STRUCTURAL AND BIOCHEMICAL STUDIES OF ALKYLPURINE DNA GLYCOSYLASE ALKD

By

Emily Holtzman Rubinson

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical and Physical Biology

May, 2011

Nashville, Tennessee

Approved:

Professor Fredrick P. Guengerich

Professor Lawrence J. Marnett

Professor Richard Armstrong

Professor D. Borden Lacy

Professor Brandt F. Eichman

For Mom and Dad,
thank you for all your love and support

ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Brandt Eichman, not only for his unending guidance, support, and encouragement but also for taking the time to teach me how to be a passionate scientist both at the bench and away from it. I am also grateful for the members of the Eichman laboratory especially Audrey Metz, who taught me everything I ever wanted to know about the "art" of crystallography and who initiated me into the world of a formerly unknown glycosylase, AlkD. Also, Briana Greer, a friend and confidant without whom I would have never successfully cloned anything. The past and present members of the lab were not only trusted colleagues but also friends and I thank them all for their help and camaraderie over the years. I also want to thank my committee members, Drs. Fred Guengerich, Larry Marnett, Richard Armstrong, and Borden Lacy for providing advice and insight into my research progress as well as my career development and I want to acknowledge the Toxicology training grant for funding my research for the past four years. I would also like to acknowledge the faculty and staff with Chemical and Physical Biology program including Lindsay Meyers for all her help during my time at Vanderbilt.

I would also like to thank my family, intermediate and beyond, especially my parents and my brothers Paul, Max, and Claude, for their love and support. I know with them in my corner I can accomplish anything. And finally, I want to thank my friends for keeping me sane throughout this process. I will carry great memories of my time at Vanderbilt and in Nashville with me forever.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1mA	*N*1-Methyladenine
3d3mA	3-Deaza-3-methyladenine
3mA	*N*3-Methyladenine
4-HNE	4-Hydroxynonenal
7mA	*N*7-Methyladenine
7mG	*N*7-Methylguanine
8-oxoG	7,8-Dihydro-8-oxoguanine
Å	Angstrom
A (Ade)	Adenine
AAG	Human alkyladenine DNA glycosylase
AGT	Alkylguanine alkyltransferase
AlkA	*E. coli* 3-methyladenine DNA glycosylase II
AP	Apurinic/apyrimidinic
AP	Aminopurine
ARM	Armadillo
ATP	Adenosine triphosphate
Aza	1-Azaribose
BER	Base excision repair
BME	2-Mercaptoethanol
C	Celsius

C (Cyt)	Cytosine
CD	Circular dichroism
CPD	Cyclobutane pyrimidine dimers
CTP	Cytidine triphosphate
d	Deuterated
DNA	Deoxyribonucleic Acid
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminotetraacetic acid
EM	Electron microscopy
EMS	Ethyl methanesulfonate
ESI	Electrospray
εA	Ethenoadenine
F	Fluorine
FAD	Flavin adenine dinucleotide
FaPy	Formamidopyrimidine
fmol	Femtomole
g	Gram
G (Gua)	Guanine
GTP	Guanosine triphosphate
H_2O	Water
H2TH	Helix-two turn-helix

HCl	Hydrochloric acid
HEAT	Huntington/Elongation/A subunit/Target of rapamycin
HhH	Helix-hairpin-helix
HPLC	High performance liquid chromatography
hr	Hour
HX	Hypoxanthine
I	Intensity
IPTG	Isopropyl thio-β-D-galactopyranoside
kcal	kilocalorie
K_d	Dissociation constant
kDa	Kilodalton
k_{st}	Single-turnover rate
L	Liter
LB	Luria broth
MAG	*S. cerevisiae* methyladenine DNA glycosylase
MagIII	*H. pylori* methyladenine DNA glycosylase
MALDI	Matrix assisted laser desorption ionization
MDA	Malondialdehyde
MeOH	Methanol
mg	Milligram
min	Minute
mL	Milliliter

mM	Millimolar
MMR	Mismatch repair
MMS	Methyl methanesulfonate
MNNG	*N*-Methyl-*N′*-nitro-*N*-nitrosoguanidine
MNU	Methylnitrosourea
mol	mole
MpgII	*N*-Methylpurine DNA glycosylase II
MpgII	*T.maritima* methylpurine DNA glycosylase II
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MTERF	Mitochondrial transcription termination factor
MutY	A/G specific adenine glycosylase
n.d.	Not determined
NaCl	Sodium chloride
NER	Nucleotide excision repair
Ni-NTA	Nickel-nitrilotriacetic acid
NNK	4-(methylnitrosamineo)-1-(3-pyridyl)-butanone
NNN	*N′*-Nitrosonornicotine
Nth	Endonuclease III
O^6mG	O^6-Methylguanine
OggI	8-oxoG DNA glycosylase I
OggII	8-oxoG DNA glycosylase II

PCR	Polymerase chain reaction
PD	Pyrimidine dimer
PDB	Protein Data Bank
PEG	Polyethylene glycol
POB	pyridyloxobutyl
Pt	Platinum
Pyr	1-pyrrolidine
rmsd	Root-mean-square deviation
rpm	Rotations per minute
s.d.	Standard deviation
SAD	Single-wavelength anomalous dispersion
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Second
ssDNA	Single-stranded DNA
T (Thy)	Thymine
TAG	*E. coli* 3-methyladenine DNA glycosylase I
TDG	Thymine DNA glycosylase
THF	Tetrahydrofuran
TOF	Time-of-flight
TRIS	Tris-(hydroxymethyl-)aminomethane
TTP	Thymidine triphosphate
UDG	Uracil DNA glycosylase

µg	Microgram
µM	Micromolar
UV	Ultraviolet
wt	Wild-type
XP	Xeroderma pigmentosum

CHAPTER I

INTRODUCTION*

In 1944 the scientists Avery, MacLeod, and McCarty published their discovery that deoxyribonucleic acid (DNA) was the "fundamental unit" of genetic information (Avery et al, 1944). This conclusion was drawn from experiments on the pathogenicity of pneumococci and was supported by *E. coli* experiments performed in 1952 by Hershey and Chase (Hershey & Chase, 1952). These results were met with cautious acceptance by the scientific community until the 1953 landmark structural model of DNA as a double helix was derived by Watson and Crick based on x-ray diffraction data from DNA fibers collected by Franklin and Wilkins. This model, which was eventually proved to be correct for B-form DNA, described two helical chains running in opposite directions held together by hydrogen bonds between pairs of bases (adenine with thymine; guanine with cytosine) with the phosphate and deoxyribose molecules on the outside of the helix (Watson & Crick, 1953a; Watson & Crick, 1953b). The structure of the double-helix suggested not only a mechanism for the replication of our genetic material but also the importance of maintaining the fidelity of DNA sequence and structure. DNA is a highly reactive molecule and thus susceptible to chemical modification by various exogenous and endogenous damaging agents. Understanding the effects of DNA damage and the various cellular responses in place to restore the damaged duplex is a crucial task that

* Part of the work presented in this chapter was published in Rubinson, E.H., Adhikary, S., and Eichman, B.F. (2010) Structural studies of alkylpurine DNA glycosylases. In Stone, M.P. (ed), *ACS Symposium Series: Structural Biology of DNA Damage and Repair*. American Chemical Society, Washington, D.C. Vol 1041, pp 29-45.

continues today and its relation to human health has been extensively studied. The persistence of DNA damage as a consequence of defective or deficient DNA repair pathways can have devastating cellular and physiological consequences, including cell-cycle arrest and apoptosis, the emergence of disease states such as cancer, and the phenomenon of aging.

Overview of DNA Damage

Endogenous DNA Damage

The aqueous environment of the nucleus and the presence of reactive oxygen lead to the spontaneous decomposition of DNA (reviewed in Lindahl, 1993). Endogenous processes like hydrolysis, oxidation, and nonenzymatic methylation of DNA contribute to the instability of the molecule with the *N*-glycosylic bond between the deoxyribose sugar and the base being particularly labile (Figure 1) (Lindahl, 1993). Hydrolytic cleavage of this bond leads to depurination or depyrimidination, with guanine and adenine being released more rapidly and frequently than thymine and cytosine. Regardless of which base is cleaved, the resulting abasic site (Figure 3a) eventually undergoes β-elimination and subsequent strand breaks of the phosphate backbone (reviewed in Lindahl, 1993). By initiating strand breaks after depurination or depyrimidination, enzymes in the base excision repair pathway take advantage of the weakened DNA backbone to restore the integrity of DNA. Hydrolysis reactions also occur at the exocyclic amino groups of the DNA bases, leading to deamination products (Figure 1). Deamination of cytosine, adenine, guanine, and 5-methylcytosine leads to uracil, hypoxanthine, xanthine, and thymine, respectively, with cytosine being the major target of this reaction (Figure 2a)

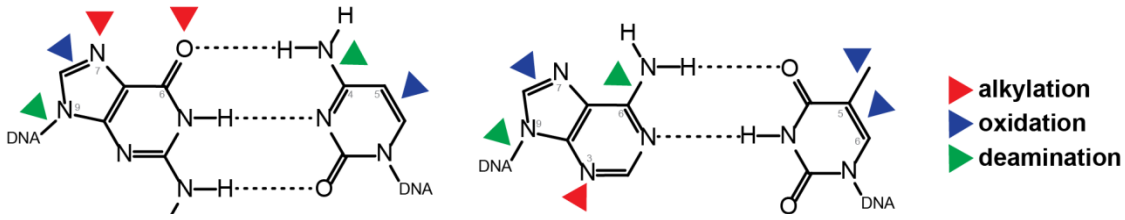


Figure 1. Major sites of DNA damage. Sites susceptible to hydrolytic depurination and deamination, oxidative damage, and alkylation damage are highlighted on the base pairs of the four DNA bases: guanine, cytosine, adenine, and thymine.

(Lindahl, 1993). The double-stranded structure of DNA provides some protection against deamination of cytosine and therefore uracil is created much more rapidly on single-stranded DNA (Lindahl & Nyberg, 1974). The lesions that persist in the DNA sequence can cause spontaneous mutagenic events, usually as a result of biological processes that involve denaturation of DNA (replication, recombination, and transcription) and must be removed by DNA repair enzymes, usually DNA glycosylases via the base excision repair pathway.

Reactive oxygen species are a ubiquitous consequence of the aerobic cellular environment and a source of endogenous DNA damage. Of the many reactive oxygen species, hydroxyl radicals are the most damaging and either add to double bonds of DNA bases or abstract hydrogens from deoxyribose sugars (Figure 2a) (Breen & Murphy, 1995). Ring saturated pyrimidines like thymine glycol and imidazole ring-opened derivatives of DNA bases like formamidopyrimidine (FaPy) lesions are products of hydroxyl radical reactions (Figure 3b). Lipid peroxidation products like malondialdehyde (MDA) and 4-hydroxynonenal (4-HNE) can lead to mutagenic exocyclic etheno adducts like pyrimido[1,2-*a*]purin-10(*3H*)-one (M_1G) and *1,N*^{6}-ethenoadenine (εA) (Figures 2a and 3c)

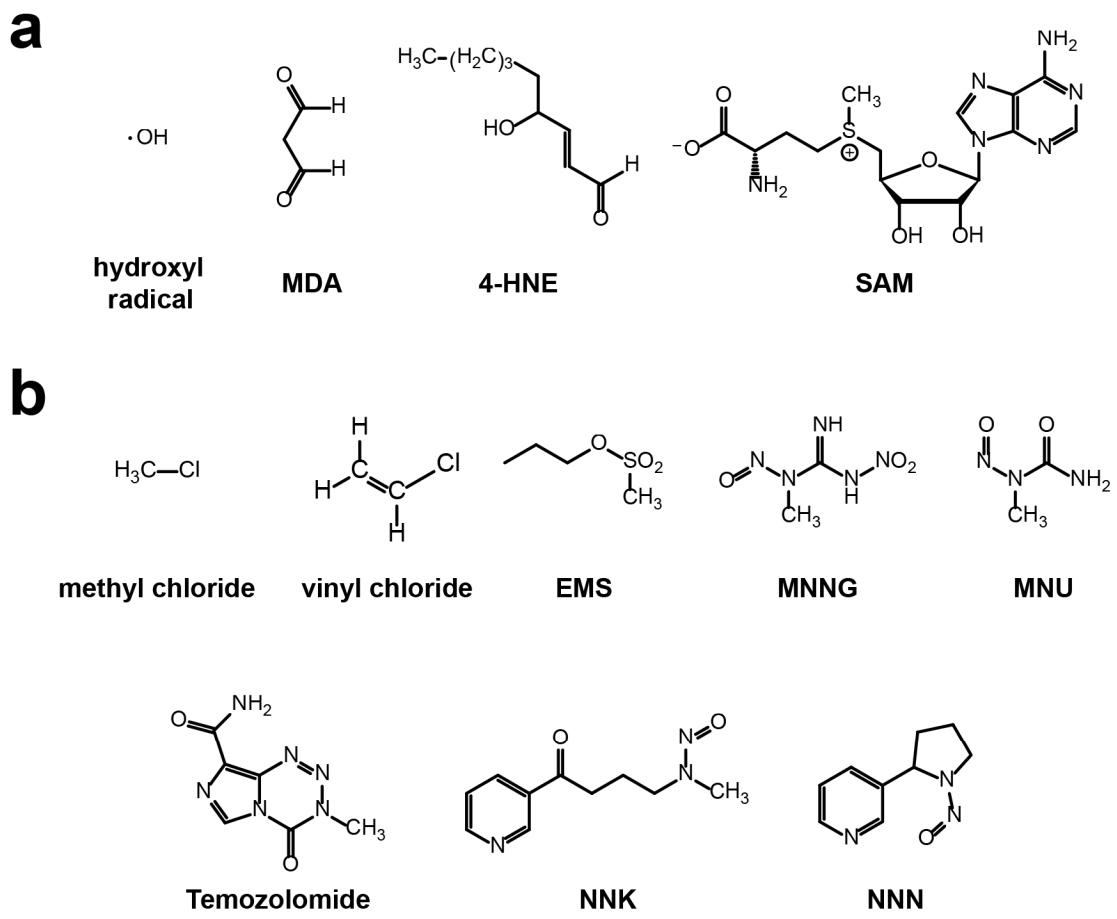


Figure 2. Common DNA damaging agents. **a**, Endogenous damaging agents may be the result of reactive oxygen species, lipid peroxidation products, or transmethylation cofactors. **b**, Exogenous damaging agents include environmental and laboratory alkylating agents.

(reviewed in Marnett, 2000). A more biologically interesting base lesion generated by reactive oxygen species is 7,8-dihydro-8-oxoguanine (8-oxoG) because it is directly mutagenic: the lesion preferentially forms a base pair with adenine instead of cytosine, leading to transversion mutations during replication (Figure 3b) (Kasai & Nishimura, 1984a; Kasai & Nishimura, 1984b; Shibutani et al, 1991). Due to this mutagenic potential, 8-oxoG must be consistently repaired in the cell, a process for which specific DNA glycosylases have evolved. Glycosylases also respond to endogenous DNA damage that results from nonenzymatic DNA methylation reactions, of which the most

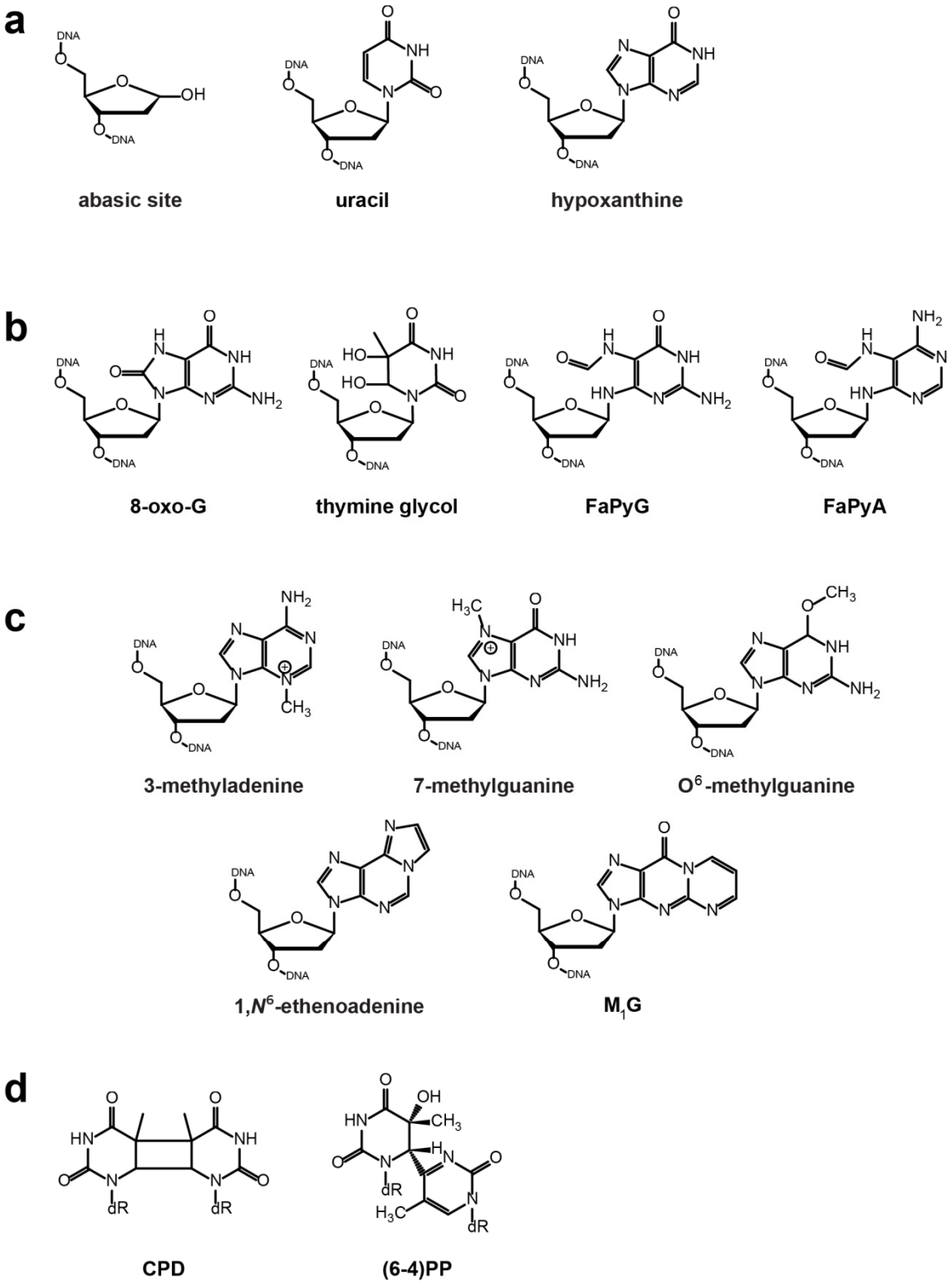


Figure 3. Common DNA lesions. DNA lesions may be the result of depurination or deamination reactions (**a**), oxidative damage (**b**), alkylation damage (**c**), or UV radiation (**d**).

prevalent source is intracellular *S*-adenosylmethionine (SAM) (Figure 2a). SAM acts as a methyl group donor cofactor in most enzymatic DNA methylation reactions but can also be a weak nonenzymatic alkylating agent (Rydberg & Lindahl, 1982). SAM preferentially alkylates the exocyclic ring nitrogen atoms of purines producing the potentially mutagenic *N*7-methylguanine (7mG) and cytotoxic *N*3-methyladenine (3mA) lesions (Rydberg & Lindahl, 1982) and to a lesser extent mutagenic O^6-methylguanine (O^6-mG), which may also be the result of nitrosation of endogenous metabolites (Barrows & Magee, 1982; Taverna & Sedgwick, 1996).

Exogenous DNA Damage

DNA damage can be the result of environmental sources including ionizing and UV radiation and chemical agents. Effects of UV radiation are biologically important since solar UV light is a constant in the environment. UV radiation leads to the formation of DNA photoproducts such as cyclobutane pyrimidine dimers (CPD), pyrimidine-pyrimidone adducts or (6-4) photoproducts [(6-4)PP] and thymine glycol (Figure 3d) (Setlow, 1966; Varghese & Wang, 1968; Yamane et al, 1967). Around a third of all UV induced photoproducts are CPDs where a cyclobutane ring is formed between the C5 and C6 carbons of each base. (6-4) photoproducts, which comprise around a fourth of photoproduct lesions, are formed through an oxetane intermediate to link the C6 of the 5′ pyrimidine to the C4 of the 3′ pyrimidine (Thompson & Sancar, 2002). These mutagenic and cytotoxic lesions interfere with DNA replication and transcription and are removed in cells by direct reversal, base excision, or nucleotide excision repair pathways.

Chemical poisons and carcinogens can have devastating consequences on DNA. Once used for chemical warfare, the effects of mustard gas and its derivatives have been co-opted for use as chemotherapeutic agents (Lawley & Phillips, 1996). These derivatives have also been used in the laboratory to study DNA repair processes. One of the most widely studied groups of chemicals includes the electrophilic compounds known as alkylating agents, which react with the many nucleophilic sites on DNA, particularly the N7 and O6 positions of guanine and the N3 position of adenine (Figure 1) (Singer & Grunberger, 1983). Laboratory methylating agents like methylnitrosourea (MNU), *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG), and ethyl methanesulfonate (EMS) are often used to study DNA glycosylases that remove 7-mG, 3-mA, and O^6-mG (Figures 2b and 3c) (Lawley, 1966). These enzymes most likely evolved to remove these lesions when they are the result of environmental mutagens and carcinogens such as methyl chloride (Vaughan et al, 1993) and vinyl chloride, which produces etheno adducts (Figure 2b) (Guengerich, 1994). Several environmental compounds can be metabolized to electrophilic reactants, which, like alkylating agents, can cause DNA damage at nucleophilic centers. Carcinogens such as aromatic amines, benzo[a]pyrene, aflatoxins, and 4-nitroquinoline 1-oxide are all metabolized by cellular processes which consequently produce damaging DNA adducts (reviewed in Brookes, 1971). Tobacco-specific *N*-nitrosamines, including 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and *N*′-nitrosonornicotine (NNN), are thought to be the major cause of smoking-related human cancers (Figure 2b) (Hecht, 1999). These carcinogens are metabolically activated and go on to form methylated and pyridyloxobutylated DNA adducts primarily at the N7 and O^6 sites of guanine residues (reviewed in Hecht, 1999).

DNA Repair of Single-stranded Adducts

In order to combat the deleterious effects of damaged or incorrect DNA bases, cells have developed many DNA repair pathways to restore the DNA (Figure 4). This review focuses on the repair of damage that affects one strand of the DNA duplex. However, damage may also occur to both strands of the DNA, resulting in inter-strand crosslinks or double-strand breaks, which are repaired by homologous recombination and non-homologous end-joining pathways. Damage to DNA may also be tolerated by the cell, leading to arrested replication, until other repair mechanisms like translesion DNA synthesis take over. In concert, these pathways prevent the cytotoxic and mutagenic effects DNA damage has on cells. A breakdown in these pathways can result in cell death and human disease, highlighting their importance in maintaining genomic integrity.

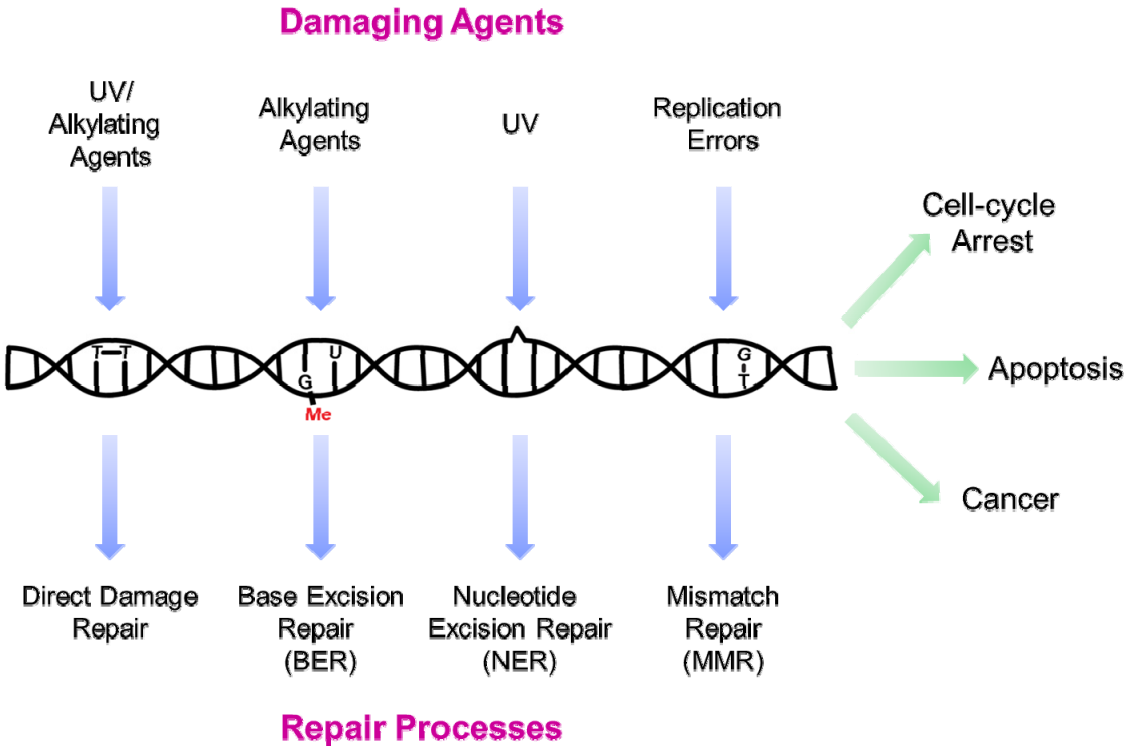


Figure 4. DNA damage repair pathways.

Direct Damage Reversal

Direct reversal of DNA damage is a highly specific and relatively error free process that involves repair by a single enzyme without synthesis of new DNA. Reversal of intra-strand DNA damage resulting from UV radiation, including cyclobutane pyrimidine dimers and (6-4) pyrimidine-pyrimidine photoproducts, is carried out by a family of enzymes known as photolyases, which are specific for a given photoproduct. The blue-light activated pyrimidine dimer-DNA photolyases (PD-photolyases) and (6-4) photoproduct-DNA photolyases [(6-4)-photolyases] utilize similar mechanisms to directly reverse intra-strand cross-links (reviewed in Sancar, 2008). Both rely on non-covalently bound chromophores, the primary flavin [flavin adenine dinucleotide (FAD)] and in most cases a secondary pterin (5,10-methenyltetrahydrofolyl polyglutamate (MTHF)] in an electron transfer cycle to restore the DNA strand to the natural bases (Sancar, 2003). While CD-photolyases simply break the bond formed between each base, (6-4)-photolyases are thought to stabilize the oxetane intermediate before restoration of the bases by cyclic electron transfer (Thompson & Sancar, 2002). As ubiquitous as these photolyases are in a variety of species, no functional photolyases have been identified in placental mammals (Sancar, 2003).

Reversal of base damage caused by alkylating agents is carried out by two families of enzymes, alkylguanine alkyltransferases (AGTs) and AlkB oxidative demethylases. Both O^6-alkylguanine-DNA alkyltransferase I, known as the ada protein in *E. coli* and AGT in humans, and O^6-alkylguanine-DNA alkyltransferase II are specific for repair of O^6-alkylguanine and O^4-alkylthymine major groove lesions (Demple et al, 1982; Potter et al, 1987). Both classes of AGTs are suicide proteins, with a single

molecule catalyzing the transfer of a single alkyl group on the DNA base by nucleophilic attack to an active site cysteine residue (Lindahl et al, 1982). Structures of hAGT bound to double-stranded DNA reveal that the enzyme uses a common base-flipping mechanism to gain access to the modified base after binding the minor groove of DNA through a helix-turn-helix motif (Daniels et al, 2004; Duguid et al, 2005).

Enzymes related to *E. coli* AlkB are 2-oxoglutarate- and iron-dependent dioxygenases which repair *N*1-methylpurines and *N*3-methylpyrimidines by oxidative demethylation (Begley & Samson, 2003). This family of enzymes, which includes the human homologs ABH1, ABH2, and ABH3, releases the oxidized methyl group as formaldehyde using α-ketoglutarate as a co-substrate and Fe(II) as a cofactor (Aas et al, 2003; Gerken et al, 2007; Ringvoll et al, 2008; Ringvoll et al, 2006). Recent structural characterization of AlkB and ABH2 in complex to double-stranded DNA illustrates how the enzymes utilize different damage recognition and base-flipping mechanisms despite their similar architectures (Yang et al, 2008). ABH2 works similarly to other base-flipping enzymes by inserting a "plug" residue into the space left by the flipped base. AlkB, however, lacks this residue and instead distorts the DNA backbone on either side of the lesioned base in order to flip it into the active site (Yang et al, 2008). This difference suggests why AlkB prefers DNA lesions in single-stranded DNA or RNA whereas ABH2 is a double-stranded DNA repair protein.

Base Excision Repair

The base excision repair pathway (BER) repairs DNA damage by removing the entire DNA base. Single base lesions, resulting from oxidation damage or chemical

modification, are recognized and removed by damage specific DNA glycosylases (Seeberg et al, 1995). Mono-functional DNA glycosylases initiate BER by cleaving the *N*-glycosylic bond, generating an abasic (AP) site which is further processed by the rest of the enzymes in the pathway (Figure 5) (Dianov & Lindahl, 1994). Apurinic/apyrimidinic (AP) endonucleases hydrolyze the phosphodiester bond 5′ to the AP site creating a nick in the DNA duplex. Several DNA glycosylases are bi-functional enzymes that have an additional AP lyase activity not present in mono-functional glycosylases. These enzymes incise the DNA duplex 3′ to the AP site via a β-elimination or β,δ-elimination mechanism. Despite the method used, a 5′ terminal phosphate residue is generated and removed by the activity of specific DNA-deoxyribophosphodiesterases (dRpase) (short-patch BER) or a flap endonuclease (long-patch BER) (Klungland & Lindahl, 1997). Long-patch BER removes an extra three nucleotides 3′ to the original lesion. The one or four nucleotide gap in the DNA is filled by a DNA polymerase: DNA polymerase I (*E. coli*) and DNA polymerase β (humans) function following the action of mono-functional glycosylases, while DNA polymerase δ or ε (mammalian cells) function in the case of bi-functional glycosylases (Kubota et al, 1996; Matsumoto et al, 1994). Finally the two ends of the DNA are sealed by species specific ATP-dependent DNA ligases. Although BER enzymes evolved to respond to endogenous DNA damage, several DNA glycosylases are essential for the removal of exogenous alkylation-damage products, including mutagenic (1,N^6-ethenoadenine) and cytotoxic (*N*3-methyladenine) lesions.

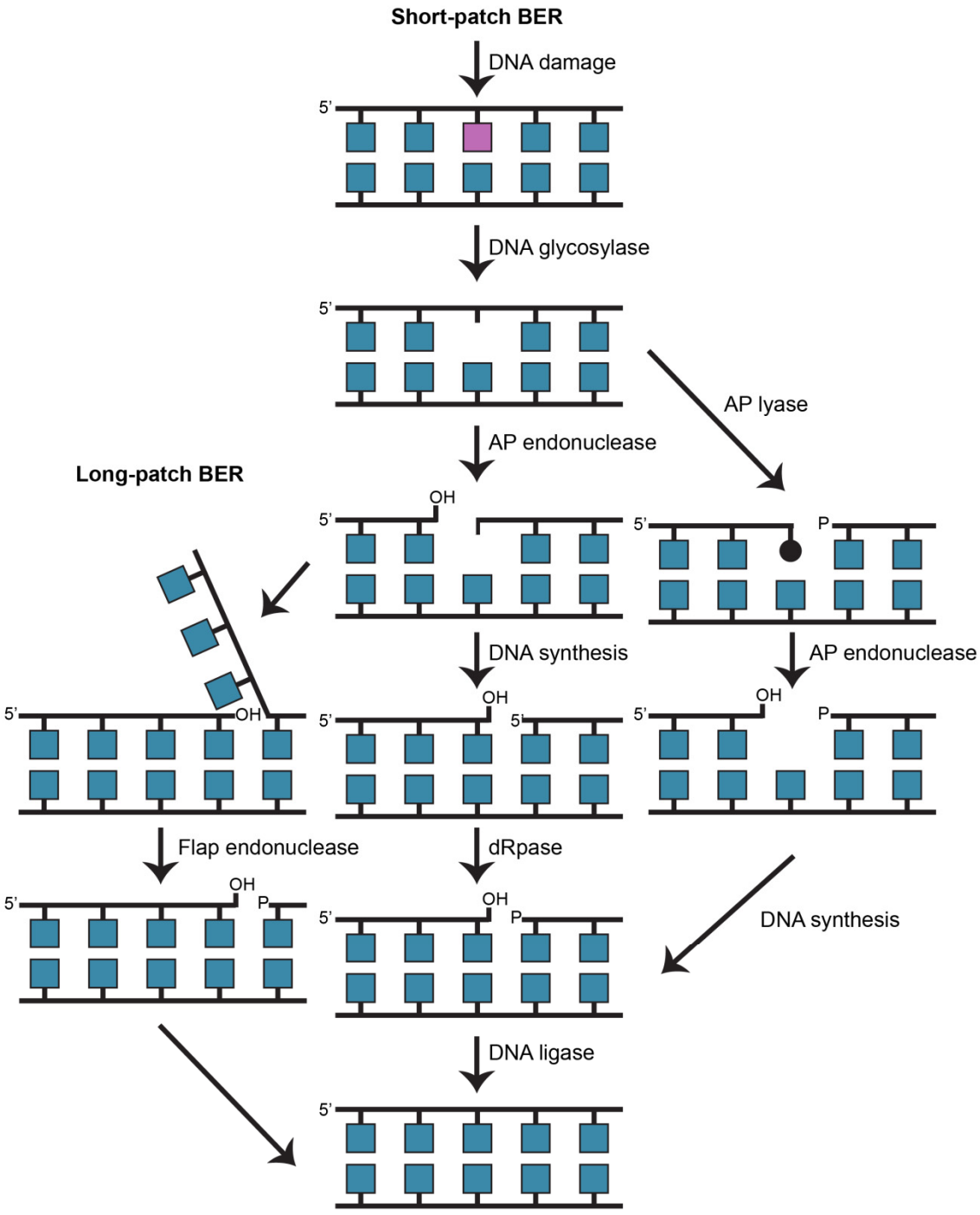


Figure 5. Base excision repair. BER pathways are initiated by damage specific mono-functional (center pathway) or bi-functional (right pathway) DNA glycosylases. The abasic site product is processed by AP endonuclease followed by repair synthesis. Finally the DNA is restored by a DNA ligase. In long-patch BER (left pathway) a 5′ flap created by strand displacement synthesis is cleaved by a Flap endonuclease and the product is repaired by a DNA ligase. [Adapted from (Scharer & Jiricny, 2001)].

Nucleotide Excision Repair

Nucleotide excision repair (NER) is a multistep process that removes bulky DNA lesions confined to one DNA strand, including UV-induced CPDs and (6-4) photoproducts. The overall process of NER is similar in prokaryotes and eukaryotes. Once the damaged base is recognized, incisions are made on either side of the damaged base and an oligonucleotide fragment is excised. The gap is filled by repair synthesis and finally covalently sealed by DNA ligase (Sancar & Reardon, 2004 302). Despite this common pathway, the individual repair proteins are very different between prokaryotes, which use a small number of enzymes, and eukaryotes, which use a large number of proteins with no orthologs in prokaryotes. In *E. coli* the recognition and incision steps are carried out by UvrA, UvrB, and UvrC proteins that sequentially excise ~12 nucleotides. UvrD (DNA helicase II), DNA polymerase I, and DNA ligase complete the repair to restore the DNA (reviewed in Reardon & Sancar, 2005). Many of the proteins involved in mammalian NER were discovered by studying the xeroderma pigmentosum (XP) syndrome (Cleaver, 1968). Seven complementation groups (XPA-XPG) are caused by mutations in genes involved in the NER pathway, and patients with this disorder are sensitive to sunlight and have a high incidence of skin cancer. In mammalian cells and other eukaryotes, NER is initiated by binding of the XPC-RAD23B heterodimer to sites of DNA distortion, specifically where base pairing is disrupted (Masutani et al, 1994). This distortion recognition allows for recruitment of XPA, XPG, replication protein A (RPA), and a helicase containing transcription initiation factor, TFIIH, which unwinds the DNA duplex in an ATP-dependent manner (reviewed in Riedl et al, 2003). Once the incision nuclease XPG is in place it cleaves the damaged strand on the 3′ side, while

ERCC1-XPF nuclease cleaves the 5′ side of the damaged strand (Mu et al, 1996; Sijbers et al, 1996). Anywhere from 25-30 nucleotides are excised and replaced by the DNA synthesis replicative machinery with DNA polymerase δ or ε, RPA, proliferating cell nuclear antigen (PCNA), a processivity clamp for the polymerases, replication factor C (RFC), and DNA ligase I (reviewed in Reardon & Sancar, 2005). The parallels in each system, distortion recognition, ATP-dependent helicase unwinding, endonuclease excision, and DNA synthesis suggest convergent evolution of the pathway.

Mismatch Repair

In order to avoid mutations resulting from errors during DNA replication and genetic recombination or the deamination of 5-methylcytosine to thymine, prokaryotes and eukaryotes have mismatch repair (MMR) pathways responsible for restoring single base mismatches to the correct base (Jiricny, 1998). MMR also responds to nucleotide insertion or deletion events during DNA synthesis. MMR systems have the unique task of identifying the newly synthesized daughter strand in order to distinguish the incorrect base from the correct base (Modrich, 1997). This is achieved in *E. coli* by recognition of methylation by dam methyltransferase at GATC sites on the template strand (reviewed in Joseph et al, 2006). This so-called methyl-directed mismatch repair relies on the *mutS*, *mutL*, *mutH*, and *uvrD* gene products. Binding of the MutS ATPase homodimer to mismatched or small insertion or deletion loops in the DNA is followed by loading of the MutL homodimer. This MutS-MutL complex relocates to the nearest hemimethylated GATC site and recruits the endonuclease MutH to nick the unmethylated strand. The UvrD DNA helicase II then unwinds the DNA, and the damaged strand is excised by an

appropriate 5′-3′ or 3′-5′ exonuclease, ExoI and ExoX or RecJ and ExoVII, respectively. DNA polymerase II fills in the resulting gap, and DNA ligase seals the nick restoring the DNA duplex. Methylation independent DNA mismatch repair occurs in gram-positive bacteria, of which the *Streptococcus pneumoniae* system is the most widely studied. While the mechanisms of eukaryotic MMR are very similar to the *E. coli* system, it remains unclear how the newly synthesized DNA strand is recognized, since eukaryotes lack methylation of GATC sequences (Modrich, 1997). The eukaryotic systems replace the MutS and MutL homodimers with a heterodimer of MutS and MutL homologues, either MutSα or MutSβ, relying on PCNA, which interacts with the MutSβ heterodimer (reviewed in Buermeyer et al, 1999; Kolodner & Marsischky, 1999). PCNA, together with Okazaki fragments on the lagging strand, has been proposed to play a role in strand discrimination in eukaryotes (Pavlov et al, 2003).

DNA Glycosylases

Our understanding of DNA glycosylases, the damage-specific enzymes that initiate base excision repair, has benefited from the numerous high-resolution structures of bacterial, eukaryotic, and mammalian enzymes. This detailed structural work has been combined with kinetic and thermodynamic experiments as well as mutational analysis to understand how glycosylases locate and recognize their lesioned substrates. Glycosylases face many challenges in the detection of DNA damage. The lesions must first be located from amongst the vast majority of unmodified DNA and then the damage must be recognized by the enzymes. Glycosylases have to gain access to the C1′ carbon of the deoxyribose sugar, which sits within the DNA duplex, in order to carry out

catalysis. Most DNA glycosylases are specific for a given lesion although some, like the alkylpurine DNA glycosylases, remove a variety of damaged substrates. Structures of DNA glycosylases bound to DNA substrates and products reveal that these enzymes, much like other DNA repair enzymes, utilize a common base-flipping mechanism for substrate recognition in which the DNA is bent at the site of the lesion and the nucleotide is flipped into an extrahelical orientation into an active site cleft where the enzyme can gain access to the *N*-glycosylic bond (Figure 6) (Slupphaug et al, 1996; Verdine & Bruner, 1997).

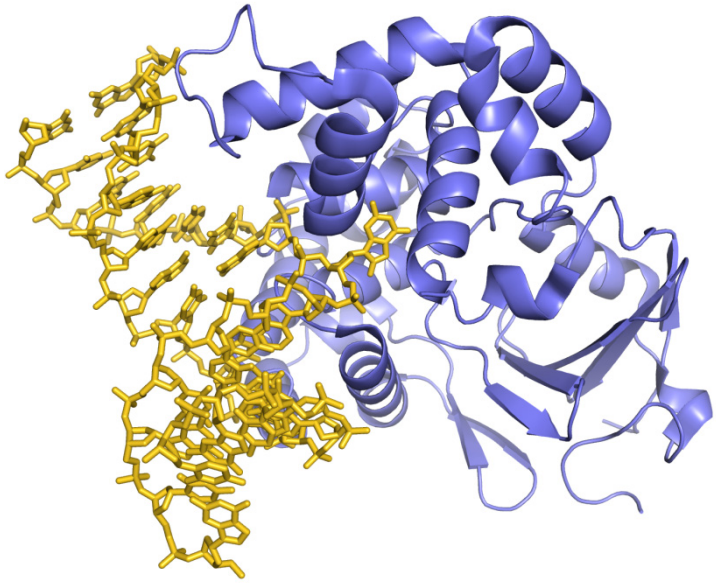


Figure 6. DNA glycosylase base-flipping. Helical representation of human 8-oxoguanine DNA glycosylase (1emb) (Bruner et al. 2000) bound to oxoG•C containing DNA. The enzyme binds DNA in the minor groove, kinks the DNA at the site of the lesion, and flips the oxoG lesion into the enzyme active site where chemistry can occur. DNA is represented as gold sticks.

Recognition of DNA Damage

Three damage-searching mechanisms have been proposed for enzymes that process lesions extrahelically: 1) the enzyme must flip every base into its active site to check for damage; 2) the enzyme detects the intrahelical lesion and then actively flips it

into its active site; or 3) the enzyme captures a transiently extrahelical lesion (Duguid et al, 2003; Verdine & Bruner, 1997). In the last two cases, the enzyme may sense instability of the damaged base pair, which seems obvious in the case of bulky lesions like εA or thymine glycol that distort the DNA helical structure; however, more subtle perturbations to the base stack in the case of mismatched base pairs or methylated lesions may also lead to instability of the base pair (O'Brien & Ellenberger, 2004b; Yang, 2006). The first scenario, in which brute force must be employed to check every base pair, is unlikely due to the costly, unfavorable energetics that would be required to flip every base. Also, experiments performed to investigate damage recognition favor the other two mechanisms. Structures of MutM cross-linked to undamaged DNA support the second mechanism by demonstrating that the enzyme senses an intrahelical lesion by actively interrogating the DNA duplex without base-flipping (Banerjee et al, 2006). Crosslinking experiments with alkyltransferases and DNA lesions support the third mechanism, in which damaged bases are located by sensing their structural perturbations and captured when they happen to be in an extrahelical orientation (Duguid et al, 2003). Structural studies on UNG also proposed a more passive role of the enzyme in trapping extrahelical pyrimidines (Parker et al, 2007). However, evidence from high-resolution crystal structures, biochemical approaches, and molecular dynamics studies suggest that most enzymes actively base flip lesions and do not simply capture an extrahelical base that may be the result of normal DNA breathing (Stivers, 2004). DNA glycosylases intercalate protein residues into the space left by the flipped nucleotide to maintain base stacking properties; however, it is still unclear if these residues are actively pushing

lesions into an extrahelical formation or are simply acting as structural supports once the lesion has been extruded.

This active mechanism is thought to follow nonspecific DNA binding of the glycosylase a small distance away from the lesion. The glycosylase then scans the DNA, either by sampling each base or identifying perturbations to the helical structure, until it reaches the lesion (Banerjee et al, 2006; Verdine & Bruner, 1997). It has been determined that DNA glycosylases T4 endonuclease V (Dowd & Lloyd, 1989; Gruskin & Lloyd, 1986), *E. coli* MutY, and MutM/Fpg (Francis & David, 2003) scan the DNA by processively sliding along the duplex. There is also evidence that UNG uses facilitated diffusion and a DNA hopping mechanism to search for uracil in long stretches of DNA (Higley & Lloyd, 1993; Porecha & Stivers, 2008). While there exist several hypotheses on the mechanism of DNA damage detection, there is no conclusion on how glycosylases find specific lesions among the vast expanse of normal DNA base pairs.

Substrate Specificity and Catalytic Mechanism of DNA Glycosylases

Regardless of whether the process is active or passive, the adducted DNA is flipped into an active site cavity on the DNA glycosylase. It is now thought that damaged or mispaired bases are identified based on size, shape, hydrogen-bonding capability, and electronic charge distribution in order to fit into the active site (Yang, 2008). Undamaged bases are most often excluded simply due to steric constraints of the active site. The specificities of these repair enzymes are determined, therefore, by the chemical and physical properties of the base-pair and the base binding pocket as well as the particular protein architectures used to probe the bases within the DNA duplex.

Following the flipping step, glycosylases utilize functional group chemistry to cleave the *N*-glycosylic bond, which liberates the lesioned base. Glycosylases may be broadly categorized as mono-functional or bi-functional enzymes. Most mono-functional glycosylases are proposed to catalyze a hydrolysis reaction by deprotonating a water molecule to generate a hydroxyl ion to attack the anomeric carbon on the lesioned nucleotide. Bi-functional DNA glycosylases, however, use an active site amine residue as a nucleophile rather than a water molecule proceeding through a Schiff base intermediate (Sun et al, 1995).

Structural Superfamilies of DNA Glycosylases

While DNA glycosylases may be classified according to reaction mechanism or substrate specificity, they may also be classified by three-dimensional structure. Examining glycosylase structure will provide a context for the results presented in this thesis. Prior to work presented in Chapter II, it was known that DNA glycosylases were comprised of five structural superfamilies (Figure 7). Although each group has unique structural characteristics, including their overall DNA binding mode, they all work by a similar base-flipping mechanism.

T4 Endonuclease V Superfamily

The first structural determination of a DNA glycosylase was of T4 endonuclease V, a pyrimidine dimer (PD) DNA glycosylase encoded by bacteriophage T4 (Vassylyev et al, 1995; Morikawa et al, 1992). Although characterized as a bi-functional DNA glycosylase and considered part of the BER pathway, T4 EndoV does not excise free

bases since only one of the two *N*-glycosylic bonds in a CPB dimer is hydrolyzed. T4 EndoV is a small monomeric protein and the crystal structures of the wild-type enzyme (Morikawa et al, 1992) and an inactive mutant complexed to a CPD-containing DNA substrate revealed a compact, all alpha-helical protein fold that supports both the PD-

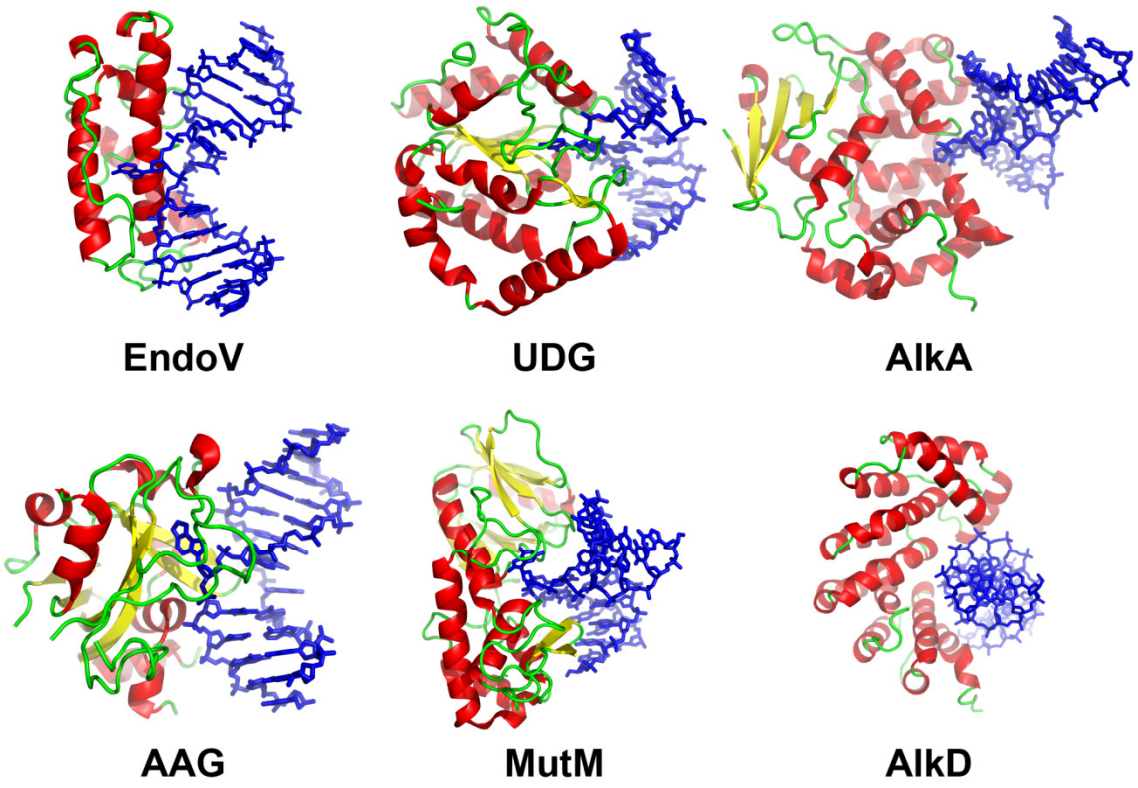


Figure 7. Six structural superfamilies of DNA glycosylases bound to DNA. Representative crystal structures shown are: T4 pyrimidine dimer DNA glycosylase EndoV (1vas) (Vassylyev et al, 1995); human uracil-DNA glycosylase, UDG (1emh) (Parikh et al, 2000); *E. coli* 3-methyladenine DNA glycosylase II, AlkA (1diz) (Hollis et al, 2000); human methylpurine DNA glycosylase, AAG (1ewn) (Lau et al, 2000); *Bacillus stearothermophilus* 8-oxoguanine DNA glycosylase, MutM (1lit) (Fromme & Verdine, 2002); and *Bacillus cereus* AlkD (3jxy) (Avery et al, 1944). Proteins colored according to secondary structure and lesion-containing DNA is shown as blue sticks.

specific glycosylase function and the AP lyase function of the protein (Vassylyev et al, 1995). Binding of the three alpha helices to the DNA creates a 60° kink at the site of the lesion. Unusually, the adenine opposite the 5′ thymine of the CPD is base-flipped into a hydrophobic pocket without making specific contacts to residues in the enzyme active

site. Active site residues bind to DNA, and a catalytically essential glutamate residue gains access to the *N*-glycosylic bond of the 5′ thymine of the CPD (Vassylyev et al, 1995).

Uracil DNA Glycosylase Superfamily

E. coli Ung was the first enzyme identified as a member of the uracil DNA glycosylases (UDGs) after it was discovered that it removed uracil, often the result of cytosine deamination or misincorporation of dUTP, from DNA by cleaving the *N*-glycosylic bond and releasing the free base (Lindahl, 1974). Much of what is known about substrate recognition and base catalysis is the result of structural and biochemical analysis of *E. coli* and human uracil DNA glycosylase; therefore, UDG has become the paradigm for understanding all glycosylases. Five UDG families are structurally related and share a common mode of DNA glycosylase action (reviewed in Pearl, 2000) and summarized below. Family 1, which contains E. coli Ung and human UNG, remove uracil from single and double stranded DNA. Family 2 is comprised of the structural orthologs bacterial mismatch-specific uracil DNA glycosylase (Mug) and eukaryotic thymine DNA glycosylase (TDG). Although TDGs remove both thymine and uracil from T•G or U•G mismatches, Mugs only remove uracil from mismatches. UDGs identified to be particularly active toward uracil in single-stranded DNA (SMUGs) make up Family 3, although SMUGs prefer double-stranded substrates. SMUG1 from vertebrates also removes 5-hydroxymethyluracil, 5-formyluracil, and 5-hydroxyuracil. Family 4 enzymes, typified by *Thermotoga maritima* UDG (TmUDG), contain iron-sulfur structural domains and remove uracil and thymine from mismatches and are distant

homologs of MUG. Family 5 UDGb enzymes are so far only present in hyperthermophiles and were found to efficiently remove hypoxanthine from DNA. Some uracil DNA glycosylases, like MIG and MBD4, as well as a family of archael UDGs, are actually members of the helix-hairpin-helix structural class of DNA glycosylases.

