genes Dev **12**(14): 2108-2113.
- Rubinson, E. H., S. Adhikary, et al. (2010). Structural Studies of Alkylpurine DNA Glycosylases. ACS Symposium Series : Structural Biology of DNA Damage and Repair. M. P. Stone. Washington, D.C., American Chemical Society. **1041**: 29-45.
- Rubinson, E. H. and B. F. Eichman (2012). "Nucleic acid recognition by tandem helical repeats." Curr Opin Struct Biol **22**(1): 101-109.
- Rubinson, E. H., A. S. Gowda, et al. (2010). "An unprecedented nucleic acid capture mechanism for excision of DNA damage." Nature **468**(7322): 406-411.
- Rubinson, E. H., A. H. Metz, et al. (2008). "A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD." J Mol Biol **381**(1): 13-23.

- Rutledge, L. R. and S. D. Wetmore (2011). "Modeling the chemical step utilized by human alkyladenine DNA glycosylase: a concerted mechanism AIDS in selectively excising damaged purines." J Am Chem Soc **133**(40): 16258-16269.
- Samet, J. M., F. Dominici, et al. (2000). "Fine particulate air pollution and mortality in 20 U.S. cities, 1987-1994." N Engl J Med **343**(24): 1742-1749.
- Santi, D. V., C. E. Garrett, et al. (1983). "On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs." Cell **33**(1): 9-10.
- Santos, F., B. Hendrich, et al. (2002). "Dynamic reprogramming of DNA methylation in the early mouse embryo." Dev Biol **241**(1): 172-182.
- Saparbaev, M., K. Kleibl, et al. (1995). "*Escherichia coli*, *Saccharomyces cerevisiae*, rat and human 3-methyladenine DNA glycosylases repair 1,*N*<sup>6</sup>-ethenoadenine when present in DNA." Nucleic Acids Res **23**(18): 3750-3755.
- Saparbaev, M., S. Langouet, et al. (2002). "1,*N*(2)-ethenoguanine, a mutagenic DNA adduct, is a primary substrate of *Escherichia coli* mismatch-specific uracil-DNA glycosylase and human alkylpurine-DNA-N-glycosylase." J Biol Chem **277**(30): 26987-26993.
- Saparbaev, M. and J. Laval (1994). "Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases." Proc Natl Acad Sci U S A **91**(13): 5873-5877.
- Saparbaev, M. and J. Laval (1998). "3,*N*4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for *Escherichia coli* double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase." Proc Natl Acad Sci U S A **95**(15): 8508-8513.
- Savva, R., K. McAuley-Hecht, et al. (1995). "The structural basis of specific base-excision repair by uracil-DNA glycosylase." Nature **373**(6514): 487-493.
- Scharer, O. D. and J. Jiricny (2001). "Recent progress in the biology, chemistry and structural biology of DNA glycosylases." BioEssays **23**(3): 270-281.
- Scharer, O. D., T. Kawate, et al. (1997). "Investigation of the mechanisms of DNA binding of the human G/T glycosylase using designed inhibitors." Proc Natl Acad Sci U S A **94**(10): 4878-4883.
- Schmitt, A., B. Malchow, et al. (2014). "The impact of environmental factors in severe psychiatric disorders." Front Neurosci **8**: 19.
- Schorderet, D. F. and S. M. Gartler (1992). "Analysis of CpG suppression in methylated and nonmethylated species." Proc Natl Acad Sci U S A **89**(3): 957-961.
- Schubeler, D., M. C. Lorincz, et al. (2000). "Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation." Mol Cell Biol **20**(24): 9103-9112.
- Sedgwick, B. (2004). "Repairing DNA-methylation damage." Nat Rev Mol Cell Biol **5**(2): 148-157.

- Setser, J., G. Lingaraju, et al. (2011). "Searching for DNA lesions: Structural evidence for lower and higher-affinity DNA binding conformations of human alkyladenine DNA glycosylase (AAG)." Biochemistry.
- Sharif, J., M. Muto, et al. (2007). "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA." Nature **450**(7171): 908-912.
- Shikazono, N., C. Pearson, et al. (2006). "The roles of specific glycosylases in determining the mutagenic consequences of clustered DNA base damage." Nucleic Acids Res **34**(13): 3722-3730.
