STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF YEAST ALKYLPURINE DNA GLYCOSYLASES

By

Suraj Adhikary

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Sciences

May, 2013

Nashville, Tennessee

Approved:

Brandt Eichman, Ph.D.

Katherine Friedman, Ph.D.

Charles Singleton, Ph.D.

Neil Osheroff, Ph.D.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to acknowledge the help and support of following people in what at times seemed like an impossible pursuit:

Aama, Dada and Mommy

Christin, Pat and Tim

Brandt, Claire, Sonja, Bri and others in the Eichman lab

Kathy, Bubba and Neil

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1mA	*N*1-Methyladenine
3d3mA	3-Deaza-3-methyladenine
3mA	*N*3-Methyladenine
3mG	*N*3-Methylguanine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
5OHC	5-hydroxycytosine
5OHU	5-hydroxyuracil
7mA	*N*7-Methyladenine
7mG	*N*7-Methylguanine
8OHG	7,8-Dihydro-8-hydroxyguanine
8oxoG	7,8-Dihydro-8-oxoguanine
Aa	Amino acid
Å	Angstrom
A (Ade)	Adenine
AAG	Human alkyladenine DNA glycosylase
AGOG	Archael GO glycosylase
AlkA	*E. coli* 3-methyladenine DNA glycosylase II
AP	Apurinic/apyrimidinic

AP	Aminopurine
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
DHT	Dihydrothymine
DHU	Dihydrocytosine
DNA	Deoxyribonucleic Acid
dRPase	Deoxyribosephodphodiesterase
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminotetraacetic acid
EM	Electron microscopy
EMS	Ethyl methanesulfonate
EMSA	Electrophoretic mobility shift assay
Endo-	Endonuclease
εA	1-N^6-ethenoadenine
εC	1-N^4-ethenocytosine
FaPy	Formamidopyrimidine

Fpg	Formamidopyrimidine DNA glycosylase
g	Gram
G (Gua)	Guanine
Gh	Guanidinohydantoin
H2TH	Helix-two turn-helix
HCl	Hydrochloric acid
HEAT	Huntington/Elongation/A subunit/Target of rapamycin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HhH	Helix-hairpin-helix
HPLC	High performance liquid chromatography
hr	Hour
Hx	Hypoxanthine
I	Intensity
Ia	Iminoallantion
IC	Interrogation Complex
IPTG	Isopropyl thio-β-D-galactopyranoside
K_d	Dissociation constant
kDa	Kilodalton
k_{st}	Single-turnover rate
L	Liter
LB	Luria broth

LRC	Lesion Recognition Complex
MAG, scMag	*S. cerevisiae* methyladenine DNA glycosylase
MagIII	*H. pylori* methyladenine DNA glycosylase
Mag1, spMag1	*S. pombe* methyladenine DNA glycosylase 1
Mag2	*S. pombe* methyladenine DNA glycosylase 2
MALDI	Matrix assisted laser desorption ionization
MES	2-(*N*-morpholino)ethanesulfonic acid
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	*T.maritima* methylpurine DNA glycosylase II
MS	Mass spectrometry
MutY	A/G specific adenine glycosylase
MUTYH	human 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
OGG1, OggI	8-oxoG DNA glycosylase I
hOGG1, hOggI	human 8-oxoG DNA glycosylase I
OGG2, OggII	8-oxoG DNA glycosylase II
PCR	Polymerase chain reaction
PCNA	Proliferating cellular nuclear antigen
PD	Pyrimidine dimer
PDB	Protein Data Bank
PEG	Polyethylene glycol
Pyr	1-pyrrolidine
RFC	Replication factor C
rmsd	Root-mean-square deviation
RNA	Ribonucleic Acid
RPA	Replication protein A
rpm	Rotations per minute
s.d.	Standard deviation

SAD	Single-wavelength anomalous dispersion
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Second
Semet	Selenomethionine
SMUG	Single-strand selective monofunctional uracil DNA glycosylase
Sp	Spiroiminodihydantoin
ssDNA	Single-stranded DNA
T (Thy)	Thymine
TAG	*E. coli* 3-methyladenine DNA glycosylase I
Tg	Thymine glycol
TDG	Thymine DNA glycosylase
THF	Tetrahydrofuran
Tris	Tris-(hydroxymethyl-)aminomethane
U	Uracil
UDC	Undamaged DNA complex
UDG, UNG	Uracil DNA glycosylase
μg	Microgram
μM	Micromolar
UV	Ultraviolet
wt	Wild-type
XP	Xeroderma pigmentosum

Chapter 1

INTRODUCTION*

1.1. DNA Damage and Repair

The central dogma of molecular biology states that DNA acts as a template for RNA, which in turn codes for protein (Crick 1958). Hence as long as the chemical integrity of DNA is maintained, it is possible for damaged or improperly synthesized RNA and proteins to be simply discarded and begin anew. However, because DNA acts as the blueprint for cellular processes, damaged DNA cannot just be discarded but needs to be repaired to maintain genomic integrity. Despite being the most indispensable macromolecule, DNA faces constant assault from both endogenous and environmental agents. These include but are not limited to ultraviolet radiation, oxidizing and alkylating agents as well as various metabolites and even some drugs used in anti-cancer therapy (Friedberg, Walker et al. 2006). These agents can induce spontaneous covalent modification of nucleobases that can be either mutagenic or cytotoxic or both (Table 1, Fig. 1.1) (Friedberg, Aguilera et al. 2006; Friedberg, Walker et al. 2006).

---

*Parts of this chapter have been previously published in the following articles:

Rubinson, E. H., **S. Adhikary**, et al. (2010). Structural Studies of Alkylpurine DNA Glycosylases. ACS Symposium Series : Structural Biology of DNA Damage and Repair. M. P. Stone. Washingon, D.C., American Chemical Society. **1041:** 29-45.

Brooks, S. C., **S. Adhikary**, et al. (2012). "Recent advances in the structural mechanisms of DNA glycosylases." Biochim Biophys Acta **1834**(1): 247-271.

Given the consequences of unrepaired DNA lesions, all organisms possess a wide array of repair mechanisms. I will use this section to introduce and discuss major sources of DNA damage and pathways and enzymes that repair the resulting modified nucleobases.

Table 1. Sources and Instances of endogenous DNA Damage

Source (Damage)	No. lesions (per-cell per-day)
Alkylation	
7-Methylguanine	6,000
3-Methyladenine	1,200
1-Methyladenine	200
O^6-Methylguanine	20-100
Oxidation	
8-oxoguanine	~1,000-2,000
Thymine glycol, cytosine hydrates	~2,000
Lipid peroxidation (εA, εC, etc.)	~1,000
Hydrolysis	
Depurination	18,000
Depyrimidation	600
Cytosine deamination	100
5-Methylcytosine deamination	10

[Adapted from (Friedberg, Walker et al. 2006)]

1.2. Response to DNA Damage

When faced with DNA damage, cells have evolved various mechanisms by which they either attempt to reverse the damage via direct reversal and excision (base, nucleotide ormismatch excision repair), bypass it (translesion synthesis, postreplicative gap filling, and replication fork progression), signal various cell cycle checkpoint pathways and/or undergo apoptosis. It is becoming increasingly apparent that response to DNA damage triggers interactions between multiple pathways and a balance between

repair and tolerance mechanisms is used to ensure genomic integrity (Fu, Calvo et al. 2012).

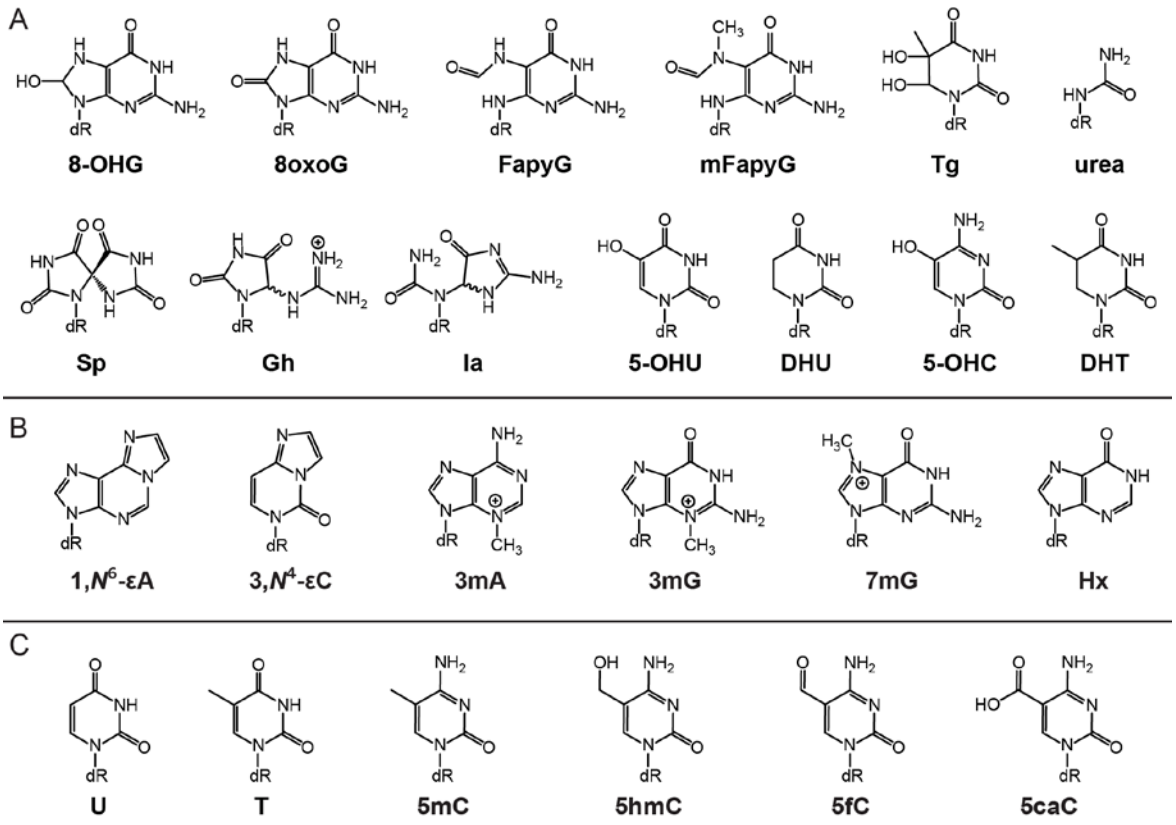


Fig. 1.1 Common DNA lesions. (A) Oxidized nucleobases. 8-OHG, 7,8-dihydro-8-hydroxyguanine; 8oxoG, 7,8-dihydro-8-oxoguanine; FapyG, 2,6-diamino-4-hydroxy-5-formamidopyrimidine; mFapyG, *N*7-methylFapyG; Tg, thymine glycol; Sp, spiroiminodihydantoin; Gh, guanidinohydantoin; Ia, iminoallantion; 5-OHU, 5-hydroxyuracil; DHU, dihydrouracil; 5-OHC, 5-hydroxycytosine; DHT, dihydrothymine. (B) Alkylated nucleobases. εA, 1,N^6-ethenoadenine; εC, 3,N^4-ethenocytosine; 3mA, *N*3-methyladenine; 3mG, *N*3-methylguanine; 7mG, *N*7-methylguanine; Hx, hypoxanthine. (C) Nucleobases repaired by the UDG/TDG family of DNA glycosylases. U, uracil; T, thymine; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine.

1.2.1 Base Excision Repair

Single base lesions resulting from oxidation, deamination and alkylation of DNA are most commonly repaired using a series of highly coordinated enzymatic reactions known as base excision repair (BER) (Lindahl and Barnes 2000; Brooks, Adhikary et al. 2012) (Fig. 1.2). BER is initiated by a class of enzymes known as DNA glycosylases. These enzymes hydrolyze the N-glycosyl bond to excise the target base and leave behind an abasic site (Duncan, Hamilton et al. 1976). Abasic sites thus produced are processed by AP-endonucleases (see section 1.3). These enzymes catalyze hydrolysis of the phosphodiester bond immediately 5' to the abasic site.Some DNA glycosylases are bi-functional with additional AP-lyase activity, and can cleave the DNA backbone 3' to abasic site. The action of AP endonucleases generates a 5' deoxyribose phosphate, which is then removed by exonucleases known as deoxyribosephospodiesterase (dRPase, short-patch BER) or a flap endonuclease (long-patch BER). Together, these enzymes create either a single (short-patch) or four-nucleotide (long-patch) gap in the duplex. This gap is then filled in with undamaged nucleotide(s) by a DNA polymerase (DNA polymerase I in *E. coli* and DNA polymerase β in humans) and the repair process is completed by a DNA ligase [reviewed in (Kim and Wilson 2012)].

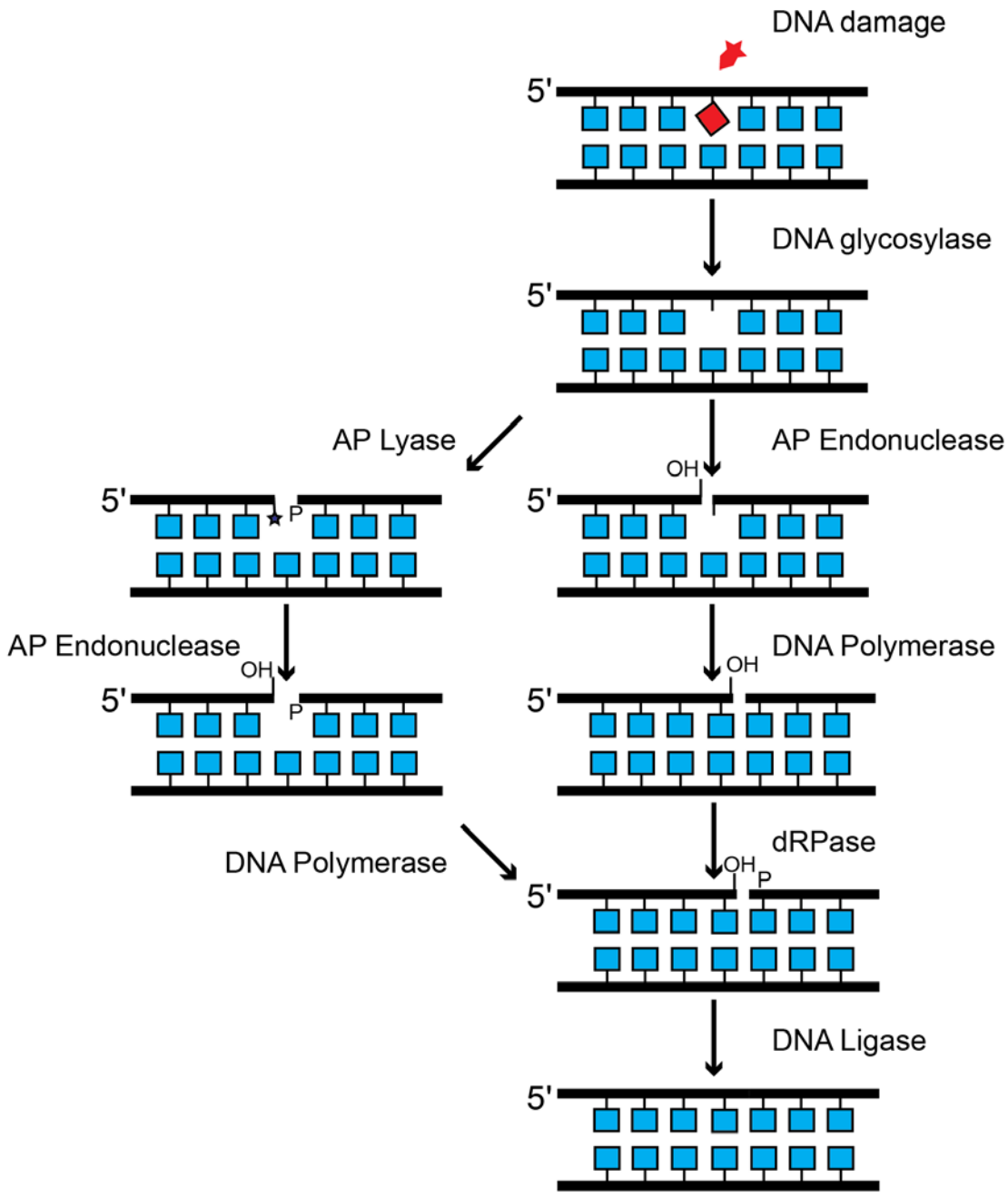


Fig. 1.2 Overview of the Base Excision Repair (BER) Pathway. The pathway is initiated by DNA glycosylases that locate, recognize and excise the damaged base and can either be monofunctional or bifunctional (Left, with additional AP lyase activity). The abasic site generated by DNA glycosylase is then acted on by AP endonuclease to create a nick in the DNA and a DNA polymerase with help from a deoxyribosephosphodiesterase (dRPase) and DNA ligase eventually fill the gap with undamaged base to complete the repair process. [Adapted from (Kim and Wilson 2012)].