Uracil DNA glycosylases are small monomeric enzymes with a highly conserved, common parallel α/β fold important for double-stranded DNA recognition (Mol et al, 1995). In *E. coli* Ung, the fold creates a positively charged groove at the C-terminal end of the β-sheet where active site residues are clustered (Mol et al, 1995). In 1996 the landmark structure of human UDG (L272R-UDG) bound to its DNA substrate (a U•G mismatch) was determined, revealing mechanistic details about substrate recognition and catalysis which are common for UDGs (Slupphaug et al, 1996). The enzyme undergoes a conformational change upon DNA binding; the DNA is kinked ~45° with Leu272 intercalating into the minor groove and promoting flipping of the uracil nucleotide into the substrate-binding pocket found in a cleft created between β1 and β3 (Parikh et al, 1998; Slupphaug et al, 1996). This binding pocket is extremely specific for uracil which differs from thymine by only a single methyl group (Mol et al, 1995). The binding pocket sterically prohibits purine bases from binding and several hydrogen bonds are created between active site residues and the exocyclic oxygen atoms of uracil. Stacking of Tyr147 sterically blocks interactions with the 5-methyl group of thymine, preventing removal of this base. Mutation of this residue results in the creation of TDGs (Dong et al, 2000).

Biochemical studies on *E. coli* Ung and human UDG have presented differing mechanisms for base removal. Initially it was thought the mechanism was driven by

acid-base catalysis from His187/His268 and Asp64/Asp145 (Slupphaug et al, 1996). This theory was replaced when evidence suggested that excision of uracil relies on Asp64 acting as a general base to activate a water molecule and His187 contributing to stabilization of uracil by becoming a hydrogen bond donor to the O2 position during the transition state (Drohat et al, 1999). Later, comparison of high-resolution crystal structures of UDG bound to an uncleaved-substrate analog and a cleaved-product DNA demonstrated an alternative catalytic mechanism and sought to reconcile the discovery that many active site mutants actually enhanced excision of uracil (Parikh et al, 2000). These structures revealed a contorted substrate and deoxyribose sugar upon base-flipping, suggesting that the enzyme places strain on the nucleoside, stretching the *N*-glycosylic bond, and couples positive stereoelectronic effects to the base distortion to promote cleavage of the *N*-glycosylic bond (Parikh et al, 2000). Moreover, a quantum-mechanical/molecular-mechanical approach demonstrated that a dissociative mechanism occurs through substrate-assisted catalysis wherein the DNA phosphate groups stabilize the positively charged oxocarbenium sugar intermediate present during the transition state (Dinner et al, 2001). Although debates continue on the mechanistic details, the UDG enzymes remain highly selective and efficient DNA glycosylases for a single substrate.

Helix-Hairpin-Helix Superfamily

Many structurally characterized DNA glycosylases contain a common DNA-binding domain known as a helix-hairpin-helix (HhH) motif, utilized by hundreds of repair proteins (Doherty et al, 1996; Thayer et al, 1995). First observed in *E. coli* Endo

III (Kuo et al, 1992), the HhH glycosylases contain two α-helical subdomains separated by an active site cleft that accommodates the flipped substrate nucleobase (Figure 8). One of these domains (helices αD-αJ) is highly conserved and contains the HhH motif (αI-αJ) (Doherty et al, 1996). The HhH anchors the protein to the DNA through electrostatic interactions between main-chain atoms from the hairpin region and the phosphoribose backbone. The HhH domain also contributes a bulky group (typically a Leu, Asn, or Gln side chain) that plugs the gap in the DNA left by the flipped-out nucleotide, and a second side chain (Phe, Tyr, Leu, or Pro) that wedges between the bases opposite the flipped out nucleotide (Figure 8a). Both plug and wedge residues are important for stabilizing the bent conformation of the DNA, and the wedge residue has been implicated in probing the DNA helix during the search process (Banerjee et al, 2006). The second domain, formed from the N- and C-termini (the N/C domain), is more varied in structure and often contains additional structural elements, including a zinc binding motif (TAG), a carbamylated lysine (MagIII), and an iron-sulfur cluster (MpgII) (Figure 8b). The precise role of the N/C domain is not clear, but it is suspected that these elements help fold this domain in order to form the active site cleft. There are six functionally diverse gene families that make up the HhH structural superfamily which includes both mono-functional and bi-functional glycosylases (Denver et al, 2003): alkyladenine DNA glycosylases (AlkD, TAG, MagIII) (Eichman et al, 2003; Labahn et al, 1996; Yamagata et al, 1996; Drohat et al, 2002), N-methylpurine DNA glycosylase II (MgpII) (Begley et al, 1999), A/G specific adenine glycosylase/mismatch glycosylase (MutY/Mig) (Guan et al, 1998; Mol et al, 2002), endonuclease III (Nth) (Kuo et al, 1992;

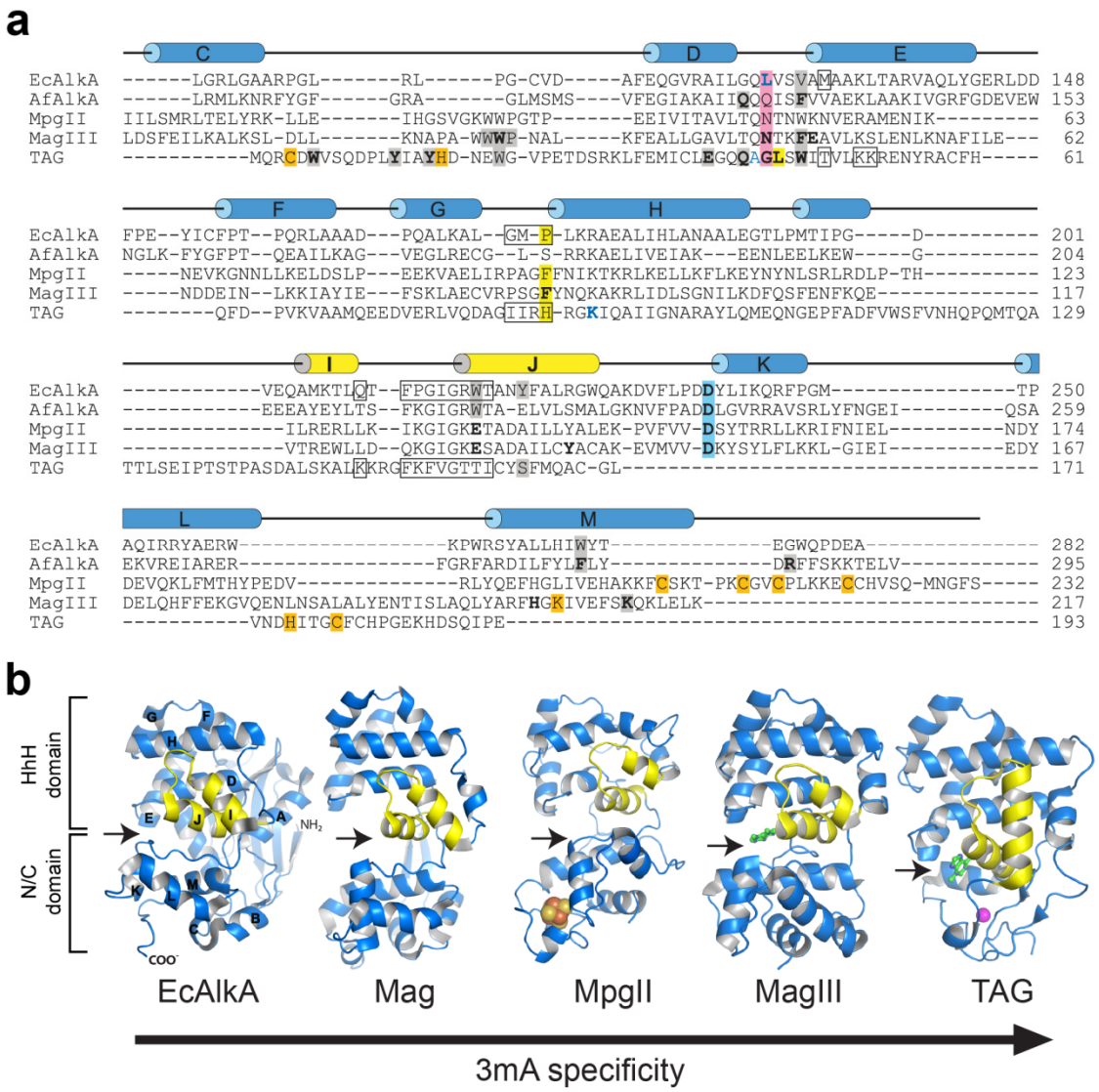


Figure 8. Alkylpurine helix-hairpin-helix superfamily. **a**, Structure based sequence alignment of alkyladenine DNA glycosylases *E. coli* AlkA, *A. fulgidus* AlkA, *B. halodurans* Mag, *H. pylori* MagIII, and *S. typhimurium* TAG shown with secondary structure from AlkA. MpgII was aligned with MagIII. Residues that contact DNA in protein/DNA complexes of AlkA and TAG are boxed, and intercalating plug and wedge residues are highlighted yellow and pink, respectively. Alkylpurine binding pocket residues and the catalytic aspartate are highlighted grey and blue, respectively. Residues that coordinate ions in TAG (Zn^{2+}), MagIII (carbamylated lysine), and MpgII (iron-sulfur cluster) are highlighted orange. **b**, Crystal structures are *E. coli* AlkA bound to 1-azaribose-DNA (1diz) (Hollis et al, 2000), *B. halodurans* Mag (2h56) (PDB, 2006), *H. pylori* MagIII bound to 3,9dma (1pu7) (Eichman et al, 2003), and *E. coli* TAG/THF-DNA/3mA complex (2ofi) (Metz et al, 2007). MpgII model was constructed using atomic coordinates from MagIII and MutY (1muy), (Guan et al, 1998) and the program SWISS-MODEL (Guex & Peitsch, 1997). HhH motifs are colored yellow and the substrate base binding pockets are marked with an arrow.

Thayer et al, 1995), 8-oxoG DNA glycosylase I (OggI) (Bruner et al, 2000) and 8-oxoG DNA glycosylase II (OggII) (Denver et al, 2003). Despite their similar structures, each family removes specific substrates, and the shape and chemical features of the active site cleft in each enzyme play a large role in defining the substrate specificity (Eichman et al, 2003). With several aromatic residues lining their active sites, it has been proposed that these enzymes use π-π or π-cation interactions to locate and recognize their substrates (Stivers, 2004).

Alkyladenine DNA Glycosylase Superfamily

The human alkyladenine DNA glycosylase, AAG falls into its own structural class and excises a broad range of alkylpurines, including 3mA and 7mG, and has a selective preference for neutral εA and Hx (O'Brien & Ellenberger, 2004a). The crystal structure of AAG in complex with DNA containing an abasic pyrrolidine transition-state analog showed that AAG is a single domain protein with a mixed α/β structure and a positively charged DNA binding surface (Lau et al, 1998; Scharer et al, 1998). The protein crystallized lacked the N-terminal 79 amino acids and thus the presence of a second domain is unknown. The DNA is bent at the damage site by ~22°, with B-form helical arms swung away from the protein. The pyrrolidine is rotated out of the DNA duplex and into a cavity on the protein surface. Tyr162 on the tip of a β-hairpin plugs the gap in the DNA left by the flipped pyrrolidine, and presumably stabilizes the distorted conformation of the extrahelical DNA. The pyrrolidine binding pocket is lined with aromatic and polar residues. A subsequent crystal structure of AAG bound to εA-

containing DNA showed the flipped εA base to be stacked between two tyrosine residues (Tyr127 and 159) and His136 inside the active site cavity (Figure 11a) (Lau et al, 2000).

Helix-Two Turn-Helix Superfamily

Bi-functional glycosylases including bacterial MutM/Fpg, endonuclease VIII/Nei, and the mammalian Nei-like NEIL enzymes are structurally related multi-domain enzymes that contain a helix-two turn-helix (H2TH) DNA-binding motif. MutM/Fpg removes oxidized purines like 8oxoG and FaPy products (Tchou et al, 1991), while the NEIL enzymes have been identified to excise oxidized pyrimidines as well as FaPy products (Bandaru et al, 2002). Endonuclease VIII/Nei also removes FaPy products but was originally identified as having activity towards thymine glycol and urea residues. Structural studies of MutM/Fpg revealed the protein contains two domains, an N-terminal domain containing an eight strand β-sandwich with two parallel alpha helices on either side, and a C-terminal domain consisting of a helix-two turn-helix motif and a zinc-finger motif which create a cleft for DNA binding (Fromme & Verdine, 2002; Gilboa et al, 2002). In all structures the protein creates a sharp bend in the DNA (~75°) and almost exclusively contacts the lesion-containing strand (Fromme & Verdine, 2002; Gilboa et al, 2002). The oxidized purine is flipped into an active site cleft and positioned for attack by the nucleophilic N-terminal proline residue which was shown to be involved in the formation of the Schiff base intermediate (Gilboa et al, 2002; Sidorkina & Laval, 2000). The reaction proceeds to incise the DNA on the inside of both the 5′ and 3′-phosphate groups leaving a one-nucleotide gap which must be processed by other enzymes in the BER pathway (Fromme & Verdine, 2002). The lesioned base is recognized by a loop in

the N-terminal domain of MutM/Fpg which becomes disordered after catalysis and the orphaned base opposite 8oxoG is detected by an arginine side chain which holds the lesion in the active site and pairs with the estranged base (Fromme & Verdine, 2002).

Tandem Helical Repeat Superfamily

Recently a sixth superfamily has been defined by two alkylpurine DNA glycosylases, AlkC and AlkD. These enzymes were identified in *Bacillus cereus* as functional complements to the alkylpurine specific DNA glycosylase *E. coli* AlkA (Alseth et al, 2006). The sequences of AlkC and AlkD are distantly related and distinct from other known proteins. Both specifically excise positively charged methylpurine bases 3mA and 7mG and have no measureable activity toward εA or Hx. This thesis describes the structure and function of this sixth superfamily in detail. Briefly, the high resolution crystal structure of *Bacillus cereus* AlkD shows that the protein adopts a C-shaped globular fold composed exclusively of helical HEAT-like repeats, and thus represents a new DNA glycosylase architecture (Rubinson et al, 2008). HEAT motifs are common protein binding domains that have been adapted by AlkD to bind DNA. The C-terminal α-helix of each HEAT repeat forms the inner, concave surface of the protein and contains lysine or arginine residues at conserved positions. Consequently, the concave surface of AlkD is positively charged and perfectly shaped to accommodate duplex DNA.

Alkylation Damage Response

Humans are exposed to alkylating agents from various environmental sources, including industrial processes, cigarette smoke, diet, and chemotherapy. These agents, in

addition to endogenous methyl donors, chemically modify the nucleobases of DNA to produce a variety of cytotoxic and mutagenic lesions that disrupt DNA replication and thus lead to heritable diseases and cancer (reviewed in Friedberg et al, 2006). To maintain genomic integrity amidst the constant threat of DNA alkylation, all organisms have devised multiple DNA repair strategies to eliminate the damage. As previously discussed, bases methylated at exocyclic substituents (e.g., O^6-methylguanine) are directly demethylated by DNA methyltransferases, including AGT in humans and ada in *E. coli*. Ring-substituted *N*1-methyladenine (1mA) and *N*3-methylcytosine (3mC) are specifically repaired through oxidative deamination by DNA dioxygenases such as *E. coli* AlkB and human homologs ABH1, ABH2, and ABH3 (reviewed in Sedgwick, 2004). The majority of alkylated bases, however, are eliminated from the genome by DNA glycosylases and the BER pathway (reviewed in Fromme & Verdine, 2004; Huffman et al, 2005). In fact, highly efficient DNA glycosylases have evolved in eukaryotes and bacteria to remove 3mA and O^6mG lesions. 7mG, however, is thought to be less readily repaired by glycosylases. Instead, it has been proposed that since the methylation increases depurination of the base, it is the resulting abasic sites which are processed by other enzymes in the BER pathway (Lindahl, 1993).

The Adaptive Response to Alkylation Damage

Because alkylation damage to DNA can promote instability of the genome and lead to heritable diseases and cancer, several DNA repair strategies described above respond to these base modifications to restore the DNA duplex. The repair of alkylation damage in bacteria, initially discovered in *E. coli*, is accomplished by a regulatory

pathway known as the adaptive response to alkylation damage. This pathway was identified by Leona Samson, a graduate student in John Cairns laboratory, as an inducible response to low levels of alkylating agents, MNNG, MNU, or MMS, and increased cellular resistance to mutagenic and lethal concentrations of these agents (Samson & Cairns, 1977). Four genes are stimulated in this pathway, *ada*, *alkA*, *alkB*, and *aidB* (Figure 9) (Sedgwick, 2004). The gene products, Ada, a methyltransferase that directly repairs O^6-alkylguanine and O^4-alkylthymine and acts as a transcriptional activator for the other three genes, AlkA, a DNA glycosylases in the BER pathway that responds to cytotoxic 3mA and mutagenic ethenoadenine, and AlkB, an oxidative demethylase that directly restores 1mA and 3mC to adenine and cytosine, respectively, have been structurally and functionally characterized. Recently, a high-resolution crystal structure of AidB has been determined, although the role of AidB in the adaptive response remains elusive (Bowles et al, 2008). It has been postulated that AidB, a redoxactive flavin containing enzyme that was shown to bind double-stranded DNA, may act to repair DNA via a dehydrogenase mechanism or more likely work as a detoxifying enzyme to deactivate alkylating agents (Landini et al, 1994; Rohankhedkar et al, 2006). Establishing the functional roles and elucidating the structures of the enzymes involved in the adaptive response pathway has led to a greater understanding of how alkylation damage is recognized and removed in cells.

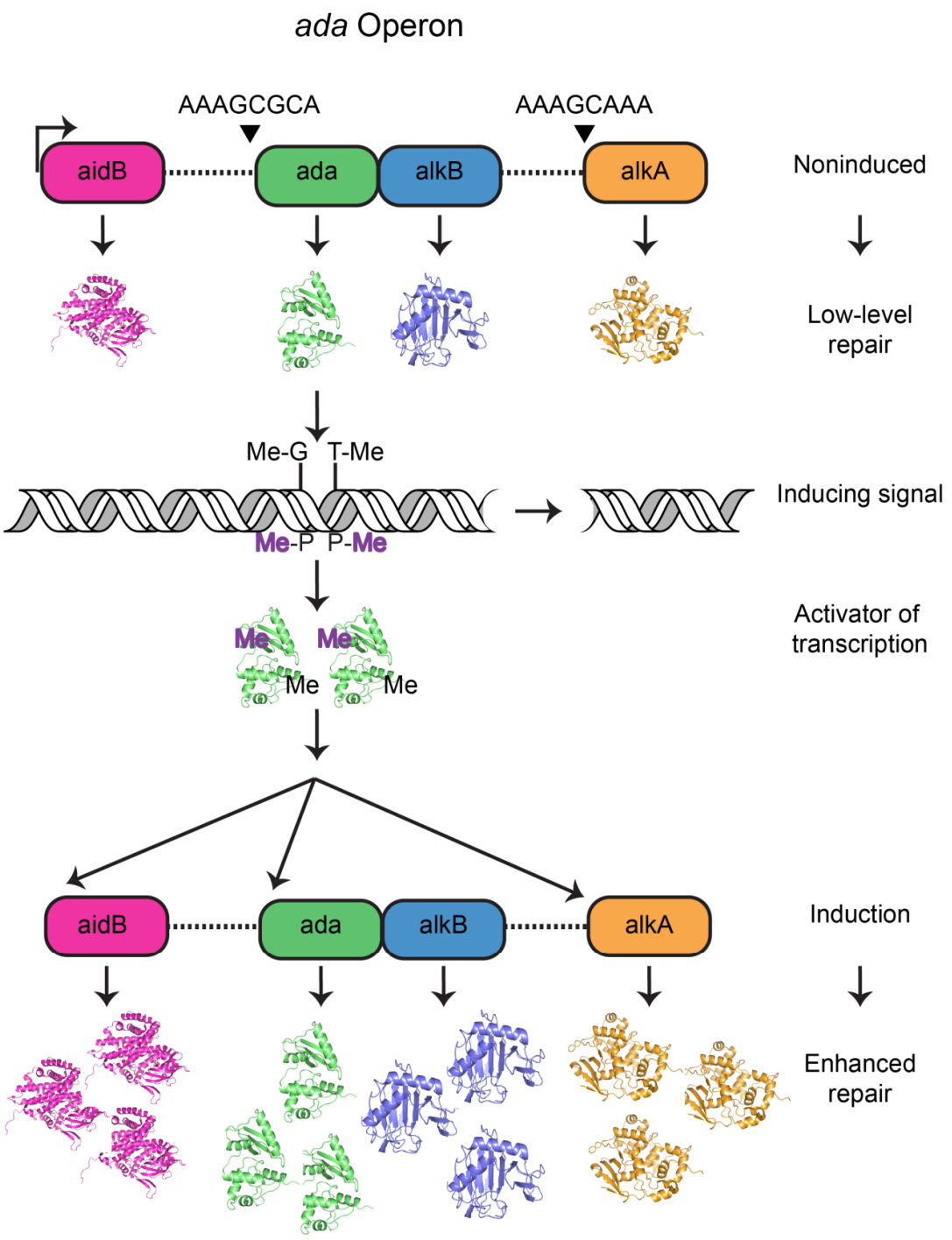


Figure 9. The adaptive response. Upon alkylation damage, the Ada protein transfers methyl groups from alkylated DNA and transcriptionally activates the other genes. Increased transcription of AidB, Ada, AlkB and AlkA increases repair of alkylated DNA and decreases the mutagenicity and toxicity of alkylating agents. The crystal structures of the gene products are rendered as cartoons [AidB (3djl) (Bowles et al, 2008), Ada (1sfe) (Moore et al, 1994), AlkB (3i3q) (Yu & Hunt, 2009), AlkA (1mpg) (Labahn et al, 1996)]. [Adapted from reference (Lindahl et al, 1988)].

Alkylpurine DNA Glycosylases

Despite progress in the field, the mechanisms by which DNA glycosylases select for a particular alkyl modification are not well understood. The importance of substrate specificity is underscored by the fact that these enzymes must locate very subtle modifications among a vast excess of normal base pairs. The diversity in their structural features despite overlapping functions presents an opportunity to understand the physical and chemical determinants of DNA alkylation damage recognition and removal. The alkylpurine DNA glycosylases, which have been characterized in eukaryotes, archaea, and bacteria, are unique because of their ability to excise a range of chemically and structurally diverse alkylpurine substrates, including cytotoxic 3mA, 7mG, and the highly mutagenic εA (Holt et al, 1998; Shuker et al, 1987; Shuker & Farmer, 1992). They include mammalian alkyladenine DNA glycosylase (AAG) (Brent, 1979; Karran et al, 1980), yeast methyladenine DNA glycosylase (*S. cerevisiae* MAG and *S. pombe* MagI) (Berdal et al, 1990; Chen et al, 1990; Memisoglu & Samson, 1996), *E. coli* 3-methyladenine (3mA) DNA glycosylase I (TAG) and II (AlkA) (Riazuddin & Lindahl, 1978; Thomas et al, 1982), *Thermotoga maritima* methylpurine DNA glycosylase II (MpgII) (Begley et al, 1999), *Helicobacter pylori* 3mA DNA glycosylase (MagIII) (O'Rourke et al, 2000), and most recently *Bacillus cereus* AlkC and AlkD (Alseth et al, 2006). TAG and MagIII are highly specific for 3mA and 3mG (Bjelland et al, 1993; O'Rourke et al, 2000), MpgII and AlkC/D are selective for positively charged lesions 3mA and 7mG (Alseth et al, 2006; Begley et al, 1999), and AlkA and AAG can excise these lesions as well as other alkylated and modified bases, including εA and hypoxanthine (Hx) (Bjelland et al, 1994; McCarthy et al, 1984; Saparbaev et al, 1995).

The alkylpurine DNA glycosylases comprise three of the six distinct superfamilies based on their three-dimensional structures (Figures 7 and 10). As described above, AAG is defined by a mixed α/β globular fold (Lau et al, 1998), which bears no structural resemblance to any other protein in the Protein Data Bank (PDB). Almost all of the bacterial enzymes including AlkA, TAG, MagIII, and MpgII (Begley et al, 1999; Drohat et al, 2002; Eichman et al, 2003; Labahn et al, 1996) fall into the helix-hairpin-helix superfamily of glycosylases. *Saccharomyces cerevisiae* MAG and *Schizosaccharomyces pombe* Mag1 likely adopt the HhH fold based on sequence similarity to AlkA (Berdal et al, 1990; Chen et al, 1990). *B. cereus* AlkC and AlkD proteins represent a third alkylpurine DNA glycosylase architecture. AlkD forms a C-shaped α-helical fold from repeating HEAT motifs, and AlkC is expected to adopt a similar fold (Dalhus et al, 2007; Rubinson et al, 2008). These differences highlight that various protein architectures can be used to create a DNA binding platform suitable for nucleobase excision.

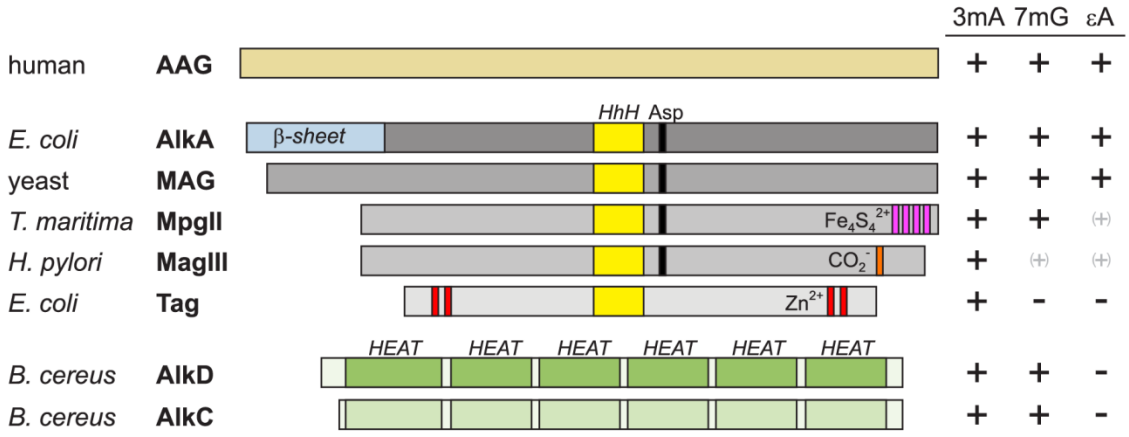


Figure 10. Domain structures of alkylpurine DNA glycosylases. The three families are defined by the structural folds of human AAG (gold), the helix-hairpin-helix (HhH) superfamily typified by AlkA (silver), and HEAT-repeat proteins AlkC and AlkD (green). Alkylpurine DNA glycosylases have a range of substrate specificities.

Structures of these enzymes have provided insight into enzymatic selection and hydrolysis of alkylpurines and despite their diversity, the alkylpurine DNA glycosylases utilize the same general strategy for DNA damage recognition as do other glycosylases, in which amino acids near the active site are able to sense an energetic difference between modified and unmodified base pairs. Nevertheless, the alkylpurine specific active sites seem to have evolved unique mechanisms for excision of relatively unstable, positively-charged bases (e.g., cytotoxic 3mA) in addition to more stable adducts like mutagenic εA. Their active sites consist of a concave pocket lined with aromatic, electron-rich side chains that base-stack with the flipped alkylpurine nucleobase. With the exception of TAG, the alkylpurine DNA glycosylases contain a conserved, catalytically essential aspartic or glutamic acid residue at the mouth of the active site (Figure 11). In all cases, intercalating side chains stabilize the extrahelical base. AlkD also contains an aromatic, electron-rich cleft and a conserved aspartate residue but lacks identifiable intercalating residues typical of the other alkylpurine glycosylases (Figure 11d).

Structural and Biochemical Highlights of Selected Alkylpurine DNA Glycosylases

The AAG/DNA complexes described above have been important for directing biochemical studies aimed at understanding the molecular basis for AAG's substrate specificity and catalytic mechanism. Discrimination against normal purines is most likely due to their proper base pairing in the DNA duplex and to unfavorable interactions with exocyclic N6 and N2 amino groups inside the active site (Figure 11a) (O'Brien & Ellenberger, 2004a). For example, His136 donates a hydrogen bond to N6 of εA,

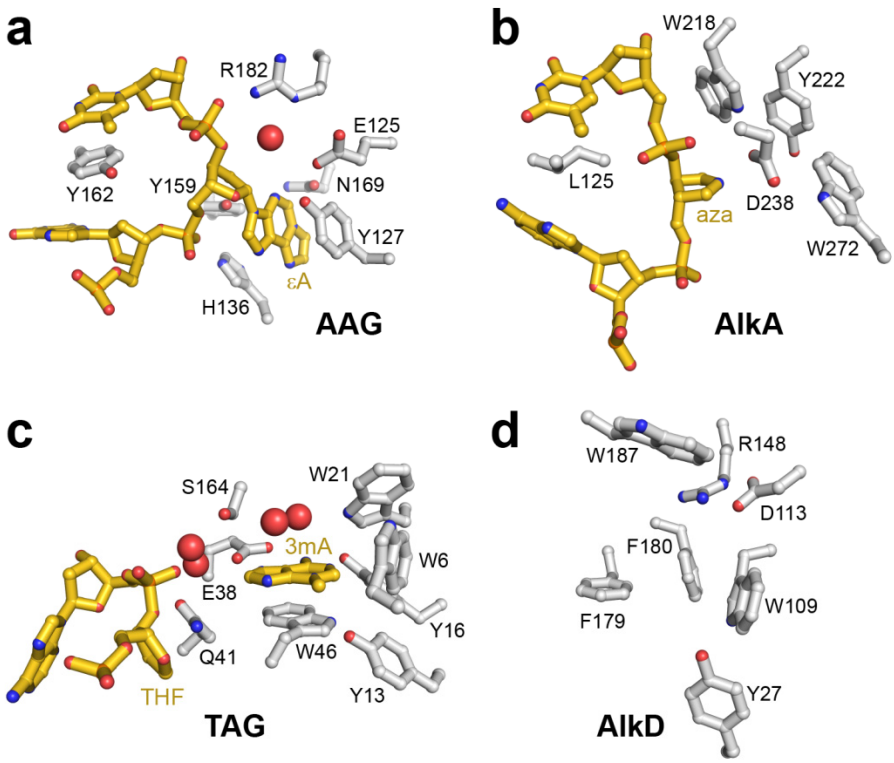


Figure 11. Active sites of alkylpurine DNA glycosylases. Protein and nucleic acid atoms are silver and gold, respectively. **a**, Human AAG in complex with εA-DNA (1ewn) (Lau et al, 2000). **b**, *E. coli* AlkA bound to 1-azaribose-DNA (1diz) (Hollis et al, 2000). **c**, *E. coli* TAG/THF-DNA/3mA complex (2ofi) (Metz et al, 2007). **d**, *B. cereus* AlkD (3bvs) (Rubinson et al, 2008).

whereas adenine cannot accept a hydrogen bond at this position. Furthermore, guanine is likely to be excluded on the basis of a steric clash between its exocyclic N2 amino group, which is absent in εA, Hx, and adenine, and the side chain of Asn169. In support of this, mutation of Asn169 gives AAG enhanced activity toward guanine (Connor & Wyatt, 2002). It has been suggested that AAG removes charged alkylpurine lesions because of their inherent instability and not through a structural recognition of the methyl group *per se*. Indeed, AAG's rate enhancement for excision of 3mA is one and three orders of magnitude less than that of εA and Hx, respectively (O'Brien & Ellenberger, 2004a). Regarding catalysis, an ordered water molecule sits adjacent to the *N*-glycosylic bond and

is hydrogen bonded to the side chains of Glu125 and Arg182, the carbonyl oxygen of Val262, and either the pyrrolidine N4′ or the O3′ of εA (Figure 11a). This arrangement is consistent with Glu125 acting as a general base to deprotonate a water molecule, which may serve as a nucleophile to attack the anomeric C1′ carbon in an S_N2 catalytic mechanism (O'Brien & Ellenberger, 2003).

AlkA lacks the iron-sulfur cluster present in EndoIII and MutY, and instead contains an amino-terminal β-sheet domain that has no identified function but presumably stabilizes the overall fold. Crystal structures of AlkA in complex with DNA containing a 1-azaribose abasic site illuminated how DNA glycosylases utilize the HhH motif to anchor the protein to the DNA (Hollis et al, 2000). Although the HhH does not directly participate in lesion recognition, it contributes most of the polar interactions between AlkA and the DNA. The DNA is highly distorted with a ~60° bend and widened minor groove around the site of the lesion. The 1-azaribose is rotated 180° around the phosphoribose backbone and points into a shallow cleft formed by several aromatic side chains (Figure 11b). Leu125 plugs the gap left by the flipped nucleotide in a manner similar to AAG Tyr162, and the hairpin between helices αG and αH wedge into the DNA strand opposite the lesion (Figures 11b and 12).

The nucleobase binding surface of AlkA is a shallow cleft that can accommodate a variety of alkylpurines. This open architecture helps explain AlkA's broad specificity. In addition, the substrate methylpurine base presumably stacks against the Trp272 indole ring, enhancing the preference of AlkA for positively charged bases. In the AlkA/DNA complex, rotation of the 1-azaribose into the active site places the N1′ nitrogen directly adjacent to the carboxylate group of the catalytic aspartate (Asp238), leaving no room for

a water nucleophile necessary for an S_N2 catalytic mechanism (Figure 11b). This close proximity between the abasic site and Asp238 has led to the suggestion that AlkA utilizes an S_N1-type mechanism, whereby the ionized carboxylate stabilizes the carbocation intermediate formed on the ribose ring during nucleobase hydrolysis (Hollis et al, 2000).

The 3mA-specific TAG enzyme is a divergent member of the HhH superfamily (Drohat et al, 2002). Despite a conserved HhH domain, TAG lacks the conserved catalytic aspartate present in all other HhH glycosylases, and the sequence and structure of the HhH motif itself is noticeably different (Figure 8). In addition, the N/C domain is devoid of any significant α-helical structure but rather contains a novel zinc binding motif that helps "snap" the N- and C-termini together (Kwon et al, 2003). NMR and base perturbation studies revealed that *E. coli* TAG binds 3mA inside a deep pocket that is sterically constrained to exclude 7mG and εA bases (Cao et al, 2003).

The crystal structure of a TAG/DNA/3mA ternary complex provided insight into how TAG achieves its high selectivity for 3mA (Metz et al, 2007). *S. typhimurium* TAG, which shares 82% sequence identity (91% similarity) to the *E. coli* protein, was crystallized in the presence of free 3mA base and DNA containing a tetrahydrofuran (THF) abasic site (Figures 11c and 12). As in the AlkA/DNA complex, the HhH hairpin contributes most of the electrostatic interactions to the DNA backbone immediately 3′ to the lesion. For the first time in a DNA glycosylase structure, however, the THF moiety is not fully flipped into the active site and does not form any polar interactions with the protein. Instead, the abasic site was observed in the electron density map to interconvert between a stacked position normally found in B-DNA and one in which the ribose is partially rotated ~90° into the minor groove. The DNA is bent by ~65° as a consequence

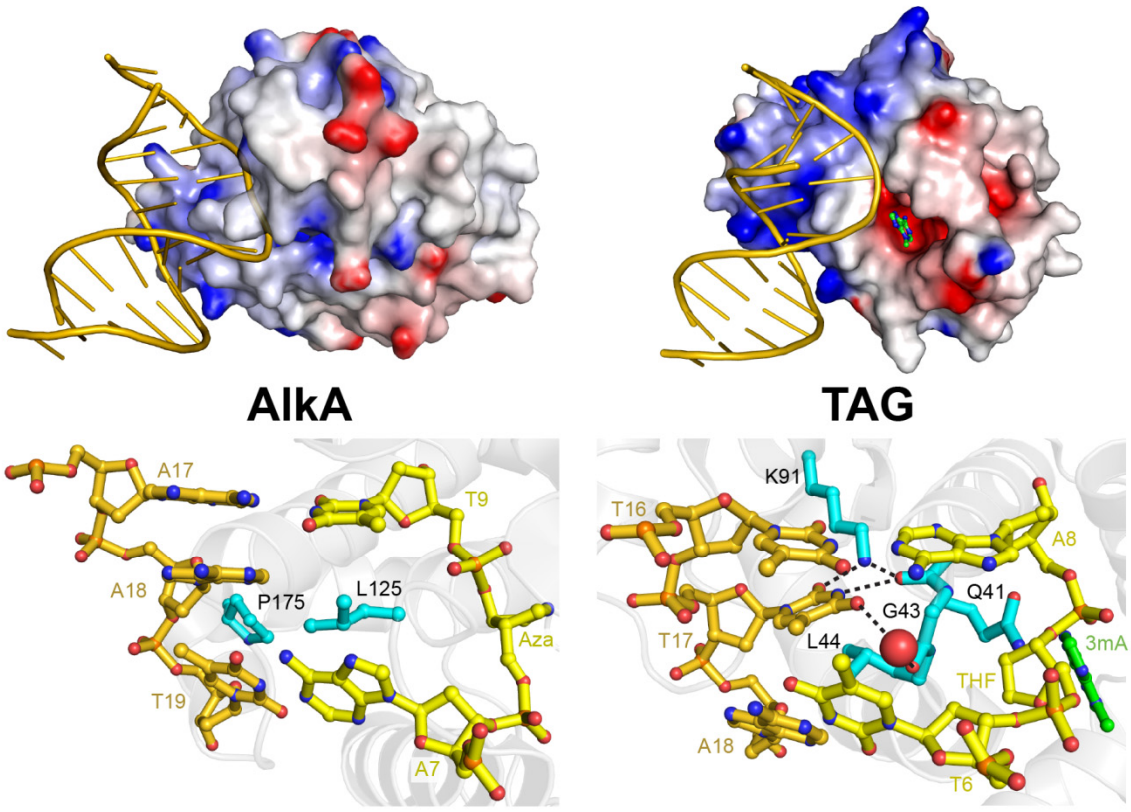


Figure 12. Comparison of TAG and AlkA DNA complexes. The overall structures of AlkA bound to 1-azaribose-containing DNA (left) and of TAG bound to THF-DNA and 3mA (right) are shown at the top, with protein rendered as an electrostatic potential surface (red, negative; blue, positive), DNA as gold cartoon, and 3mA nucleobase ball-and-stick (green carbons). At the bottom is a close-up view of the plug-and-wedge intercalation of the DNA duplexes by the proteins. A bridging water molecule in TAG is depicted as a red sphere.

of the intercalating plug and wedge interactions, which in TAG are provided by a single hairpin loop between helices αD and αE. The main-chain of Gly43 plugs the abasic gap, and the adjacent Leu44 side chain wedges between the bases across from the lesion (Figures 11c and 12). Despite the kink in the DNA, the helix remains essentially B-form as a result of the lack of specific interactions to the abasic site or to the DNA duplex on the 5′ side of the lesion. Thus, the DNA in the TAG product complex is less distorted than in the AlkA transition state complex (Figure 12).

In the TAG/DNA crystal structure, the 3mA base resides 8 Å away from the THF moiety and deep inside the active site pocket. The 3mA ring is stacked between Trp46 and several ordered water molecules and is constrained on the sides by hydrogen bonds to Glu38 and Tyr16 and by van der Waals contacts to Trp6 (Figure 11c). Substitution of Trp46 to alanine reduced the rate of 3mA excision 10-fold with respect to wild-type TAG, further highlighting the importance of base stacking on alkylpurine glycosylase activity. The hydrogen bonds between the Glu38 carboxylate and the N6 amino and N7 imino nitrogens of 3mA suggest that 7mG is sterically excluded from the TAG active site (Cao et al, 2003; Metz et al, 2007). Interestingly, mutation of Glu38 to alanine, which should relax this constraint, did not provide TAG the ability to cleave 7mG from DNA. Rather, reduction of the rate of 3mA excision by two orders of magnitude with respect to the wild-type enzyme, suggests that Glu38 is important for catalysis (Metz et al, 2007). A catalytic mechanism for 3mA excision has been proposed in which TAG provides a high-affinity base binding pocket that induces strain in the pre-catalytic TAG/DNA-3mA ground state complex (Cao et al, 2003). Release of this strain upon base hydrolysis is illustrated in the crystal structure of the TAG/DNA product complex by the large distance between 3mA and the abasic ribose and by the relatively small distortion to the conformation of the DNA helix as compared to the AlkA transition state complex (Figure 12). TAG's high specificity for 3mA, therefore, may be a result of the intrinsic instability of this lesion and the lack of a general acid or base to drive catalysis, rather than a mere steric exclusion of other bases from the active site.

Scope of this Work

The central topic of this thesis concerns alkylation-specific DNA glycosylases, with a specific focus on the recently discovered alkylpurine DNA glycosylase, AlkD. Presented in this work is a structure-function analysis of AlkD, which uses an unprecedented nucleic acid capture mechanism to excise alkylated DNA. The results presented in this dissertation provide further understanding of how alkylpurine DNA glycosylases select for and excise positively charged alkylpurines. Chapter II describes the high-resolution structural determination and biochemical characterization of *Bacillus cereus* AlkD. These studies clearly define the enzyme as the founding member of a sixth structural superclass of DNA glycosylases. Showing how tandem helical repeat architecture of AlkD binds DNA through electrostatic interactions is the subject of Chapter III. The results presented in this chapter detail the structures of AlkD bound to lesioned, abasic, and mismatched DNA and provide a structural basis for lesion recognition and substrate specificity by the enzyme. Chapter III also discusses how AlkD may catalyze the reaction through phosphate-assisted catalysis. Chapter IV contains a comprehensive analysis of DNA binding and substrate catalysis by selected AlkD mutants. Chapter V is an investigation into how tandem helical repeat motifs, usually reserved for protein-protein interactions, bind and process DNA by comparing the structure of AlkD to recent structures of MTERF1, a mitochondrial transcription factor. Chapter VI outlines the recent development of a mass spectrometry method designed to easily characterize substrate specificity of alkylpurine DNA glycosylases including AlkD, AlkC, and TAG. Finally, Chapter VII is a discussion of the implications of this research

and how the overall conclusions have expanded our knowledge of alkylpurine DNA glycosylase catalysis as well as describing future directions for this research.

CHAPTER II

A NEW PROTEIN ARCHITECTURE FOR PROCESSING ALKYLATION DAMAGED DNA: THE CRYSTAL STRUCTURE OF DNA GLYCOSYLASE AlkD*

Summary

DNA glycosylases safeguard the genome by locating and excising chemically modified bases from DNA. AlkD is a recently discovered bacterial DNA glycosylase that removes positively charged methylpurines from DNA, and was predicted to adopt a protein fold distinct from other DNA repair proteins. The crystal structure of *Bacillus cereus* AlkD presented here shows that the protein is composed exclusively of helical HEAT-like repeats, which form a solenoid perfectly shaped to accommodate a DNA duplex on the concave surface. Structural analysis of the variant HEAT repeats in AlkD provides a rationale for how this protein scaffolding motif has been modified to bind DNA. We report 7mG excision and DNA binding activities of AlkD mutants, along with a comparison of alkylpurine DNA glycosylase structures. Together, these data provide important insight into the requirements for alkylation repair within DNA and suggest that AlkD utilizes a novel strategy to manipulate DNA in its search for alkylpurine bases.

Introduction

The integrity of DNA is constantly challenged by chemical attack from endogenous metabolites and environmental agents. Chemical modification of DNA

* The work presented in this chapter was published in 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. *Journal of molecular biology* **381**, 13-23.

nucleobases by alkylation, oxidation, deamination, or hydrolysis produces mutagenic or cytotoxic lesions that can result in heritable disease, cancer, and cell death (reviewed in Friedberg et al, 2006). To safeguard against DNA damage, all organisms have evolved DNA repair mechanisms to eliminate modified bases from the genome. Base excision repair (BER) is the principal pathway by which small modifications to DNA nucleobases are removed in both prokaryotes and eukaryotes. DNA glycosylases initiate the BER pathway by catalyzing the hydrolysis of the C1′-N glycosylic bond to liberate the modified base from the phosphoribose backbone. The resulting abasic site is further processed by AP endonuclease, phosphodiesterase, DNA polymerase, and DNA ligase functions to restore the DNA to an undamaged state.

DNA glycosylases are specific for a particular type of base damage. The mechanism by which these enzymes initially locate their target substrates is believed to proceed by a processive sliding search for destabilized base pairs along DNA (Banerjee et al, 2006; Blainey et al, 2006; Francis & David, 2003; Stivers & Jiang, 2003). DNA glycosylases read out the identity of the lesion by flipping the damaged nucleotide out of the DNA helix and into a nucleobase binding pocket on the protein, the chemical and physical properties of which are complementary to the correct substrate. All glycosylases examined to date utilize a similar strategy for binding DNA and base flipping despite their structural diversity. There are five structural glycosylase superfamilies (reviewed in Fromme et al, 2004), represented by *i*) T4 pyrimidine dimer specific DNA glycosylase EndoV (Morikawa et al, 1992); *ii*) uracil DNA glycosylase (UDG, Mol et al, 1995), *iii*) bacterial 8-oxoguanine DNA glycosylase MutM/FPG (Bruner et al, 2000; Sugahara et al, 2000) ; *iv*) human alkyladenine DNA glycosylase AAG/MNPG (Lau et al, 1998); and *v*)

the helix-hairpin-helix (HhH) superfamily (Doherty et al, 1996; Nash et al, 1996). A broad range of substrate specificities exist within the HhH superfamily, which includes *E. coli* endonuclease III (EndoIII, Kuo et al, 1992), *E. coli* adenine DNA glycosylase (MutY, Guan et al, 1998), human 8-oxoguanine DNA glycosylase (hOGG1, Bruner et al, 2000), and several alkylpurine specific DNA glycosylases (Drohat et al, 2002; Eichman et al, 2003; Labahn et al, 1996).

DNA alkylation damage from endogenous methyl donors, environmental toxins, or from chemotherapeutic agents produces a chemically diverse spectrum of alkylated nucleobases (Hecht, 1999; Hurley, 2002; Rydberg & Lindahl, 1982). DNA glycosylases remove *N*3- and *N*7-substituted alkylpurines, as well as 1,N^6-ethenoadenine (εA). The variation in substrate specificities of alkylpurine DNA glycosylases from different organisms has been useful in understanding the mechanisms of base selection and removal. *E. coli* 3-methyladenine (3mA) DNA glycosylase I (TAG, Bjelland et al, 1993) and *H. pylori* methyladenine DNA glycosylase (MagIII, Begley et al, 1999; O'Rourke et al, 2000) are highly specific for 3mA. *Thermatoga maritima* methylpurine DNA glycosylase II (MpgII, Begley et al, 1999) excises 7-methylguanine (7mG) in addition to 3mA. At the other extreme, *E. coli* 3mA DNA glycosylase II (AlkA), *S. cerevisiae* MAG and human AAG have a broad, yet well-defined specificity toward 3mA, 7mG, εA, hypoxanthine, and various oxidized purines (Bjelland et al, 1994; Bjoras et al, 1995; McCarthy et al, 1984; O'Connor, 1993; Saparbaev et al, 1995; Saparbaev & Laval, 1994; Singer et al, 1992).

Several alkylpurine DNA glycosylases have recently been identified in the genomic sequences of gram-positive bacteria and lower eukaryotes *Entamoeba*

histolytica and *Dictyostelium discoideum* (Alseth et al, 2006; Dalhus et al, 2007). *Bacillus cereus* AlkC and AlkD excise 3mA and 7mG, but not εA, and thus have an intermediate specificity similar to MpgII (Alseth et al, 2006; Begley et al, 1999). These enzymes share no sequence homology with any other glycosylase, and they have been predicted to represent a new structural class of glycosylase enzymes (Dalhus et al, 2007). We present here the crystal structure of *B. cereus* AlkD at 2.1 Å resolution together with mutational analyses of 7mG excision and DNA binding. AlkD is composed of a tandem array of helical repeats reminiscent of HEAT motifs, which are known to facilitate protein-protein interactions and have not yet been associated with DNA binding or catalytic activity. A comprehensive structural analysis of the variant HEAT repeats in AlkD helps to explain how this protein interaction motif has been adapted to bind DNA. Importantly, the complementarity between AlkD's concave surface and the B-DNA backbone in our docking model raises the possibility that the enzyme uses a different strategy to engage DNA and access the damaged base.

Results

The Structure of AlkD

The crystal structure of AlkD from *B. cereus* was determined using experimental phases from a single-wavelength anomalous dispersion (SAD) experiment from a crystal derivatized with platinum tetrachloroplatinate(II) (Table 1). A crystallographic model that consists of one AlkD molecule (residues 1-225) in the asymmetric unit was built into 2.0-Å Pt-SAD electron density (Figure 13a). To avoid the possibility of structural artefacts resulting from platinum binding to the protein, the original SAD model was

refined against native diffraction data extending to 2.1 Å resolution, resulting in a crystallographic residual (R_{cryst}) of 18.9% and an R_{free} value of 22.9% (Table 1).

AlkD adopts a single α-helical domain. Twelve of the fourteen helices (αA-αN) pair in an antiparallel fashion to form 6 tandemly repeated α-α motifs (αA/αC, αD/αE, αF/αG, αH/αI, αJ/αK, and αL/αM) (Figures 13b and 15). The helical repeats are stacked into a superhelical solenoid in which helices B, C, E, G, I, K and M form a concave surface with an aromatic cleft at its center (Figure 13b). Residues within this cleft were shown previously to impair base excision activity of AlkD (Dalhus et al, 2007). The concave surface of the protein is strikingly electropositive (Figure 13c), a feature that is distinct from other helical repeat proteins and one that likely facilitates binding to DNA (discussed below). A structural homology search of the PDB using the DALI server (Holm & Sander, 1993) indicated that AlkD is most similar in tertiary structure to *Enterococcus faecalis* EF3068 (PDB ID 2B6C) and *B. cereus* BC3264 (PDB ID 1T06), two hypothetical proteins of unknown function determined from the Midwest Center for Structural Genomics (Figure A1, Table A1). EF3068 shares 35% sequence identity with AlkD (Figure A2) and was used in a previous study as the basis for a homology model of the *Bacillus* enzyme (Dalhus et al, 2007). The structures of AlkD and EF3068 are similar in overall fold (r.m.s.d. of 1.38 Å for all backbone atoms), with the most notable differences in helices αA/αC at the N-terminus (Figure A1).

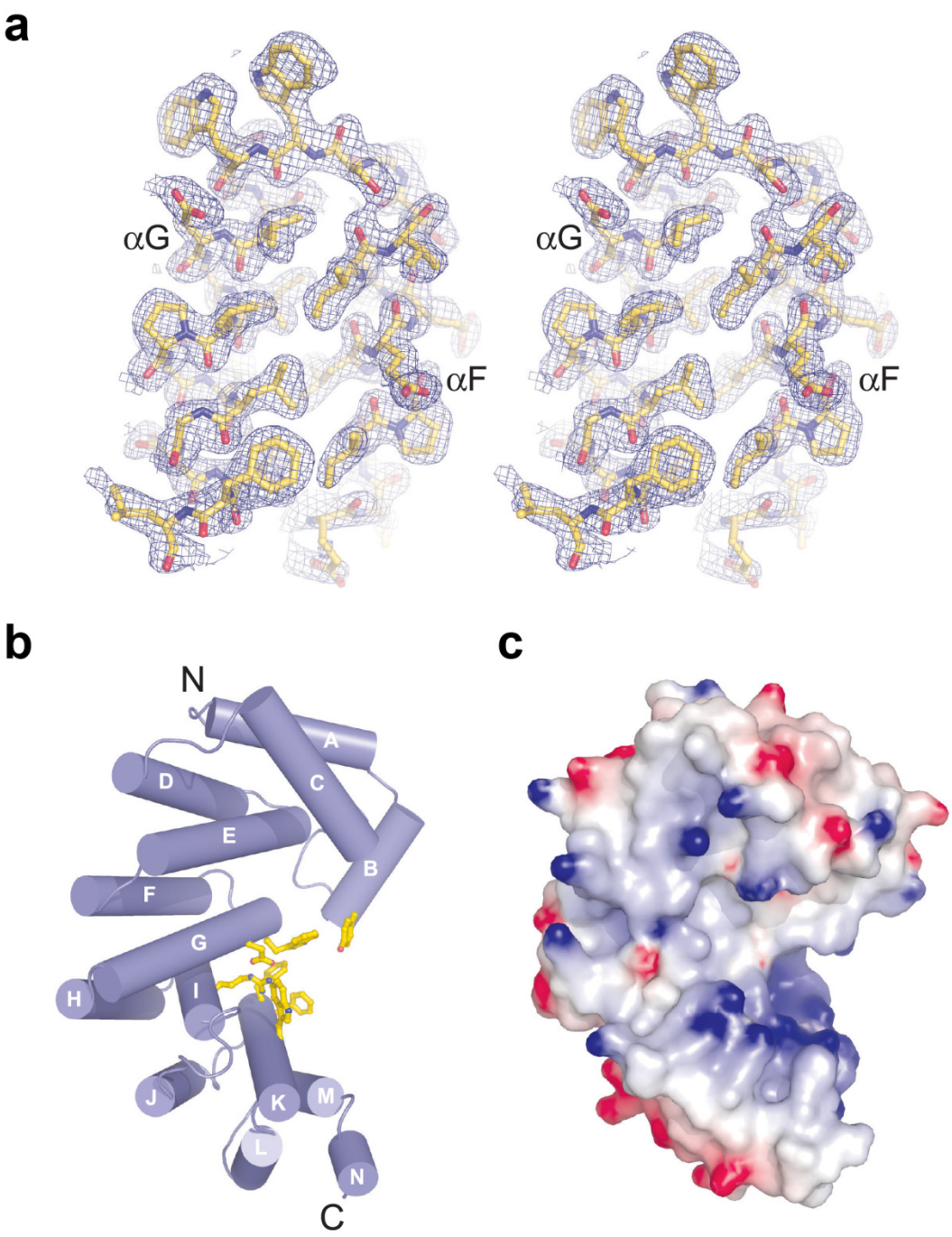


Figure 13. Structure of *B. cereus* AlkD. **a**, Stereo-view of helices αG and αF from the refined crystallographic model superimposed on a 2.0-Å Pt-SAD experimental electron density map (contoured at 1σ). **b**, AlkD forms a solenoid structure with a left-handed superhelical twist. Residues forming an aromatic cleft on the concave surface are shown as yellow sticks. **c**, Solvent accessible surface representation of AlkD in the same orientation as panel B, colored according to electrostatic potential (red negative, blue positive, -7 to +7 k_BT). Potentials in Figures 1 and 5 were calculated with the program DelPhi (Rocchia et al, 2002). All molecular images were rendered using PyMOL (http://www.pymol.org).