- Sibghat, U., P. Gallinari, et al. (1996). "Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch-specific thymine DNA-glycosylase." Biochemistry **35**(39): 12926-12932.
- Sidorenko, V. S., G. A. Nevinsky, et al. (2007). "Mechanism of interaction between human 8-oxoguanine-DNA glycosylase and AP endonuclease." DNA Repair **6**(3): 317-328.
- Singer, T., C. Yordan, et al. (2001). "Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1)." Genes Dev **15**(5): 591-602.
- Sleutels, F., R. Zwart, et al. (2002). "The non-coding Air RNA is required for silencing autosomal imprinted genes." Nature **415**(6873): 810-813.
- Slupphaug, G., C. D. Mol, et al. (1996). "A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA." Nature **384**(6604): 87-92.
- Smet-Nocca, C., J. M. Wieruszkeski, et al. (2008). "The thymine-DNA glycosylase regulatory domain: residual structure and DNA binding." Biochemistry **47**(25): 6519-6530.
- Smith, Z. D. and A. Meissner (2013). "DNA methylation: roles in mammalian development." Nat Rev Genet **14**(3): 204-220.
- Snyder, R. (2002). "Benzene and leukemia." Crit Rev Toxicol **32**(3): 155-210.
- Sobol, R. W., J. K. Horton, et al. (1996). "Requirement of mammalian DNA polymerase-beta in base-excision repair." Nature **379**(6561): 183-186.
- Song, J., O. Rechkoblit, et al. (2011). "Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation." Science **331**(6020): 1036-1040.
- Sontz, P. A., N. B. Muren, et al. (2012). "DNA charge transport for sensing and signaling." Acc Chem Res **45**(10): 1792-1800.
- Steinacher, R. and P. Schar (2005). "Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation." Curr Biol **15**(7): 616-623.
- Stivers, J. T. (2008). "Extrahelical damaged base recognition by DNA glycosylase enzymes." Chemistry **14**(3): 786-793.
- Stivers, J. T. and Y. L. Jiang (2003). "A mechanistic perspective on the chemistry of DNA repair glycosylases." Chem Rev **103**(7): 2729-2759.
- Stoger, R., P. Kubicka, et al. (1993). "Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal." Cell **73**(1): 61-71.

- Surani, M. A., S. C. Barton, et al. (1984). "Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis." *Nature* **308**(5959): 548-550.
- Szabo, P., S. H. Tang, et al. (2000). "Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function." *Curr Biol* **10**(10): 607-610.
- Tahiliani, M., K. P. Koh, et al. (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." *Science* **324**(5929): 930-935.
- Takahara, T., T. Ohsumi, et al. (1996). "Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon." *Hum Mol Genet* **5**(7): 989-993.
- Talhaoui, I., S. Couve, et al. (2014). "Aberrant repair initiated by mismatch-specific thymine-DNA glycosylases provides a mechanism for the mutational bias observed in CpG islands." *Nucleic Acids Res.*
- Tang, L. Y., M. N. Reddy, et al. (2003). "The eukaryotic DNMT2 genes encode a new class of cytosine-5 DNA methyltransferases." *J Biol Chem* **278**(36): 33613-33616.
- Tarantini, L., M. Bonzini, et al. (2009). "Effects of particulate matter on genomic DNA methylation content and iNOS promoter methylation." *Environ Health Perspect* **117**(2): 217-222.
- Thayer, M. M., H. Ahern, et al. (1995). "Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure." *EMBO J* **14**(16): 4108-4120.
- Thomas, L., C. H. Yang, et al. (1982). "Two DNA glycosylases in Escherichia coli which release primarily 3-methyladenine." *Biochemistry* **21**(6): 1162-1169.
- Tini, M., A. Benecke, et al. (2002). "Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription." *Mol Cell* **9**(2): 265-277.
- Trautner, T. A., T. S. Balganesh, et al. (1988). "Chimeric multispecific DNA methyltransferases with novel combinations of target recognition." *Nucleic Acids Res* **16**(14A): 6649-6658.
- Tudek, B. (2003). "Imidazole ring-opened DNA purines and their biological significance." *J Biochem Mol Biol* **36**(1): 12-19.
- Tweedie, S., H. H. Ng, et al. (1999). "Vestiges of a DNA methylation system in Drosophila melanogaster?" *Nat Genet* **23**(4): 389-390.
