

CHARACTERIZATION OF HEAT-LIKE REPEAT SUPERFAMILY OF DNA
GLYCOSYLASES

By

Rongxin Shi

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

Biological Sciences

June 30, 2018

Nashville, Tennessee

Approved:

Katherine L. Friedman, Ph.D.

Todd Graham, Ph.D.

Carmelo Rizzo, Ph.D.

Benjamin Spiller, Ph.D.

Brandt F. Eichman, Ph.D.

To My Beloved Husband Jun, Dear and Loving Mom, Changfen, and Dad, Feng,
Thank You for All Your Love and Support

ACKNOWLEDGMENTS

First and foremost, I want to thank my Ph.D. supervisor, Dr. Brandt Eichman for his enormous support, encouragement and guidance throughout my graduate study. His passion in science always inspires me and his talent in research tells me what a great scientist is. Besides science, I would like to express my heartfelt thanks for his support in my career development. I could not finish my graduate study and be prepared for my future career without his guidance and support. I would also like to thank my dissertation committee members: Drs. Katherine Friedman, Todd Graham, Carmelo Rizzo, and Benjamin Spiller, not only for their intellectual guidance on my research projects, but also for their supportive contributions to my personal career development. I want to thank all my wonderful current and previous lab members for project and emotional support that helps me get through the challenges in research, work and life. I especially thank Dr. Elwood Mullins for mentoring me when I first joined the lab and offering assistance on my dissertation project. The work was supported by National Science Foundation (MCB-1517695), National Institutes of Health (R01 ES019625), and Department of Biological Sciences.

I want to thank my family and friends, especially my husband Jun, mom Changfen and dad Feng. Vandy and Nashville gave me the opportunity not only to undertake my graduate study in the fantastic Eichman lab, but also to meet my husband who always inspires me. I am also so grateful to my parents for their encouragement and unconditional love.

I will always remember the great time in Nashville and everyone, without whom I could not get to this day.

TABLE OF CONTENTS

	Page
DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	xi
Chapter	
I. INTRODUCTION.....	1
Sources of DNA Damage.....	1
History of DNA Discovery and Characterization	1
DNA Damage from Replication and Physiological Hydrolysis	4
DNA Damage from Cell Metabolites and Environmental Radiation	5
DNA Damage from Exogenous Alkylation Agents	8
Overview of DNA Repair Pathways	10
Base Excision Repair	11
Direct Reversal Repair.....	14
Nucleotide Excision Repair.....	16
Alkylpurine DNA Glycosylases.....	17
Human Alkyladenine DNA Glycosylase Superfamily	18
Alkylpurine Helix-Hairpin-Helix Superfamily.....	20
Helix-Turn-Helix_42 Superfamily	23
HEAT-like Repeat Superfamily	24
Scope of this work.....	26
II. SELECTIVE BASE EXCISION REPAIR OF DNA DAMAGE BY THE NON-BASE- FLIPPING DNA GLYCOSYLASE ALKC	27
Abstract.....	27

Introduction	28
Results	32
Proteins within Two AlkC Subgroups are Functionally Distinct from AlkD	32
AlkC Encircles and Bends Damaged DNA	37
The AlkC Ig-like Domain is a Unique DNA Binding Motif in Bacteria	40
AlkC Inserts Its Active Site into the DNA Duplex in lieu of Base Flipping	45
Structural Differences between AlkC and AlkD.....	54
AlkC Catalyzes Base Excision of 3mC and 1mA from Duplex DNA	58
Discussion.....	62
Methods	62
Phylogenetic Analysis.....	62
Protein Purification	63
Base Excision Assays.....	65
X-Ray Crystallography.....	66
III. CELLULAR FUNCTION OF ALKC AND ALKD IN RESPONSE TO DIFFERENT DNA DAMAGING AGENTS.....	69
Abstract.....	69
Introduction	70
Results	73
AlkC and AlkD are Not Essential for Repair of Methylation Damage.....	73
The Expression of AlkC and AlkD upon DNA Damaging Agent Treatment	75
AlkD-Mediated BER and UvrA-Mediated NER Work Synergistically in Repair of Bulky Yatakemycin Adducts	76
Cellular Protection against 3-Methylcytosine and 1-Methyladenine by AlkC	79
Discussion.....	82
Methods	83
Preparation of $\Delta alkC$, $\Delta alkD$, and $\Delta alkC \Delta alkD$ Cells	83
Determination of Resistance to DNA Damaging Agents	84
RNA Quantification	85
Base Excision Assays.....	85

IV. STRUCTURAL BIOLOGY OF HEAT-LIKE REPEAT FAMILY OF DNA GLYCOSYLASES	87
Abstract	87
Introduction	88
Discussion	91
Phylogeny of the HLR superfamily	91
The general HLR structure	93
AlkD Uses a Non-Base-Flipping Mechanism to Excise Bulky Lesions	95
AlkC Uses a Non-Base-Flipping Mechanism to Select for Small Alkyl-adducts ...	101
AlkD2, a B-Helix-Absent HLR Superfamily Member	103
AlkF Favors Branched DNA Binding.....	103
Comparison between HLR Proteins	105
V. DISCUSSION AND FUTURE DIRECTIONS	110
Structure-Function Analysis of AlkC	110
Structure-Function Analysis of AlkD	114
Function and Activity Studies of HLR Proteins	118
REFERENCES.....	120

LIST OF TABLES

Table	Page
1. Percentage of alkylated damage	9
2. Data collection and refinement statistics	38
3. Base excision activities	59

LIST OF FIGURES

Figure	Page
1. DNA damage and consequences	3
2. Common base lesions and abasic DNA.....	6
3. Mechanisms of S _N 1 and S _N 2 nucleophilic substitution.....	8
4. Base excision repair.....	12
5. Monofunctional and bifunctional DNA glycosylases.....	13
6. Direct reversal repair in <i>E.coli</i>	15
7. Structure of AAG.....	19
8. Structures of HhH alkylpurine DNA glycosylases	22
9. Structures of HLR and HTH_42 alkylpurine DNA glycosylases	24
10. AlkC α and AlkC β are specific for N3-methyladenine (3mA).....	34
11. AlkC α and AlkC β sequence alignment.....	35
12. Substrate specificities of AlkC and AlkD	36
13. AlkC encircles damaged DNA.....	40
14. Comparison of DNA-binding Ig folds.....	42
15. The Ig-like domain is important for AlkC β function.....	43
16. AlkC α proteins contain an insertion that may stabilize DNA bend in the absence of Ig-like domain	45
17. Crystallographic features of the active site	46
18. AlkC inserts its active site into the DNA.....	48
19. Characterization of PfAlkC proteins used in this study.....	50

20. Structure of the PfAlkC/THF-DNA complex and its comparison to PfAlkC/1aR-DNA	53
21. Comparison of AlkC and AlkD.....	55
22. Glycosylase-induced DNA distortion.....	56
23. AlkC excises 3-methylcytosine (3mC) and 1-methyladenine (1mA) from DNA.....	60
24. AlkC excision of 3mC and 1mA.....	61
25. Sensitivity of $\Delta alkC$, $\Delta alkD$, and $\Delta alkC\Delta alkD$ mutants to a variety of DNA damaging agents	74
26. Expression of putative <i>B. anthracis</i> alkylpurine DNA glycosylases	76
27. Sensitivity of $\Delta alkD$ mutant and wild-type to MMS and YTM	78
28. Determination of YTM resistance.....	79
29. AlkC has no activity for 1-methyladenine in RNA.....	80
30. Cellular function characterization of AlkC	81
31. Base excision repair and alkylated lesions removed by AlkC and AlkD.....	90
32. Phylogenetic analysis of HLR superfamily proteins	93
33. Structural comparison of HLR superfamily proteins	94
34. Comparison of catalytic irrelevant and relevant AlkD structures	96
35. Structural comparison of AlkC- and AlkD-DNA complexes in context of lesion removal	100
36. Dimerization of AlkF.....	104
37. N-terminal bundle of proteins with HEAT-like repeats.....	107
38. Schematic of BcAlkC (AlkC α) and PfAlkC (AlkC β) domains	113
39. Biosynthetic gene clusters of YTM and CC-1065	115

40. AlkD homologs C10R5 and YtkR2 116

LIST OF ABBREVIATIONS

μM	Micromolar
ϵA	1,N6-ethenoadenine
1aR	1-Azaribose
1mA	N1-Methyladenine
3d3mA	3-Deaza-3-methyladenine
3mA	3-methyladenine
3mC	3-methylcytosine
3mG	3-methylguanine
5mC	5-methylcytosine
7mG	7-methylguanine
8oxoG	7,8-Dihydro-8-oxoguanine
Å	Angstrom
A	Adenine
AAG	Human alkyladenine DNA glycosylase
AlkA	<i>E.coli</i> 3mA DNA glycosylase II
AP	Apurinic/aprimidinc
APE	AP endonuclease
BER	Base excision repair
BME	2-Mercaptoethanol
°C	Degrees Celsius
C (Cyt)	Cytosine
CD	Circular dichroism
CPD	Cyclobutane pyrimidine dimers
DNA	Deoxyribonucleic acid

DSB	Double strand break
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraceticacid
EMS	Ethyl methanesulfonate
EMSA	Electrophoretic mobility shift assay
EndoIII	Endonuclease III
F	Fluorine
FAM	6-carboxyfluoroscein
FaPy	Formamidopyrimidine
g	Gram
G (Gua)	Guanine
H2TH	Helix-two turn-helix
HCl	Hydrochloric acid
HEAT	Huntington/Elongation/A subunit/Target of rapamycin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HhH	Helix hairpin helix
HPLC	High performance liquid chromatography
hr	Hour
Hx	Hypoxanthine
Ig	Immunoglobulin
IR	Ionizing radiation
IPTG	Isopropyl thio- β -D-galactopyranoside
Kd	Equilibrium dissociation constant

kDa	Kilodalton
L	Liter
LB	Luria Broth
MAG	Methyladenine DNA glycosylase
MeOH	Methanol
MES	2-(N-morpholino)-ethanesulfonic acid
mg	Milligram
min	Minute
mL	Milliliter
mM	Millimolar
MMS	Methyl methanesulfonate
MNU	Methylnitrosourea
mol	Mole
MS	Mass spectrometry
n.d.	Not determined
NaCl	Sodium chloride
NEU	N-nitroso-N-ethylurea
NER	Nucleotide excision repair
Ni-NTA	Nickel-nitrotriacetic acid
NNK	4-(methylnitrosamineo)-1-(3-pyridyl)-butanone
O ⁶ mG	O6-methyl-guanine
PCR	Polymerase chain reaction
PDB	Protein Data Bank
POB	Pyridyloxobutyl
PEG	Polyethylene glycol

qRT-PCR	Quantitative Reverse Transcription PCR polymerase chain reaction
rmsd	Root-mean-square deviation
rpm	Rotations per minute
RNA	Ribonucleic acid
s	Second
s.d.	Standard deviation
SAD	Single-wavelength anomalous dispersion
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeMet	Selenomethionine
SSB	Single strand break
ssDNA	Single-stranded DNA
ssRNA	Double-stranded DNA
T (Thy)	Thymine
TAG	3mA DNA glycosylase I
TBE	Tris Boric Acid EDTA Buffer
THF	Tetrahydrofuran
Tris	Tris-(hydroxymethyl)-aminomethane
U	Uracil
UPLC	Ultra Performance Liquid chromatography
UV	Ultraviolet
WT	Wild-type
YTM	Yatakemycin

CHAPTER I

INTRODUCTION

The double-stranded DNA (dsDNA) carries genetic information for all living cells. Its accurate translation into proteins through RNA intermediate successfully passes the genetic code to downstream events controlling cell metabolism, development and reproduction. Therefore the integrity and stability of dsDNA are essential to life (Lodish et al, 2000). However, DNA is vulnerable to damage from endogenous metabolites, environmental carcinogens and genotoxic cancer therapeutics. The damaged DNA gives rise to a variety of deleterious gene mutations and chromosomal damage, and potentially leads to inherited disease, oncogenic transformation or even cancer progression (D'Andrea, 2014). Multiple DNA damage repair pathways and proteins have evolved to protect against the toxic and mutagenic DNA lesions (Wood, 1996). Revealing the mechanism of the DNA damage repair pathway and relevant proteins is crucial to understand how deficiencies in DNA repair affect cancer progression, and provides insight for specific therapies by tailoring treatment to patients. In this chapter, I will review various types of DNA damage and their corresponding repair mechanisms with a focus on alkylation damage and repair.

Sources of DNA Damage

History of DNA Discovery and Characterization

In the 1860s, the Swiss chemist Johann Friedrich Miescher first isolated "nuclein" inside the nuclei of human leucocytes, which showed resistance to protease digestion

(Miescher-Rüsch, 1871). In 1881, the German biochemist Albrecht Kossel found that nuclein is composed of adenine (A), thymine (T), guanine (G), cytosine (C) and uracil (U) (Kossel, 1879). Nuclein was renamed DNA and RNA, which are accepted as their present chemical names. In the early and mid-1900's, the Canadian-American physician Oswald Avery established that DNA is the hereditary material, and the Austrian biochemist Erwin Chargaff and others found although the chemical composition differs among species, the total amount of purines (A and G) equals to the total amount of pyrimidines (C and G) (Chargaff et al, 1950; Tamm et al, 1952). Based on the previous work along with the X-ray crystallography diffraction by English researchers Rosalind Franklin and Maurice Wilkins in 1953 (Franklin & Gosling, 1953), Watson and Crick elucidated the 3-dimensional model of DNA, which is an antiparallel double-stranded helix (Watson & Crick, 1953). Neighboring nucleosides are linked by phosphodiester bonds and two DNA strands are held together by hydrogen bonds of pairs of bases (A and T, C and G). Determining the DNA structure has allowed for many other breakthroughs to be made, including the discovery of semiconservative DNA replication, the central dogma of molecular biology, use of recombinant DNA in protein production, as well as DNA sequencing (Dahm, 2005). Today, it is still considered very important to elucidate how DNA maintains its functions and how DNA repair proteins maintain DNA structure.

Eukaryotes and prokaryotes are the two major classes of living organisms. Prokaryotes carry circular DNA in the cytoplasm, while eukaryotic DNA is stored in a nucleus and the linear DNA is wrapped around nucleosomes to form chromosomes (Fuerst, 2010). The integrity of DNA is challenged by endogenous, environmental, and

pharmaceutical factors, which lead to accumulation of a variety of DNA damage including base adducts, intra- and inter-strand crosslinks, base mismatches and strand breaks (**Figure 1**). DNA damage needs to be specifically and efficiently repaired to maintain proper cellular metabolism and the inheritance of genetic information. Defects in DNA repair may cause a number of small mutations or large chromosome perturbations, which may subsequently result in cell death of unicellular organisms, or cancer or degenerative diseases in multicellular organism (**Figure 1**). However, the induced deleterious effects of DNA damage may serve as a target for development of therapeutics and genetic diversity (Curtin, 2012).

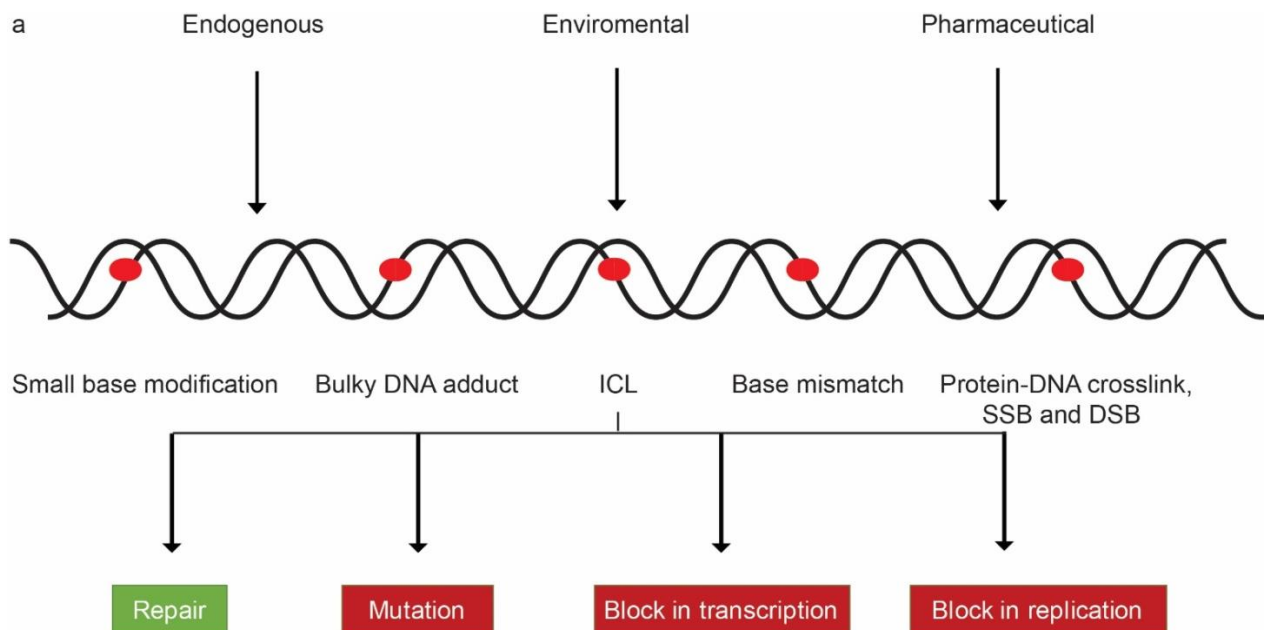


Figure 1. DNA damage and consequences. DNA damage results from endogenous, environmental, and pharmaceutical sources. The damage varies from small base modifications, bulky base adducts, interstrand crosslinks (ICLs), mismatched bases, to single-strand DNA breaks (SSBs) and double-strand breaks (DSBs). If left unrepaired, DNA damage can cause mutations in genomic DNA, and lead to cytotoxic effects on cellular metabolism by blocking DNA transcription or replication. [Adapted from (Roos et al, 2016)]

DNA Damage from Replication and Physiological Hydrolysis

The majority of DNA base damage arises from endogenous and environmental sources. Physiological processes include the introduction of errors during DNA replication, spontaneous base deamination, and modified bases caused by reactive metabolites (Pray, 2008).

Human cells replicate their DNA with high fidelity enzymes, which not only has high base selectivity but can also proofread for errors (McCulloch & Kunkel, 2008). However, DNA damage such as base substitutions, single base insertion, and deletion errors accumulate at a frequency of 10^{-6} to 10^{-8} per cell per generation (Nachman & Crowell, 2000). Approximately 5×10^{10} to 7×10^{10} cells are replicated in the human body per day with an average of 3.3×10^9 base pairs of DNA in the haploid genome (Bianconi et al, 2013). Given the enormous amount of DNA bases, misincorporated nucleotides are a major source of spontaneous mutagenesis, which can deleteriously affect cellular function. Apart from high-fidelity polymerases, topoisomerases catalyze strand scission to release superhelical tension during replication and transcription, and can also potentially cause strand breaks (Vos et al, 2011).

In addition to replication errors, spontaneous DNA damage, such as spontaneous deamination and depurination or depyrimidination, can lead to base substitution in cells. Deamination occurs where A, C, G, 5-methylcytosine (5mC) are converted to hypoxanthine (Hx), uracil (U), xanthine, and thymine respectively (**Figure 2b**) (Griffiths AJF, 2000). The deamination of cytosine alters base pairing from CG to UA, resulting in CG to TA transition mutations during DNA replication. 5mC in human cytosine-guanine (CpG) islands constitutes a hotspot for mutations. Deamination of

5mC to thymine results in a C to T transition which is responsible for one-third of hereditary diseases (Ehrlich et al, 1990; Holliday & Grigg, 1993). UV sunlight and intercalating agents such as mitomycin C enhance base deamination (Tommasi et al, 1997).

Abasic or AP (apurinic/aprimidinic) sites (**Figure 2a**) arise naturally from physiological hydrolytic processes, such as spontaneous hydrolysis of N-glycosidic bond, nucleobase adduct removal by the base excision pathway, or reactive oxygen species in cells. Depurination is more frequent than depyrimidination due to decreased stability of the purine N-glycosidic bond (Lindahl, 1993). AP sites block DNA replication and transcription, and can rapidly convert into single strand breaks (SSBs) (Boiteux & Guillet, 2004).

DNA Damage from Cell Metabolites and Environmental Radiation

Metabolism generated free radicals and other reactive species may induce spontaneous DNA damage, such as oxidization, alkylation and hydrolysis. For example, excessive reactive oxygen species cause deleterious oxidative DNA damage that is associated with the various human diseases like Alzheimer's and chronic inflammatory diseases (Kerksick & Zuhl, 2015). The electrophilic radicals react with DNA bases, elevating the level of 7,8 dihydro-8-oxoguanine (8-oxo-G) (**Figure 2b**), which generates double-strand DNA breaks (dsDNA) during replication, or causes GC to TA transversion by incorrectly pairing with adenine (Kryston et al, 2013). Thymine glycol, another major oxidation lesion, severely blocks eukaryotic DNA polymerases (Yoon et al, 2010).

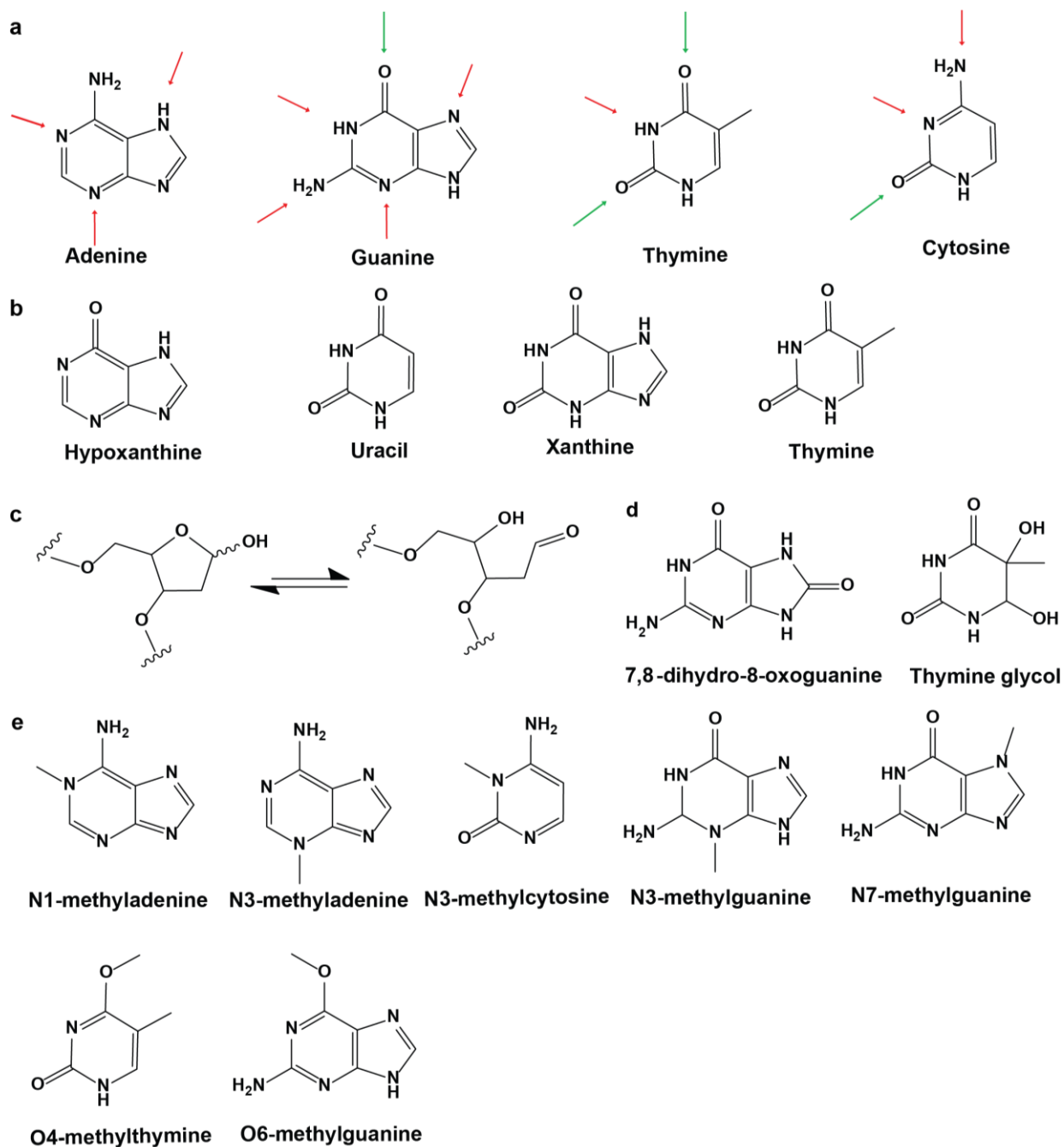


Figure 2. Common base lesions and abasic DNA. **a**, Nucleobases and alkylation sites. Major sites of N- and O-methylation are delineated with red and green arrows separately. **b**, Deaminated DNA bases. **c**, Two forms of abasic DNA: the closed sugar ring hemiacetal (left) and the open aldehyde (right) structures. **d**, Primary oxidation lesions. **e**, N-methylated and O-methylated nucleobases. [Adapted from (Chatterjee & Walker, 2017)]

Radiation is one of the most common environmental sources of DNA damage. Ultraviolet (UV) radiation is a non-ionizing type of radiation, while ionizing radiation (IR) emits high-energy particles, which induces atom ionization (Reisz et al, 2014). UV radiation causes deleterious effects on DNA by producing reactive oxygen species, or covalently crosslinking pyrimidines (Yagura et al, 2011; Yu & Lee, 2017). Our skin, as the first barrier against UV radiation, can develop skin cancer by exposure to UV-A (320-400 nm) and UV-B (290-315 nm) light (Mouret et al, 2006). UV-C (190-290 nm) is widely used in the laboratory due to its high absorption by DNA. Cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone photoproducts (6-4 PPs) are two types of DNA damage caused by UV radiation. These bulky lesions distort DNA structure, which results in blocking of DNA replication and transcription (You et al, 2001). Replication of these photoproducts by TLS polymerases will potentially lead to mutations. These lesions are cytotoxic if left unrepaired and not bypassed (Weber, 2005). Apart from UV radiation, IR is an important environmental source of DNA damage present in our daily lives. It directly reacts with DNA or attacks DNA via reactive oxygen and other species. The lesions are similar to those generated from endogenous reactive metabolites. Moreover, IR may lead to more toxic single-strand breaks (SSB) and double-strand breaks (DSB). UV and IR induced DNA damage may transform normal cells into rapidly proliferating cancer cells, and cause various cancers (Behjati et al, 2016). Despite harmful consequences, IR at low doses is used in computed tomography (CT) scanning and radiotherapy as diagnostic and cancer treatment tools.

DNA Damage from Exogenous Alkylation Agents

A variety of exogenous damage results from environmental alkylating and oxidizing agents, which include chemotherapeutic drugs, laboratory agents and tobacco smoke (**Figure 2e**). The exogenous alkylating agents are widely present in air, water, plants and food, and produce a variety of base modifications, and most commonly, methylated bases (Chakarov et al, 2014). Alkylating agents, which were among the earliest cancer therapies, are classified into five major categories: alkyl sulfonates, nitrosoureas, nitrogen mustards, triazines, and ethylenimines (Colvin, 2003).

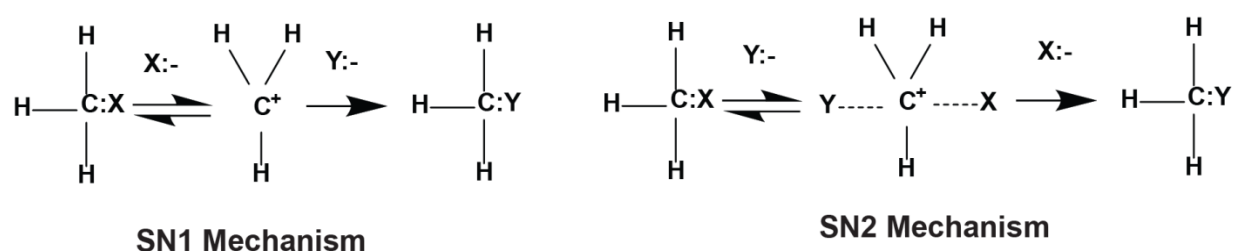


Figure 3. Mechanisms of S_N1 and S_N2 nucleophilic substitution. In S_N1 reaction, a carbenium ion intermediate is generated following by being attacked by the nucleophile from either side. In S_N2 reaction, a transition state is formed with both leaving group and nucleophile engaged. [Adapted from (Nay & O'Connor, 2013)]

The nucleophilic oxygens and nitrogens on the nucleobase are very sensitive to electrophilic attack and prone to modification by alkylating agents. The alkylation reactions proceed via either an S_N1 mechanism, which involves a positively charged carbonium ion intermediate, or an S_N2 pathway, which has both a nucleophile and an electrophile in the transition state (**Figure 3**) (Miller & Miller, 1981). Methyl methanesulfonate (MMS) and methyl nitrosourea (MNU) are commonly used mutagens in research for generating mutagenic and cytotoxic methylated bases by react at N7 of guanine and N3 of adenine. The unstable 7-methylguanine (7mG) and N3-methyladenine

(3mA) are prone to spontaneous AP site generation or formamidopyrimidine formation. These lesions may subsequently inhibit DNA replications. In addition to 7mG and 3mA, MNU treatment induces oxygen adducts, such as O⁶-methylguanine (O⁶-mG), potentially causing mutations in DNA. Because of multiple reactive sites on the nucleobases, phosphates, N1, N6 and N7 of adenine, N1, N2, and N3 of guanine, N3, N4 and O2 of cytosine, and N3, O2 and O4 of thymine can be modified as well (**Figure 2a**). The lesions occur in both ssDNA and dsDNA. As N1-adenine and N3-cytosine are less reactive because they participate in hydrogen-bonding in dsDNA, levels of these modifications are lower in dsDNA and higher in ssDNA and RNA (**Table 1**) (Nay & O'Connor., 2013; Wyatt & Pittman, 2006). Since alkylating agents serve as an important type of chemotherapy and have a long history of use, our study has been focusing on understanding alkylating damage and its repair.

Table 1. Percentage of alkylated damage [Adapted from (Nay & O'Connor, 2013)]

Alkylating Agent Alkylation	MMS (S _N 2)		MNU (S _N 1)	
	ssDNA	dsDNA	ssDNA	dsDNA
1mA	18.0	3.8	2.8	/
3mA	1.4	10.4	2.6	9.0
7mA	3.8	1.8	1.8	1.7
3mG	1.0	0.6	0.4	0.8
O ⁶ -mG	/	0.3	3.0	6.3
7mG	68.0	85.0	69.0	67.0
O ² -mT	/	/	/	0.1
3mT	/	0.8	/	0.3
O ⁴ -mC	/	/	/	0.4
O ² -mC	/	/	/	0.1
3mC	10.0	<1.0	2.3	0.6

Bifunctional alkylating agents, like nitrogen mustards, can also cause intra- and inter-strand crosslinks (ICLs) (Kohn et al, 1966). The cytotoxicity is significant as the crosslinking agent reacts with two sites on DNA, therefore the normal DNA activity is severely influenced. Similarly, crosslinking agents such as cisplatin and psoralen intercalate in DNA and cause ICL (Lopez-Martinez et al, 2016). These highly potent agents are used as chemotherapeutics in treatment of solid tumor and leukemia (Deans & West, 2011). However, due to the weak specificity for tumor cells, cancer treatment using DNA damaging drugs usually has severe side effect for patients. To prevent side effects, antibody-drug-conjugations (ADCs) are the emerging new therapies for cancer and other diseases (Kovtun & Goldmacher, 2007). For example, SYD985, in which the alkylating agent duocarmycin bearing an imidazo (1,2-a) pyridine-based DNA binding unit is conjugated to the anti-HER2 antibody trastuzumab, was highly effective and selective for multiple human cancer cell lines and increased stability in human plasma. It has been selected for Phase I clinical trials (Black et al, 2016).

DNA can also be damaged by exogenous polycyclic aromatic hydrocarbons (PAHs) containing two or more aromatic rings, other reactive electrophiles, and toxins. These bulky lesions potentially interfere with DNA replication and transcription, therefore inducing cell death or tumorigenesis (Korsh et al, 2015).

Overview of DNA Repair Pathways

DNA lesions result from both endogenous and exogenous toxicants, which potentially lead to genomic instability and apoptosis, cancer, or neurological degeneration (Hakem, 2008). Multiple DNA repair pathways have evolved in cells in

response to the frequent challenge of DNA damage. Base excision repair (BER), nucleotide excision repair (NER), direct reversal repair, double strand break repair and interstrand crosslink repair are the most common DNA-repair pathways. Although different pathways are usually associated with specific types of lesions, these processes are highly coordinated to balance the fidelity and efficiency of DNA repair. BER, direct reversal repair, and NER will be reviewed here as our study on BER reflects the cross-talks between these pathways.

Base Excision Repair

Base excision repair (BER) is the predominant pathway to repair small base adducts. This pathway includes short patch and long patch subpathways, which are initiated by a DNA glycosylase specifically recognizing and excising the damaged base from the DNA backbone (**Figure 4**).

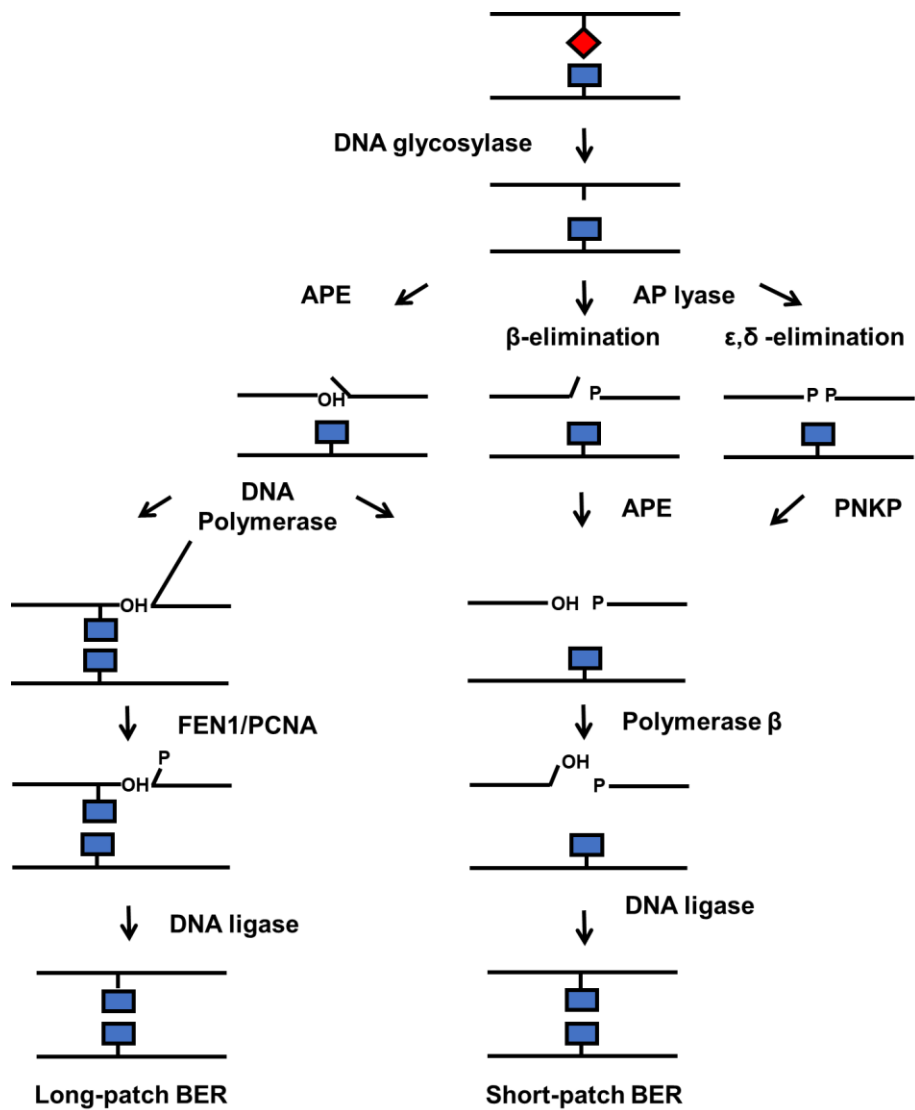
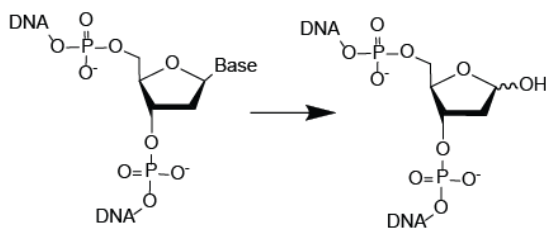


Figure 4. Base excision repair. The majority of small base modifications are repaired by base excision repair (BER). DNA glycosylase excises the damaged base and generates an abasic site. AP endonuclease (APE) and AP lyase incise the DNA backbone followed by the incised backbone processed by either long-patch or short patch BER to fill the gap and ligate the nicked strand. [Adapted from (Krokan, 2013)]

Short patch BER replaces one single nucleotide. In the case of monofunctional DNA glycosylases, an abasic site is generated upon cleavage of the N-glycosidic bond. The AP endonuclease 1 (APE1) processes the AP-DNA and leads to a one-nucleotide gap with a 5' phosphate and a 3' hydroxyl (**Figure 4 and Figure 5a**). But bifunctional glycosylases, which contain AP lyase activity in addition to base excision, cleaves the

phosphodiester backbone through β -elimination or β , δ -elimination (**Figure 4 and Figure 5b**). The product is further processed to produce a one nucleotide gap with a 5' phosphate and a 3' hydroxyl group. DNA polymerase and XRCC1/DNA ligase III complex incorporate the nucleotide and ligate the nicks. In contrast to the one nucleotide replacement reaction in short patch BER, in long-patch BER the nick generated by APE1 next to the AP site recruits polymerase β , polymerase δ , proliferating cell nuclear antigen (PCNA) and replication factor C (RFC), which synthesize a DNA fragment of 2-13 nucleotides. Flap endonuclease FEN1 removes the flapped substrate and the nick is ligated by LIG1 and PCNA (**Figure 4**) (Kim & M Wilson III, 2012; Krokan & Bjørås, 2013; Robertson et al, 2009; Schermerhorn & Delaney, 2014).

a Monofunctional DNA glycosylase



b Bifunctional DNA glycosylase

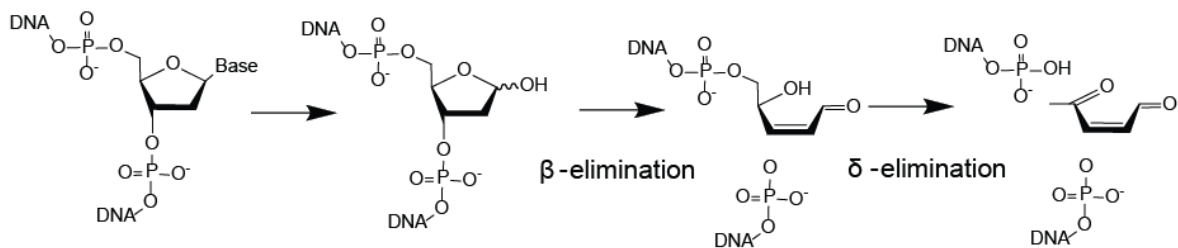


Figure 5. Monofunctional and bifunctional DNA glycosylases. Nucleobase is cleaved by monofunctional (**a**) and bifunctional (**b**) DNA glycosylases. Bifunctional glycosylase contains AP lyase activity, cleaves the phosphodiester backbone through β -elimination or β , δ -elimination in addition to N-glycosidic bond cleavage.