Table 1. Data collection, phasing and refinement statistics

	Native	K_2PtCl_4
Data collection		
Space group	$P4_3$	$P4_3$
Cell dimensions		
a, *b*, *c* (Å)	77.9 77.9 55.1	78.8 78.8 55.3
α, β, γ (°)	90.0 90.0 90.0	90.0 90.0 90.0
Wavelength	1.0000	1.0718
Resolution (Å)	50-2.08 (2.15-2.08)	50-1.85 (1.92-1.85)
R_{sym} or R_{merge}	0.046 (0.531)	0.052 (0.321)
$I / \sigma I$	26.9 (3.4)	50.1 (3.1)
Completeness (%)	99.1 (95.3)	97.4 (84.3)
Redundancy	7.1 (6.2)	7.2 (4.8)
Refinement		
Resolution (Å)	50-2.10 (2.21-2.10)	
No. reflections	18356 (2605)	
R_{work}	0.188 (0.283)	
R_{free}	0.227 (0.313)	
No. atoms		
Protein	1893	
Ion / Ligand	0	
Water	81	
B-factors		
Protein	52.7	
Water	55.2	
R.m.s. deviations		
Bond lengths (Å)	0.018	
Bond angles (°)	1.571	

Data in parentheses refer to the highest resolution shells

Putative Active Site

One striking feature of the AlkD structure is the apparent similarity in the chemical environment of the aromatic cleft to the nucleobase binding pockets of other alkylpurine DNA glycosylases. AAG, AlkA, TAG, and MagIII all feature electron-rich, aromatic pockets that accommodate the extrahelical alkylpurine base through both shape complementarity and π-π stacking interactions (Eichman et al, 2003; Hollis et al, 2000;

Hollis et al, 2001; Lau et al, 2000; Metz et al, 2007). With the exception of TAG, which exclusively excises 3mA and 3mG, the alkylpurine binding pockets also feature a conserved acidic residue that is essential for base excision activity.

AlkD contains a cluster of aromatic residues (Trp109, Trp187, Phe179, Phe180, and Tyr27) that form a shallow cleft at the center of the concave surface (Figure 14a). At the back of this cleft, Asp113 and Arg148 form an electrostatic interaction and were observed in the crystal structure to hydrogen bond to several well-ordered water molecules (Figure 14a). The relative positions of these side chains are similar to the aromatic and catalytic Asp238 residues inside the shallow nucleobase binding pocket of AlkA (Figure 14b) (Hollis et al, 2000; Labahn et al, 1996). Asp113, Arg148 and Trp109 are invariant in sequence among AlkD homologs (Figure A2) and were identified in a previous study to be important for release of methylated bases from *N*-methyl *N*-nitrosourea (MNU)-treated genomic DNA and for growth of *E. coli tag alkA* mutant cells in the presence of methylmethane sulfonate (MMS) (Dalhus et al, 2007). Because wild-type AlkD excises several alkylated nucleobases produced from MNU or MMS treatment (Alseth et al, 2006), we tested the effect of AlkD mutants on the specific excision of 7mG from a defined oligonucleotide substrate. As shown in Figures 14c and 14d, Asp113Asn and Arg148Ala each resulted in a dramatic (100-fold) decrease in the single-turnover rate of 7mG excision relative to the wild-type enzyme (Table 2), indicating that Asp113 and Arg148 play a specific role in 7mG excision by AlkD. These results show that AlkD provides catalytic assistance to liberate 7mG from DNA, and implicate the aromatic cleft

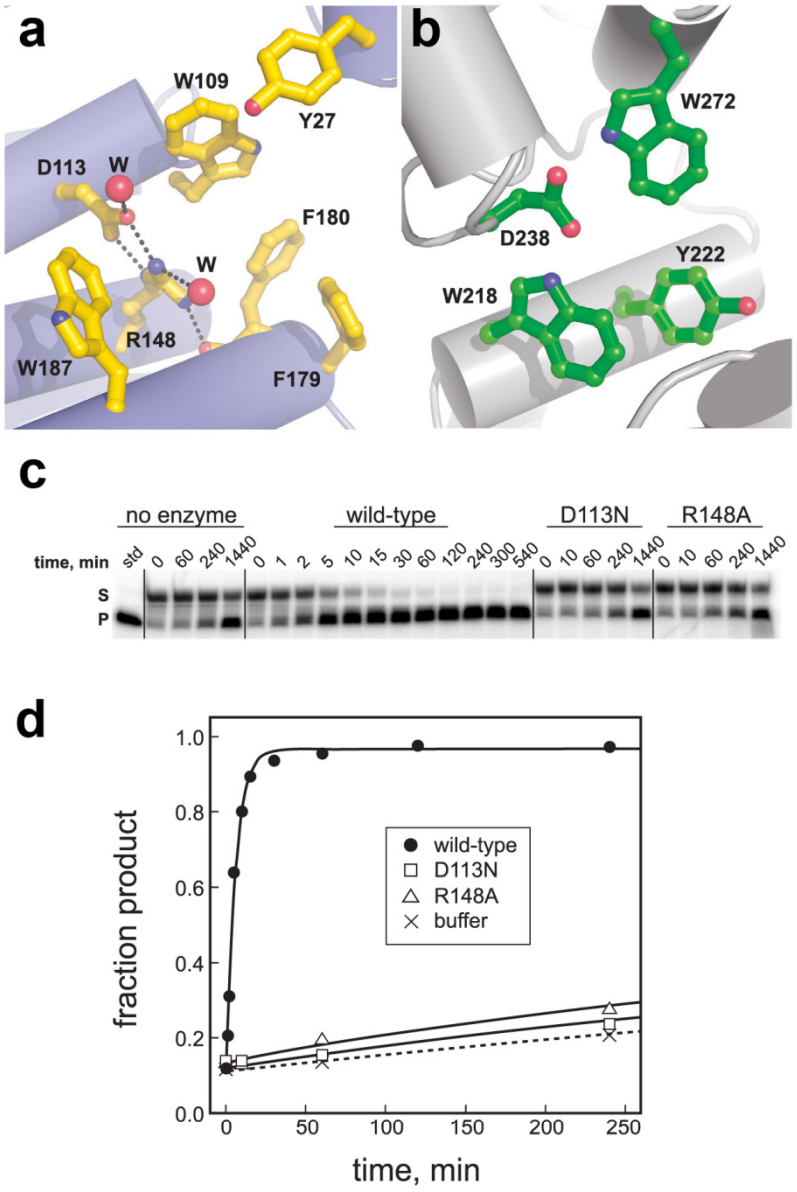


Figure 14. The putative active site of AlkD. **a**, Close-up of the concave cleft shows aromatic and charged residues implicated in nucleobase and DNA binding. Water molecules are shown as red spheres and hydrogen bonds are depicted as black dotted lines. **b**, The electron-rich active site of AlkA, showing the conserved Asp238 essential for base excision. **c**, 7-methylguanine excision by AlkD. Denaturing polyacrylamide gel showing the disappearance of 7mG-DNA substrate (S) and appearance of alkaline-cleaved abasic-DNA product (P) as a result of reaction with no enzyme, wild-type AlkD, Asp113Asn, and Arg148Ala mutants for increasing amounts of time. **d**, Quantitation of the data in panel C showing the inactivity of the Asp113Asn (squares) and Arg148Ala (triangles) mutants as compared to wild-type AlkD (closed circles). Non-enzymatic 7mG hydrolysis is shown as crosses and dotted line curve fit. Rate constants were determined from single-exponential fits.

as the active site of the enzyme.

Table 2. 7-Methylguanine excision activities for wild-type and mutants of AlkD

	k_{cat} (10^{-3} min^{-1}) [a]	Relative activity	Rate enhancement [b]
WT	171.8 ± 11.1	(1.0)	231.6
D113N	1.8 ± 0.5	0.01	2.5
R148A	1.8 ± 0.3	0.01	2.5

[a] First-order single-turnover rate constants for 7mG excision from a 25mer oligonucleotide duplex containing a 7mG·C base pair. Values represent the averages and standard deviations from three experiments.
[b] Rate enhancements are based on the fold increase above a non-enzymatic control (k_{non} = 0.7 x 10^{-3} min^{-1}).

The Variant HEAT Motif

The list of structurally homologous proteins from the DALI analysis was exclusive to proteins that contain tandem helical repeats (Table A1). Apart from EF3068 and BC3264, AlkD is most similar in structure to the PR65/A subunit of protein phosphatase 2A (PP2A) (Magnusdottir et al; Xu et al, 2006) and to the SCF ubiquitin ligase regulatory subunit Cand1 (Goldenberg et al, 2004), both of which contain Huntington/Elongation/A subunit/Target of rapamycin (HEAT) repeats (Andrade et al, 2001b). HEAT motifs are ~45-amino-acid sequences arranged into two antiparallel α-helices, which are packed together by a conserved hydrophobic interface and are tandemly repeated to form superhelical α-structures (Andrade et al, 2001b). The α-α pairs stack in a parallel arrangement to form a solenoid in which the C-terminal helix of each repeat lines the inside groove of the superhelix. The solenoid structure is typically formed from no less than 14 HEAT motifs and provides a molecular scaffold that facilitates protein-protein interactions. Out of 393 structurally similar proteins identified from the PDB search, only three were eukaryotic nucleic acid-binding proteins, including transcription factor Rcd-1 (Z-score = 5.9) (Garces et al, 2007) and the RNA-binding proteins Pumilio1 (Z-score = 4.8) (Wang et al, 2001) and Ro autoantigen (Z-score = 3.5)

(Stein et al, 2005). Ro contains HEAT repeats that are quite divergent from the canonical motif, while Rcd-1 and Pumilio-1 are comprised entirely of armadillo and pumilio repeats, respectively (Andrade et al, 2001b; Macdonald, 1992). Thus, AlkD represents the first example of a DNA-binding HEAT protein.

In contrast to the regular, repeating α-α structures found in archetypical HEAT proteins, the AlkD repeats exhibit greater structural variation despite the conservation of hydrophobic residues that stabilize the interface between helices (Figure 15). The most obvious outlier is αA/αC, which has the opposite handedness from the other repeats, as well as an inserted helix αB that contributes considerable surface area to the concave face of the protein (Figure 15a). The r.m.s.d. among the remaining five repeats αD/αE-αL/αM in AlkD is 2.48 Å, as compared with 1.76 Å among five consecutive HEAT repeats in PP2A (Groves et al, 1999). AlkD's helices are on average one turn shorter in length and, with the exception of αF/αG, do not contain the intrahelical kink characteristic of HEAT repeats (Figure 15b). Thus, the helical repeats that comprise AlkD represent a structural variant of the canonical HEAT motif and thus give rise to a unique architecture.

The most striking feature of AlkD's variant HEAT repeats is the abundance of basic residues along the C-terminal helix (Figure 15c). The positions and spacing of these basic residues are conserved among the repeats, and consequently, the concave surface of the protein is highly electropositive (Figure 13c). This positive charge along the concave surface is conserved among AlkD orthologs (Figures A1, A2), but is not observed in other helical repeat proteins (Figure A3), which faithfully utilize this surface to bind polypeptides (Conti & Kuriyan, 2000; Goldenberg et al, 2004; Matsuura & Stewart, 2004; Xu et al, 2006). Thus, it seems that AlkD's variant HEAT repeats have

evolved this additional electrostatic feature in order to stabilize the negatively charged DNA backbone at the molecular binding interface.

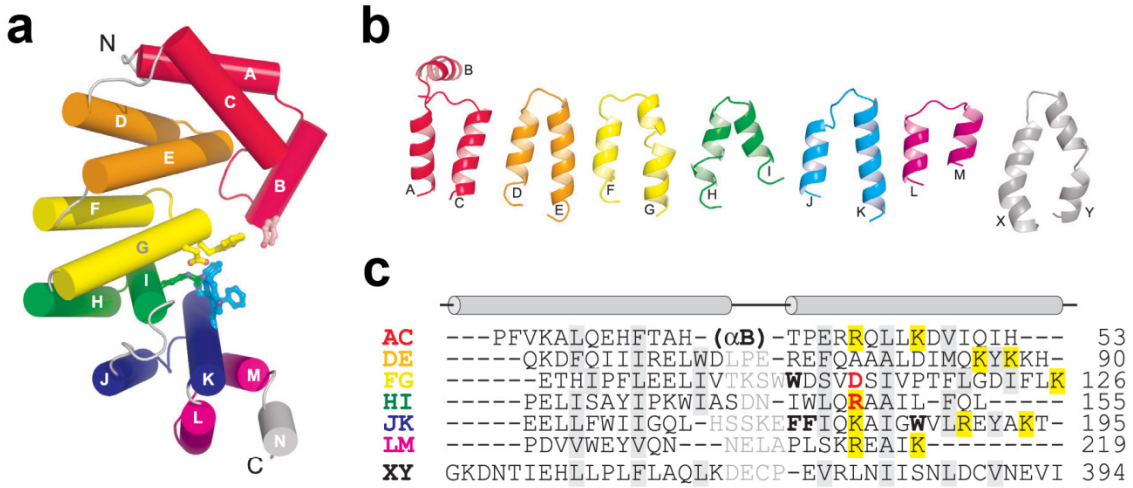


Figure 15. AlkD is composed of six variant HEAT motifs. **a**, Cylinder representation of AlkD with HEAT-like tandem repeats colored independently. Residues lining the putative active site are shown as ball-and-stick. **b**, Structures of individual AlkD α-α pairs (AC, DE, FG, HI, JK, LM) are compared to canonical HEAT repeat 10 (XY) from protein phosphatase 2A PR65/A subunit (1b3u) (Groves et al, 1999). **c**, Structure-based sequence alignment of HEAT motifs in panel B, with AlkD residues lining the aromatic cleft in boldface and Asp113/Arg148 colored red. Positively charged residues contributing to the electropositive concave surface are highlighted yellow. Interdigitating residues important for the relative orientation of paired α-helices are shaded grey.

A DNA Binding Model

Despite their structural divergence, the five existing superfamilies of DNA glycosylases (Fromme et al, 2004; Huffman et al, 2005) use a common mode of binding DNA and gaining access to the lesion via a base flipping mechanism (Figure 16). Structures of these enzymes in complex with DNA show that they all use a positively charged surface to help anchor the DNA against the enzyme. The DNA duplex is highly kinked to allow for extrusion of the damaged base into a sterically constrained active site pocket that lies adjacent to the positive surface (Figure 16). The distorted conformation

of the DNA is stabilized by a pair of side chains that intercalate into the DNA helix at the site of the lesion.

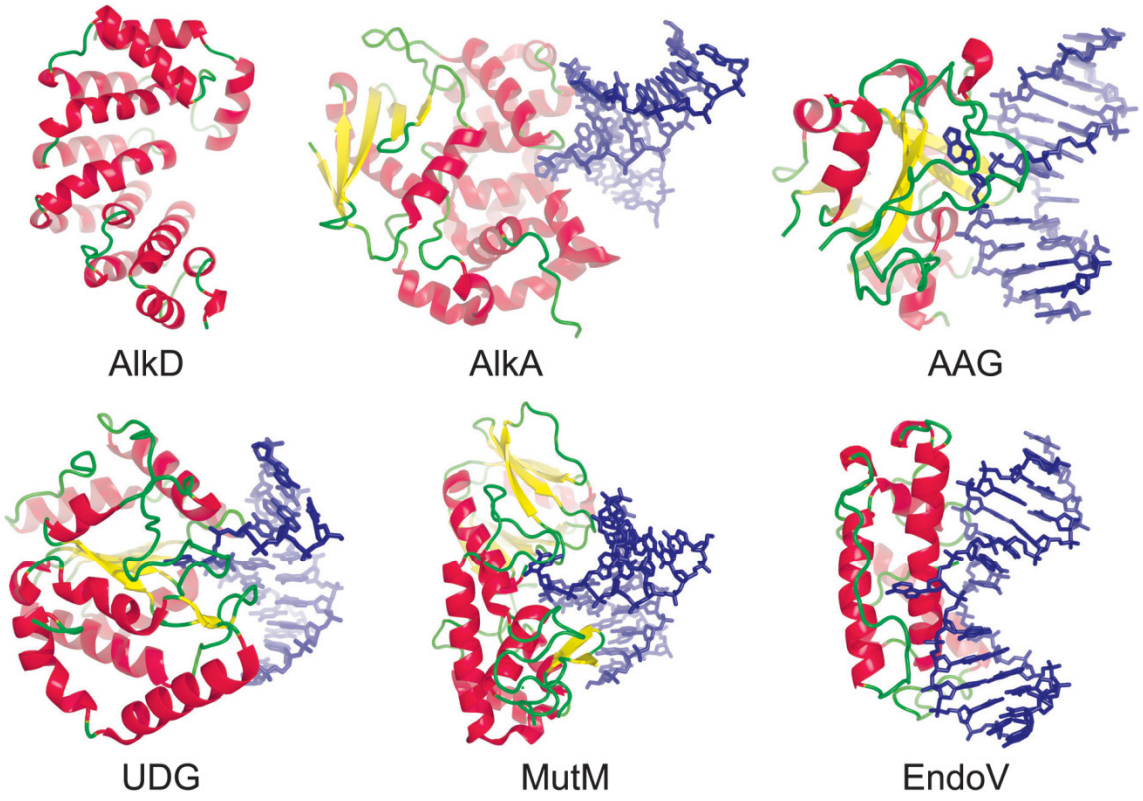


Figure 16. The six structural superfamilies of DNA glycosylases. Representative crystal structures shown are *B. cereus* AlkD; *E. coli* 3-methyladenine DNA glycosylase II, AlkA (1diz) (Hollis et al, 2000); human methylpurine DNA glycosylase, AAG (1ewn) (Lau et al, 2000); human uracil-DNA glycosylase, UDG (1emh) (Parikh et al, 2000); *Bacillus stearothermophilus* 8-oxoguanine DNA glycosylase, MutM (1lit) (Fromme & Verdine, 2002); and T4 pyrimidine dimer DNA glycosylase, EndoV (1vas) (Vassylyev et al, 1995). Proteins are colored according to secondary structure, and lesion-containing DNA is shown as blue sticks.

The solenoid structure of AlkD is distinctly different from the other DNA glycosylases (Figure 16). AlkD is the only glycosylase to form a linear, cylindrical groove along the entire width of the protein. Consequently, the environment around the putative active site is much less constrained than other DNA glycosylase architectures. To investigate how such a unique glycosylase architecture might engage DNA, we

constructed a model of DNA docked onto the structure of AlkD (Figure 17). The lack of structural similarity between AlkD and the other glycosylases precluded us from superimposing previous glycosylase-DNA complexes onto AlkD. Instead, we took advantage of the fact that the dimensions across the concave opening are ~24 Å, which can easily accommodate a B-DNA duplex. Canonical B-DNA (PDB ID 1BNA) was manually docked onto the concave surface with the minor groove of the DNA positioned toward the putative active site, consistent with structures of other glycosylase/DNA complexes. Importantly, the linear DNA model formed favorable van der Waals and electrostatic interactions along the *entire* groove of the protein with no steric collisions (Figure 17).

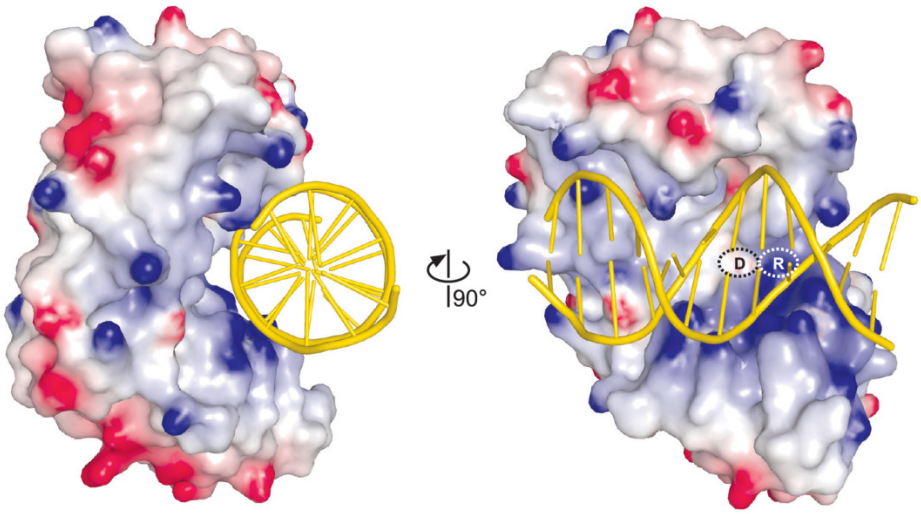


Figure 17. Theoretical model of DNA bound to AlkD. B-DNA from PDB entry 1bna manually docked onto the AlkD crystal structure shows that the DNA need not adopt a distorted conformation to access the concave surface and putative active site of the protein. The molecular surface of the protein is colored according to electrostatic potential (red negative, blue positive, -7 to +7 k_BT). Asp113 and Arg148 are labeled D and R, respectively.

The DNA docking model was used as a guide for mutational and chemical modification experiments designed to probe the nature of DNA binding to AlkD. The

DNA backbone in our docking model was in close proximity to the catalytically important Asp113 and Arg148 residues (Figure 17). The effects of these residues on DNA binding were quantified using an *in vitro* fluorescence anisotropy assay (Table 3, Figure A4). Substitution of Asp113 with asparagine increased the affinity of AlkD for DNA containing either 7mG substrate or a tetrahydrofuran (THF) abasic product analog (Table 3).

Table 3. DNA binding activities for wild-type and mutants of AlkD

	Wild-type		D113N		R148A	
	K_d (μM)	Relative affinity	K_d (μM)	Relative affinity	K_d (μM)	Relative affinity
G	2.0 ± 0.5		n.d.		n.d.	
7mG	1.8 ± 0.5	(1.0)	0.5 ± 0.2	3.5	3.8 ± 0.6	0.5
THF	3.1 ± 0.3	(1.0)	1.3 ± 0.3	2.4	6.2 ± 1.3	0.5
Pyr	3.4 ± 0.1		n.d.		n.d.	
Aza	1.8 ± 0.3		n.d.		n.d.	

Dissociation constants (K_d) for a 25mer oligonucleotide duplex containing the specified modification paired with cytosine were measured by fluorescence anisotropy as described in Experimental Procedures. n.d., not determined

Likewise, an Arg148Ala mutation resulted in a modest but significant decrease in DNA binding. These results are consistent with an electrostatic interaction between Asp113-Arg148 and the DNA backbone. To further investigate this interaction, we measured binding of wild-type AlkD to positively-charged 1-azaribose (C1'→N) and pyrrolidine (O4'→N) abasic sites, which are potent inhibitors of many DNA glycosylases and were designed to mimic the positive charge of the proposed transition state for hydrolysis of the glycosidic bond (Hollis et al, 2000; Makino & Ichikawa, 1998; Scharer et al, 1998; Scharer et al, 1995). The affinity of AlkA for pyrrolidine-DNA is three orders of magnitude greater than for THF-DNA, and AlkA binds to both pyrrolidine and 1-azaribose ~15-fold tighter than to 7mG-DNA (O'Brien & Ellenberger, 2004b). It was therefore surprising that wild-type AlkD bound with the same affinity to all DNAs tested

(Table 3), suggesting that AlkD engages DNA in the active site differently than AlkA. Taken together, the docking model and mutational data suggest that DNA need not obtain a distorted, extrahelical conformation in order to bind to the protein active site surface.

Discussion

The crystal structure of *B. cereus* alkylpurine DNA glycosylase AlkD reveals variant HEAT-like motifs that form an electropositive concave surface necessary to engage DNA. At the heart of the concave surface lies a shallow cleft of aromatic and charged residues, which we verified to be important for binding the DNA backbone and for 7mG excision. These catalytically essential residues and the unique DNA binding HEAT repeat architecture were previously predicted from an AlkD homology model (Dalhus et al, 2007), which was constructed from an unpublished structure (EF3068) determined by the Midwest Structural Genomics Center as part of the Protein Structure Initiative (Osipiuk et al). In the present study, we extend these results by a structural and mutagenic analysis of nucleic acid binding by the variant HEAT motifs in AlkD.

The AlkD solenoid is a unique DNA glycosylase fold, and therefore represents a sixth structural superfamily for these DNA repair enzymes (Figure 16). Perhaps the most unique feature of the AlkD glycosylase is its concave surface, which forms a groove perfectly complementary in shape and charge to a linear B-DNA duplex. Our DNA docking model shows an excellent fit between AlkD and an undistorted B-DNA molecule. As a comparison, we superimposed the kinked DNA from the AlkA/DNA complex (Hollis et al, 2000) onto our structure, as has been done previously for the theoretical homology model (Dalhus et al, 2007). The protein surface occupied by the

highly bent DNA was not as extensive as the undistorted B-DNA contact surface, suggesting that AlkD might bind DNA differently than other glycosylases. In order to stabilize extrahelical nucleobases and a large helical distortion in the DNA duplex, DNA glycosylases typically intercalate a pair of side chains into the DNA base stack. Interestingly, neither linear nor kinked DNA-AlkD models revealed any obvious candidate side chains that might penetrate the stacked bases in order to stabilize a flipped nucleotide. We cannot exclude the possibility that AlkD might adjust its conformation in order to engage a flipped-out nucleotide, although such a DNA-induced protein conformational change has not been observed in other glycosylases. Nevertheless, the open architecture and the apparent lack of DNA intercalating residues strongly suggests that AlkD interrogates damaged DNA differently than other glycosylases. Our model illustrates how AlkD might utilize extensive contacts with the minor groove of both DNA strands in order to sense a subtle distortion in 3mA·T and 7mG·C base pairs. This binding regime, although speculative, illustrates that this novel glycosylase architecture is well suited to scan DNA in search for a lesion without distorting the conformation of the double helix.

The distinct substrate specificity of AlkD for only positively charged alkylpurines might be explained by the Asp113-Arg148 salt bridge within the putative active site. Because it participates in an electrostatic interaction, Asp113 is not analogous to the conserved aspartate or glutamate residues found in other glycosylases (e.g., Asp238 in AlkA). The crystal structure of AlkA in complex with DNA containing 1-azaribose revealed a direct interaction between the extrahelical positively charged abasic site and Asp238 (Hollis et al, 2000). The observation that neither 1-azaribose nor pyrrolidine

enhances DNA binding by AlkD (Table 3) suggests that AlkA and AlkD use different mechanisms to catalyze nucleobase excision. Acidic residues within glycosylase active sites have been suggested to catalyze base hydrolysis by activating a water or protein nucleophile for attack of the C1' anomeric carbon (Nash et al, 1996; Thayer et al, 1995), or by directly stabilizing a carbocation intermediate formed during cleavage of the glycosylic bond (Hollis et al, 2000). However, the formal positive charges of 3mA and 7mG nucleotides make these bases favorable to hydrolysis even in the absence of direct activation by a catalytic residue. Mutation of Arg148 reduced activity by two orders of magnitude, suggesting that Asp113 alone is not sufficient for catalysis, or that the salt bridge stabilizes the local conformation of the active site. Indeed, our DNA docking model predicts that the precise orientations of the Asp113 and Arg148 side chains with respect to the lesion are critical. In this model, the phosphoribose backbones of both DNA strands are abutted against the Asp113-Arg148 pair, leaving no room for a flipped nucleotide (Figure 17). This supports the scanning mechanism described above, and raises the possibility that this aromatic cleft is not an extrahelical nucleobase binding pocket. In support of this, binding of free alkylpurine bases to AlkD could not be detected in solution by intrinsic tryptophan fluorescence quenching or 7mG excision inhibition assays, nor did we observe electron density corresponding to 3mA or 7mG bases that were soaked into AlkD crystals at 10-fold molar excess of free base (data not shown). Based on the instability of the 3mA and 7mG bases and the proximity of Asp113-Arg148 to the DNA backbone in our model, AlkD should be able to gain access to the glycosylic bond for catalysis without flipping the base into the aromatic cleft.

AlkD is the first structural example of a HEAT protein that contains catalytic activity. Based on the adaptability of the HEAT motif to bind protein and DNA, other examples are likely. Of note is deoxyhypusine hydroxylase (DOHH), an iron(II)-containing enzyme which catalyzes the final step in post-translational hypusinylation of eukaryotic initiation factor 5A (eIF5A). A theoretical model of DOHH predicts the protein to form a pair of HEAT domains similar in structure to AlkD (Park et al, 2006). Mutation of glutamate residues lining the concave surfaces of the protein impairs binding to the positively-charged deoxyhypusine-eIF5A substrate (Park et al, 1998). The apparent similarity between catalytic residues within AlkD and DOHH suggest that HEAT proteins might have evolved a common feature to couple substrate binding and catalysis.

Materials and Methods

AlkD Purification and Crystallization

The AlkD gene was PCR amplified from *Bacillus cereus* genomic DNA (ATCC 14579) and cloned into a pET27 (Novagen) derived expression vector (pBG103, Vanderbilt Center for Structural Biology) that produces a cleavable N-terminal His_6-SUMO-fusion protein. *E. coli* HMS174 cells transformed with the AlkD/pBG103 plasmid were propagated in LB media and protein was overexpressed for 3 h at 37°C upon addition of 0.5 mM IPTG. Cells were harvested in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 10% glycerol and lysed with an Avestin Emulsifier C3 homogenizer operating at ~20000 psi. AlkD-fusion protein was purified using Ni-NTA (Qiagen) affinity chromatography, followed by cleavage of the His_6-SUMO tag. AlkD was

further purified by heparin affinity and gel filtration chromatography to >99% homogeneity. Protein was concentrated to 12.5 mg/ml and stored in 20 mM Bis-Tris Propane, 100 mM NaCl, and 0.1 mM EDTA. Mutant proteins were prepared by site-directed mutagenesis using a Quik-Change Kit (Stratagene) and purified in the same manner as wild-type AlkD. Structural integrity of mutant proteins was verified by circular dichroism spectroscopy.

Crystals of unliganded AlkD were grown at 21°C by sitting-drop vapor diffusion. Drops containing equal volumes of protein (14.5 mg/ml) and reservoir (85 mM HEPES pH 7.5, 15% PEG 4000, 17% glycerol) were equilibrated against the reservoir. Single crystals with dimensions 0.2 x 0.2 x 2.0 mm^3 grew in 4 days and were flash frozen in a liquid nitrogen stream prior to X-ray data collection. Derivative crystals were prepared by soaking AlkD crystals in 1 mM K_2PtCl_4 for 72 hrs at 21°C.

X-ray Data Collection, Phasing, and Structure Refinement

X-ray diffraction data (Table 1) were collected at beamlines 21-ID (native) and 22-ID (derivative) at the Advanced Photon Source (Argonne, IL) and processed using the HKL2000 package (Otwinowski & Minor, 1997). AlkD crystals belong to space group $P4_3$ and contain one molecule in the asymmetric unit.

Experimental X-ray phases were obtained from a single-wavelength anomalous dispersion (SAD) experiment using a crystal soaked with K_2PtCl_4. Diffraction data extending to 1.85 Å (Table 1) were collected at the energy corresponding to the platinum absorption peak (11.567 keV). Significant anomalous differences between $|F^+|$ and $|F^-|$ amplitudes extended to 2 Å. The program SHARP (Vonrhein et al, 2007) was used to

locate and refine the positions of two platinum sites and to carry out phase calculation and density modification. A continuous protein chain corresponding to residues 1-225 was built into the resulting 2.0-Å electron density map. Two non-native N-terminal residues (Val -1, Pro 0) resulting from cleavage of the His_6-SUMO tag were readily identified in the experimental electron density and were included in the final model, while the C-terminal 12 residues (226-237) were unobserved.

The model was refined against both native and derivative amplitudes using experimental Pt-SAD phases and a maximum likelihood target as implemented in REFMAC 5.4 (Murshudov et al, 1997). Improvements to the model were guided by manual inspection of sigma-weighted 2mFo-Dfc and mFo-DFc electron density maps, and were judged successful by a decrease in R_{free} during refinement. Translation/libration/screw-rotation (TLS) refinement in REFMAC was used to model anisotropic motion of the protein. Refinement of the Pt-derivative model resulted in R_{cryst} and R_{free} values of 20.3% and 21.4%, which are higher than expected at 1.85 Å resolution. Efforts to resolve regions in the model corresponding to a partially chlorinated Pt(II) ion and a PEG 4000 molecule inside the putative active site and a second Pt(II) ion bound to Met 1 using both experimental and refined electron density were unsuccessful. In order to avoid any potential structural perturbation of the putative active site from Pt or PEG 4000, the native model was used in all structural analysis despite the lower resolution of the native data (Table 1). The r.m.s.d. between the native and derivative models is 0.72 Å for all atoms. The native AlkD model was validated using PROCHECK (Laskowski et al, 1993), and has been deposited along with structure factors in the Protein Data Bank under accession code 3BVS.

Glycosylase Activity Assay

AlkD glycosylase activity was measured by alkaline cleavage of the abasic DNA product of 7mG excision from a 25mer oligonucleotide duplex containing a 7mG·C base pair. 7mG was incorporated into DNA duplexes enzymatically using the previously described method (Asaeda et al, 2000). Briefly, the primer oligonucleotide 5′-GACCACTACACC was ^{32}P-labeled at the 5′-end, annealed to a 3-fold excess of the complementary strand (5′- GTTGTTAGGAAACGGTGTAGTGGTC), and extended using DNA polymerase I Klenow fragment (New England Biolabs) in the presence of deoxy-7-methylguanosine 5′-triphosphate (d7mGTP, Sigma), dCTP, dTTP, and dATP. In a 10 µl glycosylase reaction, 2 nM radiolabeled DNA duplex was incubated with 2 µM AlkD in 50 mM HEPES pH 7.5, 100 mM KCl, 10 mM DTT, and 2 mM EDTA. The reaction was stopped at various times by addition of 0.2 N NaOH, and heated at 70°C for 10 min. The 12mer product and remaining 25mer substrate DNA strands were separated by denaturing 15% polyacrylamide gel electrophoresis in 7M urea and quantitated by autoradiography.

DNA Binding Assay

DNA binding of wild-type and mutant AlkD was measured by the change in fluorescence anisotropy as protein was added to an oligonucleotide duplex that contained a site-specific modification (X) in the middle of one strand [d(GACCACTACACC***X***TTTCCTAACAAC)] and a 6-carboxyfluorescein on the 3'-end of the complementary strand [d(GTTGTTAGGAAACGGTGTAGTGGTC)-FAM].

Oligonucleotides containing abasic sites were chemically synthesized, and those containing 7mG were prepared as above but without the ^{32}P-label. Increasing concentrations of protein (0-30 uM) were added to a 50 nM DNA in 20 mM Bis-Tris Propane pH 6.5, 100 mM NaCl, 2 mM DTT, and 0.1 mM EDTA. Polarized fluorescence intensities using excitation and emission wavelengths of 485 and 538 were measured at 25 °C using a SpectraMax M5 microplate reader (Molecular Devices). Measurements were recorded within 2 minutes of adding wild-type AlkD to DNA to ensure that no more than 50% of the 7mG-DNA substrate was converted to abasic-DNA product. Dissociation constants were derived by fitting a two-state binding model to data from three independent experiments.

Acknowledgements

The authors wish to thank Zdzislaw Wawrzak for help with data collection, and the staff at the Life Sciences (LS-CAT) and Southeast Regional Collaborative Access Teams (SER-CAT) at the Advanced Photon Source (Argonne, IL). Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). This work was funded by the American Cancer Society (RSG-07-063-01-GMC). E.H.R. was supported by the Vanderbilt Training Program in Environmental Toxicology (T32 ES07028).

CHAPTER III

AN UNPRECEDENTED NUCLEIC ACID CAPTURE MECHANISM FOR EXCISION OF DNA DAMAGE*

Summary

DNA glycosylases that remove alkylated and deaminated purine nucleobases are essential DNA repair enzymes that protect the genome, and at the same time confound cancer alkylation therapy, by excising cytotoxic *N*3-methyladenine bases that are formed by many DNA targeting anticancer compounds. The basis for glycosylase specificity toward *N*3- and *N*7-alkylpurines remains unclear, and is believed to result from intrinsic instability of the modified base and not from direct enzyme functional group chemistry. Here, we present crystal structures of a recently discovered 3mA DNA glycosylase, *Bacillus cereus* AlkD, in complex with DNAs containing alkylated, mismatched, and abasic nucleotides. Unlike other glycosylases, AlkD captures the extrahelical lesion in a solvent-exposed orientation rather than an active site pocket on the protein, and provides the first illustration for how hydrolysis of unstable *N*3- and *N*7-alkylated bases can be facilitated by increased lifetime out of the DNA helix. The structures and supporting biochemical analysis of base flipping and catalysis reveal how AlkD's tandem HEAT-repeats distort the DNA backbone to sense energetic differences in non-Watson-Crick base pairs without duplex intercalation.

* The work presented in this chapter was published in Rubinson EH, Gowda AS, Spratt TED, Gold B, Eichman BF (2010) An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* **468**, 406-411.

Introduction

Alkylation of DNA by endogenous methyl donors, environmental toxins, and chemotherapeutic agents produces a diverse spectrum of cytotoxic and mutagenic lesions, including *N*3-methyladenine (3mA), *N*7-methylguanine (7mG), and 1,N^6-ethenodenine (εA), that threaten the survival of all organisms (Friedberg et al, 2006; Singer & Grunberger, 1983). These aberrant bases have all been detected in humans and rats after exposure to various carcinogens (Holt et al, 1998; Shuker et al, 1987; Shuker & Farmer, 1992). 3mA is a highly toxic lesion owing to its inhibition of DNA replication by blocking progression of DNA polymerase (Larson et al, 1985; Plosky et al, 2008), and production of such lesions is the rationale behind the use of alkylating agents in chemotherapy. *N*7-substituted guanines are generally the most prevalent alkylation lesions and display a wide range of toxic and mutagenic biological properties (Gates et al, 2004). By virtue of their positive charges at physiological pH, 7mG and 3mA, which has a pKa ~ 6.1 (Fujii & Itaya, 1999), are especially susceptible to spontaneous depurination, which generates abasic sites in DNA that can ultimately lead to single- and double-strand breaks.

DNA glycosylases initiate the base excision repair of *N*3- and *N*7-methylpurines from the genome by catalyzing hydrolysis of the glycosylic (C1′-N) bond (Figure 18a,b). Despite their structural diversity, all DNA glycosylases studied to date utilize a common base flipping mechanism to access damaged DNA and orient the substrate for the hydrolysis step, whereby the target nucleotide is rotated 180° around the phosphoribose backbone into a complementary shaped active site pocket on the protein (Stivers, 2004; Stivers, 2008). The resulting distortion to the DNA duplex and loss of base stacking is

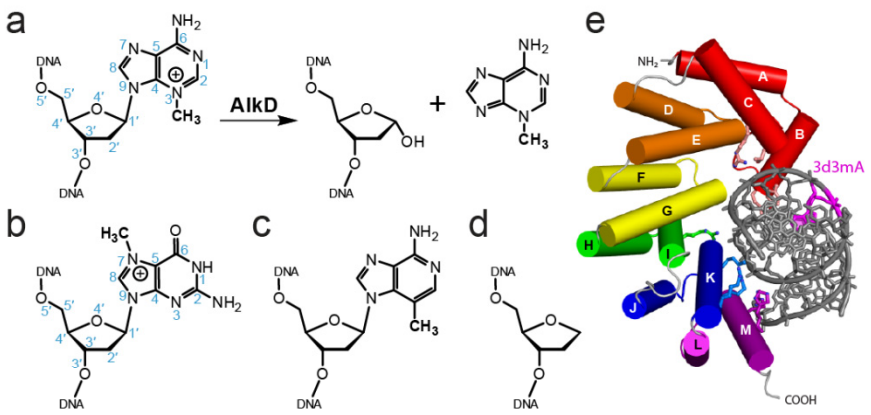


Figure 18. Base excision repair of alkylated DNA by AlkD. **a**, AlkD catalyzes the hydrolysis of the *N*-glycosylic bond to liberate an abasic site and free nucleobase. The enzyme is specific for positively charged *N*3-methyladenine (**a**) and *N*7-methylguanine (**b**). **c,d,** Structures of 3-deaza-3-methyladenosine (**c**) and tetrahydrofuran (**d**) used to trap AlkD in complex with alkylated and abasic DNA. **e**, Crystal structure of AlkD bound to 3d3mA-DNA. Each of the 6 HEAT-repeats are colored red-to-violet. The DNA is colored silver with the 3d3mA nucleotide colored magenta.

stabilized by an intercalating side chain "plug" that fills the void created by the extrahelical nucleotide. Monofunctional glycosylases (i.e., those that lack an AP lyase/phosphodiesterase activity) typically excise their substrate nucleobases by using a carboxylate side chain as a general base to activate a water nucleophile or to stabilize the carbocation transition state during base dissociation (Stivers & Jiang, 2003). Mutation of this residue, however, does not abolish catalytic activity in all cases, leading to a model in which conformational strain in the DNA arising from extensive binding energy helps to drive the reaction forward (Mol et al, 2002; Parikh et al, 2000). Similarly, the lack of a residue capable of performing general base catalysis in 3mA-specific DNA glycosylases (e.g., *E. coli* TAG, *H. pylori* MagIII) (Drohat et al, 2002; Eichman et al, 2003; Metz et al, 2007) is consistent with the idea that excision of positively charged 3mA and 7mG does not require the same level of catalytic assistance as other, more stable alkylpurines such as εA, although direct structural evidence for this has not been reported.

The AlkC/AlkD proteins, recently discovered in *Bacillus cereus* and subsequently identified in all three kingdoms of life (Figure B1), have emerged as a unique DNA glycosylase superfamily specific for positively charged *N*3- and *N*7-alkylpurines (Alseth et al, 2006; Dalhus et al, 2007). AlkD accelerates the rate of 7mG hydrolysis from DNA 100-fold over the spontaneous rate of 7mG depurination (Rubinson et al, 2008), prompting us to investigate the mechanism by which AlkD specifically catalyzes excision of destabilized alkylated bases. Here, we present crystal structures of *B. cereus* AlkD in complex with DNA damage resembling the substrate and product of the glycosylase reaction. These structures and supporting biochemical analysis of base flipping and 7mG depurination demonstrate how AlkD utilizes an unprecedented strategy to trap non-canonical base pairs that allows for specific hydrolysis of destabilized *N*-glycosylic bonds without direct chemical attack from the enzyme.

Results

A New Architecture for Binding Nucleic Acids

We previously determined the crystal structure of *B. cereus* AlkD and identified residues important for DNA binding and catalysis (Rubinson et al, 2008). AlkD is comprised entirely of HEAT (*h*untingtin, *e*longation factor 3, *A* subunit of protein phosphatase 2A and *T*OR1) repeats– tandem α-helical motifs that generate extended superhelical structures typically mediating protein, but not nucleic acid interactions (Andrade & Bork, 1995; Dalhus et al, 2007; Rubinson et al, 2008). Recently, an extended HEAT repeat ring structure was identified in the DNA-dependent protein kinase catalytic subunit, DNA-PKcs (Sibanda et al, 2010; Williams et al, 2008). To our

knowledge, the structures of AlkD-DNA complexes reported here are the first examples of a catalytic HEAT repeat protein bound to DNA. HEAT-repeats typically bind polypeptides along their concave surface, and mutation of Asp113 or Arg148 within this concave region of AlkD significantly affects 7mG excision and DNA binding activities (Rubinson et al, 2008).

To investigate the mechanisms by which this novel enzyme interacts with DNA and catalyzes base excision, we determined crystal structures of *B. cereus* AlkD in complex with DNA duplexes that resemble the substrate and product of 3mA base excision (Figure 18a). Trapping an alkylpurine DNA glycosylase onto a 3mA-containing substrate has presented a formidable challenge owing to the inherent instability of the glycosylic bond. To overcome this obstacle, we crystallized AlkD in complex with DNA containing 3-deaza-3-methyladenine (3d3mA), a structural 3mA mimetic in which the N3 nitrogen is replaced with carbon (Figure 18c). The 3d3mA base is refractory to spontaneous depurination or excision by AlkD or human alkyladenine DNA glycosylase (AAG) (Plosky et al, 2008), presumably because the 3d3mA purine ring lacks the formal positive charge associated with 3mA. Replacing N3 with carbon does not have a significant effect on duplex stability; substitution of adenine with 3-deazaA in 12mer DNA decreases the favorable ΔG^0 for duplex melting by ~1 kcal/mol, compared to ~8 kcal decrease for an A→3d3mA substitution (Hershey & Chase, 1952). Thus, the methyl substituent alone destabilizes the DNA duplex, which implies that 3d3mA and 3mA should have a similar effect on DNA. We also crystallized AlkD in complex with DNA containing a tetrahydrofuran (THF) moiety (Figure 18d), which resembles the abasic site product of base hydrolysis. The crystal structures of the AlkD/3d3mA-DNA and

AlkD/THF-DNA complexes were determined by molecular replacement using the unliganded AlkD structure as a search model. After manual fitting of the DNA molecules into the residual electron density maps, the AlkD/3d3mA-DNA model was refined to a crystallographic residual of 15.9% (R_{free} = 18.3%) at 1.6 Å, and AlkD/THF-DNA refined to 1.75 Å and an R/R_{free} of 18.5%/22.5% (Table B1, Figure B2). The 3d3mA and THF complexes, assembled from different lengths and sequences of DNA, crystallized in two unique packing arrangements (Figure 19).

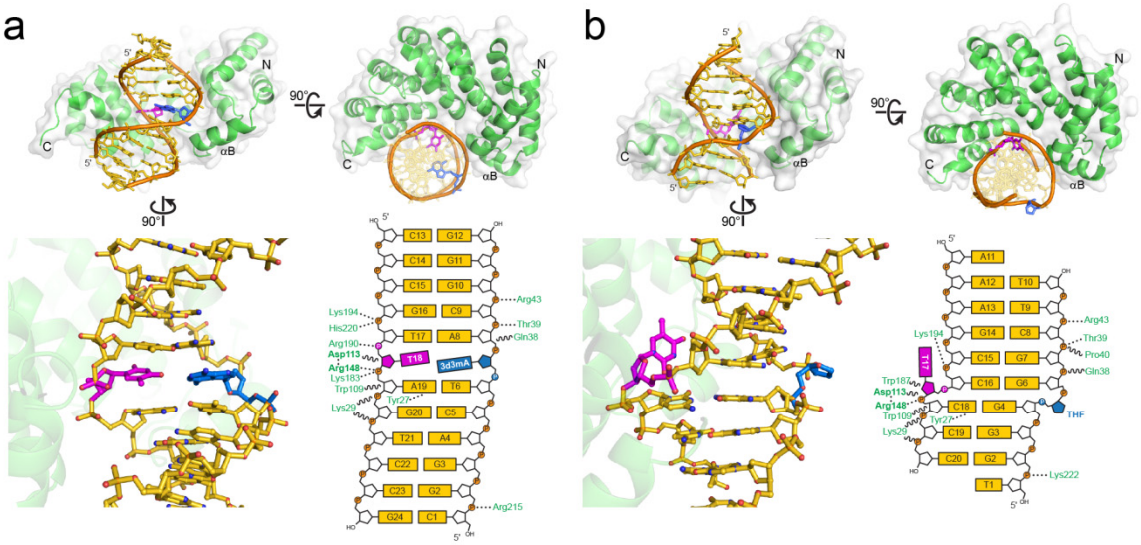


Figure 19. Crystal structures of AlkD in complex with 3d3mA-DNA (**a**) and THF-DNA (**b**). The top of each panel shows orthogonal views of the AlkD protein (green) wrapping around the DNA duplex (gold). The modified 3d3mA and THF nucleotides are colored blue, and the opposing thymine is magenta. At the bottom, a side view of the atomic model and corresponding schematic illustrates the interactions between the modified base pairs and the protein. Dashed lines represent hydrogen bonds and wavy lines represent van der Waals interactions.

Both substrate and product complexes show the same general mode of nucleic acid binding despite their unique crystal packing arrangements. The DNA duplex is positioned along AlkD's concave surface, which is lined with positively charged residues from the C-terminal α-helices of each HEAT repeat (Figures 18e and B3). The C-shaped protein wraps halfway around the DNA helix with a footprint of ~10 base-pairs (Figure

19). The contact surface is dominated by electrostatic interactions between the positively-charged side chains and the phosphoribose backbone of the non-lesioned strand in the immediate vicinity of the modified base pair. In contrast, contacts to the lesioned strand are limited to interactions between N- and C-terminal helices and base pairs further removed from the lesion (Figure 19). In both structures, the DNA helical axis is bent about 30° away from the N-terminal end of the protein as a result of helix αB (the only non-HEAT repeat in AlkD) projecting into the base stack (Figures 18e and 19). A 2-Å shift in helix αB is the only noticeable movement in the protein upon DNA binding (Figure B4).

A Novel Lesion Capture Mechanism

The most striking feature of the AlkD/DNA complexes is that both 3d3mA and THF lesions reside on the face of the DNA duplex not in contact with the protein, whereas the base opposite the lesion is nestled into a cleft on the protein's concave surface that contains catalytic residues Asp113 and Arg148 (Figures 19 and 20). The similarity in the orientation of the 12mer 3d3mA-DNA and 10mer THF-DNA is significant given the different DNA sequences and crystal packing. In the AlkD/substrate complex, the 3d3mA•T unpredictably forms a highly sheared base pair in which the 3d3mA remains stacked between T6 and A8, while the opposite thymine (T18) is displaced into the minor groove with no hydrogen bonds to the 3d3mA (Fig 20a). There are no direct contacts to the T18 base. Rather, the opposing thymine is held in this position by distortion of the T18/A19 backbone as a result of a hydrogen bond network from Asp113-Arg148 and Arg190. The protein-DNA interface is further strengthened by

van der Waals interactions between tryptophans (109 and 187) and the phosphoribose backbone flanking the damaged base pair.

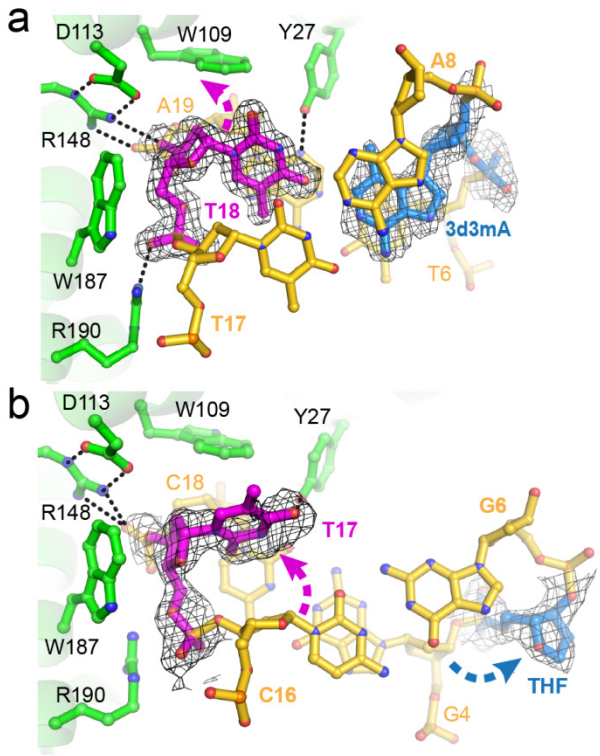


Figure 20. Recognition of DNA damage by AlkD. **a**, 3d3mA-DNA (substrate) complex; **b**, THF-DNA (product) complex. Views are down the DNA helix axis. Composite omit electron density (contoured to 1σ) for the modified base pairs is superimposed against the crystallographic models. Molecules are colored as in Fig. 2. Hydrogen bonds are shown as dashed lines. Dashed arrows denote displacement of opposing thymine and THF deoxyribose from their positions in B-DNA.