1.2.2 Nucleotide Excision Repair

UV irradiation of DNA at 254 nm produces a nucleobase adduct called cyclobutane pyrimidine dimer (CPD) (Ceska, Sayers et al. 1996). CPD and other bulky adducts (6-4-PP, 6-4 photoproduct) are repaired by excising the oligonucleotide fragment from the duplex using a process known as nucleotide excision repair (NER) (Bronner, Welker et al. 1992). Unlike BER, here the damage is excised as free nucleotides rather than bases. The overall mechanism of NER can be divided into three steps and is conserved in both prokaryotes and eukaryotes. The damage is recognized by a group of proteins, verified by another, and finally the damage is excised and the gap is filled and sealed (Kuper and Kisker 2012). In *E. coli*, the first and second steps are accomplished by two proteins - UvrA and UvrB. UvrB assists another protein, UvrC, for the second step of damage removal (Truglio, Karakas et al. 2006). The repair process is completed by UvrD (DNA helicase), DNA polymerase I and DNA ligase. Xeroderma Pigmentosum (XP), a disease characterized in humans by severe sensitivity to UV radiation has been used to identify many of the components of NER (Cleaver 1968; Cleaver 1989). NER in eukaryotes involves up to 30 proteins but the role of each component is still not fully understood. Briefly, the first step is initiated by the XPC-RAD23B complex which recruits transcription factor TFIIH, a ten subunit complex containing two helicases, XPB and XPD. XPB and XPD act to identify and verify the damage and unwind the DNA at the site. The excision step is initiated by XPG and XPF-ERCC1, which combine to remove 25-30, nucleotides and the repair process is completed by the gap filling action of the replicative machinery (DNA polymerase δ or ε, RPA, proliferating cell nuclear

antigen (PCNA, a processivity clamp), replication factor C (RFC) and DNA ligase I). [Reviewed in (Reardon and Sancar 2005) and (Kuper and Kisker 2012)].

1.2.3 Mismatch Repair

DNA base mismatches arise from failure of the proofreading activity of replicative DNA polymerases. Proteins involved in the mismatch repair (MMR) pathway act to avoid mutations and overall genomic instability that can arise from base mismatches (Modrich 1997). Individuals deficient in MMR activity are prone to Lynch syndrome (Vasen, Watson et al. 1999). Lynch syndrome (non-polyposis colorectal cancer) patients are susceptible to cancers of the small intestines, liver, stomach, etc. (Vasen, Watson et al. 1999). Unlike BER and NER proteins, MMR has to be able to identify normal, unmodified bases based solely on deviation from canonical base pairings. In γ-proteobacteria such as *E. coli*, MMR is methylation assisted. [Reviewed in (Joseph, Duppatla et al. 2006) and (Guarne 2012)]. In prokaryotes, homodimeric MutS with the help of the β subunit of polymerase δ is charged with recognizing mismatches and small insertion/deletion (indel) loops. In an ATP-dependent process, MutS then recruits MutL to the complex. The MutS-MutL heterocomplex then recruits MutH. Since all the adenines within the GATC sequence are methylated, endonuclease activity of MutH enzyme nicks the DNA at hemimethylated GATC sites on the newly synthesized strand. This nick is then acted on by UvrD helicase II, which unwinds the DNA, and the target nucleotide stretch is excised by the appropriate exonuclease (ExoI or ExoX for nicks 3’ of the mismatch and RecJ or ExoVII for nicks 5’ of the mismatch) (Burdett, Baitinger et al. 2001). This gap is then filled by DNA polymerase III and DNA ligase completes the repair by sealing the nick. In other prokaryotes and eukaryotes,

methylation-independent MMR has been documented (Zhang, Yuan et al. 2005). MutS and MutL homodimers are replaced by a heterodimer of MutS and MutL homologs (MutSα or MutSβ and MutLα) in higher eukaryotes [reviewed in (Li 2003)]. In addition to MutS and MutL homologs, eukaryotic MMR also relies on Exo1, RFC, PCNA, DNA polymerase δ and DNA ligase (Dzantiev, Constantin et al. 2004; Guarne 2012).

1.3 Spontaneous Base Hydrolysis

One of the consequences of the reduction of ribose (in RNA) to deoxyribose (in DNA) is that it weakens the N-glycosyl bond thereby making it prone to hydrolysis. This leaves behind the deoxyribose ring commonly referred to as an AP (apurinic/apyrimidinic) or abasic site. It is estimated that close to 19,000 bases are lost each day per human cell to spontaneous base hydrolysis (Table 1) (Lindahl and Barnes 2000). In general, purine bases are much more susceptible to hydrolysis than pyrimidines (Lindahl and Nyberg 1972). Even between purines, guanine is released about 1.5 times faster than adenine at acidic and neutral pH, whereas adenine is hydrolyzed much faster at alkaline pH (Lindahl and Nyberg 1972). The AP site left behind exists in equilibrium between the closed ring (*furanose*) and the ring-opened *aldehyde* form (Jones, Khan et al. 1968). The *aldehyde* form of the AP site is susceptible to *β-elimination* hydrolysis of the 3'-phosphodiester bonds at certain physiological conditions and thereby could lead to single strand breaks (Jones, Khan et al. 1968).

1.4 DNA Glycosylases and BER

DNA nucleobases are susceptible to covalent modifications from oxidation, alkylation, and deamination reactions, which produce a divergent array of DNA lesions (Fig 1.1). These lesions can be either cytotoxic or mutagenic and can lead to a number of

diseases including cancer [reviewed in (Friedberg, Aguilera et al. 2006)]. These single-base lesions are repaired by the base excision repair (BER) pathway (Lindahl 2000), which is initiated by a DNA glycosylase specialized for a particular type of chemical damage. Since the fidelity of the entire pathway depends on them, these enzymes have been the subject of a large body of work to understand their mechanism of action (Scharer and Jiricny 2001; Stivers and Jiang 2003; Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; David, O'Shea et al. 2007; Dalhus, Laerdahl et al. 2009; Friedman and Stivers 2010; Li 2010; Zharkov, Mechetin et al. 2010; Rubinson and Eichman 2012) (Table 8). The first crystal structures of DNA glycosylases were reported in 1992 for bacteriophage T4 Endonuclease V (EndoV) and *Escherichia coli* (*E. coli*) Endonuclease III (EndoIII), which remove pyrimidine dimers and oxidized pyrimidines, respectively (Kuo, McRee et al. 1992; Morikawa, Matsumoto et al. 1992). Soon thereafter, DNA or inhibitor-bound structures of EndoV and uracil DNA glycosylase (UDG) established that these enzymes use a base-flipping mechanism to gain access to modified nucleobases in DNA (Mol, Arvai et al. 1995; Mol, Kuo et al. 1995; Savva, McAuley-Hecht et al. 1995; Vassylyev, Kashiwagi et al. 1995; Slupphaug, Mol et al. 1996). Subsequent studies established that glycosylases fall into one of six structural superfamilies (Fig. 1.3). Recognition of the target modification likely proceeds in several steps, in which the protein probes the stability of the base pairs through processive interrogation of the DNA duplex, followed by the extrusion of the aberrant nucleobase into a specific active site pocket on the enzyme (Stivers and Jiang 2003; Banerjee, Santos et al. 2006). The enzyme-substrate complex is stabilized by nucleobase contacts within the active site and a pair of side chains that plug the gap in the DNA left by the

extrahelical nucleotide and wedge into the DNA base stack on the opposite strand (Scharer and Jiricny 2001; Stivers and Jiang 2003; Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; David, O'Shea et al. 2007; Dalhus, Laerdahl et al. 2009; Friedman and Stivers 2010; Li 2010; Zharkov, Mechetin et al. 2010; Rubinson and Eichman 2012).

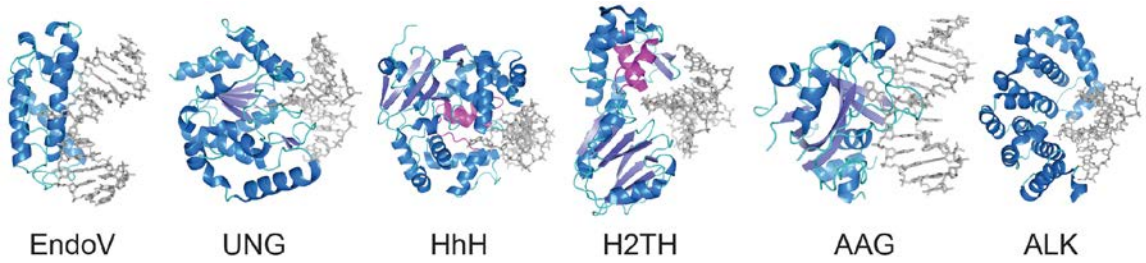


Fig. 1.3 DNA glycosylase structural superfamilies. Representative crystal structures from each class shown are: EndoV, T4 pyrimidine dimer DNA glycosylase EndoV (PDB ID 1VAS); UDG, human uracil-DNA glycosylase UDG (1EMH); Helix-hairpin-Helix (HhH), human 8-oxoguanine DNA glycosylase OGG1 (1YQK); Helix-two turn-helix (H2TH), *Bacillus stearothermophilus* 8-oxoguanine DNA glycosylase MutM (1L1T); AAG, human alkyladenine DNA glycosylase AAG/MPG (1EWN); ALK, *Bacillus cereus* alkylpurine DNA glycosylase AlkD (3JXZ). Proteins are colored according to secondary structure with the HhH and H2TH domains magenta. DNA is shown as gray sticks.

1.5 DNA Glycosylases in Alkylation Damage Repair

Alkylating agents are present in various environmental sources, including industrial processes, cigarette smoke, diet, and chemotherapy in addition to endogenous methyl donors. These agents 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, Walker et al. 2006)] (Fig. 1.4).

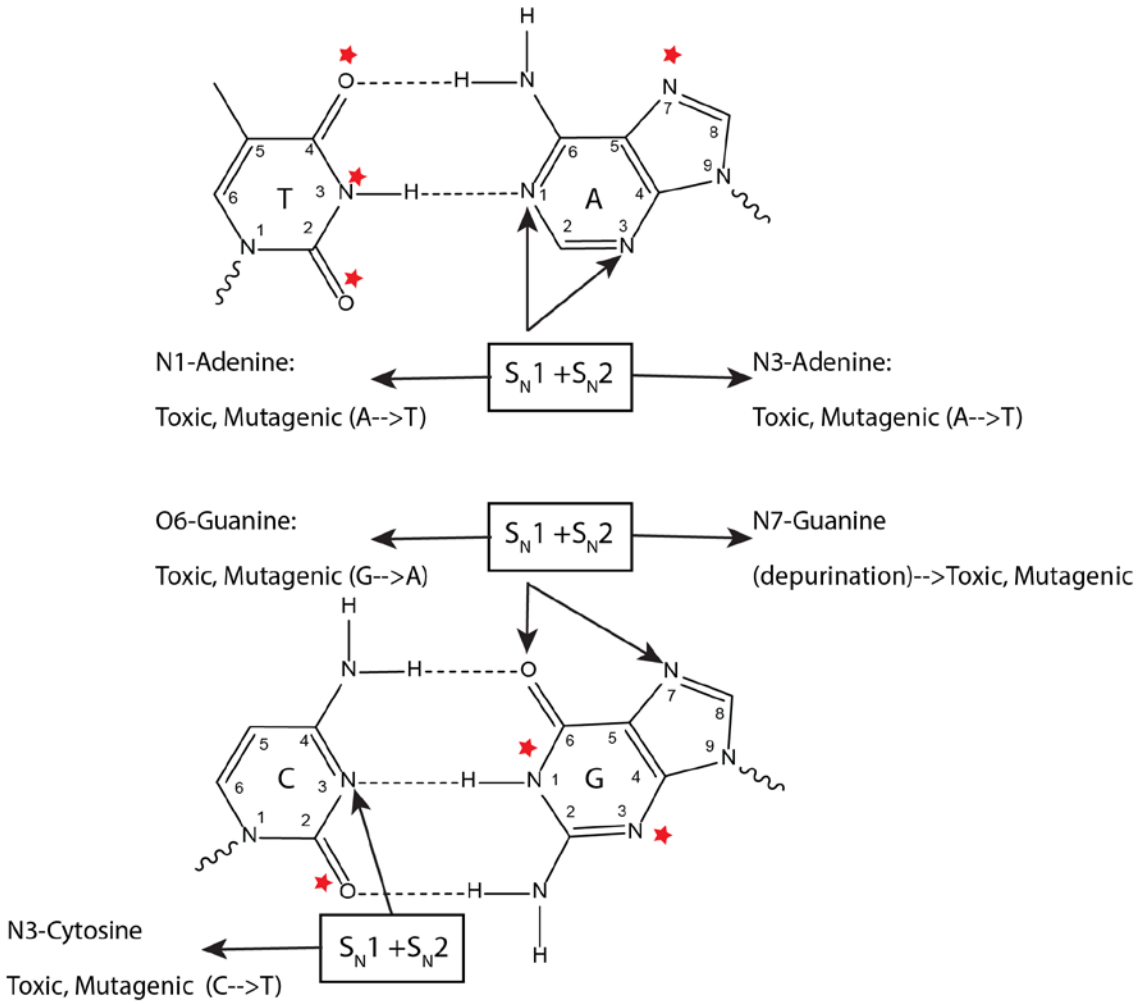


Fig. 1.4 Major sites of alkylation damage on DNA bases. Major sites of damage are delineated with arrows while minor sites are marked with red stars. [Adapted from (Fu, Calvo et al. 2012)]

Organisms have devised multiple DNA repair strategies to eliminate the damage generated by the constant assault of alkylating agents. Bases methylated at exocyclic substituents (e.g., O^6-methylguanine) are directly demethylated by DNA methyltransferases, whereas ring-substituted 1-methyladenine (1mA) and 3-methylcytosine (3mC) are specifically repaired through oxidative deamination by DNA dioxygenase (reviewed in Sedgwick 2004). The majority of alkylated bases, however,

are eliminated from the genome by the base excision repair (BER) pathway [reviewed in (Fromme and Verdine 2004; Huffman, Sundheim et al. 2005)].

1.5.1 Alkylpurine DNA Glycosylases

DNA glycosylases that remove alkylation damage have been characterized from eukaryotes, archaea, and bacteria. These include mammalian alkyladenine DNA glycosylase (AAG) (Brent 1979; Karran and Lindahl 1980), yeast methyladenine DNA glycosylase (*Saccharomyces cerevisiae* Mag and *Schizosaccharomyces pombe* Mag1) (Berdal, Bjørås et al. 1990; Chen, Derfler et al. 1990; Memisoglu and Samson 1996), *E. coli* 3-methyladenine (3mA) DNA glycosylase I (TAG) and II (AlkA) (Riazuddin and Lindahl 1978; Thomas, Yang et al. 1982), *Thermotoga maritima* methylpurine DNA glycosylase II (MpgII) (Begley, Haas et al. 1999), *Helicobacter pylori* 3mA DNA glycosylase (MagIII) (O'Rourke, Chevalier et al. 2000), and most recently *Bacillus cereus* AlkC and AlkD (Alseth, Rognes et al. 2006). Whereas most DNA glycosylases are specific for a single modification, alkylpurine DNA glycosylases can recognize a chemically diverse set of lesions (Fig 1.5A), including cytotoxic 3mA, 7-methylguanine (7mG), and the highly mutagenic 1,N^6-ethenoadenine (εA), which have been detected in humans and rats after exposure to various carcinogens (Shuker, Bailey et al. 1987; Shuker and Farmer 1992; Holt, Yen et al. 1998). TAG and MagIII are highly specific for 3mA and 3mG (Bjelland, Bjoras et al. 1993; O'Rourke, Chevalier et al. 2000), MpgII and AlkC/D are selective for positively charged lesions 3mA and 7mG (Begley, Haas et al. 1999; Alseth, Rognes et al. 2006), and AlkA and AAG can excise these lesions as well as other alkylated and modified bases, including εA and hypoxanthine (Hx) (McCarthy, Karran et al. 1984; Bjelland, Birkeland et al. 1994; Saparbaev, Kleibl et al. 1995).