Direct Reversal Repair

Direct reversal repair is an error-free pathway to eliminate DNA and RNA damage without using a nucleotide template, phosphodiester backbone breakage or DNA synthesis. There are two major types of direct reversal repair approaches: repair of alkylation damage by O⁶-alkylguanine-DNA transferases and the AlkB family of dioxygenases, and repair of UV radiation-induced photolesions by photolyases (Eker et al, 2009; Mishina et al, 2006).

Transferase and dioxygenase direct reversal pathways protect against the deleterious effects of alkylating agents. One pathway involves the transfer of the alkyl group from the modified oxygen of O⁶-methylguanine (O⁶-mG) to a cysteine residue in the catalytic site of O⁶-methylguanine-DNA methyltransferase. The active site cysteine is irreversibly converted to methylcysteine, followed by the degradation of the enzyme. O⁶-methylguanine-DNA methyltransferases have been identified in bacteria, archaea, and eukaryotes (Mitra, 2007). *E. coli* adaptive response protein Ada is the bacterial homolog of O⁶-methylguanine-DNA methyltransferase. The C-terminal domain of Ada transfers the methyl group of O⁶-mG to Cys321 while the N-terminal domain of Ada transfers the methyl group of DNA backbone methylphosphotriesters to Cys38. Upon methylation of Cys38, Ada controls the adaptive response by binding to the *ada* operon and upregulating the expression of Ada, AlkA, AlkB, and AidB (**Figure 6**). The mammalian homolog of Ada is MGMT (O⁶-methylguanine DNA methyltransferase) (Yi & He, 2013). Decreased expression of MGMT as a result of epigenetic inactivation contributes to primary human neoplasia (Feinberg et al, 2006). Besides Ada, another O⁶-alkylguanine DNA alkyltransferase is Ogt (O-linked β -N-acetylglucosamine transferase), which is constitutively expressed in *E.coli* to safeguard the genome

against low levels of methylation damage (Wilkinson et al, 1989). AGT (alanine/glyoxylate aminotransferase) is the mammalian homolog of Ogt with slightly different substrate specificity (Fang et al, 2010).

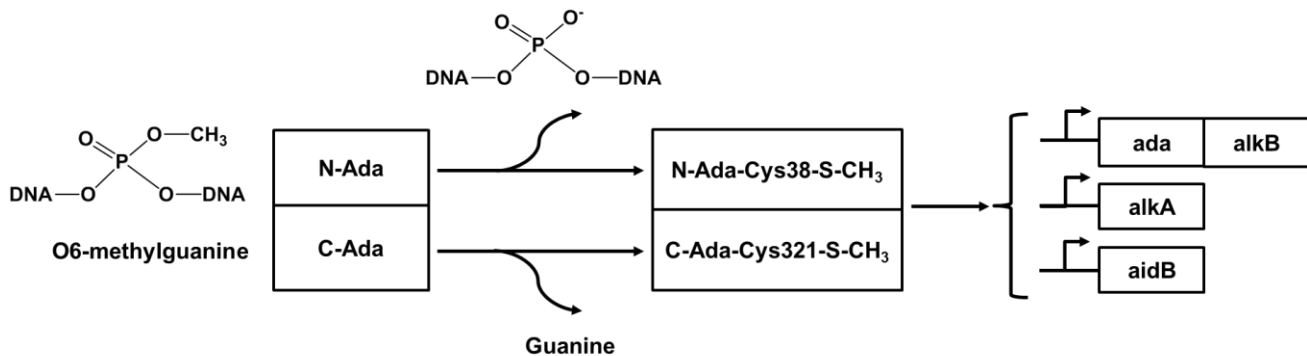


Figure 6. Direct reversal repair in *E.coli*. O⁶-methylguanine and methylated DNA backbone are repaired by *E.coli* Ada DNA methyltransferase. Both the C- and N-terminal domains of Ada are involved in transferring methyl groups. C-Ada transfers the methyl group from O⁶-MeG to Cys321 while N-Ada transfers the methyl group from DNA backbone methylphosphotriesters to Cys38. Cys38 methylation increases the expression of Ada, AlkA, AlkB, and AidB. [Adapted from (Sedgwick, 2004)]

The α -ketoglutarate/iron(II)-dependent dioxygenase *E.coli* AlkB mediates the other alkylation direct reversal repair pathway, which involves oxidative demethylation of N1-methyladenine (3mA) and N3-methylcytosine (3mC) (Pray, 2008; Trewick et al, 2002). Together with Fe(II) and α -ketoglutarate (α KG), AlkB oxidizes the modified nucleobases, followed by formaldehyde release and the nucleobase regeneration. Because the N1 and N3 positions of adenine and cytosine are involved in base pairing, these positions are less modified in dsDNA than ssDNA and single-strand RNA (ssRNA). *E.coli* AlkB prefers 1mA and 3mC in ssDNA and ssRNA over dsDNA. In addition to simple methyl adducts, AlkB has activity for etheno lesions. Compared with

E.coli AlkB, the mammalian AlkB homologs ALKBH1-8 and FTO (fat mass and obesity associated protein) exhibit a more diverse range of substrate specificity (Fedele et al, 2015).

Nucleotide Excision Repair

Nucleotide excision repair (NER) is an important repair pathway that exists from bacteria to eukaryotes. It is a multistep process that repairs photoproducts, such as CPDs and 6-4 PPs caused by UV radiation, and bulky lesions induced by mutagens or chemotherapeutic agents (Hu et al, 2015). Regardless of the source of the damage in prokaryotes or eukaryotes, NER detects and repairs the lesion through damage recognition, incision, excision, and gap filling. In bacteria, the lesion is detected by UvrA₂ and proofread by the UvrA₂UvrB₁ complex. Upon recognition and DNA duplex unwinding, UvrA is released from the complex, with UvrC subsequently recruited to UvrB. UvrC binding enables the incision of a fragment more than 10 nucleotides long flanking the damaged site (Hu et al, 2017). UvrB and UvrC hand over the nicked DNA to UvrD and DNA polymerase I to excise the damaged fragment and synthesize a new fragment using the complementary strand as a template. DNA ligase seals the nick (Kisker et al, 2013). In eukaryotes, NER is initiated by either global genomic NER (GG-NER), which eliminates helix-distorting lesions, or transcription-coupled NER (TC-NER), which allows gene expression by removing lesions that block transcription. GG-NER is initiated by the NER complex scanning through the genomic DNA and detecting the helix-distorting elements, while TC-NER is induced only when RNA polymerase II is blocked by a lesion in the template strand and the transcription pauses (Schärer, 2013). NER and cancer are closely related. Mutations in NER proteins have been shown to

cause *Xeroderma pigmentosum* (XP), an autosomal recessive disease in which deficiency in DNA repair of UV damage leads to skin carcinogenesis (Shuck et al, 2008).

Double strand break repair (DSBR) and mismatch repair (MMR) are also important repair pathways to cope with DNA lesions that arise from different sources. Highly deleterious DSBs caused by ionizing radiation or reaction intermediates in DNA repair lead to cell death or cancer. There are two major pathways to repair DSBs: nonhomologous DNA end joining (NHEJ) and homologous recombination (HR) (Chu, 1997). NHEJ is a less conservative pathway using Ku protein to recruit enzymes to the broken strand that promotes the joining of DNA ends (Lieber, 2010). HR is an error free pathway which utilizes the homologous newly replicated DNA (Jasin & Rothstein, 2013).

Alkylpurine DNA Glycosylases

Base excision repair is present in all domains of life. The proteins of this pathway recognize and repair the damaged nucleobase from the phosphodeoxyribose backbone by DNA glycosylase that specifically hydrolyzes the N-glycosidic bond. The resulting abasic site is nicked by AP endonuclease to generate a 3-hydroxyl substrate for gap repair synthesis by a DNA polymerase (Fromme & Verdine, 2004). DNA glycosylases that repair alkylated bases belong to four structural superfamilies: Helix-hairpin-Helix (HhH), human alkyladenine DNA glycosylase AAG, HEAT-like Repeat (HLR) DNA glycosylases (reviewed in Chapter IV), and Helix-turn Helix_42 (HtH_42, Pfam 06224) (Brooks et al, 2013; Mullins et al, 2017b). Alkylpurine DNA glycosylase is also called 3-methyladenine (3mA) or methylpurine DNA glycosylase, because 3mA is the first identified alkylation substrate (Lindahl, 1976). As expected, the DNA binding surface of

alkylpurine DNA glycosylase bears positive charge, as a feature of high affinity for negatively charged DNA substrates. Despite distinct glycosylase architecture, most of the glycosylases employ a base flipping mechanism (Krokan & Bjørås, 2013). In the traditional base-flipping model, the damaged base is flipped out of DNA helix and into the enzyme's active site. The base is typically recognized in this pocket through either shape complementarity or specific contacts with either the lesion or the base opposite the lesion. Meanwhile, the intrahelical base gap is filled and stabilized by specific side chains from the enzyme, which places a sharp kink in the DNA (Kim & M Wilson III, 2012). Therefore, the substrate specificity differs remarkably because of structure complementation and the stability of the N-glycosidic bond (Adhikary & Eichman, 2011; Lau et al, 2000). This thesis will describe a new type of glycosylase that acts differently than any previously studied glycosylase, so it is important to understand the molecular basis of how the canonical alkylpurine DNA glycosylase accurately discriminates its substrate against undamaged bases and other lesions.

Human Alkyladenine DNA Glycosylase Superfamily

Human AAG (also known as MPG) catalyzes the removal of a diverse range of alkylated nucleobases, not only induces cationic 3mA and 7mG, but neutral lesions such as 1,N6-ethenoadenine (ϵ A), N2,3-ethenoguanine (ϵ G) and hypoxanthine (Hx) (O'Brien & Ellenberger, 2004a). AAG plays an important role in methylation repair as it is the only 3mA glycosylase in some major tissues. *Aag*-deficient mouse cells are sensitive to alkylating agents (Engelward et al, 1997). Although AAG has activity for 8-

oxo-guanine (8oxoG), the majority of 8-oxoG lesions are repaired by 8-oxoguanine DNA glycosylase 1 (OGG1) (Bruner et al, 2000).

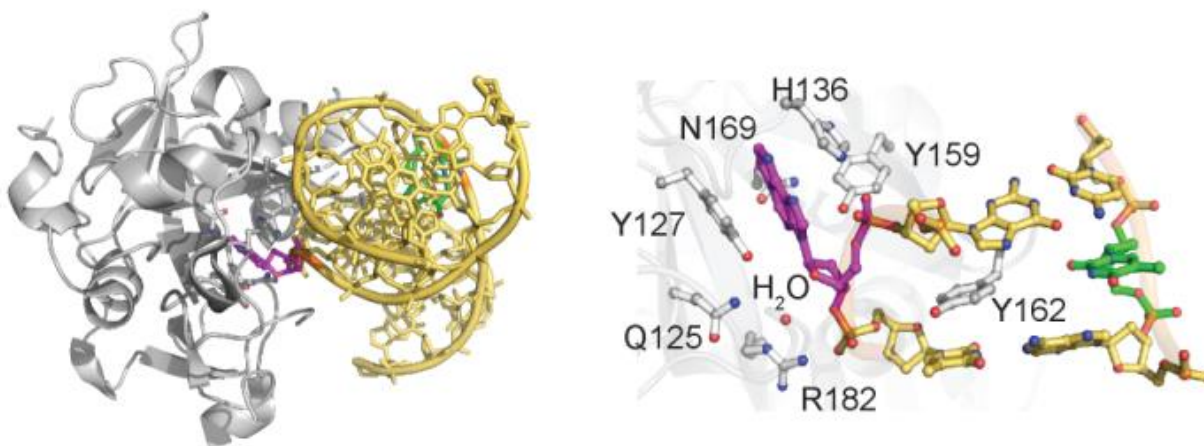


Figure 7. Structure of AAG. Structure and active site of AAG E125Q mutant in complex with 1,*N*6-ethnoadenine (ϵ A) DNA (PDB 1EWN) are shown from left to right. The protein is colored grey, DNA gold, ϵ A magenta, and nucleotide opposite lesion green. The active site residues are shown as grey sticks.

The structure of human AAG represents a unique fold with no similarity to the other DNA glycosylases. The single domain human AAG is composed of mixed α helices and β sheets (**Figure 7**) (Lau et al, 1998). In search for the aberrant bases, AAG diffuses along DNA to detect and repair them. During the lesion capture process, transient kinetic experiments of AAG with ϵ A suggested two steps involved: the first step is formation of the initial recognition complex between partially unstacked ϵ A and AAG, and the second step is formation of the stable recognition complex between fully flipped-out ϵ A and AAG (Hendershot & O'Brien, 2017; Lau et al, 2000). In addition, ϵ A flipping is highly favorable compared to normal bases (Hendershot & O'Brien, 2014; Lau et al, 2000).

In the catalytically competent conformation, DNA is only modestly bent (Lau et al, 2000). $\beta 3\beta 4$ of AAG inserts into the DNA minor groove. Tyr162, the residue at the tip of the β hairpin, is intercalated into the void generated by the flipped base to stabilize the duplex DNA and prevent reinsertion of the damaged base back into the DNA (Hendershot & O'Brien, 2017). Glu125 deprotonates the nucleophilic water to attack the anomeric C1' carbon for N-glycosidic bond cleavage via S_N2 mechanism. Arg182 stabilizes the nucleophilic water and the lesion via hydrogen bonding. The flipped-out nucleobase is stacked between Tyr159 and Tyr127. To discriminate against normal nucleobases, His136 is hydrogen bonded with N6 of ϵA , while the undamaged DNA lacks the capability to form this bond. Additionally, Asn169 clashes with N2 of guanine (**Figure 7**) (Lau et al, 2000). In the presence of the downstream endonuclease APE1, AAG will be displaced and released from the product.

Alkylpurine Helix-Hairpin-Helix Superfamily

The helix-hairpin-helix (HhH) superfamily of DNA glycosylases are widely distributed in all domains of life (Denver et al, 2003). They specifically recognize and remove diverse and distinct spectra of damaged bases. This HhH superfamily of alkylpurine DNA glycosylases includes *E.coli* 3-methyladenine DNA glycosylase II (AlkA), *E.coli* 3-methyladenine DNA glycosylase I (TAG), *S. cerevisiae* 3-methyladenine DNA glycosylase (MAG), and *S. pombe* alkylpurine DNA glycosylase (Mag1), all of which share a short, compact α -helical motif for DNA binding close to the active site (Doherty et al, 1996). Like most of the other DNA glycosylases, HhH superfamily glycosylases employ the canonical base flipping mechanism to flip out the damaged

base into the active site and remove it, with the gap left by the flipped-out base plugged by an intercalating residue (Hollis et al, 2000).

E.coli TAG is constitutively expressed and shows a narrow substrate spectrum for 3mA and 3mG (Bjelland et al, 1993). In contrast, *E.coli* AlkA, which is under control of the adaptive response pathway induced upon exposure to alkylation stress, works on a remarkably broader range of substrates, including 7mG, 7mA, O²-methylpyrimidines, the deaminated base hypoxanthine (Hx), the exocyclic base εA, 5-formyluracil, and even undamaged bases (Nakabeppu et al, 1984). The varying substrate spectra of Tag and AlkA allow *E.coli* to effectively respond to different levels of alkylation damage without producing too many cytotoxic AP sites. *E.coli* alkylpurine glycosylase *tag alkA* depleted cells are extremely sensitive to alkylating agents such as MMS.

AlkA is a monofunctional DNA glycosylase. The shallow surface of the active site allows AlkA to accommodate varying types of DNA damage. AlkA has an N-terminal β-sheet with unknown function (**Figure 8**). During catalysis, the DNA is highly distorted. The damaged nucleobase is flipped out of the DNA duplex into the catalytic pocket, leaving the gap filled by Leu125, causing a remarkable bending in DNA (Hollis et al, 2000; O'Brien & Ellenberger, 2004b). Asp238 performs direct nucleophilic attack via S_N1 mechanism. Meanwhile, Trp272 stabilizes the positively charged lesion through cation-π interaction (**Figure 8**). Pro175 contacts the lesion opposite the deoxyribose, leading to a widened minor groove. The HhH motif plays an important role in DNA binding via hydrogen bonding. Although almost all the polar protein-DNA interactions are made with the lesion-containing strand, AlkA highly prefers dsDNA (Hollis et al, 2000; Yamagata et al, 1996). AlkA uses its HhH motif to nonspecifically bind undamaged DNA. However,

the active pocket in undamaged DNA complex differs from that in damaged DNA complex with less kinking present. The arrangement of a wedge residue Leu125 and catalytic residue Asp238 do not interact with the undamaged DNA (Bowman et al, 2010).

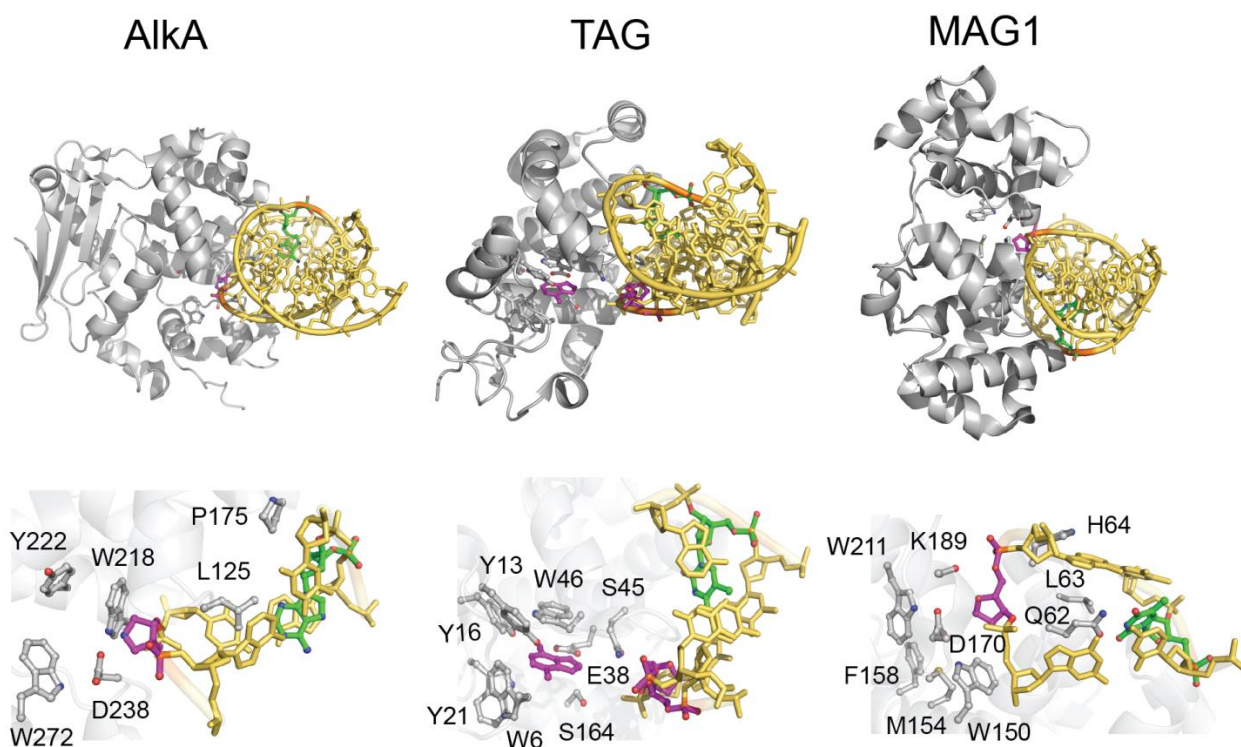


Figure 8. Structures of HhH alkylpurine DNA glycosylases. Structures and active sites of *E.coli* AlkA with 1-azaribose (1aR) DNA (PDB 1DIZ), *Salmonella typhi* TAG with tetrahydrofuran (THF) DNA (PDB 2OFI), and *S. cerevisiae* MAG1 with THF DNA (3S6I) are shown from left to right. Proteins are colored grey, DNA gold, 1aR in AlkA structure magenta, 3mA and THF in TAG structure magenta, THF in MAG1 structure magenta, and lesion opposite bases green. The active site residues are shown as grey sticks. 1aR and THF are analogs of oxocarbenium intermediate and abasic site in glycosylase reaction, correspondingly.

TAG's narrow substrate specificity for 3mA likely results from a steric clash with nucleobases other than 3mA. This hypothesis is supported by the crystal structure of DNA-*Salmonella typhi* TAG which shares high sequence similarity and identity to *E.coli* TAG (Metz et al, 2007).

MAG and Mag1 are *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* 3-methyladenine DNA glycosylases. MAG is inducible in the presence of methylating agents while Mag1 is constitutively expressed (Chen & Samson, 1991; Memisoglu & Samson, 2000). MAG excises cationic lesions 3mA, 7mG, as well as neutral bases ϵ A, Hx, and guanine, whereas Mag1 has reduced activity for ϵ A, and no activity for Hx or guanine. In the Mag1 catalysis model, Gln62 and Leu63 stabilize the gap in the damaged strand and the opposite strand, respectively. The minor groove binding residue His64 in Mag1 is substituted with Ser in MAG. Interestingly, mutation of His64 to Ser rescues Mag1 ϵ A activity, and mutation of the corresponding Ser in MAG to His remarkably decreases MAG's activity. However, these mutations have no effect on 7mG excision (Adhikary & Eichman, 2011). This observation agrees with the prediction that the substrate preference of Mag1 does not result from the steric exclusion, but due to the catalytic power of the active site to remove lesions.

Helix-Turn-Helix_42 Superfamily

AlkZ defines the most recently identified alkylpurine DNA glycosylase superfamily, which excises 7mG adducts and unhooks azinomycin B (AZB)-mediated interstrand crosslinks (ICLs) by cleaving the N-glycosidic bond, providing an alternative pathway to repair ICL (Mullins et al, 2017b). AlkZ is composed of tandem winged helix-turn-helix motifs classified as the HTH_42 superfamily with no previously determined protein structures. Gln37 and Gln39 (QxQ motif) are predicted to position a nucleophilic water to attack the N-glycosidic bond, with β 11/12 hairpin insertion into the minor groove to stabilize the repair complex (**Figure 9**) (Mullins et al, 2017b).

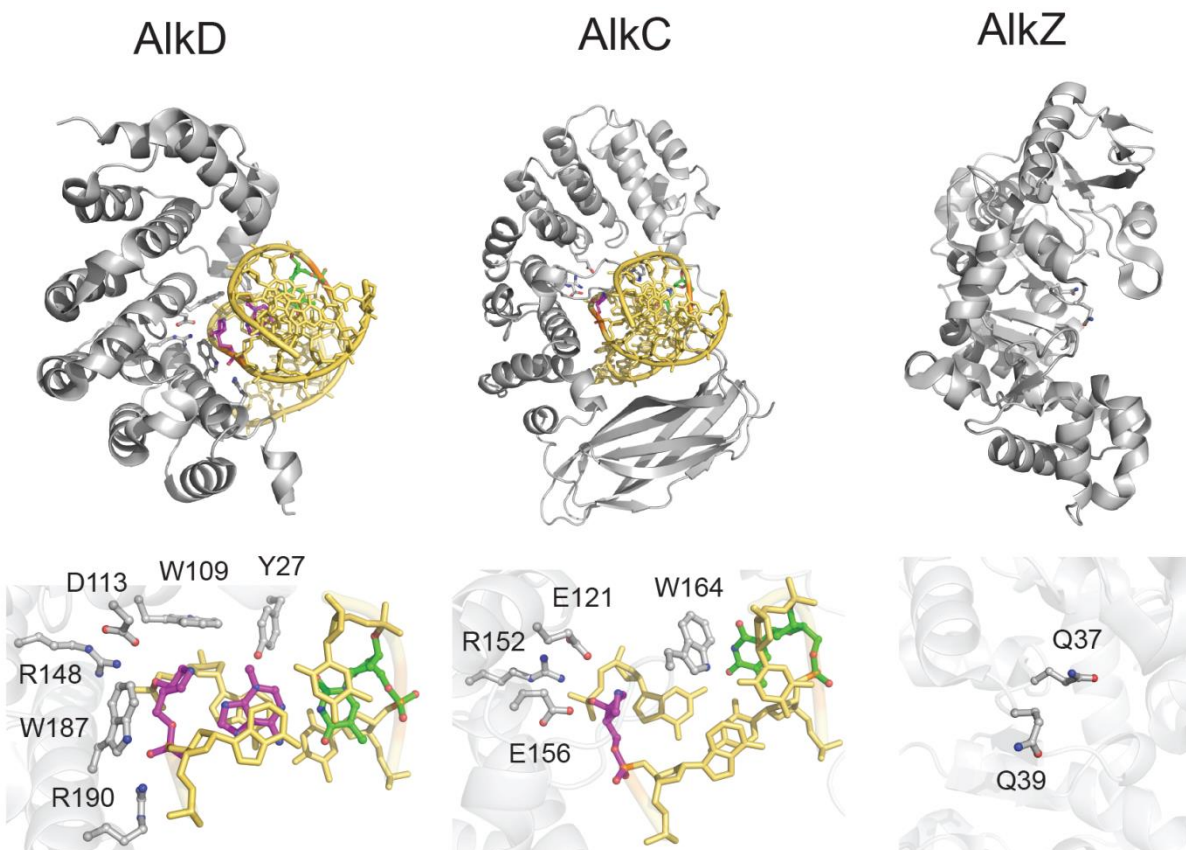


Figure 9. Structures of HLR and HTH_42 alkylpurine DNA glycosylases. Structures and active sites are shown for *B. cereus* AlkD with 1aR DNA (PDB 1DIZ), *P. fluorescens* AlkC with 1aR DNA (PDB 2OFI), and *Streptomyces sahachiroi* AlkZ (PDB 5UUJ). Proteins are colored grey, DNA gold, 1aR and 3mA in AlkD structure magenta, 1aR in AlkC structure magenta, and lesion opposite bases- green. The active site residues are shown as grey sticks.

HEAT-like Repeat Superfamily

Alkylpurine DNA glycosylases AlkC and AlkD define the HEAT-like Repeat (HLR) superfamily, and are comprised of HEAT-like α -helical repeats (HEAT is short for Huntington/Elongation/A subunit/Target-of-rapamycin) (Alseth et al, 2005; Rubinson et al, 2008). The ectopic expression of *B. cereus* AlkC and AlkD in alkylpurine glycosylase-deficient *E. coli* significantly decreased cell sensitivity. AlkD targets a relatively large range of substrates *in vitro*, including 3mA, 7mG and bulky lesions, whereas AlkC only

displays high specificity for 3mA (Alseth et al, 2006; Mullins et al, 2013). Structural studies of DNA glycosylases with DNA containing substrate, substrate-mimic, or product have provided insights into catalysis. With the exception of AlkC and AlkD, most characterized glycosylases use a similar base-flipping mechanism to distinguish and excise their substrates (Mullins et al, 2015b; Shi et al, 2018). However, the mechanism is different for AlkC and AlkD. AlkD which is composed of six tandem helical repeats, binds DNA and selectively targets destabilized base pairs using its concave surface. The lesion is held against the protein's concave surface while still stacked in the DNA duplex, which allows base removal without plug or wedge residues to stabilize the gap caused by the flipped damaged base. The catalysis is facilitated by the catalytic Glu113 and CH- π interactions between the nucleobase and Trp109 and Trp187 (Mullins et al, 2015b; Parsons et al, 2016). Meanwhile, the non-base-flipping mechanism enables AlkD to act on bulky adducts such as yatakemycin modified DNA, which provides an additional protection for cell to survive the highly potent antibiotic yatakemycin (YTM). Similar to AlkD, AlkC does not flip the lesion out of the duplex (Mullins et al, 2017a). AlkC highly kinks the DNA, exposing the damaged base into a restricted active site, therefore limiting the range of lesions that can be recognized by AlkC. 70% of AlkC homologs have the additional immunoglobulin (Ig)-like domain, which is critical in DNA binding. Due to AlkC's capability of 1mA and 3mC excision, it is proposed that AlkC is responsible for 1mA and 3mC repair *in vivo* when the cell lacks AlkB, which is critical in 1mA and 3mC lesion recovery in bacteria (Shi et al, 2018).

Scope of this work

The focus of this dissertation is on characterization of the HEAT-like repeat (HLR) alkylpurine DNA glycosylase AlkC, and its comparison to other HEAT-like repeat (HLR) superfamily proteins. Chapter II explores how AlkC specially recognizes and catalyzes cationic lesions through biochemical, structural, and bioinformatics analysis. The study reveals for the first time the distinct non-base-flipping mechanism AlkC uses for catalysis. *B. cereus* has methylpurine glycosylases AlkC and AlkD in addition to two AlkA homologs and one AAG homolog, raising the question why bacterial cells contain multiple alkylation-specific glycosylases with overlapping substrate specificities. Chapter III elucidates that AlkD has gained cellular function of excision of positively charged bulky YTM adducts besides removal of canonical BER substrates. Given that HLR superfamily proteins show distinct functions and substrate specificities, Chapter IV compares phylogenetic distribution, substrate preference and catalytic mechanisms of AlkC, AlkD, AlkF, AlkD2 families. Chapter V consists of a brief summary of some of the research regarding HLR superfamily and identifies directions that would benefit from future study.

CHAPTER II

SELECTIVE BASE EXCISION REPAIR OF DNA DAMAGE BY THE NON-BASE- FLIPPING DNA GLYCOSYLASE ALKC*

Abstract

Cellular metabolites and environmental toxins damage DNA by creating a variety of covalent DNA adducts that impair normal cellular processes and lead to heritable diseases, cancer, aging, and cell death (Friedberg et al, 2006a; Jackson & Bartek, 2009). Alkylation of nucleobase substituents and ring nitrogens is a major source of DNA damage that can directly inhibit replication and transcription or can degenerate to other forms of damage including abasic sites and strand breaks (Krokan & Bjørås, 2013; Wyatt & Pittman, 2006). The toxicity of these DNA lesions is the basis for the use of alkylating agents in cancer chemotherapy (Sedgwick, 2004).

*The work described in this chapter has been published in:

Shi, R., Mullins, E. A., Shen X-X, Lay, K. T., Yuen, P. K., David, S. S., Rokas, A., Eichman, B. F. (2018) Selective base excision repair of DNA damage by the non-base-flipping DNA glycosylase AlkC. *EMBO J.* 37(1): 63-74

DNA glycosylases preserve genome integrity and define the specificity of the base excision repair pathway for discreet, detrimental modifications, and thus the mechanisms by which they locate DNA damage is of particular interest. Bacterial AlkC and AlkD are specific for cationic alkylated nucleobases and have a distinctive HEAT-like repeat (HLR) fold. AlkD uses a unique non-base-flipping mechanism that enables excision of bulky lesions more commonly associated with nucleotide excision repair. In contrast, AlkC has a much narrower specificity for small lesions, principally N3-methyladenine (3mA). A crystal structure resembling a catalytic intermediate complex shows how AlkC uses unique HLR and immunoglobulin-like domains to induce a sharp kink in the DNA, exposing the damaged nucleobase to active site residues that project into the DNA. This active site can accommodate and excise N3-methylcytosine (3mC) and N1-methyladenine (1mA), which are also repaired by AlkB-catalyzed oxidative demethylation, providing a potential alternative mechanism for repair of these lesions in bacteria.

Introduction

To avoid the toxic effects of DNA alkylation and maintain integrity of their genomes, all organisms possess multiple repair mechanisms to remove the diversity of alkyl-DNA modifications. Nucleotide excision repair (NER) is the predominant mechanism to eliminate helix-distorting and bulky adducts, whereas small nucleobase modifications (e.g., methyl, etheno groups) are repaired by direct reversal or base excision repair (BER) pathways (Mishina et al, 2006; Reardon & Sancar, 2005; Sedgwick, 2004). N1-methyladenine (1mA) and N3-methylcytosine (3mC), predominant

in single-stranded (ss) DNA and RNA, are demethylated by the AlkB family of Fe(II)/ α -ketoglutarate-dependent dioxygenases. In contrast, N3- and N7-alkylpurines, the most abundant double-stranded DNA alkylation products, are removed from the genome by BER, whereby the damaged nucleobase is cleaved from the phosphodeoxyribose backbone by lesion-specific DNA glycosylases that hydrolyze the N-glycosidic bond. The resulting abasic site is nicked by apurinic/apyrimidinic (AP) endonuclease to generate a 3'-hydroxyl substrate for gap repair synthesis by DNA polymerase.

DNA glycosylases that remove alkylation damage are essential to eukaryotes, archaea, and bacteria (Birkeland et al, 2002; Chen et al, 1989; O'Connor & Laval, 1991; Riazuddin & Lindahl, 1978; Thomas et al, 1982). Bacteria often contain paralogs that eliminate diverse types of damage resulting from endogenous versus environmental sources. For example, in *E. coli*, the AlkA glycosylase is induced to remove a broad spectrum of alkylated and deaminated bases upon cellular exposure to alkylation agents, while the constitutively active Tag enzyme is highly specific for N3-methyladenine (3mA) formed endogenously (Bjelland et al, 1994; Bjelland et al, 1993; McCarthy et al, 1984; O'Brien & Ellenberger, 2004a; Sapparbaev et al, 1995). Damage recognition and excision by these and other glycosylases rely on a base-flipping mechanism, whereby the damaged nucleotide is rotated $\sim 180^\circ$ around the phosphate backbone and sequestered inside a nucleobase binding pocket on the protein surface (Brooks et al, 2013). This pocket contains residues that facilitate depurination by activating or stabilizing the nucleobase leaving group and/or electrostatically stabilizing an oxocarbenium intermediate that reacts with a water molecule to generate the abasic site product (**Figure 10A**) (Drohat & Maiti, 2014). The extrahelical orientation of the

substrate in the catalytic complex is stabilized by DNA-intercalating protein residues that fill the space generated by the missing base. Thus, the specificity of a DNA glycosylase for a particular substrate is achieved by a combination of duplex interrogation and complementarity of the nucleobase and the active site pocket. The bacterial AlkC and AlkD glycosylases, originally identified in *Bacillus cereus* (Bc), comprise a distinct superfamily of DNA glycosylase selective for positively charged N3- and N7-alkylpurines (Alseth et al, 2006; Dalhus et al, 2007). They specifically lack excision activity toward uncharged 1,*N*⁶-ethenoadenine or hypoxanthine nucleobases (Alseth et al, 2006), the common substrates of many previously characterized alkylpurine DNA glycosylases. Structures of BcAlkD revealed a new fold composed of tandem HEAT-like repeats (HLRs) that form a left-handed solenoid around the DNA duplex (Rubinson & Eichman, 2012; Rubinson et al, 2010; Rubinson et al, 2008). Unlike other glycosylases, AlkD does not use a base-flipping mechanism for damage recognition or base excision (Mullins et al, 2015b). Instead, the protein traps non-Watson-Crick base pairs in a sheared, base-stacked conformation with catalytic active site residues in direct contact with the deoxyribose, but not the nucleobase, of the damaged nucleoside. The lack of protein-nucleobase contacts enables AlkD to excise both major and minor groove lesions 3mA and N7-methylguanine (7mG). Similarly, by not confining the modified nucleobase inside a binding pocket, AlkD can excise bulky lesions, including pyridyloxobutyl (POB) adducts resulting from the carcinogen NNK (nicotine-derived nitrosamine ketone) (Rubinson et al, 2010), as well as N3-yatakemycinyladenine (YTMA) formed from the highly genotoxic bacterial natural product yatakemycin (YTM) (Mullins et al, 2017a;

Mullins et al, 2015b; Xu et al, 2012). Thus, AlkD excises a diverse spectrum of cationic lesions, including bulky lesions expected to be processed by NER.

AlkC from *Bacillus cereus* (BcAlkC) shares 15.8% identity and 33.7% similarity with BcAlkD and has been predicted to adopt the HLR architecture (Dalhus et al, 2007). However, in contrast to AlkD, AlkC has a strong preference for 3mA and 3mG, and only weak activity for 7mG (Alseth et al, 2006; Mullins et al, 2013). Given AlkD's unique non-base-flipping mechanism and activity for bulky lesions, we were interested to expand our understanding of this new DNA repair superfamily by determining the basis for AlkC's apparent limited substrate specificity. We carried out a comprehensive phylogenetic, biochemical, and structural comparison of AlkC and AlkD proteins and found that the majority of AlkC proteins contain an immunoglobulin (Ig)-like domain not yet observed in a DNA repair enzyme and that is essential for base excision activity in those enzymes. We also found that unlike AlkD, AlkC does not remove YTMA adducts either *in vitro* or *in vivo*. The crystal structure of *Pseudomonas fluorescens* (Pf) AlkC bound to damaged DNA revealed that the HLR and Ig-like domains wrap almost completely around the DNA duplex to recognize the damaged base pair from opposite major and minor grooves. Like AlkD, AlkC does not extrude the damaged nucleobase from the DNA helix. However, unlike AlkD, AlkC induces a sharp kink in the DNA and accesses the damage by inserting active site residues into the exposed base stack. The structure and supporting mutational analysis of base excision by AlkC provide a mechanistic basis for how AlkC selectively excises 3mA from DNA. Additionally, we show that AlkC's unique active site is also capable of robust excision activity for 3mC, which is normally repaired by AlkB-catalyzed oxidative demethylation. This work

provides a molecular basis for how a non-base-flipping glycosylase selects for discrete methylated bases, and describes an alternative mechanism for removal of 3mC in bacteria.

Results

Proteins within Two AlkC Subgroups Are Functionally Distinct from AlkD

Despite a predicted similarity in HLR secondary structure, AlkC and AlkD from *B. cereus* were previously shown to be phylogenetically and functionally distinct (Alseth et al, 2006; Mullins et al, 2013). To better understand the AlkC family of proteins and how they differ from AlkD, we obtained 779 AlkC and 764 AlkD orthologous sequences by performing an iterative PSI-BLAST search against the NCBI non-redundant protein database and then used them to reconstruct a phylogenetic tree. 93% of AlkC sequences were widespread among actinobacteria, bacteroidetes, firmicutes, and proteobacteria, and only rarely found in archaea and eukaryotes, largely consistent with a previous analysis (Backe et al, 2013). We found that AlkC proteins clustered into two distinct clades that we denote AlkC α and AlkC β (**Figure 10B**), both of which contain an HLR domain distinct from AlkD mainly within the N-terminal 60-80 residues. The AlkC β clade, which constitutes the majority (70%) of the total AlkC sequences, contains an additional ~100-residue putative immunoglobulin (Ig)-like fold at the C-terminus (**Figure 10C, Figure 11**).

The specificity of AlkC for 3mA has been characterized only from the *B. cereus* enzyme, which belongs to the AlkC α class (Alseth et al, 2006; Mullins et al, 2013). To determine whether this 3mA selectivity is a general characteristic of both AlkC α and

AlkC β proteins, we cloned and purified five AlkC α and two AlkC β orthologs and measured their ability to excise 3mA and 7mG nucleobases from N-methyl-N-nitrosourea (MNU)-treated genomic DNA, the major methylated products of which are 7mG (66%) and 3mA (8%) (Lawley, 1976). All seven AlkC proteins released the maximum amount of 3mA after one hour as judged by treatment of substrate with HCl, but exhibited only very low activity toward 7mG as compared to an AlkD control (**Figure 10D**). Consistent with this, both BcAlkC (α) and PfAlkC (β) excised 7mG from a defined oligodeoxynucleotide substrate after 24 hours, albeit with much slower kinetics than BcAlkD (**Figure 12A**). Thus, 3mA specificity seems to be a general property of all AlkC proteins, independent of the presence of the putative Ig-like domain.