In the product complex, the abasic site is rotated ~90° around the phosphoribose backbone into the major groove, and is completely solvent exposed (Figures 19b and 20b). Interestingly, the opposite thymine is slipped completely out of the base stack and into the minor groove of the DNA, as if it had continued the same extrahelical trajectory from its position in the substrate complex. This extrahelical thymine base is rotated (χ = 58°) so that the plane of the pyrimidine ring is almost co-linear with the helical axis. Unlike all other DNA glycosylase complexes, there are no intercalating amino acid side

chains that plug the gap left by the flipped base, and as a consequence, the duplex has collapsed in order to maintain base stacking interactions. Guanine G4, immediately 5′ to the THF, is now stacked with cytosine C16 on the opposite strand, with a rise of 3.3 Å (Figure 20b). Importantly, the DNA backbone, but not the helical axis, is highly distorted as a result of the large slide (4.4 Å) and twist (58°) between G4•C18 and G6•C16 base pairs (Figures 19b, 20b, and 23a). A hydrogen bonding interaction between Tyr27 at the C-terminal end of helix αB and the base 3′ to the tipped thymine is the only specific AlkD-nucleobase contact (Figure 20b). Thus, AlkD stabilizes the distortions in both substrate and product DNA—a sheared 3d3mA•T base pair and a single base THF•T bubble—through preferential interactions with the phosphoribose backbone of the non-lesioned strand.

The solvent-exposed capture of damaged DNA in the AlkD structures is both unexpected and unprecedented for a DNA glycosylase, and raises the possibilities that either AlkD utilizes an alternative mechanism to catalyze base excision, or that the 3d3mA and THF crystal lattices trapped the same non-specific, catalytically incompetent protein/DNA complexes. Indeed, the aromatic region at the center of the concave cleft loosely resembles nucleobase binding pockets of other alkylpurine DNA glycosylases (Dalhus et al, 2007; Rubinson et al, 2008) and contains highly conserved residues important for 7mG excision activity and protection against bacterial sensitivity to alkylating agents (Alseth et al, 2006; Dalhus et al, 2007; Rubinson et al, 2008). However, several important differences between this region in AlkD and the active sites of other glycosylases argue against a lesion binding pocket in AlkD. First, AlkD lacks the plug residue universally used by DNA glycosylases to prevent the flipped substrate

base from re-entering the DNA base stack. Second, an extrahelical nucleobase would be sterically prohibited from full 180° rotation into this shallow cleft (Figure 20), and full extrusion of the extrahelical base is critical for exposure of the glycosylic bond for hydrolysis. Third, high concentrations of free nucleobases do not inhibit base excision activity by AlkD as observed in other alkylpurine glycosylases (Figure B5) (Eichman et al, 2003), which argues against a specific base binding pocket in AlkD. Fourth, the electrostatic interaction between Asp113 and Arg148, not observed in other glycosylase active sites, disfavors a direct catalytic role for Asp113 as a general base in hydrolysis. Finally, mutation of residues inside a putative base binding cleft directly adjacent to the catalytic Asp113 and Arg148 residues did not affect 7mG excision activity (Figure B6).

Alkylpurine DNA glycosylases normally exhibit enhanced alkylpurine excision activity for mispaired bases, presumably because of their greater propensity to base-flip (Eichman et al, 2003; O'Brien & Ellenberger, 2004a; O'Brien & Ellenberger, 2004b). However, AlkD does not discriminate against the base opposite the lesion. AlkD excises 7mG paired with cytosine, thymine, adenine, or guanine at the same rate (Figures 22c and B7), whereas AlkA, for example, showed a 250-fold increase in rate for 7mG paired with T versus C (O'Brien & Ellenberger, 2004b). The lack of opposite base effects in AlkD is consistent with the structures, in which any of the four nucleobases can fit inside the minor-groove cavity that harbors the tipped orphaned base. Crystallographically, we observed that a cytosine opposite the lesion, which represents the product of 7mG excision, is identical to the THF•T complex (Figures B2 and B8), and that AlkD forms isomorphous crystals with THF•A and THF•G duplexes.

To determine the orientation of DNA relative to the central cleft during catalysis, we measured the rate of 7mG excision opposite a bulky nucleotide. A pyrene nucleotide wedge across from uracil has previously been shown to promote uracil flipping and consequently to enhance base excision activity of uracil DNA glycosylase (UDG) and rescue the loss of activity of UDG mutants that lack the Leu191 plug side chain (Jiang et al, 2001). In contrast to UDG, placing pyrene across from 7mG reduced AlkD's activity 10-fold relative to a 7mG•C pair (Figure 22c). Superposition of the pyrene onto the opposite thymine in the AlkD/DNA crystal structures shows that although this bulky group can fit in the minor groove cavity, the pyrene would be hindered from rotating into this tipped position. By contrast, pyrene on the face of the duplex opposite the protein would not be expected to affect the rate of 7mG excision unless AlkD utilized an intercalating plug strategy for base flipping. Thus, the consistency between the crystal structures and partial inhibition of 7mG activity by an opposing pyrene argue strongly that the crystal structures represent a catalytically competent orientation of DNA.

In a converse experiment, we tested AlkD's ability to excise bulky nucleobases given the unhindered position of the lesion in the crystal structures. Bulky pyridyloxobutyl (POB) base adducts (Figure 21a) arise in DNA upon exposure to cigarette smoke-derived nitrosamine carcinogens (Hecht, 1999). The expectation was that AlkD should excise POB-bases from DNA, whereas the tightly constrained nucleobase binding pocket of human AAG would discriminate against these bulky alkyl adducts (Lau et al, 1998). Indeed, AlkD liberated positively charged 7-POB-Gua and O^2-POB-Cyt adducts from DNA, whereas neither of these modified bases was detected after treatment with AAG or in a mock reaction containing no enzyme (Figure 21b,c). Neutral

adducts O^6-POB-Gua and O^2-POB-Thy present in the DNA were not detected in the supernatant upon reaction with AlkD, consistent with the specificity of AlkD for positively charged lesions. This result indicates that AlkD need not flip the substrate base into an active site cavity in order to excise *N*3- or *N*7-alkylpurines from DNA. In

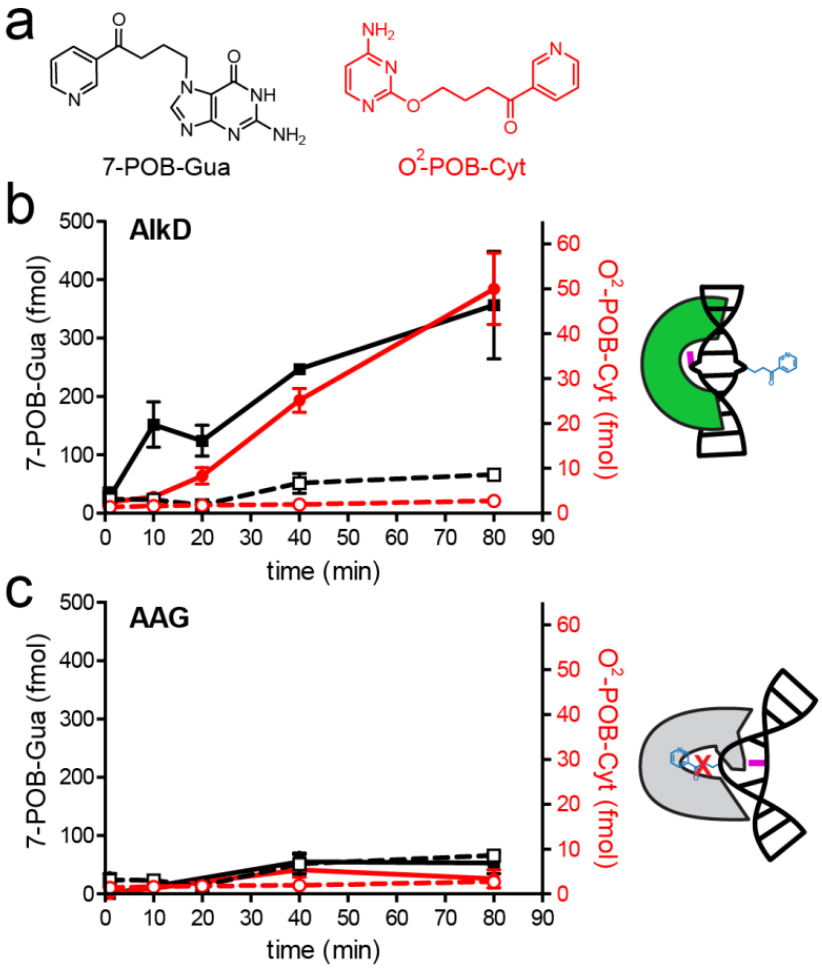


Figure 21. Excision of *N*7- and O^2- pyridyloxobutyl (POB) base adducts by AlkD. **a**, Chemical structures of *N*7-POB-deoxyguanine and O^2-POB-deoxycytosine. **b,c**, Time courses for the release of *N*7-POB-dG (black squares) and O^2-POB-dC (red circles) in the presence (closed symbols, solid lines) and absence (open symbols, dashed lines) of *Bacillus cereus* AlkD (**b**) or human AAG (**c**). Error bars represent the standard deviation from three independent measurements.

order to rule out the possibility of base excision by a second molecule of AlkD, we tested DNA binding and 7mG excision under stoichiometric conditions and found that although

AlkD is able to coat DNA nonspecifically (~2 AlkD molecules per 12 base pairs), only one molecule of AlkD per lesion is necessary for 7mG excision (Figure B9).

AlkD Traps and Restructures Destabilized Base Pairs

The mechanism by which DNA repair enzymes locate subtle chemical alterations with exquisite specificity among a vast excess of unmodified DNA remains relatively poorly understood. Recent work suggests that glycosylases detect damage by using plug side chains to probe for free energy differences between normal and modified base pairs (Banerjee et al, 2006; Banerjee et al, 2005; Yang et al, 2009; Yang et al, 2008). The lack of lesion-specific and DNA intercalating interactions in the AlkD/DNA complexes suggests that AlkD detects damage solely on the basis of DNA duplex destabilization resulting from altered base stacking or base pairing of non-canonical base pairs. In support of this, we crystallized the protein in complex with DNA containing a G•T mismatch (Figure 22a, Table B1), for which AlkD has no activity, but were unable to trap the protein onto the same oligonucleotide containing a G•C or A•T base pair at this same position. The resulting 1.5-Å AlkD/G•T-DNA structure is virtually identical to the 3d3mA•T complex, with the only difference being a slight rotation of the mismatched G toward the major groove (Figure B4).

There are significant differences between the structure of the G•T base pair bound by AlkD and in the context of DNA alone, and these differences provide a structural basis for DNA damage recognition by AlkD (Figure 22). In DNA alone, G•T mismatches form wobble base pairs with two hydrogen bonds between the Watson-Crick faces of the

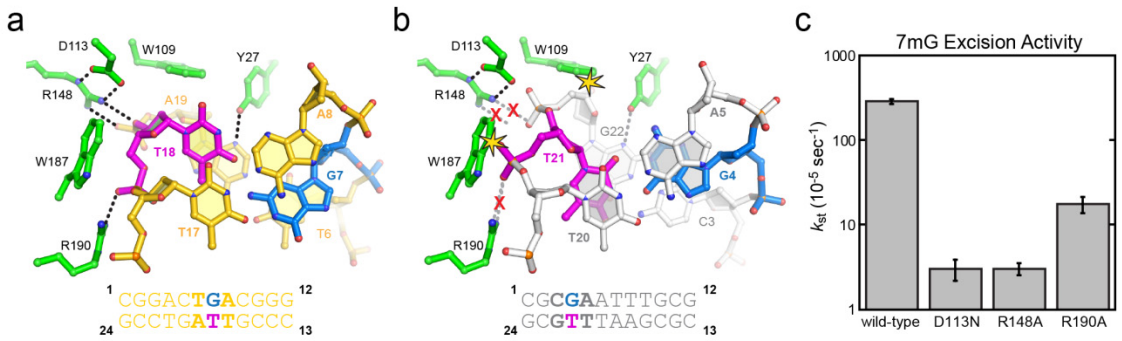


Figure 22. Remodeling of the G•T wobble base pair by AlkD. **a**, AlkD/G•T-DNA complex viewed down the helical axis. Hydrogen bonds are shown as dashed lines. **b**, The structure of a G•T wobble base pair found in DNA alone (PDB 113D) superimposed onto the AlkD/G•T complex. Steric clashes between the protein and DNA are highlighted by yellow stars, and disrupted hydrogen bonds are shown by a red X. **c**, Relative single-turnover rates (k_{st}) of 7mG excision from a 25mer oligonucleotide duplex by wild-type and mutants of AlkD. Wild-type, D113N, and R148A data from (Rubinson et al., 2008). Error bars represent the standard deviation from three independent measurements.

bases, and both bases remained well-stacked in the duplex (Hunter et al, 1987) (Figure B10). AlkD restructures the canonical G•T wobble base pair so that the two bases protrude into opposite DNA grooves with only a single hydrogen bond between guanine N^2 and thymine O^4 and with loss of base stacking on one face of the base pair (Figure 22a). Superposition of the G•T wobble DNA onto the AlkD structure reveals that the protein stabilizes this conformation by inducing a specific distortion to the DNA backbone in order to create the optimal hydrogen bonds and van der Waals interactions at the DNA capture site (Figure B10). Loss of hydrogen bonds and steric collisions would result if the wobble conformation was retained in the complex (Figure 22b). Thus, the enzyme detects the non-Watson-Crick base pair by restructuring the DNA backbone to create an optimized protein-DNA binding surface. This DNA conformation is stabilized in both 3d3mA•T and G•T complexes by specific protein-DNA contacts from Arg148-Asp113 and Arg190. Substitution of any of these residues, which are highly conserved

among AlkD orthologs (Figure B1), reduces single-turnover rates of 7mG excision by an order of magnitude (Figure 22c), highlighting the importance of these interactions to catalysis.

The similarity between the 3d3mA•T and G•T complexes suggests that AlkD detects an energetic difference in non-canonical base pairs as opposed to specifically recognizing the *N*3- or *N*7-methyl groups. Indeed, $\Delta\Delta G^0$ for formation of duplexes containing 3d3mA•T and G•T base pairs is +8 and +3.7 kcal/mol, respectively, despite the potential of these bases to hydrogen bond and to base stack (Aboul-ela et al, 1985; Hershey & Chase, 1952). Similarly, thermodynamic studies reveal that 7mG affects base stacking and hydration and modestly decreases duplex stability (Ezaz-Nikpay & Verdine, 1992), although 7mG is reported to not perturb the low temperature DNA structure (Ezaz-Nikpay & Verdine, 1994; Lee et al, 2008). Thus suggests that AlkD has the potential to detect subtle thermodynamic differences due to alterations in the electronic configuration or the hydration patterns of 7mG•C and G•C base pairs. Regardless of the mode of recognition, the AlkD-induced, unstacked DNA conformation demonstrates that the binding energy provided by the protein-DNA interface is sufficient for the protein to pause at a modified base pair.

Base Excision by Solvent Exposure

The specific orientation and conformation of the DNA trapped in the AlkD complexes provides a plausible rationale for the enzyme's specificity toward bases with a high propensity for depurination. We propose the lesion capture mechanism facilitates base hydrolysis by increasing the lifetime that the *N*-glycosylic bond is exposed to

solvent and by inducing a specific distortion to the DNA backbone. Without a direct enzymatic chemical step, only those bases with a destabilized *N*-glycosylic bond would be excised. To test this idea, we measured the rates of spontaneous 7mG depurination in single- and double-stranded DNA containing normal and mismatched base pairs. Consistent with the literature, the half-life of the 7mG glycosylic bond increases as its protection from solvent increases: d7mG•dC-dsDNA > d7mG•dT-dsDNA > d7mG-ssDNA > 5′-d7mGMP nucleotide > d7mG nucleoside (Figure B11) (Gates et al, 2004). However, the 100-fold rate enhancement of 7mG hydrolysis from duplex DNA by AlkD cannot be explained on the basis of DNA secondary structure effects alone. Close inspection of the highly distorted DNA backbone in the flipped abasic structure showed that the deoxyribose ring is positioned directly above a neighboring phosphate, and that several water molecules bridge this phosphate and the extrahelical deoxyribose C1′ carbon (Figure 23a), raising the possibility that the phosphate groups may participate in catalysis. DNA-mediated water positioning to facilitate hydrolysis is a plausible catalytic mechanism given the lack of a requirement for a general base in these likely highly dissociative reactions. Alternatively, electrostatic stabilization of an oxocarbenium ion intermediate by nearby phosphates, which has been reported for uracil DNA glycosylase (Dinner et al, 2001; Jiang et al, 2003), offers a second possible mechanism for DNA-stimulated catalysis.

Discussion

Catalysis of base excision without the use of direct functional group chemistry from the enzyme has been proposed for uracil (UDG), thymine (TDG), and 3mA

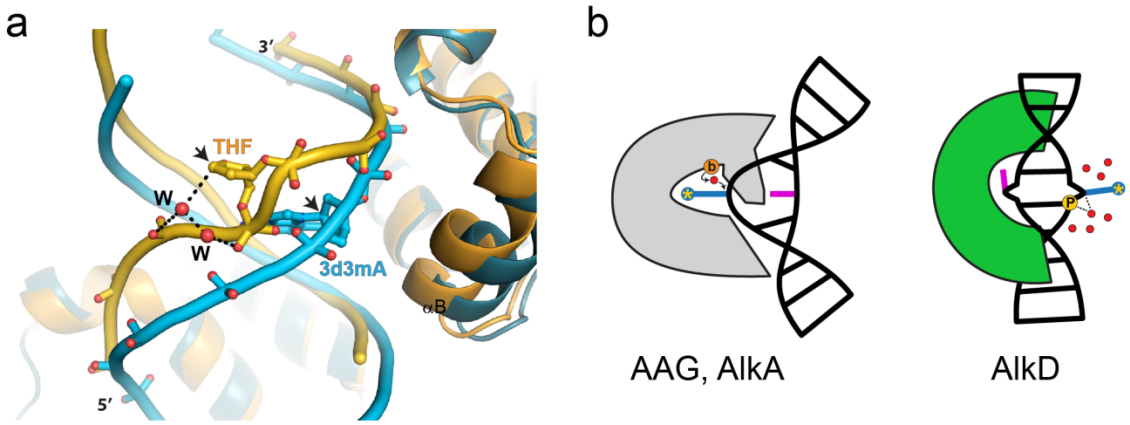


Figure 23. Proposed mechanism for how AlkD facilitates hydrolysis of *N*3 and *N*7-alkylpurines by distorting the DNA backbone. **a,** Superposition of AlkD/DNA substrate (3d3mA, blue) and product (THF, gold) complexes. The extrahelical anomeric C1′ carbon (marked with an arrow) is involved in water (W) mediated interactions with nearby DNA phosphates. **b,** Schematic showing the differences in DNA damage binding modes between AlkD and other DNA glycosylases. Typically, DNA glycosylases flip the lesioned base (blue) into a complementary active site cleft where chemistry can occur through activation of a water nucleophile (shown) or by direct stabilization of the transition state by a side chain acting as a general base (orange "b"). In these enzymes, a protein residue plugs the gap left in the DNA and interacts with the orphaned base (magenta), which remains stacked in the DNA duplex. Conversely, AlkD captures the DNA in an orientation that holds the orphaned base next to the protein and exposes the lesion to a hydrolytic environment. The distorted DNA conformation is stabilized not by a side chain plug, but by stacking of flanking base pairs as a result of both lesion and orphaned base flipping. Phosphate (orange "P") assisted hydrolysis could occur either by positioning of the water molecules adjacent to the C1' carbon in a dissociative hydrolysis reaction, or through stabilization of an oxocarbenium ion intermediate.

DNA glycosylases (TAG) (Metz et al, 2007; Mol et al, 2002; Parikh et al, 2000), but these mechanisms have remained speculative given the conservation of their protein architectures to other glycosylases that rely on specific functional groups within the active site (Huffman et al, 2005). AlkD represents a novel glycosylase found in bacteria, archaea, plants and lower eukaryotes (B. Eichman, unpublished;(Alseth et al, 2006). To our knowledge, most if not all of these organisms contain at least one AlkA or AAG alkylpurine DNA glycosylase, which remove *N*3- and *N*7-alkylpurine lesions by specifically binding to the damaged base. The fact that *Bacillus* alone contains at least

one AAG and two AlkA homologs raises the question as to why this organism has evolved an alternate mechanism for repair of alkylated DNA. The redundancy of alkylation damage repair may provide enhanced protection to organisms faced with an onslaught of methylating agents. Alternatively, AlkD may be a general DNA binding protein that coincidentally accelerates hydrolysis of unstable *N*-glycosylic bonds, or, as speculated below, may play a supporting role in generalized lesion detection as found in other DNA repair proteins. In this regard, the lack of a substrate binding pocket, a catalytic base residue, and an intercalating plug typical of other DNA glycosylases argue that AlkD may be more appropriately referred to as a "depurination facilitator protein." Regardless of the specific biological role of AlkD, our results show that AlkD's preference for *N*3- and *N*7-alkylpurines is determined by the inherent instability of these lesions coupled with the ability of the enzyme to impart a specific conformational distortion in the DNA.

In contrast to other BER enzymes, AlkD does not penetrate the DNA duplex in order to hold the lesion in an extrahelical, solvent-exposed orientation. The strategy of trapping DNA damage without the use of an intercalating plug has recently been observed in the structures of the oxidative demethylase AlkB involved in direct reversal repair of 1-methyladenine (1mA) (Yang et al, 2008). Like the AlkD complexes, the DNA bound to AlkB collapses to maintain base stacking and imparts a sharp deformation in the DNA backbone in order to flip the base out. This duplex collapse strategy in AlkB, coupled with specific contacts to the extrahelical 1mA lesion, helps explain AlkB's preference toward ssDNA substrates. In contrast, the specific capture of the *undamaged* strand and extrusion of both lesion and opposing bases in AlkD is consistent with its

preference for damaged bases in duplex DNA (Figure 20c). Thus, protein-induced DNA collapse seems to be an alternative mechanism for general extrusion of aberrant DNA bases, with substrate specificity and the type of repair dictated by the specific type of base binding pocket on the enzyme. The human AlkB homolog, ABH2, is virtually identical in structure to AlkB but has a marked preference for 1mA in dsDNA, presumably as a result of its intercalating plug strategy to maintain duplex stacking. This intercalating residue in BER glycosylases often makes specific interactions with the orphaned residue of the complementary strand. Exposure of the lesion to the solvent in AlkD, therefore, could be advantageous for AlkD because specific contacts required for the intercalation strategy would not be needed. From a thermodynamic perspective, unstacking a nucleobase from the duplex and placing it in a solvent exposed environment would be costly and thus favor base excision, while facilitating recognition of the abasic product.

This work represents the first base excision repair enzyme with activity toward bulky POB-DNA adducts normally associated with nucleotide excision repair (NER) (Graham et al, 2001; Li et al, 2009) and may be indicative of a more generalized function of AlkD in genome maintenance. AlkD's lesion capture strategy is reminiscent of Rad4, the XPC ortholog in budding yeast, which recognizes cyclopyrimidine dimers by binding to the opposing nucleotides (Min & Pavletich, 2007). Exposure of the lesion away from the protein in XPC has the biological advantage of greater accessibility of the damage by the rest of the NER machinery. Similarly, the AlkD-DNA product complex could provide a platform for recruitment of the AP endonuclease or similar protein that would presumably bind the extrahelical abasic site, as seen in human APE1-DNA complexes

(Mol et al, 2000). Interestingly, the non-enzymatic alkyltransferase-like proteins were recently found to trigger NER of O^6-alkylguanines by inducing a specific protein-DNA complex, as illustrated by the crystal structure of ATL bound to DNA containing O^6-POB-dG (Tubbs et al, 2009). It is intriguing to speculate that AlkD may participate in alternative repair pathways by virtue of its ability to non-specifically recognize and expose DNA damage.

The AlkD protein represents a new structural superfamily of DNA processing enzyme. The structures of AlkD bound to damaged DNA reveal how a HEAT repeat protein can engage nucleic acids via electrostatic and van der Waals interactions between the concave surface of the protein and the phosphoribose backbone. Comparison of this complex with the nuclear import factor importin β, which uses HEAT repeats in the same manner to bind a highly charged region of partner protein importin α (Cingolani et al, 1999) and Ran GTPase (Vetter et al, 1999), demonstrates that the concave surface of the HEAT domain is a generalized macromolecular binding platform. Related to chromosome function, HEAT repeats have been identified in several chromatin-remodeling factors, including condensins, cohesins, and some SWI2/SNF2 proteins (Neuwald & Hirano, 2000), and have been predicted to comprise the majority of DNA-damage responsive protein kinases ATM, ATR, and DNA-PK (Perry & Kleckner, 2003). Most recently, electron microscopy and crystal structures of the catalytic subunit of DNA-PK revealed HEAT repeats to form a large ring structure and a putative DNA binding element within the ring (Sibanda et al, 2010; Williams et al, 2008), raising the possibility that many other structurally uncharacterized DNA processing enzymes utilize HEAT domains to bind DNA in a manner similar to AlkD.

Methods Summary

Preparation of 3-deaza-3-methyladenine

The 3-methyl-3-deazaadenine phosphoramidite was prepared as previously described (Irani & SantaLucia, 2002). The 3-methyl-3-deazaadenine modified deoxynucleotide oligomers were synthesized at the University of Pittsburgh DNA core facility, purified by reverse phase HPLC, desalted on Sepahdex G20 and analyzed by MALDI-TOF-MS. All other oligonucleotides were synthesized by Integrated DNA Technologies.

AlkD/DNA Crystal Structure Determination

Wild-type and mutant *B. cereus* AlkD proteins were purified as described previously (Rubinson et al, 2008). AlkD/DNA complexes were assembled using a 1:1.2 molar ratio of AlkD:DNA. AlkD/THF-DNA crystals were grown at 16°C by vapor diffusion against a reservoir containing 0.1 M Bis-Tris pH 6.5, 0.1 mM NaCl, and 9% PEG 3350, and were flash frozen in a 30% glycerol/reservoir solution. AlkD/3d3mA-DNA and AlkD/G•T-DNA crystals were grown at 21°C from reservoir solutions containing 85 mM NaAcetate pH 4.6, 170 mM ammonium acetate, 25.5% PEG 4000, and 15% glycerol, and were flash frozen directly from the mother liquor. X-ray data (Table B1) were collected at the Advanced Photon Source (21-ID-D, LS-CAT). All structures were determined by molecular replacement using the unliganded AlkD structure (PDB ID 3BVS) as a search model.

Biochemical Assays

Base excision and DNA binding activity assays were performed as previously described (Rubinson et al, 2008). Kinetic data were analyzed by standard single-turnover techniques. Activity toward POB-nucleobases was measured by incubation of NNK-treated genomic DNA with AlkD or AAG, followed by DNA precipitation and mass-spectrometric detection of *N*7-POB-dG and O^2-POB-dC in the supernatant.

Acknowledgements

We thank James Stivers for providing the pyrene phosphoramidite, Zdzislaw Warzak and LS-CAT beamline staff at the Advanced Photon Source (APS) for assistance with X-ray data collection, and Tom Ellenberger, James Stivers, and Patrick O'Brien for helpful comments on the manuscript. Use of the APS was supported by the U.S. Department of Energy Office of Basic Energy Sciences. Use of LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor. This research was supported by a grant from the American Cancer Society (to B.F.E.). E.H.R. was supported in part by the Vanderbilt Training Program in Molecular Toxicology. Additional support for local crystallography facilities was provided by the Vanderbilt Center in Molecular Toxicology and the Vanderbilt-Ingram Cancer Center.

Author Contributions

E.H.R. purified and crystallized AlkD, determined crystal structures, and performed 7mG activity assays; B.G. synthesized 3d3mA oligonucleotides; A.S.P.G. and

T.E.S. performed POB activity assays; B.F.E. designed the project and analyzed data; B.F.E. and E.H.R. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 3JX7 (3d3mA•T), 3JXY (G•T), 3JXZ (THF•T), and 3JY1 (THF•C).

Methods

AlkD Purification and Crystallization

AlkD proteins were purified as described previously (Rubinson et al, 2008). Briefly, *Bacillus cereus* AlkD was overexpressed as an N-terminal His_6-SUMO-AlkD fusion protein in *E. coli* HMS174 cells for 3 h at 37°C. AlkD was isolated using Ni-NTA (Qiagen) affinity chromatography, followed by cleavage of the His_6-SUMO tag and further purification by heparin affinity and gel filtration chromatography. Protein was concentrated to 12.5 mg/mL and stored in 20 mM Bis-Tris Propane, 100 mM NaCl, 2 mM DTT and 0.1 mM EDTA. Site-directed mutagenesis of the wild-type AlkD vector was performed using a Quik-Change Kit (Stratagene), Mutant proteins were overexpressed and purified identically to wild-type AlkD, and their structure verified by circular dichroism spectroscopy.

AlkD/DNA complexes were assembled by incubating 0.45 mM protein and 0.54 mM oligonucleotide for 15 min at 4°C. Oligonucleotide sequences used were

d(TGGG(THF)GGCTT)/d(AAAGCC*Y*CCC), in which *Y* = thymine or cytosine, and d(CGGACT*X*ACGGG)/d(CCCGT*T*TCCG), in which X was either 3d3mA or G. AlkD/THF-DNA crystals were grown at 16°C by mixing 2 μL protein/DNA complex with 2 μL reservoir solution containing 0.1 M Bis-Tris pH 6.5, 0.1 mM NaCl, and 19% PEG 3350 and 2% glycerol. Crystals were soaked in 30% glycerol/reservoir solution for 1 min and flash frozen in a liquid nitrogen stream. Crystals of 3d3mA-DNA and G•T-DNA complexes were grown from reservoir solutions containing 85 mM NaAcetate pH 4.6, 170 mM ammonium acetate, 25.5% PEG 4000, and 15% glycerol at 21°C, and were flash frozen in liquid nitrogen directly from this solution.

X-ray Data Collection, Phasing, and Structure Refinement

X-ray data (Table B1) were collected at the Advanced Photon Source beamlines 21-ID-D and 21-ID-G (LS-CAT) and processed with HKL2000 (Otwinowski & Minor, 1997). Molecular replacement using unliganded AlkD (PDB ID 3BVS) as a search model in Phaser (McCoy et al, 2005) gave a clear solution for each structure. Following one round of simulated annealing refinement in CNS (Brunger et al, 1998), the entire DNA molecules could be discerned and were built into 2Fo-Fc and Fo-Fc electron density using XtalView (McRee, 1999). Atomic coordinates and B-factors for the protein/DNA models were refined in Phenix (Adams et al, 2007). TLS refinement with protein and each DNA chain defined as three separate TLS groups was carried out for each model except the GT-complex. Individual anisotropic B-factors were derived from the refined TLS parameters and held fixed during subsequent rounds of refinement, which significantly decreased the crystallographic residuals and improved the electron

density maps. Instead of TLS refinement, individual anisotropic B-factors were explicitly refined for the G•T-complex. Adjustments to the model, including addition of solvent molecules, using Coot (Emsley & Cowtan, 2004) were guided by manual inspection of 2Fo-Fc and Fo-Fc electron density maps and were judged successful by a decrease in R_{free} during refinement.

Protein and DNA models were validated using PROCHECK (Laskowski et al, 1993) and DNA parameters were quantified using CURVES 5.2 (Lavery & Sklenar, 1988). All but one out of the total 223-231 protein residues resided in the most favored (191-198 residues) or allowed (14-15 residues) regions of the Ramachandran plot. As in the unliganded structure (Rubinson et al, 2008), Thr54 in all four DNA complex structures remained in the disallowed region despite an excellent fit to 2Fo-Fc electron density maps.

Enzyme Activity

Excision of 7mG by AlkD was measured by incubating the enzyme with a 25mer oligonucleotide containing a centrally located 7mG and following the appearance of abasic DNA product after alkaline cleavage. 7mG was enzymatically incorporated into DNA duplexes using the previously described method (Asaeda et al, 2000), in which an oligonucleotide primer [d(GACCACTACACC)] was ^{32}P-labeled at the 5′-end, annealed to a 3-fold excess of the complementary strand [d(GTTGTTAGGAAACGGTGTAGTGGTC)] and extended using DNA polymerase I Klenow fragment (New England Biolabs) in the presence of 2′-deoxy-7-methylguanosine 5′-triphosphate (Sigma), dCTP, dTTP, and dATP. To create 7mG mispairs, 100-fold

excess of complementary strand with T, G, A or pyrene in place of C at position 13 was reannealed to the 7mG containing oligonucleotide. Single-stranded 7mG containing strands were obtained by reannealing to 100-fold excess of unlabeled lesion strand with G in place of 7mG [d(GACCACTACACCGTTTCCTAACAAC)].

In a 10 μL glycosylase reaction, 100 nM [^{32}P]-DNA duplex was incubated with 0-20 μM AlkD in 50 mM HEPES pH 7.5, 100 mM KCl, 10 mM DTT, and 2 mM EDTA. The reaction was quenched at various times by addition of 0.2 N NaOH and heated at 70°C for 2 min. Substrate 25mer and product 12mer DNA strands were separated by denaturing 20% polyacrylamide gel electrophoresis in 7 M urea and quantitated by autoradiography. Kinetic data were analyzed by standard single-turnover techniques (Jones et al, 2008), which have been extensively used for DNA glycosylases (Baldwin & O'Brien, 2009; Bennett et al, 2006; Lyons & O'Brien, 2009; Maher & Bloom, 2007; Maher et al, 2007; Maiti et al, 2009). Enzymatic rate constants (k) were obtained from single-exponential fits to the data ($f_P = 1 - e^{-kt}$, in which f_P is the fraction of product). For determination of the single-turnover rate constant, k_{st}, AlkD was at least 5-fold in excess over the $K_{½}$ for a particular labeled DNA substrate (e.g., 5 μM for 7mG•C). For $K_{½}$ determinations, the 7mG excision assay was performed over a range of enzyme concentrations and $K_{½}$ obtained by fitting the Michaelis-Menten plot with the equation, $k_{obs} = V_{max}[AlkD] / (K_{½} + [AlkD])$. We note that our $K_{½}$ for maximal activity may differ from the K_m value for multiple turnover because the K_m can be affected by product release. Stoichiometric 7mG excision was performed in the presence of 10 μM unlabeled 25mer DNA duplex ($K_{½}$ for this DNA was determined to be 0.9 ± 0.1 μM). Spontaneous rates of 7mG hydrolysis were determined using the sequence

d(GACCACTACACC(7mG)ATTCCTTACAAC) that had been re-annealed to 100-fold excess complementary strand d(GTTGTAAGGAAT(C/T)GGTGTAGTGGTC).

POB Adduct Excision

Materials. DNA (cat. no. D1501), alkaline phosphatase (P8361), esterase (E2884) micrococcal nuclease (N3755), and phosphodiesterase II (P9041) were purchased from Sigma. The tetra-deuterated standards were provided by Dr Stephen S. Hecht of the University of Minnesota. NNKOAc was synthesized by Dhimant Desai of the Organic Synthesis facility of the Penn State college of Medicine.

NNKOAc-damaged DNA. Five milliliters DNA (2 mg/mL) dissolved in 100 mM sodium phosphate (pH 7.0), 1 mM EDTA and 50 mM NaCl was reacted with 1 mM NNKOAc and esterase (200 units) at 37°C for 2 h. The reaction was diluted to 10 mL with H_2O and extracted with 10 mL $CHCl_3$/iso-amyl alcohol (24/1) to remove the protein and 10 mL ethyl acetate to remove any unreacted NNKOAc. The DNA was precipitated by the addition of 40 mL ethanol and washed twice 70% ethanol. Residual amounts of ethanol was removed by rotary evaporation and the DNA was dissolved in H_2O, aliquoted and stored at -80°C prior to use.

Glycosylase reactions. The damaged DNA (1 mg/mL) was incubated with 1 μM glycosylase in 400 μL buffer (50 mM HEPES (pH 7.5), 1 mM EDTA, 100 mM KCl, 1 mM DTT) at 37°C. Aliquots (100 μL) were quenched at various times by the addition of 5 μL 3 M sodium acetate (pH 5.2) and 200 μL ice cold ethanol. The mixture was centrifuged for 10 min and the supernatant decanted and saved for analysis.

HPLC-MS/MS. Deuterated standards (100 fmol each of O^2-POB-C-d_4 and 7-POB-G-d_4) were added to the ethanol supernatant and the solvent evaporated. The sample was dissolved in 50 μL methanol for MS analysis. The samples were analyzed with a MDS/Sciex 4000 QTrap instrument with electrospray ionization (ESI) coupled to an Agilent 1100 HPLC system. Samples (20 μL) were loaded onto a column (Luna C18(2) 150 x 2 mm, 3 micron) which was eluted with 10 mM ammonium formate at 0.1 mL/min. The POB−DNA adducts, along with their deuterated standards, were monitored by selected reaction monitoring (SRM). The ion transitions were as follows 7-POB-Gua, *m*/*z* 299.1 $[M + 1]^+$ to *m*/*z* 148.1 $[POB]^+$; [pyridine-D_4]7-POB-Gua, *m*/*z* 303.1 $[M + 1]^+$ to *m*/*z* 152.1 $([pyridine\text{-}D_4]POB])^+$ and $[Gua + H]^+$; O^2-POB-Cyt, *m*/*z* 259.1 $[M + 1]^+$ to *m*/*z* 148.1 $[POB]^+$; [pyridine-D_4]O^2-POB-Cyt, *m*/*z* 263.1 $[M + 1]^+$ to *m*/*z* 152.1 $([pyridine\text{-}D_4]POB])^+$. Prior to HPLC-ESI-MS/MS analysis of the samples the MS parameters were optimized for each deuterated POB−DNA adduct standard. For analysis, the MS parameters were set as follows: Curtain gas, 40 psi; Ion spray voltage, 4 kV; source temperature, 650°C; nebulizer gas (GS1), 70 psi; heater gas (GS2), 70 psi; and collision gas, 12 psi. The fragmentation potentials were optimized for each ion: 299.1 and 303.1, declustering potential (DP), 65 V, entrance potential (EP), 10 V; collision energy (CE), 20 V; collision cell exit potential (CXP), 12 V and for 259.1and 263.1: DP, 40V; EP, 8V; EP, 15V; and CE, 6V. The amount of each POB−DNA adduct was determined by comparing the MS peak area ratio of each adduct to its deuterated standard with a calibration curve. Calibration standards were prepared by spiking different amounts of each adduct with a constant amount of the corresponding internal standard in H_2O and then analyzed by LC-MS/MS without undergoing the sample

preparation procedure described above. The calibration curves were constructed by plotting concentration ratio vs MS peak area ratios of each adduct to its deuterated standard.

DNA Binding

DNA binding was monitored by a change in fluorescence anisotropy as increasing concentrations of protein were added to an oligonucleotide duplex that contained a THF abasic modification in the middle of one strand [d(TGACTACTACAT(THF)GTTGCCTACCAT)] and a 6-carboxyfluorescein (FAM) on the 3′-end of the complementary strand [d(ATGGTAGGCAACTATGTAGTAGTCA)-FAM]. For stoichiometric binding measurements, increasing concentrations of protein (0-200 μM) were added to a solution containing 50 nM FAM-DNA and 20 μM unlabeled 25mer DNA (K_d = 3.1 ± 0.3 μM) in 20 mM Bis-Tris propane pH 6.5, 100 mM NaCl, 2 mM DTT, and 0.1 mM EDTA. Polarized fluorescence intensities using excitation and emission wavelengths of 485 and 538 were measured at 25 °C using a SpectraMax M5 microplate reader (Molecular Devices). Dissociation constants were derived by fitting a two-state binding model to data from three independent experiments.

CHAPTER IV

COMPREHENSIVE ANALYSIS OF DNA BINDING AND RECOGNITION BY AlkD

Summary

Recent crystal structures of alkylpurine DNA glycosylase AlkD bound to alkylated, mismatched, and abasic nucleotides revealed how a HEAT-like repeat protein architecture can use a positively charged concave surface to bind the DNA-phosphate backbone. Once the enzyme is bound to a DNA lesion, it restructures the phosphate backbone to promote nucleotide flipping of both bases in the lesioned pair. However, the precise mechanisms by which AlkD's HEAT repeats anchor the protein to the DNA to promote catalysis remains unknown. Here, we address this gap in knowledge by a comprehensive mutational analysis of DNA binding and 7mG excision activities. The data reveal that while most of the residues mutated are involved in general DNA binding in solution, there are a subset of residues that selectively recognize distorted DNA duplexes. One in particular, tyrosine 27, can also distinguish subtle DNA perturbations like those present in a G•T mismatch. Each mutant also decreases enzymatic activity to less than 30% of wild-type. Altogether, these results suggest that AlkD residues involved in general DNA binding are important for catalytic activity.

Introduction

Prior to the availability of the AlkD/DNA crystal structures, experiments were performed by the Borjas group to assess the ability of mutant AlkD enzymes to functionally complement methyl methanesulfonate (MMS) alkylation sensitivity of *E. coli* cells that lack TAG and AlkA enzymes (Alseth et al, 2006). In a follow up publication by the same group, Dalhus *et al*. determined that while Tyr27Ala, Phe179Ala, Phe180Ala, and Trp187Ala could restore resistance to MMS exposure similar to the wild-type enzyme, mutations of Trp109, Asp113, and Arg148 were unable to complement the sensitivity to MMS (Dalhus et al, 2007). The authors proposed that Trp109, Asp113, and Arg148 were important residues involved in stacking an extrahelical base, stabilizing the phosphate backbone, and acting in the nucleophilic attack on the glycosylic bond. The crystal structures presented in Chapters II and III revealed that these residues reside on the concave, DNA binding surface of the enzyme. *In vitro* DNA binding studies from our lab show a modest increase in binding affinity for Asp113Asn and a decrease in binding affinity for Arg148Ala (Chapter II, Table 3). In addition, our previous DNA binding studies on wild-type AlkD revealed that the enzyme has no preference for mispaired or lesioned bases as it binds duplex DNA containing a 7mG•C base pair with the same affinity as a G•C base pair, a G•T base pair, and a THF•C base pair (Chapter II, Table 3) (Rubinson et al, 2008). The wild-type protein also does not distinguish DNA glycosylase transition state mimics from nonspecific DNA, (Rubinson et al, 2008) unlike traditional base flipping alkylpurine DNA glycosylases that bind much more tightly to 1-azaribose and 1-pyrollidine moieties (Makino & Ichikawa, 1998; O'Brien & Ellenberger, 2004b; Scharer et al, 1998). Taken together, these results

are consistent with an electrostatic interaction between AlkD and the DNA-phosphate backbone.

This study seeks to further examine the overall role of the unique HEAT-like repeat architecture of AlkD and the specific role of individual residues with regards to DNA binding and catalysis in solution. To accomplish this, residues shown to have an interaction with DNA in the crystal structures were mutated and assayed for *in vitro* DNA binding activity as well as *in vitro* 7mG excision activity. Each of the eleven residues mutated had an effect on DNA binding affinity regardless of its position on the concave surface of the protein or the DNA strand it contacted. Furthermore, the mutation of each residue abrogated catalytic activity by at least 60% relative to the wild-type enzyme. Taken together, the results detailed below suggest that general, nonspecific binding of DNA by AlkD is important for the removal of 7mG.

Results and Discussion

General Binding of DNA by AlkD

Residues observed in the crystal structures to be involved in hydrogen bonding or Van der Waals interactions with the phosphate backbone or DNA bases were chosen to be mutated in this study. Although some residues (Arg43, Arg148, and Arg190) most likely contribute to the overall positive charge of the DNA binding surface, others are involved in supporting the DNA backbone through hydrophobic and stacking interactions. To characterize the role of each residue in DNA binding, purified wild-type and mutant AlkD enzymes (Figure 24) were monitored as a change in fluorescence anisotropy as protein was added to an oligonucleotide duplex that contained either a G•C

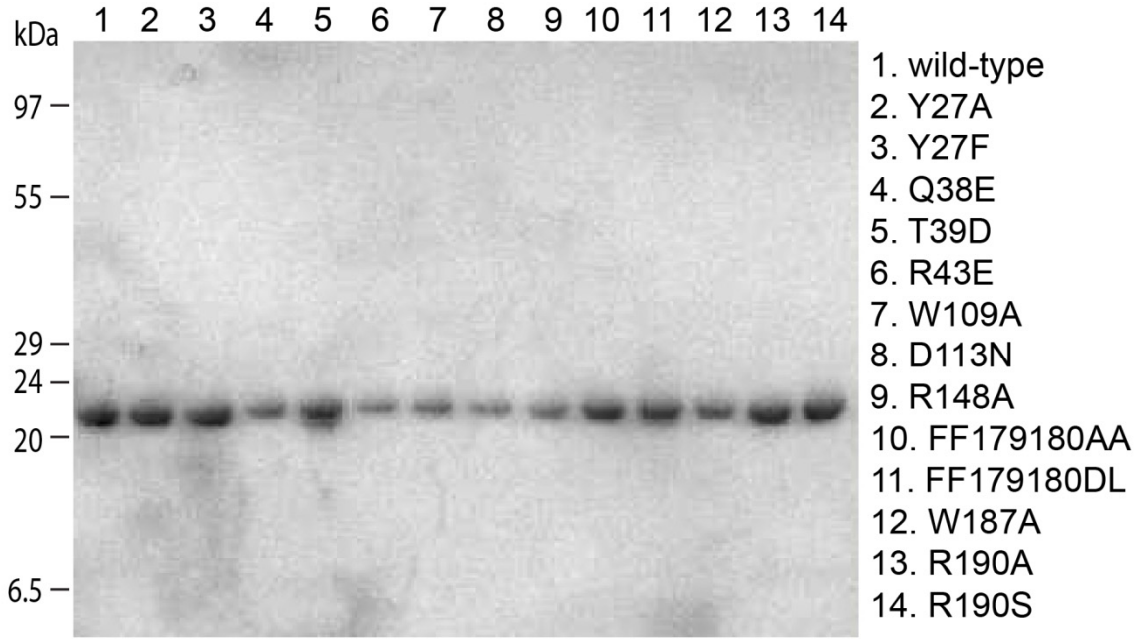


Figure 24. SDS-PAGE of purified wild-type and mutant AlkD proteins.

base pair (unmodified control), a mismatched G•T base pair, or a modified abasic site THF•C base pair. As pre dicted for an enzyme that makes mostly electrostatic interactions to DNA, wild-type and mutant AlkD enzymes bind all DNAs tested with weak micromolar affinity (K_d) and consistent with previous results, mutation of Asp113 and Arg148 increase and decrease DNA binding affinity by around two- and three- fold, respectively, as compared to wild-type enzyme (Avery et al, 1944) (Table C1). Unpredictably, Arg190Ala has either no effect, in the case of G•T-DNA, or increases DNA binding two fold, for both unmodified and THF•C-DNA, suggesting that this residue, which affects catalysis (Avery et al, 1944), did not have a substantial effect on binding.

Overall, most mutations modestly but significantly disrupted DNA binding to the modified product (THF•C) DNA duplex by as much as four fold and with the exception of Tyr27, they also negatively affect binding, albeit to a lesser extent, to unmodified

(G•C) and mismatched (G•T) DNA as compared to wild-type enzyme (Figure 25). When comparing binding affinity of individual mutants for the different DNAs, three residues (Trp109, Thr39, and Arg43) impact binding of both unmodified and modified DNA, suggesting that they are important for general DNA binding regardless of duplex distortion (Figure 25, Table C1). The reduction in binding by Thr39Asp and Arg43Glu can most likely be attributed to the electrostatic repulsion between the phosphates and the carboxyl side chains. Trp109 stacks against the sugar of the base 3′ to the thymine opposite the lesion in both the AlkD-3d3mA•T and AlkD-THF•T structures which explains why mutating this residue to an alanine negatively effects the binding affinity of both unmodified and modified DNA (Figure 25b). These results support the observations made from the crystal structures, namely, that while these residues are crucial for overall DNA binding, none sense the deformations in the product DNA (Figure 25b).

Selective Binding of Product DNA by AlkD

Conversely, many other residues (Asp113, Arg190, PhePhe179180, Gln38, Trp187, and Arg148) discriminately bind product DNA since, when mutated, binding of THF•C-DNA is significantly decreased over unmodified or mismatched DNA (Figure 25a, Table C1). Despite their different locations on the protein surface, each of these residues is shown in the crystal structures to directly interact with the distorted product DNA (Figure 25b,c). Asp113, Arg148, Trp187, and Arg190 are all involved in creating the interface between AlkD and the lesioned base pair. Both Arg148 and Trp187 are more intimately involved with binding the phosphate and sugar, respectively, of the base opposite the 3d3mA or THF lesion (either a thymine or cytosine) in the crystal structures.

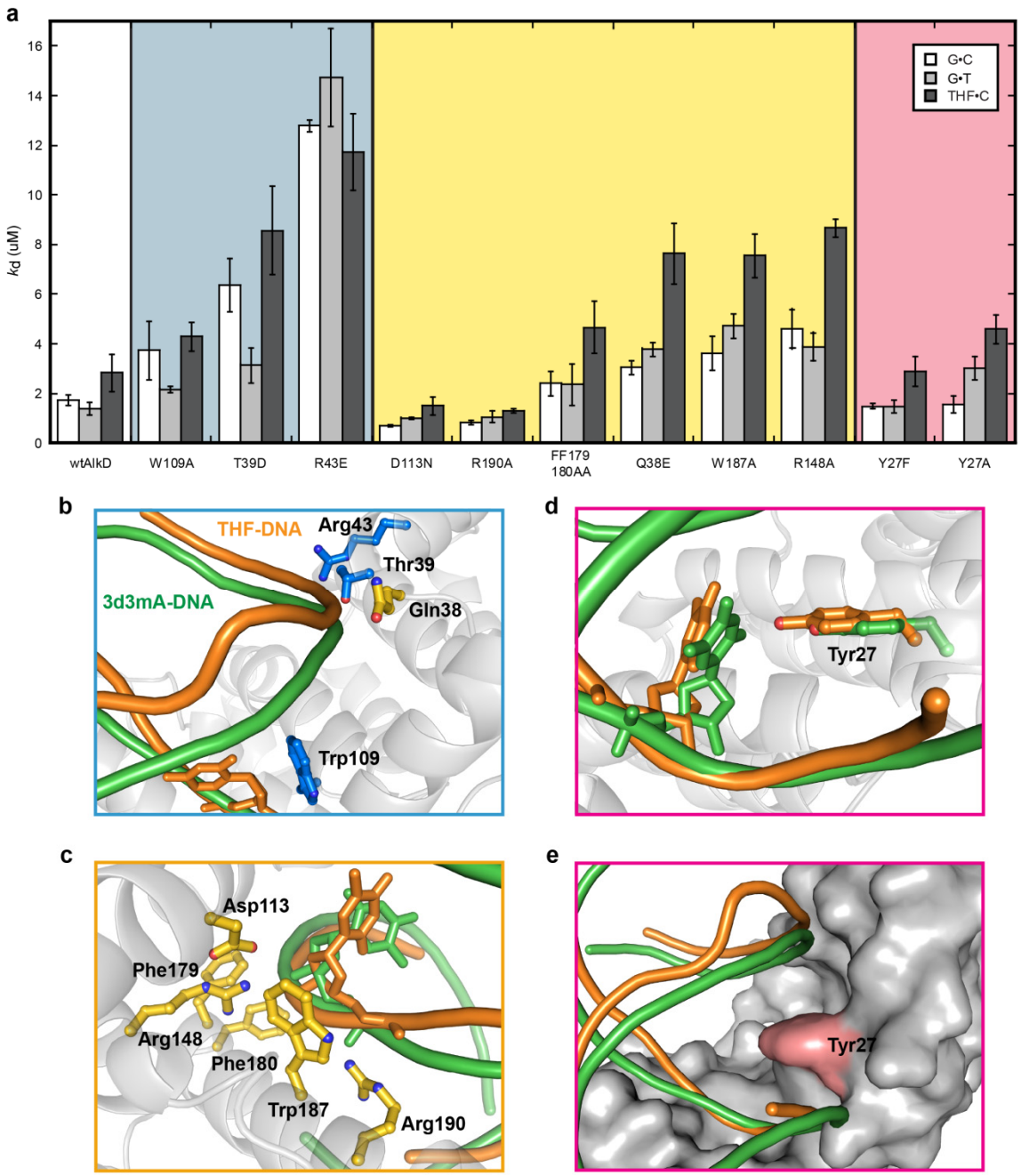


Figure 25. **a**, Bar graph of dissociation constants (K_d) for a 25mer oligonucleotide duplex containing a G•C (white), G•T (light gray), and THF•C (dark gray) base pair. Error bars represent the standard deviation from three independent measurements. **b**,**c**,**d**,**e**, Four different views of a superposition of 3d3mA•T (green) and THF•T (orange) DNA complexes on protein from AlkD/THF•T-DNA. Residues involved in general DNA binding are shown as blue sticks (**b**), while residues involved in selectively binding to distorted DNA are shown as gold sticks (**c**). **d**, Residue Tyr27 is shown as sticks and colored to match the DNA. **e**, Surface representation of AlkD with Tyr27 highlighted pink.