Alkylpurine DNA glycosylases can be classified into three structural superfamilies (Fig 1.5B). The first is defined by the mixed α/β globular fold of AAG (Lau, Scharer et al. 1998), which bears no structural resemblance to any other protein in the Protein Data Bank (PDB). The second is the helix-hairpin-helix (HhH) superfamily of glycosylases and includes AlkA, TAG, MagIII, and MpgII (Labahn, Scharer et al. 1996; Begley, Haas et al. 1999; Drohat, Kwon et al. 2002; Eichman, O'Rourke et al. 2003). These enzymes contain a HhH DNA-binding motif and a common α-helical architecture also found in bacterial endonuclease III (Endo III) and MutY, archaeal MIG, and human 8-oxoguanine (OGG1) DNA glycosylases (Kuo, McRee et al. 1992; Guan, Manuel et al. 1998; Bruner, Norman et al. 2000; Mol, Arvai et al. 2002). *S. cerevisiae* MAG and *S. pombe* Mag1 adopt the HhH fold based on sequence similarity to AlkA (Chen, Derfler et al. 1990; Berdal, Johansen et al. 1998; Chapter II). Recently discovered AlkC and AlkD proteins from *Bacillus cereus* comprise the third group. AlkD forms a C-shaped α-helical fold from repeating HEAT motifs, and AlkC is expected to adopt a smilar fold (Dalhus, Helle et al. 2007; Rubinson, Metz et al. 2008).

1.5.2 Human AAG

Also known as MPG and ANPG, AAG excises a variety of alkylated purines, including 3mA, 7mG, and εA, as well as hypoxanthine (Hx), the oxidative deamination product of adenine (Fig. 1.5A) (McCullough, Dodson et al. 1999; Wyatt, Allan et al. 1999). Biochemical studies have suggested that Hx is the predominant biological substrate of AAG, given the exceptional rate enhancement of Hx excision relative to alkylated substrates (O'Brien and Ellenberger 2004).

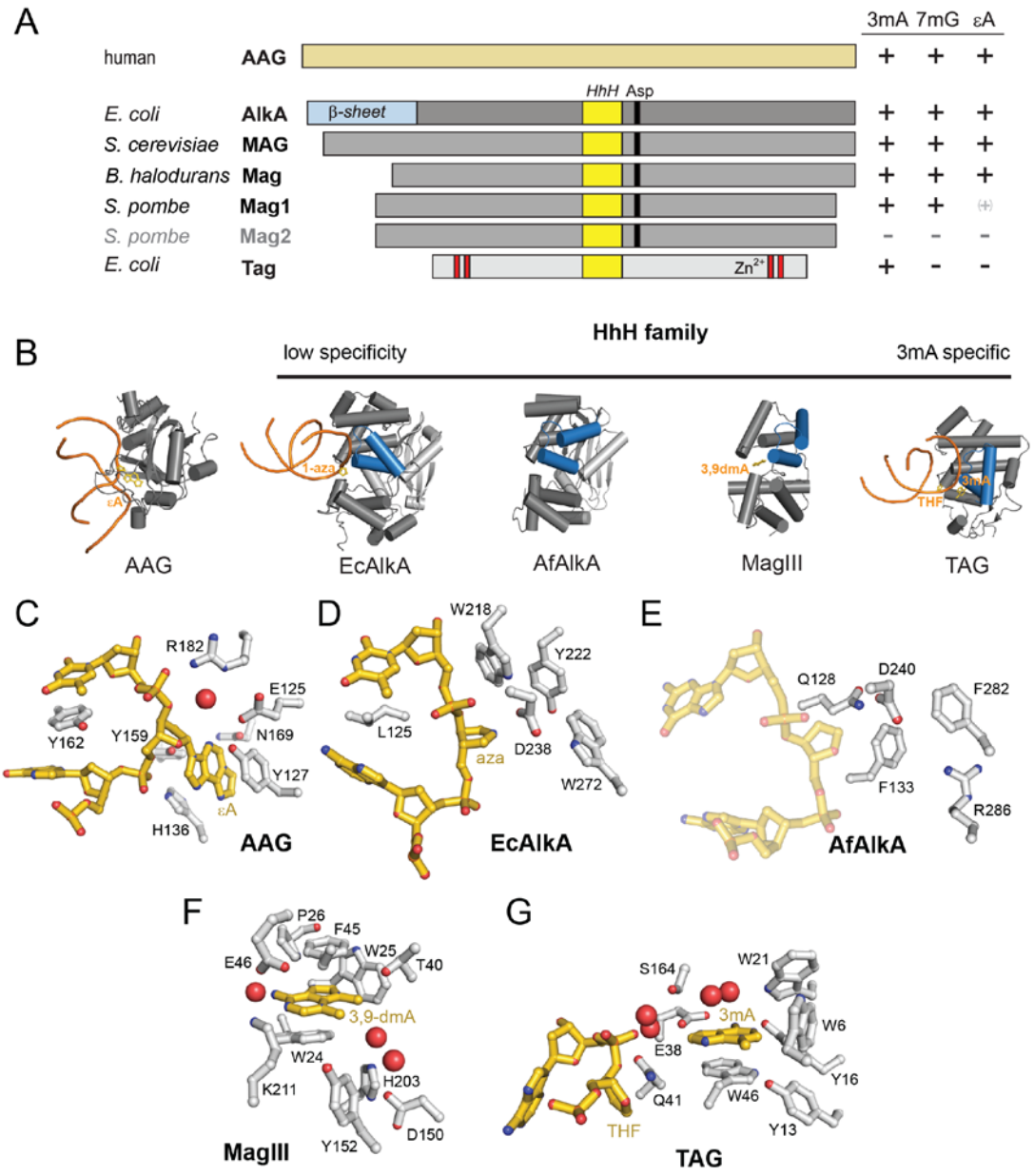


Fig. 1.5 Alkylpurine DNA glycosylase. (A) Schematic of alkylpurine DNA glycosylases from human AAG, bacterial AlkA, Tag Mag and MAG and Mag1 from yeast. The helix-hairpin-helis motif is shown in yellow and the conserved aspartate residue is shown in black. Activity against 3mA, 7mG and εA is shown as + or – on the right. (B) Overall architectures are shown on the top row, with HhH enzymes arranged in order of increasing specificity for 3mA. (C-G) Active sites. Protein and nucleic acid atoms are shaded grey and gold, respectively, and waters are shown as red spheres. (C) Human AAG/εA-DNA substrate complex (1EWN). (D) *E. coli* AlkA bound to 1-azaribose-DNA (1DIZ). (E) *A. fulgidus* AlkA (2JHJ) with THF-DNA modeled from the *S. pombe* Mag1/DNA complex (3S6I). (F) *H. pylori* MagIII/3,9-dimethyladenine (1PU7). (G) *E. coli* TAG/THF-DNA/3mA product complex (2OFI) [Adapted from (Rubinson, Adhikary et al. 2010)].

AAG also has been shown to excise 1-methylguanine and 1,N^2-εG (Saparbaev, Langouet et al. 2002; Lee, Delaney et al. 2009). Crystal structures of a catalytic fragment of AAG bound to oligonucleotides containing a pyrrolidine transition-state analog and an εA base showed that AAG is a single domain protein with a mixed α/β structure and a positively charged DNA binding surface unique to DNA glycosylases (Lau, Scharer et al. 1998; Lau, Wyatt et al. 2000). Tyr162 on the tip of a β-hairpin, plugs the gap in the DNA left by the flipped nucleotide, and the flipped εA base is stacked between two tyrosine residues (Tyr127 and Tyr159) and His136 inside the active site cavity (Fig. 1.5C). Structural studies of AAG have provided a platform for biochemical studies aimed at dissecting the catalytic mechanism, substrate preference and crosstalk with other BER proteins.

It has been suggested that substrate discrimination against undamaged bases arises from unfavorable interactions with exocyclic N6 and N2 amino groups inside the active site (O'Brien and Ellenberger 2004). For example, His136 donates a hydrogen bond to N6 of εA, whereas adenine cannot accept a hydrogen bond at this position. Furthermore, steric clash between N2 amino group of guanine makes it an unfavorable target. The latter observation was supported by the finding that mutation of Asn169 (which clashes with the exocyclic N2 amino group) which gives AAG enhanced activity toward guanine (Connor and Wyatt 2002). The rate enhancement for excision of 3mA by AAG is one and three orders of magnitude less than that of εA and Hx, respectively (O'Brien and Ellenberger 2004). This hints that AAG may remove positively charged lesions because of their inherent lability and not through molecular recognition of the methyl group. The structures also offer a glimpse into a possible catalytic mechanism. An ordered water

molecule is located 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. 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 SN2 catalytic mechanism (O'Brien and Ellenberger 2003).

1.5.3 AlkC and AlkD

AlkC and AlkD from *Bacillus cereus* represent a new structural class of alkylpurine glycosylases specific for 3mA and 7mG (Alseth, Rognes et al. 2006; Rubinson, Metz et al. 2008). AlkD is composed exclusively of HEAT repeats (Fig. 1.3) — short α-helices in tandem pairs that come together to form extended, non-enzymatic scaffolds that generally mediate protein-protein interactions (Rubinson, Metz et al. 2008). AlkD is the first HEAT repeat protein that interacts with nucleic acids and/or to contain enzymatic activity (Rubinson and Eichman 2012). The HEAT repeats form a solenoid whose concave face is positively-charged, perfectly suited to interact with the negatively charged DNA backbone. The protein-DNA interface is also lined with highly conserved residues that are important for DNA binding, catalysis and protection against alkylation sensitivity in alkylation damage repair-impaired bacteria (Dalhus, Helle et al. 2007; Rubinson, Metz et al. 2008; Rubinson, Gowda et al. 2010).

Recently, high resolution crystal structures of AlkD in complex with DNA containing substrate (3-deaza-3-methyladenine, 3d3mA) and product (THF) mimics have begun to provide insights into a novel mechanism of damage capture and catalysis (Rubinson, Metz et al. 2008; Rubinson, Gowda et al. 2010). Unlike other DNA glycosylases, AlkD positions the 3d3mA and THF moieties away from the protein. In the

case of the substrate complex, the 3d3mA•T base pair is broken and the thymine is pushed toward the protein and into minor groove. The product structure showed a more dramatic distortion around the abasic site with THF moiety flipped 180° out of the helix. The orphaned base was also extruded from the helix toward the minor groove and stacked against the protein. This creates a single base bulge in the duplex with the adjacent bases maintaining base stacking. The structural and biochemical work show that AlkD lacks the traditional plug and wedge residues and is not inhibited by high concentrations of free base. Although these characteristics represent a novel mode of alkylation damage repair, the source of substrate specificity is still poorly understood as the enzyme itself doesn not make any contacts with the lesion.

1.5.4 HhH superfamily

As mentioned earlier, the HhH fold is by far the most common structural fold among glycosylases with the majority of bacterial, archaeal and fungal proteins adopting the fold (Denver, Swenson et al. 2003). Notable exceptions to this rule are AlkC/AlkD and bacterial orthologs of human AAG (Aamodt, Falnes et al. 2004; Rubinson, Metz et al. 2008). The HhH glycosylases contain two α-helical domains with the active site cleft located at their interface. The domain containing the HhH motif and DNA intercalating residues is formed from an internal region of the primary structure and has a relatively conserved tertiary structure. The HhH anchors the protein to the DNA through a series of hydrogen bonds between main-chain atoms of the hairpin and the phosphoribose backbone downstream of the lesion. At the damage site, bulky side chains from neighboring loops fill the void left by the extrahelical nucleobase target and wedge into the base stack opposite the flipped out nucleotide. Both plug and wedge residues are

important for stabilizing the bent conformation of the DNA and have been implicated in probing the DNA helix during the search process (Bowman, Lee et al. 2010). The second domain, formed from the N- and C-termini, is more structurally divergent and often contains additional structural elements, such as a zinc ion (TAG), iron-sulfur cluster (MpgII), or carbamylated lysine (MagIII) (Rubinson, Gowda et al. 2010).

Comparative analysis of the HhH alkylpurine glycosylases has been instrumental in deciphering the physical and chemical determinants of substrate recognition (Rubinson, Adhikary et al. 2010). On the one hand, we have learned that the HhH scaffold accommodates a diverse array of nucleobase binding pockets that discriminate between lesions on the basis of shape complementarity. For example, the nucleobase binding surface of AlkA is a shallow cleft that can accommodate a variety of alkylpurines, whereas the active sites of TAG and MagIII are constrained and perfectly shaped for 3mA. On the other hand, this steric selection is not the only determinant of specificity, because some active sites can accommodate nucleobases that they do not excise (e.g., Mag1) (Chapter II). In addition, the catalytic requirements for excision of cationic lesions 3mA and 7mG differ from the uncharged alkylpurines (e.g., εA) by virtue of their weaker N-glycosidic bonds (Stivers and Jiang 2003). Hence, the inherent instability of these lesions render their excision highly dissociative, and recent reports suggest that cationic lesions may be removed and even detected within DNA differently than neutral lesions (Metz, Hollis et al. 2007; Rubinson, Gowda et al. 2010; Chapter II)

1.5.4.1 *E. coli* AlkA

Crystal structures of unliganded AlkA identified the enzymes as a member of the HhH superfamily and revealed a shallow nucleobase binding surface that can

accommodate a variety of alkylpurines, a feature that helped to explain its broad specificity (Labahn, Scharer et al. 1996; Yamagata, Kato et al. 1996) (Fig. 1.4D). In addition to the two-domain HhH architecture, AlkA contains an amino-terminal β-sheet domain of unknown function that is also present in bacterial and eukaryotic 8oxoG glycosylase OGG1 (section 1.6.1). A structure of AlkA bound to DNA containing 1-azaribose, which mimics the oxocarbenium reaction intermediate, has contributed greatly to our understanding of these enzymes (Hollis, Ichikawa et al. 2000; Hollis, Lau et al. 2000). The HhH anchors the protein to the DNA and does not directly participate in lesion recognition. The DNA is kinked by ~60° around the 1-azaribose, is rotated 180° around the phosphoribose backbone, and is pointed into the active site cleft (Fig. 1.5D). Leu125 plugs the gap in the DNA. Rotation of the 1-azaribose into the active site places the N1′ nitrogen directly adjacent to the carboxylate group of the catalytic Asp238, which likely stabilizes the oxocarbenium intermediate (Hollis, Ichikawa et al. 2000). In addition to this lesion-specific binding mode, AlkA has the ability to bind to DNA ends (Zhao and O'Brien 2011). This can lead to heterogeneous complex population which may explain why a structure of AlkA bound to a substrate DNA has not been determined. Nonetheless, this feature was exploited to develop a host-guest crystallization strategy to determine structures of various lesions in DNA where the oligonucleotide containing the target lesion is suspended between two protein molecules (Bowman, Lee et al. 2008).