We also found that AlkC's specificity for N3-alkyladenine adducts is limited to small adducts. Neither BcAlkC nor PfAlkC were able to cleave YTMA lesions for which AlkD exhibits robust activity (Mullins, 2017) (**Figure 12B**). We previously showed that in addition to AlkD excision of YTMA *in vitro*, *B. anthracis* cells lacking AlkD exhibited a sensitivity to YTM (Mullins, 2015b). We therefore tested growth of *alkC* deficient *B. anthracis* strains in the presence of YTM. Consistent with our *in vitro* results, YTM sensitivity of a $\Delta alkC$ strain did not differ from that of wild-type *B. anthracis*, nor did a $\Delta alkC\Delta alkD$ double mutant show additional sensitivity compared to $\Delta alkD$ (**Figure 12C**). Thus, substrate specificities of both α and β subgroups of AlkC are distinct from AlkD, suggesting that AlkC utilizes a different strategy to recognize damaged DNA despite its predicted structural similarity to AlkD.

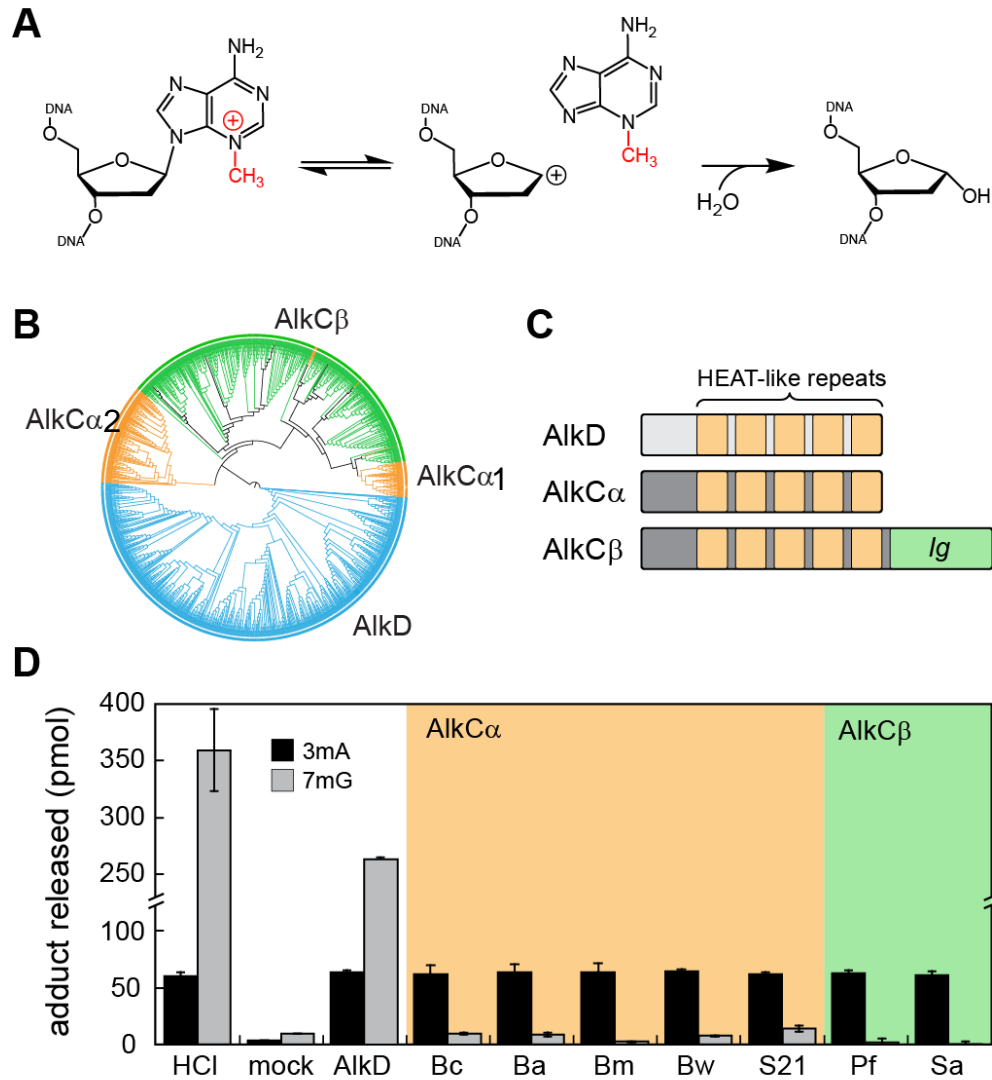


Figure 10. AlkC α and AlkC β are specific for N3-methyladenine (3mA). **A**, 3mA excision reaction catalyzed by AlkC. **B**, Phylogenetic tree of 779 AlkC (orange, green) and 764 AlkD (blue) protein sequences visualized using the iTOL web server (Letunic & Bork, 2011). **C**, Schematic of AlkC and AlkD protein domains. **D**, Release of 3mA (black bars) and 7mG (grey bars) from methylated genomic DNA after a 1-hour incubation with either HCl, no enzyme (mock), *Bacillus cereus* AlkD, or one of seven AlkC orthologs. Bc, *Bacillus cereus*; Ba, *Bacillus anthracis*; Bm, *Bacillus mycoides*; Bw, *Bacillus weihenstephanensis*; S21, *Shingobacterium sp. 21*; Pf, *Pseudomonas fluorescens*; Sa, *Streptomyces albus*. Values are the mean \pm SD (n=3 for each). HCl and no-enzyme controls provide upper and lower limits of 3mA and 7mG detection.

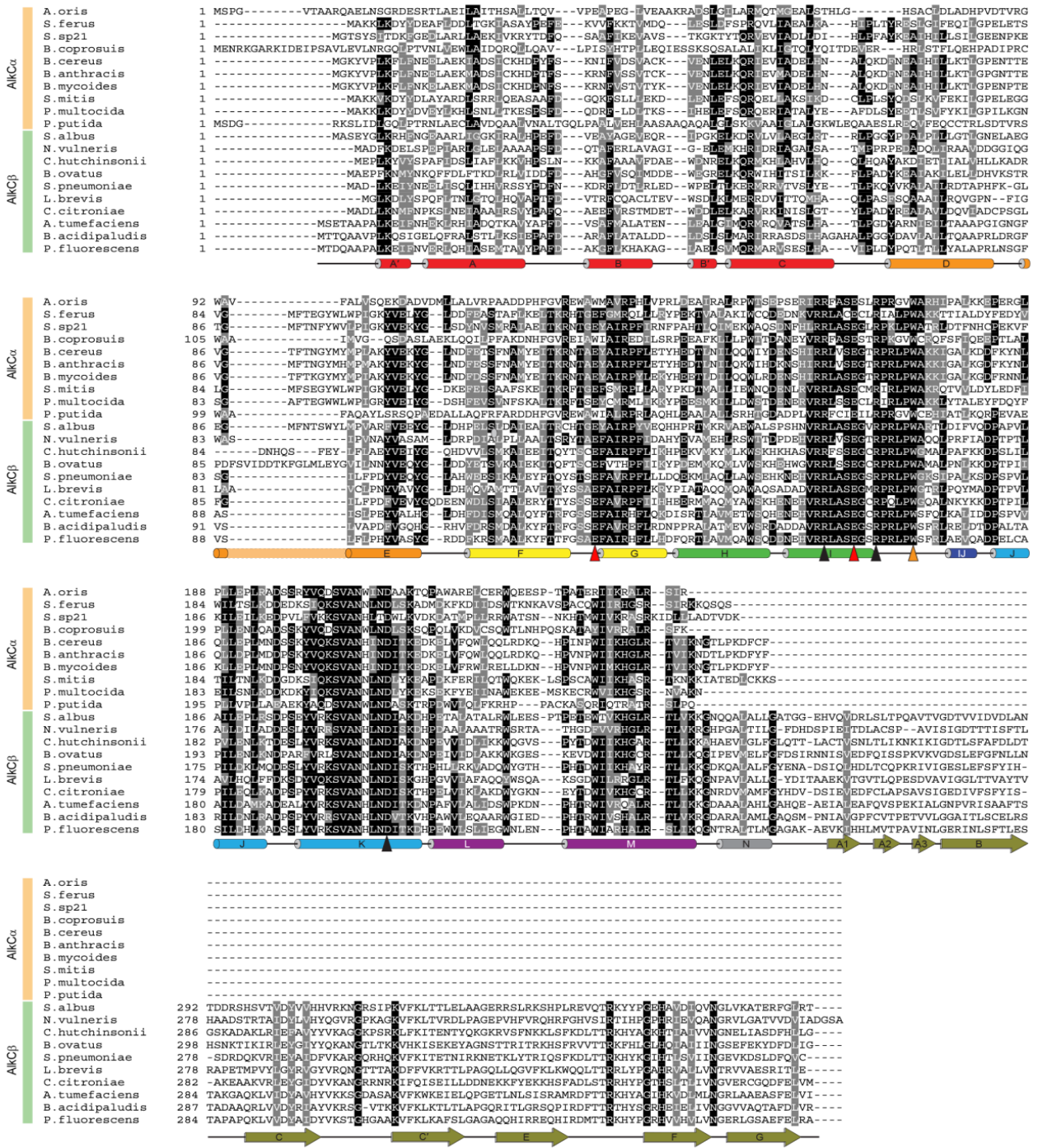


Figure 11. AlkC α and AlkC β sequence alignment. Twenty selected sequences from four representative phyla (Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria) were aligned using Clustal Omega and annotated using BoxShade (http://www.ch.embnet.org/software/BOX_form.html). Shaded residues have >50% sequence identity (black) and similarity (grey). Secondary structural elements identified from the PfAlkC crystal structure are shown below the sequences. Triangles designate PfAlkC residues important for base excision activity (red), nucleobase binding (orange) and for stabilizing the DNA backbone in the vicinity of the lesion (black).

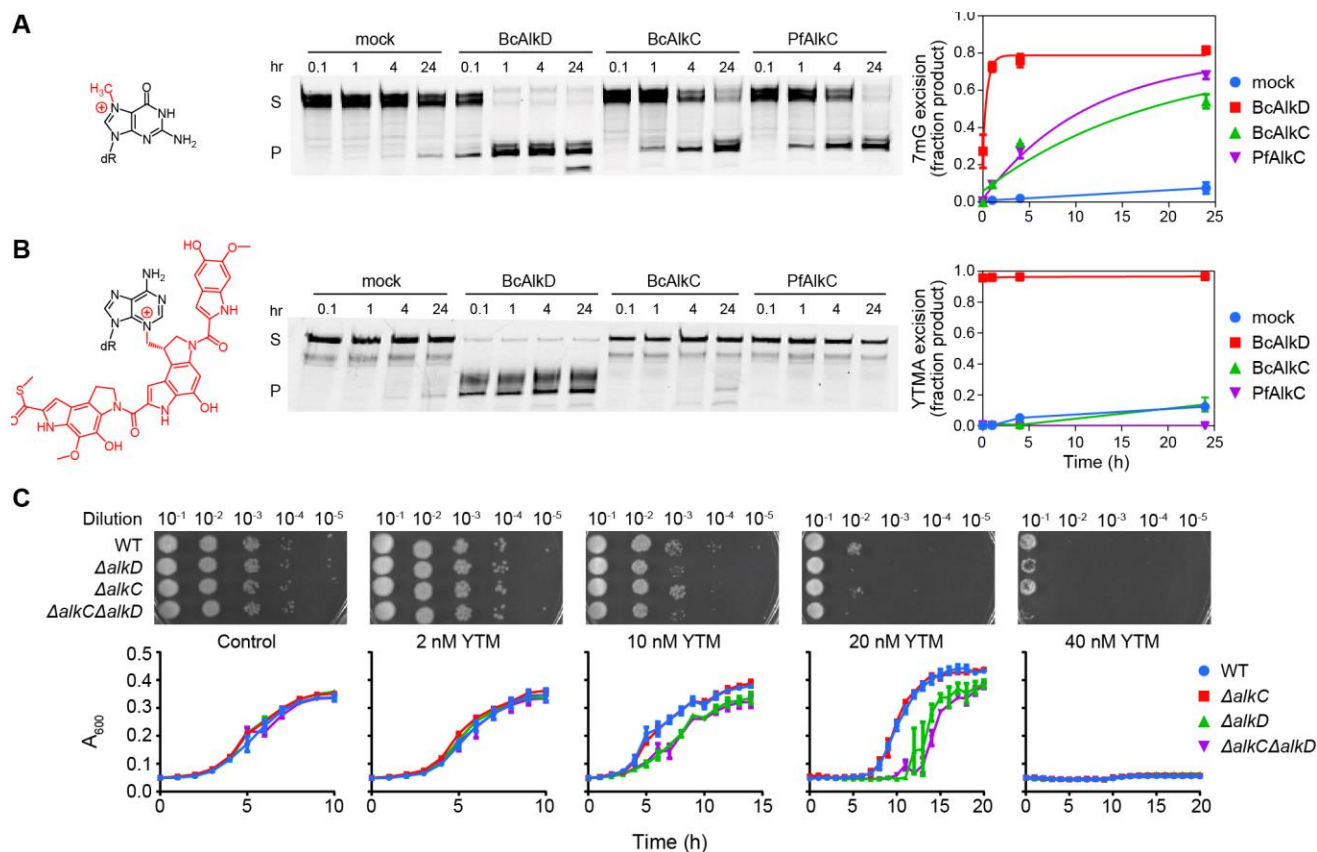


Figure 12. Substrate specificities of AlkC and AlkD. **A,B**, Excision activity of BcAlkD and PfAlkC against oligodeoxynucleotides containing 7mG (**A**) and YTMA (**B**). Representative denaturing polyacrylamide gels show substrates (S) and hydroxide-nicked products (P) as a function of time. Quantification of the data from three independent experiments is shown on the right (average \pm s.d.). The extra band below the 12-mer product in panel A corresponds to a nuclease contaminant in the AlkD preparation. The smearing in the AlkD-YTMA product band is a result of incompletely denatured GC-rich duplex DNA after hydroxide nicking. **C**, Growth of *B. anthracis* wild-type (blue), $\Delta alkC$ (red), $\Delta alkD$ (green), and $\Delta alkC\Delta alkD$ (purple) in the presence of varying concentrations of yatakemycin (YTM). *Bacillus anthracis* $\Delta alkC$ cells were generated and their resistance to YTM assayed as described for $\Delta alkD$ (Mike et al, 2014; Mullins et al, 2015b; Stauff & Skaar, 2009). Briefly, growth curves were obtained by growing cell cultures in the presence or absence of YTM and recording cell density every hour for 20 hours. Spot assays were performed by serial dilution of early-mid-log phase cells on LB plates prepared with or without YTM. Growth curves and spot assays were performed in triplicate.

AlkC Encircles and Bends Damaged DNA

To determine the structural basis for AlkC's preference for 3mA and whether a non-base-flipping mechanism is a common feature among HLR glycosylases, we determined a crystal structure of PfAlkC in complex with DNA containing 1'-aza-2',4'-dideoxyribose (1aR), which mimics the oxocarbenium ion intermediate formed upon nucleobase dissociation (**Figure 10A**) (Chu et al, 2011; Hollis et al, 2000; Schramm, 2011). The PfAlkC/1aR-DNA model was refined against X-ray diffraction data extending to 1.8 Å resolution to a crystallographic residual of 14.1% ($R_{\text{free}}=16.4\%$) (Table 2) and contained two crystallographically unique protein-DNA complexes, each with 361 of 369 amino acids and all 22 nucleotides clearly defined by the electron density. The structures of the two complexes in the asymmetric unit were virtually identical (r.m.s.d. = 0.42 Å for all atoms), differing only in the location of a statically disordered 5'-overhanging adenosine A1 on the undamaged strand that forms a crystal packing contact with the same nucleotide in an adjacent protein/DNA complex.

Table 2. Data collection and refinement statistics

	SeMet-AIkC/THF-DNA	AIkC/1aR-DNA
Data collection		
Space group	$P2_12_12_1$	$P6_1$
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	80.6, 94.9, 134.0	198.4, 198.4, 60.2
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0
Resolution (Å)	50.00–2.40 (2.49–2.40) ^a	100.00–1.80 (1.86–1.80)
R_{sym}	0.109 (0.429)	0.092 (0.490)
Avg. $I/\sigma I$	23.0 (5.8)	24.8 (5.3)
Completeness (%)	99.7 (99.8)	99.1 (97.1)
Redundancy	9.8 (9.7)	9.7 (9.1)
Wilson <i>B</i> -factor (Å ²)	25.2	19.8
Refinement		
Resolution (Å)	40.42–2.40 (2.46–2.40)	49.29–1.80 (1.82–1.80)
No. reflections	40,519 (2,602)	124,478 (3,819)
R_{work}	0.168 (0.219)	0.141 (0.177)
R_{free}^b	0.225 (0.333)	0.164 (0.202)
No. atoms ^c		
Protein	5,685	5,718
DNA	870	870
Solvent ^d	573	1,383
Avg. <i>B</i> -factors ^{c,e} (Å ²)		
Protein	25.6	23.0
DNA	25.6	25.1
Solvent ^d	29.2	37.5
R.m.s. deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	0.967	1.059
Ramachandran distribution (%)		
Favored	97.6	97.6
Allowed	2.4	2.2
Disallowed	0.0	0.1

^a Statistics for the highest resolution shell are shown in parentheses.

^b R_{free} was determined from the 5% of reflections excluded from refinement.

^c Riding hydrogen atoms were not included in no. atoms or avg. *B*-factors.

^d In addition to water molecules, values for solvent include one PEG 4000 molecule in the THF structure and two Na⁺ ions, two pentaerythritol propoxylate, two glycerol, and three MES molecules in the 1aR structure.

^e Equivalent isotropic *B*-factors were calculated in conjunction with TLS-derived anisotropic *B*-factors.

The PfAlkC crystal structure confirmed the presence of distinct HLR and Ig-like domains, which together wrap almost completely around a highly bent DNA duplex (**Figure 13**). As a consequence of its interactions with the HLR and Ig-like domains, the DNA is sharply bent at the 1aR lesion by 60° away from the minor groove and the minor groove widened by 7 Å. The HLR domain (helices $\alpha A'$ - αN) is similar to that of AlkD in that it contains an N-terminal α -helical bundle ($\alpha A'$ - αC) followed by five HLRs— αD - αE , αF - αG , αH - αI , αJ - αK , and αL - αM —that together form a left-handed solenoid that wraps around one-half of the DNA duplex from the minor groove side (**Figure 13A,B**). The N-terminal α -helical bundle contacts the backbone of the undamaged strand from the minor groove using polar side chains and the helix dipoles of helices $\alpha A'$ and αC , which point directly at thymines T6 (opposite the 1aR) and T7, respectively (**Figure 13A**). All HLRs except HLR1 directly contact the DNA, primarily along the phosphate backbones. The C-terminal helices of each HLR adorn basic and polar side chains that bind either strand from the minor (HLR2, HLR3) or major (HLR4, HLR5) groove sides (**Figure 13C**). HLR2 and HLR3 also contain conserved glutamate side chains that directly contact the 1aR. Between HLR3 and HLR4, an 11-residue loop/ α -helix insertion (αJ -loop) not found in AlkD penetrates the minor groove at the lesion. The HLR domain is tethered by a 9-residue linker to the Ig-like fold, which is composed of nine β -strands ($\beta A1$ - βG) and an extended loop that contacts the DNA from the major groove side (**Figure 13A**). Both domains form a positively charged, highly conserved concave surface that engulfs the DNA (**Figure 13B**). Extensive polar and van der Waals contacts are formed with the DNA backbone 5' to the 1aR and along the entire undamaged strand (**Figure 13C**).

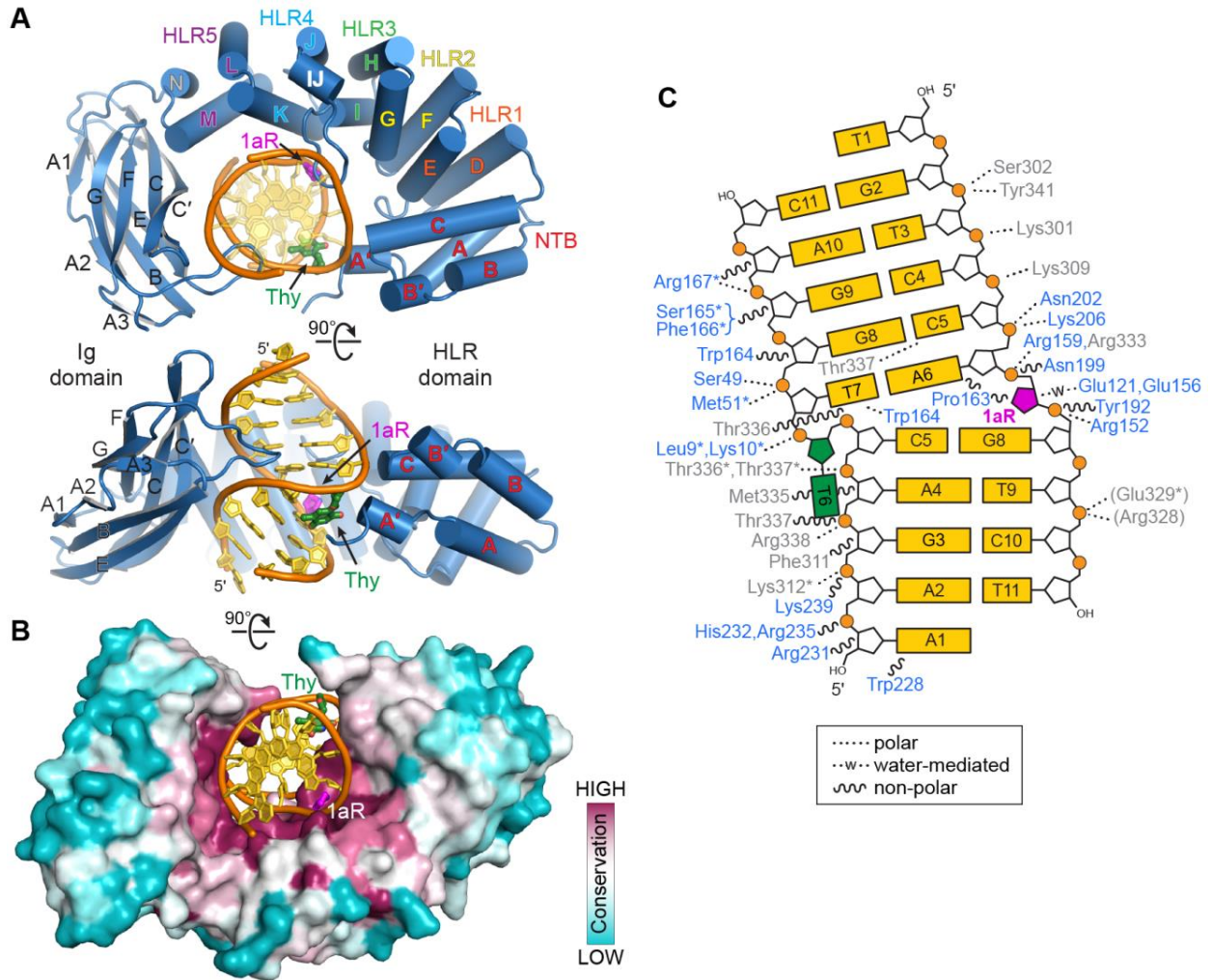


Figure 13. AlkC encircles damaged DNA. **A**, Two orthogonal views of the PfAlkC/1aR-DNA complex crystal structure. The protein is colored blue, DNA gold, 1'-aza-2',4'-dideoxyribose (1aR) magenta, and opposite thymine green. **B**, AlkC sequence conservation (purple, high; cyan, low) superposed onto the protein surface. **C**, Schematic illustration of AlkC-DNA interactions. Dashed and wavy lines denote polar and non-polar interactions, respectively. Residues from HLR and Ig-like domains are blue and grey, respectively. Contacts to the backbone are marked with an asterisk, and symmetry-related contacts are in parentheses.

The AlkC Ig-like Domain is a Unique DNA Binding Motif in Bacteria

Ig-like domains are prevalent in both bacteria and eukaryotes as a generic scaffold (Bodelón et al, 2013; Bork et al, 1994; Halaby & Mornon, 1998; Halaby et al, 1999) and are important for sequence-specific DNA binding in some eukaryotic

transcription factors (Becker et al, 1998; Bravo et al, 2001; Chen et al, 1998b; Cho et al, 1994; Cramer et al, 1997; Lamoureux et al, 2002; Nagata et al, 1999; Rohs et al, 2010; Rudolph & Gergen, 2001; Tahirov et al, 2001). To our knowledge, no bacterial Ig-like domains have been reported to bind DNA or to function in DNA repair. The Ig-like domain of PfAlkC is composed of a nine-strand (β A1- β G) antiparallel β -sandwich, in which β A1, β A2, β A3, β B and β E form one β -sheet packed against a second sheet formed by strands β C, β C', β F and β G (**Figure 13A**). The topology is consistent with the C2 subtype (**Figure 15B**) but with very low sequence similarity, a highly kinked β A strand (β A1- β A2- β A3), and a longer β C' strand than other C2-type Ig-like domains (Bodelón et al, 2013; Halaby et al, 1999). The closest structural homolog of AlkC's Ig-like domain as judged by a Dali search (Holm & Laakso, 2016) is found in a bacterial β -glucosidase and has been proposed to contribute to substrate binding and dimerization of that enzyme (McAndrew et al, 2013).

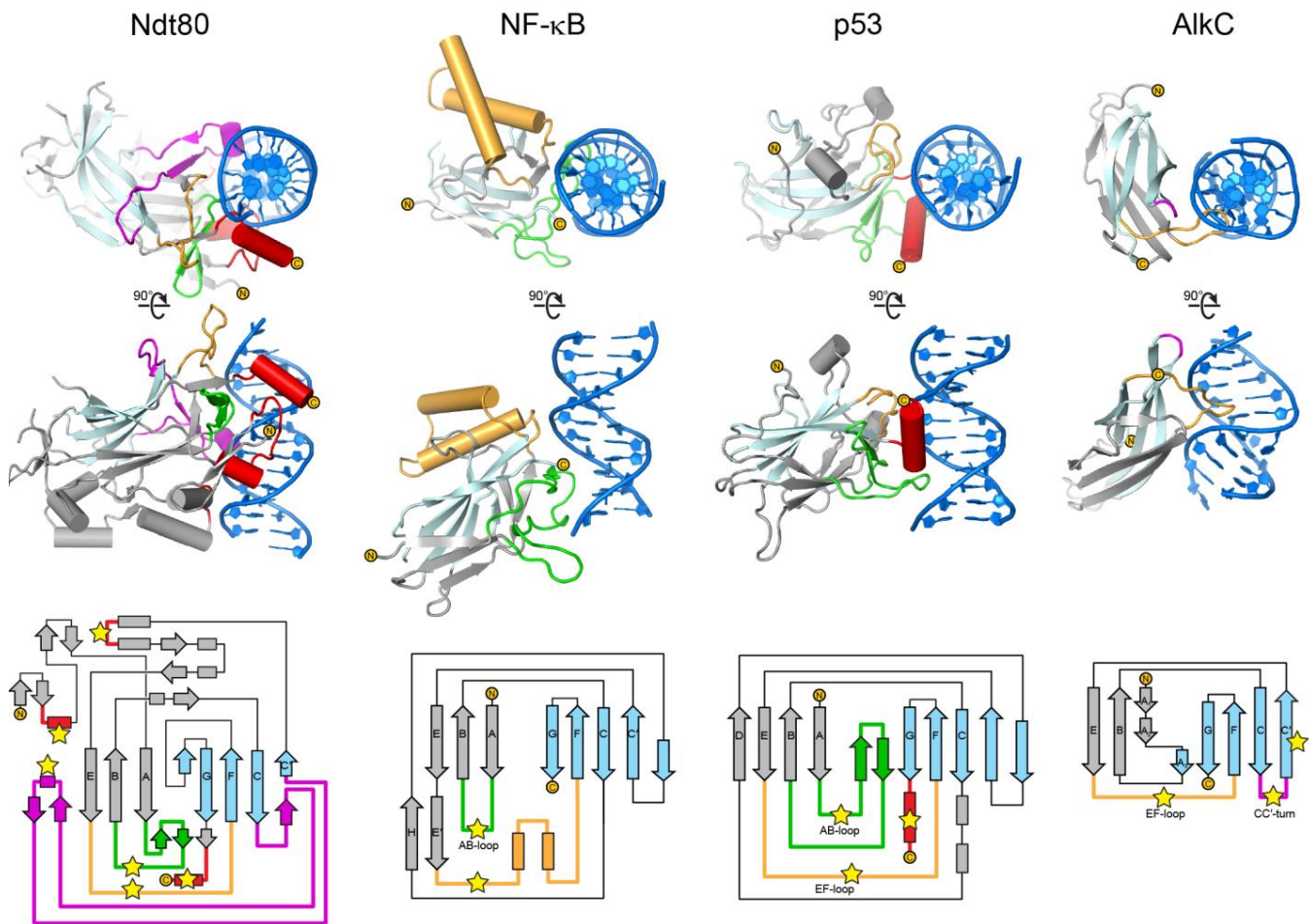


Figure 14. Comparison of DNA-binding Ig folds. The top two rows show orthogonal views of DNA-bound structures of Ig-like domains from three eukaryotic transcription factors Ndt80 (PDB 1MNN), NF- κ B (PDB 1A3Q), and p53 (PDB 1TUP), compared to that of PfAlkC (this work). Topology diagrams of each structure is shown at the bottom and are colored the same as the crystal structures. The two β -sheets within the Ig core are colored grey (A, B, D, E, E', and H) and pale cyan (C, C', F, and G). Regions contacting the DNA are colored green (AB-loop), magenta (CC'-loop), orange (EF-loop), and red. Points of contact with the DNA in are marked with yellow stars in the topology diagram.

The Ig-like domains of eukaryotic transcription factors mediate DNA binding through the loops between strands β A- β B and β E- β F (AB- and EF-loops) and the C-terminal tail regions, all of which emanate from the β -sandwich core (**Figure 14**) (Becker et al, 1998; Bravo et al, 2001; Chen et al, 1998a; Cho et al, 1994; Cramer et al, 1997;

Lamoureux et al, 2002; Nagata et al, 1999; Rohs et al, 2010; Rudolph & Gergen, 2001; Tahirov et al, 2001). Although we have no evidence for sequence-specific binding by AlkC, AlkC's Ig-like domain also mediates DNA binding by the EF-loop, which in PfAlkC contains a conserved Met-Thr-Thr-Arg motif (residues 336-338) that contacts nucleobases on both strands in the major groove (**Figure 15A**). Side- and main-chain

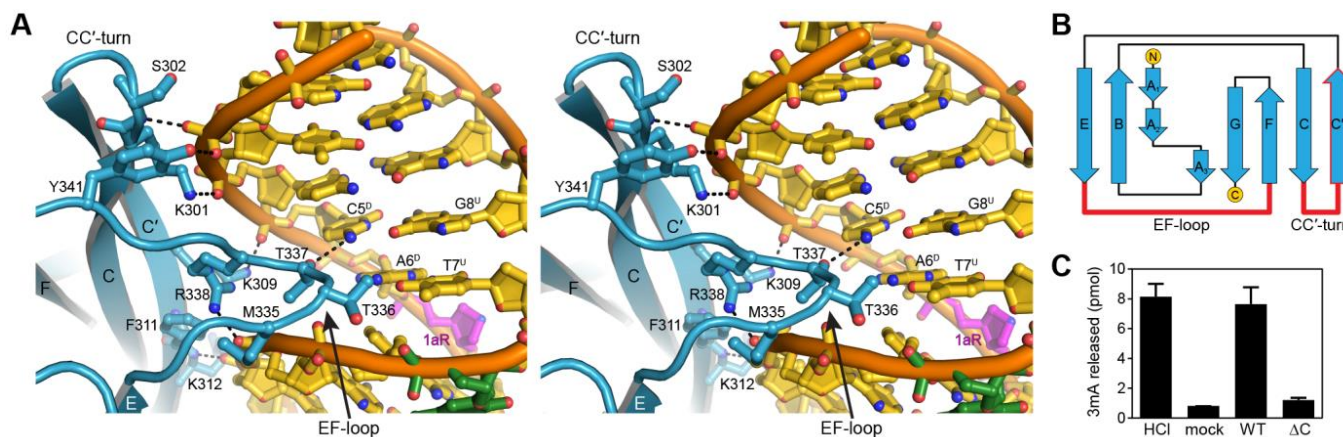


Figure 15. The Ig-like domain is important for AlkC β function. **A**, Stereoview of the Ig-like domain (blue) interactions with the major groove of DNA (gold). DNA interacting side chains are shown as sticks, and hydrogen bonds are shown as dashed lines. Superscripts in nucleotide labels refer to damaged (D) and undamaged (U) strands. **B**, Topology of AlkC β Ig-like domain. Regions contacting the DNA are highlighted in red. **C**, Release of 3mA from methylated genomic DNA after a 5-min incubation with either HCl, no enzyme (mock), PfAlkC (WT), or PfAlkC Δ C (Δ C).

groups within this motif interact with the phosphoribose groups on the undamaged strand immediately 5' to the orphaned thymidine T6, and Thr337 is engaged with cytosine C5 on the damaged strand. In contrast to the eukaryotic transcription factors, AlkC's AB-loop does not contact DNA, despite its orientation toward the DNA (**Figure 14**). Instead, PfAlkC makes unique DNA binding interactions using the β C' strand and the preceding CC'-turn (**Figure 15A,B**). Side chains of Lys309, Phe311, and Lys312 on the β C' strand interact with the backbones of both strands, and Lys301 and Ser302 on

the CC'-turn form polar interactions with phosphates on the 5'-side of the lesion strand. As a consequence of these contacts to both DNA strands, the Ig-like domain in AlkC plays an important role in stabilizing the kinked DNA conformation toward the major groove.

Consistent with a functional importance of the Ig-like domain, deleting it from PfAlkC (PfAlkC Δ C, **Figure 19A,B**) abrogated base excision activity (**Figure 15C**), likely owing to a severe decrease in DNA binding affinity relative to the wild-type protein (**Figure 19C**). The dependence of PfAlkC activity on the Ig-like domain is interesting given the absence of this domain in the fully functional AlkC α proteins, which suggests that the AlkC α HLR domain contains a structural feature that compensates for DNA binding in the absence of the Ig-like domain. Perhaps the most distinguishing feature of the AlkC α HLR primary structure is an 8-10 residue insertion within helix α E of HLR1 (**Figure 11**). Whereas HLR1 does not contact the DNA in the PfAlkC structure, a homology model of BcAlkC suggests that the extra residues endow AlkC α proteins with a novel DNA binding interaction to the damaged strand (**Figure 16**). Together with potential additional interactions between the N-terminal helical bundle and the undamaged strand, these extra HLR1-DNA contacts would help stabilize the DNA kink. Although the precise manner in which AlkC α proteins compensate for the lack of the Ig-like domain has yet to be determined, it is clear that DNA binding by the Ig motif is essential for AlkC β function.

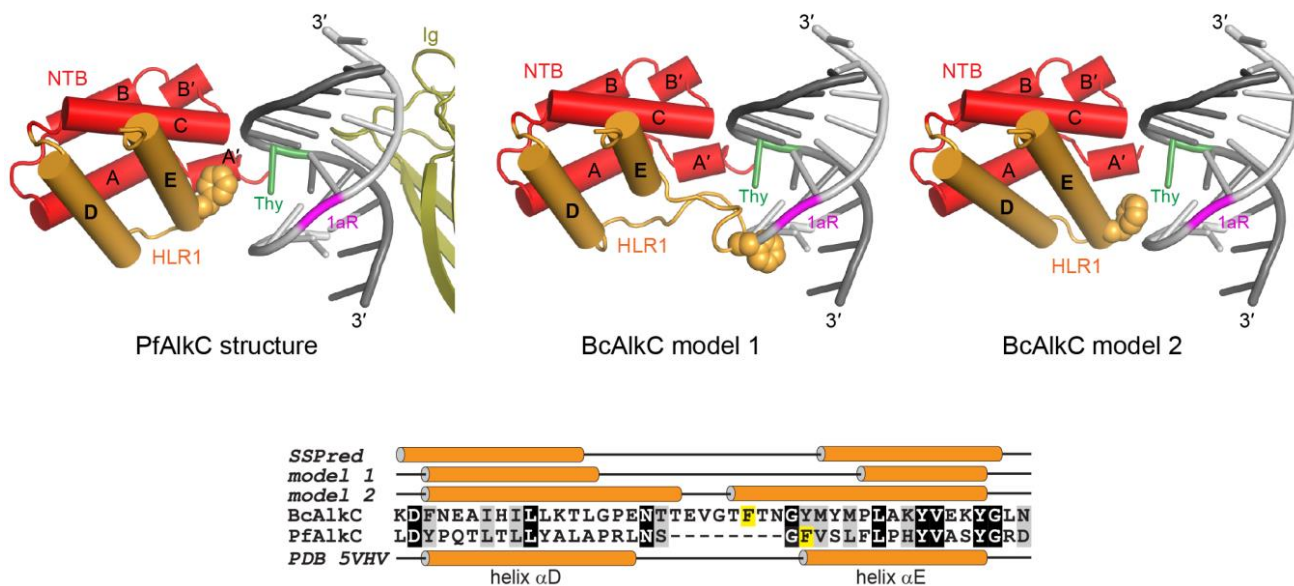


Figure 16. AlkC α proteins contain an insertion that may stabilize DNA bend in the absence of Ig-like domain. Two homology models for BcAlkC (center and right) are shown superimposed against the 1aR-DNA from the PfAlkC/1aR-DNA crystal structure (left). Only the N-terminal helical bundle (NTB, red), HEAT-like repeat 1 (HLR1, orange), and Ig-like domain (olive) are shown for clarity. A phenylalanine side chain at the N-terminus of helix α E in PfAlkC and in the 8-residue insertion of BcAlkC is shown in spheres and highlighted yellow in the sequence alignment at the bottom. The secondary structural elements from the three models and from a secondary structure prediction of BcAlkC are shown against the sequences for each protein. The homology models were generated in SWISS-MODEL (<https://swissmodel.expasy.org/>) using either the sequence alignment shown in Fig. S1 (model 1) or generated by SWISS-MODEL (model 2). In both models, the insertion makes contacts to the 1aR strand, either as a loop (model 1) or as an N-terminal extension to helix α E (model 2).