In fact, Trp187 does not make any direct contacts to the substrate- or mismatched-DNA duplexes in the crystal structures (Chapter III, Figure 19b). Glutamine 38 and phenylalanines 179 and 180 are located farther away from the lesioned base pair. Despite its distant location, Gln38Glu has the largest effect on product DNA binding. This can be explained by comparing the substrate- and product-bound AlkD structures since Gln38 makes more specific contacts to the phosphate backbone when the enzyme is bound to the abasic site structure (Chapter III, Figure 19). Although the phenylalanines make no direct contacts to the DNA, these residues create a hydrophobic pocket that stacks against the DNA backbone one to two residues 3′ to the nonlesioned base-pair and thus PhePhe179180AlaAla has a more modest effect on selective product DNA binding. On the whole, these residues appear to play an important role in selective recognition of gross duplex distortion, although they do not appear to distinguish between the altered base-pairing of a G•T mismatch versus unmodified DNA.

Recognition of Duplex Distortion by Tyr27

The crystal structures reveal that tyrosine 27 is the only residue that makes specific contacts to a DNA base (Figure 25d). Interestingly, mutating this residue to a similarly bulky phenylalanine does not affect binding to any DNAs tested compared to wild-type enzyme and further mutating the residue to an alanine has no effect on binding unmodified (G•C) DNA (Figure 25a, Table C1). However, the Tyr27Ala mutation does significantly affect binding to both G•T-DNA and THF•T-DNA by two- and three-fold, respectively. The decrease in binding to product DNA can be explained by the crystal structures of AlkD bound to DNA. Tyrosine 27 is located on helix αB which is the only

noticeable difference in the protein between the substrate and product bound structures (Chapter III, Figure B4). Consequently, this residue lies closer to the DNA duplex in the AlkD/THF-DNA structure than in the AlkD/3d3mA-DNA or AlkD/G•T-DNA structures (Figure 25d). The surprising difference this residue has on G•T versus unmodified DNA binding suggests that this bulky residue may play an important role in recognition of subtle intrahelical duplex distortions, like those created by lesioned bases, through its interaction with the minor groove (Figure 25e).

7mG Excision by AlkD Mutants

The effect of each of these mutations on 7mG excision activity was determined by measuring the alkaline cleavage of the abasic DNA product of lesion excision from a 25mer oligonucleotide duplex containing a 7mG•C base pair. In the course of performing these experiments it was discovered that the enzyme displays sequence dependence. Prior to this mutational analysis study, the glycosylase assay was performed with a DNA containing a 7mGT/AC base step. For the present analysis, the duplex was remade to contain a 7mGA/TC base step to more closely match DNAs used in the crystal structures. A striking difference was discovered in a side-by-side comparison between the two different DNA contexts: wild-type AlkD and Asp113Asn remove 7mG from the duplex containing the 7mGA/TC base step 11 and 17-fold faster than from the 7mGT/AC containing duplex (Figure 26a). Interestingly, the spontaneous rate of 7mG release is slightly decreased, down to 1.1×10^{-5} s^{-1} from 1.5×10^{-5} s^{-1}, in the 7mGA/TC base step, implying that the 7mG lesion may be more stabilized in this context. This may be explained by the greater stacking associated with GT/AC versus GT/TC base steps, as

a

	k_{st} (10^{-3} s^{-1})		Rate enchancement
	7mGT/AC	**7mG**A/TC	
wtAlkD	2.32	23.7	10.9
D113N	0.0135	0.226	16.7
no enzyme	0.0149	0.0115	0.8

b

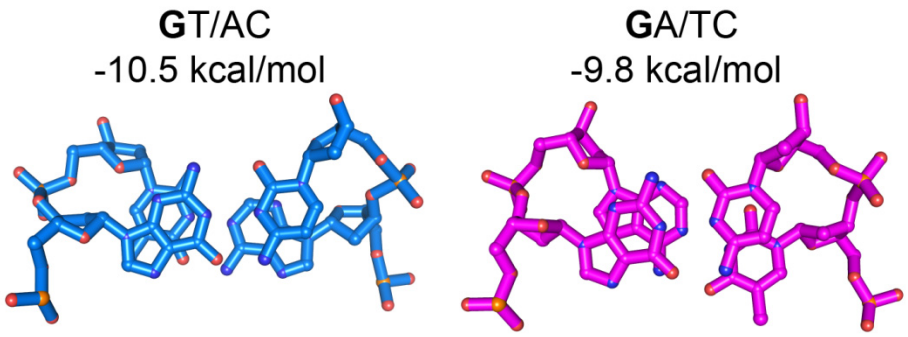


Figure 26. Stacking difference for GA/TC vs GT/AC base steps. **a**, First-order single-turnover rates (k_{st}) for 7mG excision from a 25mer oligonucleotide duplex containing a 7mGT/AC or 7mGA/TC base step. Rate enhancements are calculated as the fold-increase of 7mG excision from the 7mGA/TC base step over 7mGT/AC base step. **b**, Free energy differences for stacking of GA/TC (blue) vs GT/AC (magenta) base steps.

quantified by their free energy values, -10.5 kcal/mol for a GT/AC step versus -9.8 kcal/mol for a GA/TC step (Figure 26b) (Singer & Grunberger, 1983). In order to directly compare the single-turnover excision rates (k_{st}) all the mutants were tested at equal concentrations for their ability to remove 7mG from 7mGA/TC-DNA (Table 4). Every mutant decreased enzyme activity by no less than three-fold. However Tyr27Phe, Gln38Glu, and Tyr27Ala have the least effect on catalysis (Table 4). The remainder of the mutants tested (Thr39Asp, Arg43Glu, Trp109Ala, Asp113Asn, Arg148Ala, PhePhe179180AlaAla, Trp187Ala, and Arg190Ala) were all less than 3% active. The role of Asp113, Arg148, and Arg190 in creating an optimal protein-DNA binding surface

by distorting the phosphate-backbone of the DNA at a lesioned site has been discussed in Chapter 3 (Avery et al, 1944) and was recapitulated by these experiments. Asp113, which plays a slightly more indirect role, was the most active as compared to Arg148 and Arg190 (Table 4). All of the residues tested disrupt not only catalysis of 7mG but also binding to unmodified DNA, modified DNA, or both, suggesting that nonspecific DNA binding is important for the removal of positively-charged alkylpurine lesions by AlkD.

Table 4. 7-Methylguanine excision activities for wild-type and mutants of AlkD

Enzyme	k_{st} ± s.d. (10^{-3} s^{-1})	Relative Activity
wtAlkD	24 ± 4	1.000
W109A	0.08 ± 0.02	0.003
T39D	0.2 ± 0.1	0.009
R43E	0.7 ± 0.1	0.029
D113N	0.292 ± 0.008	0.012
R190A	0.068 ± 0.009	0.003
R190S	0.1 ± 0.05	0.004
F179A,F180A	0.628 ± 0.006	0.027
F179D,F180L	0.17 ± 0.03	0.007
Q38E	7 ± 2	0.277
W187A	0.31 ± 0.01	0.013
R148A	0.051 ± 0.005	0.002
Y27A	3.1 ± 0.1	0.130
Y27F	9 ± 4	0.392
No Enzyme	0.0008 ± .0002	3.57×10^{-5}

First-order single-turnover rate constants for 7mG excision from a 25mer oligonucleotide duplex containing a 7mG•C base pair. Values represent the averages and standard deviations from three experiments.

Future Directions

While wild-type AlkD does not distinguish between DNAs containing unmodified (G•C), mismatched (G•T), and modified (THF•C) base pairs, specific enzyme point

mutations including Gln38Glu, Arg148Ala, and Trp187Ala, have a greater effect on binding distorted DNA duplexes like those observed in the AlkD/THF•T- and THF•C-DNA crystal structures. This comprehensive mutational analysis not only supports what is observed in the crystal structures but also reveals that AlkD may selectively recognize distorted DNA duplexes and subsequently base lesions that perturb DNA base pairing or base stacking. In addition, Tyr27 appears to discriminate between all three DNAs tested, implying that it may be able to recognize very small structural deformities between a G•T mismatch prior to base extrusion. Residue Tyr27 may act as a probe to detect DNA distortion and it is possible to imagine that it "tickles" the DNA ladder as the enzyme processively slides along the DNA duplex. It may be necessary, then, to calculate binding constants for mutant AlkDs with a 7mG•C substrate to determine which residues, if any, are necessary for recognition of a duplex perturbed by the formal positive charge on the lesioned base. If the enzyme's binding energy for grossly distorted DNA is what is driving flipping of the lesioned base, then presumably the 7mG lesion will be flipped out of the duplex as it is predicted from the AlkD/THF-DNA structures. Therefore, residues with preferentially binding to THF-containing DNA are expected to be important for binding toward the 7mG•C containing DNA. In order to prevent catalysis from occurring, these experiments will have to be performed with mutations created in an Asp113Asn or Arg148Ala background.

Since the binding energy is dominated by electrostatic interactions between AlkD and the DNA-phosphate backbone, any specific DNA binding by the wild-type enzyme is lost. One way to reveal specificity is to determine the real-time dissociation rate constants (k_{on} and k_{off}) for wild-type AlkD using stopped-flow fluorescence techniques.

This experiment is performed by pre-incubating the enzyme with fluorescently labeled DNA ligands and monitoring the change in fluorescence intensity after adding an excess of unlabeled DNA (Allan et al, 1998a; Drohat et al, 2002; Hiller et al, 2003; Jiang et al, 2002). AlkD is hypothesized to have increased off rates for both 7mG•C and G•T containing DNAs since they would be recognized equally as structural perturbations. Furthermore, the off rate for a THF containing DNA may be increased over potential substrates since more contacts are created between the enzyme and DNA after product release. However, since almost all the interactions between AlkD and DNA are electrostatic in nature, as suggested by the crystal structures, it is possible that k_{off} will remain equal for each DNA.

As a corollary to these experiments, it will be interesting to investigate the mode of base-flipping promoted by AlkD in solution via time-resolved fluorescence spectroscopy. This technique takes advantage of the fluorescent properties of 2-aminopurine (2-AP) and its sensitivity to its molecular environment (Stivers et al, 1999). Quenched by the DNA duplex, 2-AP fluorescence is increased when base-pairing or base stacking is disrupted, presumably upon binding by the enzyme. By placing the 2-AP moiety in place of the lesion, across from the lesion, and adjacent to the lesion we can observe changes in fluorescence when bound to AlkD. Since both the lesion and the base opposite the lesion are in an extrahelical conformation upon enzyme binding, when the 2-AP is positioned across from a guanine or a 7mG base it should fluoresce after incubation with AlkD. A burst in fluorescence may also be observed when the 2-AP is placed next to either the lesion or the base opposite the lesion since base-stacking is perturbed upon enzyme binding, however the fluorescence should remain quenched when 2-AP is placed

far away from the lesion in the presence and absence of AlkD because it will always remain in the DNA duplex.

Methods

AlkD Mutant Purification

The AlkD gene was PCR amplified from *B. cereus* genomic DNA (ATCC 14579) and cloned into a pET27 (Novagen) derived expression vector (pBG103, Vanderbilt Center for Structural Biology) that produces a cleavable N-terminal His_6-SUMO-fusion protein. *E. coli* HMS174 cells transformed with the AlkD/pBG103 plasmid were propagated in LB media and protein was overexpressed for 3 h at 37°C upon addition of 0.5 mM IPTG. Cells were harvested in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 10% glycerol and lysed with an Avestin Emulsifier C3 homogenizer operating at ~20000 psi. AlkD-fusion protein was purified using Ni-NTA (Qiagen) affinity chromatography, followed by cleavage of the His_6-SUMO tag. AlkD was further purified by heparin affinity and gel filtration chromatography to >99% homogeneity. Protein was concentrated to 12.5 mg/ml and stored in 20 mM Bis-Tris Propane, 100 mM NaCl, and 0.1 mM EDTA. Mutant proteins were prepared by site-directed mutagenesis using a Quik-Change Kit (Stratagene) and purified in the same manner as wild-type AlkD.

Circular Dichroism Spectroscopy

Structural integrity of mutant proteins was verified by far-UV circular dichroism (CD) spectroscopy using a Jasco J-810 CD with attached Peltier temperature control. Table of melting temperatures T_m were derived from fits to the CD data using the

equation $\Theta = 1/(1+e^{(Tm-T)/k})$, where T_m corresponds to the temperature at 50% denaturation and k describes the cooperativity of the transition. Calculation of T_m revealed that while almost all of these mutations retained the fully folded structure of the enzyme at 37 °C, Asp113Asn (T_m = 30.7 °C) is partially unfolded at this temperature (Figure C1). Since the DNA binding assay was performed at 25 °C binding to this mutant should not be affected but to ensure accurate calculation of base catalysis by this mutant, the base excision assays were performed at lower temperatures to rule out inactivity due to the unfolded protein.

DNA Binding Assay

DNA binding of wild-type and mutant AlkD was measured by the change in fluorescence anisotropy as protein was added to an unmodified oligonucleotide duplex (X = G) or a duplex that contained a site-specific modification (X = THF) in the middle of one strand [d(GACCACTACACC***X***ATTCCTAACAAC)] and a 6-carboxyfluorescein on the 3'-end of the complementary strand [d(GTTGTAAGGAATCGGTGTAGTGGTC)-FAM]. Increasing concentrations of protein (0-30 uM for Asp113Asn and Arg148Ala; 0-75 for wild-type AlkD and all other mutants) were added to a 50 nM DNA in 20 mM Bis-Tris Propane pH 6.5, 100 mM NaCl, 2 mM DTT, and 0.1 mM EDTA. Polarized fluorescence intensities using excitation and emission wavelengths of 485 and 538 were measured at ambient temperature using a SpectraMax M5 microplate reader (Molecular Devices). Equilibrium dissociation constants (K_d) were derived by fitting a two-state binding model to data from three independent experiments.

Glycosylase Activity Assay

AlkD glycosylase activity was measured by alkaline cleavage of the abasic DNA product of 7mG excision from a 25mer oligonucleotide duplex containing a 7mG·C base pair. 7mG was enzymatically incorporated into DNA duplexes using the method previously described (Asaeda et al, 2000). Briefly, the primer oligonucleotide 5′-GACCACTACACC was ^{32}P-labeled at the 5′-end, annealed to a 3-fold excess of the complementary strand (5′- GTTGTAAGGAATCGGTGTAGTGGTC), and extended using DNA polymerase I Klenow fragment (New England Biolabs) in the presence of deoxy-7-methylguanosine 5′-triphosphate (d7mGTP), dCTP, dTTP, and dATP. In a 10 µl glycosylase reaction, 2 nM radiolabeled DNA duplex was incubated at 37°C with 20 µM AlkD in 50 mM HEPES pH 7.5, 100 mM KCl, 10 mM DTT, and 2 mM EDTA. The reaction was stopped at various times by addition of 0.2 N NaOH, and heated at 70°C for 10 min. The 12mer product and remaining 25mer substrate DNA strands were separated by denaturing 15% polyacrylamide gel electrophoresis in 7M urea and quantitated by autoradiography.

CHAPTER V

DNA PROCESSING BY VARIANT TANDEM HELICAL REPEAT PROTEINS

Summary

Proteins that contain tandem helical repeat motifs are typically associated with protein-protein interactions or scaffolding functions. Recent structural elucidation of two helical repeat proteins, AlkD and MTERF1, bound to DNA reveals how this structural motif can be used for DNA binding. Here, the details of DNA interactions by the AlkD and MTERF1 structures are reviewed in detail in order to draw general conclusions regarding DNA interactions by tandem α-helical motifs. Although their individual structural units differ, both proteins bind DNA through extensive electrostatic interactions. They are also thought to actively promote extrahelical nucleobases in order to perform their functional roles. Comparison of each structure reveals important similarities in how they are ideally suited to bind DNA but highlights differences in their mode of DNA recognition.

Introduction

Proteins that contain internal tandem α-helical repeats of small structural units are found in all domains of life and thought to evolve through tandem duplication within a gene (Bjorklund et al, 2006). Helical repeats are typically short 25-45 residue motifs that align in tandem arrays or superhelical structures between three and 25 repeating motifs (Groves & Barford, 1999). The individual repeating units are held together by hydrogen bonds created by a conserved hydrophobic core and stack parallel to form extended

solenoid structures (Figure 27) (reviewed in Groves & Barford, 1999). Traditionally, helical repeat motifs have been identified as protein scaffolding domains owing to the large surface area achieved through their structural repetition. Several important eukaryotic cellular proteins contain tandem helical repeats and depending on how each repeat is structured and linked together, specific functionalities emerge for each arrangement (Groves & Barford, 1999). For example, repeat and packing arrangements can dictate a protein's specificity for a particular ligand (Grove et al, 2008). Three-dimensional structures have been reported for several classes of tandem α-helical repeats and extensive reviews have detailed their characteristics (Andrade et al, 2001a; Grove et al, 2008; Groves & Barford, 1999; Main et al, 2003).

There are five typical helical repeat structures, the tetratricopeptide (TPR) repeats, armadillo (ARM) repeats, HEAT repeats, leucine-rich variant (LRV) repeats, and ankyrin (ANK) repeats, are summarized below (Figure 27). TPR proteins are arrangements of two antiparallel α-helices of 34 degenerate amino acids each that form a tandem array (3-16 motifs) which almost always generates a right-handed superhelix with a channel for ligand or protein binding (Blatch & Lassle, 1999; Das et al, 1998). TPR motifs are found in large protein assemblies including protein chaperones, cell-cycle complexes and transcription complexes (Blatch & Lassle, 1999). ARM repeats, found in proteins involved in cell adhesion, like β-catenin, and signaling pathways, consist of a short two-turn α-helix that lies perpendicular to two longer antiparallel α-helices (~3 turns each), stacking into a right-handed superhelix (Andrade et al, 2001b; Huber et al, 1997). HEAT (Huntington/Elongation/A subunit/Target of rapamycin) motifs, often utilized for protein-

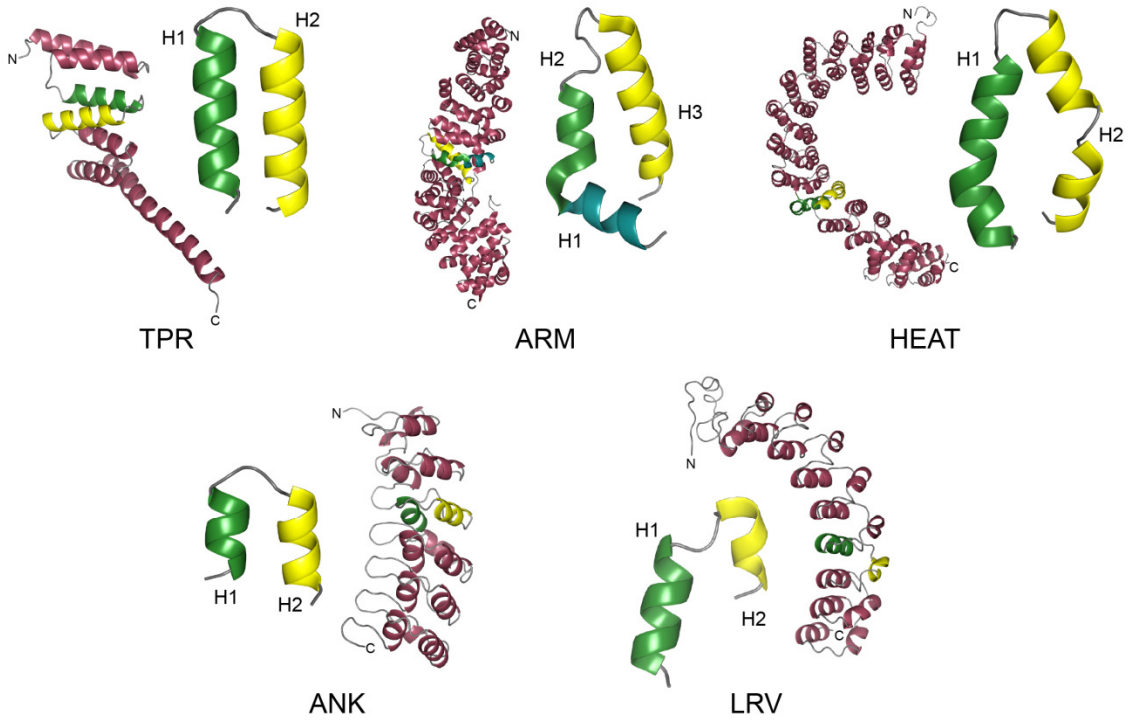


Figure 27. Tandem helical repeat proteins. Ribbon representations of five classical repeat structures and their corresponding individual repeating unit colored by helix: tricopeptide repeats (TPR) from protein phosphatase 5 (PDB ID 1a17) (Das et al, 1998), armadillo (ARM) repeats from β-catenin (PDB ID 1jdh) (Graham et al, 2001), HEAT repeats from protein phosphatase 2A (PP2A) (PDB ID 1b3u) (Groves et al, 1999), ankyrin (ANK) repeats of IκBα (PDB ID 1nfi) (Jacobs & Harrison, 1998), and leucine-rich variant (LRV) motif (PDB ID 1lrv) (Peters et al, 1996).

protein interactions within nuclear transport molecular assemblies or protein synthesis complexes or found in proteins involved in chromosome dynamics, are characterized by helical arrays of two antiparallel alpha helices averaging 40 amino acids in length (Groves et al, 1999). The first helix in a HEAT repeat usually contains a kink, mimicking the first two helices of ARM repeats (Andrade et al, 2001a). Several conserved hydrophobic residues line the core of both ARM and HEAT repeats (Andrade et al, 2001b). A variant of the leucine-rich repeat (LRR), LRV, is very similar to the HEAT motif and they pack in a similar way. In LRRs, an α-helix is linked to a β-strand

and they stack so the β-strands line the concave surface. In the case of LRVs, its motif is a short 3_{10} helix packed against a three turn α-helix that superimposes onto the C-terminal end of H1 and N-terminal end of H2 of a HEAT motif (Groves & Barford, 1999). Each motif is stacked in parallel and twisted relative to each other so that they form a curved structure. The mixed α/β structure of ANK repeats is highly conserved and the motif is the most common protein-protein interaction motif in nature. Thirty-three amino acid ankyrin repeats are two antiparallel α-helices joined by a perpendicular β-hairpin/loop forming an L-shape which packs into an antiparallel β-sheet where proteins bind (Main et al, 2003). Binding interactions may also take place along the α-helices that line the concave surface of proteins that contain this motif. Stacked in parallel, the motifs create a left-handed twist (Main et al, 2003).

Comparing the different topological folds created by a pair of α-helices in each of these proteins described above illustrates how each provides a unique interface for a particular protein binding partner. Additionally, some repeat proteins have been identified to bind molecules other than proteins, including RNA. Pumilio is the founding member of the Puf family of proteins from *Drosophila melanogaster* that act as translational repressors of certain genes by binding specific mRNA sequences. Structures of Pumilio in the absence and presence of an RNA ligand revealed that the protein is made up of repeating α-helical units similar to ARM repeats (Edwards et al, 2001; Wang et al, 2002; Wang et al, 2001). In this case, residues that line the concave surface of the protein make direct contacts to the RNA bases, which are also intercalated between amino acid side chains from individual repeats. The phosphate backbone is almost exclusively exposed to solvent (Wang et al, 2002). The HEAT repeats of Ro autoantigen

have also been shown to bind misfolded, small RNAs (Stein et al, 2005). Other proteins, including the ARM-like repeat protein human Rcd-1, have been shown biochemically to bind nucleic acids. Within the last year, however, the structures of two proteins made up of variant tandem helical repeat motifs were determined and shown for the first time to bind DNA (Avery et al, 1944; Jimenez-Menendez et al, 2010; Yakubovskaya et al, 2010). Presented in this paper is a comparison of the recent crystal structures of two nucleic acid binding proteins, AlkD and MTERF1, in complex with DNA, which allows a first glimpse at how helical repeat proteins use tandem arrays of α-helices to bind, recognize, and process DNA.

AlkD and MTERF1 are Variant Tandem Helical Repeat Proteins

Bacillus cereus AlkD is an alkylpurine DNA glycosylase initially identified as a functional complement to *E. coli* AlkA and shown to be involved in the cellular response to methylating agents (Alseth et al, 2006). AlkD and its homolog AlkC have no sequence similarity to any known protein and are specific for the removal of alkylated bases (3mA, 3mG, and 7mG) in the base excision repair (BER) pathway (Alseth et al, 2006). The atomic resolution crystal structure of AlkD and a computational study revealed AlkD belonged to a novel structural superfamily of DNA glycosylases (Dalhus et al, 2007; Rubinson et al, 2008). AlkD is composed exclusively from HEAT-like repeats which form into a small, singular solenoid domain with a positively charged concave surface (Figure 28a). Thirteen α-helices form into six repeating antiparallel α-helices (αA/αC,

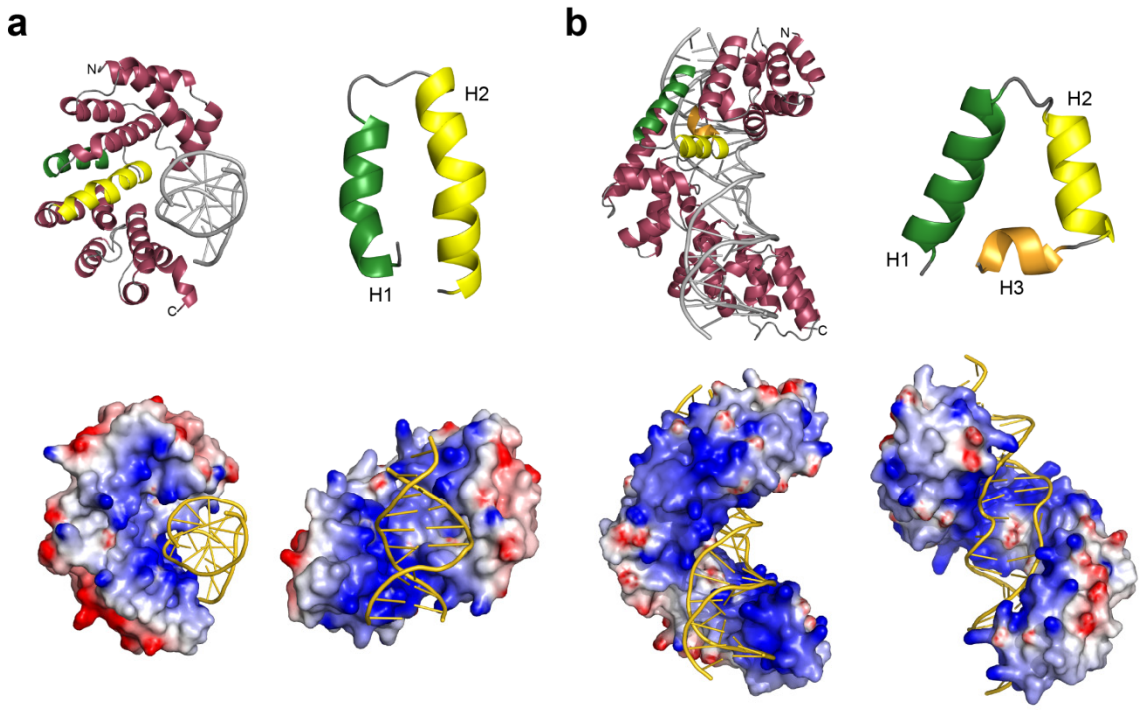


Figure 28. Crystal structures of AlkD (3jxz) (**a**) and MTERF1 (3mva) (**b**) in complex to DNA. The top panel shows ribbons representations of each protein bound to DNA (silver) and their corresponding individual repeating unit colored by helix. The bottom panel shows two orientations of solvent accessible surface representations for each protein colored according to electrostatic potential (red negative, blue positive, -7 to +7 k_BT) bound to DNA (gold).

αD/αE, αF/αG, αH/αI, αJ/αK, and αL/αM), around 35 residues each, similar to tandem HEAT repeats, with the N-terminal repeat containing an α-helical insertion (αB helix) (Figure 28a) (Rubinson et al, 2008). The pairs pack in a parallel arrangement so that a ladder of the N-terminal helices (H1) forms the convex surface of the protein while the concave surface is lined with the C-terminal helix (H2) of each repeat. Despite the conservation of hydrophobic resides that form the core of each repeat, AlkD's repeats are structural variations of canonical HEAT repeats. They are approximately one helical turn

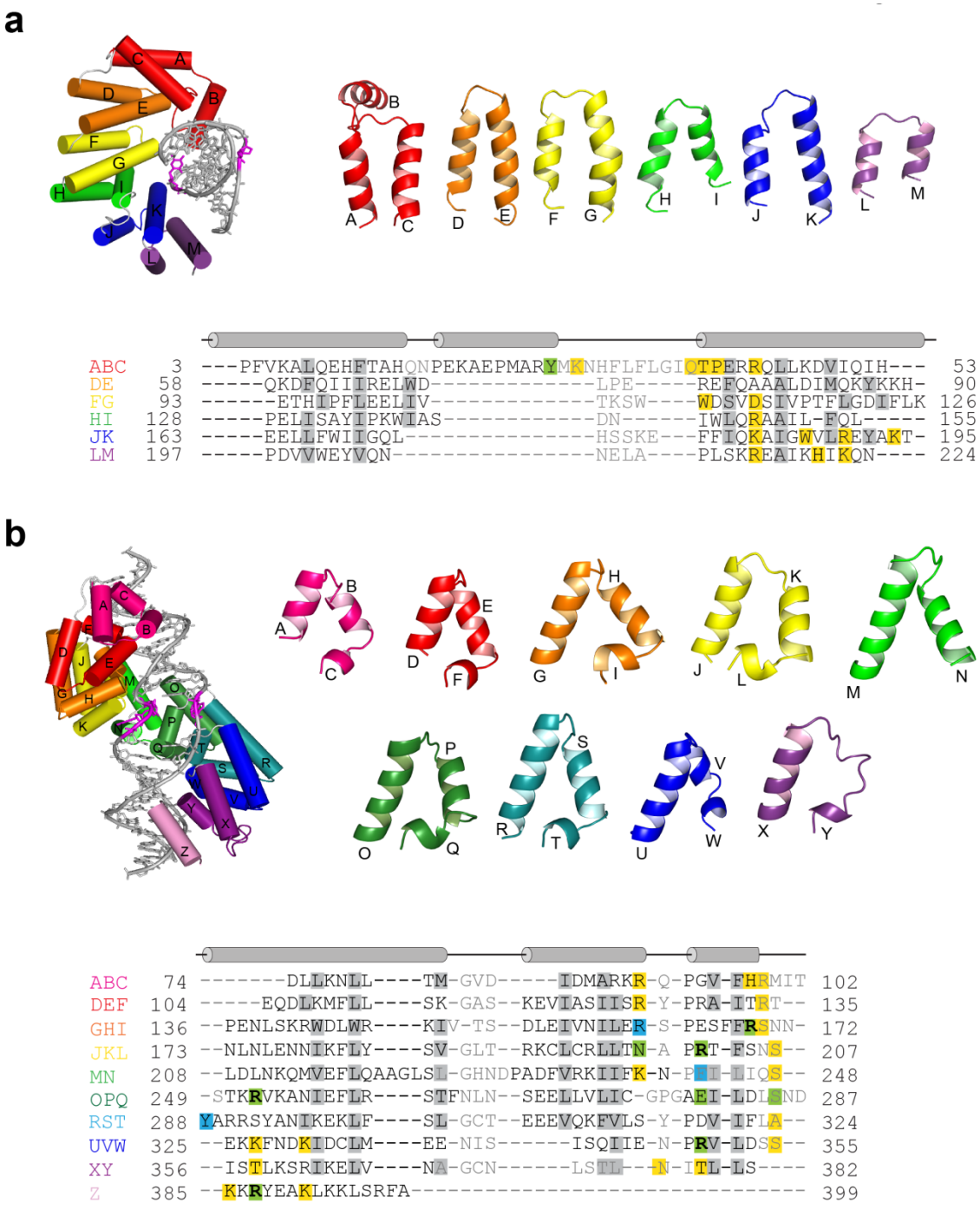


Figure 29. The variant HEAT motifs of AlkD (**a**) and terf repeats of MTERF1 (**b**). On the left are cylindrical representations of each protein colored by repeat which are then represented as individual ribbons colored independently. On the bottom are structure-based sequence alignments of individual repeats, with hydrophobic residues that line the interface between helices highlighted gray, residues that contact the phosphate-backbone are highlighted yellow, base binding residues are colored green, and MTERF1 residues involved in stacking flipped bases are colored cyan. MTERF1 guanine residues involved in sequence recognition are in boldface.

shorter and lack the characteristic kink found in the N-terminal helix (Figure 29a). Unlike other HEAT repeat proteins, several positively-charged basic residues line the concave groove of AlkD and are conserved among the repeats as well as AlkD orthologs (Figures 28a and 29a).

The MTERF family of proteins (MTERF1-4) has been shown to be regulators of eukaryotic mitochondrial gene transcription and replication. While the roles of human MTERF1, a transcription termination factor, and MTERF3, a negative regulator of transcription, have been elucidated, the specific functions of MTERF2 and MTERF4 remain elusive (reviewed in Roberti et al, 2009). MTERF1 has been shown both to bind a sequence specific 28 nucleotide region near the 3' end of the 16S rRNA gene and promote transcription termination using a reconstituted transcription system (Asin-Cayuela et al, 2005; Kruse et al, 1989). MTERF3, on the other hand, appears to have no sequence specificity but has been shown to interact with the mitochondrial promoter region (Park et al, 2007). Initially, sequence characterization of this family of proteins suggested that they used leucine-zipper motifs to bind DNA but recent structures of MTERF3 alone and MTERF1 complexed to DNA reveal the proteins to be single domains made up of repetitions of a left-handed variant ARM motif, which the authors term "mterf" or "Zurdo" motif (Jimenez-Menendez et al, 2010; Yakubovskaya et al, 2010). MTERF1 contains nine structural TERF repeats (αA/αB/αC, αD/αE/αF, αG/αH/αI, αJ/αK/αL, αM/αN, αO/αP/αQ, αR/αS/αT, αU/αV/αW, and αX/αY) made up of around 30 residues that form two antiparallel α-helices (H1 and H2) separated by a three residue loop, followed by a smaller 3_{10} helix or a short α-helix with 3_{10} characteristics (H3) (Figure 28b). One repeat (αM/αN), in the middle of the sequence, is lacking the 3_{10}

helix but is longer overall and separated by five residues. The most C-terminal repeat (αX/αY) lacks the H2 α-helix (Figure 29b). The protein has both an N-terminal segment and a C-terminal helical extension (Jimenez-Menendez et al, 2010). A moderately conserved glycine residue between H1 and H2 confers, unexpectedly, a left-handed character to each repeat and a conserved proline residue at the start of H3 causes the C-terminal helix to lie perpendicular to the antiparallel H1 and H2 helices giving each repeat a triangular shape similar to that of ARM repeats (Jimenez-Menendez et al, 2010). The concave surface of the protein is formed by the H3 helix while H1 and H2 line the convex side of the protein. Like AlkD, the core of each repeat in MTERF1 is lined with hydrophobic residues (Figure 28b), which provides rigidity to each individual repeat while flexibility is retained between units.

DNA Binding and Base-Flipping

Structures of AlkD in complex with DNAs containing a variety of centrally-located modified nucleotides revealed that DNA binds along the positively charged inner surface through electrostatic interactions (Avery et al, 1944). The HEAT-like repeats of AlkD distort the DNA backbone to detect non-Watson-Crick base pairs without duplex intercalation (Avery et al, 1944). AlkD bound to DNA containing an abasic site showed that this enzyme captures modified bases by flipping opposing bases out of the duplex (Figure 30a). The structure of MTERF1 bound to DNA containing the termination sequence illustrated how the protein also uses these unique motifs to bind to DNA and recognize its target sequence by melting the DNA and capturing extrahelical bases (Figure 30b) (Yakubovskaya et al, 2010). Both AlkD and MTERF1 bind DNA through

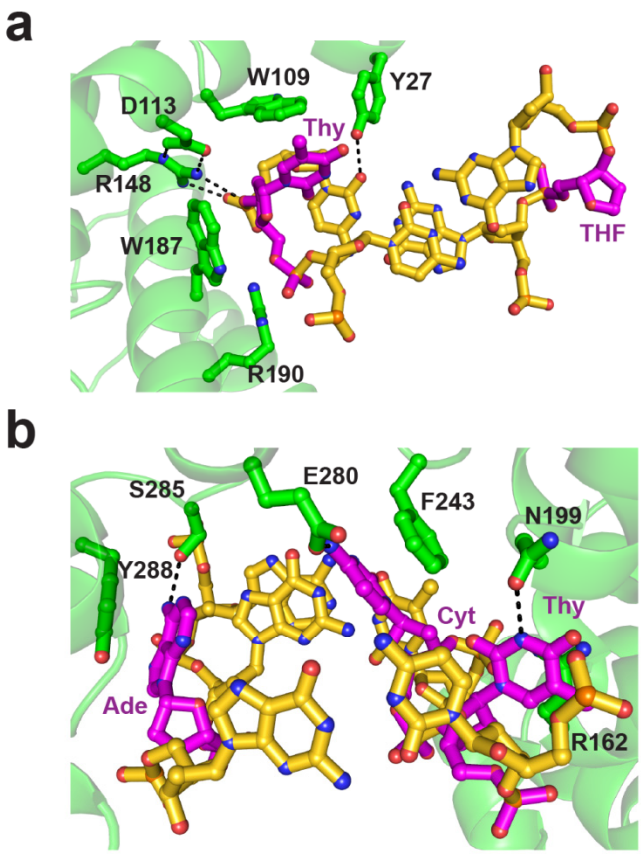


Figure 30. DNA binding pockets of wild-type AlkD/THF-DNA (**a**) and MTERF1-DNA (**b**). **a**, Orthogonal view of AlkD protein (green) wrapped around the DNA duplex (gold) showing the THF•T base pair (magenta) is flipped out of the duplex. **b**, Orthogonal view of MTERF1 protein (green) wrapped around the DNA duplex (gold) where the adenine, cytosine, and thymine (magenta) have been flipped out of the duplex. Although neither protein intercalates residues into the space left by the flipped bases, in the case of AlkD the duplex collapses on itself to maintain base stacking properties. Important DNA recognition or base stacking protein residues are shown as green sticks. Hydrogen bonds are shown as dashed lines.

electrostatic interactions between positively charged residues through the C-terminal helices of each repeat. In AlkD, H2 stacks along the DNA while in MTERF1, the short H3 creates the binding surface. While AlkD interacts with the minor groove of DNA, MTERF1 interacts with the major groove of DNA and this difference appears to be the result of the handedness of each repeat. The right-handed repeats of AlkD form a left-handed superhelix around double-stranded DNA and MTERF1 is made up of left-handed

repeats which wrap a right-handed superhelix around the DNA (Figure 31). With AlkD, the six helical repeats make a half turn around 12 DNA bases with a pitch of around 44 Å (88 Å for a full turn) and all repeats but αD/αE contribute to DNA binding. All nine bundles of MTERF1 supply DNA binding contacts and they make almost a complete full turn around the DNA duplex with a footprint of 22 nucleotides and a pitch of around 70 Å.

When bound to AlkD, the DNA essentially retains its B-form structure in both complexes resembling the substrate and product conformations of the DNA glycosylase reaction, with the exception of the lesioned base pair. In the substrate structure the base pair that contains the lesion is sheared, resulting in a loss of hydrogen bonding as the base

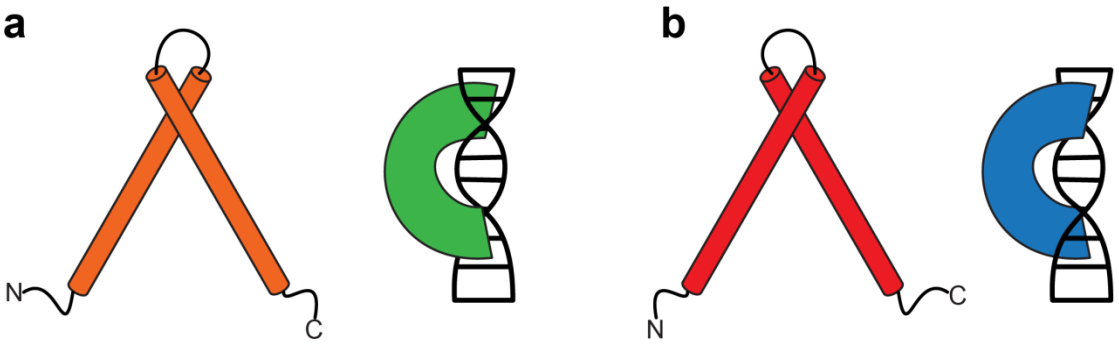


Figure 31. Helical repeats. **a**, AlkD's right-handed motifs form a left-handed superhelix around DNA. **b**, The left-handed repeats of MTERF1 form a right-handed superhelix around DNA.

opposite the lesion is rotated into the minor groove. In the product complex this base is completely displaced from the duplex while the lesioned base is also rotated around the phosphate backbone into an extrahelical, solvent exposed orientation. To stabilize both flipped bases the duplex collapses in order to retain base stacking interactions. In both

AlkD structures the phosphate backbone is distorted and stabilized by optimal protein hydrogen bonding interactions from Asp113, Arg148, and Arg190 (Figure 30a). Mutation of these residues leads to decreased catalytic activity by the enzyme. MTERF1 also binds DNA in a B-form structure aside from a slight 25° bend created upon binding in the central portion of the duplex. Here, MTERF1 decreases the twist of the DNA and unwinds the duplex (Yakubovskaya et al, 2010). In the case of the wild-type protein, three nucleotides are rotated out of the duplex and are stabilized by stacking interactions between residues Arg162, Phe243, and Tyr288 (Yakubovskaya et al, 2010) (Figure 30b). Mutating these residues and disrupting these stacking interactions prevents complete base-flipping of the residues. However, the helical unwinding still exists, suggesting that it is independent of nucleotide eversion and that the wild-type protein actively base-flips three nucleotides (Yakubovskaya et al, 2010). Similar to AlkD, MTERF1 does not intercalate residues into the space left by the flipped bases, instead the unwound duplex is simply maintained through other binding interactions, while in the AlkD-DNA structures the duplex collapses on itself in order to preserve stacking interactions.

DNA Recognition

Nucleic acid recognition by tandem helical repeats may be elucidated by the crystal structures of AlkD and MTERF1. Although both AlkD and MTERF1 bind DNA in a similar manner, through a series of electrostatic interactions made between positively charged protein residues and the negatively charged phosphate backbone, each appears to have a unique mode of DNA recognition. AlkD does not recognize a particular sequence context since almost all of the interactions between AlkD and DNA are to the phosphate

backbone as opposed to the nucleobases. AlkD is proposed to detect the thermodynamic differences between unmodified and modified DNA, which creates altered stacking and/or pairing of non-canonical base pairs (Avery et al, 1944). The structure of AlkD bound to a G•T mismatch, for which it has no activity, provides a structural basis for lesion recognition. Once AlkD traps a lesion or non-canonical base pair it restructures the DNA phosphate backbone around the base pair to promote flipping of each nucleotide into the extrahelical space to create optimal protein-DNA interactions at the capture site, which when disrupted leads to decreased lesion excision rates (Avery et al, 1944).

Despite the fact that the majority of DNA binding by MTERF1 occurs via electrostatic interactions, sequence recognition is thought to come about through five arginine residues that make specific contacts to guanine bases in the protein's singular termination sequence (Yakubovskaya et al, 2010) (Figure 32). This is despite evidence that not all individual arginine mutants lose specific binding (Yakubovskaya et al, 2010). The importance of these residues in maintaining sequence specificity is not equal to their role in tight DNA binding, however, as each arginine is necessary for transcriptional termination activity (Yakubovskaya et al, 2010). Unexpectedly, structures of MTERF1 with mutations in stacking residues revealed that sequence recognition is not a result of nucleotide flipping. However, this mutant is unable to promote termination, suggesting that base-flipping is required for the protein to function. Taken together these results suggest that DNA binding affinity is correlated to protein activity for both AlkD and MTERF1 and that base-flipping is also a critical step in the reaction.

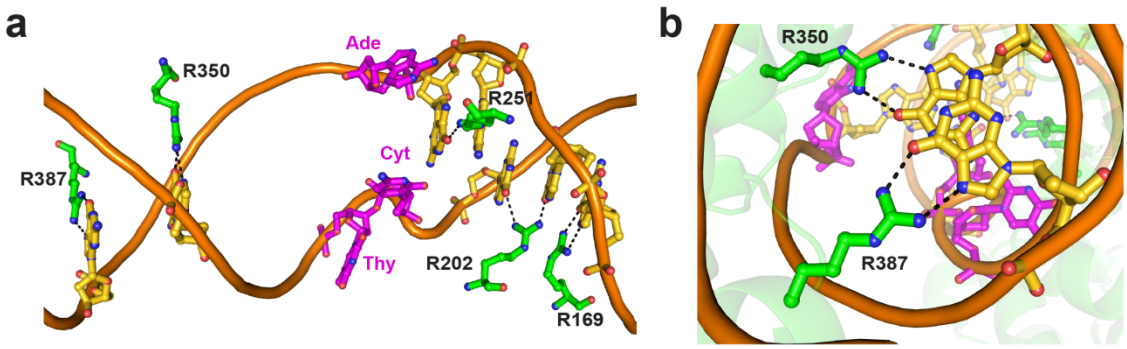


Figure 32. DNA recognition by MTERF1 arginine residues. **a**, Five arginine residues (green sticks) determine sequence recognition by binding to specific guanine residues (gold sticks). **b**, Close-up of two arginine-guanine partners. Flipped bases are shown as magenta sticks. Hydrogen bonds are shown as dashed lines.

Conclusions

Structural studies of AlkD and MTERF1 bound to specific DNAs illustrate the first examples of how these proteins utilize variant tandem helical repeat structures to bind and manipulate DNA. Despite their structural differences the repeats of both AlkD and MTERF1 have evolved to recognize DNA through electrostatic interactions between positively charged protein residues and the negatively charged duplex backbone. Initially, each protein may probe the DNA randomly and can interact with DNA in a lesion or sequence independent manner. Correct recognition relies on different mechanisms for each protein. For AlkD, the small stacking disruptions created by a lesioned base are presumably recognized by the enzyme, followed by rearrangement of the DNA backbone to force the lesion out of the DNA duplex. MTERF1 is thought to halt transcription by base-flipping three nucleotides, although this step is thought to be independent of DNA binding and bending. Each protein distorts the helical backbone of DNA to promote the extrahelical orientation of specific DNA bases. Interestingly, neither AlkD nor MTERF1 intercalate residues into the space left by the flipped

nucleotides, a property observed in virtually all other base-flipping enzymes. The Yakubovskaya group proposes that base-flipping may occur despite the sequence context if the DNA duplex is previously distorted (Yakubovskaya et al, 2010). If this concept is true then both proteins may recognize structural perturbations and use this to promote base-flipping to perform their biological function.

Other proteins have also been predicted to use variant repeat structures to bind DNA. Noticeably, AlkC, a homolog of AlkD is predicted to have a similar HEAT-like repeat structure (Dalhus et al, 2007) and the other MTERF family members have also been shown to contain "terf" repeats, verified in part by a partial crystal structure of MTERF3 (Spahr et al, 2010). Although they have yet to be structurally characterized, HEAT repeat motifs in particular have been identified in proteins involved in chromatin-remodeling, including condensins and cohesions, and in the DNA-damage response, like ATM, ATR and DNA-PK (Neuwald & Hirano, 2000; Perry & Kleckner, 2003). Recently, a low-resolution crystal structure of DNA-PKcs suggested the protein uses HEAT repeats motifs to bind DNA (Sibanda et al, 2010). Apart from discovering more proteins and enzymes that bind DNA through tandem helical repeats there is potential to design *de novo* proteins with these properties. The successful design of tandem helical repeat proteins based on TPR, ANK, LRR/LRV, ARM, and HEAT motifs using consensus design has flourished in the past ten years (Forrer et al, 2004; Hausrath & Goriely, 2006; Main et al, 2003; Parmeggiani et al, 2008; Stumpp et al, 2003; Wetzel et al, 2010). These techniques have led to the creation of proteins that bind other proteins and/or peptide ligands. The knowledge gleaned from the AlkD and MTERF1 structures about how they use their repeat architecture to bind and process DNA could potentially

be integrated into the design of new proteins which would be able to recognize and bind DNA.

CHAPTER VI

SIMULTANEOUS DETECTION OF MULTIPLE METHYLPURINE ADDUCTS IN DNA RELEASED BY ALKYLPURINE DNA GLYCOSYLASES USING HPLC-TANDEM MASS SPECTROMETRY

Summary

This work describes a method for the simultaneous detection and quantitation of five alkylpurine adducts released from *N*-methyl-*N*-nitrosourea (MNU)-treated DNA upon exposure to alkylpurine DNA specific glycosylases. Separation of adducts by reversed-phase HPLC was followed by detection with electrospray tandem mass spectrometry by multiple reaction monitoring (MRM). The limits of detection were determined to be approximately 100 fmol for 1mA, 3mA, 7mA, 7mG and O^6mG. The lower limits of quantitation were found to be approximately 500 fmol for 1mA, 3mA, 7mA and 7mG and 100 fmol for O^6mG. The calibration curves for all calibration standards were linear over the concentration range of 0.5-50 pmol with all correlation coefficients (R^2) over 0.95. Preliminary adduct profiles for *E. coli* TAG, *B. cereus* AlkD and AlkC, and *S. cerevisiae* MAG have been determined.

Introduction

Owing to the extreme lability of *N*3-methyladenine (3mA) in DNA, incorporating a single, centrally located 3mA lesion on a short oligonucleotide to be used for alkylpurine glycosylase activity assays is problematic. Instead, the enzymatic properties of 3mA DNA glycosylases are traditionally investigated by examining the abilities of purified enzymes to remove alkylated bases from genomic DNA treated with a

radiolabeled laboratory methylating agent such as *N*-[^{3}H]-methyl-*N*-nitrosourea (MNU), followed by separation of the radiolabelled excision products by HPLC (Boiteux et al, 1984; Karran et al, 1982; McCarthy et al, 1984; O'Connor et al, 1988; O'Rourke et al, 2000). The advantage of this more traditional method is the simultaneous quantitation of all methyl bases liberated by glycosylase activity, but suffers from laborious scintillation counting and the use of radioactivity. In order to eliminate the need for expensive radioactive material and to provide a more comprehensive analysis of substrate specificity by these enzymes, we have begun to develop a mass spectrometric method to simultaneously detect five methylated bases excised by alkylpurine DNA glycosylases. Reverse-phase HPLC separation of nucleobase products is followed by positive electrospray [ESI(+)] mass spectrometric detection in multiple reaction monitoring (MRM) mode.

Several analytical techniques have been used to quantify the release of oxidative or methylated adducts from DNA including immunoassays, HPLC combined with electrochemical detection, gas-chromotography/mass spectrometry (GC/MS), and ^{32}P-postlabeling experiments although each technique has limitations (Cadet et al, 1999; Jaruga et al, 2000; Medeiros, 2009). Expanding on GC/MS techniques, combining liquid chromatography with tandem mass spectrometry (HPLC-MS/MS) has been shown to be highly selective and sensitive for the separation and quantitation of nucleosides and nucleotides (Abdel-Hamid et al, 2000; Cai et al, 2004; Medeiros, 2009; Zhang et al, 2006). This method has previously been used to identify enzymatically released DNA adducts including products of oxidative damage (Cadet et al, 2002; Podmore et al, 2000) and alkylating agents (Chadt et al, 2008). The method designed here will contribute to

these recent studies by comparing the profiles of released adducts for wild-type and, eventually, mutant glycosylases. Once completely validated, this method will help identify the roles specific protein residues play in substrate specificity and catalysis. Method development has been ongoing and the five most common methylpurines formed by reaction of DNA with MNU can be fully resolved based on their chromatographic retention times and MRM transitions. Calibration curves have been constructed and the extraction recovery of adducts, limits of detection, and limits of quantitation have been determined. Preliminary adduct profiles of available alkylpurine DNA glycosylases have been evaluated.

Experimental Design

Chemicals and Reagents

Calf thymus DNA, salmon testis DNA, *N*-methyl-*N*-nitrosourea, sodium cacodylate, sodium perchlorate, Na HEPES, Trizma base, potassium chloride, sodium chloride, ammonium acetate (NH_4OAc), dithiothreitol (DTT), ethylenediaminetetraacetic acid (EDTA), bovine serum albumin (BSA), 7-methyladenine (7mA), O-methylguanine (O^6mG), 7-methylguanine (7mG), ethanol, and perfluoropentanoic acid were purchased from Sigma-Aldrich Chemicals (Saint Louis, MO). 3-Methyladenine (3mA) was purchased from EMD Biosciences (Darmstadt, Germany). Hydrochloric acid and 1-methyladenine were purchased from Thermo Fisher Scientific (Waltham, MA). HPLC grade methanol and water were from J.T. Baker (Phillisburg, NJ). 3-Methyl-d_3-adenine (d_3-3mA) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). 7-

Methyl-d_3-guanine (d_3-7mG) was synthesized by Plamen Christov at Vanderbilt University (Nashville, TN).