High resolution structures of AlkA cross-linked to undamaged DNA bases provided insight into how the enzyme detects damage within the context of unmodified DNA (Bowman, Lee et al. 2010). Not surprisingly, the most notable differences between these undamaged DNA complexes (UDCs) and the 1-azaribose lesion recognition

complex (LRC) are centered around the lesion. The UDCs do not exhibit the kink present in the LRC DNA. The domain containing most of the catalytically important residues, including Asp238, is shifted 2.4 Å toward the lesion strand in the LRC compared to the UDCs. This movement, combined with a modest 1-Å shift of the Leu125 plug residue toward the lesion strand, clamps the lesion between the two domains and creates additional protein contacts that stabilize the LRC. In contrast, the HhH motif makes the same DNA contacts in LRC and UDC structures, providing additional evidence that the HhH motif is a non-specific DNA binding motif and is not involved in distorting the DNA for catalysis. Leu125 in the UDCs does not interact with the DNA, although it is still present in the minor groove. The phosphate backbone in the LRC is significantly (~9 Å) closer to the protein, which allows the Leu125 side-chain to intercalate into the DNA base stack in that structure. A 3mA base modeled in place of a centrally located cytosine indicates that Leu125 likely makes van der Waals contacts with the *N*3-methyl group of the 3mA (Bowman, Lee et al. 2010). These observations suggest that AlkA employs a passive scanning mechanism along the minor groove and uses the Leu125 side chain to detect abnormal bases and flip them into the active site.

1.5.4.2 Archaeal AlkA

An AlkA ortholog from the archaeon *Archaeoglobus fulgidus* (AfAlkA), has been shown to excise 1mA and 3mC in addition to 3mA, 7mG, εA and Hx from DNA (Birkeland, Anensen et al. 2002; Mansfield, Kerins et al. 2003; Leiros, Nabong et al. 2007). The crystal structure of this ortholog shows that the nucleobase binding pockets of AfAlkA and *E. coli* AlkA are strikingly different despite the similarity in overall enzyme architectures (Leiros, Nabong et al. 2007) (Fig. 1.5D, E). The substrate nucleobase is

predicted to stack between Phe133 and Phe282, similar to stabilization of 3mA by MagIII (section 1.5.4.3, Fig, 1.5F). In support of this, substitution of Phe133 or Phe282 with alanine diminishes εA and 1mA base excision, and the double mutant abrogates activity. Arg286 is predicted to orient εA in the active site through hydrogen bonding, but would potentially repel the protonated amine groups of 1mA and 3mC which are normally repaired via direct demethylation (Leiros, Nabong et al. 2007). Mutation of the catalytic Asp240 (Asp238 in EcAlkA) completely eliminates base excision activity in AfAlkA.

1.5.4.3 *H. pylori* MagIII and *T. maritima* MpgII

MagIII and MpgII are related alkylpurine glycosylases identified by their sequence similarity to EndoIII (Begley, Haas et al. 1999; O'Rourke, Chevalier et al. 2000). MagIII is highly specific for 3mA but can excise mispaired 7mG, whereas MpgII can excise both 3mA and 7mG (Begley, Haas et al. 1999; O'Rourke, Chevalier et al. 2000). The crystal structure of MagIII showed a unique feature in the N/C-terminal domain, which contains a carbamylated lysine (Lys205) that neutralizes an otherwise highly positively charged region of the protein (Eichman, O'Rourke et al. 2003). The preference MagIII for 3mA can be explained by the snug fit of 3mA inside the active site, which partially excludes *N*7-substituted purines. Structures of MagIII bound to positively charged 3,9-dimethyladenine (3,9-dmA) and neutral εA bases showed the nucleobases to stack between Phe45 and Trp24 and to be bound on three sides by Trp25, Pro26, and Lys211 (Fig. 1.5F). Other than these van der Waals and π-stacking interactions, there are no specific hydrogen bonded or polar contacts to the adenine ring like those observed in TAG (see section 1.5.5.4). Similar to spMag1, mutation of the putative catalytic aspartate Asp150 in MagIII does not completely abrogate base excision activity. This suggests that

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

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

1.5.4.4 *E. coli* TAG

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

1.5G). Binding studies and NMR investigation of 3mA in the active site led to the suggestion that TAG enhances the rate of 3mA depurination by binding tightly to the nucleobase, thereby destabilizing the ground state of the enzyme-substrate complex (Cao, Kwon et al. 2003). This idea was illustrated by crystal structures of a TAG/abasic-DNA/3mA product complex using the *Salmonella typhi* ortholog, which is 82% identical and 92% conserved overall with *E. coli* TAG (Metz, Hollis et al. 2007). In that structure, the bound DNA is more B-form when compared to the highly distorted 1-azaribose DNA bound to AlkA, and there was a large (7 Å) separation between the THF, which is not fully engaged inside the active site, and 3mA, which is buried deep inside the cleft. These observations indicated that the DNA undergoes significant relaxation upon breakage of the N-glycosidic bond, and it was suggested that steric strain may contribute to bond cleavage (Metz, Hollis et al. 2007). A recent structure of *Staphylococcus aureus* TAG recapitulates the structural features observed in the *E. coli* and *S. typhi* structures, and the authors suggested that tautomerization of 3mA contributes to its recognition by TAG (Zhu, Yan et al. 2012).

1.5.5.5 Yeast Mag/Mag1/Mag2

S. cerevesiae Mag *and S. pombe* Mag1 are 42% and 47% similar in sequence to *E. coli* AlkA, respectively, but have a more restricted substrate specificity (Fig. 1.5A) (Rubinson, Adhikary et al. 2010). Mag excises 3mA, 7mG, εA, Hx, and guanine, but not oxidized substrates (e.g., O^2-methylthymine) from DNA, while Mag1 is restricted to 3mA, 3mG, and 7mG and has only a modest activity toward εA (Saparbaev and Laval 1994; Bjørås, Klungland et al. 1995; Saparbaev, Kleibl et al. 1995; Berdal, Johansen et al. 1998; Alseth, Osman et al. 2005; Lingaraju, Kartalou et al. 2008) Chapter II) .

Differences in substrate preferences are reflected in cells. For example, Mag deletion strains are more sensitive to alkylating agents than are *S. pombe* Mag1 deletion strains. In addition, fold induction of Mag is higher than Mag1 upon exposure to alkylation agents (Chen and Samson 1991; Memisoglu and Samson 2000). These phenotypic differences suggest that these proteins have different roles in protecting cells against alkylation damage (Memisoglu and Samson 2000; Memisoglu and Samson 2000).

The *S. pombe* genome codes for a second Mag1-like protein, Mag2, which despite being 41% identical and 69% similar to Mag1 shows no discernible glycosylase activity (Alseth, Osman et al. 2005). Like *mag1* deletion mutants, *mag2*Δ does not increase alkylation sensitivity. Interestingly, deleting *mag2* in an *nth1*Δ (AP lyase-deficient) background increased the resistance the *nth1*Δ strain. Moreover, exogenous expression of either Mag1 or Mag2 in a *rad16*Δ (NER)/*mag1*Δ/*mag2*Δ triple mutant also restored MMS resistance lost as a result of the deletions. Together, genetic studies have hinted that there may be considerable amount of cross-talk in the response of yeast to alkylation damage (Memisoglu and Samson 2000; Kanamitsu, Tanihigashi et al. 2007). Regardless, the lack of base excision activity of Mag2 is confounding given that it contains all the hallmarks of an alkylpurine glycosylase.

There was a relative dearth of structural and biochemical information available on yeast alkylpurine glycosylases at the time of my dissertation proposal. Mag1, Mag and Mag2, despite remarkable sequence similarity, represent the range of substrate specificity seen in the broader family of alkylpurine glycosylases. This observation provided me with an opportunity to embark on a detailed structure-function study of these

glycosylases to parse out the source of substrate preference between Mag and Mag1 and understand the structural basis of inactivity of Mag2 (Chapters II-IV).

Glycosylases involved in oxidative DNA damage and uracil repair

In the following two sections I introduce a few DNA glycosylases involved in repair of oxidative damage of DNA and removal of uracil that will be referred to in the following chapters. The last two decades have seen a sharp rise in studies focused on understanding how glycosylases translocate on DNA and search and identify a particular lesion. Most of these studies have focused on bacterial and human enzymes that repair oxidative damage (OGG1, MutM and MutY) or remove uracil (UNG) (Drohat, Jagadeesh et al. 1999; Jiang, Kwon et al. 2001; Fromme, Bruner et al. 2003; Fromme, Banerjee et al. 2004; Banerjee, Yang et al. 2005; Banerjee, Santos et al. 2006; Banerjee and Verdine 2006; Parker, Bianchet et al. 2007; Blainey, Luo et al. 2009; Qi, Spong et al. 2009; Friedman and Stivers 2010). For a comprehensive review of recent progress in structural and biochemical studies of DNA glycosylases the reader is directed to (Brooks, Adhikary et al. 2012).

1.6 Oxidation Damage

DNA bases undergo oxidative damage from chemical oxidants, free radicals and reactive oxygen species (ROS) produced from cellular respiration, inflammatory responses, and ionizing radiation (Klaunig and Kamendulis 2004; Valko, Rhodes et al. 2006; van Loon, Markkanen et al. 2010). Oxidized bases often are used as biomarkers for oxidative stress and cancer (Klaunig and Kamendulis 2004; Kryston, Georgiev et al. 2011). Guanines are especially susceptible to oxidation, leading to a number of lesions

that are substrates for BER (Fig. 1.1A) (Neeley and Essigmann 2006). Attack of a hydroxyl radical at the C8 position of guanine produces 7,8-dihydro-8-hydroxyguanine (8-OHG), which tautomerizes to 8-oxo-7,8-dihydroguanine (8oxoG), or the ring-opened 2,6-diamino-5-formamido-4-hydroxy-pyrimidine (FapyG), two of the most abundant oxidative DNA adducts (Burrows and Muller 1998; Evans, Dizdaroglu et al. 2004). 8oxoG is a particularly insidious lesion because of its dual coding potential by replicative polymerases. This leads to G→T transversion mutations likely as a result of its ability to form both 8oxoG(syn)•A(anti) and 8oxoG(anti)•C(anti) base pairs (Cheng, Cahill et al. 1992; Brieba, Eichman et al. 2004; Hsu, Ober et al. 2004; Klaunig and Kamendulis 2004; van Loon, Markkanen et al. 2010). Oxidation of guanine and 8oxoG also produces a variety of ring-opened purines in addition to FapyG, including hydantoin lesions, spiroiminodihydantoin (Sp), guanidinohydantoin (Gh), and its isomer iminoallantoin (Ia) (Fig. 1.1A) (Luo, Muller et al. 2001; Burrows, Muller et al. 2002; Henderson, Delaney et al. 2003). While most Fapy lesions inhibit DNA polymerases some (FapyA and Fapy7mA) are potentially mutagenic and can lead to A→G transitions (Tudek 2003). Hydantoin lesions have been suggested to lead to an increase in G→T and G→C transversions and stall the replication machinery (Leipold, Muller et al. 2000; Burrows, Muller et al. 2002; Henderson, Delaney et al. 2003; Delaney, Neeley et al. 2007). In addition to purines, reaction of hydroxyl radicals at positions 5 or 6 of thymine produces 5,6-dihydroxy-5,6-dihydrothymine (thymine glycol, Tg), a cytotoxic lesion that distorts the DNA duplex and can inhibit replication (Evans, Dizdaroglu et al. 2004; Aller, Rould et al. 2007). Other potentially harmful pyrimidines include dihydrothymine (DHT), dihydrouracil (DHU), 5-hydroxyuracil (5-OHU), 5-hydroxycytosine (5-OHC), 5-

hydroxymethyluracil (5hmU), and 5-formyluracil (5fU) (Boorstein and Teebor 1988; Purmal, Kow et al. 1994; Yoshida, Makino et al. 1997; Kreutzer and Essigmann 1998; Liu and Doetsch 1998; Shikazono, Pearson et al. 2006).

DNA glycosylases that remove oxidative DNA damage can be categorized on the basis of their preferences for purine or pyrimidine lesions and their structural folds (APPENDIX D). Oxidized purines, including 8oxoG and FapyG, are removed from DNA by 8oxoG DNA glycosylase (OGG1) in eukaryotes and MutM (also known as FapyG DNA glycosylase, Fpg) in bacteria (recently reviewed in van Loon, Markkanen et al. 2010). Oxidized pyrimidines are removed by endonuclease III (EndoIII, or Nth) and endonuclease VIII (Endo VIII, or Nei), and their eukaryotic orthologs, NTH1 and NEIL1 (Nei-like1), respectively. Despite their different substrates, OGG1 and EndoIII/Nth adopt a common architecture characteristic of the Helix-hairpin-Helix (HhH) superfamily of DNA glycosylases (Nash, Bruner et al. 1996). MutM/Fpg and EndoVIII/Nei also are structurally similar, with helix-two turn-helix (H2TH) and antiparallel β-hairpin zinc finger motifs. They also share a common bifunctional catalytic mechanism involving both base excision and AP lyase activities (Bailly, Verly et al. 1989; Jiang, Hatahet et al. 1997; Sugahara, Mikawa et al. 2000; Zharkov, Golan et al. 2002; Zharkov, Shoham et al. 2003).

Eukaryotic OGG1 and bacterial MutM/Fpg preferentially catalyze removal of 8oxoG across from C (Castaing, Geiger et al. 1993; van der Kemp, Thomas et al. 1996). Recent work has aimed to understand how these enzymes locate base damage amidst the sea of unmodified DNA (for an excellent review, see ref. David, O'Shea et al. 2007). Both enzymes are bifunctional in that they contain base excision and AP lyase activities,

although recent reports suggest that human OGG1 (hOGG1) may function as a monofunctional glycosylase under physiological conditions (Nash, Bruner et al. 1996; Bruner, Norman et al. 2000; Dalhus, Forsbring et al. 2011). The OGG enzymes can be subdivided into three structural families (Fig. 1.6): (1) OGG1, including human OGG1 and the recently discovered *Clostridium acetobutylicum* (CaOGG) enzyme (Fig. 1.6A-C) (Aburatani, Hippo et al. 1997; Arai, Morishita et al. 1997; Bjørås, Luna et al. 1997; Nagashima, Sasaki et al. 1997; Radicella, Dherin et al. 1997; Roldan-Arjona, Wei et al. 1997; Rosenquist, Zharkov et al. 1997; Robey-Bond, Barrantes-Reynolds et al. 2008; Faucher, Robey-Bond et al. 2009; Faucher, Wallace et al. 2009), (2) archaeal OGG2 (Fig. 1.6D-F) (Gogos and Clarke 1999; Faucher, Duclos et al. 2009), and (3) archaeal 8oxoG glycosylase (AGOG), represented by the *Pyrobaculum aerophilum* enzyme (Fig. 1.6G-H) (Sartori, Lingaraju et al. 2004). Structural studies of the various OGG orthologs (Faucher, Doublié et al. 2012) and of MutM (Fig. 1.6I-J) have enhanced our understanding of 8oxoG recognition and excision from two distinct protein architectures. Each of these three classes is discussed below.