AlkC Inserts Its Active Site into the DNA Duplex in lieu of Base Flipping

Contacts between the HLR domain and the lesion help to explain AlkC's specificity for 3mA and its mechanism of base excision. At the DNA bend, both the 1aR lesion and its opposing thymidine on the undamaged strand (T6U) are displaced into the widened minor groove (**Figure 18A**). The 1aR is slightly rotated toward the protein to make electrostatic interactions and water-bridging hydrogen bonds to the carboxylate side chains of Glu121 and Glu156 (**Figure 18B**). Contrary to base-flipping glycosylases, there are no protein residues filling the void left by the missing nucleobase on the

damaged strand. On the undamaged strand, the opposing thymidine T6U is displaced into the minor groove as a single nucleotide bulge as a consequence of the sharp kink in the DNA. The thymine base is no longer stacked with the flanking nucleotides and is tethered to the damaged strand by a hydrogen bond to guanine G8D and via Na⁺ coordination to T9D and C10D nucleobases (**Figure 18A** and **Figure 17A**). The minor

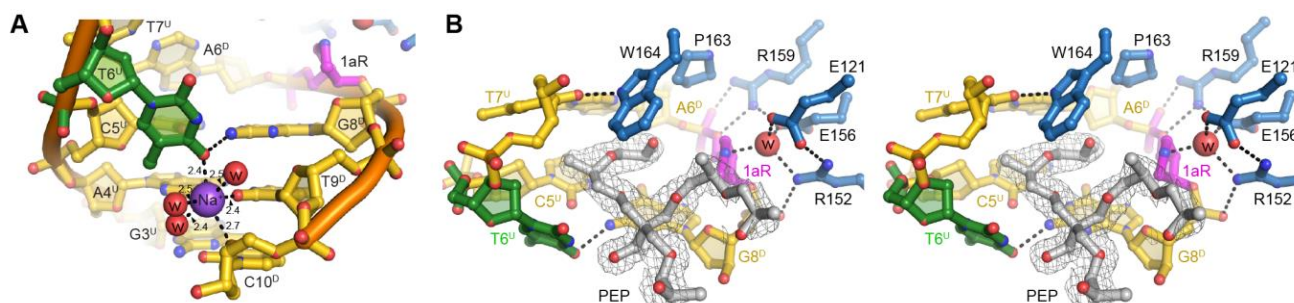


Figure 17. Crystallographic features of the active site. **A**, Na⁺ ion coordination in the 1aR-DNA structure. Na⁺ and water oxygens are shown as purple and red spheres, respectively. 1aR is magenta and the opposite thymidine T6 is green. Superscripts in nucleotide labels refer to the 1aR-containing, damaged (D) strand or the opposite, undamaged (U) strand. Hydrogen bonds are shown as dashed lines with interatomic distances in Ångstroms. **B**, Stereo-views of the PfAlkC/1aR-DNA active site. Active site residues are blue and DNA is gold/magenta/green. Pentaerythritol propoxylate (PEP) sequestered from the solvent is in silver and superimposed against composite omit electron density contoured at 1 σ . The putative catalytic water is shown as a red sphere. One arm of the PEP projects into the DNA kink between the 1aR and the flanking base pairs, and thus may limit rotation of the 1aR ring back toward the DNA. The other two polymeric PEP arms project outward to solvent. The fourth PEP arm is not present; its terminal hydroxyl group forms a hydrogen bond to the displaced T6^U thymine base. The position of the thymine is thus affected by the presence of the PEP in addition to its coordination by the Na⁺ ion. However, even in the absence of these stabilizing contacts, the T6^U thymidine would be displaced into the minor groove as a result of the kink in the DNA. The space occupied by the PEP molecule would be occupied by the excised base both before and immediately after cleavage.

groove is held in this opened conformation by the AlkC-specific α J-loop, at the tip of which Pro163 and Trp164 form van der Waals and hydrogen bonding contacts to the edges of the A6D•T7U base pair just 5' to the lesion (**Figure 18A**). Residues on the α J-

loop and at the N-terminal end of helix α G form a cavity at the hinge point in the DNA that is appropriately shaped for a 3mA nucleobase (**Figure 18C**). In the crystal structure, this cleft is filled by a pentaerythritol propoxylate solvent molecule from the crystallization buffer (**Figure 17B**), but presumably would be occupied by the modified nucleobase prior to and immediately following nucleobase excision (**Figure 10A**).

To gain a better idea of how the protein selects for 3mA, we modeled the nucleotide in the active site using 1aR as a guide (**Figure 18C**). Holding the deoxyribose ring in a fixed position, the 3mA nucleobase could adopt only a limited range of positions that varied within a 15° torsional rotation about the N-glycosidic bond, constrained by π -stacking against either its 3'-neighbor guanine G8D on one side or the protein surface on the other. Against the protein, the N1-C2 edge of the 3mA nucleobase abuts the indole side chain of Trp164 to form a stabilizing cation- π interaction. In this conformation, the N3-methyl group resides in a pocket formed by Phe122 and Glu121 (**Figure 18B**). These contacts to the damaged nucleobase explain AlkC's inability to excise bulky minor groove Y-TMA lesions. Moreover, in this model, the Hoogsteen face of the purine ring is snugly nestled between the flanking nucleobases at the DNA kink, suggesting that purine N7-adducts would be sterically disfavored from this collapsed major groove, providing a rationale for AlkC's low activity toward 7mG.

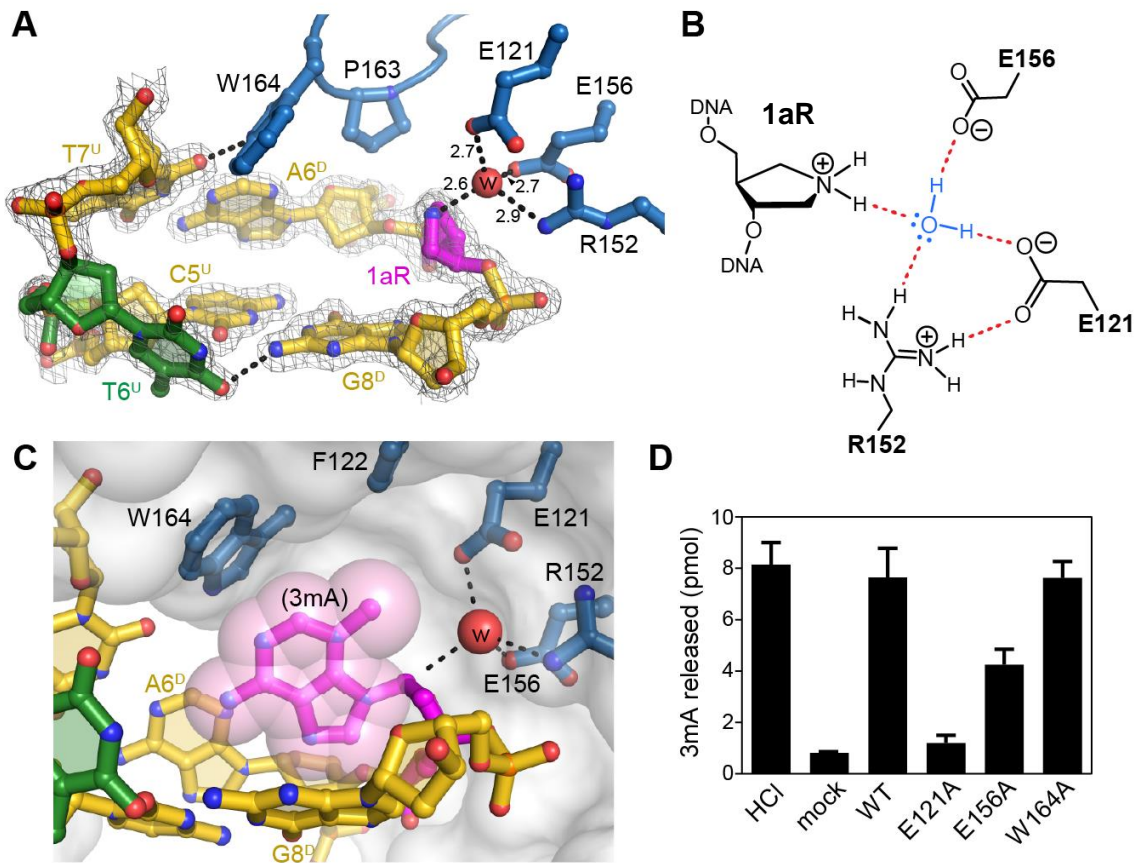


Figure 18. AlkC inserts its active site into the DNA. **A**, Close-up view of the AlkC active site (blue) bound to 1aR-DNA (gold). The 1aR and opposite thymine are magenta and green, respectively. Water is shown as a red sphere and hydrogen bonds are depicted as dashed lines. Composite omit electron density contoured to 1σ is superimposed against only the DNA for clarity. Superscripts in nucleotide labels refer to the 1aR-containing, damaged (D) strand or the opposite, undamaged (U) strand. The pentaerythritol propoxylate molecule that occupies the active site has been omitted for clarity. **B**, Schematic of the alignment of a catalytic water molecule (blue) against the 1aR oxocarbenium mimetic by AlkC active site residues. Hydrogen bonds are shown in red. **C**, A hypothetical model for AlkC bound to a 3mA-DNA substrate was generated by superimposing the 3mA deoxyribose ring onto that of 1aR in the crystal structure, followed by rotating about the 3mA χ (N-glycosidic bond) torsion angle to maximize van der Waals interactions. The solvent accessible surface of AlkC is shown as a transparent white envelope. **D**, Release of 3mA from methylated genomic DNA after a 5-min incubation with either HCl, no enzyme (mock), PfAlkC (WT), E121A, E156A, or W164A.

The position of the 1aR deoxyribose in the crystal structure provides a model for catalysis of base excision by AlkC. The N1 nitrogen of 1aR, which mimics the anomeric C1' carbon in the oxocarbenium intermediate, is positioned perfectly for nucleophilic attack by the water molecule held in place by Glu121, Arg152, and Glu156 (**Figure 18A,B**). These putative catalytic protein-DNA contacts are made without flipping the modified nucleotide into an active site pocket. In fact, the nucleotide is sterically constrained from further rotation around the DNA backbone. A network of alternating charges between Glu121-Arg152-Glu156-Arg159-Asp203 spans the backbone in a 3'-5' direction across the damaged nucleotide. Arg152 and Arg159 electrostatically stabilize the two phosphates flanking the 1aR and hold the carboxylate side chains of Glu121 and Glu156 in close proximity to the 1aR sugar ring (**Figure 17B**). Thus, Glu121 and Glu156 are positioned to electrostatically stabilize the positive charge that develops on the deoxyribose during base excision and to orient and deprotonate the catalytic water nucleophile for attack of the anomeric C1' carbon, without the need to flip the damaged nucleotide out of the DNA.

To validate this region of the protein as the active site, we tested the contribution of Glu121, Glu156, and Trp164 to catalysis by substituting individual residues with alanine and comparing the ability of the mutants to excise 3mA from methylated genomic DNA. We verified that the mutations did not compromise protein integrity or DNA binding (**Figure 19**). The Glu121Ala substitution abrogated 3mA excision activity and Glu156Ala significantly decreased 3mA excision relative to wild-type (**Figure 18C**), consistent with a catalytic role for these residues. In contrast, removal of the indole side chain from Trp164 had no effect on the total amount of 3mA excised after 5 minutes,

suggesting that this residue either does not participate in catalysis or has a less detectable role in this assay. These structural and biochemical data clearly designate the region surrounding the 1aR as the active site and demonstrate that like AlkD, the AlkC protein uses a non-base-flipping mechanism to access its substrate.

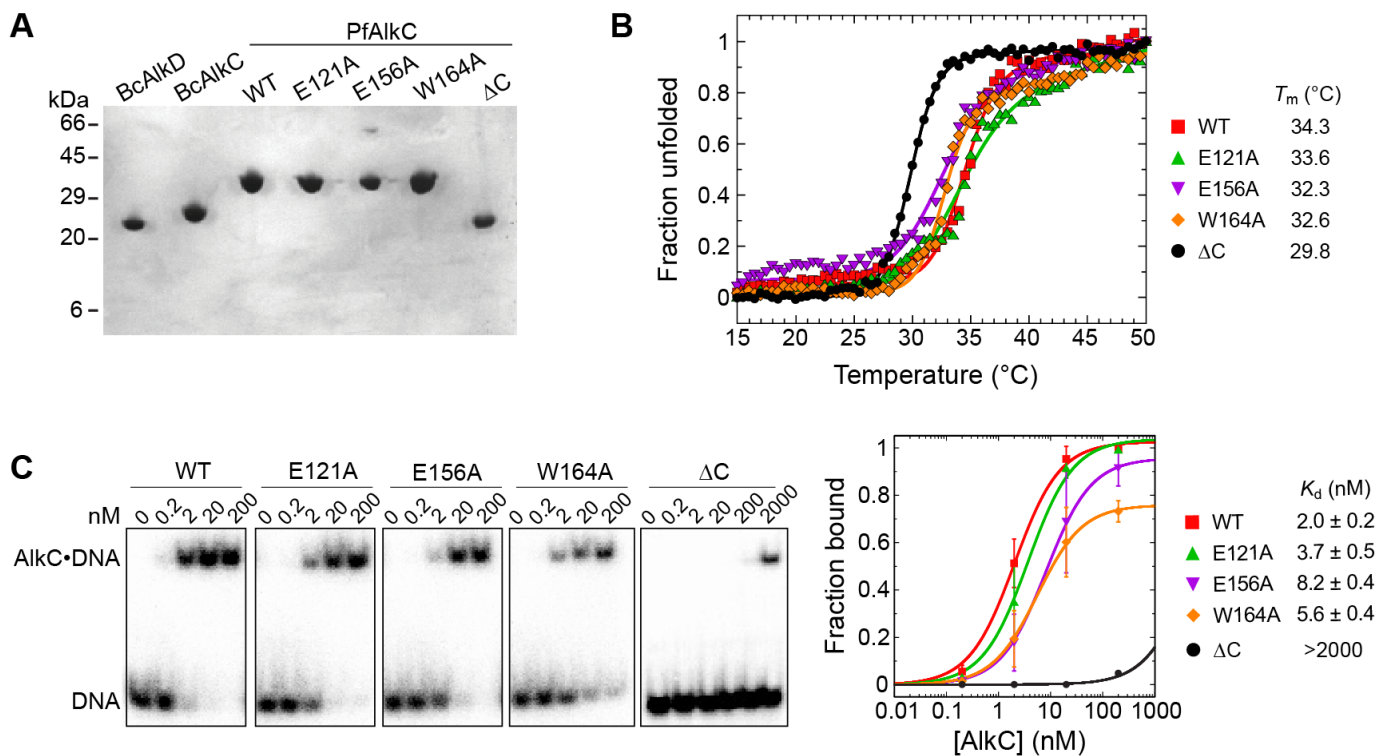


Figure 19. Characterization of PfAlkC proteins used in this study. **A**, SDS-PAGE of purified proteins, stained with Coomassie Blue. **B**, Thermal denaturation of wild-type and PfAlkC mutants monitored by circular dichroism. The fraction of unfolded protein was expressed as the normalized molar ellipticity at 222 nm. Melting temperatures (T_m) were derived by fitting the data to the equation $(fraction\ unfolded) = 1 / (1 + e^{(T_m - T)/k})$, where k denotes the cooperativity of the transition. Although PfAlkC ΔC has a lower T_m than the full-length proteins, it is properly folded at 21 °C at which biochemical assays were performed (**Figure 19**). **C**, Electrophoretic mobility shift assay for PfAlkC mutants binding to 1 nM THF-DNA of the sequence ^{32}P -d(GACCACTACTACT(THF)ATTCCTAACAAC) / d(GTTGTTAGGAAT(T)AGTGTAGTGGTC) in 25 mM HEPES (pH 7.5), 50 mM KCl, 5 mM DTT, 5% glycerol, and 0.05 mg/mL BSA at 21 °C for 30 min. Concentrations of PfAlkC are shown at the top of the representative gels. Electrophoretic separation was carried out on a Novex™ TBE gel (ThermoFisher Scientific). Quantitation is plotted on the right, in which each value is the mean \pm SD ($n=3$). Equilibrium dissociation constants (K_d) were extracted by fitting the data to a two-state binding model.

DNA glycosylase product complexes are often remarkably similar in conformation to their substrate or intermediate structures, and consequently these enzymes are typically inhibited by their abasic site products, which we recently showed to be true for the non-base-flipping glycosylase AlkD (Brooks, 2013; Mullins, 2017; Mullins, 2015b). In an effort to trap an AlkC product complex, we crystallized PfAlkC with DNA containing a tetrahydrofuran (THF) abasic site mimetic. The PfAlkC/THF-DNA crystal structure was refined to a crystallographic residual of 16.8% ($R_{\text{free}}=22.5\%$) against X-ray data extending to 2.4 Å (**Table 2, Figure 20**). The asymmetric unit of the THF structure contained two protein-DNA complexes with HLR and Ig-like domains wrapping around highly bent DNA duplexes as in the 1aR structure. However, unlike the virtually identical protein-DNA complexes in the 1aR structure, each protein-DNA complex in the THF structure differed significantly in the relative position of their Ig-like and HLR domains as well as in the conformation of the DNA, with a 60° bend angle in one and an 85° bend in the other (**Figure 20G**). Most notably, there was a substantial difference in the positions of THF and 1aR bound to AlkC. The THF moieties were slipped out of the DNA helix as single nucleotide bulges, each adopting a distinct conformation and a set of contacts to AlkC outside of the active site (**Figure 20D-F**). As a consequence, adenine A6D, immediately 5' to the THF on the damaged (D) strand, formed an opportunistic base pair with thymine T6U opposite THF on the undamaged (U) strand, leaving T7U (instead of T6U) unpaired. The mismatched A6D•T6U base pair was observed in both Watson-Crick and Hoogsteen orientations in the two complexes in the asu (**Figure 20F**). Because neither THF contacted the active site residues, similar to that observed in a non-catalytic complex of THF-DNA bound to AlkD (Rubinson et al, 2010), we conclude

that the PfAlkC/THF-DNA structure is not representative of the true PfAlkC product complex. Rather, the structural heterogeneity observed in this structure highlights a key aspect of the non-base-flipping mechanisms used by AlkC and AlkD. These enzymes lack the intercalating side chain used by base-flipping enzymes to plug the gap left in the DNA as a result of the everted or excised base, and thus AlkC and AlkD do not directly stabilize the conformation of the abasic DNA on their own. Rather, we previously showed that the excised nucleobase stabilizes intermediate and product complexes by remaining stacked in the DNA duplex from its position in the substrate (Mullins et al, 2015b). Attempts to trap ternary PfAlkC/1aR-DNA/3mA and PfAlkC/THF-DNA/3mA complexes containing free 3mA nucleobase were unsuccessful, likely owing to the lack of base stacking at the sharp kink in the DNA, and suggesting that the nucleobase readily dissociates following N-glycosidic bond cleavage.

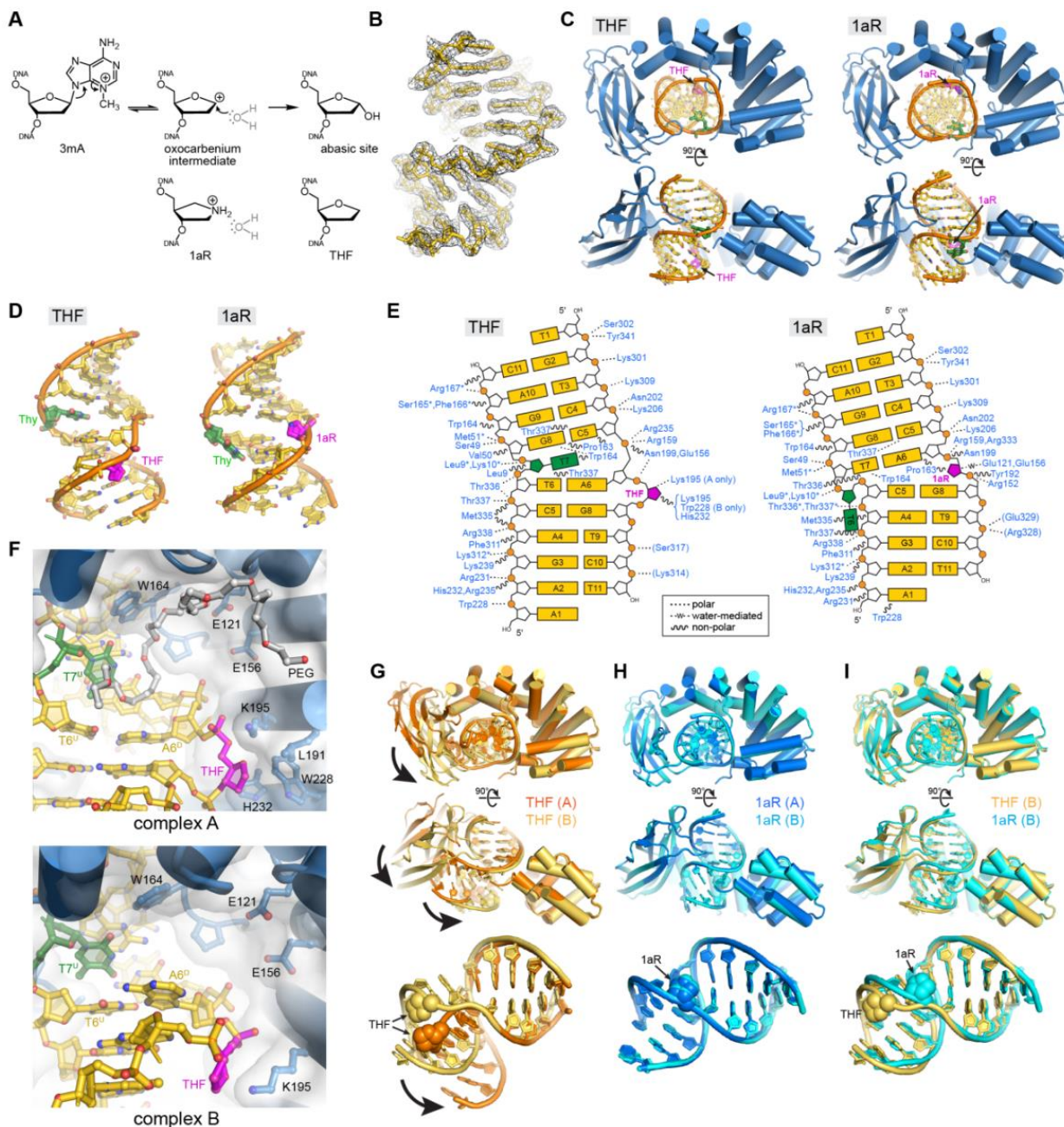


Figure 20. Structure of the PfAlkC/THF-DNA complex and its comparison to PfAlkC/1aR-DNA. **A**, DNA glycosylase-catalyzed reaction together with chemical structures of intermediate and product mimetics used in this study. **B**, Annealed composite omit electron density contoured to 1 σ is superposed onto the THF-DNA crystallographic model. **C**, Two views of the crystal structures, with protein in blue, DNA in gold, THF/1aR in magenta, and the thymidine opposite THF/1aR in green. **D**, DNA structures extracted from the complexes. **E**, Schematic of protein-DNA contacts. PfAlkC maintains similar contacts with the DNA in THF and 1aR complexes relative to the position of the DNA bend and not to the position of the 1aR/THF. **F**, Details of THF-DNA bound outside of the PfAlkC active site. Both complexes in the crystallographic asymmetric unit are shown. A PEG 4000 molecule (white carbons) fills the void in the

catalytic pocket in complex A. **G**, Superposition of the two protomers in the THF complex. Bold arrows highlight the difference in positions of the Ig-domain relative to the HLR domain and in the two DNA conformations. **H**, Superposition of the two 1aR complexes. **I**, Superposition of one protomer from each of the THF (gold) and 1aR (cyan) complexes.

Structural Differences between AlkC and AlkD

With crystal structures for both AlkC and AlkD in hand, we are able to understand the basis for their differences in locating and discriminating DNA alkylation damage. Despite the low sequence conservation between the HLR domains of BcAlkD and PfAlkC (16.3% identity and 27.1% similarity), they adopt the same general C-shaped architecture that complements the DNA helix (**Figure 21A-C**). Neither protein flips the target nucleobase out of the helix, but they use different mechanisms to interrogate the minor groove for lesion recognition. The modified nucleobase in the BcAlkD complex does not contact the protein, but remains stacked within a DNA duplex that is only modestly bent (**Figure 22**), which allows the enzyme to tolerate different positions and sizes of alkyl substituents. The bulky YTMA lesion in particular resides in a cleft between the enzyme and the widened minor groove (Mullins et al, 2017a). In contrast, PfAlkC imposes a much sharper bend in the DNA duplex, which disrupts base stacking and exposes the modified nucleobase to protein contacts from the minor groove (**Figure 22**). A defining feature of AlkC that distinguishes it from AlkD is the highly conserved α J-loop insertion that projects into the kink and traps the lesion between the surface of the protein and the kinked DNA (**Figure 21A-C**). This loop occupies the space in the minor groove that would be occupied by YTMA, and explains why AlkC is unable to excise bulky minor groove lesions. Because of its direct contact to the modified

nucleobases, the presence of the α J-loop also restricts the size of the lesion that can be accommodated (**Figure 18C**).

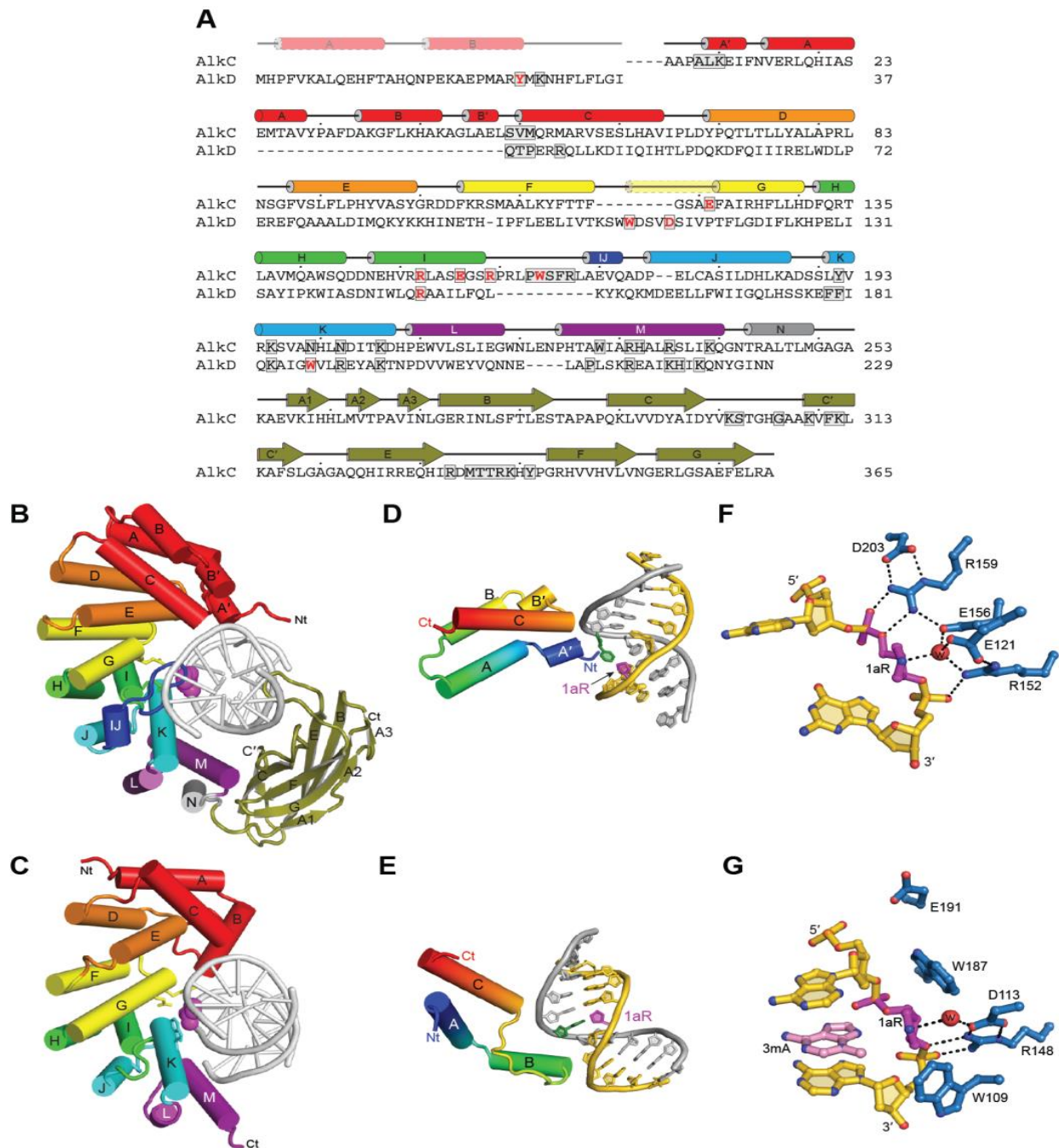


Figure 21. Comparison of AlkC and AlkD. **A**, Structure-based sequence alignment of PfAlkC and BcAlkD. Active site residues are red and DNA-interacting residues are boxed. Secondary structures derived from the crystal structures are shown above the sequences. The transparent N-terminal segment of helix α G (yellow) is unique to AlkD and helix α J (dark blue) is unique to AlkC. **B**, **C**, Crystal structures of 1aR-DNA

complexes of PfAlkC (B) and BcAlkD (C, PDB 5CLD). The N-terminal helical bundles are red; HEAT-like repeats are orange, yellow, green, cyan, and purple; and the Ig-like domain of AlkC is olive. Side-chains of active site residues are shown as sticks, and the 1aR moieties are shown as magenta spheres. **D, E**, Interactions between the N-terminal helical bundles of PfAlkC (D) and BcAlkD (E) and DNA. Protein is colored rainbow from N- (blue) to C-terminus (red). DNA strands are colored gold and silver, and 1aR and opposite thymine are magenta and green, respectively. **F, G**, Active sites of PfAlkC (F) and BcAlkD (G). Protein residues are colored blue, DNA gold, 1aR magenta, and 3-deaza-3mA (3d3mA) nucleobase pink. The catalytic water is shown as a red sphere and hydrogen bonds are depicted as dashed lines.

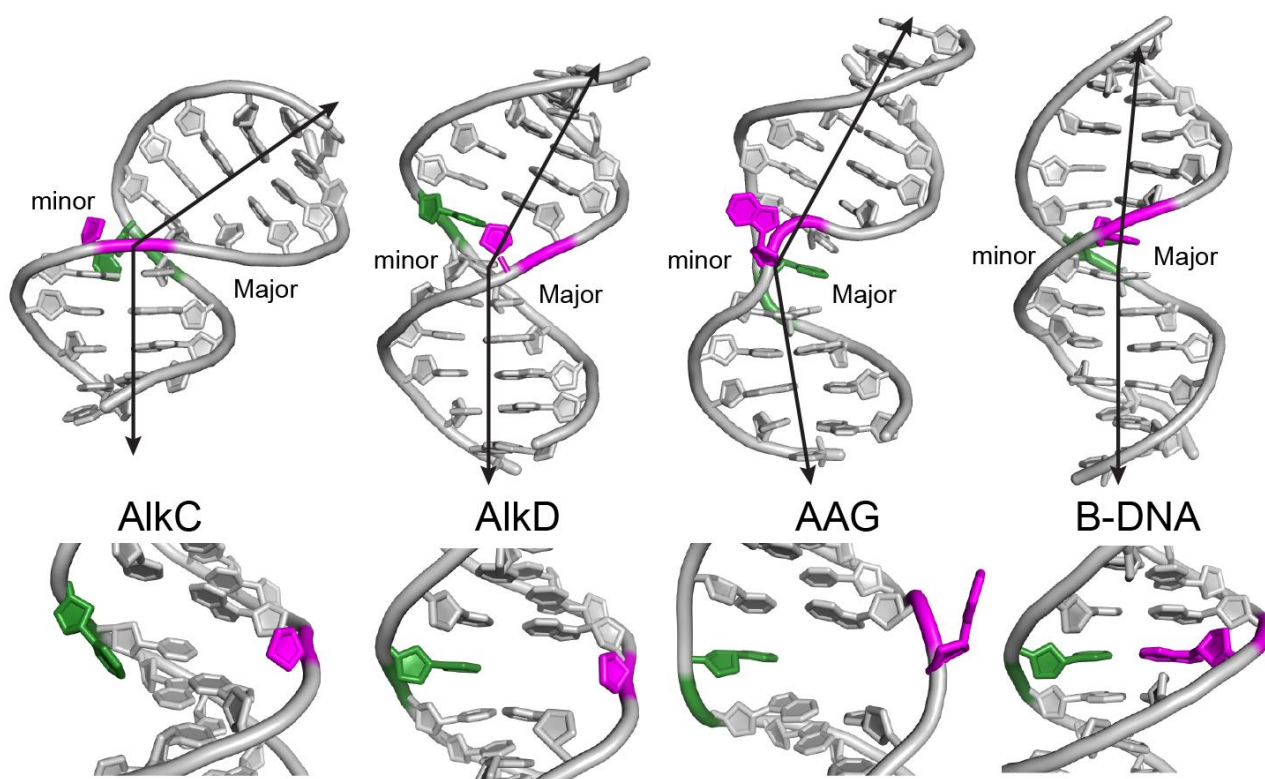


Figure 22. Glycosylase-induced DNA distortion. DNA models from glycosylase-DNA co-crystal structures of PfAlkC/1aR-DNA (this work), BcAlkD/1aR-DNA (PDB 5KUB), and base-flipping human AAG/1,*N*⁶-ethenoadenine-DNA (PDB 1EWN) compared to unbound B-DNA (PDB 1BNA). The modified and partner nucleotides are colored magenta and green, respectively. Black arrows depict the helical axes.

The highly kinked DNA important for lesion recognition in AlkC is stabilized by both the Ig-like domain as well as the N-terminal helical bundle, the conformation and

topology of which is distinct from AlkD and other HLR proteins AlkD2 and AlkF/AlkG (**Figure 21D,E**) (Backe et al, 2013; Rubinson et al, 2008). In AlkD's three-helical bundle (α A- α C), helix α B interacts with the damaged DNA strand and is essential for binding affinity (Mullins et al, 2015a). In contrast, PfAlkC's N-terminal helical bundle does not contact the damaged strand. PfAlkC's helix α A and α B, which are broken into 2-helix segments to give α A', α A, α B, and α B', run antiparallel to one another and with opposite polarity with respect to BcAlkD so that α B does not contact the DNA. Instead, the N-terminal end of helix α A' points toward the undamaged strand DNA, similar to the helix α C in both PfAlkC and BcAlkD (**Figure 21D,E**). As noted above, this region of AlkC may work together with the Ig-like domain to stabilize the bent DNA conformation.

The active sites of AlkC and AlkD are surprisingly different. AlkC contains a network of alternating charged residues along the damaged DNA backbone, whereas AlkD uses a Trp-Asp-Trp motif to cradle the backbone around the damaged nucleotide (**Figure 21F,G**). Each of the charged AlkC residues that span the damaged DNA backbone are invariant except for Glu121 (**Figure 11**). Glu121 is positioned closest to the anomeric C1' carbon and is spatially aligned with AlkD's catalytic Asp113 on helix α G (**Figure 21A**), consistent with its essential role in PfAlkC activity. Interestingly, among AlkC orthologs, the position of Glu121 is almost always either a glutamate or a tryptophan. We recently established that the two tryptophan side chains that flank the catalytic Asp113 in AlkD contribute to catalysis, presumably by further stabilizing the development of positive charge on the deoxyribose as the N-glycosidic bond is broken (Mullins et al, 2015b; Parsons et al, 2016). In contrast, invariant Glu156 is positioned closer to the O1' of the deoxyribose and is sandwiched between two arginine side

chains from the ionic network, significantly lowering its predicted pKa (2.9) relative to that of Glu121 (4.3). We therefore speculate that Glu121 deprotonates the water nucleophile and that Glu156 plays more of a role in stabilizing positive charge that develops on the deoxyribose during base excision.

AlkC Catalyzes Base Excision of 3mC and 1mA from Duplex DNA

During our phylogenetic analysis, we noticed that most AlkC-containing bacteria lacked an AlkB ortholog, implying that these bacteria would be unable to repair 1-methyladenine (1mA) and 3-methylcytosine (3mC) lesions by oxidative demethylation. Out of 834 completely sequenced and annotated bacterial genomes that contain either an AlkB or AlkC ortholog, only 6% contained both proteins (**Figure 23A**). Given this distribution and the apparent specificity of AlkC for cationic methylbases, we tested the ability of purified BcAlkC (α) and PfAlkC (β) proteins to excise 1mA and 3mC from a double-stranded oligodeoxynucleotide substrate under single-turnover conditions. Compared to a no-enzyme control, BcAlkC and PfAlkC exhibited robust activity for 3mC and modest activity for 1mA, while AlkD had no activity toward either 3mC or 1mA (**Figs. 6B,C and Figure 24A,B**). AlkC excision of 1mA was comparable to its weak activity for 7mG (Table 3), suggesting that 3mC is a more biologically relevant substrate for AlkC than is 1mA. Consistent with this, we were able to easily model 3mC, but not 1mA, into our crystal structure. Positioning 3mC in the active site places the N3-methyl group in van der Waals contact with Trp164 and the O2 oxygen in the Phe122/Glu121 pocket, whereas the N1-methyl group on 1mA sterically clashed with Trp164. 3mC and 1mA excision was abrogated by Glu121Ala and Glu156Ala mutants and impaired by

Trp164Ala substitution (**Figure 24C,D**), consistent with a catalytic role for the glutamates and a possible role for Trp164 in discrimination of different alkylated substrates. AlkC repair of these unusual lesions and the low percentage of bacteria that contain both AlkB and AlkC raises the interesting possibility that AlkC repairs 3mC and 1mA in bacteria that do not contain the AlkB oxidative demethylase (**Figure 23D**). Similarly, AlkA orthologs from the archaeon *Archaeoglobus fulgidus* and the archaea-related bacterium *Deinococcus radiodurans*, neither of which contain an AlkB homolog, are capable of 1mA and 3mC excision (Leiros et al, 2007; Moe et al, 2012). We did not observe 3mC or 1mA excision activity from ssDNA, nor did we observe excision of 1mG and 3mT (**Figure 24E**), all of which are substrates for AlkB (Delaney & Essigmann, 2004; Falnes et al, 2004; Falnes et al, 2002; Koivisto et al, 2004; Trewick et al, 2002).

Table 3. Base excision activities ^a

	7mG	3mC	1mA
BcAlkD	1.3×10^{-2} ^b	6.1×10^{-7}	n.d. ^c
BcAlkC	1.5×10^{-5}	1.7×10^{-4}	6.9×10^{-6}
PfAlkC	2.4×10^{-5}	5.2×10^{-4}	2.4×10^{-5}
PfAlkC E121A		n.d.	n.d.
PfAlkC E156A		n.d.	n.d.
PfAlkC W164A		6.1×10^{-5}	5.6×10^{-6}

^a Values are single-turnover rate constants (s^{-1}), averaged from three independent measurements. Errors (s.d.) were between 10-20% for each.

^b Value from Parsons (2016) *J. Am. Chem. Soc.*, 138: 11485-8.

^c n.d., none detected.