- Ulbert, S., M. Cross, et al. (2002). "Expression of the human DNA glycosylase hSMUG1 in Trypanosoma brucei causes DNA damage and interferes with J biosynthesis." *Nucleic Acids Res* **30**(18): 3919-3926.
- Um, S., M. Harbers, et al. (1998). "Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase." *J Biol Chem* **273**(33): 20728-20736.
- Urnov, F. D. and A. P. Wolffe (2001). "Above and within the genome: epigenetics past and present." *J Mammary Gland Biol Neoplasia* **6**(2): 153-167.

- Valko, M., C. J. Rhodes, et al. (2006). "Free radicals, metals and antioxidants in oxidative stress-induced cancer." Chem Biol Interact **160**(1): 1-40.
- van Leeuwen, F., M. de Kort, et al. (1998). "The modified DNA base beta-D-glucosylhydroxymethyluracil confers resistance to micrococcal nuclease and is incompletely recovered by 32P-postlabeling." Anal Biochem **258**(2): 223-229.
- van Leeuwen, F., R. Kieft, et al. (1998). "Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*." Mol Cell Biol **18**(10): 5643-5651.
- van Leeuwen, F., M. C. Taylor, 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**(5): 2366-2371.
- van Leeuwen, F., E. R. Wijsman, et al. (1997). "Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*." Genes Dev **11**(23): 3232-3241.
- van Loon, B., E. Markkanen, et al. (2010). "Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine." DNA Repair **9**(6): 604-616.
- Varriale, A. (2014). "DNA Methylation, Epigenetics, and Evolution in Vertebrates: Facts and Challenges." Int J Evol Biol **2014**: 475981.
- Vassilyev, D. G., T. Kashiwagi, et al. (1995). "Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition." Cell **83**(5): 773-782.
- Veron, N. and A. H. Peters (2011). "Epigenetics: Tet proteins in the limelight." Nature **473**(7347): 293-294.
- Vineis, P. and K. Husgafvel-Pursiainen (2005). "Air pollution and cancer: biomarker studies in human populations." Carcinogenesis **26**(11): 1846-1855.
- Vinkenoog, R., M. Spielman, et al. (2000). "Hypomethylation promotes autonomous endosperm development and rescues postfertilization lethality in *ma1* mutants." Plant Cell **12**(11): 2271-2282.
- Wagenknecht, H. A. (2005). Charge Transfer in DNA: From Mechanism to Application. Weinheim, Germany, Wiley-VCH.
- Wang, J., S. Hevi, et al. (2009). "The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation." Nat Genet **41**(1): 125-129.
- Waters, T. R., P. Gallinari, et al. (1999). "Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1." J Biol Chem **274**(1): 67-74.
- Waters, T. R. and P. F. Swann (1998). "Kinetics of the action of thymine DNA glycosylase." J Biol Chem **273**(32): 20007-20014.
- Weber, M., J. J. Davies, et al. (2005). "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." Nat Genet **37**(8): 853-862.

- Weinfeld, M., A. Rasouli-Nia, et al. (2001). "Response of base excision repair enzymes to complex DNA lesions." Radiat Res **156**(5 Pt 2): 584-589.
- Wen, S., N. Wen, et al. (2012). "Structural genes of wheat and barley 5-methylcytosine DNA glycosylases and their potential applications for human health." Proc Natl Acad Sci U S A **109**(50): 20543-20548.
- White, M. F. and M. S. Dillingham (2012). "Iron-sulphur clusters in nucleic acid processing enzymes." Curr Opin Struct Biol **22**(1): 94-100.
- Wibley, J. E., T. R. Waters, et al. (2003). "Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1." Mol Cell **11**(6): 1647-1659.
- Wigler, M., D. Levy, et al. (1981). "The somatic replication of DNA methylation." Cell **24**(1): 33-40.
- Wilke, K., E. Rauhut, et al. (1988). "Sequential order of target-recognizing domains in multispecific DNA-methyltransferases." EMBO J **7**(8): 2601-2609.
- Williams, R. T. and Y. Wang (2012). "A density functional theory study on the kinetics and thermodynamics of N-glycosidic bond cleavage in 5-substituted 2'-deoxycytidines." Biochemistry **51**(32): 6458-6462.