1.6.1 hOGG1

A battery of recent structures of hOGG1 in complex with DNA containing an 8oxoG•C base pair (Lesion Recognition Complex, LRC) or a normal G•C base pair (Interrogation Complex, IC) from the Verdine group has been invaluable in understanding how DNA glycosylases recognize and discriminate their substrates from normal DNA (Bruner, Norman et al. 2000; Banerjee, Yang et al. 2005; Banerjee and Verdine 2006; Radom, Banerjee et al. 2007). The original hOGG1 LRC structure was obtained from a catalytically inactive Lys249Gln mutant bound to DNA containing an

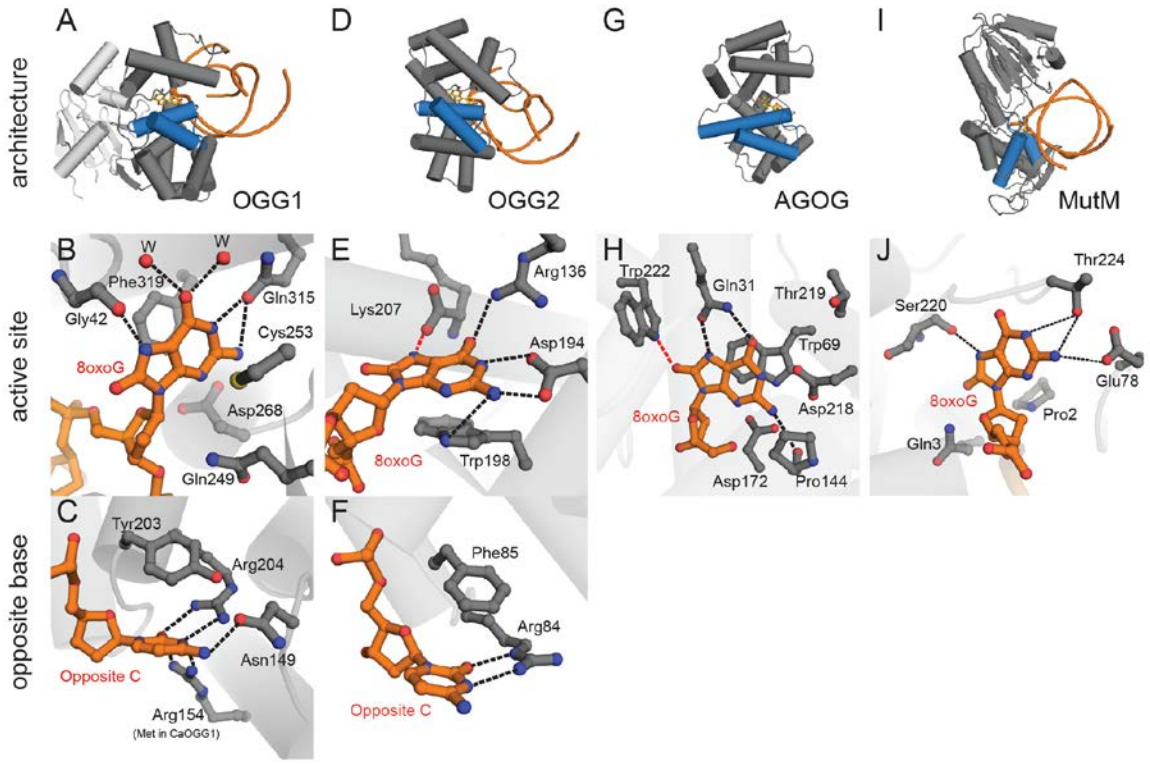


Fig. 1.6 Oxidative DNA glycosylases. (A-C) OGG1, represented by human OGG1 (PDB ID 1EBM), (D-F) OGG2, represented by MjOGG (PDB ID 3KNT), (G-H) *Pyrobaculum aerophilum* AGOG (PDB ID 1XQP) and (I-J) *Geobacillus stearothermophilus* MutM. The overall folds of each enzyme are shown on the top row (blue HhH motif), active sites on the second row, and opposing base on the bottom row. In the close-up views, the protein side-chains are grey and the DNA orange. Water molecules are represented by red spheres and hydrogen bonds are shown as dashed lines. (B) The human OGG1 8oxoG recognition pocket. The only 8oxoG specific contact is the hydrogen bond from the carbonyl group of Gly42 to the protonated N7 of 8oxoG. (C) The high specificity of hOGG1 for 8oxoG•C base pairs can be rationalized by the 5 hydrogen bonds between the opposite cytosine and 3 side chains. (E) In MjOGG, the 8oxoG N7 donates a hydrogen bond (red dashed line) to the C-terminal Lys207 carboxylate. (F) The opposite cytosine in MjOGG is contacted by only one side chain. (H) 8oxoG nucleoside bound inside the AGOG active site, with a unique 8oxoG-specific contact to Trp222 (red dashed line). (J) Active site of MutM (PDB ID 1R2Y) shows multiple contacts to 8oxoG but lacks the aromatic residues seen in the OGG1, OGG2, and AGOG enzymes. [Adapted from (Brooks, Adhikary et al. 2012))].

8oxoG•C base pair (Bruner, Norman et al. 2000). It revealed how hOGG1 utilizes the HhH architecture to kink the DNA duplex, disrupt the 8oxoG•C base pair, and extrude the 8oxoG out of the helix and into a base binding pocket (Bruner, Norman et al.

2000). Of the multiple contacts to the extrahelical 8oxoG, only one—between the carbonyl oxygen of Gly42 and the N7 hydrogen of 8oxoG—is specific to 8oxoG (Fig. 1.6B) and was thus proposed to account for the ability of OGG1 to distinguish 8oxoG from G. The position of the backbone and the integrity of the 8oxoG-specific hydrogen bond are not dependent on glycine in this position, as a Gly42Ala substitution did not alter the protein backbone conformation or disrupt the hydrogen bond (Radom, Banerjee et al. 2007).

In the hOGG1 IC structure, which used a disulfide crosslinking strategy to trap the enzyme bound to a G•C base pair, the extrahelical guanine was situated in a pocket adjacent to the active site that the authors termed the 'exo' site (Banerjee, Yang et al. 2005). In a subsequent IC structure, in which the enzyme was forcibly presented with a G•C base pair adjacent to 8oxoG, the extrahelical guanine was not observed in the active or exo sites, likely as a result of steric and electrostatic clashes imposed by the 8oxoG (Banerjee and Verdine 2006). In both of these ICs, the protein (Asn149Cys) was crosslinked to the cytosine opposite the extrahelical G. In a more recent structure of a catalytically active hOGG1/G•C-DNA complex that was crosslinked (Ser292Cys) at a more remote location from the lesion, the target guanine was fully engaged inside the active site in a virtually identical position as 8oxoG in the LRC. In the IC, however, the guanine remained uncleaved, presumably because it lacks the N7 hydrogen that 8oxoG uses to form a specific hydrogen bond with the carbonyl of Gly42 (Crenshaw, Nam et al. 2012). The alignment of active site residues other than Gly42 also are important for catalysis, as observed in a phototrapped, uncleaved hOGG1/8oxoG-DNA complex that showed an intact 8oxoG-Gly42 interaction amidst a collection of side chain conformers

that differed from their position in the LRC (Lee, Radom et al. 2008). Taken together, these data demonstrated that hOGG1 recognition of 8oxoG within DNA occurs in multiple steps, and that 8oxoG excision relies on precise chemical compatibility within the base binding pocket.

hOGG1 has been regarded as a bifunctional DNA glycosylase involving two key catalytic residues, Asp268 and Lys249 (Nash, Lu et al. 1997; Bruner, Norman et al. 2000; Bjørås, Seeberg et al. 2002; Norman, Chung et al. 2003). The proposed catalytic mechanism involves Asp268-dependent deprotonation of the Lys249 ε-amino group, which forms a Schiff base with ribose C1′ of the 8oxoG nucleotide, resulting in β-elimination. However, various groups have reported monofunctional glycosylase activity for hOGG1 *in vivo* (Zharkov, Rosenquist et al. 2000; Hill, Hazra et al. 2001; Vidal, Hickson et al. 2001; Kuznetsov, Koval et al. 2005; Morland, Luna et al. 2005). Recently, Dalhus and colleagues used structural and mutational analysis to show that the weak AP lyase activity in hOGG1 is an artifact of the proximity of Lys249 to the C1′ and may not reflect a physiological role (Dalhus, Forsbring et al. 2011). A double Lys↔Cys swap mutant (Lys249Cys/Cys253Lys) abrogated AP lyase activity while maintaining 8oxoG excision activity, and a Lys249Cys/Cys253Lys/Asp268Asn triple mutant also eliminated the base excision activity. A crystal structure of the triple mutant revealed that Lys253 was too far (4.7Å) away from the incoming C1′ to form the Schiff base, whereas Asn268 was in the same position as Asp268 in the wild-type enzyme. These results provided additional evidence for hOGG1 acting as monofunctional enzyme, in which Asp268 stabilizes an oxocarbenium intermediate during base hydrolysis (Norman, Bruner et al. 2001; Norman, Chung et al. 2003) and Lys249 helps to position 8oxoG in the active site.

The specificity of hOGG1 for 8oxoG•C base pairs likely results from the 'pentad' of hydrogen bonds between the enzyme (Arg204, Asn149 and Arg154) and the opposing cytosine base (Bruner, Norman et al. 2000) (Fig. 1.6C). Structures of an OGG ortholog from the bacterium *Clostridium acetobutylicum* CaOGG provided further insight into specificity for the opposing base (Robey-Bond, Barrantes-Reynolds et al. 2008; Faucher, Robey-Bond et al. 2009; Faucher, Wallace et al. 2009). Whereas hOGG1 displays a very high preference for C opposite 8oxoG (Bjørås, Luna et al. 1997), CaOGG can excise 8oxoG opposite any base. Structures of CaOGG in complex with DNA containing 8oxoG•C and 8oxoG•A showed that the bacterial protein maintains the fold and general DNA interactions as hOGG1. However, it lacks two of the five hydrogen bonds with the opposing base as a result of Met132 in place of the Arg154 in hOGG1 (Fig. 1.6C). In addition, the Asn149-cytosine hydrogen bond in hOGG1 is stabilized by the interaction of Asn149 with the hydroxyl group of Tyr203, which is missing in CaOGG (Phe179 at this position). Thus, the fewer number of stabilizing contacts with and around the opposite base in CaOGG creates an environment that can accommodate other nucleobases at this position.

1.6.2 MutM/Fpg

MutM/Fpg excises a number of oxidized nucleobases in addition to 8oxoG, including FapyG, hydantoins, Tg, DHU, and 5-OHU (Tchou, Kasai et al. 1991; Wiederholt, Delaney et al. 2003; Zharkov, Shoham et al. 2003). The crystal structure of *Thermus thermophilus* MutM/Fpg defined the structural architecture as distinct N- and C-terminal domains separated by a flexible hinge (Sugahara, Mikawa et al. 2000) (Fig. 1.6I,J). The N-terminal domain is comprised of a two layer β-sandwich flanked by α-

helices on either side and contains the catalytically important N-terminal proline and glutamate residues. The predominantly α-helical C-terminal domain contains the hallmark H2TH motif essential for DNA binding (Zharkov, Shoham et al. 2003). DNA-bound structures of MutM/Fpg from *Lactococcus lactis* (Serre, Pereira de Jesus et al. 2002) and *Geobacillus stearothermophilus* (Fromme and Verdine 2002) revealed that the DNA was severely kinked by ~75° with the lesion flipped into the active site similar to other DNA glycosylases (Fig. 1.6J). Subsequent structures detailed the interactions of the enzyme with various substrates and abasic analogs, including 8oxoG, FapyG, DHU, tetrahydrofuran (THF), 1,3-propanediol (Pr), hydroxy propanediol and hydantoin carbanucleoside (Gilboa, Zharkov et al. 2002; Fromme and Verdine 2003; Coste, Ober et al. 2004; Pereira de Jesus, Serre et al. 2005). These structures illustrated that even though specific amino acids contacting the base in the active site may differ, the orientation of the backbone deoxyribose remains relatively unchanged. This suggests that catalysis proceeds by properly positioning the deoxyribose ring (Pereira de Jesus, Serre et al. 2005). In addition, these MutM/abasic-DNA complexes suggested that β-elimination occurs concurrently with depurination, as opposed to sequential depurination-β-elimination reactions proposed previously for hOGG1, based on the fact that the enzyme sterically clashes with the cyclic, but not the ring-opened form of the deoxyribose (Fromme, Bruner et al. 2003; Pereira de Jesus, Serre et al. 2005).

More recently, a series of crystal structures of *Geobacillus stearothermophilus* MutM/Fpg from the Verdine laboratory provided significant insights into how the enzyme differentiates between 8oxoG and guanine in the context of duplex DNA (Fromme and Verdine 2002; Fromme and Verdine 2003; Banerjee, Santos et al. 2006).

MutM ICs crosslinked with DNA containing normal A•T or G•C base pairs showed Phe114 probing the minor groove, with the interrogated base pairs severely buckled but remaining intrahelical (Fromme and Verdine 2002). In the LRC structure of MutM crosslinked to 8oxoG-DNA, the Phe114 residue fully penetrates the base stack and helps to induce a severe kink in the DNA that allows the target 8oxoG to become extrahelical. The side chains of Met77 and Arg112 fill the space vacated by the flipped 8oxoG, with the Arg112 guanidinium moiety interacting with the Watson-Crick face of the estranged cytosine (Banerjee, Santos et al. 2006). A third set of so-called encounter complexes (ECs) with 8oxoG-DNA or Gua-DNA were determined using a variant form of *G. stearothermophilus* MutM that has an altered or absent oxoG capping loop, which normally interacts with 8oxoG in the active site (Qi, Spong et al. 2009; Qi, Spong et al. 2010). These complexes showed that MutM can detect the presence of intrahelical 8oxoG in the duplex based on local steric effects that influence the surrounding phosphate backbone. Recent data from the *E. coli* enzyme showed that the interaction with the 8oxoG capping loop is specific for 8-oxoG, since an EcMutM/Fpg variant lacking the tip of the capping loop can efficiently excise mFapyG, DHU, Sp, and Gh but not 8oxoG (Duclos, Aller et al. 2012). Furthermore, a recent study showed that hydrophobic isosteres of 8oxoG are good, and in some cases better, substrates for Fpg, demonstrating that hydrogen bonding to the base is not important for efficient excision by Fpg (McKibbin, Kobori et al. 2012). Taken together, these studies have provided detailed snapshots along the reaction pathway that illustrate how MutM/Fpg actively interrogates the DNA duplex to locate the 8oxoG lesion and how the base is extruded into the active site pocket for catalysis.

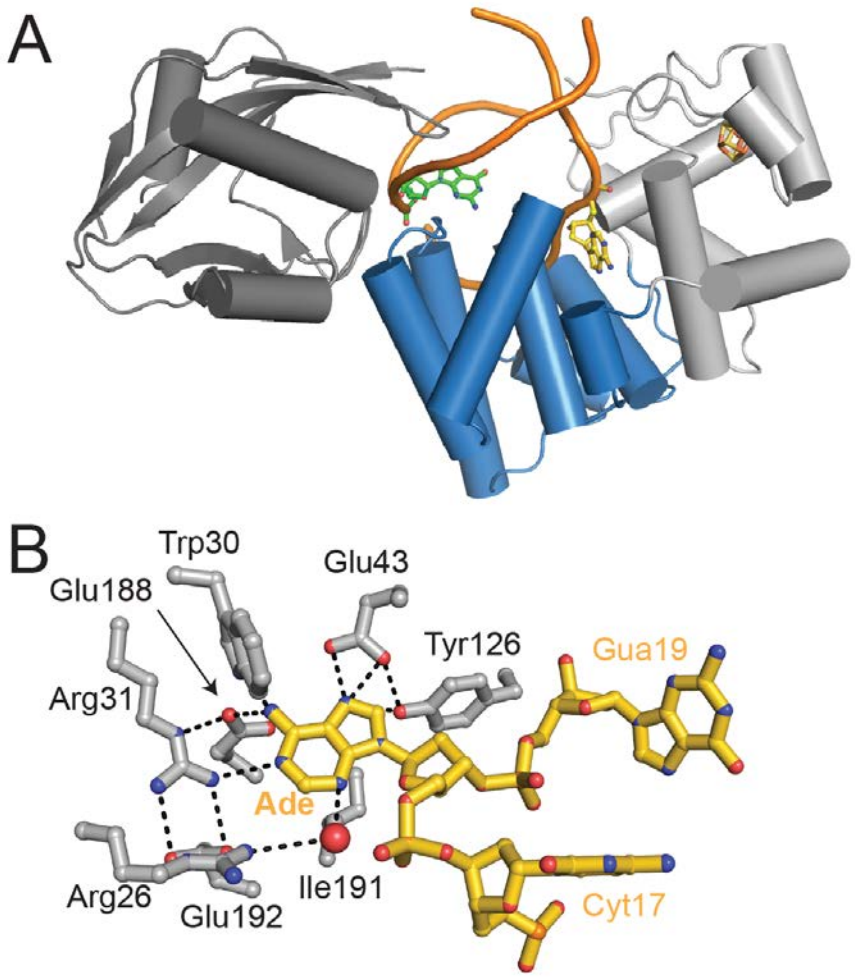


Fig. 1.7 Crystal structure of *Bacillus stearothermophilus* MutY. (A) Overall structure of BsMutY (PDB ID 1RRQ) colored by domain (silver, iron-sulfur cluster domain; blue, catalytic domain; gray, C-terminal (8oxoG recognition) domain. The DNA is colored orange with flipped-out adenine substrate in gold and opposite 8oxoG in green. (B) Active site details of the MutY fluorinated lesion recognition complex (FLRC) bound to adenine (3G0Q). Protein (silver) and nucleic acid (gold) atoms are shown as sticks, water molecules are shown as red spheres. Hydrogen bonds are shown as dashed lines. (Adapted from (Fromme, Banerjee et al. 2004) and (Lee and Verdine 2009))

1.6.3 MutY/MUTYH

Failure of MutM/OGG1 to excise 8oxoG from an 8oxoG•C base pair prior to replication results in 8oxoG•A mispairs. The adenine of the 8oxoG•A mismatch is the substrate for MutY/MUTYH glycosylase (Nghiem, Cabrera et al. 1988; Tominaga,

Ushijima et al. 2004). BER of the resulting AP site restores the 8oxoG•C pair, providing another chance for MutM/OGG1 to eliminate the 8oxoG from the DNA (reviewed in Michaels and Miller 1992). Structures of the catalytic domain of *E. coli* (Ec) MutY bound to adenine base revealed a HhH-FeS architecture similar to EndoIII and provided details of the active site and a proposed catalytic mechanism for adenine excision (Guan, Manuel et al. 1998; Manuel, Hitomi et al. 2004). Transition state analysis from kinetic isotope effect measurements confirmed a stepwise, dissociative (S_N1) reaction mechanism whereby Glu43 acts as a general acid to protonate adenine N7, which facilitates cleavage of the N-glycosidic bond. The resulting oxocarbenium ion is likely stabilized by nearby Asp144 and converted to the product AP site upon nucleophilic attack by water (McCann and Berti 2008). A high-resolution crystal structure of EcMutY bound to adenine provided evidence that MutY-catalyzed β-elimination, which involves Lys142, Lys20 and possibly Glu161, and is an activity secondary to and separable from the depurination reaction, similar to that observed in hOGG1 (see section 1.6.1.1) (Manuel, Hitomi et al. 2004).