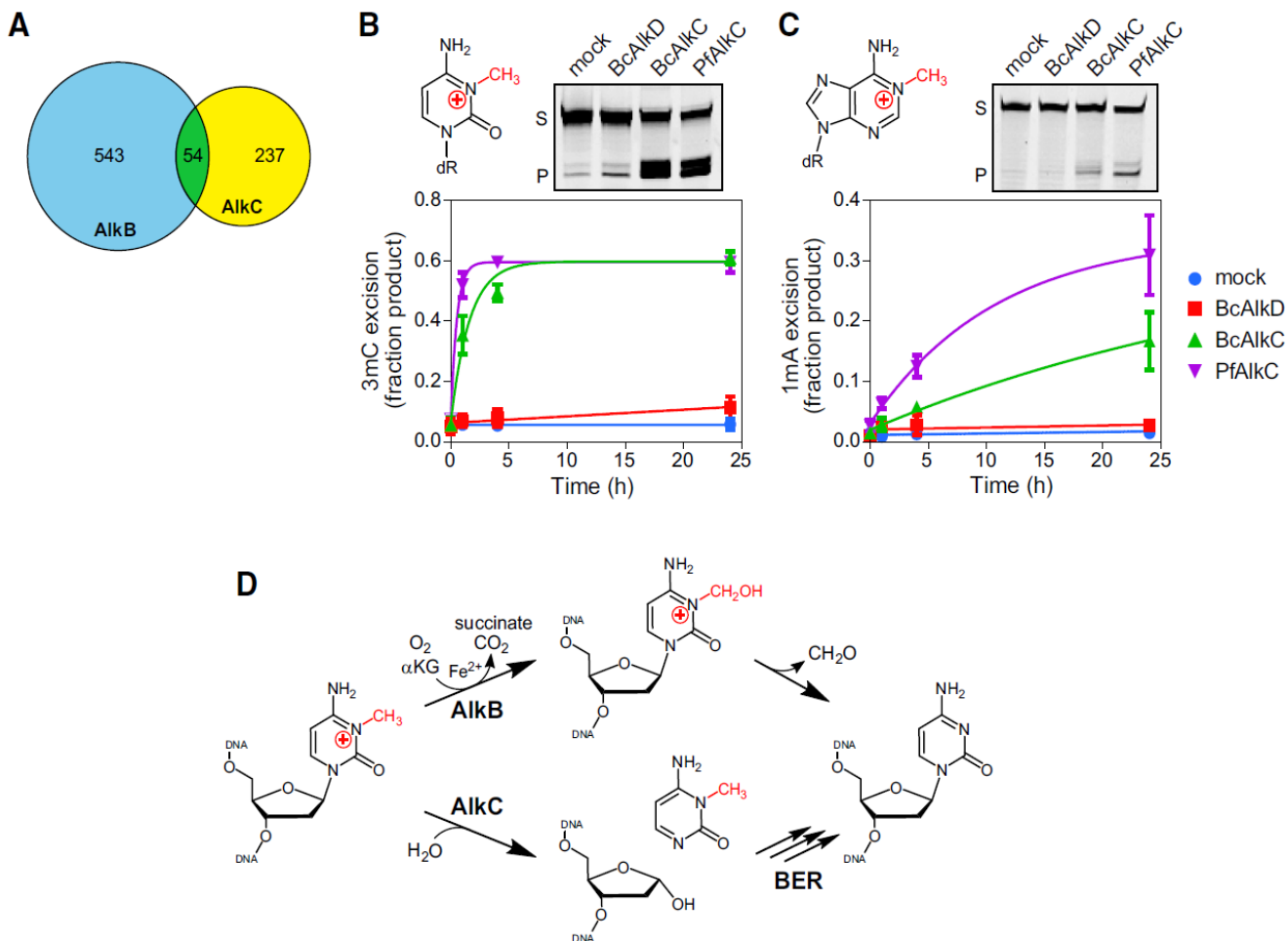


Figure 23. AlkC excises 3-methylcytosine (3mC) and 1-methyladenine (1mA) from DNA. **A**, Venn diagram of the numbers of bacterial species containing either AlkB (blue), AlkC (yellow), or both (green). **B**, **C**, Chemical structures and *in vitro* base excision of 3mC(**B**) and 1mA (**C**) from 25-mer double-stranded oligodeoxyribonucleotides. Representative denaturing electrophoresis gels show substrate (S) and product (P) after a 24-hour incubation with either no enzyme (mock), BcAlkD, BcAlkC (AlkC α), or PfAlkC (AlkC β). Plots show quantified time courses from three experiments (values are mean \pm SD). **D**, 3mC may be repaired in bacteria by either AlkB-catalyzed oxidative demethylation or AlkC-catalyzed base excision. α KG, α -ketoglutarate.

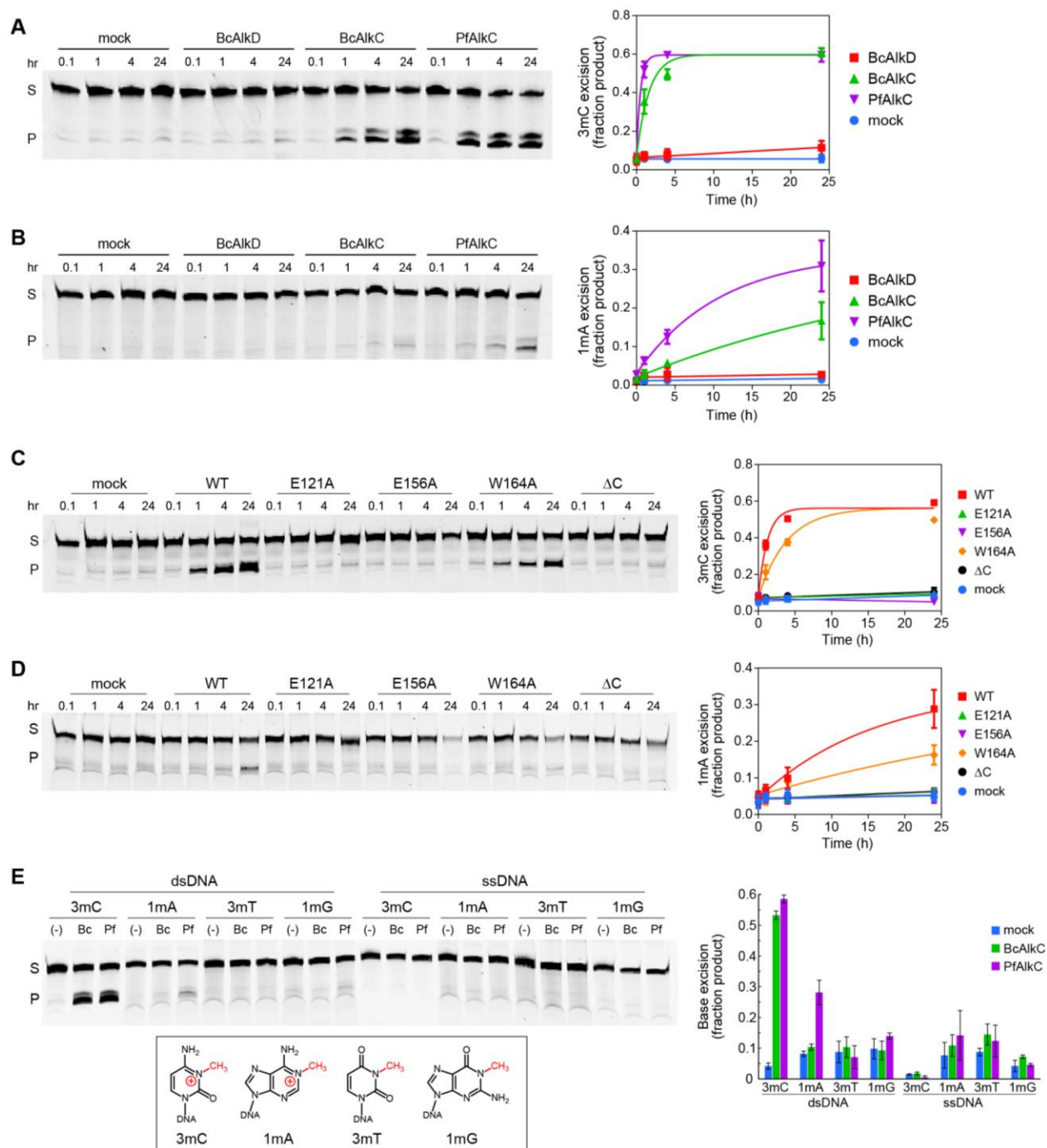


Figure 24. AlkC excision of 3mC and 1mA. Representative denaturing electrophoresis gels show substrate (S) and NaOH-nicked abasic-DNA product (P) over a 24-hour incubation. The plots show averages and s.d. from three experiments. **A,B**, 3mC-DNA (a) and 1mA-DNA (b) excision by no enzyme (mock), BcAlkD, BcAlkC, or PfAlkC. **C,D**, 3mC (c) and 1mA (d) excision by no enzyme (mock), wild-type PfAlkC (WT), or PfAlkC mutants (E121A, E156A, W164A, ΔC). **E**, 24-hour incubation of BcAlkC or PfAlkC with single-stranded (ss) or double-stranded (ds) oligodeoxynucleotides containing a single 3mC, 1mA, 1mG, or 3mT modification.

Discussion

Previous work defined the HLR domain as a new enzymatic scaffold for duplex DNA containing bulky lesions (Rubinson & Eichman, 2012; Rubinson et al, 2010; Rubinson et al, 2008). We now show how decoration of this DNA repair fold with additional motifs adapts it for specific, discreet lesions. The majority of AlkC orthologs extend the DNA binding surface beyond the HLR domain using an Ig-like domain, which works together with the HLR N-terminal helical bundle to severely distort the DNA duplex and expose the nucleobase lesion. Projection of the AlkC-unique α J-loop into this kinked DNA helix limits the size of alkylbases that can be excised, while two glutamate side chains enable AlkC to excise 3mC and 1mA, providing an alternative means of removing these modified nucleobases by BER, in addition to oxidative demethylation. In contrast, AlkD utilizes a Trp-Asp-Trp triad to recognize DNA damage without contacting the nucleobase or disrupting the DNA base stack, enabling AlkD to excise both major and minor groove lesions of various sizes. Thus, although AlkC and AlkD are related evolutionarily through their HLR architectures, and neither enzyme flips its target substrate into an active site pocket, they have developed different strategies to recognize and excise specific sets of alkylated lesions.

Methods

Phylogenetic Analysis

AlkC and AlkD orthologs were identified by the PSI-BLAST search against the NCBI non-redundant protein sequence database (last accessed on March 14, 2016) using PfAlkC (NCBI reference WP_012723400.1) and BcAlkD (NCBI reference

CAJ31885.1) as queries with a cutoff e-value of 10^{-5} . If multiple protein sequences were included for a single species, a simple neighbor joining tree and the corresponding HLR domain were used to remove the potentially redundant and paralogous sequences. All AlkC and AlkD orthologous sequences were aligned using Clustal Omega (Sievers et al, 2011) and the phylogenetic tree constructed using FASTTREE (Price et al, 2010) initiated with 100 random starting trees using a Whelan and Goldman + GAMMA amino acid model of substitution. The distribution of AlkB and AlkC sequences in bacteria was culled from the complete genomes using RefSeq representative genomes of NCBI TBLASTN (last accessed on March 14, 2016) with a cutoff e-value of 10^{-5} . The corresponding HLR and AlkB conserved domain were used to manually remove species that have potential paralogous sequences.

Protein Purification

Cloning and purification of *Bacillus cereus* (ATCC 14579) AlkC and AlkD were described previously (Mullins et al, 2013; Rubinson et al, 2008). AlkC genes from *Bacillus anthracis* (NCBI WP_000521720.1), *Bacillus mycoides* (WP_003190106.1), *Bacillus weihenstephanensis* (WP_012261200.1), *Sphingobacterium* sp. 21 (WP_013668283.1), *Pseudomonas fluorescens* SBW25 (WP_012723400.1), and *Streptomyces albus* (WP_003946460.1) were synthesized into expression vector pJ434 (DNA2.0). Constructs used in the genomic DNA base excision assay were overexpressed in *E. coli* C41(DE3) cells at 16 °C overnight upon induction with 500 μ M IPTG. For all other biochemical assays and X-ray crystallography, *P. fluorescens alkC* was sub-cloned into pBG100 (Vanderbilt Center for Structural Biology) encoding a

cleavable N-terminal hexahistidine tag and overexpressed in *E. coli* HMS174(DE3) at 16 °C overnight with 500 μM IPTG. Cells were lysed by sonication at 4 °C in buffer A (50 mM Tris•HCl pH 8.5, 500 mM NaCl, and 10% (v/v) glycerol). AlkC orthologs were purified from soluble lysate by Ni-NTA (Qiagen) affinity chromatography by sequentially washing and eluting in buffer A containing 20 mM and 500 mM imidazole, respectively. Pooled fractions were digested with Rhinovirus 3C (PreScission) Protease overnight at 4 °C to remove the N-terminal hexahistidine tag. Cleaved protein was diluted 10-fold in buffer B (50 mM Tris•HCl pH 8.5, 10% (v/v) glycerol, 2 mM DTT, and 0.1 mM EDTA) and purified over heparin sepharose (GE Healthcare) using a 50-1000 mM NaCl linear gradient. The protein was further purified using gel filtration chromatography (Superdex 200, GE Healthcare) in buffer C (20 mM Tris•HCl pH 8.5, 150 mM NaCl, 10% (v/v) glycerol, 2 mM DTT, and 0.1 mM EDTA). Purified proteins were concentrated, flash-frozen, and stored at -80 °C in buffer C.

PfAlkC mutants were constructed using the Q5 Site-Directed Mutagenesis Kit (New England BioLabs, Inc.). PfAlkC Δ C (residues 1-249) was generated by mutating the codon corresponding to residue 250 to a stop codon. Purification of PfAlkC mutants was conducted in the same manner as wild-type PfAlkC except that PfAlkC Glu121Ala was overexpressed in *E. coli* ArcticExpress (DE3) cells. Selenomethionine (SeMet)-incorporated PfAlkC was overexpressed and purified the same as wild-type PfAlkC, except overexpression was carried out in M9 minimal media supplemented with 0.4% (w/v) dextrose; 1 mM MgSO₄; 0.1 mM CaCl₂; 1 mg/L thiamine; 60 mg/L selenomethionine; 50 mg/L each of leucine, isoleucine, and valine; and 100 mg/L each of phenylalanine, lysine, and threonine. Integrity of mutant proteins was verified by

thermal denaturation, monitored by circular dichroism molar ellipticity at 222 nm on a Jasco J-810 spectropolarimeter using 7.5 μ M protein in 50 mM HEPES pH 8.5, 100 mM KCl, and 10% (v/v) glycerol.

Base Excision Assays

Base excision using genomic DNA and oligonucleotide substrates were carried out as previously described (Mullins et al, 2013). Briefly, to measure release of methylbases from genomic DNA by AlkC orthologs, 5 μ M protein was incubated with 10 μ g of MNU-treated calf thymus DNA at 37 °C for 1 h in 50 mM HEPES pH 7.5, 100 mM KCl, 5% (v/v) glycerol, 10 mM DTT, 2 mM EDTA, and 0.1 mg/mL BSA in a 50 μ L reaction and products quantified by HPLC-MS/MS. The activity of PfAlkC and mutants were tested in the same manner except that the assay was carried out at 21 °C for 5 min in a 30 μ L reaction consisting of 10 μ M protein, 50 mM HEPES pH 8.5, 100 mM KCl, 5% (v/v) glycerol, and 6 μ g of MNU-treated calf thymus DNA from a different batch. For oligonucleotide-based assays, 7mG, 1mA, 3mC, 1mG, or 3mT were incorporated into the sequence d(GACCACTACACCXATTCCTTACAAC) at the underlined position either enzymatically (7mG) (Mullins et al, 2013) or by solid-phase synthesis by Midland Certified Reagent Company (1mA, 1mG, and 3mT) or ChemGenes Corporation (3mC), and annealed to the complementary strand in annealing buffer (10 mM MES pH 6.5 and 40 mM NaCl). The YTMA lesion was generated in the sequence [d(CGGGCGGCGGCA(YTMA)AGGGCGCGGGCC)/d(GGCCCCGCGCCCTTTGCCGCGCCCG) as described previously (Mullins et al, 2015b). Glycosylase reactions contained 100 nM 6-carboxyfluorescein (FAM) labelled-

DNA and 10 μ M protein, and were carried out either at 21 °C in 50 mM HEPES pH 8.5, 100 mM KCl and 10% (v/v) glycerol (PfAlkC) or at 35 °C in 25 mM HEPES pH 7.5, 50 mM KCl and 5% (v/v) glycerol (BcAlkC and BcAlkD).

X-Ray Crystallography

Crystallization of the seven purified AlkC orthologs were first screened against a library of double-stranded oligonucleotides (Integrated DNA Technologies) ranging in length from 8-15 nucleotides and containing a centralized THF across from thymine. Diffracting crystals were obtained from the PfAlkC protein grown in the presence of an 11-mer THF-DNA [d(TGTCCA(THF)GTCT)/d(AAGACTTGGAC)]. The PfAlkC/THF-DNA structure was determined using single-wavelength dispersion (SAD) phases from SeMet-incorporated protein and refined to 2.4-Å resolution (Table 2, Supplemental **Figure 20**). Crystals of the 1aR-DNA complex were obtained by replacing THF with 1aR in the same 11-mer sequence, which was synthesized as previously described (Chu et al, 2011). The PfAlkC/1aR-DNA structure was determined by molecular replacement using the refined protein coordinates from the PfAlkC/THF-DNA complex as a search model.

Protein-DNA complexes were assembled by incubating 0.12 mM PfAlkC with 0.15 mM DNA on ice for 30 min. Crystals were grown using the hanging-drop vapor diffusion method at 21 °C. For the THF complex, 1 μ L of SeMet-PfAlkC/THF-DNA solution was mixed with 1 μ L of reservoir solution (100 mM Tris•HCl pH 8.5, 18% PEG 4,000, and 5% (v/v) glycerol). For the 1aR complex, 1 μ L of wild-type PfAlkC/1aR-DNA solution was mixed with 1 μ L of reservoir solution (18% pentaerythritol propoxylate and

100 mM MES pH 6.1). Drops were equilibrated against 500 μ L of reservoir solution. Crystals were flash frozen in liquid N₂ in either mother liquor supplemented with 30% (v/v) glycerol (THF complex) or in mother liquor alone (1aR complex).

X-ray diffraction data were collected at Advanced Photon Source beamline 21-ID-F (THF complex) and Advanced Light Source SIBYLS beamline (1aR complex). All data were processed with HKL2000 (Otwinowski & Minor, 1997). Phases for the PfAlkC/THF-DNA complex were determined by Se-SAD from positions of 18 Se atoms using AutoSHARP (Vonrhein et al, 2007). PfAlkC/1aR-DNA phases were determined by molecular replacement using the refined coordinates of the SeMet-PfAlkC protein as a search model in the program Phaser (McCoy et al, 2007). Atomic models were built in Coot (Emsley & Cowtan, 2004) and atomic positions, individual B-factors, TLS parameters, and occupancies were refined using Phenix (Adams et al, 2010). The final models were validated with MolProbity (Davis et al, 2007) and contained no residues (THF complex) or one residue (1aR complex) in the disallowed regions of the Ramachandran plot (Table 2). The identity of the sodium ion was verified by ligand distances, coordination, and geometry using the CheckMyMetal web server (Zheng et al, 2014). Structure factors and coordinates were deposited in the Protein Data Bank under accession codes 5VI0 (THF complex) and 5VHV (1aR complex), and the corresponding X-ray diffraction images deposited in the SBGrid Data Bank (Meyer et al, 2016).

Structural biology software was curated by SBGrid (Morin et al, 2013). Structure figures were prepared in PyMOL (Schrödinger, 2016). Conservation from 527 AlkC β sequences was mapped onto the structure of PfAlkC using the ConSurf server (http://consurftest.tau.ac.il/index_full_form_example.php). DNA geometric parameters

were measured using Web 3DNA (Zheng et al, 2009). Structure-based pKa calculations were carried out in Rosetta (Kilambi & Gray, 2012; Lyskov et al, 2013).

CHAPTER III

CELLULAR FUNCTION OF ALKC AND ALKD IN RESPONSE TO DIFFERENT DNA DAMAGING AGENTS

Abstract

DNA damage repair is essential to safeguard the stability and integrity of the genomic DNA against endogenous and environmental DNA damaging agents. Among the diverse repair pathways, direct reversal and base excision repair (BER) are two fundamental pathways that occur in cells to repair DNA methylation damage, while nucleotide excision repair (NER) is responsible for the majority of bulky helix-distorting lesions. *Bacillus cereus* contain methylpurine DNA glycosylases AlkC and AlkD, which exhibit robust *in vitro* excision activity for 3-methyladenine, in addition to three other putative methylpurine DNA glycosylases. The reason for such an apparent redundancy is unclear, and suggests an alternate role for AlkC and AlkD in repairing other types of damage. Indeed, we previously showed that AlkC excises 3-methylcytosine (3mC) and 1-methyladenine (1mA), which are usually repaired by the direct reversal enzyme AlkB.

*Some of the work described in this chapter has been published in:

Mullins, E. A., Shi, R., Parsons, Z. D., Yuen, P. K., David, S. S., Igarashi, Y., Eichman, B. F. (2015). The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. *Nature*, 527(7577), 254.

Mullins, E. A., Shi, R., Eichman, B. F. (2017). Toxicity and repair of DNA adducts produced by the natural product yatakemycin. *Nat. Chem. Biol.*, 13(9), 1002.

Shi, R., Mullins, E. A., Shen X-X, Lay, K. T., Yuen, P. K., David, S. S., Rokas, A., Eichman, B. F. (2018) Selective base excision repair of DNA damage by the non-base-flipping DNA glycosylase AlkC. *EMBO J.* 37(1): 63-74

Here, we used bacterial genetics to understand the cellular roles of AlkC and AlkD. Cell growth data from *Bacillus anthracis* strains lacking *alkC* or *alkD* genes indicated that AlkC and AlkD are not critical for protection of cells against methylating or other common DNA damaging agents. Consistent with its *in vitro* excision activity, we found that AlkD mediated BER and UvrA-mediated NER work synergistically in repair of YTMA, the antifungal and antibiotic compound produced by *Streptomyces*. Despite the excision activity of AlkC toward 3mC and 1mA *in vitro*, we did not observe significant effect of AlkC on repair of these adducts in cells. These results indicate that AlkC and AlkD may have evolved to contribute to repair of non-BER substrates, and give an example of how the cell uses one scaffold for diverse biological roles.

Introduction

The cell encounters constant environmental and endogenous challenges that cause genomic instability (Friedberg et al, 2006b). In bacteria, two major pathways play key roles in removal of alkylated lesions (Mishina et al, 2006). Base excision repair (BER) is the predominant pathway to deal with the majority of methylated nucleobases. BER is initiated with recognition and excision of the damaged nucleobase by a DNA glycosylase, followed by incision of the abasic (AP) site, DNA synthesis and ligation (Fromme & Verdine, 2004). DNA glycosylases are widely spread in all domains of life, consistent with their important role in damage repair (Scharer & Jiricny, 2001). 3-methyladenine DNA glycosylases are responsible for two of the most abundant N-methylated lesions in cells, 3-methyladenine (3mA) and 7-methylguanine (7mG). 3mA is highly toxic as it inhibits DNA replication while 7mG is prone to produce labile AP sites

(Wyatt et al, 1999). The expression levels of methylpurine DNA glycosylases are tightly regulated in cells to prevent, rather than confer alkylation sensitivity (Posnick & Samson, 1999; Troll et al, 2014). Human alkyl-adenine DNA glycosylase (AAG) preferentially eliminates hypoxanthine (Hx) and its expression largely depends on the tissue type and the age (Allan et al, 1998). *E.coli* TAG, the 3-methylpurine DNA glycosylase specific for 3mA, is constitutively expressed to repair a basal level of endogenous alkylation damage, while AlkA, which is part of the adaptive response and has a broader substrate range, is inducible upon methylating drug treatment (Lindahl et al, 1988; Sedgwick & Lindahl, 2002). Rapid removal of methylated bases by overexpressing methylpurine DNA glycosylase increases cell sensitivity to damaging agents as the process generates an elevated amount of AP sites and strand breaks, which are more cytotoxic than the neutral 7mG lesions (Glassner et al, 1998; Posnick & Samson, 1999; Troll et al, 2014).

The other fundamental pathway to repair methylation damage is direct reversal repair. One form of direct reversal is oxidative demethylation, which is mediated by the dioxygenase AlkB (Falnes et al, 2002). AlkB oxidizes the erroneous methyl group on 3-methyl-cytosine (3mC) and 1-methyl-adenine (1mA) to formaldehyde, and reverts the damaged bases to cytosine (C) and adenine (A) (Landini & Volkert, 2000; van den Born et al, 2009). Despite bearing similar positive charge with 3mA and 7mG at physiological pH, 3mC and 1mA bases with high abundance in ssDNA and RNA have more stable N-glycosidic bonds, and are repaired by direct reversal repair instead of base excision repair (Dalhus et al, 2009). Interestingly, AlkA2 from *Deinococcus radiodurans* and AlkC

from *Bacillus anthracis* process *in vitro* activities toward the AlkB substrates 3mC and 1mA (Moe et al, 2012; Shi et al, 2018).

YtkR2, a homolog of 3-methyladenine DNA glycosylases AlkD from HEAT-like repeat superfamily, has been identified in the gene cluster of yatakemycin synthesis pathway in *Streptomyces* sp. TP-A0356. Besides removing 3mA and 7mG, YtkR2 shows activity for N3-base adducts modified by the highly potent antifungal drug yatakemycin (YTM), which belongs to the family of duocarmycin drug family (Xu et al, 2012). Unlike the majority of glycosylases that flip the modified base into the active site, AlkD adopts a non-base flipping mechanism in which the damaged nucleotide remains stacked in the DNA duplex during catalysis. This lack of constraints on enzyme substrate size allows YtkR and AlkD to efficiently recognize and excise bulky positively charged N3-yatakemicyladenine (YTMA) in *in vitro* glycosylase assays (Mullins et al, 2015b).

In *Bacillus cereus*, five 3-methylpurine DNA glycosylases coexist in cells (Alseth et al, 2006). Together with robust *in vitro* activities of AlkC for 3mC and AlkD for YTMA, this high redundancy leads to the hypothesis that AlkC and AlkD may have gained activity for noncanonical base adducts besides 3mA and 7mG removal. Here we demonstrate, though AlkC and AlkD exhibit high activity for 3mA (Alseth et al, 2006; Mullins et al, 2013), knockout of *alkC* and *alkD* genes imparts no sensitivity to the methylating agents. However, AlkD-mediated BER and UvrA-mediated NER contribute to the repair of bulky YTM-adducts *in vivo*.

Results

AlkC and AlkD are Not Essential for Repair of Methylation Damage

In *E.coli*, AlkA and TAG are two critical methylpurine DNA glycosylases to excise methylated nucleobases (Evensen & Seeberg, 1982; Wyatt & Pittman, 2006). Depletion of these two 3-methyladenine glycosylases sensitizes cells to methylating agents, such as methyl methane sulfonate (MMS). The $\Delta alkA\Delta tag$ double-knockout strain exhibits remarkable susceptibility to MMS, whereas complementation with AlkC or AlkD significantly rescues its sensitivity (Alseth et al, 2006). Consistent with this, AlkC and AlkD exhibit robust *in vitro* activities toward 3-methyladenine (3mA) (Mullins et al, 2013). The cellular function of AlkC and AlkD were characterized through ectopic expression in *E. coli*, and no AlkC or AlkD homolog has been discovered in *E.coli*. In order to precisely understand the cellular function of AlkC and AlkD, it is necessary to work on a model organism that contains *alkC* and *alkD* genes. Alkylpurine DNA glycosylases AlkC and AlkD were originally discovered in *Bacillus cereus*, which is nearly identical (>90%) to that of *Bacillus anthracis*. The avirulent Sterne strain (34F2) of *Bacillus anthracis* has lost the plasmid that codes one virulence factor, which allows it to be widely used in scientific research and animal immunization (Welkos, 1991; Welkos & Friedlander, 1988).

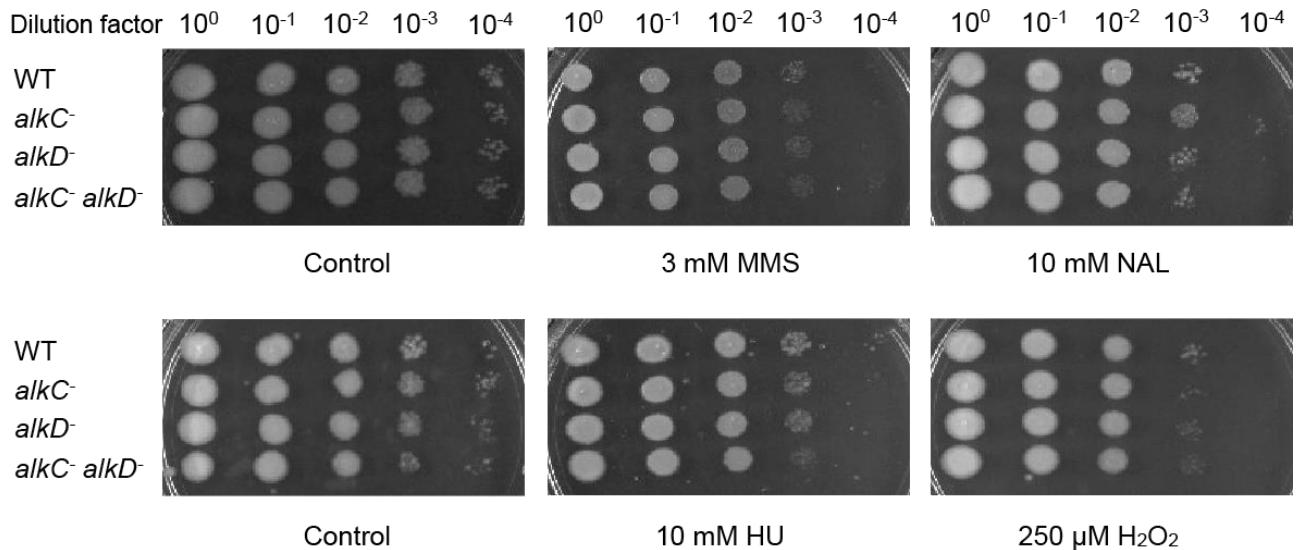


Figure 25. Sensitivity of $\Delta alkC$, $\Delta alkD$, and $\Delta alkC\Delta alkD$ mutants to a variety of DNA damaging agents. The logarithmically growing cells of wild-type (WT), $\Delta alkC$, $\Delta alkD$, and $\Delta alkC\Delta alkD$ -knockout strains of *B. anthracis* were diluted and grown on LB plates containing 3 mM MMS, 10 mM NAL, 10 mM HU and 250 μ M H₂O₂.

To characterize the cellular function of AlkC and AlkD in repair of various DNA damage, we treated *alkC* and *alkD* depleted cells with a variety of DNA damaging agents in *B. anthracis* Sterne. No noticeable difference in growth was observed for single- and double- knockouts compared with wild-type after treatment with hydroxyurea (HU), nalidixic acid (NAL), MMS, N-methyl-N-nitrosourea (MNU) (data not shown), and hydrogen peroxide (H₂O₂) (**Figure 25**). HU is the inhibitor of ribonucleotide reductase, which causes replication stress by blocking DNA synthesis and decreasing the level of dNTP pool (Koç et al, 2004). NAL inhibits DNA gyrase, leading to double-strand DNA breaks and inducing the SOS response (McQueen et al, 1989). H₂O₂ generates reactive hydroxyl radicals to damage DNA (Petersen et al, 2000). As expected, the results indicate that AlkC and AlkD are not involved in repair of lesions induced by HU, NAL, and H₂O₂. Alkylating agents and carcinogens MMS and MNU produce different spectra of methylated bases. MMS primarily causes 3mA and N7-methylguanine (7mG) adducts

while MNU generate O⁶-methylguanine in addition to N-methylated products (Nay & O'Connor, 2013). Compared to a broad range of substrates (3mA, 7mG, and POB-adducts) removed by AlkD, AlkC is more specific for 3mA *in vitro* (Mullins et al, 2013; Rubinson et al, 2010). Interestingly, the presence of AlkC and AlkD does not render the cell higher resistance to methylation damage (**Figure 25**). Because of the coexistence of multiple alkylpurine DNA glycosylases in *Bacillus anthracis*, AlkC and AlkD are not essential for cell survival when challenged with MMS and MNU. Similar results were obtained when the mutant tolerance and persistence were tested using different drug concentrations of MMS and MNU (data not shown).

The Expression of AlkC and AlkD upon DNA Damaging Agent Treatment

Bacillus cereus has two AlkA homologs and one AAG homolog presumably to control the level of alkylated DNA. To investigate the expression of AlkC and AlkD in response to DNA damage and understand the regulatory mechanism of balance maintenance between different alkylpurine DNA glycosylases, we compared the expression of alkylpurine DNA glycosylases AlkA, AAG, AlkC, AlkD, and NER enzyme UvrA in the presence and absence of DNA damaging agents, with particular attention to the methylated and bulky lesions produced by methyl methanesulfonate (MMS) and NNK precursor 4-[(acetoxymethyl) nitrosamino]-1-(3-pyridyl)-1-butanone (NNKOAc-). NNK is a potent tobacco-specific carcinogen that results in bulky pyridyloxobutyl (POB)-modified DNA (Xue et al, 2014). The cytotoxic POB adducts are repaired by NER *in vivo* (Brown et al, 2008), while AlkD also releases cationic 7-POB-G and O²-POB-T adducts *in vitro* (Rubinson et al, 2010). As expected, one of the AlkA homologs was upregulated

following MMS stimulation. The human AAG homolog was not induced after MMS or NNKOAc exposure, whereas the *E.coli* UvrA homolog remarkably increased its expression in the presence of MMS. AlkC and AlkD also showed inducible expression with MMS (**Figure 26**). The data suggest that under normal circumstances, methylpurine DNA glycosylases maintain basal levels of expression to protect against endogenous DNA damage, whereas these glycosylases (except for AAG) are upregulated in response to elevated levels of damaged bases. This qRT-PCR experiment was performed only once, and needs to be repeated to ensure the accuracy of the results.

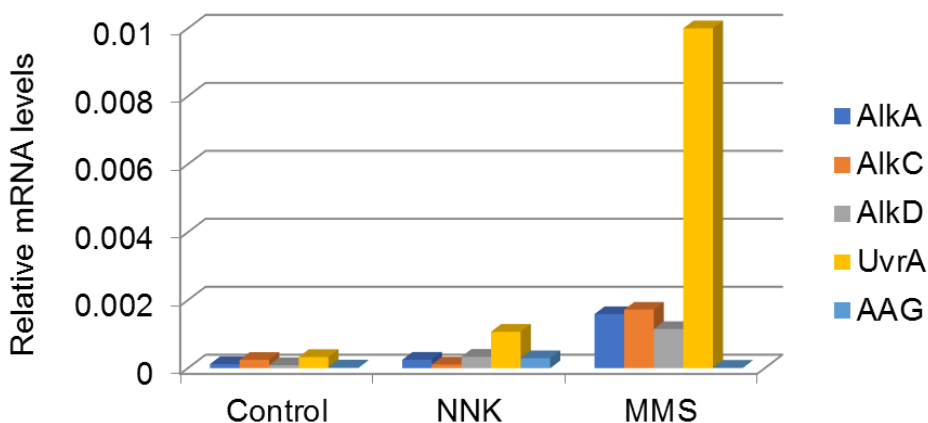


Figure 26. Expression of putative *B. anthracis* alkylpurine DNA glycosylases. mRNA levels of *B. anthracis* AlkA (blue), AlkC (orange), AlkD (grey), UvrA (yellow), and AAG (light blue) with and without 4-(methylnitrosamino)-1-(3-pyridyl)-butanone (NNKOAc) and MMS were determined by qRT-PCR and normalized to the internal 16sRNA control. This experiment was conducted only once.

AlkD-Mediated BER and UvrA-Mediated NER Work Synergistically in Repair of Bulky Yatakemycin Adducts

Unlike alkylpurine DNA glycosylases AAG, MAG and AlkA, AlkD is the first glycosylase that does not rely on the base-flipping mechanism and excises bulky YTM *in vitro* (Mullins et al, 2015b; Rubinson et al, 2010). YtkR2, the homolog of AlkD, has been identified in the YTM biosynthesis gene cluster as providing self-resistance to this extremely cytotoxic antitumor antibiotic YTM (Xu et al, 2012). We wanted to know whether AlkD plays an important role in repair of YTM in *Bacillus anthracis*. Therefore, we constructed a *B. anthracis* $\Delta alkD$ knockout and tested the cell sensitivity to YTM. Consistent with previous results, the lack of *alkD* showed no effect on cell sensitivity toward MMS compared to the wild-type due to the redundancy of multiple alkylpurine DNA glycosylases (**Figure 27a**). However, deletion of *alkD* did increase cell susceptibility in the presence of YTM, indicating that AlkD is important in removal of bulky lesions *in vivo* (**Figure 27b**). Consistent with the lack of the ability to remove YTM *in vitro*, *alkC*-depleted *B. anthracis* strains did not show additional sensitivity in response to YTM compared with the wild-type (**Figure 12c**).

Bulky lesions, including those formed from a natural product akin to YTM and CC-1065, are known to be repaired via nucleotide excision repair (Kiakos et al, 2007; Selby & Sancar, 1988). To understand how AlkD-mediated BER and UvrA-mediated NER work synergistically in resistance to YTM, we constructed $\Delta alkD$ and $\Delta uvrA$ single knockouts and $\Delta alkD\Delta uvrA$ double knockout strains, and treated them with increasing concentrations of YTM. $\Delta uvrA$ showed slightly more growth defect than $\Delta alkD$ after YTM treatment, while $\Delta alkD\Delta uvrA$ exhibited cumulative sensitivity (**Figure 28 a,b**). The high sensitivity of $\Delta alkD$ and $\Delta uvrA$ single knockouts at low drug concentration suggests NER and AlkD provide low levels of protection *in vivo*. Single-

and multiple-turnover excision of YTM by AlkD have been tested and showed that the YTM elimination efficiency is restricted by product inhibition, thus correlating the robust *in vitro* activity of AlkD with its low cellular protection (Mullins et al, 2017a).