Preparation of Methylated DNA Substrate

A methylated genomic DNA substrate is prepared by treating 5 mg of calf thymus DNA prepared in 0.3 M Na-cacodylate/0.1 M $NaClO_4$ pH 8.3 with 1 mM (MNU) in 0.1 M acetic acid for 8 hours at room temperature. The methylated substrate was precipitated with a final concentration of 1 M NaCl and 3 volumes of ice-cold, 100% ethanol. DNA was washed, dried and resuspended in TE buffer and dialyzed against the same buffer to remove any hydrolyzed methyladducts. An absorbance revealed a final concentration of 0.77 mg/mL methylated DNA in TE buffer.

Preparation of Calibration Standards

Calibration of the HPLC-MS/MS response was performed with standard solutions of 3mA, 7mG, O^6mG, 1mA, and 7mG over a range of 50-5000 nM for each compound. 1mA and 7mA were prepared in stock solutions at 1 mM in NH_4OAc. A stock solution of 3mA was prepared at 1 mM in water. An 800 uM stock solution of O^6mG and a 39 uM stock solution of 7mG were prepared in methanol. Serial dilutions of each stock were prepared in water to provide standard calibration solutions at the appropriate levels. In order to account for matrix effects, calibration standards were prepared in mock reaction conditions using AlkD Asp113Asn mutant protein owing to its low enzymatic activity. An internal standard stock solution of d_3-3mA was prepared by dilution in water (100 uM). A working solution of the internal standard was prepared by dilution of the

stock solution into stop buffer (0.5 mg/mL salmon DNA, 1 M NaCl, 1 mg/mL BSA, 10 uM d_3-3mA). Full scan MS in positive ion electrospray mode gave a major protonated molecular ion with *m/z* 150 $[M + H]^+$ for each adenine analogue, *m/z* 166 $[M + H]^+$ for each guanine analogue, and *m/z* 153 $[M + H]^+$ for d_3-3mA. Collision induced dissociation (CID) of *m/z* 150 of protonated 1mA, 3mA, and 7mA gave product ions with *m/z* 109 $[M + H - 41]^+$, 123 $[M + H - 27]^+$, and 106 $[M + H - 44]^+$, respectively. CID of *m/z* 166 of protonated 7mG and O^6mG gave product ions with *m/z* 124 $[M + H - 42]^+$ and 134 $[M + H - 32]^+$, respectively. CID of *m/z* 153 of protonated d_3-3mA gave a product ion with *m/z* 109 $[M + H - 44]^+$ (Figure 33).

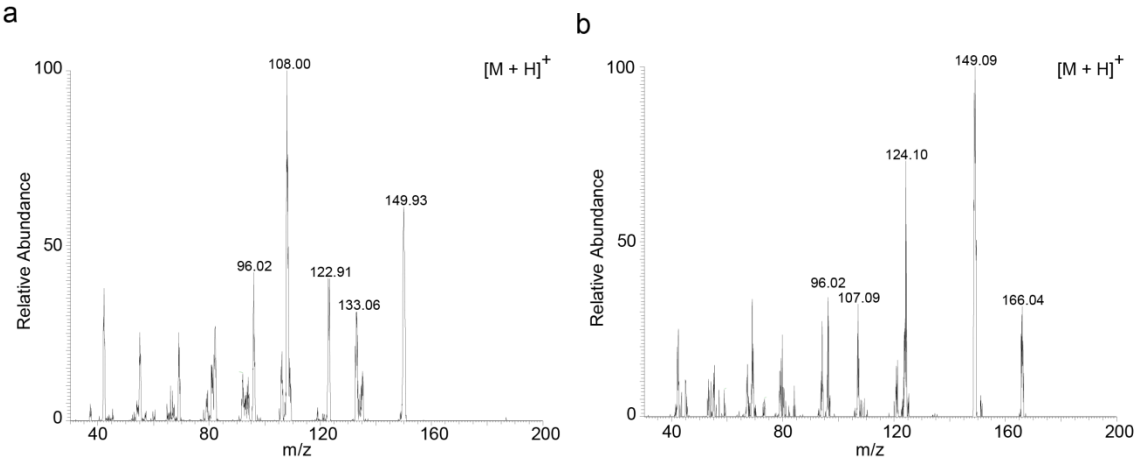


Figure 33. MS/MS-ESI(+) product ion spectra of $[M + H]^+$ ions of *N*3-methyladenine (**a**) and *N*7-methylguanine (**b**).

Instrumentation

Sample analyses were carried out using a Waters Acquity UPLC system (Milford, MA), made up of a binary solvent manager, refrigerated sample manager, and a heated column manager. Tandem mass spectrometric detection was performed using a ThermoElectron TSQ Quantum Ultra triple-stage quadrupole mass spectrometer (San

Jose, CA) equipped with an *Ion Max* heated electrospray (HESI) source and a 50 μM I.D. stainless steel capillary.

Chromatography Methods

A SymmetryShield RP_{18} column [2.1 X 150 mm, 3.5 uM particle size (Waters)], equipped with an Acquity UPLC in-line stainless steel filter unit (0.2 μM, Waters) was used for all chromatographic separations. The column and autosampler tray temperatures were both set to 25 °C. Mobile phases were made up of 0.2% perfluoropentanoic acid in (A) H_2O and in (B) MeOH:H_2O (95:5). Gradient conditions were as follows: 0-1 min, B = 0 %; 1-8 min, B = 0-25 %; 8-9 min, B = 25-100 %; 9-10 min, B = 100-0%; 10-15 min, B = 0 %. The flow rate was maintained at 0.4 mL/min. The total chromatographic run time was 15 minutes and a software-controlled divert valve was used to transfer eluent from 0-8 minutes and from 11-15 minutes of each run to waste. The sample injection volume was 10 μL. The autosampler injection valve and the sample injection needle were flushed and washed sequentially with mobile phase B (1 mL) and mobile phase A (2 mL) between each injection.

Tandem Mass Spectrometry Methods, Data Acquisition and Processing

The mass spectrometer was operated in positive ion mode and quantitation was based on MRM detection of methylated adenines and guanines. Product ion mass spectra were acquired in 30-200 *m/z* range; each methylated base was readily identified by its characteristic precursor ion and major product ion. The mass transitions (precursor to product) monitored were 150 →109 for 1mA, 150 →123 for 3mA, 150 →106 for 7mA,

166 →124 for 7mG, 166 →134 for O^6mG, and 153 →109 for d_3-3mA. Although the five most prevalent methylated bases that result from MNU treatment (1mA, 3mA, 7mA, 7mG, and O^6mG) do not have unique *m/z* transitions, they can be chromatographically separated (Figure 2). The following optimized parameters were used for the detection of analytes and internal standards: spray voltage, 4500 eV; capillary temperature, 300 °C; sheath gas, 30 psi; source CID on; and varied collision energies. Data acquisition and quantitative spectral analysis were performed using Thermo-Finnigan Xcalibur version 1.3 and Thermo-Finnigan LCQuan version 2.5.5, respectively. Standard curves were prepared by plotting observed peak area ratios (analyte peak area / internal standard peak area) against the corresponding molar concentration ratios for a series of methylated adenine and guanine standards.

Alkylpurine DNA glycosylase purification

TAG. *S. typhimurium* TAG protein was expressed with an N-terminal His_{10}-fusion from a pET-19b plasmid (Novagen) into *E. coli* C41 cells and propagated in LB media supplemented with 5 mM ZnSO4. Protein was overexpressed for 4 h at 25°C following addition of 0.5 mM IPTG. Cells were harvested in 50mM Tris-HCl (pH 7.5), 500mM NaCl, and 10% glycerol and lysed with an Avestin Emulsifier C3 homogenizer operating at ~20000 psi. TAG-fusion protein was purified using Ni-NTA (Qiagen) affinity chromatography. After cleavage of the His_{10} tag, TAG was further purified by heparin affinity and gel filtration chromatography to >99% homogeneity as estimated by Coomassie staining of an SDS gel. Protein was concentrated to 8 mg/ml and stored in 20 mM Tris (pH 8.5), 5% glycerol, 100 mM NaCl, 2 mM DTT, and 0.1 mM EDTA.

AlkD. The AlkD gene was PCR amplified from *Bacillus cereus* genomic DNA (ATCC 14579) and cloned into a pET27 (Novagen) derived expression vector (pBG103, Vanderbilt Center for Structural Biology) that produces a cleavable N-terminal His_6-SUMO-fusion protein. *E. coli* HMS174 cells transformed with the AlkD/pBG103 plasmid were propagated in LB media and protein was overexpressed for 3 h at 37°C upon addition of 0.5 mM IPTG. Cells were harvested in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 10% glycerol and lysed with an Avestin Emulsifier C3 homogenizer operating at ~20000 psi. AlkD-fusion protein was purified using Ni-NTA (Qiagen) affinity chromatography, followed by cleavage of the His_6-SUMO tag. AlkD was further purified by heparin affinity and gel filtration chromatography to >99% homogeneity. Protein was concentrated to 12.5 mg/ml and stored in 20 mM Bis-Tris Propane, 100 mM NaCl, and 0.1 mM EDTA.

AlkC. *Bacillus cereus* AlkC was cloned into a pET-19b expression vector (Novagen) that contains a cleavable N-terminal His_{10}-Tag. The vector was transformed into *E. coli* Rosetta cells and was overexpressed over night at 16°C upon addition of 0.1 mM IPTG. Cells were harvested in 50 mM Tris-HCl (pH 8.5), 500 mM NaCl, and 10 % glycerol and lysed with an Avestin Emulsifer C3 homogenizer operating at ~20000 psi. AlkC-fusion protein was purified using Ni-NTA (Qiagen) affinity chromatography. Following cleave of the His_{10} tag, AlkC was further purified by heparin affinity and gel filtration chromatography to >99% homogeneity. Protein was concentrated to 2.4 mg/ml and stored in 20 mM TRIS-HCl (pH 8.5), 200 mM NaCl, 10% glycerol, 2 mM DTT, and 0.1 mM EDTA.

MAG. *Saccharomyces cerevisiae* MAG was cloned into a pBG100 expression vector (Vanderbilt Center for Structural Biology) that produces a cleavable N-terminal His_6-fusion protein. The protein was overexpressed overnight in *E. coli* BL21 cells at 16°C upon addition of 0.1 mM IPTG. Cells were harvested in 50 mM TRIS-HCl (pH 7.5), 500 mM NaCl, and 10 % glycerol and lysed with an Avestin Emulsifer C3 homogenizer operating at ~20000 psi. ScMAG was purified by Ni-NTA (Qiagen) affinity chromatography and cleaved with Precission Protease to remove the tag. Purification continued with SP Sepharose affinity and gel filtration chromatography to >99% homogeneity. Protein was concentrated to ~7 mg/ml and stored in 20 mM TRIS-HCl (pH 7.5), 150 mM NaCl, 25 % glycerol, 2 mM DTT and 0.01 mM EDTA.

Alkylpurine DNA Glycosylase Assay and Sample Preparation

The rate of methylated purine base excision was measured after wild-type alkylpurine DNA glycosylase was incubated over time with MNU-treated genomic DNA. A 50 μL reaction containing the methylated genomic DNA substrate was incubated with the enzyme (up to 20 uM) in activity buffer (250 mM HEPES pH 7.5, 500 mM KCl, 50 mM DTT, 10 mM EDTA, 0.5 mg/mL BSA) at 37°C and terminated at a given time point by addition of stop buffer spiked with internal standard (0.5 mg/mL salmon DNA, 1 M NaCl, 1 mg/mL BSA, 10 uM d_3-3mA) followed by three volumes of ice-cold, 100% ethanol to precipitate the DNA. Acid hydrolysis samples were treated with 0.5 N HCl instead of enzyme and no enzyme controls were substituted with glycosylase buffer (20 mM Bis-Tris Propane, 100 mM NaCl, and 0.1 mM EDTA). Following a 20 minute incubation at -20 °C the reactions were centrifuged for 15 minutes and the supernatant,

containing any released bases, was evaporated under a steady stream of nitrogen gas at 30 °C. The residue was reconstituted in 50 μL ultrapure water and transferred to a 200 μL silanized autosampler vial equipped with Teflon-lined bonded rubber septa, followed by analysis by HPLC-MS/MS.

Results and Discussion

Previous reports have shown how LC/ESI-MS/MS methods may be used to directly and simultaneously detect alkylated DNA adducts (Chadt et al, 2008; Roberts et al, 2001; Yang et al, 2002; Zhang et al, 2006) after exposure to various laboratory methylating agents. The method developed here will expand on these reports by examining the release of methylated adducts by alkylpurine DNA glycosylases. While some of these enzymes like *E. coli* TAG are very specific for only 3mA, others like *S. cerevisae* MAG and *B. cereus* AlkD and AlkC are more promiscuous and can remove additional methylated bases (Alseth et al, 2006; O'Rourke et al, 2000; Riazuddin & Lindahl, 1978). Treatment of double-stranded DNA with the methylating agent MNU produces a variety of methylated bases with *N*7-methylguanine comprising 67%, followed by *N*3-methyladenine at 9%, O^6-methylguanine at 6.3% and *N*1-methyladenine and *N*7-methyladenine at 1.3 and 1.7%, respectively (Figure 34) (Singer & Grunberger, 1983). Following incubation of this substrate with alkylpurine DNA glycosylases, free methylated bases were released from DNA and extracted from the samples. The released methylated DNA nucleotides were then separated and monitored with reversed-phase

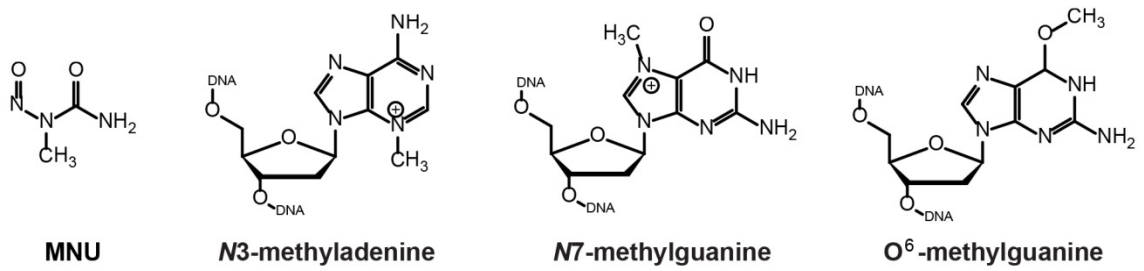


Figure 34. Structures of the methylating agent *N*-methyl-*N*-nitrosourea (MNU) and the three most prevalent methylated bases, *N*3-methyladenine (3mA), *N*7-methylguanine (7mG), and O^6-methylguanine (O^6mG).

high performance liquid chromatography and positive electrospray tandem mass spectrometry (Figure 35). The HPLC separation benefited greatly from the addition of perfluoropentanoic acid to the mobile phase. This ion-pairing agent also increased the retention times of the methylated bases. In order to obtain the highest possible selectivity, mass spectrometry was performed in MRM mode. Direct infusion of *N*1-, *N*3-, *N*7-methyladenine and *N*7- and O^6-methylguanine allowed for tuning of the ESI source and optimization of MRM parameters (150 →109 for 1mA, 150 →123 for 3mA, 150 →106 for 7mA, 166 →124 for 7mG, 166 →134 for O^6mG). To date, the measure of released bases has been quantified using a deuterated 3mA standard.

Extraction Recovery, Calibration Curve, Limit of Detection, and Limit of Quantitation

In order to investigate the extraction recovery of the assay for each analogue, triplicate samples were pre-spiked with 100 nM of each analogue by addition to the reaction buffer during the glycosylase assay and compared to samples that were spiked

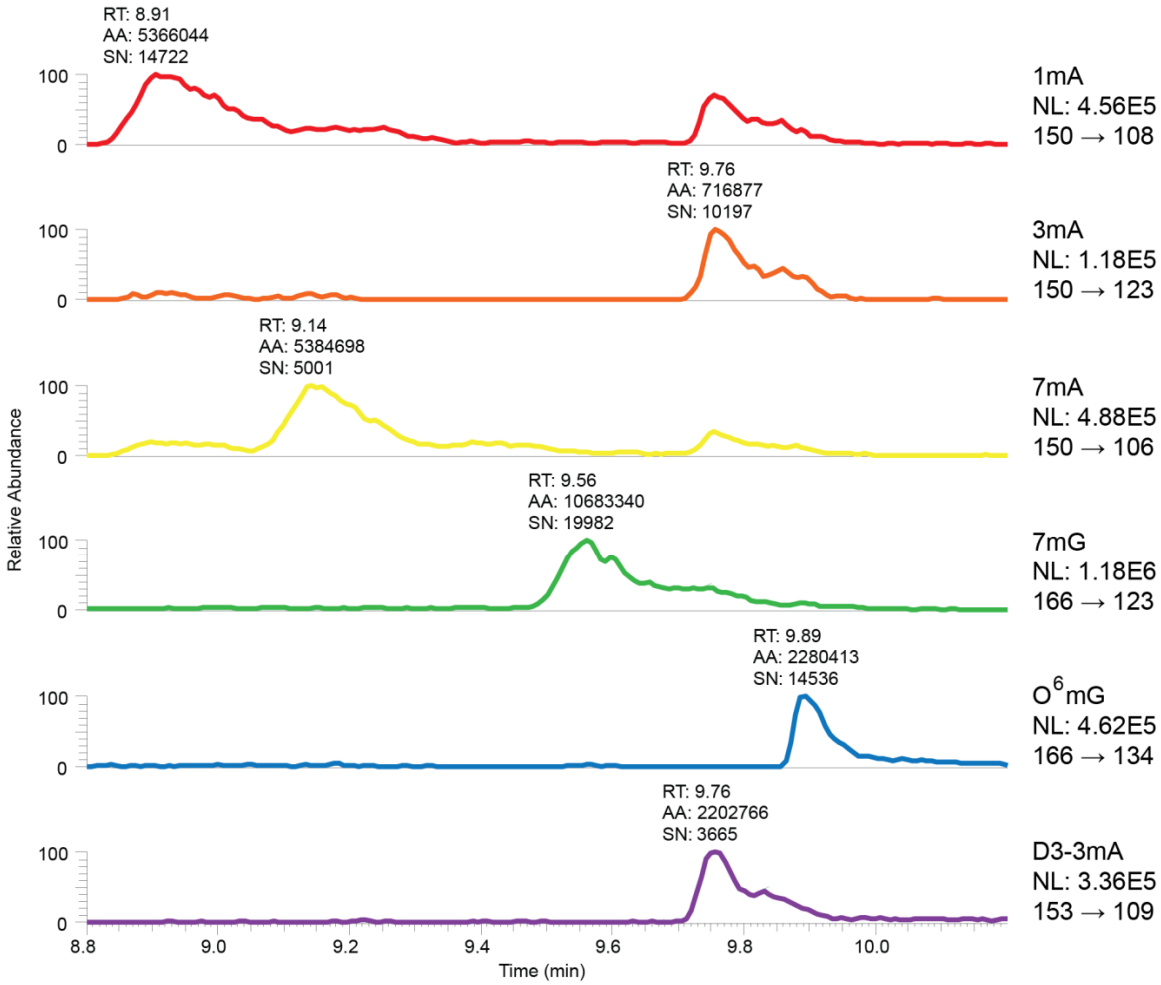


Figure 35. HPLC chromatogram of methylated bases (1mA, 3mA, 7mA, 7mG, O^6mG, and d_3-3mA) in MRM mode.

with 100 nM of each analogue during the final reconstitution step. These samples included the inactive glycosylase AlkD Asp113Asn as a surrogate matrix. Extraction recovery values were determined to be 85 ± 11 % for 3mA, 73 ± 14 % for 7mG and 82 ± 83 % for d_3-3mA. While these recovery values are quite good, there is room for improvement. Around neutral pH the hydrophobic purine bases are nearly insoluble in water, therefore, the recovery could benefit from adding a small amount of methanol to during the reconstitution step. Calibration curves have been constructed by preparing samples with various concentrations of standards and using a fixed concentration of d_3-3mA as an internal standard. The calibration curves for all methylpurines were linear

over the concentration range of 0.5-50 pmol with all correlation coefficients (R^2) over 0.95 (Table 5). In order to completely calibrate the technique, deuterated standards for each methylated base observed should be synthesized or purchased and used to determine

Table 5. Calibration parameters of the method tested with standards, 1mA, 3mA, 7mA, 7mG and O^6mG.

Component	Calibration curve equation	R^2
1mA	y = 0.2917x-0.0012	0.9505
3mA	y = 0.0205x-0.0002	0.9834
7mA	y = 0.2556x-0.0005	0.9853
7mG	y = 0.2235x-0.0003	0.9879
0^6mG	y = 0.0892x-0.0010	0.9936

calibration curves. To this end, d_3-7mG has already been synthesized. The limits of detection have been determined to be 100 fmol for 1mA, 3mA, 7mA, 7mG and O^6mG. The lower limits of quantitation have been found to be 500 fmol for 1mA, 3mA, 7mA and 7mG and 100 fmol for O^6mG. A comprehensive evaluation of this method will include intra- and inter-day accuracy and precision measurements. To determine intra-day precision of the method, quality control samples will be analyzed multiple times in one day. Processing and analyzing the calibration standards over the course of several days will allow for the determination of inter-day precision and accuracy. Accuracy will be expressed as percent error and precision will be expressed as percent relative standard deviation.

Acid hydrolysis of MNU-treated DNA

To determine the upper limit of adducts produced in calf thymus DNA by the MNU reaction, samples were incubated with a final concentration of 0.5 N HCl in order to fully depurinate all methylated bases. The release of 1mA, 3mA, 7mA, 7mG, and

O^6mG was monitored at four different times (five, 30, 120 minutes and 24 hours) with three different amounts of DNA [3.85 μg (1X), 7.7 μg (2X), and 15.4 μg (4X)]. After a 24 hour exposure of the highest amount of DNA with HCl all monitored DNA adducts could be observed (Table 6). Out of all adducts observed, *N7*-methylguanine made up the highest percentage of total released bases measured at ~84% while 3mA was ~14%, O^6mG was ~2% and both 1mA and 7mA were under 1%. This spectrum is consistent with previously measured amounts of adducts formed by MNU treatment in double stranded DNA, at least with respect to the rank order of each adduct measured (Singer & Grunberger, 1983). While small amounts of O6mG could be detected in some samples at

Table 6. Quantitation of adducts 1mA, 3mA, 7mA, 7mG and O^6mG, released by acid hydrolysis over time.

DNA (μg)	Time (m)	Concentration of Adduct [μM]				
		1mA	3mA	7mA	7mG	O^6mG
3.85 (1X)	5	n.d.	0.157	n.d.	n.d.	n.d.
	30	n.d.	0.069	n.d.	0.340	n.d.
	120	n.d.	0.310	n.d.	0.607	0.041
	1440	n.d.	0.074	n.d.	1.186	n.d.
7.70 (2X)	5	n.d.	0.335	n.d.	0.162	n.d.
	30	n.d.	0.307	n.d.	0.603	0.025
	120	n.d.	0.223	n.d.	1.669	n.d.
	1440	n.d.	0.262	n.d.	2.278	0.041
15.4 (4X)	5	n.d.	0.642	0.046	0.232	n.d.
	30	n.d.	0.649	n.d.	0.818	0.040
	120	n.d.	0.570	0.008	2.694	0.099
	1440	0.010	0.645	n.d.	3.900	0.073

each concentration of DNA, 7mA and 1mA were only observed after incubation with the highest amount of DNA (Table 6). Levels of 3mA remained steady over the time course for a given DNA amount, but increased from an average of 0.15 μM at 1X DNA to 0.28 μM at 2X DNA to 0.63 μM at 4X DNA. This implies that almost all of the 3mA adducts

are released within five minutes. Roughly, each time the amount of DNA is doubled, the amount of released 3mA observed also doubles, suggesting that a five minute, 0.5 N HCl treatment is sufficient to remove all 3mA adducts from MNU-treated DNA (Figure 36). The same doubling trend holds true for observed 7mG adducts for individual time points, but unlike 3mA, the level of 7mG continually increases over time (Figure 36). This result may suggest that not all 7mG adducts are released by acid hydrolysis within 24 hours. This experiment will be repeated in triplicate in order to determine accurate concentrations of bases released and access the error. Examining acid hydrolysis adduct profiles under different substrate methylation conditions will provide insight into how many lesions are produced on the DNA during the reaction.

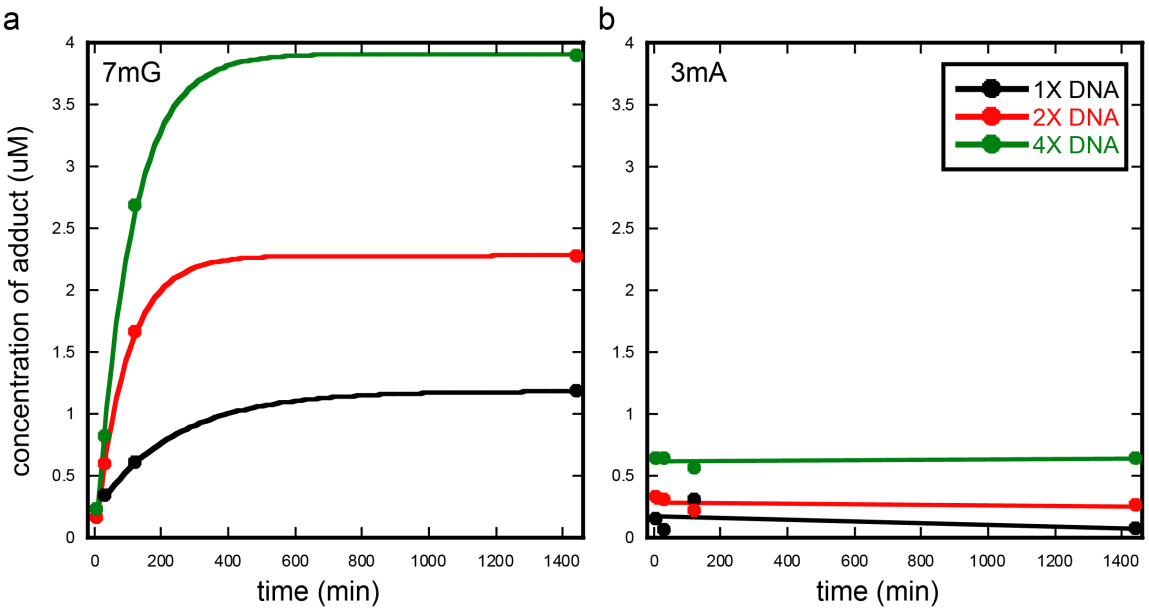


Figure 36. Quantitation of 7mG (**a**) and 3mA (**b**) adducts released by acid hydrolysis over time. Data was fit to a single exponential (7mG) or linear equation (3mA).

Quantification of Released Adducts by Alkylpurine DNA Glycosylases

In a preliminary experiment to determine the efficacy of this method, adduct profiles of wild-type alkylpurine DNA glycosylases, TAG, AlkD, and MAG were analyzed after incubation with MNU-treated DNA for one hour. The results were

compared to an HCl treated sample, which provided an upper limit of adduct release, and a sample that lacked enzyme but was still subjected to a heated incubation in order to determine the lower limit of adduct release. Each of these enzymes has been shown to remove cytotoxic 3mA but unlike TAG, both AlkD and MAG are also able to remove 7mG from DNA (Figure 37). As predicted, all three glycosylases, as well as a sample treated with HCl, remove approximately equal concentrations of 3mA, at least three times the amount released in the no enzyme negative control. Relative to the acid hydrolyzed positive control, MAG removes almost twice the amount of 3mA than 7mG suggesting that this enzyme may be more specific for 3mA. Unexpectedly, AlkD removes more than twice the amount of 7mG over the HCl treated sample. Although it is tempting to conclude that AlkD is more specific for 7mG, the fact is that the amounts of 7mG are imprecise since the calibration is based on only a 3mA internal standard. Calibration of the experiment with a d_3-7mG internal standard, as well as repetition of this experiment should produce more accurate results.

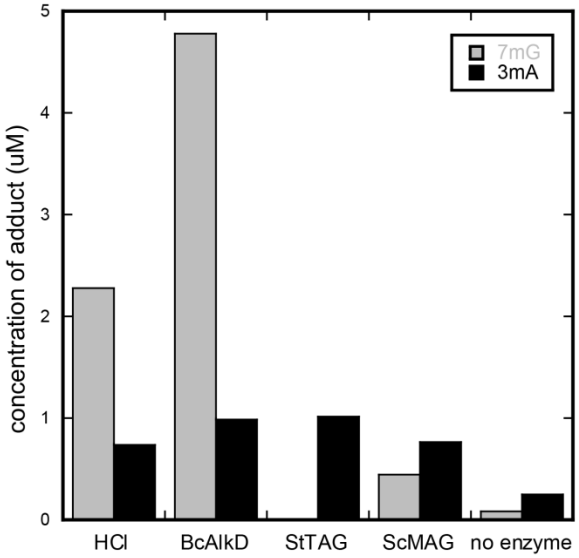


Figure 37. DNA glycosylases release 7mG and 3mA. Quantitation of 7mG (grey bars) and 3mA (black bars) adducts released after a 1h incubation with alkylpurine DNA glycosylases.

Time-course of Adduct Release by Alkylpurine DNA Glycosylases

So far, a single time-course experiment (five minutes, 30 minutes, two hours, six hours, 24 hours) has been performed to examine the amounts of adducts releases by wild-type TAG, AlkD, and AlkC (Figure 38). Overall more 7mG bases are released for all enzymes tested with the exception of TAG. Interestingly a small amount of 7mG is observed in the TAG 24 hour time point may actually be within error of the no enzyme control and is most likely due to the lack of an internal 7mG standard. The majority of 7mG appears to be removed within five minutes of AlkD incubation, while AlkC appears to remove the same lesion more slowly and only after at least a thirty minute incubation. Overall, smaller amounts of 3mA are removed by each enzyme but still well above the no enzyme controls for every time point. The results of this time course also suggest that

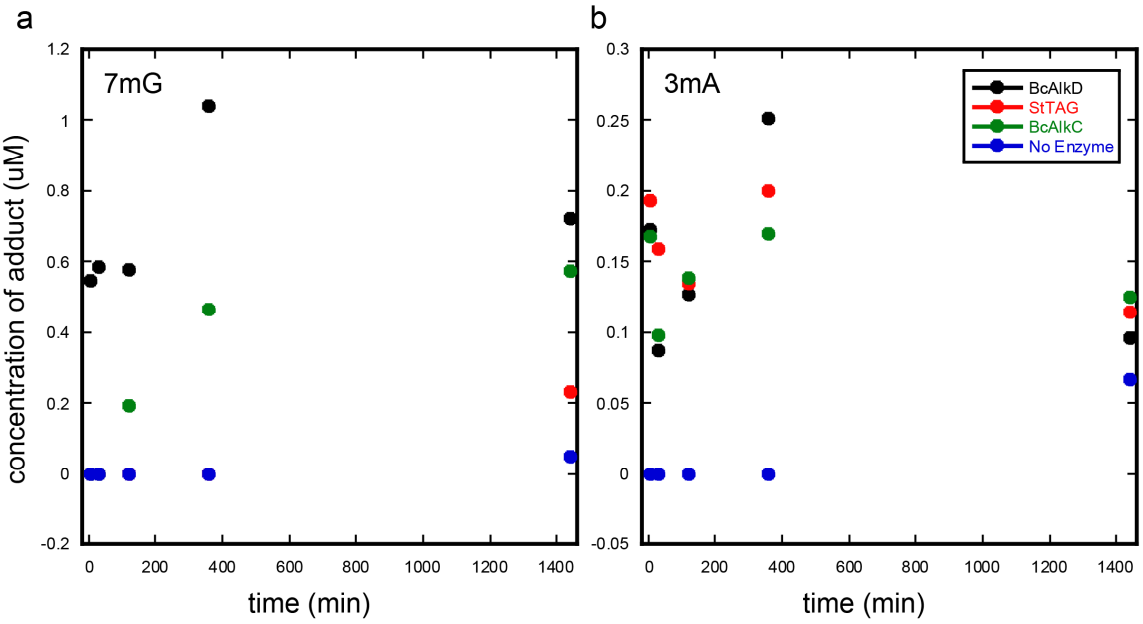


Figure 38. Time course of 7mG and 3mA release by DNA glycosylases. Quantitation of 7mG (**a**) and 3mA (**b**) adducts released over time after incubation with alkylpurine DNA glycosylases with 7.7 ug of methylated DNA.

over time. However, in this assay the enzyme concentration is always at a sub-saturating

level with respect to the DNA, most likely due to nonspecific binding of the proteins to the vast excess of unmodified bases in the genomic DNA substrate.

Conclusion

Once the method has been completely developed and verified the results of experiments to examine the biological response of glycosylases to alkylation damage will be used to provide detailed conclusions about the substrate specificity of different alkylpurine DNA glycosylases. This method is powerful because not only will it eventually allow us to determine the true substrate specificity of alkylpurine DNA glycosylases but also the role specific residues play in substrate specificity by analyzing adduct profiles of mutant enzymes. For example, structural and functional studies of TAG show that Glu38 sterically excludes N7-substituted methylpurine bases from the enzyme. Mutation of this residue abolishes activity for a 3mA substrate but is predicted to allow for 7mG excision. This mutation will be compared to wild-type enzyme using the method described here and will shed light onto the exact role of this residue.

Acknowledgements

We thank Wade Calcutt at the Vanderbilt University Mass Spectrometry Core for assistance with method development.

CHAPTER VII

DISCUSSION AND FUTURE DIRECTIONS

Understanding the complex mechanisms that underlie how damaged DNA bases are repaired in the cell has been the focus of decades of research. High-resolution structural techniques have been combined with biochemical approaches to discover how DNA glycosylases, damage specific enzymes that initiate the base excision repair pathway, locate their substrates and perform catalysis. Prior to the research presented in this thesis, all DNA glycosylases were known to utilize a base flipping mechanism for catalysis. After locating and recognizing the lesioned base, the enzymes rotate the base out of the DNA helix and into an active site cleft to gain access to the *N*-glycosylic bond, which must be cleaved to liberate the base. It is proposed that these enzymes use a combination of aromatic π-π or π-cation interactions and protein residue functional group chemistry to recognize and remove their substrates (Stivers & Jiang, 2003). Each DNA glycosylase is specific for a particular lesioned base and the specificities of these repair enzymes are determined, therefore, by the chemical and physical properties of the base binding pocket, and by the particular protein architectures used to probe the bases within the DNA duplex. The alkylpurine specific DNA glycosylases are unique since they have a range of substrate specificities. In addition to removing cytotoxic *N*3-methyladenine (3mA) lesions some remove other positively charged methylated lesions (e.g. 7mG) and still others remove uncharged alkyl-lesions (e.g. εA, HX). Regardless of the wealth of structural information available on these enzymes, the mechanisms of alkylpurine

recognition, flipping, and catalysis remain poorly understood. The focus of this work was to address fundamental questions about how alkylpurine glycosylases locate and select for positively charged methylated bases and to understand the catalytic mechanism of nucleobase excision. This was accomplished by studying a recently identified alkylpurine DNA glycosylase from *Bacillus cereus*, AlkD. Specifically, I have combined structure determination of the enzyme in the absence and presence of DNA by X-ray crystallography with *in vitro* mutational analysis of ligand binding and substrate catalysis to investigate interactions between glycosylases and DNA ligands. The results of this work support an emerging mechanistic hypothesis that alkylpurine DNA glycosylases take advantage of the inherent instability of charged alkylpurines to excise them from the genome.

HEAT Repeats as a DNA Binding Platform

The high-resolution crystal structure of AlkD, presented in Chapter II revealed that the protein adopts a C-shaped globular fold composed exclusively of helical HEAT-like repeats. This architecture, which is also predicted to be found in an AlkD homolog, AlkC, separates these enzymes from all other DNA glycosylases and thus AlkD represents the sixth DNA glycosylase structural superfamily. Normally, HEAT motifs are protein binding domains and AlkD is the first structural example of an enzyme that uses a unique tandem helical repeat structure to bind DNA. The C-terminal α-helix of each HEAT repeat forms the inner, concave surface of the protein and contains lysine or arginine residues at conserved positions. The enzyme is not only perfectly shaped, but is the appropriate charge, to accommodate a DNA duplex. DNA binding experiments

highlighted in Chapters II and IV confirm AlkD binds DNA and the structures of AlkD bound to DNA showcased in Chapter III provide the first structural example of a HEAT motif enzyme bound to DNA.

The discovery of HEAT repeats in AlkD is significant given the large number of DNA processing proteins that have been predicted to contain these motifs. Several chromatin-remodeling factors, including condensins, cohesins, along with other proteins involved in chromosome dynamics have been identified to contain HEAT repeat domains. Initially proposed to facilitate protein interactions within their large multimeric complexes, it is possible to imagine that they may instead be used to bind and protect DNA during the dynamic organization of the nucleus (Neuwald & Hirano, 2000). Because the interactions are mainly electrostatic in nature, the DNA binding may be more transient in nature, allowing for the proteins to come off and on DNA quickly. Several DNA-damage response kinases in the PIK-like protein superfamily (ATMs, ATRs, TORs, DNA-PKs) have also been predicted to contain HEAT repeat domains outside of their kinase motifs and recent low resolution structures of the catalytic subunit of DNA-PK showed HEAT repeats forming a large ring structure that contained potential DNA binding elements (Perry & Kleckner, 2003; Sibanda et al, 2010; Williams et al, 2008). Recently the structure of a mitochondrial transcription factor, MTERF1, bound to its specific DNA termination sequence was determined. MTERF1 was also determined to use tandem helical repeats to bind DNA, much like AlkD. Comparison of the AlkD/DNA and MTERF/DNA structures in Chapter V reveals the common features of these tandem helical repeat proteins including creating electrostatic and van der Waals interactions to the phosphate backbone of nucleic acids. Similar to AlkD, MTERF1 also

extrudes nucleic acids out of the DNA duplex, most likely through an active mechanism, in order to perform its function.

DNA Damage Recognition by AlkD

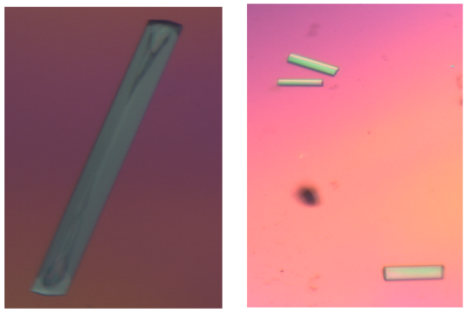
The results presented in Chapters II and III reveal that the binding surface of AlkD is ideally suited to bind B-form DNA and the positively charged surface of AlkD allows for favorable interaction with unmodified DNA. These properties predict that the enzyme will use a processive duplex scanning mechanism to locate substrate nucleotides, in a similar manner to T4 EndoV and UDG. However, instead of specifically recognizing an intrahelical or extrahelical lesioned base, the AlkD/3d3mA•T-DNA and AlkD/G•T-DNA structures suggest that the protein recognizes a structurally destabilized base pair or disrupted stacking interactions within the DNA duplex. AlkD also lacks intercalating residues which further suggests that the enzyme is detecting altered base stacking or pairing over a base lesion. Although this mode of DNA damage recognition has yet to be observed in other DNA glycosylases, there is precedent in other base repair enzymes like AlkB (Yang et al, 2008) and proteins that interact with DNA including some restriction enzymes like Ecl18kl (Tamulaitis et al, 2007). Although both proteins flip nucleotides (one in the case of AlkB, two in the case of Ecl18kl) neither intercalate protein residues into the space left by the nucleotides, instead, the protein squeeze or collapse the DNA duplex to maintain base stacking properties much like what is observed for AlkD. Although wild-type AlkD does not appear to distinguish between unmodified and modified DNA, a comprehensive mutational analysis of AlkD presented in Chapter IV reveals the role individual residues play in binding DNA and lesion recognition. The

overall electropositive surface of the enzyme is responsible for general DNA binding however key residues (Gln38, Asp113, Arg148, Trp187, and Arg190) that line the concave cleft of AlkD are more important for detecting duplex distortion. Interestingly, tyrosine 27 seems to sense perturbations as subtle as those created by a G•T mismatch, which suggests that this residue may act as a probe to recognize destabilized base pairs.

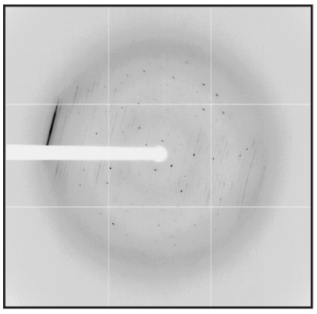
To further investigate base-pair remodeling by AlkD it is necessary to understand the structure of the base-pairs in the context of DNA alone. Although there are DNA structures of G•T mismatches available for comparison none contain a similar sequence context as what was crystallized in complex with AlkD. Furthermore, there is no high-resolution structural information available for the 3d3mA moiety outside of the structures with AlkD presented in Chapter III. To this end, G•T-DNA and 3d3mA•T-DNA have been recently been crystallized within a sequence context similar to the G•T/3d3mA•T context in the original AlkD/G•T and AlkD/3d3mA•T structures: d(CGT*X*AATTCGCG)/d(GCA*T*TTAAGCGC) (DNA1) and self-complementary sequence d(CGT*G*AAT*T*TACG) (DNA2) where *X* is G or 3d3mA (Figure 39a,b). Although the diffraction data collected using these crystals is incomplete (Figure 39c), preliminary structures of G•T-DNA were solved by molecular replacement using modified Dickerson dodecamers, d(CGCGAATTTGCG) (PDB ID: 113D) and d(CGCGAATTCGCG) (PDB ID: 463D), respectively and are undergoing further data processing and model building steps (Figure 39d). It appears that the G•T base pair forms a wobble structure, unlike what is observed in the enzyme-DNA structures, implying that the enzyme induces distortion to the DNA duplex. Additional crystal trials will be undertaken to optimize the diffraction data. In lieu of using molecular

replacement to determine phases, heavy atoms may be introduced into the crystal by replacing thymine residues with 5-bromo- or 5-iodo-uracil during DNA synthesis. Importantly, the 3d3mA•T-DNA structure will provide the first structural information about 3d3mA•T in the context of DNA alone and allow for detailed comparison to the base pair bound to AlkD.

a



b



c

	G•T-DNA
Cell Parameters	
Space group	R3
a,b,c (Å)	43.4, 43.3, 98.9
α, β , γ (°)	90.0, 90.0, 120
Data Collection	
Wavelength (Å)	1
Resolution (Å)	20-2.65
Completeness (%)	98.4 (100)
Redundancy	5.5 (5.6)
Refinement	
Resolution (Å)	20-2.65
R_{work}/R_{free}	0.258
R.m.s.d. bonds (Å)	0.007
R.m.s.d. angles (°)	1.493

d

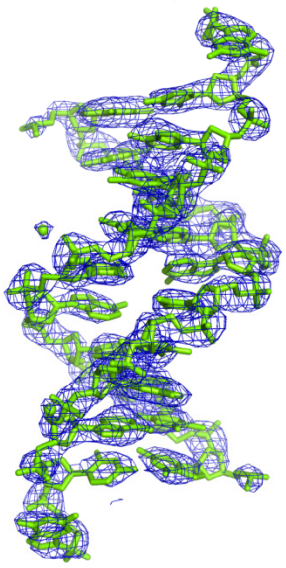


Figure 39. Crystallization of G•T and 3d3mA•T DNA. **a**, Crystal of G•T-DNA on the left, 3d3mAT-DNA on the right. **b**, Diffraction pattern from 3d3mA•T-DNA. **c**, Current data collection and refinement statistics for G•T-DNA2. **d**, Model of G•T-DNA2 (green) built into an electron density map (blue mesh).

Substrate Specificity of AlkD

Co-crystal structures of AlkD bound to DNA, discussed in Chapter III, combined with biochemical studies of the enzyme, illustrate why AlkD is specific for only positively charged alkylpurines and provides insight into the mechanism of catalysis of these lesions. Structures of AlkD bound to DNAs containing an isosteric substrate (3d3mA) and product (THF) mimic show that there are no contacts made to the lesioned base, instead, the 3d3mA and THF lesions are positioned on the opposite face of the DNA helix from the protein binding surface. In the substrate structure, the 3d3mA•T base pair is disrupted so that the thymine is partially rotated toward the protein, while in the product structure, both the abasic THF site and the base opposite the lesion are rotated out of the helix, causing the DNA duplex on either side to collapse to maintain base stacking. This technique is unlike all other DNA glycosylases studied to date, which flip the DNA damage into a protein active site and use a set of side chains to fill the gap created by a flipped-out modified base. After recognition, AlkD remodels the DNA phosphate backbone to create an optimal protein-DNA binding interface and promotes the lesion into solvent and therefore provides indirect catalytic assistance to liberate charged alkylpurines from the DNA duplex.

This unique DNA structure helps explain why AlkD only removes positively-charged purines. These lesions, including 3mA and 7mG, do not require a great amount of catalytic assistance to undergo hydrolysis due to their high rates of spontaneous depurination. Because AlkD does not recognize a particular lesion through interactions with the protein active site, it is also able to remove larger purines and even pyrimidines, such as pyridyloxobutylated (POB) lesions, as long as they have a weakened *N*-

glycosylic bond that results from a formal positive charge. Since the substrate specificity of AlkD is more a byproduct of inherently unstable positively-charged lesions it is reasonable to assume that AlkD recognizes and exposes many different types of damage to the solvent as long as it affects base pairing or base stacking in the DNA duplex. Much like Rad4, the XPC ortholog in *Saccharomyces cerevisiae*, exposes cyclopyrimidine dimers to the rest of the NER machinery, AlkD may expose more stable lesions to the solvent and allow for other DNA glycosylases or possibly NER enzymes to remove these lesions. In fact, organisms that contain AlkD orthologs also contain at least one AlkA or AAG alkylpurine DNA glycosylase. AlkD may assist in scanning and probing DNA without the energy cost of kinking the DNA to promote lesions toward other DNA glycosylases.

Base Excision by AlkD

Analysis of the structures of AlkD bound to DNA presented in Chapter III generates the hypothesis that AlkD excises alkylpurines through solvent exposure. Unfortunately, the work presented in this thesis does not answer the question for how AlkD interacts with a positively charged lesioned base. Although 3d3mA is an excellent structural mimetic to 3mA, it lacks the formal positive charge of the true substrate. In fact,

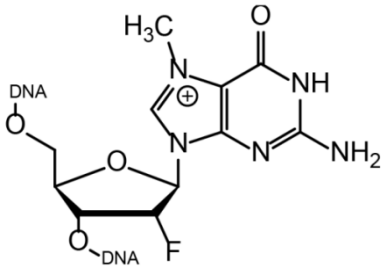


Figure 40. 2'-Fluoro-7-methylguanine.

there are no structures of a glycosylase bound to a charged alkylpurine substrate due to the inherent instability of the lesions. Recently, however, the Verdine group synthesized a charged alkylpurine mimic, 2'-fluoro-7mG (2'F-7mG) (Figure 40) and determined the crystal structure of this lesion incorporated into a duplex DNA strand (Lee et al, 2008). We are in the process of synthesizing the 2'F-7mG phosphoramidite and incorporating it into a 12mer DNA duplex of the same sequence as the 3d3mA•T duplex used in the co-crystal structures. Unlike the uncharged 3d3mA moiety, the 2'F-7mG retains the formal positive charge present in AlkD's preferred substrates, so this substrate mimic will be used for co-crystal trials with wild-type AlkD. Upon production of crystals they will be optimized for quality, size, and reproducibility. They will then be screened and evaluated for high-resolution diffraction with Vanderbilt's in-house X-ray source and high quality crystals will be flash frozen and stored for synchrotron data collection. The previously determined structure of unliganded AlkD will be used as a search model for molecular replacement to accomplish phasing of the protein-DNA complex. Once a model has been built into the density corresponding to the DNA, phase improvement and refinement will be carried out to produce a high quality atomic model of AlkD bound to DNA containing the 2'-F-7mG. A structure of this complex will illuminate how AlkD interacts with a charged substrate. We predict the ribose ring of the 7mG will overlay with the THF abasic site in the AlkD/THF-DNA structures so that they 7mG base will be in an extrahelical orientation. This would strengthen the hypothesis that solvent exposure of the substrate base is necessary for catalysis.

The details of the AlkD/DNA structures also raise the possibility that the phosphate groups flanking the lesions may participate in catalysis by positioning water

molecules necessary for hydrolysis of the *N*-glycosylic bond. The highly distorted DNA backbone in the AlkD/THF-DNA structure shows that the flipped ribose ring is positioned directly above a neighboring phosphate (Chapter III, Figure 23). Several water molecules bridge the extrahelical ribose C1′ carbon and the phosphate, raising the possibility that the phosphate groups may serve to activate a water nucleophile. Since the highly dissociative reaction mechanism lacks the requirement for a general base, using a DNA phosphate to position a water nucleophile near the *N*-glycosylic bond of the lesioned base is a plausible catalytic mechanism. DNA phosphate effects can be measured by eliminating the phosphate charge by replacement with a methylphosphonate linkage. The methylphosphonamidites will be purchased from Glen Research and synthesized onto the 3' end of 12mer primer (5'-GACCACTACACC). This primer will be ^{32}P-labeled at the 5′-end, annealed to a 3-fold excess of the complementary strand (5′-GTTGTAAGGAATCGGTGTAGTGGTC), and extended using DNA polymerase I Klenow fragment (New England Biolabs) in the presence of deoxy-7-methylguanosine 5′-triphosphate (d7mGTP, Sigma), dCTP, dTTP, and dATP. This substrate, which will contain a methylphosphonate linkage on the 5' end of the base 5' to the 7mG lesion (N-1), will be incubated with wild-type or mutant AlkD over time to determine the corresponding k_{cat}. A decreased rate of excision of this 7mG substrate versus the unmodified 7mG substrate would suggest that the DNA phosphate plays a role in catalysis. The methylphosphonate linkage could be moved around 5' to the lesion or to the complementary strand to provide measurements of individual phosphate effects. This type of experiment has been successfully performed in the lab of James Stivers to measure the effects of the DNA phosphate backbone on the uracil DNA glycosylase

reaction (Parker & Stivers, 2008). DNA phosphates may also play a role in stabilizing a potential oxocarbenium intermediate, which has been observed for UDG (Dinner et al, 2001; Jiang et al, 2003).

Mutation of several residues effects catalytic activity of AlkD (Chapter IV). For some residues (Asp113, Arg148, and Arg190) this may be a consequence of disrupting favorable interactions between the protein and the DNA at the site of the lesioned base pair. To understand the structural basis for the behavior of the catalytic mutants, a structure of mutant AlkD Asp113Asn or AlkD Arg148Ala in complex to DNA containing alkylated, mismatched, or abasic nucleotides is underway. Although far-UV circular dichroism verified the secondary structural elements of the mutant protein as equal to that of wild-type, a near-UV scan was inconclusive in identifying if there were any local perturbations to the structure as a result of Asp113Asn or Arg148Ala mutations. The structures are predicted to show at least small conformational changes at the site of the mutation including the dissolution of the Asp113-Arg148 salt bridge. In the presence of DNA the structures may show a lack of base-pair distortion as seen in the AlkD/3d3mA•T-DNA and AlkD/G•T-DNA structures and may appear more like the simulation of a G•T wobble superimposed on the structure. This is because optimal hydrogen bonds will be destroyed and prevent the backbone distortion necessary to rearrange the base-pair.

The Putative Role of AlkD in Base Excision Repair

AlkD's novel architecture and the way the enzyme interacts with DNA suggests possible roles for the enzyme within the context of the base excision repair (BER)

pathway as well as DNA repair pathways in general. Since the DNA binding activity is found along the concave surface of the enzyme, the convex surface is free to possibly bind other enzymes in the BER pathway. Crystal studies and biochemical experiments suggest that DNA glycosylases are very product inhibited and most likely remain on the DNA until it can be passed to the next enzyme in the pathway, an AP endonuclease. Still, researchers have been unable to detect an interaction between glycosylases and endonucleases and it is unclear how the endonucleases could access the abasic site if it is buried within a glycosylase active site. The unexpected structure of AlkD bound to a DNA containing an isosteric product mimic (Chapter III) showed that the enzyme promotes the abasic site into an extrahelical orientation away from the protein and into the solvent. From this orientation, the DNA could easily be passed to an endonuclease. In order to test whether an interaction may occur between AlkD and an endonuclease or any other protein, co-immunoprecipitation assays could be performed by adding an antibody specific for AlkD to a cell lysate and precipitating the antibody to determine if there are any proteins that interact with AlkD. This experiment may also be performed by tagging exogenous AlkD and the *B. cereus* AP endonuclease to see if either enzyme can pull down the other in the presence of abasic site DNA.