A similar disulfide crosslinking strategy employed in the OGG1 and MutM structures was used to obtain structures of the full-length *B. stearothermophilus* homolog (BsMutY) anchored to 8oxoG•A-DNA (Fromme, Banerjee et al. 2004; Lee and Verdine 2009) (Fig. 1.7A). In this structure, the adenine is flipped into the glycosylase active site but remains uncleaved as a result of mutation of the catalytic aspartate (Asp144Asn) (Fromme, Banerjee et al. 2004). Surprisingly, no direct hydrogen bonds were observed between the catalytic domain and the extrahelical adenine substrate. A subsequent structure of a catalytically proficient (Asp144) BsMutY crosslinked to DNA containing a

non-hydrolyzable 2′-fluorinated deoxyadenosine showed adenine deeper into the active site and directly hydrogen bonded to Gln43, Tyr126, Arg31, Glu188, and Trp30 (Lee and Verdine 2009) (Fig. 1.7B). Mutation of the Glu188 residue in EcMutY (Gln182) decreased binding and activity for 8oxoG•A and G•A mismatches but increased binding affinity toward 8oxoG•T and G•T mismatches, which are not normal substrates for MutY (Chang, Madabushi et al. 2009). Cellular repair assays on the *E. coli* enzyme confirmed the importance of Asp138 (BsMutY144) and Glu37 (BsMutY Glu43) for the excision of adenine opposite 8oxoG (Brinkmeyer, Pope et al. 2012).

The C-terminal domain contributes specific contacts to the stacked 8oxoG lesion that are functionally important for lesion recognition and enzyme activity. Tyr88 intercalates the duplex and stacks against the 8oxoG nucleobase, and Gly260 contacts the phosphate 5′ to 8oxoG (Fromme, Banerjee et al. 2004). Inherited mutations at these positions in MUTYH (Tyr165Cys and Gly382Asp) have been implicated in the development of colorectal cancer (Al-Tassan, Chmiel et al. 2002). Substitution of the analogous residues in EcMutY (Tyr82Cys and Gly253Asp) reduce the DNA binding and base excision activities relative to the wild-type enzyme and the glycine has been implicated in discrimination of 8oxoG from G (Chmiel, Livingston et al. 2003; Livingston, Kundu et al. 2005). Furthermore, *in vivo* enzymatic studies with modified substrates demonstrated that MutY cannot effectively process adenine paired with guanine or modified forms of 8oxoG, whereas changes made to the target adenine are tolerated (Livingston, O'Shea et al. 2008), implying that recognition of the 8oxoG by the C-terminal domain is necessary for locating the misincorporated adenine.

A crystal structure of a human MUTYH consisting of the catalytic domain and the interdomain connector (IDC) that tethers the catalytic and C-terminal domains was determined recently (Luncsford, Chang et al. 2010). The human IDC sequence, which is not conserved in prokaryotic MutY, has been reported to recruit the Rad9, Rad1, Hus1 (9-1-1) complex involved in genome maintenance in eukaryotes (Chang and Lu 2005; Lu, Bai et al. 2006; Shi, Chang et al. 2006). Mutations in the IDC disrupted the MUTYH-9-1-1 interaction and decreased DNA repair of oxidative lesions *in vivo*, suggesting that structural studies of the human enzyme will reveal insights into its broader role in maintaining genome integrity (Shi, Chang et al. 2006; Luncsford, Chang et al. 2010).

1.7 Removal of Uracil

G•U mismatches arise from deamination of cytosine and lead to A•T transition mutations (Coulondre, Miller et al. 1978; Duncan and Miller 1980). Uracil is excised in eukaryotes by uracil DNA glycosylase (UNG, also known as UDG), single-stranded monofunctional uracil glycosylase (SMUG), and to a lesser extent by thymine DNA glycosylase (TDG). In bacteria, uracil is removed by the UNG ortholog, Ung, and mispaired uracil glycosylase (MUG) (Lindahl 1974; Gallinari and Jiricny 1996; Haushalter, Todd Stukenberg et al. 1999; Kavli, Sundheim et al. 2002). With the exception of MBD4 and MIG, which remove thymine from G•T mismatches and belong to the HhH superfamily, the UNG/TDG glycosylases adopt a highly conserved α/β fold (Fig. 1.8A) and can be divided into 4 subfamilies on the basis of sequence similarity and substrate specificity (Mol, Arvai et al. 1995; Mol, Arvai et al. 2002; Wu, Qiu et al. 2003).

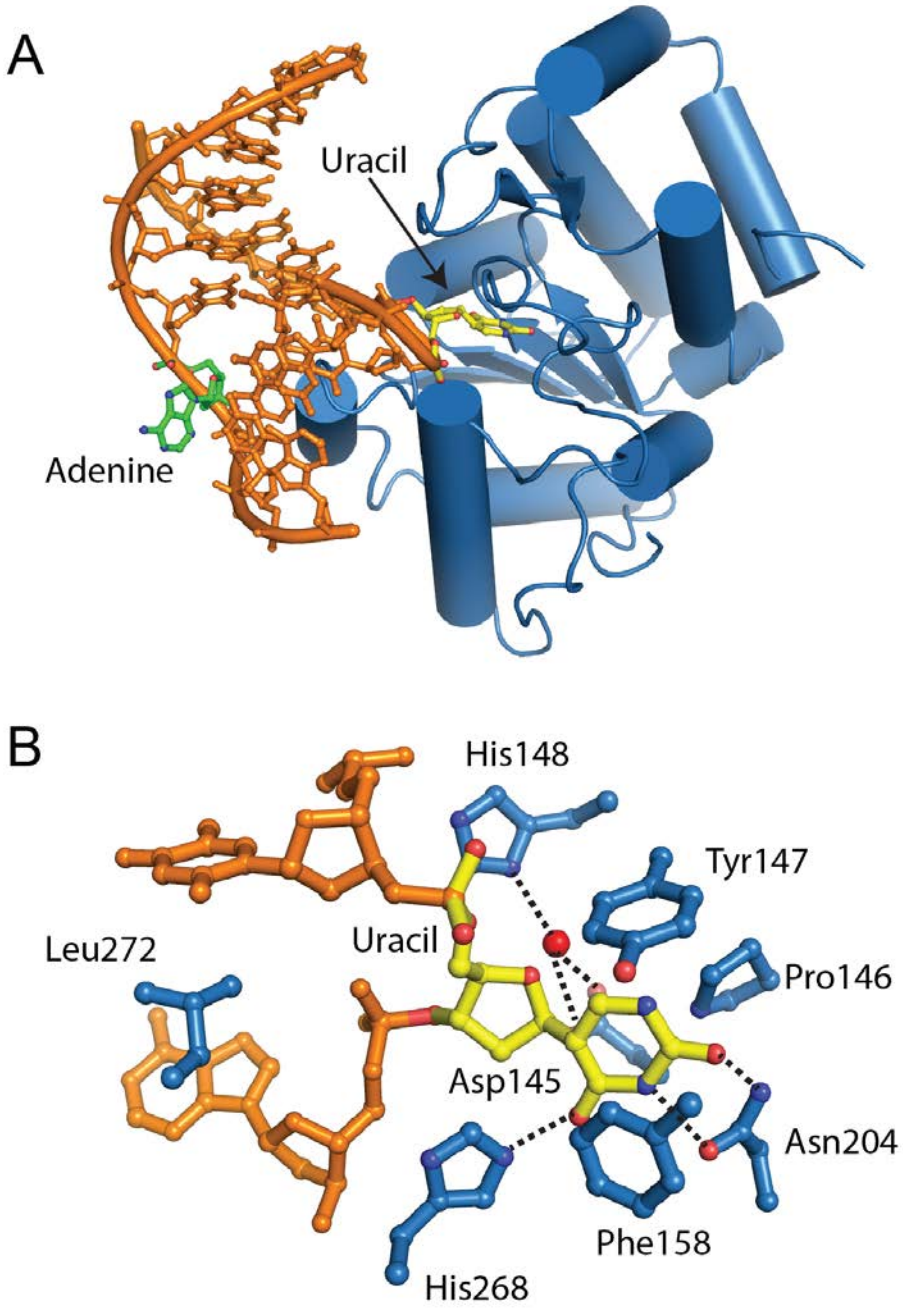


Fig. 1.8 Human Uracil DNA Glycosylase. (A) Crystal structure of human uracil DNA glycosylase (blue cylinders and loops; PDB ID 1EMH) in complex with uracil (yellow sticks) containing DNA (orange sticks). Adenine opposite the uracil is shown as green sticks. (B) Close-up of the active site. Protein side chains are represented as blue sticks. DNA follows the same color scheme as (A). Note the base specific contacts (dashed lines) made by Asn204 and His268 and Leu272 that plugs the gap in the duplex left by the flipped out base. (Adapted from (Parikh, Walcher et al. 2000)).

UDG family 1 contains UDG/UNG and is defined by the landmark structures of the human and viral enzymes in various states, which revealed mechanistic details about substrate recognition and catalysis common to the entire superfamily (Mol, Arvai et al. 1995; Mol, Kuo et al. 1995; Savva, McAuley-Hecht et al. 1995; Slupphaug, Mol et al. 1996). Family 2 is composed of thymine-specific TDG and MUG, which are homologous to UNG in structure but not sequence (Barrett, Savva et al. 1998; Barrett, Savva et al. 1998; Barrett, Scharer et al. 1999; Maiti, Morgan et al. 2008). The third family is defined by SMUG, and the fourth by *Thermus thermophilus* TDG.

1.7.1 Human Uracil DNA Glycosylase

UNG has served as a model for understanding the structural and biochemical functions of DNA glycosylases in general, and recent work has focused on the mechanism by which the enzyme locates uracil amidst undamaged DNA and maintains specificity for uracil.

This collective body of work on UNG has been the subject of several recent reviews (Krokan, Drablos et al. 2002; Fromme, Banerjee et al. 2004; Huffman, Sundheim et al. 2005; O'Brien 2006; Stivers 2008; Friedman and Stivers 2010; Zharkov, Mechetin et al. 2010). The common structural fold of UNG enzymes consists of four β-sheets sandwiched between four α-helices (Mol, Arvai et al. 1995; Zharkov, Mechetin et al. 2010) (Fig. 1.8A). Unlike HhH glycosylases, the overall structure is not divided into separate domains and the secondary structure elements combine to form a shallow DNA binding cleft that contains the uracil-binding pocket (Fig. 1.7A). (Parikh, Mol et al. 1998; Parikh, Walcher et al. 2000). Although there is a modest conformational change in the protein upon DNA binding, the structure of UNG in complex with DNA shows a

dramatic shift in DNA structure, especially around the uracil (Parikh, Mol et al. 1998; Parikh, Walcher et al. 2000). Consistent with other DNA glycosylase structures, the DNA is kinked ~45° with the uracil moiety flipped out of the duplex. The void left by the extruded base is filled by Leu272 in the minor groove (Fig. 1.8B). Interestingly mutating this leucine to alanine severely reduced the catalytic efficiency of the enzyme. This suggests that DNA interacting residues away from the active site are important in flipping and positioning the target base in the active site (Parikh, Mol et al. 1998). UNG derives its extraordinary specificity for a uracil through an active site that is exquisitely sculpted to accommodate uracil with several base-specific contacts and exclude purines, thymine and other substituted pyrimidines (Fig. 1.8B) (Parikh, Mol et al. 1998; Parikh, Walcher et al. 2000). A recent crystal structure of UNG in complex with DNA containing a T•methylindole base pair provides insights into the overall mechanism of uracil search and capture (Parker, Bianchet et al. 2007). Methylindole is a chemical variant of adenine that cannot hydrogen bond with thymine. Authors used this weakened base pair to trap a complex where the thymine opposite methylindole is unstacked but not as severely as uracil and is placed not in the uracil binding pocket but in an 'exo site'. This complex is thought resemble an early recognition complex where the enzyme can verify the identity of the base before it is flipped into the active site. A comparable 'exo site' has also been reported for hOGG1 (Banerjee, Yang et al. 2005) (see section 1.6.1.1).

1.8 Locating Damaged Bases

DNA glycosylases are charged with a 'needle in a haystack' problem where they need to locate a relatively small number of damaged bases in a sea of unmodified DNA (Friedberg, Walker et al. 2006). It has been calculated that a DNA glycosylase is 70,000-

times more likely to encounter a normal base compared to a damaged one (Friedman and Stivers 2010). Moreover they also need to be able to discern individual DNA lesions from each other, which most of the time differ only by a few atoms. The quest to understand the mechanism of how these proteins locate and maintain specificity for a particular damage has inspired some ingenious experiments and collaborations between geneticists, biochemists, structural biologists and chemists (Bruner, Norman et al. 2000; Banerjee, Yang et al. 2005; Banerjee, Santos et al. 2006; Radom, Banerjee et al. 2007; Lee, Bowman et al. 2008; Friedman, Majumdar et al. 2009).

One of the modes that DNA binding proteins can utilize to scan the DNA is simple diffusion either along the phosphodiester backbone or major/minor grooves ('sliding'). Indeed it was reported early on that these proteins display a processive mechanism of transfer along the DNA, thus facilitating interrogation through non-specific binding (Lloyd, Hanawalt et al. 1980). However, facilitated diffusion alone has been proven to be uneconomical and unlikely given the number of glycosylases per cell and packing of DNA into chromatin, which may render certain areas of the genome inaccessible to the protein (reviewed in (Zharkov and Grollman 2005)). An alternative mechanism of 'hopping' has been proposed to help explain the overall efficiency of scanning (Halford 2001). Using restriction endonucleases, Steve Halford and colleagues showed that a 3-D rather than a simple one dimensional diffusion would be better equipped to explain strand switching and speed of removal of multiple substrates (Halford 2001; Halford and Szczelkun 2002; Gowers and Halford 2003). It is currently accepted that a combination of facilitated diffusion (sliding) over short range with hopping (3D diffusion) is used to translocate along the DNA in search of damage

(Porecha and Stivers 2008) and [reviewed in (Friedman and Stivers 2010)and (Zharkov and Grollman 2005)].