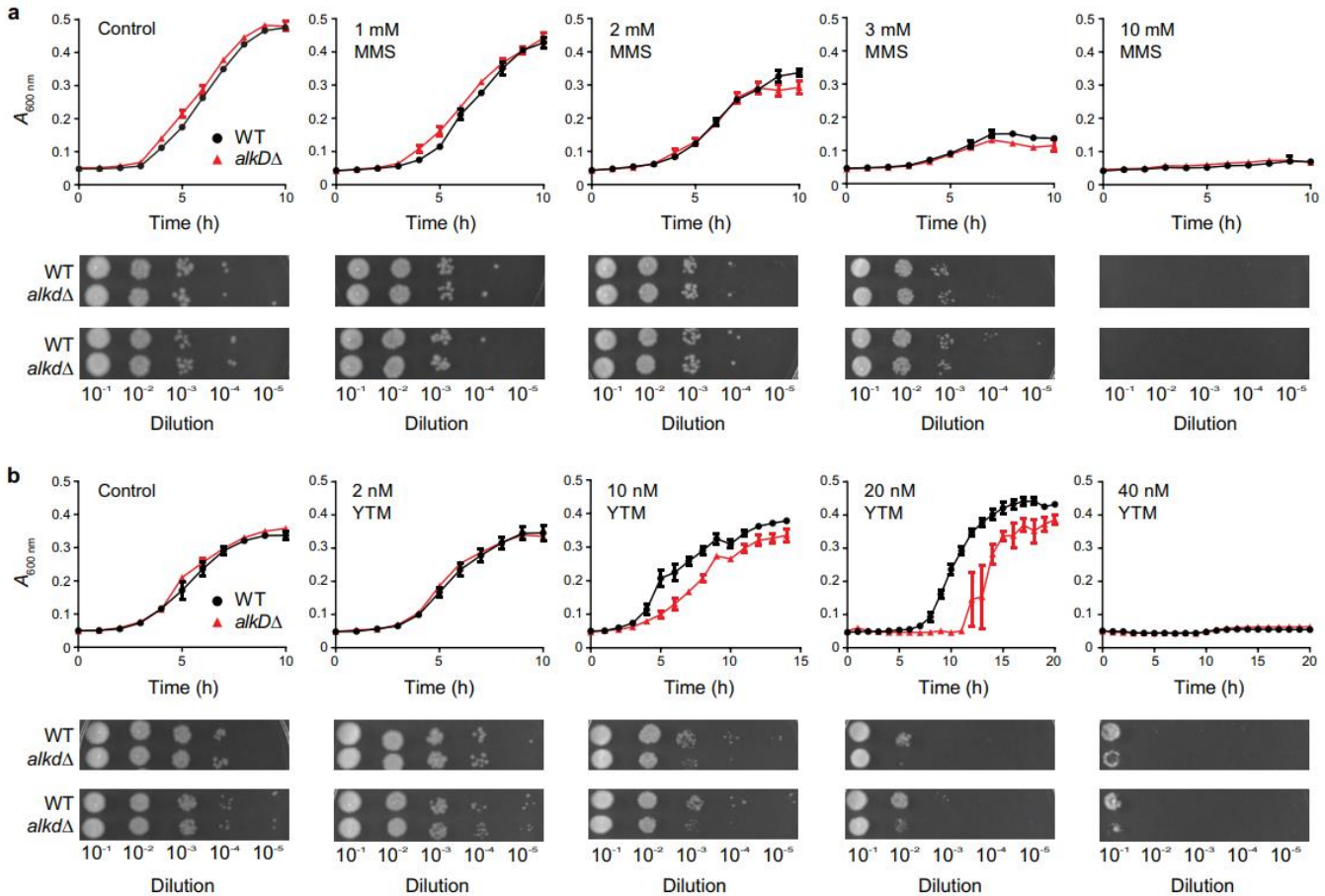


Figure 27. Sensitivity of $\Delta alkD$ mutant and wild-type to MMS and YTM. Wild-type (black) and $\Delta alkD$ knockout (red) strains of *B. anthracis* were treated with varying concentrations of MMS and YTM. Control experiments contained no drug. Deletion of AlkD caused no observable phenotype with MMS but resulted in increased sensitivity to YTM. **a**, MMS treatment. **b**, YTM treatment. Error bars represent the s.e.m. from three replicate growth curves. Spot assays were performed in duplicate. (Mullins, 2015)

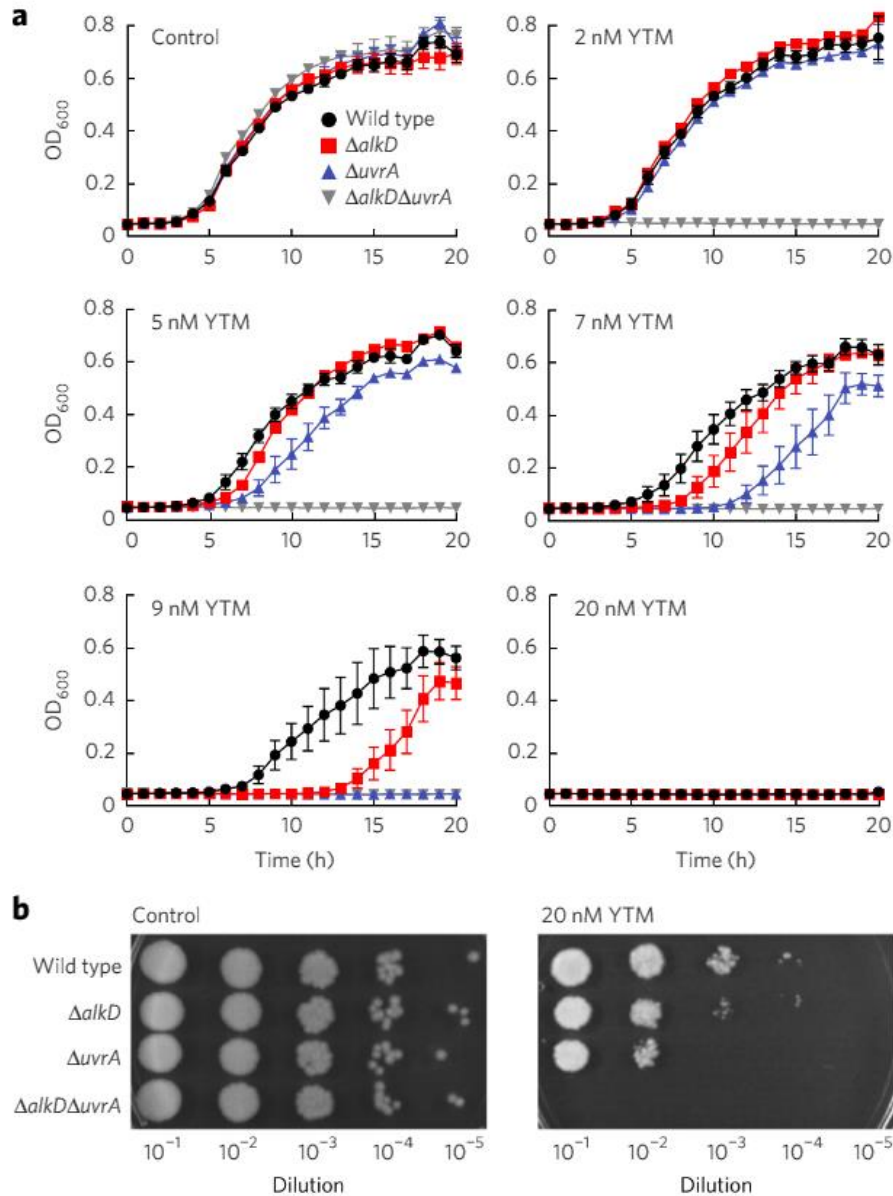


Figure 28. Determination of YTM resistance. **a**, Growth of *B. anthracis* deletion strains in liquid medium. Error bars represent s.e.m. from four replicate experiments. OD600, optical density at 600 nm. **b**, Growth of *B. anthracis* deletion strains on solid medium. YTM was omitted from all control experiments. (Mullins, 2017)

Cellular Protection against 3-Methylcytosine and 1-Methyladenine by AlkB

The predominant pathway for repair of 3-methylcytosine (3mC) and 1-methyladenine (1mA) in DNA and RNA is mediated by the dioxygenase AlkB through

oxidative demethylation. We found that AlkC specially removes 1mA and 3mC damage in dsDNA. Because the majority of 3mC and 1mA are present in single-strand (ss) DNA and RNA, we were interested to know whether AlkC removes these lesions from ssDNA, ssRNA and dsRNA, and provides an alternative pathway to eliminate 3mC and 1mA *in vivo*.

To examine the activity, we incubated dsRNA and ssRNA containing 1mA with *Pseudomonas fluorescens* AlkC (PfAlkC) and *B. cereus* AlkC (BcAlkC), and quantified the released nucleobases by HPLC-MS. No 1mA nucleobase was liberated following a 24-hour incubation with AlkC, suggesting AlkC discriminates against lesion in ssDNA (Figure 24E) and RNA (Figure 29).

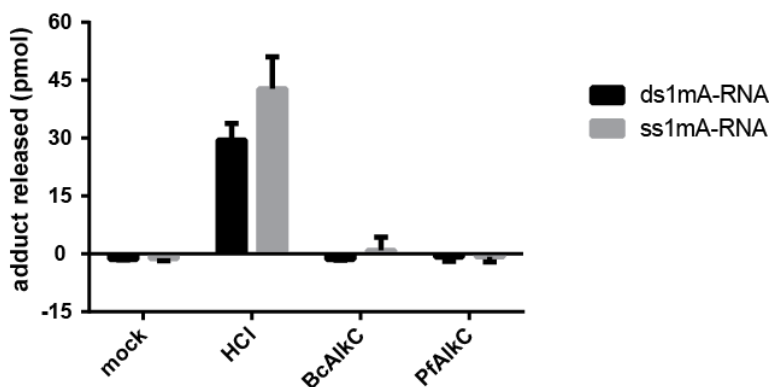


Figure 29. AlkC has no activity for 1-methyladenine in RNA. 1-methyladenine (1mA) was released from double-strand RNA (black bars) and single-strand RNA (grey bars) after a 24-hour incubation with either no enzyme (mock), positive control (HCl), *B. cereus* AlkC, or *P. fluorescens* AlkC. Error bars represent the s.e.m. from three replicate experiments.

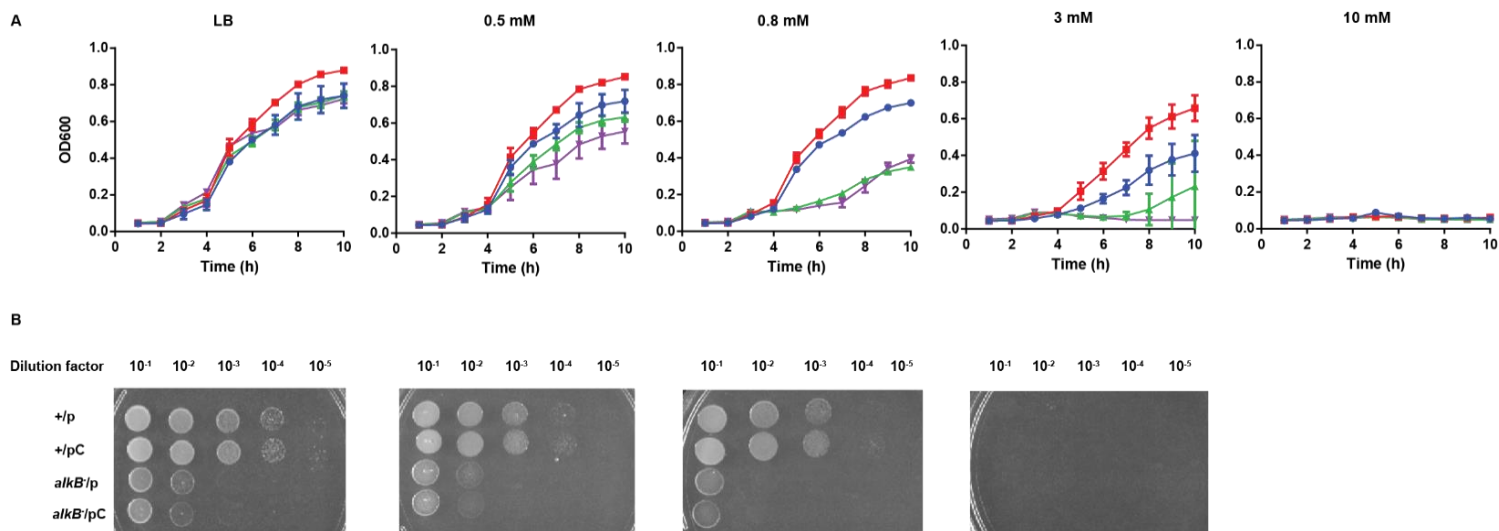


Figure 30. Cellular function characterization of AlkC. **A**, Growth of wild-type *E.coli* (MV1161) with empty vector pBR322 (+p; red) and pBR322-bcAlkC (+pC; blue), *E.coli* $\Delta alkB$ (MV2029) with empty vector pBR322 ($\Delta alkB/p$; green) and pBR322-bcAlkC ($\Delta alkB/pC$; purple) was determined in the presence of varying concentrations of N-methyl-N-nitrosourea (MNU). Cell densities were recorded every hour for 10 hours. **B**, Early-mid-log phase cells were plated on LB plates prepared with 0, 0.5, 0.8, and 3 mM MMS. Growth curves and spot assays were performed in triplicate. Error bars represent the s.e.m. from three replicate growth curves.

AlkB prevents cell death from exposure to MMS. We speculated AlkC might provide cellular protection against toxicity of 1mM and 3mM in the absence of *alkB*. We therefore asked whether overexpression of AlkC would suppress the sensitivity of *alkB*-deleted cells to MMS. To control the expression level of AlkC, we expressed AlkC either from the “leaky” *trc* promoter or the endogenous *alkB* promoter. $\Delta alkB$ mutants were more sensitive to MMS than wild-type, whereas additional expression of AlkC did not enhance MMS resistance of either strain (**Figure 30 A,B**). In contrast, expression of ectopic AlkC in wild-type and $\Delta alkB$ mutant decreased the cell viability. This is probably because AlkC is highly active in the cell to remove methylated lesions, thus increasing

abasic sites and strand breaks, whereas 1mA and 3mC lesions in dsDNA themselves are of low toxicity to cells and only comprise a small portion of DNA damage.

Discussion

Here we report that although AlkC and AlkD catalyze repair of a broad range of positively charged lesions *in vitro*, the relative cellular repair efficiency for each lesion varies dramatically in *B. anthracis*. The non-base-flipping glycosylase AlkD provides *in vivo* activity for NER substrates YTMA, but not for classical BER substrates 3mA and 7mG. Similarly, in contrast to high 3mA activity *in vitro*, lack of AlkC has no effect on survival with methylation damage. Note that no detectable effect on cellular repair of 3mA was observed in the absence of AlkC and AlkD compared to wild-type, suggesting that this specificity may have been developed as a result of the overlapping methylation repair of alkylpurine DNA glycosylases in *B. anthracis*.

A host cell reactivation assay can be performed to determine the cellular function of AlkC in 3mC repair. In this experiment, the survival of methylated lambda M13 duplex DNA is assayed in *B. anthracis* $\Delta alkC$, $\Delta alkB$ and $\Delta alkC\Delta alkB$ mutants. Specifically, M13 DNA pre-treated with MMS is transformed into *B. anthracis* mutants. Failure to repair the methylated bases will hinder the transformation efficiency of DNA. Thus, repair efficiency can be monitored by counting plaque-forming units. If AlkC is highly effective in repair of 3mC and 1mA in dsDNA, $\Delta alkC\Delta alkB$ mutant will show no or only modest deficiency in reactivating M13 compared to single mutants.

The global transcriptional upregulation of multiple methylpurine DNA glycosylases and UvrA in *B. anthracis* following MMS and NNKOAc treatment suggests

BER and NER may play a minor role in removal of lesions induced by these two chemicals. The discrepancy in expression indicates the relative levels of DNA glycosylases are fine-tuned for balance between eliminating toxic lesions and maintaining low levels of harmful repair intermediates.

More work needs to be done to fully understand the regulation of methylpurine DNA glycosylases. AlkD excises bases modified by YTM, thus mRNA levels can be monitored with YTM treatment. Prior to examination of the expression levels of AlkC and AlkD upon MMS and YTM treatment, the substrate spectra of putative AlkA and AAG homologs should be determined to verify that they are not non-functional proteins and have identical specificities with *E.coli* AlkA and human AAG. Furthermore, AlkC and AlkD are upregulated simultaneously with AlkA after MMS treatment, it would be interesting to characterize the regulatory pathway of AlkC and AlkD and test if the operons of *alkC* and *alkD* genes are transcriptionally activated by Ada.

Methods

Preparation of $\Delta alkC$, $\Delta alkD$, and $\Delta alkC \Delta alkD$ Cells

The 1-kilobase flanking regions surrounding *alkD* were inserted into the knockout-plasmid pLM4 using standard molecular biology techniques. The modified plasmid was propagated in nonmethylating *E. coli* K1077 and dialyzed against electroporation buffer for 1-4 hours before introduction by electroporation into *B. anthracis* Sterne cells in pre-chilled 1mm BioRad cuvette. *B. anthracis* colonies containing the plasmid were grown on LB plates supplemented with 20 $\mu\text{g ml}^{-1}$ kanamycin at 42.5 °C for 1–2 days to generate merodiploids with plasmid DNA

integrated into their chromosomal DNA. Merodiploids were then grown in LB medium for 1 day at 30 °C to facilitate elimination of redundant DNA from their genomes. Cultures were serially diluted and grown on LB plates without kanamycin for 1 day at 30 °C so that colonies lacking *alkD* could be identified by PCR screening.

Determination of Resistance to DNA Damaging Agents

Overnight cultures of *B. anthracis* Sterne (wild-type, $\Delta alkD$, $\Delta alkC$, $\Delta uvrA$ and $\Delta alkC \Delta alkD$) grown at 30 °C were diluted 1:100 in 100 μ l of LB medium in the presence or absence of varying concentrations of MMS, H₂O₂, HU, NAL, NNKOAc, YTM in a 96-well flat-bottom plate. The plate was incubated at 30 °C with shaking for 20 h, and cell density was measured at 600 nm every hour using a Synergy 2 multi-detector microplate reader. Before each measurement, the plate was gently vortexed to ensure full resuspension of sedimented cells. Experiments were performed in triplicate. Overnight cultures of *B. anthracis* Sterne (wild-type, $\Delta alkD$, $\Delta alkC$, $\Delta uvrA$ and $\Delta alkC$) grown at 30 °C were diluted 1:100 in 5 ml of LB medium and incubated at 30 °C with shaking until early logarithmic phase. Culture aliquots (5 μ l) were then tenfold serially diluted (1:10⁰/10⁻¹–1:10⁻⁴/10⁻⁵) and spotted on LB plates prepared with or without MMS, H₂O₂, HU, NAL, NNKOAc, YTM. Plates were incubated at 37 °C and imaged after 2 days. Experiments were performed in duplicate.

Growth curve and spot assay of *E. coli* were performed in the similar manner as *B. anthracis*. *E. coli* wild-type MV1161 and $\Delta alkB$ MV2029 were transformed by empty vectors (pTrc19A; pBR322) and AlkC expression constructs (pTrc19A-bcAlkC; pBR322-bcAlkC with *alkB* promoter). The cells were grown in LB medium or LB plates containing

MNU. Cell densities were recorded and plates were imaged at the indicated time. *E. coli* wild-type MV1161, $\Delta alkB$ MV2029, and vector pTrc99a were gifts from Michael Volkert (University of Massachusetts Medical School). pBR322 was gift from Neil Osheroff (Vanderbilt University). DNA fragment of *alkC* with *alkB* promoter was synthesized and subcloned into pBG322 by GeneScript.

RNA Quantification

Cell culture of *B. anthracis* wild-type was reseeded at 1:100 from an overnight culture and spiked with MMS or NNKOAc at mid-log growth phase and incubated for 30 minutes. Cells were harvested by adding equal volume of ice cold acetone-ethanol. Cell pellet was resuspended in RNA trizol and lysed in bead beater tubes containing Lysing Matrix B. RNA was isolated using chloroform and Qiagen RNeasy Prep Kit. DNA contamination and DNase were removed using RQ1 (Promega) and Qiagen RNeasy miniprep RNA Cleanup protocol. RNA was quantified by Nanodrop on BioTek and used to generate cDNA. cDNA was amplified and quantified by Bio-Rad iQ SYBR Green Supermix and 3-step with melt curve program.

Base Excision Assays

Base excision using oligonucleotide substrates were carried out as previously described (Mullins, 2013). Briefly, to measure release of methylbases from double-strand RNA (dsRNA) and single-strand RNA (ssRNA) by *Pseudomonas fluorescens* AlkC and *Bacillus cereus* AlkC, 1mA, or 1mA, 3mC, 1mG, and 3mT, were incorporated into the RNA sequence (GACCACTACACCXATTCCTTACAAC) or DNA sequence d(GACCACTACACCXATTCCTTACAAC) at the underlined position by solid-phase

synthesis by Midland Certified Reagent Company, and annealed to the complementary strand in annealing buffer (10 mM MES pH 6.5 and 40 mM NaCl) to generate dsRNA or dsDNA. The lesion DNA strand has 6-carboxyfluorescein (FAM). 10 μ M protein was incubated with ssRNA, dsRNA, ssDNA, dsDNA for 24 hours at 21 °C in 50 mM HEPES pH 8.5, 100 mM KCl and 10% (v/v) glycerol (PfA1kC) or at 35 °C in 25 mM HEPES pH 7.5, 50 mM KCl and 5% (v/v) glycerol (BcA1kC) in a 50 μ L reaction. Products released from RNA and DNA were quantified by HPLC-MS/MS and gel electrophoresis as described previously (Mullins et al, 2013; Shi et al, 2018).

CHAPTER IV

STRUCTURAL BIOLOGY OF HEAT-LIKE REPEAT FAMILY OF DNA GLYCOSYLASES

Abstract

DNA glycosylases maintain genome integrity by removing aberrant nucleobases as the first step of the base excision repair pathway. The HEAT-like repeat (HLR) DNA glycosylase superfamily contains DNA glycosylases that specifically remove a variety of cationic alkylated DNA bases, as well as non-enzymatic proteins with varied DNA binding functions. In contrast to glycosylases in other superfamilies, HLR AlkC and AlkD, engage damaged DNA and excise the nucleobase without flipping the base lesion into the active site and stabilizing the DNA duplex with plug residues. Owing to the non-base flipping mechanisms, AlkD efficiently excises the bulky yatakemycin-modified DNA adduct (YTMA), while AlkC significantly enhances the catalysis of 3-methylcytosine (3mC), and 3-methyladenine. In contrast to AlkC and AlkD, BER activity for HLR proteins AlkD2 and AlkF has not been identified. Apart from their catalytic site, the most notable differences for individual family of HLR proteins are the N-terminal cap. This superfamily exhibits a diverse range of substrates. This chapter focuses on reviewing the chemical mechanisms of AlkC and AlkD in the context of structures, with emphasis on the comparison of the catalytic pocket and therefore substrate selectivity. Together with a thorough phylogenetic and functional comparison between different families, we outline the implications of this novel protein folds composed of short α -helical repeats.

*This chapter will be published as a review article in BioEssays.

Introduction

A number of cellular metabolites and environmental toxins generate a diverse spectrum of alkylated nucleobase modifications that thwart normal DNA processes (Friedberg, 2008; Lindahl, 1993). The cytotoxic and mutagenic effects of alkylation damage can lead to genomic instability and disease, while also making some genotoxic DNA alkylating agents effective anticancer and antimicrobial chemotherapeutics (Sedgwick, 2004; Tubbs & Nussenzweig, 2017; Wyatt & Pittman, 2006). Alkylation damage is repaired in the cell by several modification-specific pathways. Bulky, crosslinked, and helix-distorting lesions typically are processed by nucleotide excision repair (NER), whereas small modifications are removed by direct reversal and base excision repair (BER) mechanisms. BER is initiated by a lesion specific DNA glycosylase that liberates the modified nucleobase from the DNA backbone via hydrolysis of the N-glycosidic bond (**Figure 31A**). This reaction generates an apurinic/apyrimidinic (AP) site that is subsequently excised by an AP endonuclease and replaced by a gap-filling DNA polymerase.

Alkylation of DNA nucleobases is one of the most common forms of DNA damage. Consequently, alkylation specific DNA glycosylases are widely distributed across all domains of life. These enzymes can be classified into one of three distinct structural superfamilies with overlapping substrate specificities (Brooks et al, 2013). Human AAG adopts a unique $\alpha\beta$ -fold (Lau et al, 1998), while those from unicellular organisms, including yeast MAG and Mag1 and *E. coli* AlkA and Tag, adopt the helix-hairpin-helix (HhH) fold observed in other oxidation-specific glycosylases (Adhikary & Eichman, 2011; Brooks et al, 2013; Drohat et al, 2002; Labahn et al, 1996; Rubinson et

al, 2009). Like other previously defined glycosylase superfamilies, both AAG and HhH architectures scaffold a general DNA binding surface and an active site pocket, the latter of which binds the target nucleobase that has been extruded out of the DNA helix in a process known as base flipping (Roberts & Cheng, 1998). The remodeled helix is stabilized by intercalation of one or more side chains into the void generated by the everted base. The nucleobase binding pocket in large part defines the enzyme's substrate specificity and places a limit on the size of the chemical modification that can be accommodated. Whereas AAG, AlkA, and MAG recognize a wide spectrum of alkylated and even deaminated nucleobases, the constitutively active Tag enzyme has a tight specificity for 3-methyladenine generated endogenously (Rubinson et al, 2009).

A third superfamily of DNA glycosylases was recently defined by bacterial AlkC and AlkD, which have a distinct specificity for cationic N3- and N7-alkylguanines and a unique construction from tandem α -helical repeats related to HEAT motifs (Huntington/Elongation/A-subunit/Target-of-rapamycin) (Alseth et al, 2006; Dalhus et al, 2007). HEAT repeats are pairs of antiparallel α -helices that stack in parallel arrays to form extended superhelical or C-shaped structures, which often form scaffolds for large multiprotein assemblies, including those involved in chromatin maintenance and remodeling (Neuwald & Hirano, 2000; Perry & Kleckner, 2003; Sibanda et al, 2010). HEAT domains are also involved in binding and transporting protein ligands within the inner channel of the superhelix (Andrade et al, 2001; Grove et al, 2008; Yoshimura & Hirano, 2016). More recently, this channel has emerged as a nucleic acid binding surface in proteins with diverse functions, including RNA nuclear export, regulation of

mRNA stability, and chromosome segregation (Kschonsak et al, 2017; Lahr et al, 2015; Okada et al, 2009).

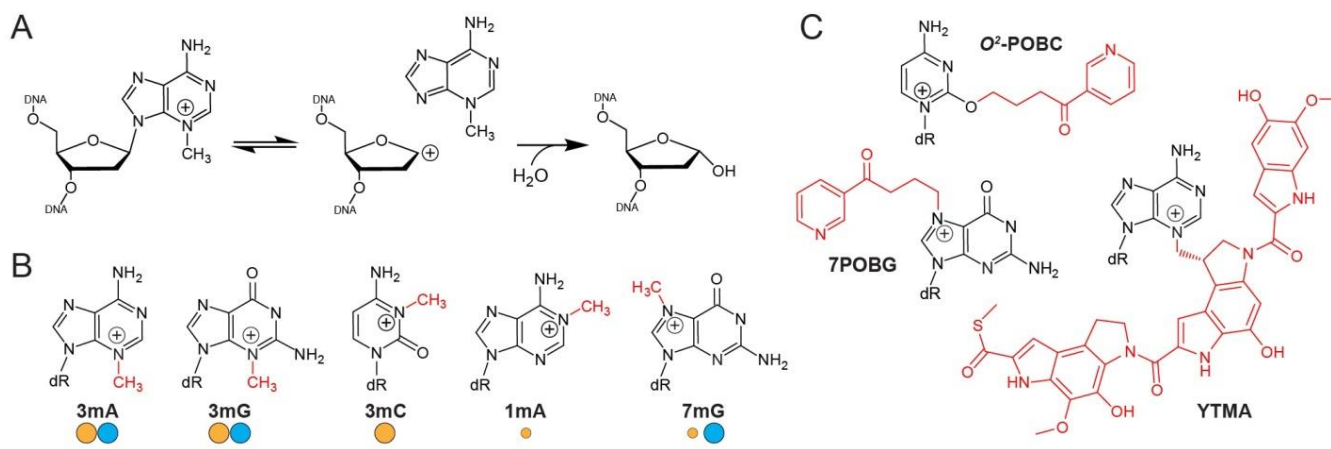


Figure 31. Base excision repair and alkylated lesions removed by AlkC and AlkD. **A**, N-glycosidic bond is cleaved to release cationic 3mA nucleobase from DNA backbone followed by nucleophilic water attacking the oxocarbenium intermediate to generate an abasic (AP) DNA. **B**, Methylation substrates of AlkC and AlkD. The relative enzyme activities of AlkC and AlkD toward different substrates are shown as yellow and blue circles, respectively. 3mA, 3-methyladenine; 3mG, 3-methylguanine; 3mC, 3-methylcytosine; 1mA, 1-methyladenine; 7mA, 7-methyladenine. **C**, Bulky lesions excised by AlkD. O²-POBC, O²-[4-3-(pyridyl)-4-oxobut-1-yl]-cytosine; 7POBG, 7-[4-3-(pyridyl)-4-oxobut-1-yl] guanine; YTMA, N3-yatakemycyladenine.

The HEAT-like repeat (HLR) glycosylases AlkC and AlkD have illustrated how this versatile helical repeat architecture can serve as a DNA damage sensor in addition to a non-specific DNA binding platform (Rubinson & Eichman, 2012), and have provided new insight into repair of DNA alkylation by multiple pathways. In contrast to the base-flipping AAG and HhH glycosylases, AlkC and AlkD are unique in that they do not use a base flipping mechanism to recognize and cleave their substrates, and instead, interact with damaged nucleotides while stacked in the DNA helix. Consequently, these enzymes recognize a chemically different set of alkyl substituents previously associated with direct reversal and NER (**Figure 31B,C**). AlkC is capable of excising 3-

methylcytosine (3mC) and 1-methyladenine (1mA) (Shi et al, 2018), which are also directly repaired by oxidative demethylation by the AlkB family of dioxygenases (Fedeles et al, 2015; Mishina & He, 2006) (**Figure 31B**). The non-base flipping mechanism of AlkD enables it to remove bulky adducts, including pyridyloxobutyl (POB) adducts of guanine and cytosine (**Figure 31C**). Likewise, *Bacillus cereus* AlkD and its ortholog in *Streptomyces* sp. TP-A0356, YtkR2, excise adenine modified by yatakemycin, a bulky and highly toxic *Streptomyces* secondary metabolite that belongs to the duocarmycin/CC-1065 family of natural products, which are known NER substrates (Jin et al, 2001; Kiakos et al, 2007; Selby & Sancar, 1988) (**Figure 31C**).

In addition to AlkC and AlkD enzymes, two distinct HLR proteins lacking base excision activity—AlkD2 and AlkF—have been identified. AlkD2 has no detectable DNA affinity, whereas AlkF has an affinity for branched DNA structures (Backe et al, 2013; Mullins et al, 2015a). In this chapter, we will discuss the structural basis for the unique specificities of the non-base flipping mechanisms of AlkC and AlkD, as well as the structural features that alter DNA binding properties of AlkD2 and AlkF.

Phylogeny of the HLR superfamily

The HLR superfamily consists of four distinct clades typified by AlkC, AlkD, AlkD2, and AlkF (**Figure 32**). We searched for AlkC, AlkD, AlkD2 and AlkF homologs from a PSI-BLAST search against the NCBI non-redundant protein sequence database using *Pseudomonas fluorescens* AlkC, *Bacillus cereus* AlkD, *Streptococcus mutans* AlkD2, and *Bacillus cereus* AlkF as queries. The phylogenetic tree shown in **Figure 32A** was constructed from multiple sequence alignment after redundant sequences were

excluded using the neighbor-joining tree (Shi et al, 2018). All homologs were primarily confined to bacteria, although a small percentage of AlkD homologs were found in eukaryotes. The AlkC and AlkD clades comprise the majority of HLR sequences, with 43% and 32% of the total, respectively (**Figure 32A**). Over 16% of AlkD orthologs are found in *Proteobacteria*, 69% are distributed among *Firmicutes*, *Actinomycetes*, and *Bacteroidetes*, and only 5% are distributed in eukaryotes and archaea (**Figure 32B**). Similarly, 94% of AlkC orthologs are predominantly found in *Actinomycetes*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. AlkF is the third most populated clade, comprising 21% of the total HLR sequences. The original identification of AlkF also defined a separate family of AlkG proteins (Backe et al, 2013). AlkF and AlkG are closely related (the two paralogs from *Bacillus cereus* share 35% similarity and 52.7% identity), and thus were grouped together in the AlkF clade in our analysis. The vast majority (94%) of AlkF/AlkG orthologs are distributed among *Firmicutes* and *Actinobacteria*. In contrast to the other groups, AlkD2 represents only 4% of the total HLR sequences and is primarily found in *Bacteroidetes*. Thus, the enzymes AlkC and AlkD are the most abundant and are distributed among bacterial phyla, whereas the non-enzymatic AlkF and AlkD2 sequences represent ~25% of the total HLR family and are rarely found in *Proteobacteria*.

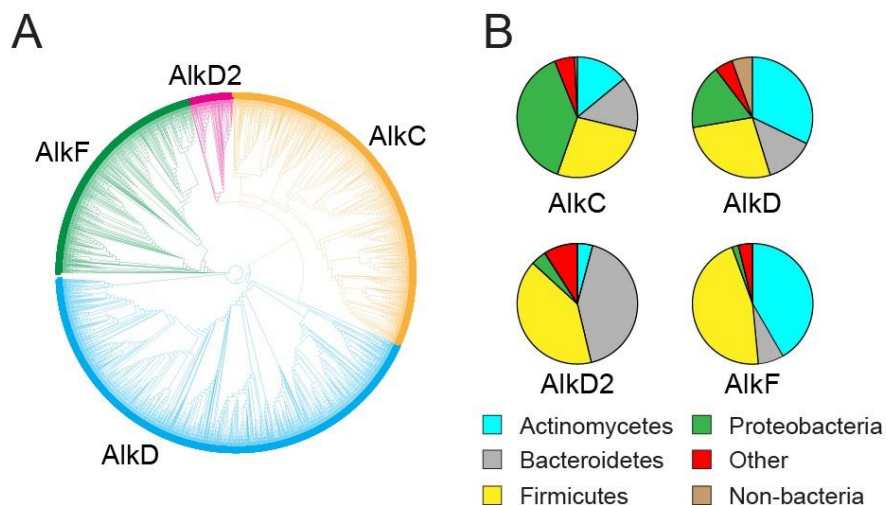


Figure 32. Phylogenetic analysis of HLR superfamily proteins. **A**, Phylogenetic tree composed of 968 AlkC (yellow), 1299 AlkD (blue), 123 AlkD2 (magenta) and 629 AlkF (green) generated and visualized using Clustal Ω and iTOL. **B**, The distribution of HEAT-like repeats (HLR) proteins in diverse phyla is shown in pie chart.

The general HLR structure

Each member of the HLR superfamily is composed of an N-terminal helical bundle (NTB) followed by 5 HLR motifs, which together form a C-shaped structure (**Figure 33**). The C-terminal helices of each HLR line the concave surface, which is the most conserved region (**Figure 33A**). In AlkC and AlkD this surface is highly positively charged as a result of the distribution of lysine and arginine side chains important for DNA binding (Rubinson et al, 2008). In contrast, the positive charges on the surfaces of AlkD2 and AlkF are not concentrated within the concave cleft. Each of the five HLR motifs (HLR1-5) have a distinct structure (Rubinson et al, 2008) that is largely conserved in each of the four HLR proteins, with a few notable exceptions (described in the following sections). The HLR families differ in the relative positions and stacking of adjacent HLR motifs, which affects the overall solenoid structure. For example, AlkD2 is a hybrid of AlkC and AlkD families in that HLR1-2 shows a high degree of structural

similarity to AlkD, and HLR3-5 are more similar to AlkC (**Figure 33A**). In addition to differences in HLR1-5 packing, some HLR families are also distinguished by inserted motifs within the HLRs (**Figure 33B**). The AlkC family is largely defined by the presence of an inserted DNA binding loop (DBL) between HLR3 and HLR4 that is present in both AlkC α and AlkC β subtypes, which differ in the presence of an additional 100-residue immunoglobulin (Ig)-like domain at the C-terminus. AlkF contains an inserted β -hairpin motif between helix α L and α M of HLR5 that affects its ability to bind branched DNA structures (Backe et al, 2013).

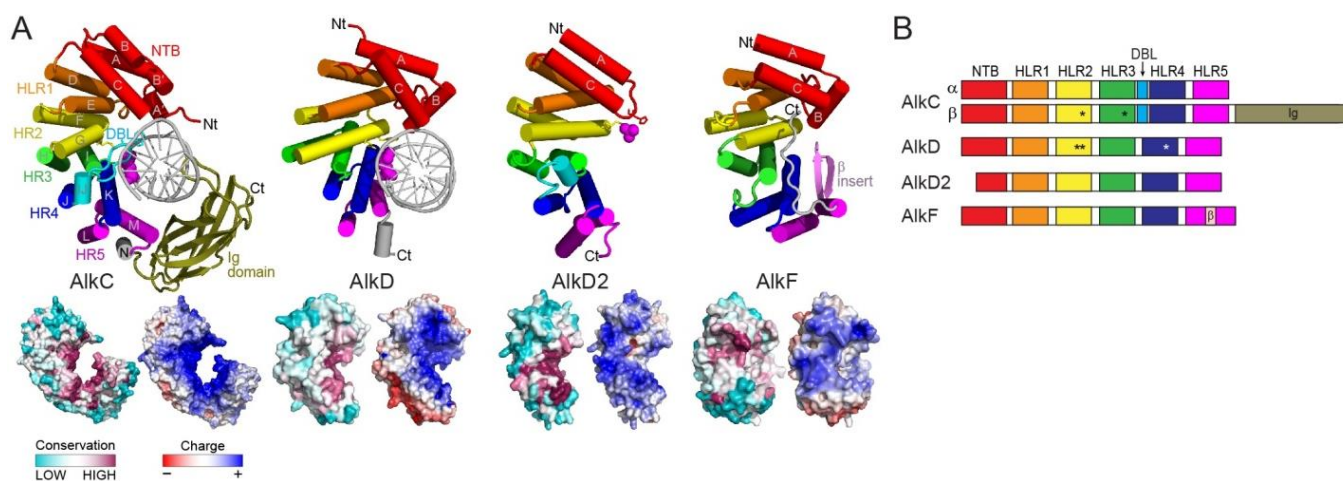


Figure 33. Structural comparison of HLR superfamily proteins. **A**, Structures of BcAlkD in complex with 1aR-DNA (PDB 5CLB), PfAlkC in complex with 1aR-DNA (PDB 5VHV), AlkD2 (PDB 4X8Q), and AlkF (PDB 3ZBO). N-terminal bundle, red; HEAT-like repeats, orange, yellow, green, navy and purple; β -sheet, paleyellow; Other helix or loop, grey. HLR protein sequence conservation (purple, high; cyan, low) and electrostatic surface potential (blue, positive; red, negative) are mapped onto the protein surface. **B**, Schematic of protein domains of AlkC, AlkD, AlkD2, and AlkF. Catalytic residues are highlighted with asterisk. Ig, immunoglobulin-like domain.

The largest structural differences between the four HLR family members are found in the NTB, which also plays a role in DNA binding. The NTB of AlkD and AlkF

are composed of three helices (α A, α B, α C), in which antiparallel helices α A and α C stack onto HLR1, leaving helix α B to protrude into the concave surface. In AlkD, helix α B contacts DNA (Mullins et al, 2014; Mullins et al, 2015b; Rubinson et al, 2010). The NTB of AlkD2 has the same antiparallel arrangement and packing of α A and α C, but lacks the α B helix, which greatly reduces DNA binding affinity of AlkD (Mullins et al, 2015a). In contrast, the NTB of AlkC is highly divergent from AlkD, AlkD2, and AlkF, largely in the orientation of the α A and α B helices (Shi et al, 2018). In AlkC, α A and α B are each split into two co-linear helices that pack together with helix α C. Interestingly, helix α A resides in the same location as in the other three families, but runs in the opposite direction. Consequently, the N-terminus of helix α A in AlkC points into the inner surface of the solenoid and contacts DNA (**Figure 33B**) (Shi et al, 2018). These structural differences endow each HLR protein with a particular substrate specificity, which we describe in more detail below.

AlkD Uses a Non-Base-Flipping Mechanism to Excise Bulky Lesions

AlkD excises a variety of cationic alkylated DNA lesions. The *Bacillus cereus* enzyme (BcAlkD) was characterized initially to remove 3mA, 3-methylguanine (3mG), and 7-methylguanine (7mG), and to provide Tag and AlkA deficient *E. coli* with resistance to methylating agents (Alseth et al, 2006). BcAlkD was subsequently shown to also remove bulky, cationic POB adducts (Rubinson et al, 2010). More recent work revealed that AlkD proteins likely evolved specifically to repair YTM-DNA damage. The *Streptomyces* AlkD homolog, YtkR2, excises YTM-adenine adducts and protects *E. coli* challenged with either YTM and the methylating agent methyl methane sulfonate

(MMS). Deletion of YtkR2 from *Streptomyces* sp. TP-A0356 reduces YTM production by 40% (Xu et al, 2012). BcAlkD also excises YTM-adenine lesions *in vitro*, and deletion of AlkD from *Bacillus anthracis* ($\Delta alkD$) sensitized cells to low levels of YTM toxicity (Mullins et al, 2017a). In contrast, $\Delta alkD$ cells were not sensitive to MMS, likely because of the AAG, AlkA, and AlkC orthologs present in *Bacillus* (Alseth et al, 2006).