- Wolfe, A. E. and P. J. O'Brien (2009). "Kinetic mechanism for the flipping and excision of 1,N(6)-ethenoadenine by human alkyladenine DNA glycosylase." Biochemistry **48**(48): 11357-11369.
- Wolffe, A. P. and M. A. Matzke (1999). "Epigenetics: regulation through repression." Science **286**(5439): 481-486.
- Wu, H. and Y. Zhang (2011). "Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation." Genes Dev **25**(23): 2436-2452.
- Wu, H. and Y. Zhang (2014). "Reversing DNA methylation: mechanisms, genomics, and biological functions." Cell **156**(1-2): 45-68.
- Wu, P., C. Qiu, et al. (2003). "Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4." J Biol Chem **278**(7): 5285-5291.
- Wu, S. C. and Y. Zhang (2010). "Active DNA demethylation: many roads lead to Rome." Nat Rev Mol Cell Biol **11**(9): 607-620.
- Wyatt, M. D., J. M. Allan, et al. (1999). "3-methyladenine DNA glycosylases: structure, function, and biological importance." Bioessays **21**(8): 668-676.
- Xiao, G., M. Tordova, et al. (1999). "Crystal structure of Escherichia coli uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited." Proteins **35**(1): 13-24.
- Yamagata, Y., M. Kato, et al. (1996). "Three-dimensional structure of a DNA repair enzyme, 3-methyladenine DNA glycosylase II, from Escherichia coli." Cell **86**(2): 311-319.
- Yoder, J. A. and T. H. Bestor (1998). "A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast." Hum Mol Genet **7**(2): 279-284.

- Yoder, J. A., N. S. Soman, et al. (1997). "DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe." J Mol Biol **270**(3): 385-395.
- Yoder, J. A., C. P. Walsh, et al. (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet **13**(8): 335-340.
- Yoshida, M., K. Makino, et al. (1997). "Substrate and mispairing properties of 5-formyl-2'-deoxyuridine 5'-triphosphate assessed by in vitro DNA polymerase reactions." Nucleic Acids Res **25**(8): 1570-1577.
- Yu, Z., P. A. Genest, et al. (2007). "The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase." Nucleic Acids Res **35**(7): 2107-2115.
- Zhang, H. and J. K. Zhu (2012). "Active DNA demethylation in plants and animals." Cold Spring Harb Symp Quant Biol **77**: 161-173.
- Zhang, L., X. Lu, et al. (2012). "Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA." Nat Chem Biol **8**(4): 328-330.
- Zhang, W., Z. Liu, et al. (2011). "Crystal structure of the mismatch-specific thymine glycosylase domain of human methyl-CpG-binding protein MBD4." Biochem Biophys Res Commun **412**(3): 425-428.
- Zhang, X., J. Yazaki, et al. (2006). "Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis." Cell **126**(6): 1189-1201.
- Zhao, B. and P. J. O'Brien (2011). "Kinetic mechanism for the excision of hypoxanthine by Escherichia coli AlkA and evidence for binding to DNA ends." Biochemistry **50**(20): 4350-4359.
- Zharkov, D. O. and A. P. Grollman (2005). "The DNA trackwalkers: principles of lesion search and recognition by DNA glycosylases." Mutat Res **577**(1-2): 24-54.
- Zharkov, D. O., G. V. Mechetin, et al. (2010). "Uracil-DNA glycosylase: Structural, thermodynamic and kinetic aspects of lesion search and recognition." Mutat Res **685**(1-2): 11-20.
- Zhu, B., Y. Zheng, et al. (2000). "5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex." Proc Natl Acad Sci U S A **97**(10): 5135-5139.
- Zhu, J., A. Kapoor, et al. (2007). "The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis." Curr Biol **17**(1): 54-59.
- Zhu, J. K. (2009). "Active DNA demethylation mediated by DNA glycosylases." Annu Rev Genet **43**: 143-166.
- Zhu, X., X. Yan, et al. (2012). "A model for 3-methyladenine recognition by 3-methyladenine DNA glycosylase I (TAG) from Staphylococcus aureus." Acta Crystallogr F Struct Biol Cryst Commun **68**(Pt 6): 610-615.
- Zilberman, D., M. Gehring, et al. (2007). "Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription." Nat Genet **39**(1): 61-69.