Once a DNA glycosylase has located a site of damage, the target base then needs to be flipped out of the duplex into an active site pocket. This phenomenon can be understood better as two distinct but correlated processes – (1) bending of the DNA to facilitate the flipped base and (2) physical extrusion of the base from the helix. Recently, structural and biochemical studies have started to offer insights into this fascinating process. More specifically, work on UNG by Jim Stivers and on MutM and OGG1 by Greg Verdine and colleagues have been instrumental in painting a detailed picture of interrogation, recognition and excision by DNA glycosylases (Bruner, Norman et al. 2000; Banerjee, Santos et al. 2006; Blainey, van Oijen et al. 2006; Parker, Bianchet et al. 2007; Porecha and Stivers 2008).

Elegant NMR studies on UNG showed that the spectra of the peptide chain changed upon binding a 10-bp long DNA and that the change was observed around the known DNA binding region including the minor-groove probing Leu272 residue (Friedman, Majumdar et al. 2009). With help of imino proton exchange NMR techniques and high resolution structure of UNG-DNA complexes, it has been suggested that UNG use the intrinsic opening/closing (‘breathing’) of DNA bases to initiate base flipping and formation of the interrogation complex (IC) (Cao, Jiang et al. 2004; Cao, Jiang et al. 2006; Parker, Bianchet et al. 2007). However, in the absence of a UNG-DNA late stage reaction complex structure, the source of high degree of specificity shown by the enzyme toward uracil over thymine and the molecular details of the late stage of substrate verification is still poorly understood.

In addition to high resolution structural studies of MutM and OGG1 discussed earlier, single molecule imaging have lent to a more complete understanding of how these enzymes search for 8oxoG. One such study showed that OGG1 can slide processively on λ-DNA (under flow) for ~400 base pairs (Blainey, van Oijen et al. 2006). In another study the same group was able show that hOgg1 and MutM spin around the DNA as they slide (Blainey, Luo et al. 2009). By using proteins tagged with streptavidin to increase the radius of protein 'sliding' along the DNA, the same group was able to show that the rate of sliding follows a $1/r^3$ ratio (r=protein radius), thus suggesting that these proteins translocate along the minor groove, major groove or the phosphodiester backbone (Blainey, Luo et al. 2009).

SCOPE OF THIS WORK

Work presented in this dissertation details structural and biochemical characterization of three yeast alkylpurine DNA glycosylases – Mag1 and Mag2 from *S. pombe* and Mag from *S. cerevisiae*. Chapter II presents a crystal structure of spMag1 bound to DNA containing an abasic site. The structure is then used to parse out the basis of substrate preference in alkylpurine DNA glycosylases with help of extensive mutational biochemistry. Data presented in this chapter provide novel evidence that protein-DNA interactions away from the active site can modulate the specificity of DNA glycosylases and a single substitution in the minor-groove interrogating loop is sufficient to alter the preference of one alkylpurine glycosylase for εA to that of a homologous alkylpurine glycosylase. Chapter III addresses the lack of base excision activity in another alkylpurine DNA glycosylase-like protein from *S. pombe* (spMag2) despite almost 70% sequence similarity with Mag1. The crystal structure of spMag2 in complex with DNA is presented along with phylogenetic analysis of the Mag genes in *S. pombe*, *S. cerevisiae* and other related fungal species and biochemical experiments. Together these studies provide evidence that spMag2 cannot form a catalytically competent complex with DNA due to differences in the minor-groove interrogating loop and overall electrostatic surface potential compared to spMag1. Phylogenetic analysis also suggests that spMag2 may have evolved to carry out a yet uncharacterized function. Chapter IV is a structural and functional study of scMag and investigates the role of the extra N-terminal domain and an internal loop seen in scMag compared to spMag1 and other related DNA glycosylases. In addition to confirming the results seen for spMag1, the chapter provides evidence that the N-terminal extension seen in scMag is important in

maintaining structural integrity of the protein. Chapter V discusses the results and findings presented in chapters II-IV in the context of the broader DNA repair field.

Chapter 2

ALTERING SUBSTRATE SPECIFICITY OF *SCHIZOSACCHAROMYCES POMBE* ALKYLPURINE DNA GLYCOSYLASE MAG1*

Abstract

DNA glycosylases specialized for alkylation damage must identify, with exquisite specificity, a diverse array of subtle modifications within DNA. The current mechanism involves damage sensing through interrogation of the DNA duplex followed by more specific recognition of the target base inside the active site pocket. To better understand the physical basis for alkylpurine detection, we determined the crystal structure of *S. pombe* Mag1 in complex with DNA and performed a mutational analysis of spMag1 and the close homolog from *S. cerevisiae* Mag. Despite strong homology, spMag1 and scMag differ in substrate specificity and cellular alkylation sensitivity, although the enzymological basis for their functional differences is unknown. We show that Mag preference for 1,N^6-ethenoadenine (εA) is influenced by a minor groove interrogating residue more than the composition of the nucleobase binding pocket. Exchanging this residue between Mag proteins swapped their εA activities, providing evidence that residues outside of the extrahelical base binding pocket play a role in identification of a particular modification in addition to sensing damage.

*The work presented in this chapter was published in **Adhikary, S.** and B. F. Eichman (2011). "Analysis of substrate specificity of Schizosaccharomyces pombe Mag1 alkylpurine DNA glycosylase." EMBO Rep **12**(12): 1286-1292.

2.1 Introduction

DNA is susceptible to alkylation damage from environmental toxins and from endogenous lipid peroxidation products and methyl donors (Friedberg, Walker et al. 2006). These agents produce a chemically diverse array of detrimental alkylated nucleobases that threaten genome integrity by causing mutations, DNA replication arrest, and single- and double-strand breaks (Barnes and Lindahl 2004). The toxic effects of alkylating agents are the rationale for their use in cancer chemotherapy, while the mutagenic potential of DNA alkylation damage leads to genomic instability and increases cancer risk. Alkylated DNA bases account for ~23 percent of nucleobase damage in the genome (Friedberg, Walker et al. 2006) and have been detected in humans and rats after exposure to various carcinogens (Shuker, Bailey et al. 1987; Holt, Yen et al. 1998).

A large number of toxic and mutagenic alkylpurines, including 3-methyladenine (3mA), 7-methylguanine (7mG), and εA (Fig. 2.1), are eliminated by DNA glycosylases, which initiate the base excision repair pathway by locating the modified bases and catalyzing the hydrolysis of the *N*-glycosidic bond. DNA glycosylases specialized for alkylpurine lesions are found in all organisms and exhibit an exceptionally broad substrate range. For example, in *E. coli*, the constitutively active TAG enzyme is highly specific for cytotoxic 3mA lesions, while alkylation damage inducible AlkA recognizes a wide range of mutagenic substrates, including εA [reviewed in (Rubinson, Adhikary et al. 2010)]. Similarly, the human AAG enzyme is a functional counterpart to AlkA and exhibits a robust activity toward etheno and oxidized DNA adducts (Saparbaev, Kleibl et al. 1995). Structural studies have illustrated how these enzymes utilize a common base flipping mechanism to gain access to the lesion inside an active site pocket on the surface

of the enzyme (Rubinson, Adhikary et al. 2010). In all cases, the extrahelical DNA conformation is stabilized by surface residues that intercalate into the DNA base stack and plug the gap left by the flipped base.

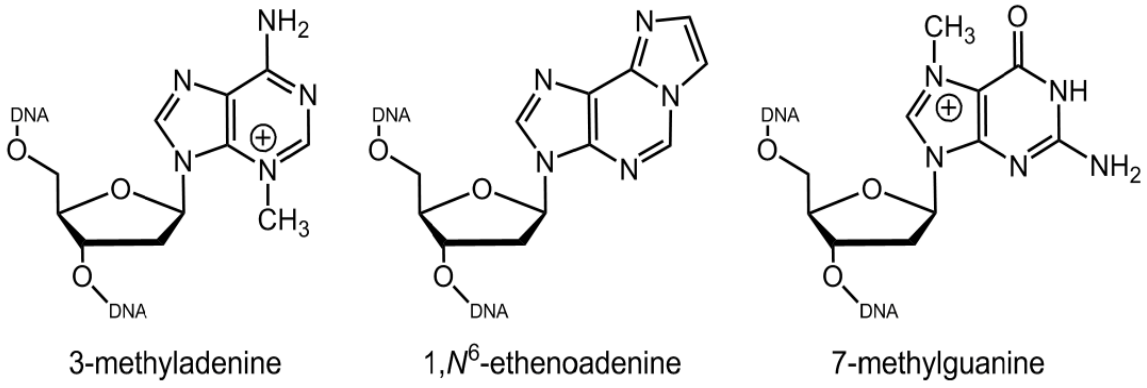


Fig. 2.1 Structures of alkylated bases relevant to this study.

The molecular basis for alkylpurine discrimination remains poorly understood, but it is believed to be a consequence of shape and chemical complementarity between the extrahelical nucleobase substrate and the active site pocket (Lau, Wyatt et al. 2000; Eichman, O'Rourke et al. 2003; Metz, Hollis et al. 2007). Recent work, however, has revealed that some DNA glycosylases use the DNA plug residues as damage sensors by interrogating undamaged DNA prior to base flipping (Banerjee, Santos et al. 2006; Qi, Spong et al. 2009), suggesting that these interrogating residues may also be important for selection of a particular substrate. In addition, the specific catalytic mechanism of base excision and the thermodynamic stability of the lesion have been shown to influence the choice of substrates (Parikh, Walcher et al. 2000; Stivers 2004; Rubinson, Adhikary et al. 2010).

In an attempt to understand the molecular basis for the selection of alkylation damage in particular, we carried out a structure-function analysis of two closely related yeast alkylpurine DNA glycosylases. Despite their extensive sequence homology, scMag and spMag1 have different DNA repair phenotypes and substrate preferences analogous to AlkA and TAG, respectively. Like AlkA, scMag is induced by exposure to DNA damaging agents, exhibits a strong mutator phenotype when overexpressed, and excises a broad spectrum of alkylpurines, including εA (Chen, Derfler et al. 1990; Chen and Samson 1991; Saparbaev, Kleibl et al. 1995; Lingaraju, Kartalou et al. 2008). SpMag1, on the other hand, is constitutively expressed, has a much weaker mutator phenotype, and has a restricted substrate preference (Memisoglu and Samson 1996; Memisoglu and Samson 2000). Specifically, spMag1 has been reported to lack εA excision activity (Alseth, Osman et al. 2005).

Here, we report the crystal structure of spMag1 bound to DNA, together with a mutational analysis of εA and 7mG excision, which enabled identification of the residues responsible for the yeast Mag substrate specificity differences. SpMag1 contains a unique histidine that contacts the minor groove outside of the nucleobase binding pocket. Substitution of this histidine with the corresponding serine residue in other Mag homologs resulted in an exchange of their relative εA activities while not severely affecting 7mG activity. Surprisingly, mutation of residues in the extrahelical nucleobase binding pockets had no effect on substrate specificity, challenging the previous notion that substrate recognition is based solely on steric exclusion of a lesion from the active site pocket. These data provide evidence for how DNA glycosylases discriminate among

different types of damage outside of the active site, and suggest that alkylpurine selection may begin prior to base flipping.

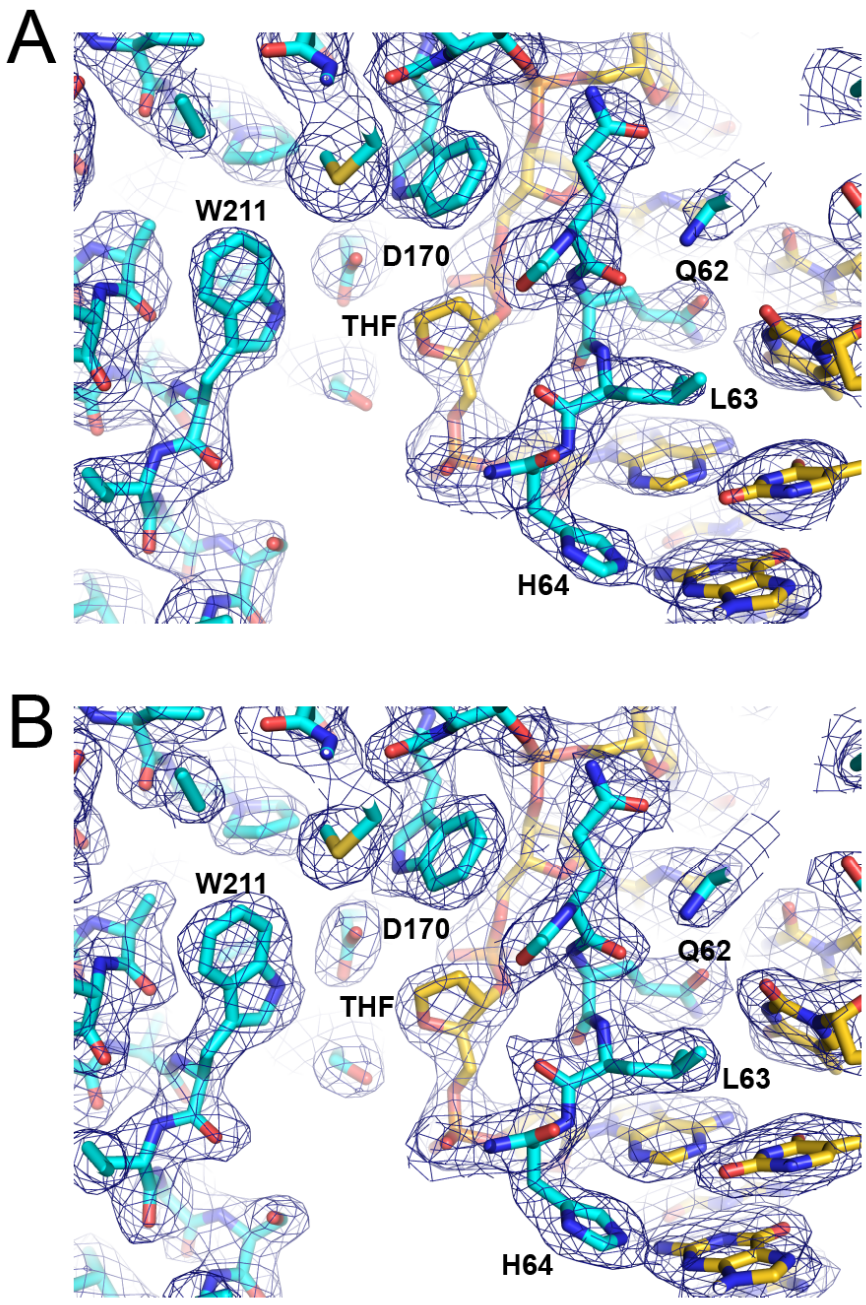


Fig. 2.2 SpMag1-DNA structural data. A cross-section of the final protein-DNA model is shown superimposed onto experimental (solvent-flattened) SAD (A) and refined 2Fo-Fc (B) electron density contoured at 1.5σ. Protein and DNA carbon atoms are colored cyan and gold, respectively.