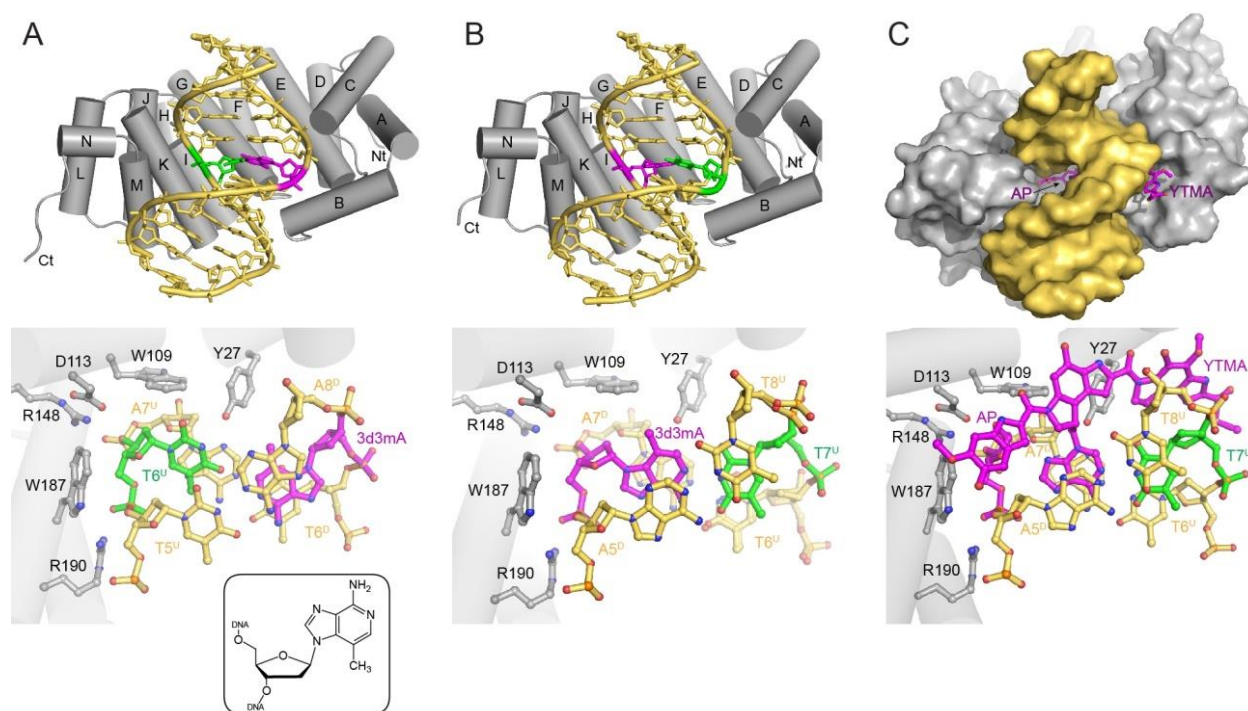


Figure 34. Comparison of catalytic irrelevant and relevant AlkD structures. **A**, **B**, Structure and active site of BcAlkC in complex with 3-deaza-3-methyladenine (3d3mA) DNA in catalytic irrelevant (A, PDB 3JX7) and relevant (B, PDB 5CL8) conformations. The protein is colored in grey, DNA gold, 3d3mA magenta, and opposite thymine green. The active site residues are shown as grey sticks. **C**, Ternary product complex and close view of catalytic active site of AlkD, AP-DNA and yatakemycynyladenine (YTMAdenine) (PDB 5UUF). AlkD is colored grey, DNA gold, AP-site and YTMAdenine magenta, and lesion opposite thymine green.

A homology model of BcAlkD, constructed from a crystal structure of an ortholog from *Enterococcus faecalis* (PDB 2b6c) determined by the Midwest Center for Structural

Genomics consortium, enabled identification of functionally important residues, which were corroborated by a subsequent BcAlkD crystal structure (Rubinson et al, 2008). The active site consists of electrostatically paired Asp113-Arg148 side chains surrounded by a several aromatic residues (Tyr27, Trp109, Phe179, Phe180, and Trp187) clustered on the concave surface of the protein (**Figure 34**). Mutational analysis confirmed a role for Asp113, Arg148, Trp109, Trp187 in glycosylase activity in vitro and in cells, providing direct evidence for catalytic activity by an HLR protein (Dalhus et al, 2007; Rubinson et al, 2008). Structures of AlkD bound to DNA containing either a 3-deaza-3-methyladenine (3d3mA) substrate analog (**Figure 34A**), a tetrahydrofuran (THF) abasic site product analog, or a G•T mismatch illustrated how AlkD encircles half of the DNA duplex to place the shallow active site surface against the DNA backbone of the aberrant base pair (Rubinson et al, 2010). These initial structures revealed that the AlkD active site is distinctly different from traditional base flipping glycosylases in that there was neither a putative extrahelical nucleobase binding pocket nor an obvious helix intercalating residue. Instead, the enzyme was able to form a specific complex with a variety of non-Watson-Crick base pairs without contacting the damage itself. However, the protein-DNA contacts important for catalysis were not discernable because the damaged nucleotides resided on the solvent exposed strand and not adjacent to the active site.

In order to trap a catalytically relevant complex containing the damaged nucleotide in contact with the protein, we took advantage of a nearest neighbor effect on AlkD catalysis in which base excision activity was weakest with thymine 3' to the lesion (E.H. Rubinson and B.F. Eichman, unpublished). Modeling thymine immediately 3' to

the aberrant base pair in the crystal structures ($A7^U$ in **Figure 34A**) generated a potential steric clash between the thymine methyl group and Arg190, providing a structural rationale for the sequence effect and a strategy for preventing binding of the enzyme to the strand opposite the lesion. Indeed, altering the original $5'T(3d3mA)A^{3'}/5'TTA^{3'}$ sequence to $5'A(3d3mA)A^{3'}/5'TTT^{3'}$ resulted in a structure with the modified nucleotide positioned against the active site (**Figure 34B**) (Mullins et al, 2015b). Although 3d3mA was refractory to excision by AlkD at neutral pH, it was clear from the electron density that hydrolysis of the N-glycosidic bond occurred in the crystals to produce a mixture of intact 3d3mA-DNA substrate and AP-DNA + 3d3mA nucleobase product, thus confirming the catalytic relevance of this new DNA orientation. As in the previous mismatch structures, the 3d3mA nucleotide remains stacked within the DNA helix in a sheared orientation, with the 3d3mA displaced toward the minor groove (**Figure 34B**). As a consequence, the 3d3mA deoxyribose is cradled by active site residues Asp113, Trp109, and Trp187, along with a water molecule positioned by the Asp113 carboxylate for in-line attack of the deoxyribose C1' carbon. In addition to substrate and product complexes generated by harvesting crystals at various times after protein-DNA mixing, an AlkD-DNA complex mimicking the oxocarbenium intermediate was determined using 1-azaribose (1aR)-DNA and 3mA nucleobase. Strikingly, the positions of the protein, DNA, and free nucleobase did not differ significantly among the substrate, intermediate, and product complexes, indicating that the protein structure is tailor-made to place minimal constraint on a non-Watson-Crick base pair for N-glycosidic bond hydrolysis without base flipping.

Perhaps the most striking feature of the structures of AlkD in complex with methylated DNA is that the active site residues do not directly contact the target nucleobase. Instead, the protein facilitates base excision exclusively through interactions with the lesion deoxyribose. The position of the Asp113 carboxylate is consistent with a role of this side chain in positioning the water nucleophile for in-line attack of the anomeric C1' carbon, and in stabilization of the developing positive charge on the oxocarbenium reaction intermediate, as proposed for other DNA glycosylases (Werner & Stivers, 2000). In addition, the CH- π interactions between Trp109 and Trp187 indole side chains with the lesion C2' and C4'/C5', respectively, suggested that these residues also play a role in stabilization of positive charge on the deoxyribose (Mullins et al, 2015b). Aromatic residues in base flipping glycosylases either bind the extrahelical base or plug the gap through π - π stacking, and thus interfere with substrate binding (Hollis et al, 2000). In contrast, mutation of the highly conserved Trp109 and Trp187 in AlkD resulted in a 100-fold decrease in catalysis without compromising DNA binding (Parsons et al, 2016).

Due to the non-base-flipping catalytic mechanism, AlkD is capable of repairing non-duplex-distortion bulky lesions *in vitro* (Mullins et al, 2015b). YTM, is highly toxic, and belongs to the antimicrobial and antitumor agent of CC-1065 and duocarmycin antibiotic family (Igarashi et al, 2003). YTMA increases the melting temperature of duplex DNA by 36 °C through intensive CH- π interactions, and has extraordinarily slow spontaneous rate. The enhanced barrier is overcome by AlkD, which pries open the minor groove around the lesion, allowing catalytic residues Asp113, Trp109, Trp187, and the nucleophilic water to access to the lesion's N-glycosidic bond (**Figure 34C**).

Single mutations of Asp113, Trp109 and Trp187 to alanine, though with significantly reduced activity, remain at least 10^4 -fold higher than the spontaneous depurination (Mullins et al, 2017a).

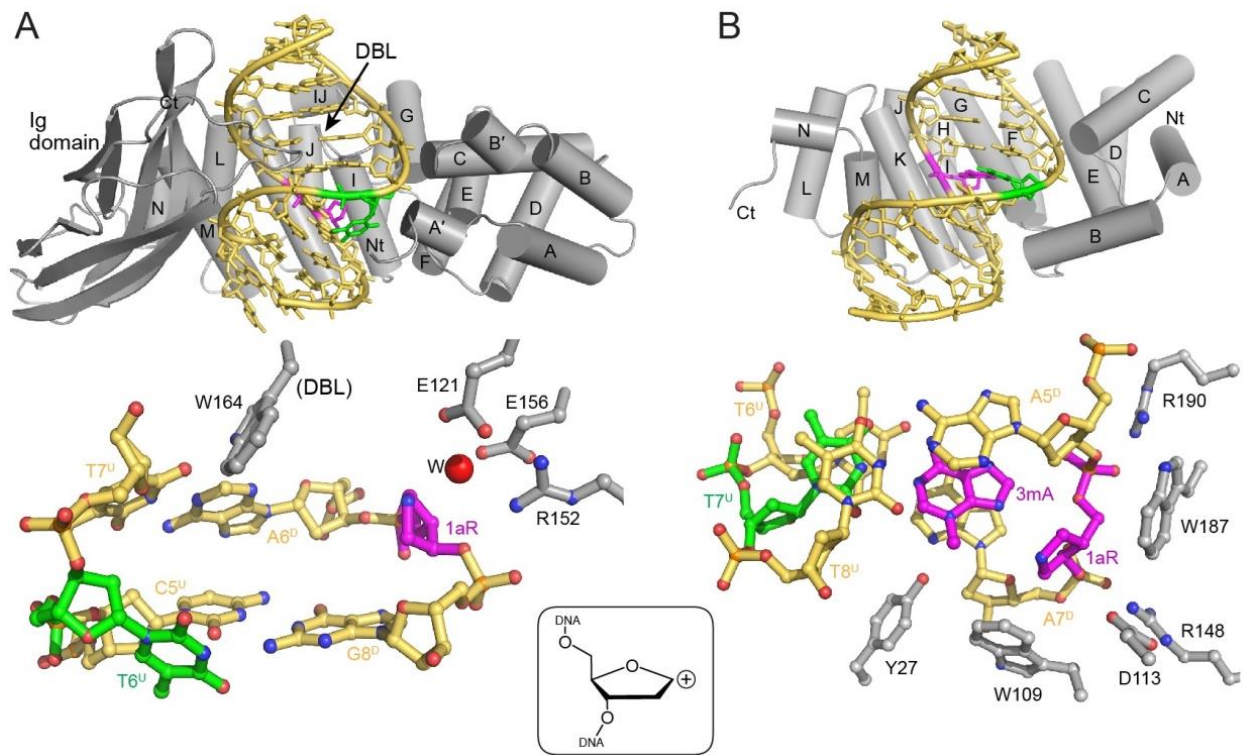


Figure 35. Structural comparison of AlkC- and AlkD-DNA complexes in context of lesion removal. Structures and close views of catalytic active sites in AlkC (A) and AlkD (B) in complex with 1aR-DNA. AlkD is colored grey, DNA gold, 1aR and 3mA magenta, lesion opposite thymine green, and water red.

We have shown that both AlkD-involved BER and UvrA-involved NER pathways confer resistance to YTM in *B. cereus* (Mullins et al, 2017a; Mullins et al, 2015b). The double knockout $\Delta alkD \Delta uvrA$ strain show additive sensitivity toward YTM compared to single-knockouts. However, given the weak resistance of wild-type bacteria to YTM, none of these pathways repairs lesions sufficiently *in vivo*. It raises the question that why toxicity of YTM is only slightly alleviated by AlkD even though AlkD exhibits high efficiency of YTM removal *in vitro*. That is explained by strong product inhibition of AlkD

to AP-DNA that prevents AP-DNA being further processed by downstream endonuclease. Interestingly, YtkR2 from *Streptomyces* sp. TP-A0356, an ortholog of AlkD, is in the YTM biosynthetic gene cluster. Its robust activity for YTMA is considered as part of a self-defense mechanism to safeguard genomic against endogenous DNA modification drugs. It is suggested that *ytkR2* together with four other genes in the biosynthetic gene cluster, including an AP endonuclease, are responsible for biologically antibiotic resistance to YTM. It is possible that this gene cluster which is adapted to YTM production may work differently from AlkD to efficiently hand over the abasic site product to the subsequent step in the pathway. Repair of the non-traditional lesion by AlkD family of proteins helps cells outcompete other soil bacteria by providing additional antimicrobial resistance and safeguarding genomic against naturally DNA damaging chemicals (Mullins et al, 2017a).

AlkC Uses a Non-Base-Flipping Mechanism to Select for Small Alkyl-adducts

AlkC cleaves a narrow range of protonated alkylpurines, including the minor groove lesion 3mA, as well as 3-methyl-cytosine (3mC) and 1-methyl-adenine (1mA) which are usually repaired by AlkB involved direct reversal repair. In contrast to a broader range of substrates that AlkD can remove, AlkC displays low activity for 7mG and no activity for bulky lesions (Alseth et al, 2006; Shi et al, 2018).

Unlike the other HLR superfamily DNA glycosylases comprised of single domain, 70% of AlkC have acquired the additional Immunoglobulin (Ig) -like domain, which is essential for substrate binding (Shi et al, 2018). The structure of AlkC in complex with transition state analog 1'-aza-2',4'-dideoxyribose (1aR) shows that like AlkD, AlkC is a

non-base-flipping glycosylase (**Figure 35A**). Despite of the similar HLR architecture and non-base flipping mechanism between AlkC and AlkD, they utilize different non-base-flipping strategies (Shi et al, 2018). AlkD recognizes and catalyzes damaged base cleavage without disrupting the DNA base stack. However, in contrast to the modest DNA bend by AlkD, AlkC bends 1aR-containing DNA by 60° from the minor groove. This sharp bending pushes the positively charged lesion against the concave cleft of the protein, thus exposing the lesion adjacent to the active site pocket. Despite of the severe kink at the damage site, the lesion is not flipped out the DNA duplex. Instead of using a Trp-Asp-Trp motif to only contact DNA backbone, AlkC inserts its active site of Glu-Arg cluster into DNA to access to the lesion. AlkC employs a catalytic pocket mediated by a network of alternating charge residues Glu121-Arg152-Glu156-Arg159-Asp203. The phosphates flanking the damaged base are electrostatically stabilized by positively charge residues Arg152 and Arg159, while Glu121 and Glu156 are well positioned to stabilize the developed positive charge on the transition state intermediate in addition to activating the nucleophilic water (**Figure 35A**). Substitution of either of two glutamate residues with alanine in PfAlkC greatly decreases 3mA activity, and abolishes the catalytic activities for 1mA and 3mC lesions. This is consistent with the observation that alanine mutation of catalytic active site residues can distort the pocket conformation, therefore affecting the catalytic activity. Different from AlkD, a highly conserved loop between helices I and J in AlkC occupies the catalytic active site, precluding the access of any bulky lesion. The substrate specificity of AlkC is speculated to be constrained by not only π -stacking between the damaged base and DNA or protein surface, but also cation- π interaction between the lesion and conserved

loop residues like Trp164. The strict arrangement of the active site limits substrate specificities of AlkC (Shi et al, 2018).

AlkD2, a B-Helix-Absent HLR Superfamily Member

AlkD2 is not a DNA glycosylase and does not exhibit any significant DNA binding affinity (ref). While maintaining more sequence similarity to AlkC, the overall structure of AlkD2 resembles the structure of AlkD with the major discrepancy in the absence of B-helix in AlkD2 (Mullins et al, 2015a). AlkD2 is the only HLR protein that lacks B-helix (**Figure 33A**), an essential element in HLR DNA binding proteins. Instead, it has a substituted short loop. Although AlkD2 only displays fairly weak DNA binding affinity for various types of DNA, it has a rather conserved and positive charged protein-ligand interface. Lack of the penetrating B-helix which directly strengthens the interaction between the minor groove of DNA and the protein, is proposed to lead to loss of affinity for DNA. The *alkD2* gene in *S. mutants* locates in the gene cluster of purine synthesis pathway, suggesting the its potential role in purine synthesis. The loose DNA binding may be relevant to its biological role of passing the reaction intermediate to the following step (Mullins et al, 2015a). Agreeing with this hypothesis, we successfully solved the structure of SmAlkD2 in complex with inosine-monophosphate (IMP), which is an intermediate ribonucleoside monophosphate in purine metabolism (not published).

AlkF Favors Branched DNA Binding

AlkF and its close ortholog AlkG do not display glycosylase activity toward alkylated, oxidized, or other modified DNA nucleotides. However, AlkF shows a specific DNA binding activity toward branched structures (Backe et al, 2013).

As the first solved structure in HLR superfamily, AlkF from *Bacillus cereus* was initially deposited as a hypothetical protein by Midwest Center for Structural Genomics in 2004 (PDB 1t06). Bjørås group redefined and characterized the AlkF family when discovering AlkF to be AlkC and AlkD homologs (Backe et al, 2013). AlkF and AlkG display low millimolar affinity for branched DNA, like Holliday and three-way junctions, albeit the binding is fairly weak compared with low micromolar binding affinity of AlkC and AlkD for lesioned DNA (Backe et al, 2013; Shi et al, 2018).

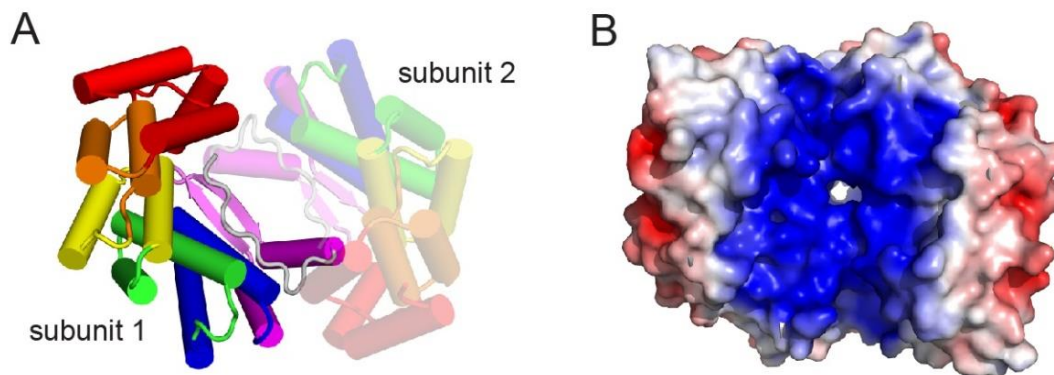


Figure 36. Dimerization of AlkF. **A**, AlkF crystallizes in a dimeric form (subunit 1 and subunit 2). **B**, Electrostatic surface potential (blue, negative; red, positive) of AlkF mapped onto the dimer surface.

The overall concave surface of AlkF is not as negatively charged as other DNA binding HLR proteins, possibly explaining the weak DNA preferences even for branched structures. The structure is rigid except for the flexible 12-amino acid β -hairpin between L-helix and M-helix (**Figure 33A**). This unique β -hairpin covers the concave surface

corresponding to the DNA binding site of AlkD. Substitutions of Arg203, Lys206, and Lys207 with alanine within the β -hairpin greatly reduce preference of AlkF to branched DNA. It is not clear if this unique variant β -hairpin acts as a platform for DNA binding or engages in a novel feature suitable for other unknown function. This β -hairpin is only conserved in some of the AlkF homologs. However, it is difficult to predict from sequences if this structural element is present in other AlkF homologs. Few residues are invariant in AlkF family. They are buried in the inner surface of AlkF and none of the invariant residues are predicted to directly contact with DNA. Intriguingly, in the AlkF structure, the flexible C-terminal region folds back onto the concave surface of the protein, presumably making it in a locked self-inhibitory conformation (**Figure 36A**) (Backe et al, 2013). The cellular function of AlkF has not been well characterized. Despite of the moderate sensitivity of double knockout $\Delta alkF \Delta alkG$ to MMS and NAL, single-knockout mutants $\Delta alkF$ and $\Delta alkG$ only show negligible sensitivity to various types of genotoxic stresses. Besides, they are not essential for cell growth under normal conditions. No evidence has been found that it takes part in DNA damage (Backe et al, 2013). Higher affinity of AlkF family of proteins for branched DNA indicates AlkF may play a role in the cellular process involving branched structures. AlkF was crystallized as a dimer with a tunnel formed by the concave surfaces of the protein, indicating it may function as a dimer and utilize the tunnel for cargo transportation (**Figure 36**).

Comparison between HLR Proteins

Despite of high structural similarity in HLR superfamily, glycosylases within different families display distinct substrate preferences (Backe et al, 2013; Mullins et al,

2015a; Mullins et al, 2015b; Shi et al, 2018). *B. cereus* AlkC and AlkD catalyze repair of DNA adducts. However, *B. cereus* AlkF and *S. mutants* AlkD2, without any detectable DNA glycosylase activity, are structural, but not functional homologs of AlkC and AlkD. Therefore it raises the question about the structural determinants for enzyme specificity.

Prior to the discovery of AlkC and AlkD, a base-flipping mechanism was the generally accepted catalytic mechanism for all known glycosylases, by which removal of the methyl moiety is accompanied by damaged base flipping into the active pocket (Brooks et al, 2013). Unexpectedly, AlkD is capable of removing DNA adducts stacked in the DNA duplex, thus making AlkD competent for bulky adducts like yatakemycin-adducts and pyridyloxobutyl (POB)-adducts *in vitro* in addition to removal of positively charged 3mA and 7mG. Similarly, AlkC accommodates the damaged base without base-flipping. Its substrate selectivity relies on the exposure of lesion at the kinked DNA to a deep protein pocket. AlkC is able to repair 3mC and 1mA besides 3mA. The unique structural features contribute to the repair of non-traditional BER substrates by AlkD and AlkC (Mullins et al, 2015b; Rubinson & Eichman, 2012; Shi et al, 2018). Although complementation of HLR glycosylases in *E. coli* methylpurine-glycosylase depleted $\Delta tag\Delta alkA$ strain rescues methyl methanesulfonate (MMS) sensitivity (Alseth et al, 2006), neither *B. anthracis* $\Delta alkC$ or $\Delta alkD$ is more susceptible to the drug compared to wild-type. Lack of sensitivity to alkylating agents in *alkC* and *alkD* depleted bacteria may be due to the redundancy of five methylpurine DNA glycosylases present in *Bacillus anthracis*.

AlkD2 and AlkF are distributed in diverse bacteria, but none of them exhibit any glycosylase activity and their cellular functions have not been well characterized yet.

AlkD2, the evolutionary intermediate of AlkC and AlkD, shares similar HLR architecture but lacks the critical B-helix for substrate binding. Though AlkD2 has a positive charged concave surface favorable for DNA binding (**Figure 33A**), it does not retain DNA binding specificity. AlkF has binding preferences for branched DNA. The presence of the unique β -hairpin in AlkF was proposed to engage DNA in its major groove.

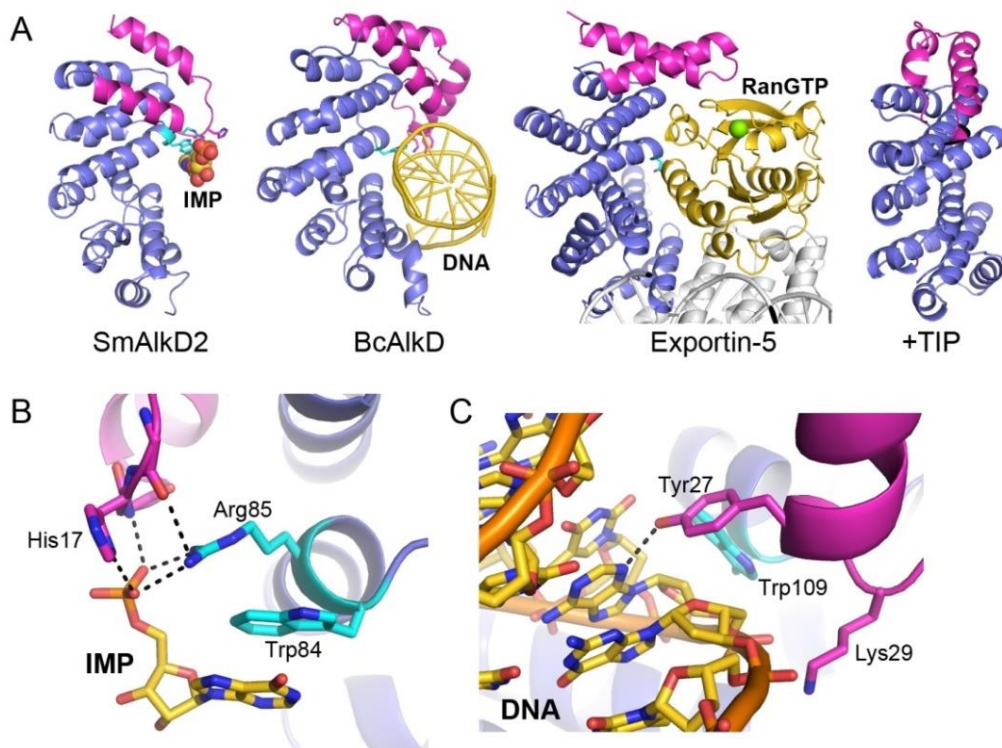


Figure 37. N-terminal bundle of proteins with HEAT-like repeats. **A**, Structures of SmAlkD2 in complex with inosine monophosphate (IMP), BcAlkD in complex with DNA, Exportin-5 in complex with RanGTP, and TIP. N-terminal bundle is colored magenta, HEAT-like repeat blue, and ligand gold. Close views of the interaction between substrate and N-terminal bundle in SmAlkD2 (**B**) and BcAlkD (**C**).

N-terminal helical bundle (NTB) is one critical element for HLR proteins to distinguish substrates. NTB is usually composed of 2- or 3-helix segments, dependent on its substrate preference (**Figure 37A**). In the SmAlkD2–IMP complex, the protein

engages IMP with the NTB. His17, in proximity to the phosphate group of IMP, is in the loop which is substituted with B-helix in other DNA binding HLR proteins. To bind IMP instead of DNA, guanidino group of Arg85 in SmAlkD2 is hydrogen bonded with the phosphate group of IMP. This interaction is further enhanced by the electrostatic interaction between these two functional groups, in which the indole ring of Trp84 is positioned by π - π interaction with the ribose ring of IMP (**Figure 37B**). Interestingly to note, not only for HLR proteins, the NTB also plays important role in holding a variety of substrates in superhelical structures. For example, Exportin-5, which consists of 20 HEAT repeats and transports pre-miRNA, forms U-like structure with RanGTP using the NTB without B-helix (**Figure 37A**). Although Tyr27 in B-helix of AlkD significantly contributes to the binding affinity of damaged DNA, this contact is not conserved and essential for AlkC and AlkF (**Figure 37C**).

AlkC, AlkD, AlkD2 and AlkF are structural homologs, they do not share the same functionalities. With more structural study on HLR superfamily proteins, it will help explicitly explain the mechanism how DNA repair enzymes are evolved into similar architecture but with distinct features, and employ them as a platform to protect against a variety of endogenous and exogenous DNA damaging agents. Interestingly, a fourth superfamily was recently defined by the tandem winged helix-turn-helix structure of bacterial AlkZ, a DNA glycosylase with the surprising ability to unhook interstrand crosslinks derived from the natural product azinomycin B (Mullins et al, 2017b; Wang et al, 2016). The active site of AlkZ does not form an obvious nucleobase binding pocket but is instead located on the surface of the protein so that catalytic residues are predicted to interact with a modified nucleotide that remains stacked in the DNA helix.

The cross-talks between BER and other DNA repair pathways suggest the involvement of catalytic mechanisms associated with broadening of enzyme substrate specificity.

CHAPTER V

DISCUSSION AND FUTURE DIRECTIONS

Glycosylases play a key role in base excision repair (BER) of DNA damaged by endogenous and exogenous toxins and mutagens. AlkC and AlkD belong to the HEAT-like repeat (HLR) family of glycosylases with interesting catalytic and DNA binding activities. In this thesis, I have described work to characterize the selective base excision repair mechanism of alkylpurine DNA glycosylase AlkC (Chapter II) and the cellular function of AlkD (Chapter III). However, several interesting questions remain regarding the AlkC subclades (AlkC α and AlkC β), the role of AlkD homologs YtkR2 and C10R5 in BER in *Streptomyces* to provide resistance against yatakemycin/CC-1065 natural products, and the cellular function of other HLR proteins. In this chapter, I will discuss some of the interesting findings of AlkC and other HLR proteins and propose future work in this area.

Structure-Function Analysis of AlkC

The bioinformatics analysis identified two AlkC subfamilies: the single-domain AlkC α subfamily which resembles the overall HEAT-like repeat (HLR) architecture of AlkD and the two-domain AlkC β subfamily which acquires an additional C-terminal immunoglobulin (Ig)-like domain (Shi et al, 2018). The crystal structure of *Pseudomonas fluorescens* (Pf) AlkC, which belongs to the AlkC β subfamily, provided clues into the differences between AlkC subfamilies (see Chapter II) in addition to the DNA-binding

capabilities of the Ig-like domain of AlkC β . The Ig-like domain of PfAlkC donates nearly 40% of the DNA contacting residues in the protein, and truncation of this domain (PfAlkC Δ C) diminished DNA binding. Similar to DNA binding mediated by the EF-loop in other Ig domains, a Met-Thr-Thr-Arg motif within the EF-loop of PfAlkC contacts both strands of DNA. The β C' strand and the CC'-turn also contribute to interacting and stabilizing the kinked DNA (Shi et al, 2018). To fully understand the role of these interactions in DNA binding, we can mutate the DNA contacting residues in the Ig-like domain to alanine and test the effect on binding. It would also be interesting to address the how the Ig-like domain works with the HLR domain. In the crystal structure of PfAlkC in complex with tetrahydrofuran (THF)-DNA, PfAlkC distorts and restructures the DNA into two different conformations (**Figure 20**). The difference in DNA distortion results from the flexible linker between relative rigid HLR and Ig-like domains, and the tight fit between PfAlkC and DNA. It is therefore interesting to speculate this Ig-like domain anchors DNA onto HLR domain and enhances DNA binding. This also agrees with our analysis that PfAlkC has higher activity for 1mA and 3mC than BcAlkC, and AlkC β is the dominant population in AlkC family. Because the Ig-like domain has not been observed in other HLR superfamily proteins, it is possible that the ancestor of AlkC containing species rearranged the existing HLR domain and acquired this Ig-like domain during evolution to improve their antimicrobial resistance (Trivedi et al, 2015).

Because AlkC α is able to function without the Ig-like domain, we were interested to know whether AlkC α contains specific structural elements that compensate for DNA binding in the absence this domain. Based on the multiple sequence alignment between AlkC α and AlkC β homologs and the structural comparison between the PfAlkC structure

and a homology model of *B. cereus* AlkC (BcAlkC, AlkC α), we identified a unique insertion of *TEVGTFITNG* in BcAlkC, which is absent in PfAlkC and predicted to stabilize DNA bend (**Figure 16**). Consistent with our hypothesis, deletion of this short sequence in BcAlkC severely reduced the 1-methyladenine (1mA) and 3-methylcytosine (3mC) excision activities. However, the activity was not recovered by insertion of this sequence in PfAlkC Δ C (**Figure 38**). It is likely that this BcAlkC deletion adversely destabilized and inactivated the protein. Alternatively, multiple structural elements in the HLR domain of BcAlkC could account for the binding functionality, explaining why the PfAlkC Δ C insertion did not rescue its activity.



Figure 38. Schematic of BcAlkC (AlkC α) and PfAlkC (AlkC β) domains. The 9-residue insertion of BcAlkC is shown in blue. The predicted and observed enzymatic activity towards 1mA and 3mC are annotated as “active” (Y) and “not active” (N).

We still have no structural information for the AlkC α subfamily. The AlkC α subfamily is divided into AlkC α 1 and AlkC α 2 groups, the latter of which constitutes the majority of AlkC proteins (**Figure 10**). The AlkC α homologs characterized in Chapter II belong to the AlkC α 1 group, which is the closer relative to AlkC β than is AlkC α 2. The activity of AlkC α 2 has not yet been determined. Therefore, we selected four AlkC α 2 proteins (*Aquimarina latercula* AlkC α 2, *Bacillus thuringiensis* AlkC α 2, *Actinomyces*

viscosus AlkCa2 and *Pseudomonas fluorescens* AlkCa2) and cloned them into protein expression vector pBG102 for future structure-function analysis.

PfAlkC uses the non-base-flipping mechanism where the damaged base is slightly rotated and exposed to the protein active site through a sharp bend in DNA. Catalytic residues Glu121 and Glu156 stabilize the positive charge developed on the sugar and deprotonate the nucleophilic water to attack the N-glycosidic bond. The presence of two glutamate residues remarkably elevates its catalytic power and enables AlkC to cleave the more stable 1-methyladenine (1mA) and 3-methylcytosine (3mC) lesions from DNA. Modeling 3-methyladenine (3mA) into the PfAlkC structure illustrates that the conformational constraints lead to the selectivity of PfAlkC for 3mA (**Figure 18C**). It will be interesting to know if substitution of Glu121 or Glu156 with Gln or Asp will affect substrate catalysis to the same degree as Glu to Ala mutations.

Five putative methylpurine DNA glycosylases are present in *B. anthracis*. This high redundancy reflects the high load of exogenous DNA damage bacteria encounters in the environment. It also suggests possible alternative functions of AlkC. Although no cellular evidence has been found yet, it is likely that AlkC removes 3mC and 1mA, the direct reversal repair substrates, from dsDNA *in vivo*.

Structure-Function Analysis of AlkD

YTM and CC-1065 of duocarmycin family are potent natural antibiotics produced by *Streptomyces*. The remarkable cytotoxicity of this cyclopropylpyrroloindole family of drugs for tumor cells makes them great candidates for anticancer therapy development. The biosynthetic gene clusters of YTM and CC-1065 have been characterized (**Figure**

39), which expand our knowledge in how cells generate these toxic chemicals. Although the YTM and CC-1065 biosynthetic gene clusters are evolutionarily closely related, only 12 of 28 genes in CC-1065 cluster express proteins homologous to proteins from YTM cluster (Wu et al, 2017; Xu et al, 2012). YtkR2 from *Streptomyces* sp. TP-A0356, an ortholog of AlkD, is identified in this cluster. Its high activity for YTMA is considered to be part of a self-defense mechanism to safeguard genomics against endogenous DNA modification drugs (Xu et al, 2012). Likewise, C10R5 from *Streptomyces zelensis*, another ortholog of AlkD, is identified in the CC-1065 biosynthetic gene cluster (Wu et al, 2017). We have shown that AlkD confers resistance to YTM in *B. anthracis*. However, since YtkR2, C10R5, and AlkD in cells encounter different levels and types of damage induced by the corresponding secondary metabolites, it is not clear whether they have evolved specific catalytic mechanisms and activities to adapt to these antibiotics. Future work can focus on biochemical and genetic characterization of C10R5 and YtkR2, and comparison of YtkR2, C10R5 with AlkD.

G,

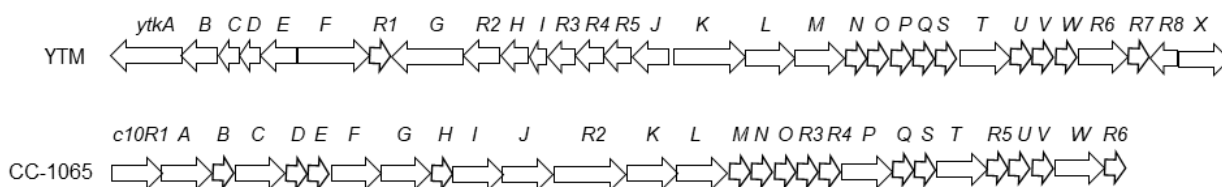


Figure 39. Biosynthetic gene clusters of YTM and CC-1065. *YtkR2*, *I*, *J*, *K*, *L*, *N*, *Q*, *S*, *T*, *U*, *V*, *R7* genes in YTM cluster are homologous to *c10R5*, *H*, *R1*, *C*, *N*, *K*, *G*, *B*, *P*, *Q*, *L*, *R6* genes in YTM cluster, respectively.

Pairwise sequence alignment showed that YtkR2 shares 24% identity and 44% similarity with AlkD, and C10R5 shares 20% identity and 34% similarity with AlkD. The

relatively high homology with AlkD indicates YtkR2 and C10R5 are composed of HEAT repeats. The catalytic residues Trp109, Asp113, and Trp187 of AlkD are conserved in these two enzymes (**Figure 40**). The non-base-flipping mechanism allows AlkD to remove YTMA. Given the robust activity of YtkR2 for YTMA and the homology of C10R5 and YtkR2 to AlkD, YtkR2 and C10R5 should also utilize non-base-flipping mechanisms. To validate our hypothesis and understand the catalytic mechanism of YtkR2 and C10R5, we can determine the crystal structures of these two glycosylases in complex with DNA containing the reaction intermediate analog 1aR-DNA, substrate YTMA-DNA, or CC-1065-DNA. Previous studies revealed that AlkD has high activity for 3mA, 7mG, POB-adducts, and YTMA, and that YtkR2 has high activity for 3mA and YTMA (Alseth et al, 2006; Mullins et al, 2015b; Rubinson et al, 2010; Xu et al, 2012). It remains unclear whether YtkR2 and C10R5 have similar substrate selectivity and catalytic efficiency. To understand how they distinguish distinct substrates, we can compare the activities of YtkR2 and C10R5 toward methylated bases (3mA and 7mG) and bulky adducts (POB-adducts, YTMA, and CC-1065). Furthermore, it is intriguing to know if AlkD is critical for excision of CC-1065 adducts *in vivo* and *in vitro*. Only 6 of 16 DNA contacting residues in AlkD are conserved in C10R5. Thus the binding and substrate preference of C10R5 may be different from AlkD. To test this, we can perform CC-1065 excision assay by AlkD, solve the structure of AlkC with CC-1065 containing DNA, as well as test sensitivity of *alkD*-depleted *B. anthracis* in response to CC-1065.