Structure-function Analysis of AlkC

One way to discover if the mechanism of substrate orientation and catalysis by AlkD is a general mode of catalysis by HEAT-repeat enzymes is to determine the structure of AlkC. Based on secondary structure predictions of AlkC, we expect the protein to adopt a similar HEAT-like repeat architecture to AlkD and thus AlkC may

interact with DNA in a similar manner to AlkD. However, the initial functional characterization of AlkC revealed that the enzyme preferentially removes 3mA over the other positively charged methylpurines (Alseth et al, 2006). This difference in substrate specificity may be explained by the differences in amino acids that in AlkD show an effect on catalysis or DNA binding. A structure of AlkC in complex to DNA will reveal how the protein binds DNA and a comparison of the two proteins may reveal how differences in substrate specificity are obtained regardless of base excision mechanism and further explain what role the unique HEAT-like repeat structure has on DNA binding. Prior to crystallization, the role of specific AlkC residues on substrate specificity can be determined by screening mutants with the HPLC-MS/MS technique to determine an adduct profile. Analysis of wild-type protein will determine the substrate specificity of the enzyme. Alignment of AlkC with AlkD allows for the identification of residues that may be involved in catalysis and binding of alkylpurine-DNA substrates (Figure 41). Individual constructs containing eight different AlkC mutations have been created and these mutants will be expressed, purified, and tested with this assay to determine individual adduct profiles. This will help determine the roles of individual residues in substrate excision.

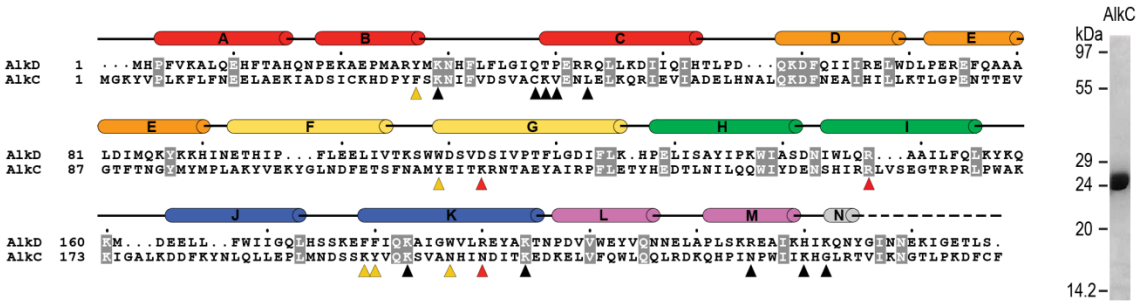


Figure 41. AlkC/AlkD sequence alignment. Triangles denote DNA binding (black), active site (yellow), and catalytic residues (red) in AlkD. SDS-PAGE of purified AlkC is shown on the right.

B. cereus AlkC can be expressed and purified to milligram quantities (Figure 41). Co-crystallization trials of AlkC in complex to oligonucleotides of differing length containing a centrally located THF abasic site analog and free 3mA nucleobase have been underway but have yet to produce diffraction quality protein or protein/DNA crystals. Once a high-resolution native data set is obtained from diffraction-quality crystals, experimental phases will be determined using heavy-atom derivitized crystals which will allow for the determination of electron density and subsequent model building, phase improvement and refinement. AlkC will also be co-crystallized in the presence of an oligonucleotide containing a centrally located non-hydrolyzable 3d3mA substrate analog for structural determination. The results of which AlkD and AlkC mutants affect substrate specificity by the HPLC-MS/MS method will round out the biochemical analysis of each enzyme.

AlkD as a Molecular Tool for Cancer Prevention and Therapeutics

DNA has been a molecular target for cancer chemotherapies since it was discovered, paradoxically, that exogenous alkylating agents, many of which cause cancers, could be used to treat lymphomas due to their cytotoxic effects on the highly proliferative cells, a marker of cancers (Hurley, 2002). Simple methylating agents, like temozolomide or streptozotocin, and chloroethylating agents, like carmustine or lomustine, continue to be used in the clinic as chemotherapies. Under normal circumstances the base excision repair pathway defends against the toxic and carcinogenic effects of alkylating agents and it has been shown that mouse embryonic stem cells lacking AAG are sensitive to alkylating agents (Engelward et al, 1997). In

cancer patients, however, this protective mechanism hampers the effect of alkylating agent-derived cancer treatments (Allan et al, 1998b). Recently, the inhibition of BER to sensitize cancer cells to alkylating drugs has been explored, (Fishel et al, 2007; Liu et al, 2002) and it has been shown that depleting HeLa cells, a cervical carcinoma cell line, of AAG increases sensitivity to several laboratory and chemotherapeutic alkylating agents including temozolomide (Paik et al, 2005). Understanding how cytotoxic lesions like 3mA are selected for over mutagenic lesions (e.g., εA and 7mG) needs to be considered so alkylation therapy can be used effectively in the clinic.

Structural analysis of AlkD bound to DNA strengthens the idea that some DNA glycosylases promote steric strain of these labile lesions to select for potentially toxic damaged bases. Combined with structural studies of TAG and UDG, the structures of AlkD suggest that positively charged methylpurines, specifically the cytotoxic 3mA lesion, are processed differently from other alkylpurine lesions such as ethenoadenine and hypoxanthine and is based on the enzyme's ability to further weaken the glycosylic bond. It is therefore possible that promiscuous base-flipping enzymes like AlkA and AAG use multiple modes of catalysis to excise different alkyl-lesions. This difference may lead to the development of alkylpurine DNA glycosylase inhibitors to be used in conjunction with chemotherapeutic alkylating agents. Since the success of these therapies depends on the cytotoxic effects of 3mA lesions, it is more beneficial to target 3mA specific catalysis of these enzymes over catalysis of mutagenic lesions. This mechanistic difference may have other benefits as well; it was shown that AlkD is able to remove bulky, positively-charged POB-lesions that are the result of tobacco-derived carcinogens. Therefore, one

can imagine that specially designed AlkD-like molecules could be used to remove these lesions from DNA within cells in the hopes of reducing their mutagenic potential.

In summary, research on AlkD has helped shed light on how the intrinsic biochemical and biophysical properties of positively charged alkylpurines help determine how they are removed from the genome. The determination of AlkD's tandem helical repeat architecture shows how a motif normally used for protein-protein interactions binds DNA, which may help identify DNA binding regions of other HEAT-repeat enzymes involved in DNA processing. Further examination of how AlkD, and eventually AlkC, interact with DNA will add to the knowledge of how DNA glycosylases scan and probe DNA for base lesions. Analysis of the AlkD/DNA crystal structures provide a rationale for why positively charged methylpurines can be selected for over uncharged DNA damaging lesions and that enhanced binding energy may be a reasonable mechanism of 3mA and 7mG catalysis by other DNA glycosylases. This body of work may one day allow for the development of molecular tools to help prevent and treat cancer.

APPENDIX A*

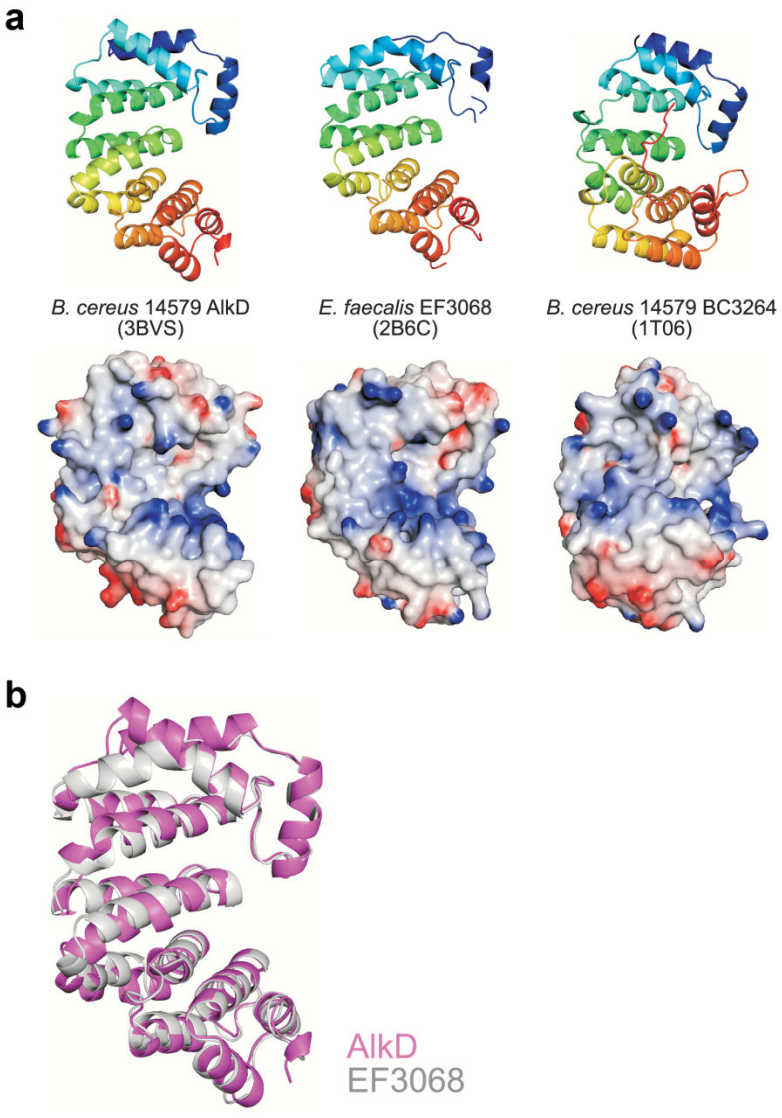


Figure A1. Comparison of the closest AlkD structural orthologs. (**a**) Ribbon and electrostatic potential surface representations of *B. cereus* AlkD, *E. faecalis* EF3068 (PDB code 2B6C) and *B. cereus* BC3264 (PDB code 1T06) are shown. (**b**) Superposition of AlkD and EF3068.

* The work presented in this chapter was published in 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. *Journal of molecular biology* **381**, 13-23.

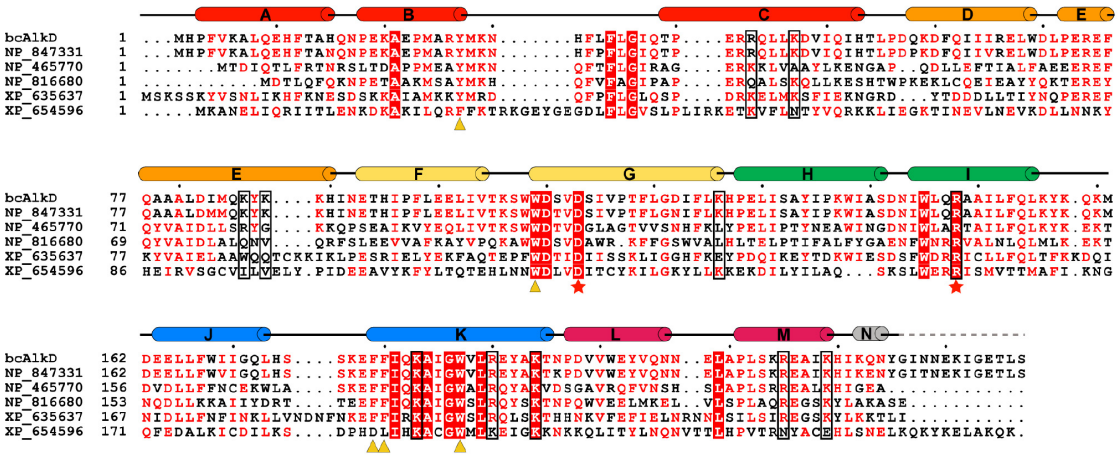


Figure A2. Sequence alignment of five putative AlkD homologs. The sequences of *Bacillus cereus* AlkD, *Bacillus anthracis* (NP_847331), *Listeria monocytogenes* (NP_465770), *Enterococcus faecalis* (NP_816680), *Dictyostelium discoideum* (XP_635637), and *Entamoeba histolytica* (XP_654596) are aligned. The secondary structure identified in the crystal structure is shown schematically above the sequence. Aromatic residues inside the putative active site are marked with a yellow triangle, and Asp113 and Arg148 residues important for 7mG excision and DNA binding are marked with a red asterisk. Positively charged residues lining the concave surface are boxed.

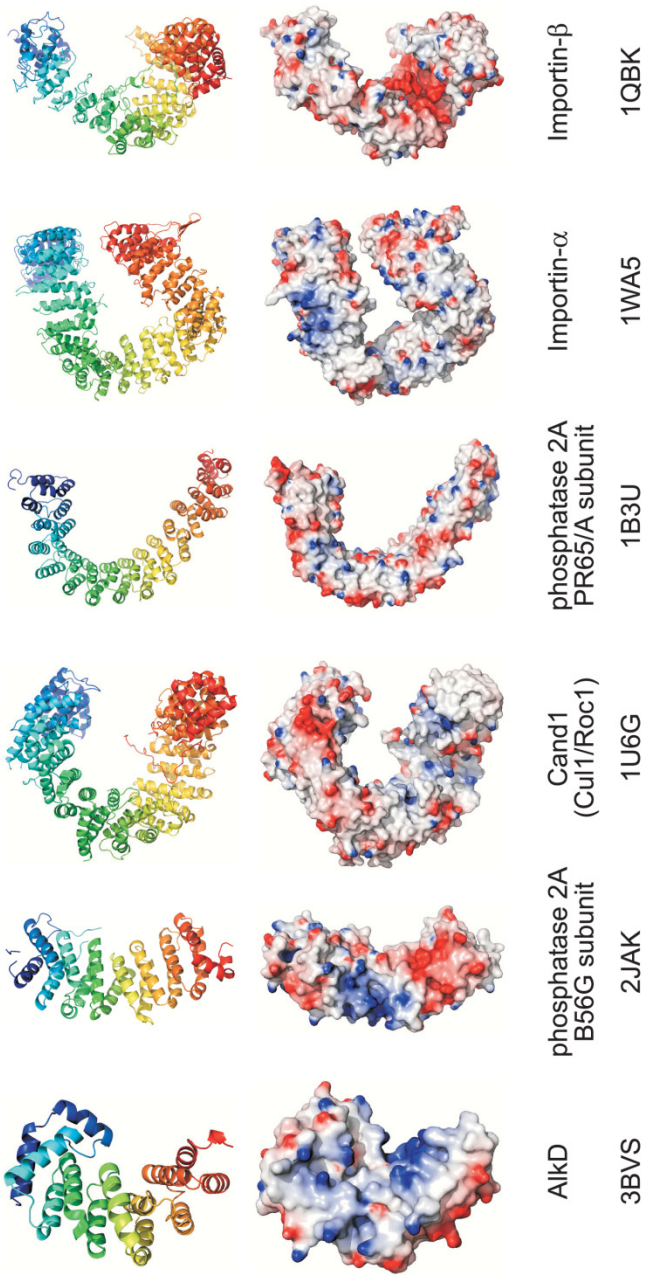


Figure A3. Helical repeat proteins identified by the DALI server as being closely related in structure to AlkD (see Table A1). Shown are HEAT proteins—protein phosphatase 2A (PP2A) 56kD regulatory subunit (2JAK, 2NPP), SCF ubiquitin ligase regulatory subunit Cand1 (1U6G), PP2A 65kD scaffolding subunit (1B3U)—and Armadillo repeat proteins Importin-α (1WA5) and Importin–β (1QBK). Proteins are rendered as ribbon (top) and electrostatic potential surface representations (bottom). Electrostatic potentials (red negative, blue positive, -7 to +7 k_BT) were calculated with the program DelPhi (Rocchia et al, 2002). Note that AlkD is the only structure to contain a positively charged concave surface.

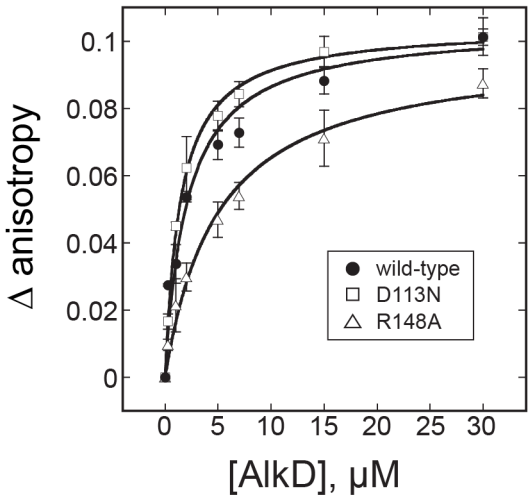


Figure A4. AlkD-DNA binding. Shown is the binding curve for AlkD against tetrahydrofuran (THF)-DNA. Binding of AlkD to abasic DNA was measured by the change in fluorescence anisotropy as AlkD was titrated into a solution containing 25mer oligonucleotide containing an internal THF residue and a 6-carboxyfluorescein moiety at the 3′-end.

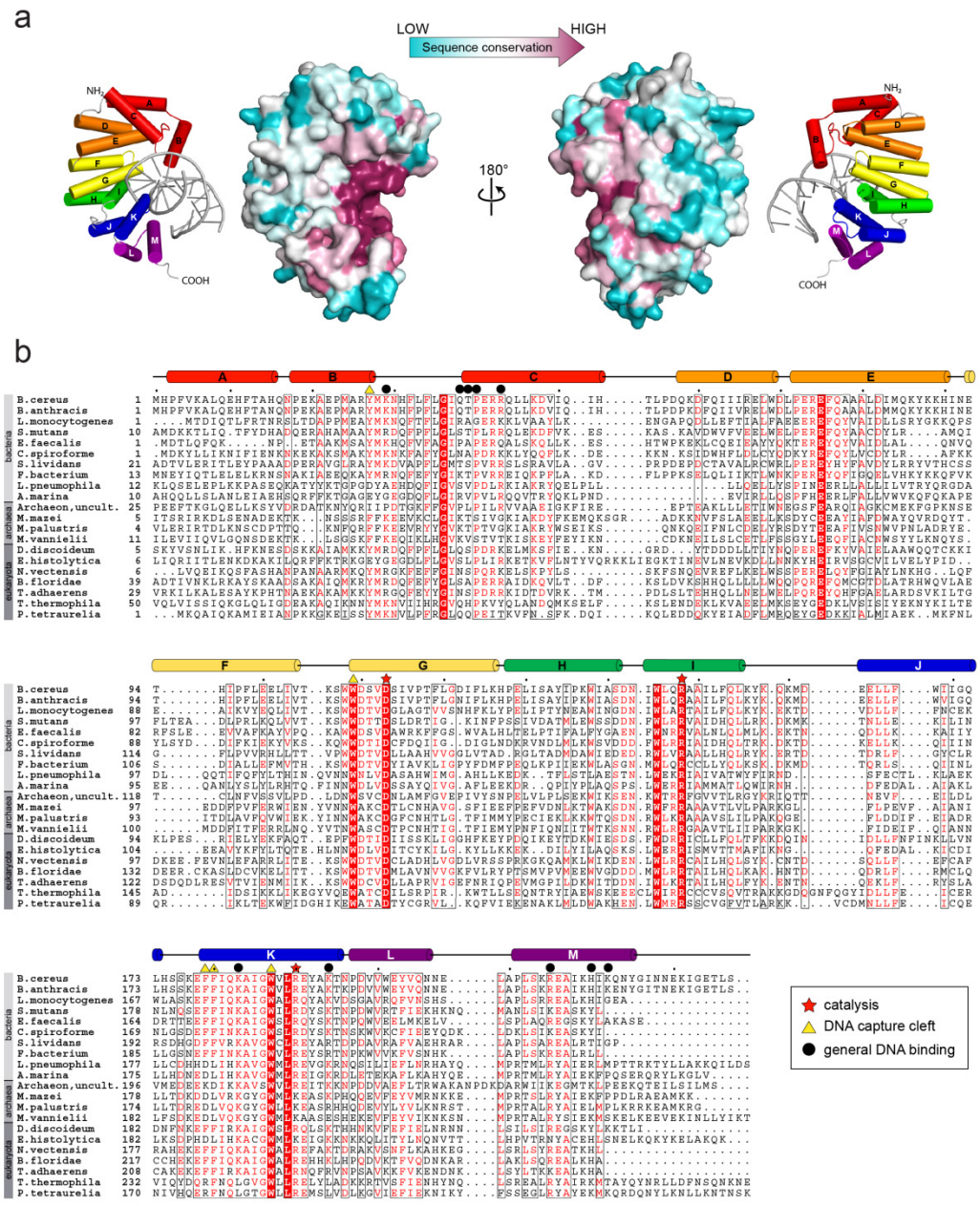
Table A1. Top 20 DALI hits for BcAlkD.

Rank	PDB ID	Chain	Z Score	RMSD	Name	Reference
1	2b6c	A	23.9	2.2	Hypothetical protein, unknown function	(Osipiuk et al.)
2	1t06	A	13	3.8	Hypothetical protein, unknown function	(Zhang et al.)
3	2jak	A	8.6	6.3	Protein phosphatase 2A 56kD regulatory subunit	(Magnusdottir et al.)
4	1u6g	C	8.5	3.9	Cullin interacting protein, Cand1	(Goldenberg et al., 2004)
5	1b3u	A	8.1	3.5	Protein phosphatase 2A 65kD scaffolding subunit	(Groves et al., 1999)
6	2db0	A	7.8	2.9	253aa long hypothetical protein	(Handa et al.)
7	2npp	B	7.6	6.2	Protein phosphatase 2A 56KDa subunit	(Xu et al., 2006)
8	1wa5	C	7.5	4.1	Importin-α	(Matsuura and Stewart, 2004)
9	1qbk	B	7.2	4.2	Importin-β2	(Chook and Blobel, 1999)
10	2vgl	A,B	7	8.9	Adaptor-related protein complex 2 (AP2) α2 subunit	(Collins et al., 2002)
11	1qgr	A	6.4	4.2	Importin-β subunit	(Cingolani et al., 1999)
12	1dvp	A	6.1	3.8	Membrane trafficking protein Hrs	(Mao et al., 2000)
13	3bct	A	6	4.5	β-catenin (*Mus musculus*)	(Huber et al., 1997)
14	1oyz	A	6	3.6	Hypothetical protein, unknown function	(Kuzin et al.)
15	2fv2	A	5.9	7.3	Required for cell differentiation protein, Rcd-1	(Garces et al., 2007)
16	1ee4	A	5.9	4.7	Importin-α	(Conti and Kuriyan, 2000)
17	2gl7	A	5.7	9.4	β-catenin	(Sampietro et al., 2006)
18	2ilr	A	5.5	7.9	Fanconi Anemia Group E protein, FANCE	(Nookala et al., 2007)
19	1upk	A	5.3	5.6	Mouse protein 25α, MO25α	(Milburn et al., 2004)
20	1ft1	A	5.3	3.3	Protein farnesyltransferase, FTase	(Park et al., 1997)

APPENDIX B*

Figure B1. Evolutionary conservation of AlkD. **a**, The sequence conservation across 266 organisms from eukaryota, archaea, and bacteria are superimposed on the molecular surface of AlkD. The degree of conservation ranges from cyan (little or no conservation) to purple (highly conserved). Front and back views are shown and are compared to a cylinder representation for reference. The figure was prepared using ConSeq (http://conseq.tau.ac.il) and ConSurf (http://consurf.tau.ac.il) programs. **b**, Sequence alignment of 21 representative AlkD orthologs from eukaryota, archaea, and bacteria. Invariant residues are highlighted with red boxes and conserved residues are shown in red text. Secondary structure is shown schematically and colored by HEAT repeat. Residues that decrease enzymatic activity are marked with red stars, those forming the concave cleft are marked with yellow triangles, and those contacting the DNA in the crystal structures are marked with black circles.

* The work presented in this chapter was published in Rubinson EH, Gowda AS, Spratt TED, Gold B, Eichman BF (2010) An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* **468**, 406-411.


a
LOW
HIGH
Sequence conservation
NH2
COOH
180°
b
A
B
C
D
E
F
G
H
I
J
K
L
M
bacteria
archaea
eukaryota
catalysis
DNA capture cleft
general DNA binding

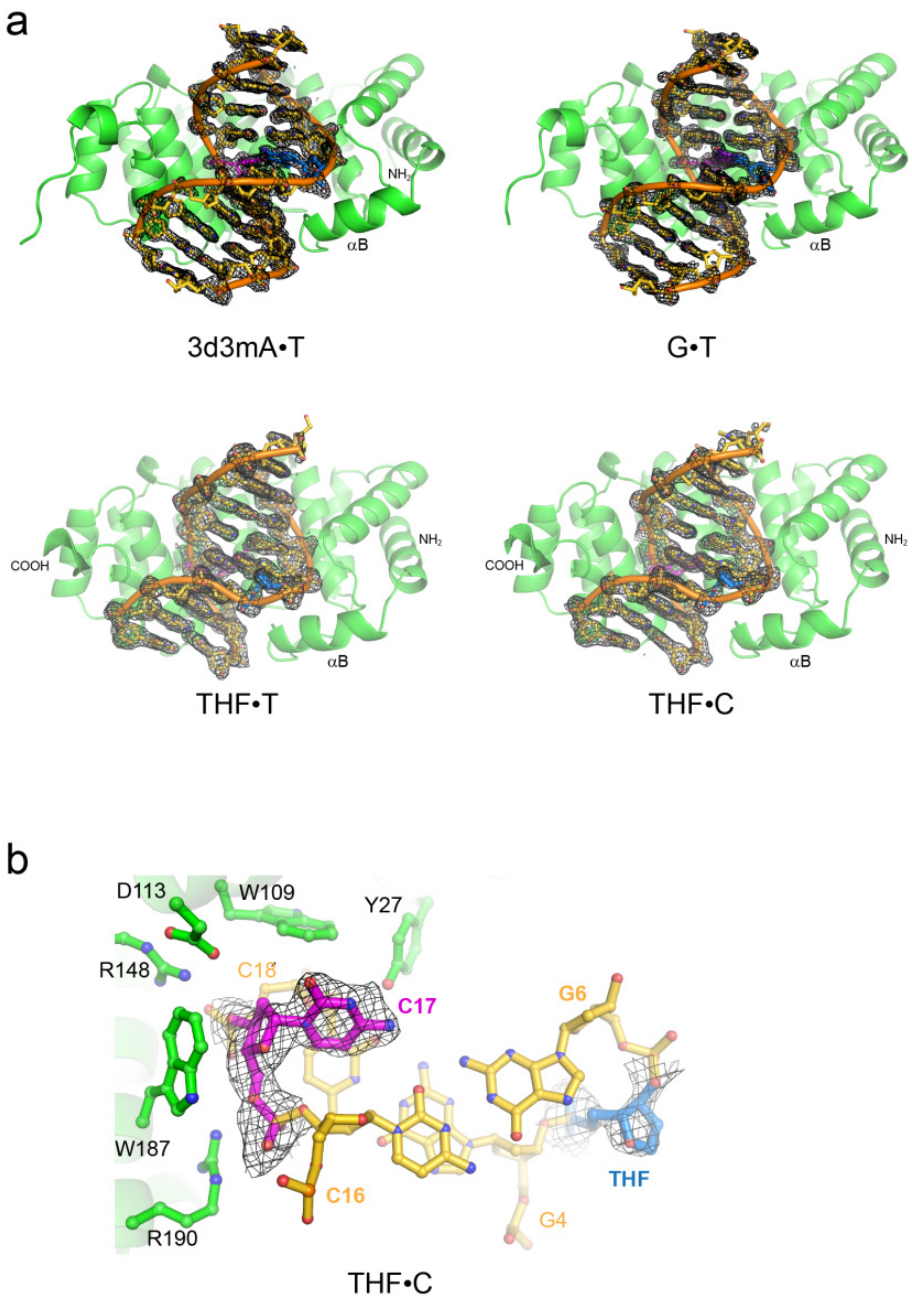


Figure B2. Molecular replacement solutions of AlkD complexes. **a**, Composite omit DNA electron density, contoured to 1σ, was calculated from the refined AlkD/DNA models using the program CNS. For clarity, only density corresponding to the DNA is superimposed onto the structures of each AlkD/DNA complex. Protein is colored green, DNA is gold with orange backbone trace, lesioned nucleotides (3d3mA, G, THF) are blue; opposite nucleotides (T or C) are magenta. **b**, Close-up view of the THF•C pair, with composite omit electron density superimposed. The view is identical to Fig. 3.

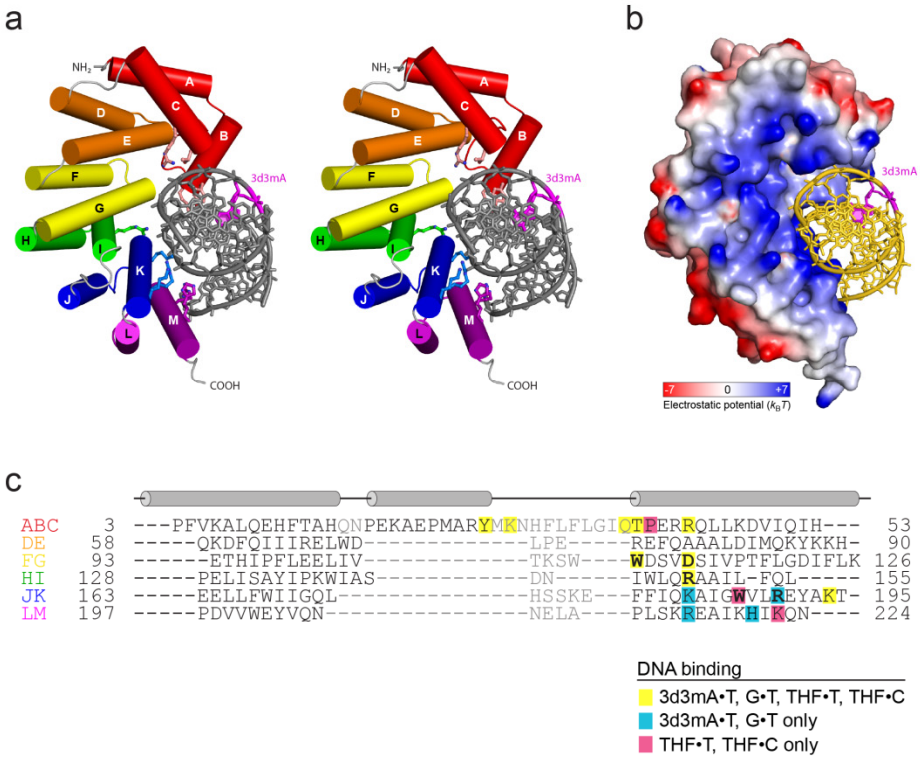


Figure B3. DNA binding by the HEAT repeats of AlkD. **a**, Stereoimage of the AlkD-3d3mA complex shown in Figure 1e. AlkD contains six tandem HEAT-like repeats (colored independently). Side chains that contact the DNA are shown as ball and stick. The 3d3mA nucleotide is colored magenta. **b**, Electrostatic surface potential (blue, positive; red, negative) of AlkD showing a high degree of positive charge within the concave DNA binding cleft. **c**, Structure based sequence alignment of HEAT repeats. Residues that contact the DNA in both substrate (3d3mA•T, G•T) and product (THF•T, THF•C) complexes are highlighted yellow, residues contacting the DNA in only substrate or only product are highlighted blue and magenta, respectively.

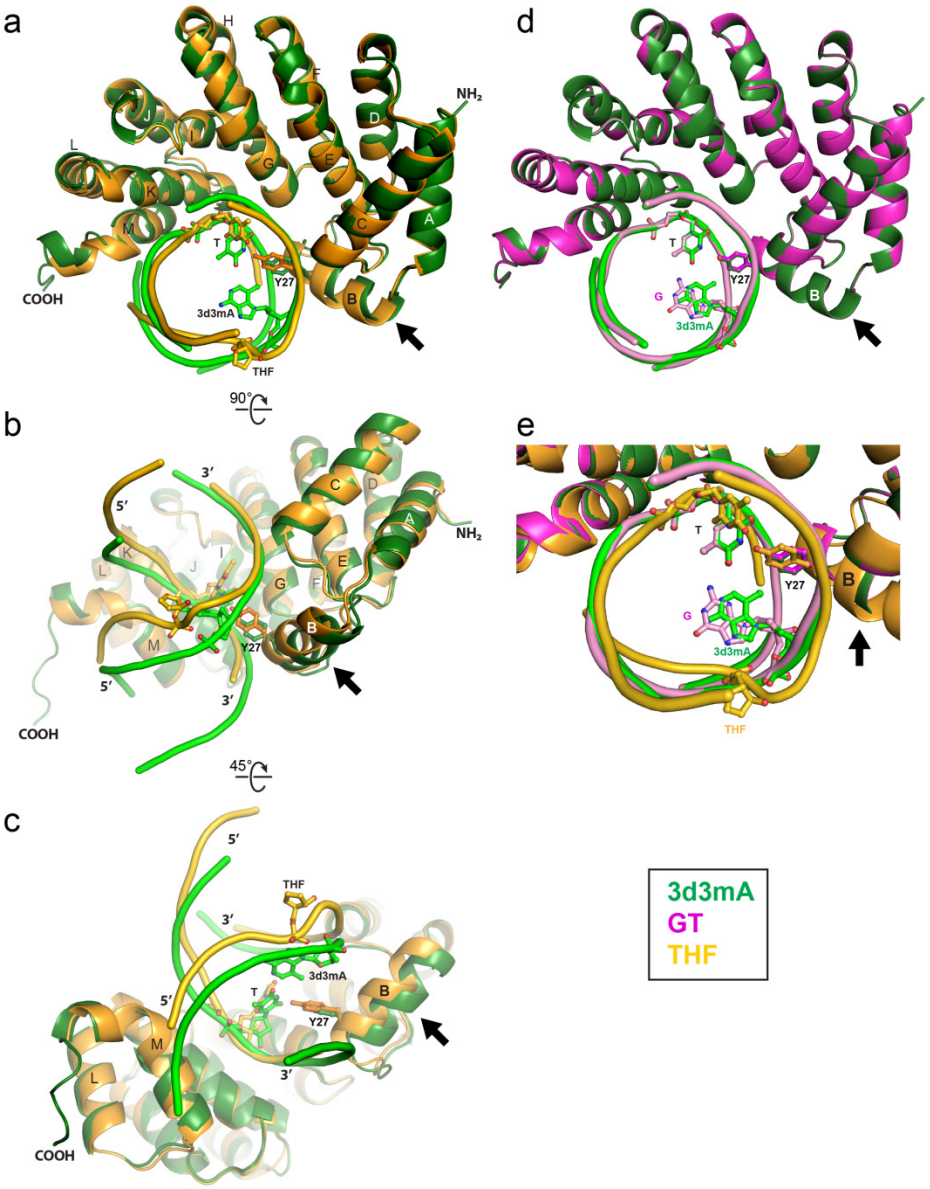


Figure B4. Distortion to substrate and product DNA by AlkD. **a,b,c**, Three different views of a superposition of 3d3mA•T (green) and THF•T (gold) complexes. Only the lesioned base pairs and DNA backbone are shown for clarity. Protein residue Y27 in both structures is shown as sticks. Helix αB (highlighted with a black arrow) is the only noticeable difference in the protein in the two structures. **d**, Superposition of 3d3mA•T (green) and G•T (magenta) structures. **e**, Close-up view of the three AlkD/DNA complexes with thymine opposite the lesion.

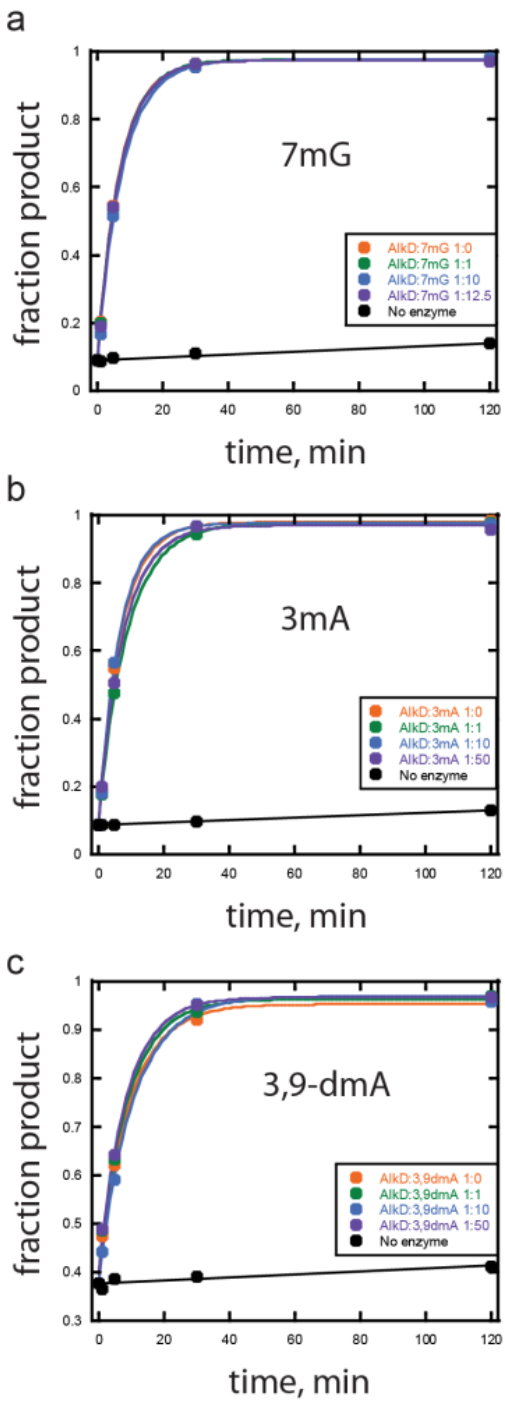


Figure B5. Excision of 7mG by AlkD is not inhibited by free nucleobases. The rates of 7mG release from 25mer DNA were measured after pre-incubating AlkD with various concentrations of (**a**) 7mG, (**b**) 3mA, and (**c**) 3,9-dimethyladenine (3,9-dmA) up to the solubility limit. The data were repeated in triplicate and is shown for one representative experiment only.

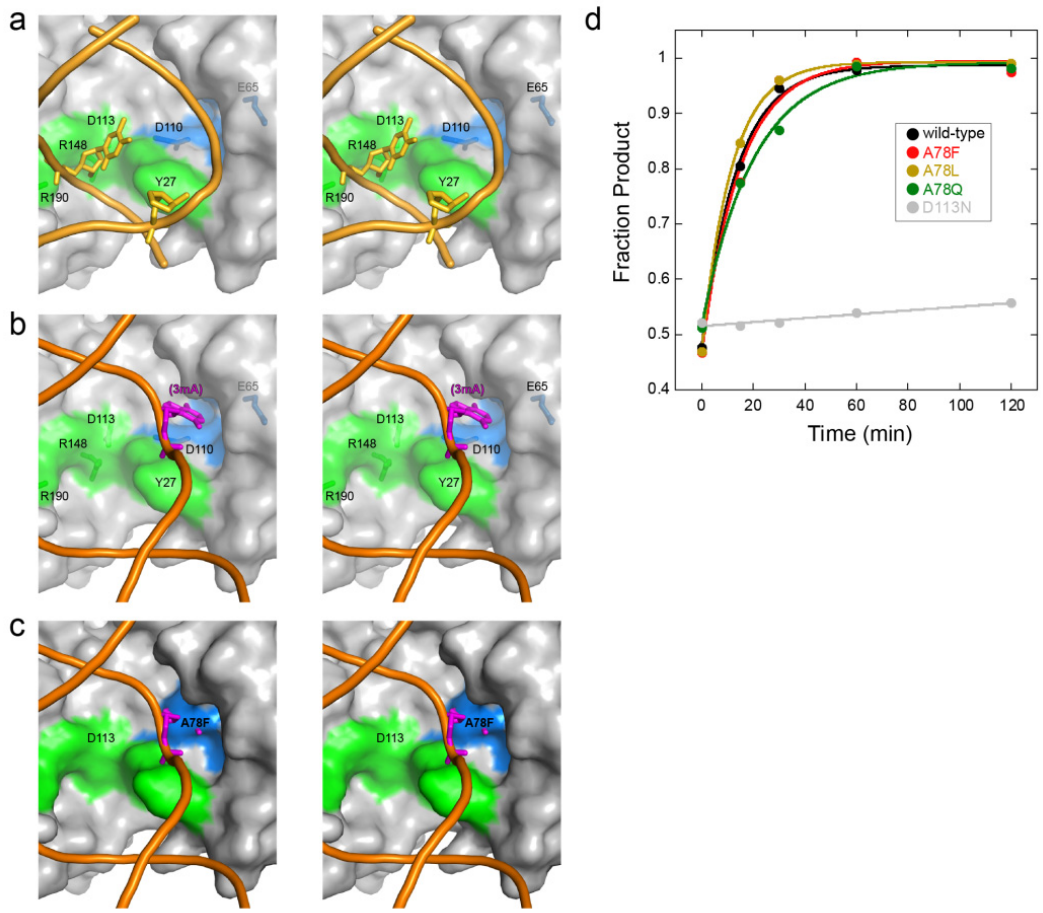


Figure B6. A putative nucleobase binding pocket does not affect base excision activity. A surface-exposed cleft bounded in part by residues D110, A78, F34, and E75 (shaded blue) and potentially suitable for binding nucleobases was located using PocketFinder (http://www.modelling.leeds.ac.uk/pocketfinder/). In panels **a**, **b**, and **c**, AlkD from the THF•T co-crystal structure is shown as a molecular surface, with green marking the DNA capture site identified in the crystal structures (Y27, D113, R148, W187, R190) and blue marking the putative base binding pocket. Each image is shown in stereo, with the view looking into the concave cleft of the protein. **a**, The AlkD/THF•T-DNA complex, with the DNA backbone colored gold and THF•T nucleotides shown as sticks. **b**, DNA (orange) from the human AAG/εA-DNA co-crystal structure (PDB id 1EWN) was manually docked on the AlkD crystal structure, and the εA base substituted with 3mA (magenta). Despite a snug fit of 3mA, the DNA could not be positioned against the protein surface without large-scale conformational rearrangement to alleviate steric clashes to the N- and C-terminal regions of the protein. **c**, Molecular model of the A78F mutation. Substitution of Ala78 with phenylalanine fills the pocket and blocks 3mA access. **d**, The 7mG excision activity of A78F (red), A78L (gold), and A78Q (green) AlkD mutants are shown relative to wild-type (black) and D113N (grey). The data were repeated in triplicate and is shown for one representative experiment only.

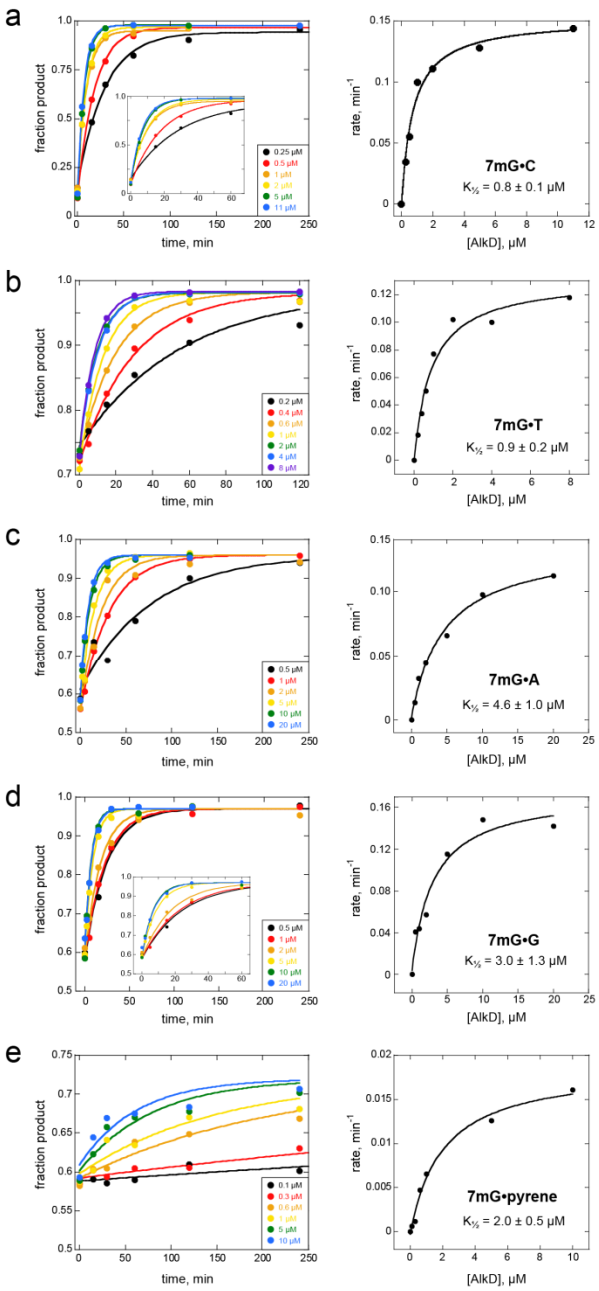


Figure B7. Determination of $K_{½}$ for AlkD catalyzed excision of 7mG paired with C, T, A, G, or pyrene on the opposite DNA strand. *Left panels,* Rates of 7mG release from 25mer DNA at varying concentrations of AlkD are shown for one representative experiment only. The insets show the initial burst phase of the reactions. *Right panels*, rate constants (k_{obs}) obtained from single-exponential fits to the kinetic traces were plotted against enzyme concentration, and the $K_{½}$ was determined by fitting the data to the equation, $k_{obs} = V_{max}[AlkD] / (K_{½} + [AlkD])$. Plots represent data from one experiment, and $K_{½}$ values are the averages and standard deviations from 3-4 separate experiments.

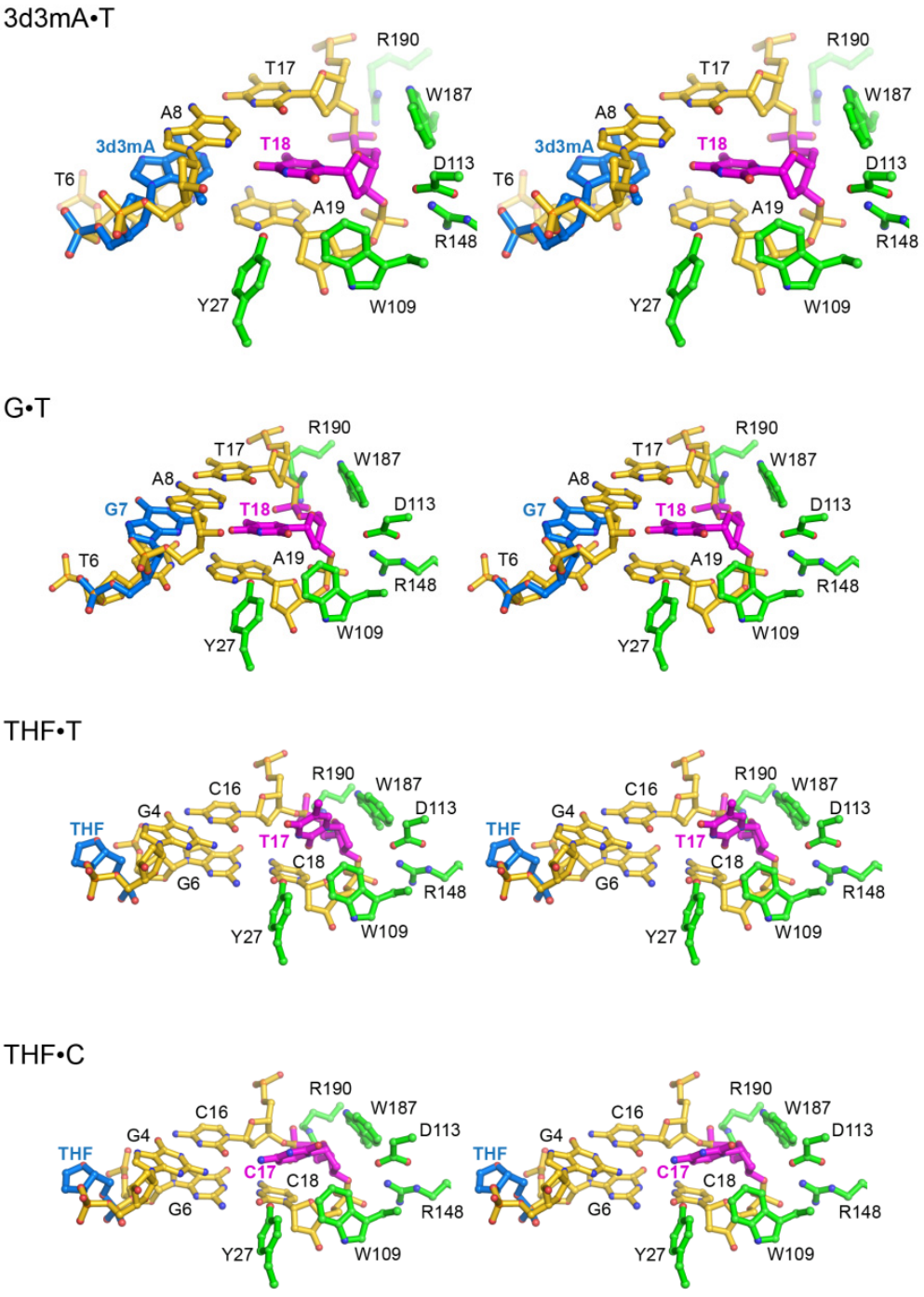


Figure B8. Atomic details of the AlkD-DNA damage interaction. Stereo views of each of the four complexes are shown, with protein residues colored green, DNA gold, lesioned bases (3d3mA, G, and THF) blue, and opposite bases (T and C) magenta.

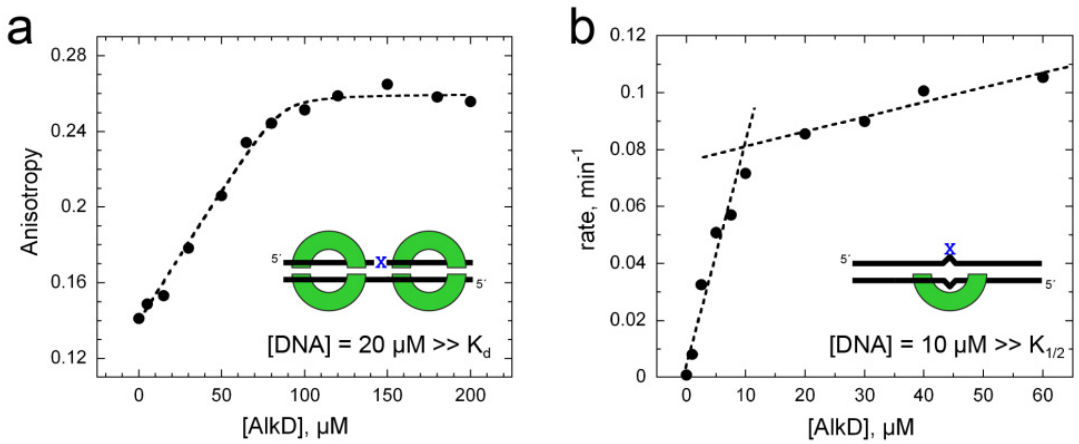


Figure B9. Stoichiometry of AlkD/DNA binding and catalysis. **a**, Stoichiometric binding of AlkD to 25mer G•T-DNA monitored by fluorescence anisotropy in the presence of excess unlabeled DNA (20 uM). Saturation of binding at 85 μM AlkD shows that four AlkD molecules can bind a 25mer length of DNA. **b**, Excision of 7mG-DNA under stoichiometric conditions, in which the concentration of DNA was held 10-fold above the $K_{1/2}$ for the reaction. Saturation of the rate at ~10 uM AlkD demonstrates that only one AlkD per lesion is needed for catalysis.