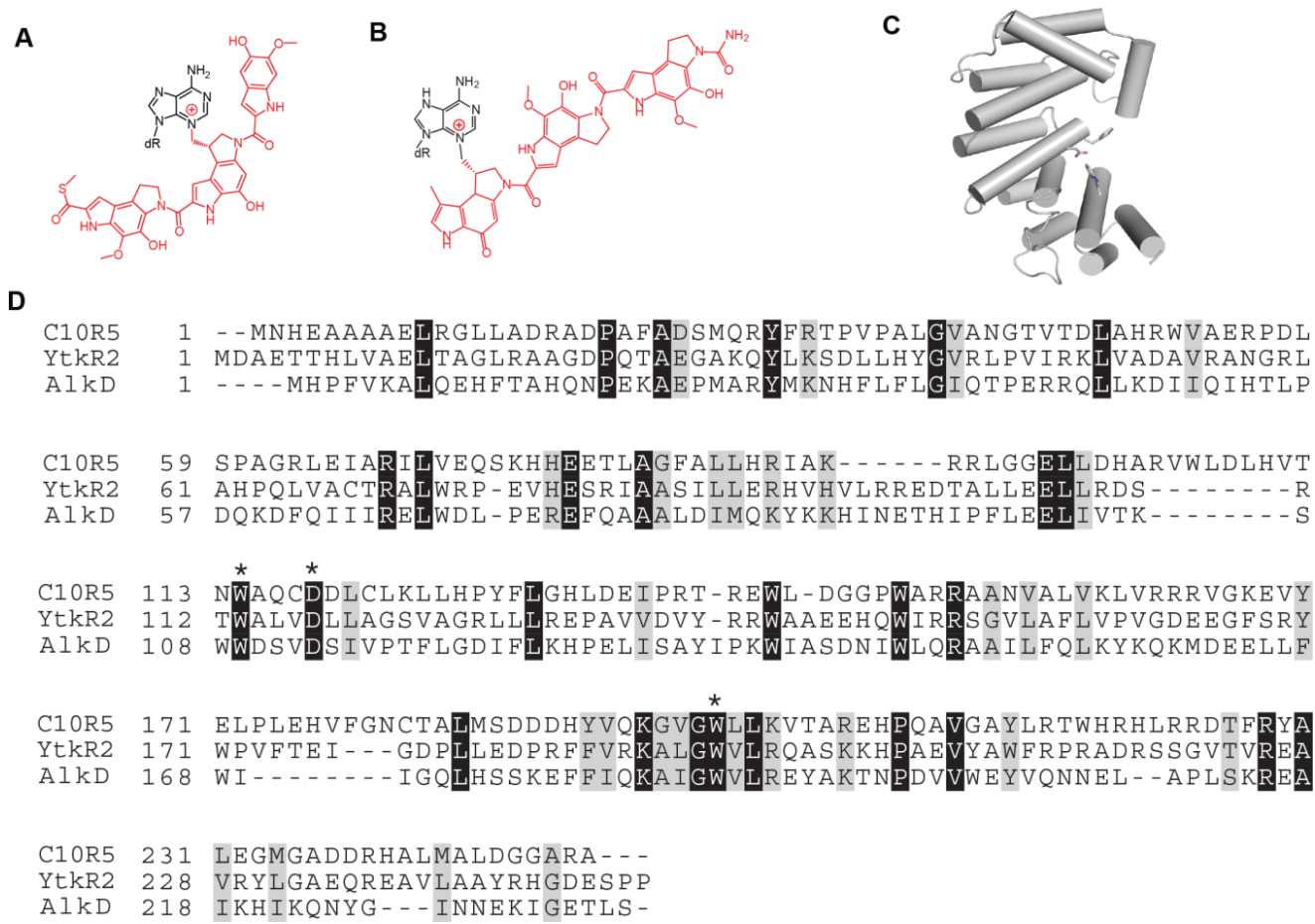


Figure 40. AlkD homologs C10R5 and YtkR2. **A**, Yatakemycin-adenine DNA. **B**, CC1065-adenine DNA. **C**, C10R5 model generated using SWISS-MODEL. **D**, Multiple sequence alignment of C10R5, YtkR2, and AlkD. Sequences were aligned using Clustal Omega and annotated using BoxShade. Shaded residues have 100% sequence identity (black) and similarity (grey). Catalytic residues are marked with an asterisk.

The presence of an endonuclease gene in the YTM biosynthetic gene cluster raises an interesting question whether this endonuclease has co-evolved with YtkR2 to efficiently remove YTM. The glycosylase remains bound to the AP site until displaced with the AP endonuclease. Presumably, this tight binding of the glycosylase to the AP site is to prevent the toxicity of AP lesion (Pope et al, 2002). In *E.coli*, exonuclease III (Exo III) and endonuclease IV (Endo IV) are AP endonucleases responsible for the backbone nick step following the N-glycosidic bond cleavage. The high affinity of AlkD

for abasic DNA prevents the handover of the AP site from DNA glycosylase to the subsequent endonucleases Exo III and Endo IV. This presumably causes the low YTM repair efficiency (Mullins et al, 2017a). The robust activity of YtkR2 for YTM and the high sensitivity of *B. anthracis* wild-type strain to YTM raise the question of how the AP site is efficiently dissociated from YtkR2 and processed by the rest of BER proteins in *Streptomyces* for YTM production. We hypothesize that YtkR2 and the endonuclease YtkR5 in the YTM biosynthetic gene cluster are adapted to YTM production and work differently from AlkD and endonucleases ExoIII and Endo IV to efficiently hand over the abasic DNA product to the subsequent step in BER. Consistent with this, YtkR5 shares relatively low sequence similarity and identity with Exo III (15% identity and 25% similarity) and Endo IV (21% identity and 34% similarity). 13 of 16 AlkD-DNA interaction residues are conserved in YtkR2 (**Figure 39**). To characterize if YtkR5 enhances the AP site release from YtkR2, we can examine the YTM excision rate under multiple turnover conditions in the presence of YtkR5. Together with the biochemical analysis of YtkR2 we proposed earlier, we will understand better how the BER pathway in antibiotic-producing bacteria develops self-resistance to avoidance of harmful effects of secondary metabolites. Similar characterization can be performed regarding the CC-1065 biosynthetic gene cluster.

Function and Activity Studies of Other HLR Proteins

Four HEAT-like repeat family proteins, AlkC, AlkD, AlkD2 and AlkF share similar α -helical domains, whereas the arrangement of first two or three helices in the N-terminal bundle may be important for differentiating the enzyme substrates. In addition

to characterizing AlkC and AlkD proteins, efforts have been made to understand the roles of AlkF and AlkD2. We found AlkD2 lacks B-helix which is present in the other three families and this may directly lead to loss of affinity for DNA (Mullins et al, 2015a). The *alkD2* gene in *S. mutants* is in the gene cluster of purine synthesis pathway, suggesting its potential role in purine biosynthesis. Therefore, we wanted to know if *alkD2* genes in other species locate in the same gene cluster. However, the majority of genes located adjacent to *alkD2* are annotated to produce hypothetical proteins, making this bioinformatics-function analysis very challenging. We were also interested to determine the binding affinity of SmAlkD2 to inosine monophosphate (IMP), a purine synthesis intermediate, but we were not able to determine the protein-ligand binding using isothermal titration calorimetry because of the high dilution heat released when IMP was added to SmAlkD2. To understand how AlkF binds to three-way and four-way DNA junctions, we incubated AlkF with DNA four-way junction and observed a secondary shift at high protein concentration in addition to the primary shift using electrophoretic mobility shift assay, which may be consistent with our assumption that AlkF dimerizes when acting on the substrate.

In summary, we studied the versatility of the novel HEAT-like repeat structure in substrate recognition and removal. The non-base-flipping mechanisms of alkylpurine DNA glycosylases AlkC and AlkD elucidate their novel selective repair of direct reversal and NER repair substrates. Characterization of the HLR proteins reveals the evolutionary dynamics that HLR architecture serves a flexible binding platform for diverse substrates, and multiple methylpurine DNA glycosylases tightly coordinate repair to prevent genomic damage. Additionally, the protection of AlkD from YTM

toxicity not only advances our understanding of self-defense mechanism provided by BER, but also sheds light on its potential application in increasing the yield of the antitumor agents via elimination deleterious lesions.

REFERENCES

- Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**(2): 213-221
- Adhikary S, Eichman BF (2011) Analysis of substrate specificity of *Schizosaccharomyces pombe* Mag1 alkylpurine DNA glycosylase. *EMBO Rep* **12**(12): 1286-1292
- Allan JM, Engelward BP, Dreslin AJ, Wyatt MD, Tomasz M, Samson LD (1998) Mammalian 3-methyladenine DNA glycosylase protects against the toxicity and clastogenicity of certain chemotherapeutic DNA cross-linking agents. *Cancer Res* **58**(17): 3965-3973
- Alseth I, Osman F, Korvald H, Tsaneva I, Whitby MC, Seeberg E, Bjoras M (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, Rognes T, Lindback T, Solberg I, Robertsen K, Kristiansen KI, Mainieri D, Lillehagen L, Kolsto AB, Bjoras M (2006) A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol Microbiol* **59**(5): 1602-1609
- Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P (2001) Comparison of ARM and HEAT protein repeats. *J Mol Biol* **309**(1): 1-18
- Backe PH, Simm R, Laerdahl JK, Dalhus B, Fagerlund A, Okstad OA, Rognes T, Alseth I, Kolsto AB, Bjoras M (2013) A new family of proteins related to the HEAT-like repeat DNA glycosylases with affinity for branched DNA structures. *J Struct Biol* **183**(1): 66-75
- Becker S, Groner B, Muller CW (1998) Three-dimensional structure of the Stat3 β homodimer bound to DNA. *Nature* **394**(6689): 145
- Behjati S, Gudem G, Wedge DC, Roberts ND, Tarpey PS, Cooke SL, Van Loo P, Alexandrov LB, Ramakrishna M, Davies H (2016) Mutational signatures of ionizing radiation in second malignancies. *Nature communications* **7**: 12605
- Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, Vitale L, Pelleri MC, Tassani S, Piva F (2013) An estimation of the number of cells in the human body. *Annals of human biology* **40**(6): 463-471

Birkeland NK, Anensen H, Knaevelsrud I, Kristoffersen W, Bjoras M, Robb FT, Klungland A, Bjelland S (2002) Methylpurine DNA glycosylase of the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *Biochemistry* **41**(42): 12697-12705

Bjelland S, Birkeland NK, Benneche T, Volden G, Seeberg E (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, Bjoras M, Seeberg E (1993) Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of *Escherichia coli*. *Nucleic Acids Res* **21**(9): 2045-2049

Black J, Menderes G, Bellone S, Schwab CL, Bonazzoli E, Ferrari F, Predolini F, De Haydu C, Cocco E, Buza N (2016) SYD985, a novel duocarmycin-based HER2-targeting antibody-drug conjugate, shows antitumor activity in uterine serous carcinoma with HER2/Neu expression. *Mol Cancer Ther* **15**(8): 1900-1909

Bodelón G, Palomino C, Fernández LÁ (2013) Immunoglobulin domains in *Escherichia coli* and other enterobacteria: from pathogenesis to applications in antibody technologies. *FEMS Microbiol Rev* **37**(2): 204-250

Boiteux S, Guillet M (2004) Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae*. *DNA repair* **3**(1): 1-12

Bork P, Holm L, Sander C (1994) The immunoglobulin fold: structural classification, sequence patterns and common core. *J Mol Biol* **242**(4): 309-320

Bowman BR, Lee S, Wang S, Verdine GL (2010) Structure of *Escherichia coli* AlkA in complex with undamaged DNA. *J Biol Chem* **285**(46): 35783-35791

Bravo J, Li Z, Speck NA, Warren AJ (2001) The leukemia-associated AML1 (Runx1)–CBF β complex functions as a DNA-induced molecular clamp. *Nat Struct Biol* **8**(4): 371-378

Brooks SC, Adhikary S, Rubinson EH, Eichman BF (2013) Recent advances in the structural mechanisms of DNA glycosylases. *Biochim Biophys Acta* **1834**(1): 247-271

Brown PJ, Bedard LL, Massey TE (2008) Repair of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-induced DNA pyridyloxobutylation by nucleotide excision repair. *Cancer Lett* **260**(1): 48-55

Bruner SD, Norman DP, Verdine GL (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**(6772): 859-866

Chakarov S, Petkova R, Russev GC, Zhelev N (2014) DNA damage and mutation. Types of DNA damage. *BioDiscovery* **11**

Chargaff E, Magasanik B, Vischer E, Green C, Doniger R, Elson D (1950) Nucleotide composition of pentose nucleic acids from yeast and mammalian tissues. *J Biol Chem* **186**(1): 51-67

Chatterjee N, Walker GC (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* **58**(5): 235-263

Chen J, Derfler B, Maskati A, Samson L (1989) Cloning a eukaryotic DNA glycosylase repair gene by the suppression of a DNA repair defect in *Escherichia coli*. *Proc Natl Acad Sci USA* **86**(20): 7961-7965

Chen J, Samson L (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, Glover JM, Hogan PG, Rao A, Harrison SC (1998a) Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **392**(6671): 42-48

Chen L, Glover JN, Hogan PG, Rao A, Harrison SC (1998b) Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **392**(6671): 42-48

Cho Y, Gorina S, Jeffrey PD, Pavletich NP (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* **265**(5170): 346-355

Chu AM, Fettingner JC, David SS (2011) Profiling base excision repair glycosylases with synthesized transition state analogs. *Bioorg Med Chem Lett* **21**(17): 4969-4972

Chu G (1997) Double strand break repair. *J Biol Chem* **272**(39): 24097-24100

Colvin M (2003) *Alkylating Agents*, 6th edn.: Hamilton (ON): BC Decker.

Cramer P, Larson CJ, Verdine GL, Muller CW (1997) Structure of the human NF- κ B p52 homodimer-DNA complex at 2.1 Å resolution. *EMBO J* **16**(23): 7078-7090

D'Andrea AD (2014) DNA repair pathways and human cancer. In *The Molecular Basis of Cancer* 4th edn, pp 47-66. Elsevier

Dahm R (2005) Friedrich Miescher and the discovery of DNA. *Dev Biol* **278**(2): 274-288

Dalhus B, Helle IH, Backe PH, Alseth I, Rognes T, Bjoras M, Laerdahl JK (2007) Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats. *Nucleic Acids Res* **35**(7): 2451-2459

Dalhus B, Laerdahl JK, Backe PH, Bjørås M (2009) DNA base repair-recognition and initiation of catalysis. *FEMS Microbiol Rev* **33**(6): 1044-1078

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**(suppl 2): W375-W383

Deans AJ, West SC (2011) DNA interstrand crosslink repair and cancer. *Nature Reviews Cancer* **11**(7): 467

Delaney JC, Essigmann JM (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in *alkB* *Escherichia coli*. *Proc Natl Acad Sci U S A* **101**(39): 14051-14056

Denver DR, Swenson SL, Lynch M (2003) An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. *Mol Biol Evol* **20**(10): 1603-1611

Doherty AJ, Serpell LC, Ponting CP (1996) The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res* **24**(13): 2488-2497

Drohat AC, Kwon K, Krosky DJ, Stivers JT (2002) 3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member. *Nat Struct Biol* **9**(9): 659-664

Drohat AC, Maiti A (2014) Mechanisms for enzymatic cleavage of the N-glycosidic bond in DNA. *Org Biomol Chem* **12**(42): 8367-8378

Ehrlich M, Zhang X-Y, Inamdar NM (1990) Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. *Mutat Res Genet Toxicol Environ Mutagen* **238**(3): 277-286

Eker A, Quayle C, Chaves I, Van der Horst G (2009) DNA repair in mammalian cells. *Cell Mol Life Sci* **66**(6): 968-980

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(12): 2126-2132

Engelward BP, Weeda G, Wyatt MD, Broekhof JL, de Wit J, Donker I, Allan JM, Gold B, Hoeijmakers JH, Samson LD (1997) Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase. *Proc Natl Acad Sci USA* **94**(24): 13087-13092

Evensen G, Seeberg E (1982) Adaptation to alkylation resistance involves the induction of a DNA glycosylase. *Nature* **296**(5859): 773

- Falnes P, Bjørås M, Aas PA, Sundheim O, Seeberg E (2004) Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res* **32**(11): 3456-3461
- Falnes PO, Johansen RF, Seeberg E (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **419**(6903): 178-182
- Fang Q, Kanugula S, Tubbs JL, Tainer JA, Pegg AE (2010) Repair of O⁴-alkylthymine by O⁶-alkylguanine-DNA alkyltransferases. *J Biol Chem* **285**(11): 8185-8195
- Fedeles BI, Singh V, Delaney JC, Li D, Essigmann JM (2015) The AlkB family of Fe (II)/ α -ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J Biol Chem* **290**(34): 20734-20742
- Feinberg AP, Ohlsson R, Henikoff S (2006) The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**(1): 21
- Franklin RE, Gosling RG (1953) Molecular configuration in sodium thymonucleate. *Nature* **171**(4356): 740
- Friedberg EC (2008) A brief history of the DNA repair field. *Cell Res* **18**(1): 3
- Friedberg EC, Aguilera A, Gellert M, Hanawalt PC, Hays JB, Lehmann AR, Lindahl T, Lowndes N, Sarasin A, Wood RD (2006a) DNA repair: from molecular mechanism to human disease. *DNA Repair* **5**(8): 986-996
- Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T (2006b) *DNA Repair and Mutagenesis*, 2nd edn. Washington, DC: ASM Press.
- Fromme JC, Verdine GL (2004) Base excision repair. *Adv Protein Chem* **69**: 1-41
- Fuerst JA (2010) Beyond prokaryotes and eukaryotes: planctomycetes and cell organization. *Nature Education* **3**(9, Article No. 44): Online
- Glassner BJ, Rasmussen LJ, Najarian MT, Posnick LM, Samson LD (1998) Generation of a strong mutator phenotype in yeast by imbalanced base excision repair. *Proc Natl Acad Sci* **95**(17): 9997-10002
- Griffiths AJF MJ, Suzuki DT, et al. (2000) Spontaneous mutations. *An Introduction to Genetic Analysis* **7th edition**
- Grove TZ, Cortajarena AL, Regan L (2008) Ligand binding by repeat proteins: natural and designed. *Curr Opin Struct Biol* **18**(4): 507-515
- Hakem R (2008) DNA-damage repair; the good, the bad, and the ugly. *The EMBO journal* **27**(4): 589-605

- Halaby DM, Mornon JP (1998) The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. *J Mol Evol* **46**(4): 389-400
- Halaby DM, Poupon A, Mornon J (1999) The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng* **12**(7): 563-571
- Hendershot JM, O'Brien PJ (2014) Critical role of DNA intercalation in enzyme-catalyzed nucleotide flipping. *Nucleic Acids Res* **42**(20): 12681-12690
- Hendershot JM, O'Brien PJ (2017) Search for DNA damage by human alkyladenine DNA glycosylase involves early intercalation by an aromatic residue. *J Biol Chem* **292**(39): 16070-16080
- Holliday R, Grigg G (1993) DNA methylation and mutation. *Mutat Res Fund Mol Mech Mut* **285**(1): 61-67
- Hollis T, Ichikawa Y, Ellenberger T (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
- Holm L, Laakso LM (2016) Dali server update. *Nucleic Acids Res* **44**(W1): W351-W355
- Hu J, Adar S, Selby CP, Lieb JD, Sancar A (2015) Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes Dev* **29**(9): 948-960
- Hu J, Selby CP, Adar S, Adebali O, Sancar A (2017) Molecular mechanisms and genomic maps of DNA excision repair in *Escherichia coli* and humans. *J Biol Chem* **292**(38): 15588-15597
- Igarashi Y, Futamata K, Fujita T, Sekine A, Senda H, Naoki H, Furumai T (2003) Yatakemycin, a novel antifungal antibiotic produced by *Streptomyces* sp. TP-A0356. *J Antibiot* **56**(2): 107-113
- Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* **461**(7267): 1071
- Jasin M, Rothstein R (2013) Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* **5**(11): a012740
- Jin S-G, Choi J-H, Ahn B, O'Connor TR, Mar W, Lee C-S (2001) Excision Repair of Adozeiesin-N3 Adenine Adduct by 3-Methyladenine-DNA Glycosylases and UvrABC Nuclease. *Mol Cells (Springer Science & Business Media BV)* **11**(1)

Kerksick CM, Zuhl M. (2015) Mechanisms of oxidative damage and their impact on contracting muscle. In M L (ed.), *Antioxidants in sport nutrition*. Boca Raton (FL): CRC Press/Taylor & Francis, Vol. Chapter 1, pp. 1-16.

Kiakos K, Sato A, Asao T, McHugh PJ, Lee M, Hartley JA (2007) DNA sequence-selective adenine alkylation, mechanism of adduct repair, and in vivo antitumor activity of the novel achiral seco-amino-cyclopropylbenz [e] indolone analogue of duocarmycin AS-I-145. *Mol Cancer Ther* **6**(10): 2708-2718

Kilambi KP, Gray JJ (2012) Rapid calculation of protein pK_a values using Rosetta. *Biophys J* **103**(3): 587-595

Kim Y-J, M Wilson III D (2012) Overview of base excision repair biochemistry. *Curr Mol Pharmacol* **5**(1): 3-13

Kisker C, Kuper J, Van Houten B (2013) Prokaryotic nucleotide excision repair. *Cold Spring Harb Perspect Biol* **5**(3): a012591

Koç A, Wheeler LJ, Mathews CK, Merrill GF (2004) Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. *J Biol Chem* **279**(1): 223-230

Kohn KW, Spears CL, Doty P (1966) Inter-strand crosslinking of DNA by nitrogen mustard. *J Mol Biol* **19**(2): 266-288

Koivisto P, Robins P, Lindahl T, Sedgwick B (2004) Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J Biol Chem* **279**(39): 40470-40474

Korsh J, Shen A, Aliano K, Davenport T (2015) Polycyclic aromatic hydrocarbons and breast cancer: a review of the literature. *Breast Care* **10**(5): 316-318

Kossel A (1879) Ueber das Nuclein der Hefe. *Zeitschrift für physiologische Chemie* **3**(4): 284-291

Kovtun YV, Goldmacher VS (2007) Cell killing by antibody-drug conjugates. *Cancer Lett* **255**(2): 232-240

Krokan HE, Bjørås M (2013) Base excision repair. *Cold Spring Harbor Perspect Biol* **5**(4): a012583

Kryston TB, Georgiev AB, Pissis P, Georgakilas AG (2013) Role of oxidative stress and DNA damage in human carcinogenesis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **711**(1): 193-201

- Kschonsak M, Merkel F, Bisht S, Metz J, Rybin V, Hassler M, Haering CH (2017) Structural basis for a safety-belt mechanism that anchors condensin to chromosomes. *Cell* **171**(3): 588-600. e524
- Labahn J, Scharer OD, Long A, Ezaz-Nikpay K, Verdine GL, Ellenberger TE (1996) Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **86**(2): 321-329
- Lahr RM, Mack SM, Héroux A, Blagden SP, Bousquet-Antonelli C, Deragon J-M, Berman AJ (2015) The La-related protein 1-specific domain repurposes HEAT-like repeats to directly bind a 5' TOP sequence. *Nucleic Acids Res* **43**(16): 8077-8088
- Lamoureux JS, Stuart D, Tsang R, Wu C, Glover JM (2002) Structure of the sporulation-specific transcription factor Ndt80 bound to DNA. *EMBO J* **21**(21): 5721-5732
- Landini P, Volkert MR (2000) Regulatory responses of the adaptive response to alkylation damage: a simple regulon with complex regulatory features. *J Bacteriol* **182**(23): 6543-6549
- Lau AY, Schärer OD, Samson L, Verdine GL, Ellenberger T (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 AY, Wyatt MD, Glassner BJ, Samson LD, Ellenberger T (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc Natl Acad Sci USA* **97**(25): 13573-13578
- Leiros I, Nabong MP, Grosvik K, Ringvoll J, Haugland GT, Uldal L, Reite K, Olsbu IK, Knaevelsrud I, Moe E, Andersen OA, Birkeland NK, Ruoff P, Klungland A, Bjelland S (2007) Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA. *EMBO J* **26**(8): 2206-2217
- Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79**: 181-211
- Lindahl T (1976) New class of enzymes acting on damaged DNA. *Nature* **259**(5538): 64
- Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**(6422): 709
- Lindahl T, Sedgwick B, Sekiguchi M, Nakabeppu Y (1988) Regulation and expression of the adaptive response to alkylating agents. *Annu Rev Biochem* **57**: 133-157
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2000) DNA Damage and Repair and Their Role in Carcinogenesis. In *Molecular cell biology* 4th edition edn, Section 12.4. New York: W. H. Freeman

Lopez-Martinez D, Liang C-C, Cohn MA (2016) Cellular response to DNA interstrand crosslinks: the Fanconi anemia pathway. *Cell Mol Life Sci* **73**(16): 3097-3114

Lyskov S, Chou F-C, Conchúir SÓ, Der BS, Drew K, Kuroda D, Xu J, Weitzner BD, Renfrew PD, Sripakdeevong P (2013) Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PLoS One* **8**(5): e63906

McAndrew RP, Park JI, Heins RA, Reindl W, Friedland GD, D'haeseleer P, Northen T, Sale KL, Simmons BA, Adams PD (2013) From soil to structure, a novel dimeric β -glucosidase belonging to glycoside hydrolase family 3 isolated from compost using metagenomic analysis. *J Biol Chem* **288**(21): 14985-14992

McCarthy TV, Karran P, Lindahl T (1984) Inducible repair of O-alkylated DNA pyrimidines in *Escherichia coli*. *EMBO J* **3**(3): 545-550

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**(4): 658-674

McCulloch SD, Kunkel TA (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell research* **18**(1): 148

McQueen CA, Rosado RR, Williams GM (1989) Effect of nalidixic acid on DNA repair in rat hepatocytes. *Cell Biol and Toxicol* **5**(2): 201-206

Memisoglu A, Samson L (2000) Base excision repair in yeast and mammals. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **451**(1): 39-51

Metz AH, Hollis T, Eichman BF (2007) DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG). *EMBO J* **26**(9): 2411-2420

Meyer PA, Socias S, Key J, Ransey E, Tjon EC, Buschiazzi A, Lei M, Botka C, Withrow J, Neau D (2016) Data publication with the structural biology data grid supports live analysis. *Nat Commun* **7**: 10882

Miescher-Rüsch F (1871) *Ueber die chemische Zusammensetzung der Eiterzellen*.

Mike LA, Choby JE, Brinkman PR, Olive LQ, Dutter BF, Ivan SJ, Gibbs CM, Sulikowski GA, Stauff DL, Skaar EP (2014) Two-component system cross-regulation integrates *Bacillus anthracis* response to heme and cell envelope stress. *PLoS Pathog* **10**(3): e1004044

Miller EC, Miller JA (1981) Mechanisms of chemical carcinogenesis. *Cancer* **47**(S5): 1055-1064

Mishina Y, Duguid EM, He C (2006) Direct reversal of DNA alkylation damage. *Chem Rev* **106**(2): 215-232

Mishina Y, He C (2006) Oxidative dealkylation DNA repair mediated by the mononuclear non-heme iron AlkB proteins. *J Inorg Biochem* **100**(4): 670-678

Mitra S (2007) MGMT: a personal perspective. *DNA Repair* **6**(8): 1064-1070

Moe E, Hall DR, Leiros I, Monsen VT, Timmins J, McSweeney S (2012) Structure-function studies of an unusual 3-methyladenine DNA glycosylase II (AlkA) from *Deinococcus radiodurans*. *Acta Crystallogr D* **68**(6): 703-712

Morin A, Eisenbraun B, Key J, Sanschagrín PC, Timony MA, Ottaviano M, Sliz P (2013) Collaboration gets the most out of software. *Elife* **2**: e01456

Mouret S, Baudouin C, Charveron M, Favier A, Cadet J, Douki T (2006) Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proc Natl Acad Sci U S A* **103**(37): 13765-13770

Mullins EA, Rubinson EH, Eichman BF (2014) The substrate binding interface of alkylpurine DNA glycosylase AlkD. *DNA Repair* **13**: 50-54

Mullins EA, Rubinson EH, Pereira KN, Calcutt MW, Christov PP, Eichman BF (2013) An HPLC-tandem mass spectrometry method for simultaneous detection of alkylated base excision repair products. *Methods* **64**: 59-66

Mullins EA, Shi R, Eichman BF (2017a) Toxicity and repair of DNA adducts produced by the natural product yatakemycin. *Nat Chem Biol* doi: **10.1038/nchembio.2439**

Mullins EA, Shi R, Kotsch LA, Eichman BF (2015a) A new family of HEAT-like repeat proteins lacking a critical substrate recognition motif present in related DNA glycosylases. *PLoS One* **10**(5): e0127733

Mullins EA, Shi R, Parsons ZD, Yuen PK, David SS, Igarashi Y, Eichman BF (2015b) The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. *Nature* **527**(7577): 254-258

Mullins EA, Warren GM, Bradley NP, Eichman BF (2017b) Structure of a DNA glycosylase that unhooks interstrand cross-links. *Proc Natl Acad Sci U S A* **114**(17): 4400-4405

Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**(1): 297-304

Nagata T, Gupta V, Sorce D, Kim W-Y, Sali A, Chait BT, Shigesada K, Ito Y, Werner MH (1999) Immunoglobulin motif DNA recognition and heterodimerization of the PEBP2/CBF Runt domain. *Nat Struct Mol Biol* **6**(7): 615-619

Nakabeppu Y, Miyata T, Kondo H, Iwanaga S, Sekiguchi M (1984) Structure and expression of the alkA gene of *Escherichia coli* involved in adaptive response to alkylating agents. *J Biol Chem* **259**(22): 13730-13736

Nay SL, O'Connor. TR (2013) Direct repair in mammalian cells. In *New Research Directions in DNA Repair*. InTech

Neuwald AF, Hirano T (2000) HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome Res* **10**(10): 1445-1452

O'Brien PJ, Ellenberger T (2004a) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J Biol Chem* **279**(11): 9750-9757

O'Brien PJ, Ellenberger T (2004b) The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J Biol Chem* **279**(26): 26876-26884

O'Connor TR, Laval J (1991) Human cDNA expressing a functional DNA glycosylase excising 3-methyladenine and 7-methylguanine. *Biochem Biophys Res Commun* **176**(3): 1170-1177

Okada C, Yamashita E, Lee SJ, Shibata S, Katahira J, Nakagawa A, Yoneda Y, Tsukihara T (2009) A high-resolution structure of the pre-microRNA nuclear export machinery. *Science* **326**(5957): 1275-1279

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307-326

Parsons ZD, Bland JM, Mullins EA, Eichman BF (2016) A catalytic role for C-H/ π interactions in base excision repair by *Bacillus cereus* DNA glycosylase AlkD. *J Am Chem Soc* **138**(36): 11485-11488

Perry J, Kleckner N (2003) The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell* **112**(2): 151-155

Petersen AB, Gniadecki R, Vicanova J, Thorn T, Wulf HC (2000) Hydrogen peroxide is responsible for UVA-induced DNA damage measured by alkaline comet assay in HaCaT keratinocytes. *J Photochem Photobiol* **59**(1-3): 123-131

Pope MA, Porello SL, David SS (2002) *Escherichia coli* apurinic-apyrimidinic endonucleases enhance the turnover of the adenine glycosylase MutY with G: A substrates. *J Biol Chem* **277**(25): 22605-22615

Posnick LM, Samson LD (1999) Imbalanced Base Excision Repair Increases Spontaneous Mutation and Alkylation Sensitivity in *Escherichia coli*. *J Bacteriol* **181**(21): 6763-6771

Pray L (2008) DNA replication and causes of mutation. *Nature Education* **1**(1): 214

Price MN, Dehal PS, Arkin AP (2010) FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One* **5**(3): e9490

Reardon JT, Sancar A (2005) Nucleotide excision repair. *Prog Nucleic Acid Res Mol Biol* **79**: 183-235

Reisz JA, Bansal N, Qian J, Zhao W, Furdai CM (2014) Effects of ionizing radiation on biological molecules-mechanisms of damage and emerging methods of detection. *Antioxid Redox Signal* **21**(2): 260-292

Riazuddin S, Lindahl T (1978) Properties of 3-methyladenine-DNA glycosylase from *Escherichia coli*. *Biochemistry* **17**(11): 2110-2118

Roberts RJ, Cheng X (1998) Base flipping. *Annu Rev Biochem* **67**: 181-198

Robertson AB, Klungland A, Rognes T, Leiros I (2009) Base excision repair: the long and short of it. *Cell Mol Life Sci* **66**(6): 981-993

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

Roos WP, Thomas AD, Kaina B (2016) DNA damage and the balance between survival and death in cancer biology. *Nature Reviews Cancer* **16**(1): 20

Rubinson EH, Adhikary S, Eichman BF (2009) Structural studies of alkylpurine DNA glycosylases. In *ACS Symposium Series: Structural Biology of DNA Damage and Repair*, Stone MP (ed), Vol. to be published. Washington, D.C.: American Chemical Society

Rubinson EH, Eichman BF (2012) Nucleic acid recognition by tandem helical repeats. *Curr Opin Struct Biol* **22**: 101-109

Rubinson EH, Gowda AS, Spratt TE, Gold B, Eichman BF (2010) An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* **468**(7322): 406-411

Rubinson EH, Metz AH, O'Quin J, Eichman BF (2008) A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD. *J Mol Biol* **381**(1): 13-23

Rudolph MJ, Gergen JP (2001) DNA-binding by Ig-fold proteins. *Nat Struct Biol* **8**(5): 384-386

Saparbaev M, Kleibl K, Laval J (1995) *Escherichia coli*, *Saccharomyces cerevisiae*, rat and human 3-methyladenine DNA glycosylases repair 1,N⁶-ethenoadenine when present in DNA. *Nucleic Acids Res* **23**(18): 3750-3755

Schärer OD (2013) Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol* **5**(10): a012609

Scharer OD, Jiricny J (2001) Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *Bioessays* **23**(3): 270-281

Schermerhorn KM, Delaney S (2014) A chemical and kinetic perspective on base excision repair of DNA. *Acc Chem Res* **47**(4): 1238-1246

Schramm VL (2011) Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu Rev Biochem* **80**: 703-732

Schrödinger L (2016) The PyMOL molecular graphics system, Version 1.7.4 Schrödinger, LLC.

Sedgwick B (2004) Repairing DNA-methylation damage. *Nat Rev Mol Cell Biol* **5**(2): 148-157

Sedgwick B, Lindahl T (2002) Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene* **21**(58): 8886

Selby CP, Sancar A (1988) ABC excinuclease incises both 5' and 3' to the CC-1065-DNA adduct and its incision activity is stimulated by DNA helicase II and DNA polymerase I. *Biochemistry* **27**(19): 7184-7188

Shi R, Mullins EA, Shen X-X, Lay KT, Yuen PK, David SS, Rokas A, Eichman BF (2018) Selective base excision repair of DNA damage by the none-base-flipping DNA glycosylase AlkC. *EMBO J* **37**(1): 63-74

Shuck SC, Short EA, Turchi JJ (2008) Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res* **18**(1): 64

Sibanda BL, Chirgadze DY, Blundell TL (2010) Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature* **463**(7277): 118-121

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**(1): 539

- Stauff DL, Skaar EP (2009) *Bacillus anthracis* HssRS signalling to HrtAB regulates haem resistance during infection. *Mol Microbiol* **72**(3): 763-778
- Tahirov TH, Inoue-Bungo T, Morii H, Fujikawa A, Sasaki M, Kimura K, Shiina M, Sato K, Kumasaka T, Yamamoto M (2001) Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBF β . *Cell* **104**(5): 755-767
- Tamm C, Hodes M, Chargaff E (1952) The formation of apurinic acid from the desoxyribonucleic acid of calf thymus. *J Biol Chem* **195**(1): 49-63
- Thomas L, Yang CH, Goldthwait DA (1982) Two DNA glycosylases in *Escherichia coli* which release primarily 3-methyladenine. *Biochemistry* **21**(6): 1162-1169
- Tommasi S, Denissenko MF, Pfeifer GP (1997) Sunlight induces pyrimidine dimers preferentially at 5-methylcytosine bases. *Cancer research* **57**(21): 4727-4730
- Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**(6903): 174-178
- Trivedi MK, Patil S, Shettigar H, Gangwar M, Jana S (2015) Antimicrobial sensitivity pattern of *Pseudomonas fluorescens* after biofield treatment. *Infect Dis Ther* **3**(3)
- Troll CJ, Adhikary S, Cueff M, Mitra I, Eichman BF, Camps M (2014) Interplay between base excision repair activity and toxicity of 3-methyladenine DNA glycosylases in an *E. coli* complementation system. *Mutat Res Fund Mol Mech Mut* **763**: 64-73
- Tubbs A, Nussenzweig A (2017) Endogenous DNA damage as a source of genomic instability in cancer. *Cell* **168**(4): 644-656
- van den Born E, Bekkelund A, Moen MN, Omelchenko MV, Klungland A, Falnes PØ (2009) Bioinformatics and functional analysis define four distinct groups of AlkB DNA-dioxygenases in bacteria. *Nucleic Acids Res* **37**(21): 7124-7136
- Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* **364**: 215-230
- Vos SM, Tretter EM, Schmidt BH, Berger JM (2011) All tangled up: how cells direct, manage and exploit topoisomerase function. *Nat Rev Mol Cell Biol* **12**(12): 827
- Wang S, Liu K, Xiao L, Yang L, Li H, Zhang F, Lei L, Li S, Feng X, Li A (2016) Characterization of a novel DNA glycosylase from *S. sahachiroi* involved in the reduction and repair of azinomycin B induced DNA damage. *Nucleic Acids Res* **44**(1): 187-197

Watson JD, Crick FH (1953) Molecular structure of nucleic acids. *Nature* **171**(4356): 737-738

Weber S (2005) Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Biochim Biophys Acta, Bioenerg* **1707**(1): 1-23

Welkos S (1991) Plasmid-associated virulence factors of non-toxigenic (pX01-) *Bacillus anthracis*. *Microb Pathog* **10**(3): 183-198

Welkos S, Friedlander A (1988) Pathogenesis and genetic control of resistance to the Sterne strain of *Bacillus anthracis*. *Microb Pathog* **4**(1): 53-69

Werner RM, Stivers JT (2000) Kinetic isotope effect studies of the reaction catalyzed by uracil DNA glycosylase: evidence for an oxocarbenium ion-uracil anion intermediate. *Biochemistry* **39**(46): 14054-14064

Wilkinson M, Potter P, Cawkwell L, Georgiadis P, Patel D, Swann P, Margison GP (1989) Purification of the *E. coli* ogt gene product to homogeneity and its rate of action on O⁶-methylguanine, O⁶-ethylguanine and O⁴-methylthymine in dodecadeoxyribnucleotides. *Nucleic Acids Res* **17**(21): 8475-8483

Wood RD (1996) DNA repair in eukaryotes. *Annu Rev Biochem* **65**(1): 135-167

Wu S, Jian X-H, Yuan H, Jin W-B, Yin Y, Wang L-Y, Zhao J, Tang G-L (2017) Unified biosynthetic origin of the benzodipyrrole subunits in CC-1065. *ACS Chem Biol* **12**(6): 1603-1610

Wyatt MD, Allan JM, Lau AY, Ellenberger TE, Samson LD (1999) 3-methyladenine DNA glycosylases: structure, function, and biological importance. *Bioessays* **21**(8): 668-676

Wyatt MD, Pittman DL (2006) Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem Res Toxicol* **19**(12): 1580-1594

Xu H, Huang W, He QL, Zhao ZX, Zhang F, Wang RX, Kang JW, Tang GL (2012) Self-resistance to an antitumor antibiotic: a DNA glycosylase triggers the base-excision repair system in yatakemycin biosynthesis. *Angew Chem* **51**(42): 10532-10536

Xue J, Yang S, Seng S (2014) Mechanisms of cancer induction by tobacco-specific NNK and NNN. *Cancers* **6**(2): 1138-1156

Yagura T, Makita K, Yamamoto H, Menck CF, Schuch AP (2011) Biological sensors for solar ultraviolet radiation. *Sensors* **11**(4): 4277-4294

Yamagata Y, Kato M, Odawara K, Tokuno Y, Nakashima Y, Matsushima N, Yasumura K, Tomita K-i, Ihara K, Fujii Y (1996) Three-dimensional structure of a DNA repair

enzyme, 3-methyladenine DNA glycosylase II, from *Escherichia coli*. *Cell* **86**(2): 311-319

Yi C, He C (2013) DNA repair by reversal of DNA damage. *Cold Spring Harb Perspect Biol* **5**(1): a012575

Yoon J-H, Bhatia G, Prakash S, Prakash L (2010) Error-free replicative bypass of thymine glycol by the combined action of DNA polymerases κ and ζ in human cells. *Proceedings of the National Academy of Sciences* **107**(32): 14116-14121

Yoshimura SH, Hirano T (2016) HEAT repeats-versatile arrays of amphiphilic helices working in crowded environments? *J Cell Sci* **129**(21): 3963-3970

You Y-H, Lee D-H, Yoon J-H, Nakajima S, Yasui A, Pfeifer GP (2001) Cyclobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. *J Biol Chem* **276**(48): 44688-44694

Yu S-L, Lee S-K (2017) Ultraviolet radiation: DNA damage, repair, and human disorders. *Mol Cell Toxicol* **13**(1): 21-28

Zheng G, Lu X-J, Olson WK (2009) Web 3DNA-a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res* **37**(suppl_2): W240-W246

Zheng H, Chordia MD, Cooper DR, Chruszcz M, Müller P, Sheldrick GM, Minor W (2014) Validating metal binding sites in macromolecule structures using the CheckMyMetal web server. *Nat Protoc* **9**(1): 156