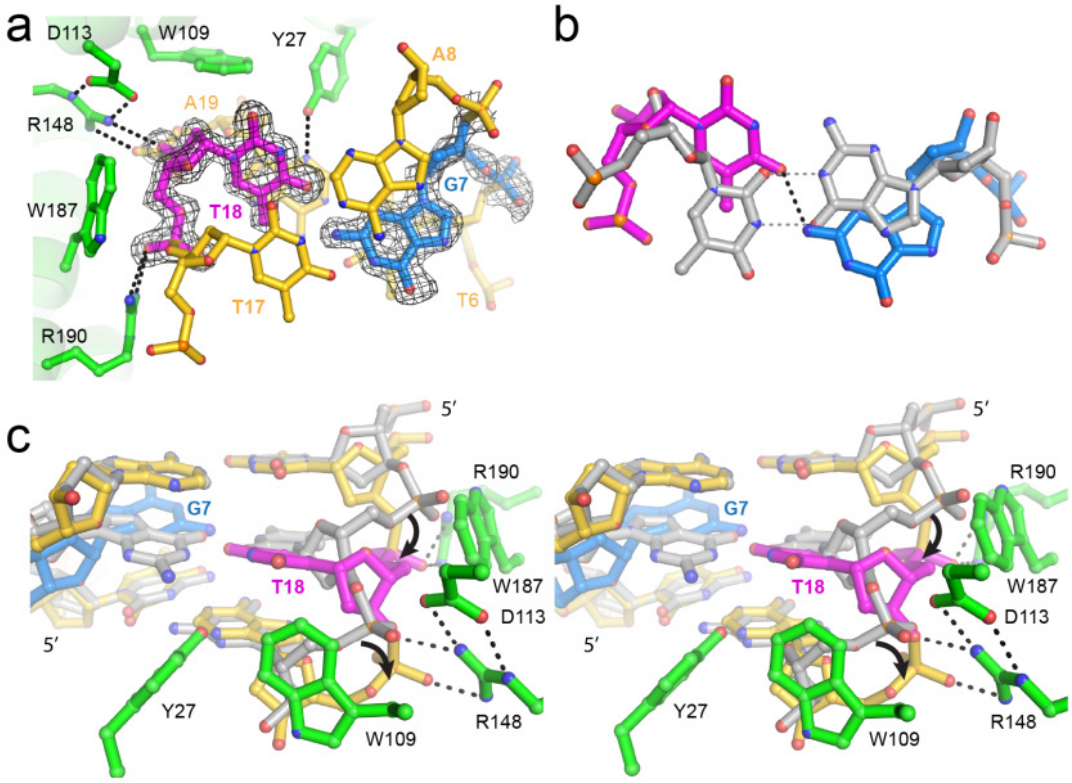
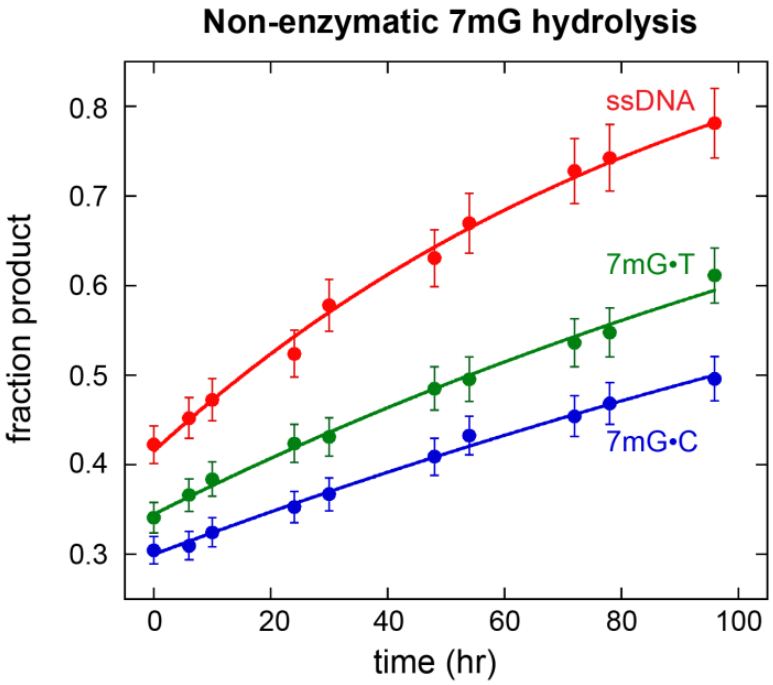


Figure B10. AlkD manipulation of a G•T mismatch. **a**, The AlkD/G•T-DNA crystal structure, viewed down the DNA helix axis and colored as in Fig. 3. Composite omit electron density (1σ contour) is shown for the G•T base pair. **b**, Superposition of the G•T base pair in the AlkD complex (blue G, magenta T) onto a canonical wobble G•T mispair (grey, PDB code 113D) (Hunter et al, 1987). **c**, Stereo view of the wobble G•T-DNA (grey) superimposed onto the AlkD/G•T-DNA complex (colored). Dashed lines represent hydrogen bonds and black arrows highlight the differences between DNA backbone conformations.



Rates of spontaneous d7mG hydrolysis from DNA

	Substrate		k_{non} (10^{-3} h^{-1})	$t_{1/2}$ (h)
●	d7mG•dC	dsDNA	3.5 ± 0.2 [a]	198
	d7mG	dsDNA		192 [b]
●	d7mG•dT	dsDNA	5.0 ± 0.2 [a]	138
●	d7mG	ssDNA	10.2 ± 0.5 [a]	68
	5′-d7mGMP	nucleotide		23 [c]
	d7mG	nucleoside		6.5 [d]

[a] Non-enzymatic rates (k_{non}) were measured at 37° and pH 6.5 as described in *Methods*

[b] 37°, pH 7.2 (Margison, G. P., *et al* (1976) *Biochem. J. 157*, 627-634)

[c] 37°, pH 7.4 (Kriek, E. and Emmelot, P. (1964) *Biochim. Biophs. Acta 91*, 59-66)

[d] 37°, pH 7 (Singer, B., and Grunberger, D. (1983) *Molecular Biology of Mutagens and Carcinogens*, Plenum, New York)

Figure B11. Non-enzymatic hydrolysis of 7mG from DNA. Top panel shows the kinetic traces for spontaneous release of 7mG from ssDNA (red) and dsDNA containing a 7mG•C (blue) or 7mG•T (green) base pair. Reactions were performed at 37° C, pH 6.4, and 100 mM ionic strength.. The bottom panel shows the quantitation from the top panel together with data reported in the literature for release of 7mG from dsDNA, 7-methyldeoxyguanosine-5′-monophosphate (5′-d7mGMP), and 7-methyldeoxyguanosine (d7mG). Error bars represent the standard deviation from three independent measurements.

Table B1. Data collection and refinement statistics

	3d3mA•T	G•T	THF•T	THF•C
Data collection				
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions				
a , *b* , *c* (Å)	37.76, 92.97, 47.77	38.20, 93.59, 48.12	52.33, 57.45, 53.65	52.13, 57.49, 54.07
a, b, g (°)	90, 113.01, 90	90, 112.60, 90	90, 101.88, 90	90, 101.67, 90
Wavelength	0.9785	0.9785	0.9785	0.9785
Resolution (Å)	44.0-1.60 (1.66-1.60)	35.3-1.50 (1.55-1.50)	28.9-1.75 (1.81-1.75)	41.1-1.75 (1.81-1.75)
R_{sym}	0.051 (0.247)	0.043 (0.186)	0.056 (0.467)	0.049 (0.351)
I / sI	24.8 (3.5)	22.0 (4.9)	13.5 (2.9)	30.8 (3.0)
Completeness (%)	95.8 (75.3)	93.2 (87.2)	95.5 (94.8)	92.3 (65.7)
Redundancy	4.5 (3.4)	3.8 (3.8)	3.3 (3.3)	3.6 (3.2)
Refinement				
Resolution (Å)	1.6	1.5	1.75	1.75
No. reflections	38,310	46,341	30,134	28,987
R_{work} / R_{free}	0.159 / 0.183	0.130 / 0.174	0.185 / 0.225	0.183 / 0.204
No. atoms				
Protein	1967	1949	1902	1888
DNA	487	487	393	392
Water	446	525	274	220
B-factors				
Protein/DNA	21.1	20.1	31.2	43.3
Water	38.7	43.2	43	50.7
R.m.s. deviations				
Bond lengths (Å)	0.006	0.005	0.006	0.005
Bond angles (°)	1.091	0.986	1.029	0.997
Ramachandran Statistics				
Favored (%)	92.9	92.9	91.8	92.8
Allowed (%)	6.1	6.1	7.2	6.2
Generously allowed (%)	0.5	0.5	0	0.5
Disallowed (%)	0.5	0.5	1	0.5
PDB ID code	3JX7	3JXY	3JXZ	3JY1

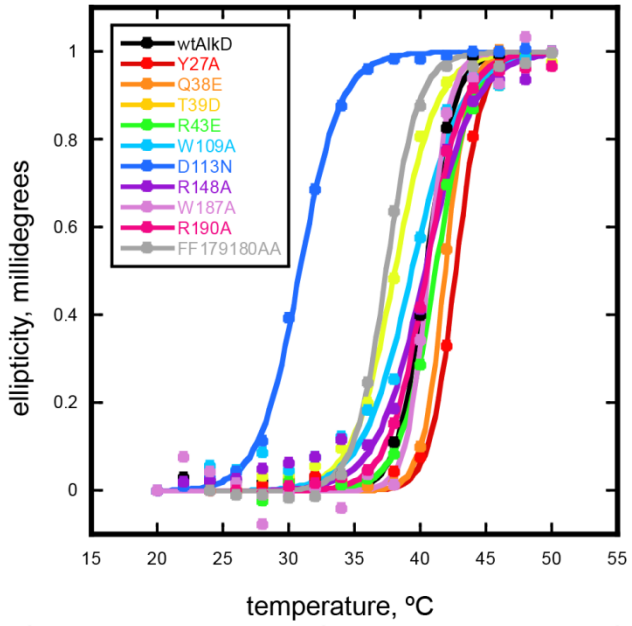
Table B2. Opposite base effects on 7mG excision by AlkD

Substrate	kst (10-5 s-1)	K1/2 (μM)	kst/K1/2 (s-1 M-1)
7mG ssDNA	0.6 ± 0.02		
7mG•C	248 ± 12	0.8 ± 0.1	3.0×10^3
7mG•T	209 ± 17	0.9 ± 0.2	2.3×10^3
7mG•A	260 ± 34	3.0 ± 1.3	8.6×10^2
7mG•G	206 ± 27	4.6 ± 1.1	4.5×10^2
7mG•pyrene	24 ± 4	2.0 ± 0.5	1.2×10^2

Single-turnover excision rates (k_{st}) and the concentration of enzyme at half-maximal activity ($K_{1/2}$) for 7mG across from different bases within a 25mer oligonucleotide were measured at pH 7.5 and 100 mM ionic strength. Errors represent the standard deviation from three independent measurements. Raw data is shown in Figure B8.

APPENDIX C

a



b

Enzyme	T_m (°C)
wtAlkD	40.4
Y27A	42.6
Q38E	41.8
T39D	38.0
R43E	41.1
W109A	39.3
D113N	30.7
R148A	40.3
W187A	40.5
R190A	40.4
FF179180AA	37.4

Figure C1. Thermal denaturation of AlkD mutants. Protein unfolding was monitored by circular dichroism spectroscopy and following the change in molar elipticity at 222 nm as a function of temperature. **a**, Melting curves for AlkD mutants. **b**, Table of melting temperatures (T_m).

Table C1. DNA binding activities for wild-type and mutants of AlkD

Enzyme	K_d (μM)					
	G•C	Relative	G•T	Relative	THF•C	Relative
wtAlkD	1.73 ± 0.22	1.00	1.38 ± 0.26	1.00	2.82 ± 0.76	1.00
Y27A	1.53 ± 0.34	1.13	3.01 ± 0.46	0.46	4.59 ± 0.59	0.61
Y27F	1.48 ± 0.12	1.17	1.45 ± 0.26	0.95	2.88 ± 0.61	0.98
Q38E	3.03 ± 0.26	0.57	3.78 ± 0.28	0.37	7.63 ± 1.24	0.37
T39D	6.37 ± 1.06	0.27	3.12 ± 0.71	0.44	8.56 ± 1.80	0.33
R43E	12.8 ± 0.23	0.14	14.7 ± 1.98	0.09	11.7 ± 1.54	0.24
W109A	3.72 ± 1.19	0.47	2.15 ± 0.13	0.64	4.28 ± 0.57	0.66
D113N	0.68 ± 0.03	2.54	0.99 ± 0.06	1.39	1.26 ± 0.12	2.23
R148A	4.59 ± 0.78	0.38	3.86 ± 0.57	0.36	8.67 ± 0.37	0.33
FF179180AA	2.39 ± 0.50	0.72	2.34 ± 0.85	0.59	4.65 ± 1.06	0.61
FF179180DL	12.9 ± 0.95	0.13	6.13 ± 2.29	0.22	8.05 ± 1.59	0.35
W187A	3.62 ± 0.70	0.48	4.71 ± 0.49	0.29	7.55 ± 0.88	0.37
R190A	0.83 ± 0.07	2.08	1.05 ± 0.23	1.31	1.29 ± 0.09	2.18
R190S	0.86 ± 0.22	2.01	1.09 ± 0.30	1.27	2.38 ± 0.33	1.18

Dissociation constants (K_d) for a 25mer oligonucleotide duplex containing the specified modification paired with cytosine were measured by fluorescence anisotropy as described in Chapter IV.

REFERENCES

Aas PA, Otterlei M, Falnes PO, Vagbo CB, Skorpen F, Akbari M, Sundheim O, Bjoras M, Slupphaug G, Seeberg E, Krokan HE (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421:** 859-863

Abdel-Hamid M, Novotny L, Hamza H (2000) Stability study of selected adenosine nucleosides using LC and LC/MS analyses. *Journal of pharmaceutical and biomedical analysis* **22:** 745-755

Aboul-ela F, Koh D, Tinoco I, Jr., Martin FH (1985) Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A,C,G,T). *Nucleic acids research* **13:** 4811-4824

Adams PD, Afonine PV, Grosse-Kunstleve RW, Moriarty NW, Sauter NK, Zwart PH, Gopal K, Ioerger TR, Kanbi L, McKee E, Pai RK, Hung L-W, Radhakannan T, McCoy AJ, Read RJ, Storoni LC, Romo TD, Sacchettini JC, Terwilliger. TC (2007) Automated structure determination with Phenix. In *Evolving Methods for Macromolecular Crystallography*, Read RJ, Sussman JL (eds), pp 101-109. Dordrecht: Springer

Allan BW, Beechem JM, Lindstrom WM, Reich NO (1998a) Direct real time observation of base flipping by the EcoRI DNA methyltransferase. *The Journal of biological chemistry* **273:** 2368-2373

Allan JM, Engelward BP, Dreslin AJ, Wyatt MD, Tomasz M, Samson LD (1998b) Mammalian 3-methyladenine DNA glycosylase protects against the toxicity and clastogenicity of certain chemotherapeutic DNA cross-linking agents. *Cancer Res* **58:** 3965-3973

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. *Molecular microbiology* **59:** 1602-1609

Andrade MA, Bork P (1995) HEAT repeats in the Huntington's disease protein. *Nature Genetics* **11:** 115-116

Andrade MA, Perez-Iratxeta C, Ponting CP (2001a) Protein repeats: structures, functions, and evolution. *Journal of structural biology* **134:** 117-131

Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P (2001b) Comparison of ARM and HEAT protein repeats. *Journal of molecular biology* **309:** 1-18

Asaeda A, Ide H, Asagoshi K, Matsuyama S, Tano K, Murakami A, Takamori Y, Kubo K (2000) Substrate specificity of human methylpurine DNA N-glycosylase. *Biochemistry* **39:** 1959-1965

Asin-Cayuela J, Schwend T, Farge G, Gustafsson CM (2005) The human mitochondrial transcription termination factor (mTERF) is fully active in vitro in the non-phosphorylated form. *The Journal of biological chemistry* **280:** 25499-25505

Avery OT, Macleod CM, McCarty M (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *Journal of Experimental Medicine* **79:** 137-158

Baldwin MR, O'Brien PJ (2009) Human AP endonuclease 1 stimulates multiple-turnover base excision by alkyladenine DNA glycosylase. *Biochemistry* **48:** 6022-6033

Bandaru V, Sunkara S, Wallace SS, Bond JP (2002) A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. *DNA repair* **1:** 517-529

Banerjee A, Santos WL, Verdine GL (2006) Structure of a DNA glycosylase searching for lesions. *Science* **311:** 1153-1157

Banerjee A, Yang W, Karplus M, Verdine GL (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434:** 612-618

Barrows LR, Magee PN (1982) Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro. *Carcinogenesis* **3:** 349-351

Begley TJ, Haas BJ, Noel J, Shekhtman A, Williams WA, Cunningham RP (1999) A new member of the endonuclease III family of DNA repair enzymes that removes methylated purines from DNA. *Current Biology* **9:** 653-656

Begley TJ, Samson LD (2003) AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. *Trends in biochemical sciences* **28:** 2-5

Bennett MT, Rodgers MT, Hebert AS, Ruslander LE, Eisele L, Drohat AC (2006) Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *Journal of the American Chemical Society* **128:** 12510-12519

Berdal KG, Bjoras M, Bjelland S, Seeberg E (1990) Cloning and expression in Escherichia coli of a gene for an alkylbase DNA glycosylase from Saccharomyces cerevisiae; a homologue to the bacterial alkA gene. *The EMBO journal* **9:** 4563-4568

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*. *The Journal of biological chemistry* **269:** 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 research* **21:** 2045-2049

Bjoras M, Klungland A, Johansen RF, Seeberg E (1995) Purification and properties of the alkylation repair DNA glycosylase encoded the MAG gene from *Saccharomyces cerevisiae*. *Biochemistry* **34:** 4577-4582

Bjorklund AK, Ekman D, Elofsson A (2006) Expansion of protein domain repeats. *PLoS Computational Biology* **2:** e114

Blainey PC, van Oijen AM, Banerjee A, Verdine GL, Xie XS (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proceedings of the National Academy of Sciences USA* **103:** 5752-5757

Blatch GL, Lassle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21:** 932-939

Boiteux S, Huisman O, Laval J (1984) 3-Methyladenine residues in DNA induce the SOS function sfiA in Escherichia coli. *The EMBO journal* **3:** 2569-2573

Bowles T, Metz AH, O'Quin J, Wawrzak Z, Eichman BF (2008) Structure and DNA binding of alkylation response protein AidB. *Proceedings of the National Academy of Sciences USA* **105:** 15299-15304

Breen AP, Murphy JA (1995) Reactions of oxyl radicals with DNA. *Free Radical Biology and Medicine* **18:** 1033-1077

Brent TP (1979) Partial purification and characterization of a human 3-methyladenine-DNA glycosylase. *Biochemistry* **18:** 911-916

Brookes P (1971) On the interaction of carcinogens with DNA. *Biochemical Pharmacology* **20:** 999-1003

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

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallography D Biological Crystallography* **54 (Pt 5):** 905-921

Buermeyer AB, Deschenes SM, Baker SM, Liskay RM (1999) Mammalian DNA mismatch repair. *Annual Reviews in Genetics* **33:** 533-564

Cadet J, Delatour T, Douki T, Gasparutto D, Pouget JP, Ravanat JL, Sauvaigo S (1999) Hydroxyl radicals and DNA base damage. *Mutation research* **424:** 9-21

Cadet J, Douki T, Frelon S, Sauvaigo S, Pouget JP, Ravanat JL (2002) Assessment of oxidative base damage to isolated and cellular DNA by HPLC-MS/MS measurement. *Free Radical Biology and Medicine* **33:** 441-449

Cai Z, Qian T, Yang MS (2004) Ion-pairing liquid chromatography coupled with mass spectrometry for the simultaneous determination of nucleosides and nucleotides. *Se Pu* **22:** 358-360

Cao C, Kwon K, Jiang YL, Drohat AC, Stivers JT (2003) Solution structure and base perturbation studies reveal a novel mode of alkylated base recognition by 3-methyladenine DNA glycosylase I. *The Journal of biological chemistry* **278:** 48012-48020

Chadt J, Sykora D, Nilsson R, Vodicka P (2008) Monitoring of dimethyl sulphate-induced N3-methyladenine, N7-methylguanine and O(6)-methylguanine DNA adducts using reversed-phase high performance liquid chromatography and mass spectrometry. *Journal of Chromatography B Analytical Technologies in Biomedical and Life Sciences* **867:** 43-48

Chen J, Derfler B, Samson L (1990) Saccharomyces cerevisiae 3-methyladenine DNA glycosylase has homology to the AlkA glycosylase of E. coli and is induced in response to DNA alkylation damage. *The EMBO journal* **9:** 4569-4575

Cingolani G, Petosa C, Weis K, Muller CW (1999) Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature* **399:** 221-229

Cleaver JE (1968) Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218:** 652-656

Connor EE, Wyatt MD (2002) Active-site clashes prevent the human 3-methyladenine DNA glycosylase from improperly removing bases. *Chemistry & Biology* **9:** 1033-1041

Conti E, Kuriyan J (2000) Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. *Structure* **8:** 329-338

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 research* **35:** 2451-2459

Daniels DS, Woo TT, Luu KX, Noll DM, Clarke ND, Pegg AE, Tainer JA (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nature structural & molecular biology* **11:** 714-720

Das AK, Cohen PW, Barford D (1998) The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *The EMBO journal* **17:** 1192-1199

Demple B, Jacobsson A, Olsson M, Robins P, Lindahl T (1982) Repair of alkylated DNA in Escherichia coli. Physical properties of O6-methylguanine-DNA methyltransferase. *The Journal of biological chemistry* **257:** 13776-13780

Denver DR, Swenson SL, Lynch M (2003) An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. *Molecular Biology and Evolution* **20:** 1603-1611

Dianov G, Lindahl T (1994) Reconstitution of the DNA base excision-repair pathway. *Current Biology* **4:** 1069-1076

Dinner AR, Blackburn GM, Karplus M (2001) Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **413:** 752-755

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 research* **24:** 2488-2497

Dong J, Drohat AC, Stivers JT, Pankiewicz KW, Carey PR (2000) Raman spectroscopy of uracil DNA glycosylase-DNA complexes: insights into DNA damage recognition and catalysis. *Biochemistry* **39:** 13241-13250

Dowd DR, Lloyd RS (1989) Biological consequences of a reduction in the non-target DNA scanning capacity of a DNA repair enzyme. *Journal of molecular biology* **208:** 701-707

Drohat AC, Jagadeesh J, Ferguson E, Stivers JT (1999) Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase. *Biochemistry* **38:** 11866-11875

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

Duguid EM, Mishina Y, He C (2003) How do DNA repair proteins locate potential base lesions? a chemical crosslinking method to investigate O6-alkylguanine-DNA alkyltransferases. *Chemistry & Biology* **10:** 827-835

Duguid EM, Rice PA, He C (2005) The structure of the human AGT protein bound to DNA and its implications for damage detection. *Journal of molecular biology* **350:** 657-666

Edwards TA, Pyle SE, Wharton RP, Aggarwal AK (2001) Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell* **105:** 281-289

Eichman BF, O'Rourke EJ, Radicella JP, Ellenberger T (2003) Crystal structures of 3-methyladenine DNA glycosylase MagIII and the recognition of alkylated bases. *The EMBO journal* **22:** 4898-4909

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallography D Biological Crystallography* **60:** 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. *Proceedings of the National Academy of Sciences USA* **94:** 13087-13092

Ezaz-Nikpay K, Verdine GL (1992) Aberrantly methylated DNA: site-specific introduction of N-7-methyl-2'-deoxyguanosine into the Dickerson/Drew dodecamer. *Journal of the American Chemical Society* **114:** 6562-6563

Ezaz-Nikpay K, Verdine GL (1994) The effects of N7-methylguanine on duplex DNA structure. *Chemistry & Biology* **1:** 235-240

Fishel ML, He Y, Smith ML, Kelley MR (2007) Manipulation of base excision repair to sensitize ovarian cancer cells to alkylating agent temozolomide. *Clinical Cancer Research* **13:** 260-267

Forrer P, Binz HK, Stumpp MT, Pluckthun A (2004) Consensus design of repeat proteins. *Chembiochem* **5:** 183-189

Francis AW, David SS (2003) Escherichia coli MutY and Fpg utilize a processive mechanism for target location. *Biochemistry* **42:** 801-810

Friedberg EC, Aguilera A, Gellert M, Hanawalt PC, Hays JB, Lehmann AR, Lindahl T, Lowndes N, Sarasin A, Wood RD (2006) DNA repair: from molecular mechanism to human disease. *DNA repair* **5:** 986-996

Fromme JC, Banerjee A, Verdine GL (2004) DNA glycosylase recognition and catalysis. *Current opinion in structural biology* **14:** 43-49

Fromme JC, Verdine GL (2002) Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature Structural Biology* **9:** 544-552

Fromme JC, Verdine GL (2004) Base excision repair. *Advances in protein chemistry* **69:** 1-41

Fujii T, Itaya T (1999) Systematic tables of mono- and poly-N-methylated adenines: Acid dissociation constants and UV and NMR spectral data. *Heterocycles* **51:** 2255-2277

Garces RG, Gillon W, Pai EF (2007) Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with in vitro nucleic acid binding properties. *Protein Science* **16:** 176-188

Gates KS, Nooner T, Dutta S (2004) Biologically relevant chemical reactions of N7-alkylguanine residues in DNA. *Chemical Research in Toxicology* **17:** 839-856

Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS, Yeo GS, McDonough MA, Cunliffe S, McNeill LA, Galvanovskis J, Rorsman P, Robins P, Prieur X, Coll AP, Ma M, Jovanovic Z, Farooqi IS, Sedgwick B, Barroso I, Lindahl T, Ponting CP, Ashcroft FM, O'Rahilly S, Schofield CJ (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318:** 1469-1472

Gilboa R, Zharkov DO, Golan G, Fernandes AS, Gerchman SE, Matz E, Kycia JH, Grollman AP, Shoham G (2002) Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *The Journal of biological chemistry* **277:** 19811-19816

Goldenberg SJ, Cascio TC, Shumway SD, Garbutt KC, Liu J, Xiong Y, Zheng N (2004) Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. *Cell* **119:** 517-528

Graham TA, Ferkey DM, Mao F, Kimelman D, Xu W (2001) Tcf4 can specifically recognize beta-catenin using alternative conformations. *Nature Structural Biology* **8:** 1048-1052

Grove TZ, Cortajarena AL, Regan L (2008) Ligand binding by repeat proteins: natural and designed. *Current opinion in structural biology* **18:** 507-515

Groves MR, Barford D (1999) Topological characteristics of helical repeat proteins. *Current opinion in structural biology* **9:** 383-389

Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D (1999) The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell* **96:** 99-110

Gruskin EA, Lloyd RS (1986) The DNA scanning mechanism of T4 endonuclease V. Effect of NaCl concentration on processive nicking activity. *The Journal of biological chemistry* **261:** 9607-9613

Guan Y, Manuel RC, Arvai AS, Parikh SS, Mol CD, Miller JH, Lloyd S, Tainer JA (1998) MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nature Structural Biology* **5:** 1058-1064

Guengerich FP (1994) Mechanisms of formation of DNA adducts from ethylene dihalides, vinyl halides, and arylamines. *Drug Metabolism Reviews* **26:** 47-66

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18:** 2714-2723

Hausrath AC, Goriely A (2006) Repeat protein architectures predicted by a continuum representation of fold space. *Protein Science* **15:** 753-760

Hecht SS (1999) DNA adduct formation from tobacco-specific N-nitrosamines. *Mutation research* **424:** 127-142

Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* **36:** 39-56

Higley M, Lloyd RS (1993) Processivity of uracil DNA glycosylase. *Mutation research* **294:** 109-116

Hiller DA, Fogg JM, Martin AM, Beechem JM, Reich NO, Perona JJ (2003) Simultaneous DNA binding and bending by EcoRV endonuclease observed by real-time fluorescence. *Biochemistry* **42:** 14375-14385

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. *The EMBO journal* **19:** 758-766

Hollis T, Lau A, Ellenberger T (2001) Crystallizing thoughts about DNA base excision repair. *Progress in Nucleic Acid Research & Molecular Biology* **68:** 305-314

Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *Journal of molecular biology* **233:** 123-138

Holt S, Yen TY, Sangaiah R, Swenberg JA (1998) Detection of 1,N6-ethenoadenine in rat urine after chloroethylene oxide exposure. *Carcinogenesis* **19:** 1763-1769

Huber AH, Nelson WJ, Weis WI (1997) Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **90:** 871-882

Huffman JL, Sundheim O, Tainer JA (2005) DNA base damage recognition and removal: new twists and grooves. *Mutation research* **577:** 55-76

Hunter WN, Brown T, Kneale G, Anand NN, Rabinovich D, Kennard O (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-A resolution. *The Journal of biological chemistry* **262:** 9962-9970

Hurley LH (2002) DNA and its associated processes as targets for cancer therapy. *Nature reviews* **2:** 188-200

Irani RJ, SantaLucia J, Jr. (2002) The synthesis of anti-fixed 3-methyl-3-deaza-2'-deoxyadenosine and other 3H-imidazo[4,5-c]pyridine analogs. *Nucleosides Nucleotides Nucleic Acids* **21:** 737-751

Jacobs MD, Harrison SC (1998) Structure of an IkappaBalpha/NF-kappaB complex. *Cell* **95:** 749-758

Jaruga P, Speina E, Gackowski D, Tudek B, Olinski R (2000) Endogenous oxidative DNA base modifications analysed with repair enzymes and GC/MS technique. *Nucleic acids research* **28:** E16

Jiang YL, Ichikawa Y, Song F, Stivers JT (2003) Powering DNA repair through substrate electrostatic interactions. *Biochemistry* **42:** 1922-1929

Jiang YL, Kwon K, Stivers JT (2001) Turning On uracil-DNA glycosylase using a pyrene nucleotide switch. *The Journal of biological chemistry* **276:** 42347-42354

Jiang YL, Stivers JT, Song F (2002) Base-flipping mutations of uracil DNA glycosylase: substrate rescue using a pyrene nucleotide wedge. *Biochemistry* **41:** 11248-11254

Jimenez-Menendez N, Fernandez-Millan P, Rubio-Cosials A, Arnan C, Montoya J, Jacobs HT, Bernado P, Coll M, Uson I, Sola M (2010) Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat. *Nature structural & molecular biology* **17:** 891-893

Jiricny J (1998) Replication errors: cha(lle)nging the genome. *The EMBO journal* **17:** 6427-6436

Jones BN, Quang-Dang DU, Oku Y, Gross JD (2008) A kinetic assay to monitor RNA decapping under single- turnover conditions. *Methods in enzymology* **448:** 23-40

Joseph N, Duppatla V, Rao DN (2006) Prokaryotic DNA mismatch repair. *Progress in Nucleic Acid Research & Molecular Biology* **81:** 1-49

Karran P, Hjelmgren T, Lindahl T (1982) Induction of a DNA glycosylase for N-methylated purines is part of the adaptive response to alkylating agents. *Nature* **296:** 770-773

Karran P, Lindahl T, Ofsteng I, Evensen GB, Seeberg E (1980) Escherichia coli mutants deficient in 3-methyladenine-DNA glycosylase. *Journal of molecular biology* **140:** 101-127

Kasai H, Nishimura S (1984a) Hydroxylation of deoxy guanosine at the C-8 position by polyphenols and aminophenols in the presence of hydrogen peroxide and ferric ion. *Gann* **75:** 565-566

Kasai H, Nishimura S (1984b) Hydroxylation of deoxyguanosine at the C-8 position by ascorbic acid and other reducing agents. *Nucleic acids research* **12:** 2137-2145

Klungland A, Lindahl T (1997) Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). *The EMBO journal* **16:** 3341-3348

Kolodner RD, Marsischky GT (1999) Eukaryotic DNA mismatch repair. *Current Opinions in Genetics and Development* **9:** 89-96

Kruse B, Narasimhan N, Attardi G (1989) Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. *Cell* **58:** 391-397

Kubota Y, Nash RA, Klungland A, Schar P, Barnes DE, Lindahl T (1996) Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *The EMBO journal* **15:** 6662-6670

Kuo CF, McRee DE, Fisher CL, O'Handley SF, Cunningham RP, Tainer JA (1992) Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. *Science* **258:** 434-440

Kwon K, Cao C, Stivers JT (2003) A novel zinc snap motif conveys structural stability to 3-methyladenine DNA glycosylase I. *The Journal of biological chemistry* **278:** 19442-19446

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:** 321-329

Landini P, Hajec LI, Volkert MR (1994) Structure and transcriptional regulation of the Escherichia coli adaptive response gene aidB. *Journal of Bacteriology* **176:** 6583-6589

Larson K, Sahm J, Shenkar R, Strauss B (1985) Methylation-induced blocks to in vitro DNA replication. *Mutation research* **150:** 77-84

Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **26:** 283-291

Lau AY, Scharer 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:** 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. *Proceedings of the National Academy of Sciences USA* **97:** 13573-13578

Lavery R, Sklenar H (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *Journal of Biomolecular Structure & Dynamics* **6:** 63-91

Lawley PD (1966) Effects of some chemical mutagens and carcinogens on nucleic acids. *Progress in Nucleic Acid Research & Molecular Biology* **5:** 89-131

Lawley PD, Phillips DH (1996) DNA adducts from chemotherapeutic agents. *Mutation research* **355:** 13-40

Lee S, Bowman BR, Ueno Y, Wang S, Verdine GL (2008) Synthesis and structure of duplex DNA containing the genotoxic nucleobase lesion N7-methylguanine. *Journal of the American Chemical Society* **130:** 11570-11571

Li L, Perdigao J, Pegg AE, Lao Y, Hecht SS, Lindgren BR, Reardon JT, Sancar A, Wattenberg EV, Peterson LA (2009) The influence of repair pathways on the cytotoxicity and mutagenicity induced by the pyridyloxobutylation pathway of tobacco-specific nitrosamines. *Chemical Research in Toxicology* **22:** 1464-1472

Lindahl T (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proceedings of the National Academy of Sciences USA* **71:** 3649-3653

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

Lindahl T, Demple B, Robins P (1982) Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase. *The EMBO journal* **1:** 1359-1363

Lindahl T, Nyberg B (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* **13:** 3405-3410

Lindahl T, Sedgwick B, Sekiguchi M, Nakabeppu Y (1988) Regulation and expression of the adaptive response to alkylating agents. *Annual Reviews in Biochemistry* **57:** 133-157

Liu L, Nakatsuru Y, Gerson SL (2002) Base excision repair as a therapeutic target in colon cancer. *Clinical Cancer Research* **8:** 2985-2991

Lyons DM, O'Brien PJ (2009) Efficient recognition of an unpaired lesion by a DNA repair glycosylase. *Journal of the American Chemical Society* **131:** 17742-17743

Macdonald PM (1992) The Drosophila pumilio gene: an unusually long transcription unit and an unusual protein. *Development (Cambridge, England)* **114:** 221-232

Magnusdottir A, Stenmark P, Flodin S, Nyman T, Kotenyova T, Nilsson P, Ogg D, Nordlund P Crystal Structure of the Human Pp2A Regulatory Subunit, B56G. *PDB ID: 2JAK-A*

Maher RL, Bloom LB (2007) Pre-steady-state kinetic characterization of the AP endonuclease activity of human AP endonuclease 1. *The Journal of biological chemistry* **282:** 30577-30585

Maher RL, Vallur AC, Feller JA, Bloom LB (2007) Slow base excision by human alkyladenine DNA glycosylase limits the rate of formation of AP sites and AP endonuclease 1 does not stimulate base excision. *DNA repair* **6:** 71-81

Main ER, Jackson SE, Regan L (2003) The folding and design of repeat proteins: reaching a consensus. *Current opinion in structural biology* **13:** 482-489

Maiti A, Morgan MT, Drohat AC (2009) Role of two strictly conserved residues in nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase. *The Journal of biological chemistry* **284:** 36680-36688

Makino K, Ichikawa Y (1998) Synthesis of a 2-deoxy-ribose type 1-N-iminosugar. *Tetrahedron Letters* **39:** 8245-8248

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

Masutani C, Sugasawa K, Yanagisawa J, Sonoyama T, Ui M, Enomoto T, Takio K, Tanaka K, van der Spek PJ, Bootsma D, et al. (1994) Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *The EMBO journal* **13:** 1831-1843

Matsumoto Y, Kim K, Bogenhagen DF (1994) Proliferating cell nuclear antigen-dependent abasic site repair in Xenopus laevis oocytes: an alternative pathway of base excision DNA repair. *Molecular Cell Biology* **14:** 6187-6197

Matsuura Y, Stewart M (2004) Structural basis for the assembly of a nuclear export complex. *Nature* **432:** 872-877

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

McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallography D Biological Crystallography* **61:** 458-464

McRee DE (1999) XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *Journal of structural biology* **125:** 156-165

Medeiros MH (2009) Exocyclic DNA adducts as biomarkers of lipid oxidation and predictors of disease. Challenges in developing sensitive and specific methods for clinical studies. *Chemical Research in Toxicology* **22:** 419-425

Memisoglu A, Samson L (1996) Cloning and characterization of a cDNA encoding a 3-methyladenine DNA glycosylase from the fission yeast *Schizosaccharomyces pombe*. *Gene* **177:** 229-235

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

Min JH, Pavletich NP (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449:** 570-575

Modrich P (1997) Strand-specific mismatch repair in mammalian cells. *The Journal of biological chemistry* **272:** 24727-24730

Mol CD, Arvai AS, Begley TJ, Cunningham RP, Tainer JA (2002) Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases. *Journal of molecular biology* **315:** 373-384

Mol CD, Arvai AS, Slupphaug G, Kavli B, Alseth I, Krokan HE, Tainer JA (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80:** 869-878

Mol CD, Izumi T, Mitra S, Tainer JA (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* **403:** 451-456

Moore MH, Gulbis JM, Dodson EJ, Demple B, Moody PC (1994) Crystal structure of a suicidal DNA repair protein: the Ada O6-methylguanine-DNA methyltransferase from E. coli. *The EMBO journal* **13:** 1495-1501

Morikawa K, Matsumoto O, Tsujimoto M, Katayanagi K, Ariyoshi M, Doi T, Ikehara M, Inaoka T, Ohtsuka E (1992) X-ray structure of T4 endonuclease V: an excision repair enzyme specific for a pyrimidine dimer. *Science* **256:** 523-526

Mu D, Hsu DS, Sancar A (1996) Reaction mechanism of human DNA repair excision nuclease. *The Journal of biological chemistry* **271:** 8285-8294

Murshudov GN, A.A.Vagin AA, Dodson EJ (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica* **D53:** 240-255

Nash HM, Bruner SD, Scharer OD, Kawate T, Addona TA, Spooner E, Lane WS, Verdine GL (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Current Biology* **6:** 968-980

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

O'Brien PJ, Ellenberger T (2003) Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines. *Biochemistry* **42:** 12418-12429

O'Brien PJ, Ellenberger T (2004a) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *The Journal of biological chemistry* **279:** 9750-9757

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

O'Connor TR (1993) Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic acids research* **21:** 5561-5569

O'Connor TR, Boiteux S, Laval J (1988) Ring-opened 7-methylguanine residues in DNA are a block to in vitro DNA synthesis. *Nucleic acids research* **16:** 5879-5894

O'Rourke EJ, Chevalier C, Boiteux S, Labigne A, Ielpi L, Radicella JP (2000) A novel 3-methyladenine DNA glycosylase from helicobacter pylori defines a new class within the endonuclease III family of base excision repair glycosylases. *The Journal of biological chemistry* **275:** 20077-20083

Osipiuk J, Hatzos C, Moy S, Collart F, Joachimiak A X-ray structure of predicted DNA alkylation repair enzyme from *Enterococcus faecalis*. *PDB ID: 2B6C-A*

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

Paik J, Duncan T, Lindahl T, Sedgwick B (2005) Sensitization of human carcinoma cells to alkylating agents by small interfering RNA suppression of 3-alkyladenine-DNA glycosylase. *Cancer Research* **65:** 10472-10477

Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE, Tainer JA (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *The EMBO journal* **17:** 5214-5226

Parikh SS, Walcher G, Jones GD, Slupphaug G, Krokan HE, Blackburn GM, Tainer JA (2000) Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proceedings of the National Academy of Sciences USA* **97:** 5083-5088

Park CB, Asin-Cayuela J, Camara Y, Shi Y, Pellegrini M, Gaspari M, Wibom R, Hultenby K, Erdjument-Bromage H, Tempst P, Falkenberg M, Gustafsson CM, Larsson NG (2007) MTERF3 is a negative regulator of mammalian mtDNA transcription. *Cell* **130:** 273-285

Park JH, Aravind L, Wolff EC, Kaevel J, Kim YS, Park MH (2006) Molecular cloning, expression, and structural prediction of deoxyhypusine hydroxylase: a HEAT-repeat-containing metalloenzyme. *Proceedings of the National Academy of Sciences USA* **103:** 51-56

Park MH, Joe YA, Kang KR (1998) Deoxyhypusine synthase activity is essential for cell viability in the yeast *Saccharomyces cerevisiae*. *The Journal of biological chemistry* **273:** 1677-1683

Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **449:** 433-437

Parker JB, Stivers JT (2008) Uracil DNA glycosylase: revisiting substrate-assisted catalysis by DNA phosphate anions. *Biochemistry* **47:** 8614-8622

Parmeggiani F, Pellarin R, Larsen AP, Varadamsetty G, Stumpp MT, Zerbe O, Caflisch A, Pluckthun A (2008) Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *Journal of molecular biology* **376:** 1282-1304

Pavlov YI, Mian IM, Kunkel TA (2003) Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. *Current Biology* **13:** 744-748

PDB (2006) Crystal structure of DNA-3-methyladenine glycosidase (10174367) from Bacillus halodurans at 2.55 Å resolution. *Joint Center for Structural Genomics, PDB ID 2H56*

Pearl LH (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutation research* **460:** 165-181

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

Peters JW, Stowell MH, Rees DC (1996) A leucine-rich repeat variant with a novel repetitive protein structural motif. *Nature Structural Biology* **3:** 991-994

Plosky BS, Frank EG, Berry DA, Vennall GP, McDonald JP, Woodgate R (2008) Eukaryotic Y-family polymerases bypass a 3-methyl-2'-deoxyadenosine analog in vitro and methyl methanesulfonate-induced DNA damage in vivo. *Nucleic acids research* **36:** 2152-2162

Podmore ID, Cooper D, Evans MD, Wood M, Lunec J (2000) Simultaneous measurement of 8-oxo-2'-deoxyguanosine and 8-oxo-2'-deoxyadenosine by HPLC-MS/MS. *Biochemical and Biophysical Research Communications* **277:** 764-770

Porecha RH, Stivers JT (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proceedings of the National Academy of Sciences USA* **105:** 10791-10796

Potter PM, Wilkinson MC, Fitton J, Carr FJ, Brennand J, Cooper DP, Margison GP (1987) Characterisation and nucleotide sequence of ogt, the O6-alkylguanine-DNA-alkyltransferase gene of E. coli. *Nucleic acids research* **15:** 9177-9193

Reardon JT, Sancar A (2005) Nucleotide excision repair. *Progress in Nucleic Acid Research & Molecular Biology* **79:** 183-235

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

Riedl T, Hanaoka F, Egly JM (2003) The comings and goings of nucleotide excision repair factors on damaged DNA. *The EMBO journal* **22:** 5293-5303

Ringvoll J, Moen MN, Nordstrand LM, Meira LB, Pang B, Bekkelund A, Dedon PC, Bjelland S, Samson LD, Falnes PO, Klungland A (2008) AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer Research* **68:** 4142-4149

Ringvoll J, Nordstrand LM, Vagbo CB, Talstad V, Reite K, Aas PA, Lauritzen KH, Liabakk NB, Bjork A, Doughty RW, Falnes PO, Krokan HE, Klungland A (2006) Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *The EMBO journal* **25:** 2189-2198

Roberti M, Polosa PL, Bruni F, Manzari C, Deceglie S, Gadaleta MN, Cantatore P (2009) The MTERF family proteins: mitochondrial transcription regulators and beyond. *Biochimica et Biophysica Acta* **1787:** 303-311

Roberts DW, Churchwell MI, Beland FA, Fang JL, Doerge DR (2001) Quantitative analysis of etheno-2'-deoxycytidine DNA adducts using on-line immunoaffinity chromatography coupled with LC/ES-MS/MS detection. *Analytical chemistry* **73:** 303-309

Rocchia W, Sridharan S, Nicholls A, Alexov E, Chiabrera A, Honig B (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *Journal of computational chemistry* **23:** 128-137

Rohankhedkar MS, Mulrooney SB, Wedemeyer WJ, Hausinger RP (2006) The AidB component of the Escherichia coli adaptive response to alkylating agents is a flavin-containing, DNA-binding protein. *Journal of Bacteriology* **188:** 223-230

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. *Journal of molecular biology* **381:** 13-23

Rydberg B, Lindahl T (1982) Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *The EMBO journal* **1:** 211-216

Samson L, Cairns J (1977) A new pathway for DNA repair in Escherichia coli. *Nature* **267:** 281-283

Sancar A (2003) Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chemical reviews* **103:** 2203-2237

Sancar A (2008) Structure and function of photolyase and in vivo enzymology: 50th anniversary. *The Journal of biological chemistry* **283:** 32153-32157

Sancar A, Reardon JT (2004) Nucleotide excision repair in E. coli and man. *Advances in protein chemistry* **69:** 43-71

Saparbaev M, Kleibl K, Laval J (1995) *Escherichia coli*, *Saccharomyces cerevisiae*, rat and human 3-methyladenine DNA glycosylases repair 1,N^6-ethenoadenine when present in DNA. *Nucleic acids research* **23:** 3750-3755

Saparbaev M, Laval J (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. *Proceedings of the National Academy of Sciences USA* **91:** 5873-5877

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

Scharer OD, Nash HM, Jiricny J, Laval J, Verdine GL (1998) Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases. *The Journal of biological chemistry* **273:** 8592-8597

Scharer OD, Ortholand JY, Ganesan A, Ezaznikpay K, Verdine GL (1995) Specific Binding of the DNA-Repair Enzyme Alka to a Pyrrolidine-Based Inhibitor. *Journal of the American Chemical Society* **117:** 6623-6624

Sedgwick B (2004) Repairing DNA-methylation damage. *Nature Reviews Molecular and Cellular Biology* **5:** 148-157

Seeberg E, Eide L, Bjoras M (1995) The base excision repair pathway. *Trends in biochemical sciences* **20:** 391-397

Setlow RB (1966) Cyclobutane-type pyrimidine dimers in polynucleotides. *Science* **153:** 379-386

Shibutani S, Takeshita M, Grollman AP (1991) Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* **349:** 431-434

Shuker DE, Bailey E, Parry A, Lamb J, Farmer PB (1987) The determination of urinary 3-methyladenine in humans as a potential monitor of exposure to methylating agents. *Carcinogenesis* **8:** 959-962

Shuker DE, Farmer PB (1992) Relevance of urinary DNA adducts as markers of carcinogen exposure. *Chemical Research in Toxicology* **5:** 450-460

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

Sidorkina OM, Laval J (2000) Role of the N-terminal proline residue in the catalytic activities of the Escherichia coli Fpg protein. *The Journal of biological chemistry* **275:** 9924-9929

Sijbers AM, de Laat WL, Ariza RR, Biggerstaff M, Wei YF, Moggs JG, Carter KC, Shell BK, Evans E, de Jong MC, Rademakers S, de Rooij J, Jaspers NG, Hoeijmakers JH, Wood RD (1996) Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **86:** 811-822

Singer B, Antoccia A, Basu AK, Dosanjh MK, Fraenkel-Conrat H, Gallagher PE, Kusmierek JT, Qiu ZH, Rydberg B (1992) Both purified human 1,N6-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N6-

ethenoadenine and 3-methyladenine. *Proceedings of the National Academy of Sciences USA* **89:** 9386-9390

Singer B, Grunberger D (1983) *Molecular Biology of Mutagens and Carcinogens: Intrinsic Properties of Nucleic Acids*, New York: Plenum Press.

Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384:** 87-92

Spahr H, Samuelsson T, Hallberg BM, Gustafsson CM (2010) Structure of mitochondrial transcription termination factor 3 reveals a novel nucleic acid-binding domain. *Biochemical and Biophysical Research Communications* **397:** 386-390

Stein AJ, Fuchs G, Fu C, Wolin SL, Reinisch KM (2005) Structural insights into RNA quality control: the Ro autoantigen binds misfolded RNAs via its central cavity. *Cell* **121:** 529-539

Stivers JT (2004) Site-specific DNA damage recognition by enzyme-induced base flipping. *Progress in Nucleic Acid Research & Molecular Biology* **77:** 37-65

Stivers JT (2008) Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chemistry* **14:** 786-793

Stivers JT, Jiang YL (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chemical reviews* **103:** 2729-2759

Stivers JT, Pankiewicz KW, Watanabe KA (1999) Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase. *Biochemistry* **38:** 952-963

Stumpp MT, Forrer P, Binz HK, Pluckthun A (2003) Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *Journal of molecular biology* **332:** 471-487

Sugahara M, Mikawa T, Kumasaka T, Yamamoto M, Kato R, Fukuyama K, Inoue Y, Kuramitsu S (2000) Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8. *The EMBO journal* **19:** 3857-3869

Sun B, Latham KA, Dodson ML, Lloyd RS (1995) Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. *The Journal of biological chemistry* **270:** 19501-19508

Tamulaitis G, Zaremba M, Szczepanowski RH, Bochtler M, Siksnys V (2007) Nucleotide flipping by restriction enzymes analyzed by 2-aminopurine steady-state fluorescence. *Nucleic acids research* **35:** 4792-4799

Taverna P, Sedgwick B (1996) Generation of an endogenous DNA-methylating agent by nitrosation in Escherichia coli. *J Bacteriol* **178:** 5105-5111

Tchou J, Kasai H, Shibutani S, Chung MH, Laval J, Grollman AP, Nishimura S (1991) 8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proceedings of the National Academy of Sciences USA* **88:** 4690-4694

Thayer MM, Ahern H, Xing D, Cunningham RP, Tainer JA (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *The EMBO journal* **14:** 4108-4120

Thomas L, Yang CH, Goldthwait DA (1982) Two DNA glycosylases in Escherichia coli which release primarily 3-methyladenine. *Biochemistry* **21:** 1162-1169

Thompson CL, Sancar A (2002) Photolyase/cryptochrome blue-light photoreceptors use photon energy to repair DNA and reset the circadian clock. *Oncogene* **21:** 9043-9056

Tubbs JL, Latypov V, Kanugula S, Butt A, Melikishvili M, Kraehenbuehl R, Fleck O, Marriott A, Watson AJ, Verbeek B, McGown G, Thorncroft M, Santibanez-Koref MF, Millington C, Arvai AS, Kroeger MD, Peterson LA, Williams DM, Fried MG, Margison GP, Pegg AE, Tainer JA (2009) Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* **459:** 808-813

Varghese AJ, Wang SY (1968) Thymine-thymine adduct as a photoproduct of thymine. *Science* **160:** 186-187

Vassylyev DG, Kashiwagi T, Mikami Y, Ariyoshi M, Iwai S, Ohtsuka E, Morikawa K (1995) Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* **83:** 773-782

Vaughan P, Lindahl T, Sedgwick B (1993) Induction of the adaptive response of Escherichia coli to alkylation damage by the environmental mutagen, methyl chloride. *Mutation research* **293:** 249-257

Verdine GL, Bruner SD (1997) How do DNA repair proteins locate damaged bases in the genome? *Chemistry & Biology* **4:** 329-334

Vetter IR, Arndt A, Kutay U, Gorlich D, Wittinghofer A (1999) Structural view of the Ran-Importin beta interaction at 2.3 A resolution. *Cell* **97:** 635-646

Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods in molecular biology (Clifton, NJ*) **364:** 215-230

Wang X, McLachlan J, Zamore PD, Hall TM (2002) Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110:** 501-512

Wang X, Zamore PD, Hall TM (2001) Crystal structure of a Pumilio homology domain. *Molecular cell* **7:** 855-865

Watson JD, Crick FH (1953a) Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171:** 964-967

Watson JD, Crick FH (1953b) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171:** 737-738

Wetzel SK, Ewald C, Settanni G, Jurt S, Pluckthun A, Zerbe O (2010) Residue-resolved stability of full-consensus ankyrin repeat proteins probed by NMR. *Journal of molecular biology* **402:** 241-258

Williams DR, Lee KJ, Shi J, Chen DJ, Stewart PL (2008) Cryo-EM structure of the DNA-dependent protein kinase catalytic subunit at subnanometer resolution reveals alpha helices and insight into DNA binding. *Structure* **16:** 468-477

Xu Y, Xing Y, Chen Y, Chao Y, Lin Z, Fan E, Yu JW, Strack S, Jeffrey PD, Shi Y (2006) Structure of the protein phosphatase 2A holoenzyme. *Cell* **127:** 1239-1251

Yakubovskaya E, Mejia E, Byrnes J, Hambardjieva E, Garcia-Diaz M (2010) Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription. *Cell* **141:** 982-993

Yamagata Y, Kato M, Odawara K, Tokuno Y, Nakashima Y, Matsushima N, Yasumura K, Tomita K, Ihara K, Fujii Y, Nakabeppu Y, Sekiguchi M, Fujii S (1996) Three-dimensional structure of a DNA repair enzyme, 3-methyladenine DNA glycosylase II, from Escherichia coli. *Cell* **86:** 311-319

Yamane T, Wyluda BJ, Shulman RG (1967) Dihydrothymine from UV-irradiated DNA. *Proceedings of the National Academy of Sciences USA* **58:** 439-442

Yang CG, Garcia K, He C (2009) Damage detection and base flipping in direct DNA alkylation repair. *Chembiochem* **10:** 417-423

Yang CG, Yi C, Duguid EM, Sullivan CT, Jian X, Rice PA, He C (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452:** 961-965

Yang W (2006) Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA repair* **5:** 654-666

Yang W (2008) Structure and mechanism for DNA lesion recognition. *Cell Research* **18:** 184-197

Yang Y, Nikolic D, Swanson SM, van Breemen RB (2002) Quantitative determination of N7-methyldeoxyguanosine and O6-methyldeoxyguanosine in DNA by LC-UV-MS-MS. *Analytical chemistry* **74:** 5376-5382

Yu B, Hunt JF (2009) Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proceedings of the National Academy of Sciences USA* **106:** 14315-14320

Zhang F, Bartels MJ, Pottenger LH, Gollapudi BB, Schisler MR (2006) Simultaneous quantitation of 7-methyl- and O6-methylguanine adducts in DNA by liquid chromatography-positive electrospray tandem mass spectrometry. *Journal of Chromatography B Analytical Technologies in Biomedical and Life Sciences* **833:** 141-148