2.2 Results and Discussion

2.2.1 Structure of the Mag1-DNA complex

Mag1 from *S. pombe* (spMag1) was crystallized in complex with DNA containing a tetrahydrofuran (THF) abasic site analog, and the structure determined by single-wavelength anomalous dispersion (SAD) from selenomethionine (SeMet)-substituted spMag1-DNA crystals (Fig. 2.2). The resulting crystallographic model consisting of two spMag1-DNA complexes in the asymmetric unit was refined against 2.2 Å native diffraction data (Table 2) to a crystallographic residual of 18.5% (R_{free} = 22.5%). The overall structure and DNA binding mode of spMag1 is consistent with the helix-hairpin-helix (HhH) superfamily of DNA glycosylases (Huffman, Sundheim et al. 2005; Rubinson, Adhikary et al. 2010). Two α-helical subdomains pack together to form the extrahelical nucleobase binding cleft at their interface (Fig. 2.3). The DNA is anchored to the protein from the minor groove side primarily through electrostatic interactions between the HhH domain (helices αC-αJ) and the phosphate backbones of both strands (Fig. 2.3A,B). The HhH motif (helices αI-αJ) binds the phosphate backbone of the damaged strand immediately downstream from the THF abasic site (Appendix, Fig. A1), while helices αF-αG engage the strand opposite the lesion (Fig. 2.3C). The damaged strand is buried in the cleft between the two domains with the THF abasic site fully rotated 180° around the backbone into the extrahelical base binding pocket. Importantly, the αC-αD loop intercalates into the duplex at the damage site, resulting in a 70° kink in the DNA. The arms of the duplex are primarily B-form DNA and are swung away from the protein toward the major groove (Fig. 2.3B). The DNA binding mode of spMag1 is similar to that observed in the structure of AlkA bound to DNA containing a 1-azaribose

transition state analog (Fig. 2.3D and A3) (Hollis, Ichikawa et al. 2000). Despite the lack of the N-terminal β-sheet extension present in AlkA, the spMag1 structure is highly similar to AlkA residues 89-282 with an r.m.s.d. of 1.45 Å for main chain atoms (Fig. 2.4 and A2). Superposition of the two proteins results in a remarkable agreement in positions of the damaged DNA strands and the 1-azaribose and THF moieties. The only noticeable difference in the two DNA complexes is the trajectory of the duplex arms (Fig. A3). With the structures of spMag1 and AlkA in hand, we were able to pinpoint putative active site and DNA binding residues in both spMag1 and scMag for the purposes of explaining substrate specificity differences in the yeast proteins.

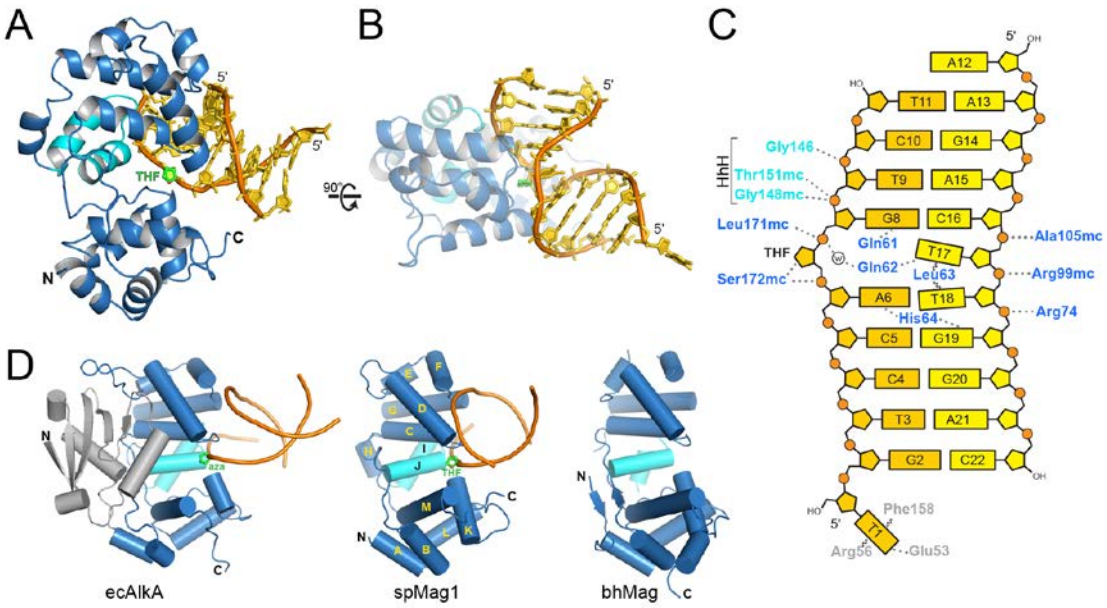


Fig. 2.3 The spMag1-DNA complex crystal structure. (A,B) Orthogonal views of spMag1 (blue ribbons) bound to DNA (gold) containing a THF abasic site analog (green). The HhH motif is light blue. (C) Schematic of spMag1-DNA interactions, with protein residues in blue, THF-DNA strand in gold, undamaged DNA strand in yellow, and phosphates depicted as orange circles. Dotted and wavy lines represent hydrogen bonds and van der Waals interactions, respectively. (D) Structural alignment of 3-methyladenine DNA glycosylases: *E. coli* AlkA bound to 1-azaribose (aza) DNA (PDB ID 1DIZ), *S. pombe* Mag1, and *B. halodurans* Mag (PDB ID 2H56).

Table 2. Crystallographic data collection and refinement statistics

	Native	SeMet
Data collection*		
Wavelength (Å)	0.9787	0.9792
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	54.3, 112.3, 61.1	54.1, 111.3, 61.2
α, β, γ (°)	90, 99.6, 90	90, 99.1, 90
Resolution (Å)	29.10-2.28 (2.38-2.28)	50.00-2.80 (2.90-2.80)
R_{sym}	0.109 (0.404)	0.103 (0.326)
I / σ_I	11.50 (2.50)	12.60 (3.20)
Completeness (%)	99.6 (98.1)	98.9 (93.0)
Redundancy	4.8 (3.4)	3.8 (3.0)
Refinement		
Resolution	2.28	
No. reflections	32,188	
R_{work} / R_{free}	0.185 / 0.225	
No. of atoms		
Protein	3184	
DNA	845	
Solvent	114	
B-factors (Å^2)		
Protein	38.3	
DNA	56.2	
Solvent	40.4	
R.m.s. deviations		
Bond lengths (Å)	0.008	
Bond angles (°)	1.185	

*Values in parentheses refer to the highest resolution shell.

2.2.2 Base excision activity

A DALI search against the Protein Data Bank revealed spMag1 to be most similar to an unpublished structure of a putative Mag ortholog from *Bacillus halodurans* (bhMag) (PDB ID 2H56). SpMag1 and bhMag superimpose with an r.m.s.d. of 1.42 Å for main chain atoms and share 27% sequence identity and 65% overall similarity. The high sequence and structural similarity among spMag1, scMag, and bhMag prompted us to compare their base excision activities in order to understand Mag functional differences (Table 3).

Table 3. Sequence and structural similarity between Mag orthologs

		Identity (%)	Similarity (%)	RMSD (Å)	Q-score	PDB ID
spMag1	bhMag	26.5	65.2	1.78	0.56	2H56
spMag1	scMag	22.2	48.5			
spMag1	ecAlkA	17.4	46.1	1.81	0.37	1MPG
scMag	bhMag	22.0	49.3			
scMag	ecAlkA	19.9	41.7			
bhMag	ecAlkA	17.0	40.7			

Sequence identity and overall similarity were obtained by CLUSTALW from the PBIL server (http://npsa-pbil.ibcp.fr). RMSD and Q-scores for Cα atoms only were calculated using the PDBeFold server at EMBL-EBI (http://www.ebi.ac.uk). Q-score (Quality of Alignment) is defined by $Q=(N_{align}^2)/[(1+(RMSD/3Å)^2)*N_{res1}*N_{res2}]$, where N_{align} is the number of aligned residues and N_{res} is the number of total residues (http://www.ebi.ac.uk/msd-srv/ssm/rl_qscore.html). Abbreviations: sp, *Schizosaccharomyces pombe*; sc, *Saccharomyces cerevisiae*; bh, *Bacillus halodurans*; ec, *Escherichia coli*.

We measured the single-turnover kinetics of εA and 7mG excision from oligonucleotides containing a single lesion. Under the conditions of our assay, all three enzymes removed 7mG at equal rates (Table 5), whereas their activities toward εA

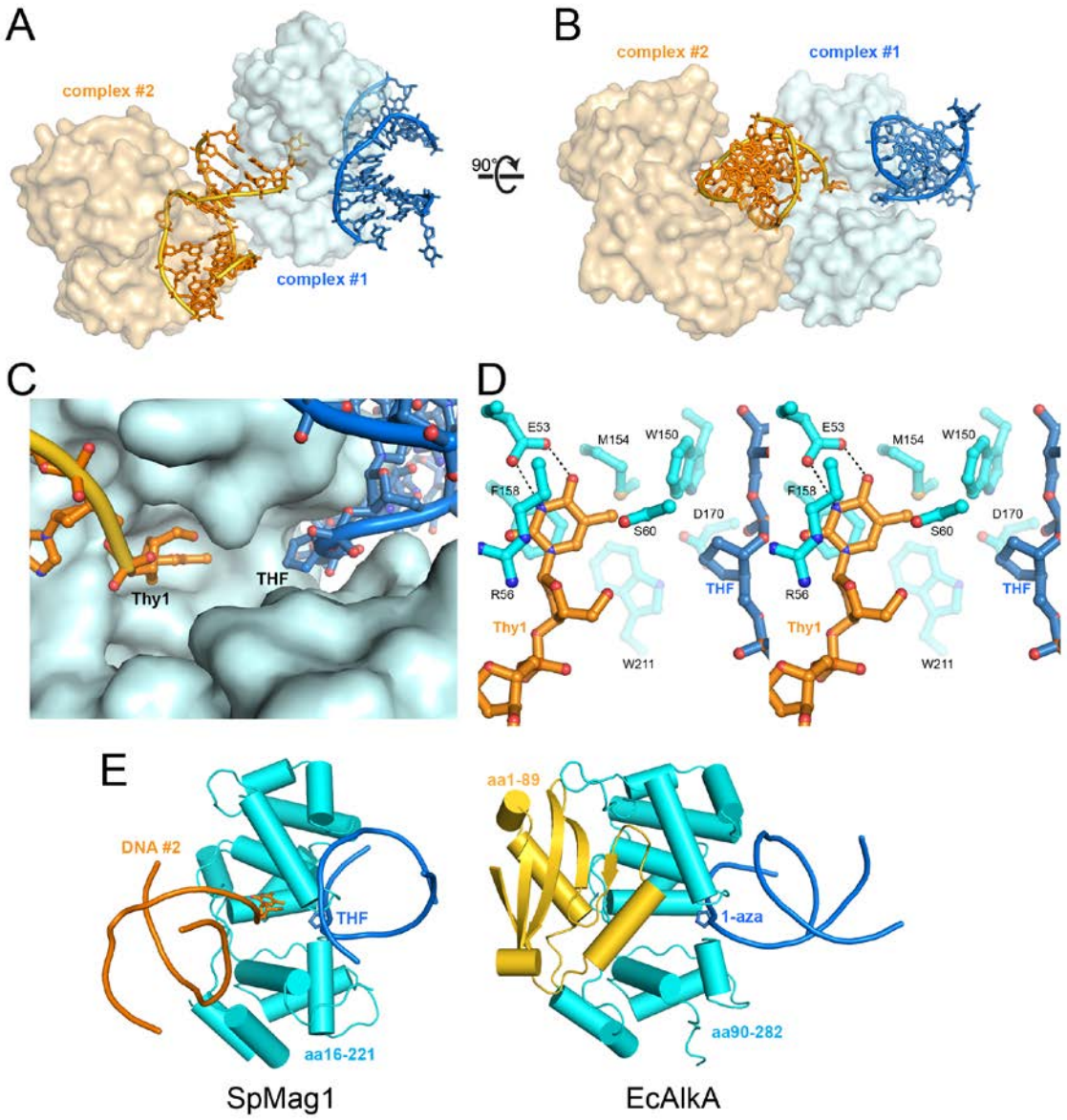


Fig. 2.4 Packing of spMag1-DNA crystals. (A) Two spMag1-DNA complexes in the asymmetric unit are colored blue and orange. Protein molecules are shown as van der Waals surfaces and DNA is shown as sticks. (B) A 90° rotation along the vertical axis of the view shown in A. (C) Close-up view of the base binding pocket of spMag1 from complex #1 (blue), with Thy1 nucleotide from complex #2 (orange) inserted into the cavity. (D) Stereo view of the Thy1 interactions in the active site. Protein and DNA residues from complex #1 are colored cyan and blue, and the DNA from complex #2 is orange. (E) The position of the invading DNA (orange) from the adjacent molecule in spMag1 is in the same location as the N-terminal β-sheet domain in AlkA (gold).

differed (Fig. 2.5). Contrary to a previous report (Alseth, Osman et al. 2005), we found that spMag1 can indeed excise εA at a low level of activity (Fig. 2.5A). This discrepancy is most likely due to the specific conditions or DNA sequence used to test activity. Under the same conditions, scMag removed εA at a rate 2-3-fold greater than spMag1—a modest but significant difference (Fig. 2.5B) (Table 5). The rate constant of 12.8 ± 1.7 x 10^{-5} sec^{-1} for the scMag/εA-DNA reaction is consistent with values previously reported for scMag and AlkA (O'Brien and Ellenberger 2004; Lingaraju, Kartalou et al. 2008). Surprisingly, bhMag excised εA at a rate comparable to scMag despite its stronger similarity to spMag1 (Fig. 2.5C-D; Table 5).

2.2.3 The nucleobase binding pocket

To reconcile spMag1's low εA excision activity relative to its close homologs, we first compared the structural details of the base binding pockets since preference for a particular substrate is determined in large part by its fit within the active site (Eichman, O'Rourke et al. 2003). Consistent with our biochemical results, the spMag1 binding pocket is comparable in size to that of AlkA and can easily accommodate an εA base (Fig. 2.6A,D). In fact, although we crystallized Mag1 in complex with an abasic site, we observed a nucleobase in this pocket from insertion of the 5′-terminal thymine base from an adjacent DNA molecule in the crystal through the large opening at the rear of the active site (Fig. 2.4)

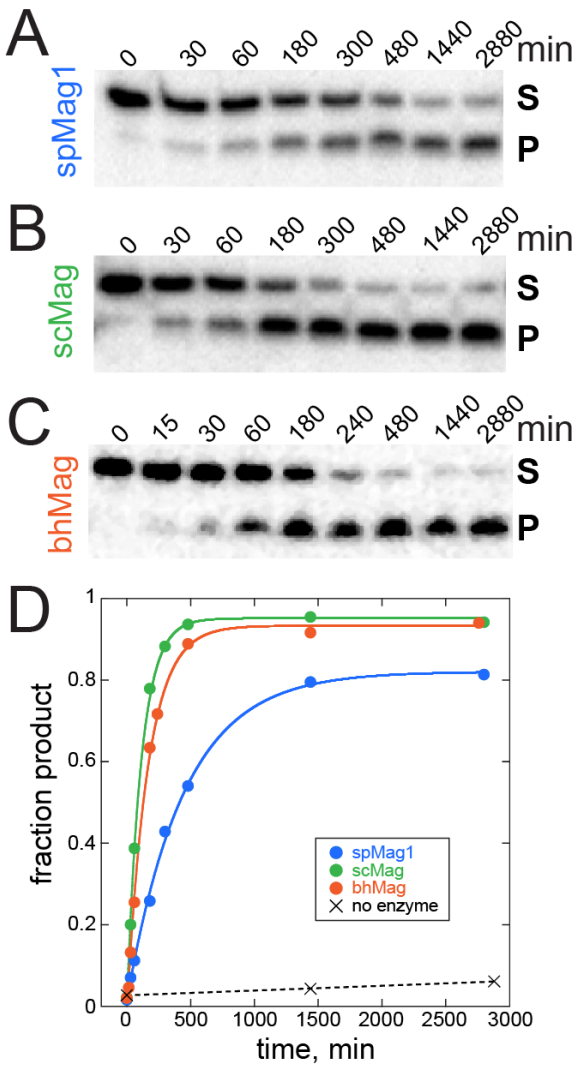


Fig. 2.5 1,N^6-ethenoadenine (εA) excision activity of Mag orthologs. Denaturing polyacrylamide gels showing the disappearance of εA-containing DNA substrate (S) and appearance of alkaline-cleaved abasic-DNA product (P) as a function of time after addition of spMag1(A), scMag (B) and bhMag (C). (D) Quantitation of the panels shown in A-C. Blue, spMag1; green, scMag; orange, bhMag. Rate constants calculated from the single-exponential fits to the data are shown in Table 5.

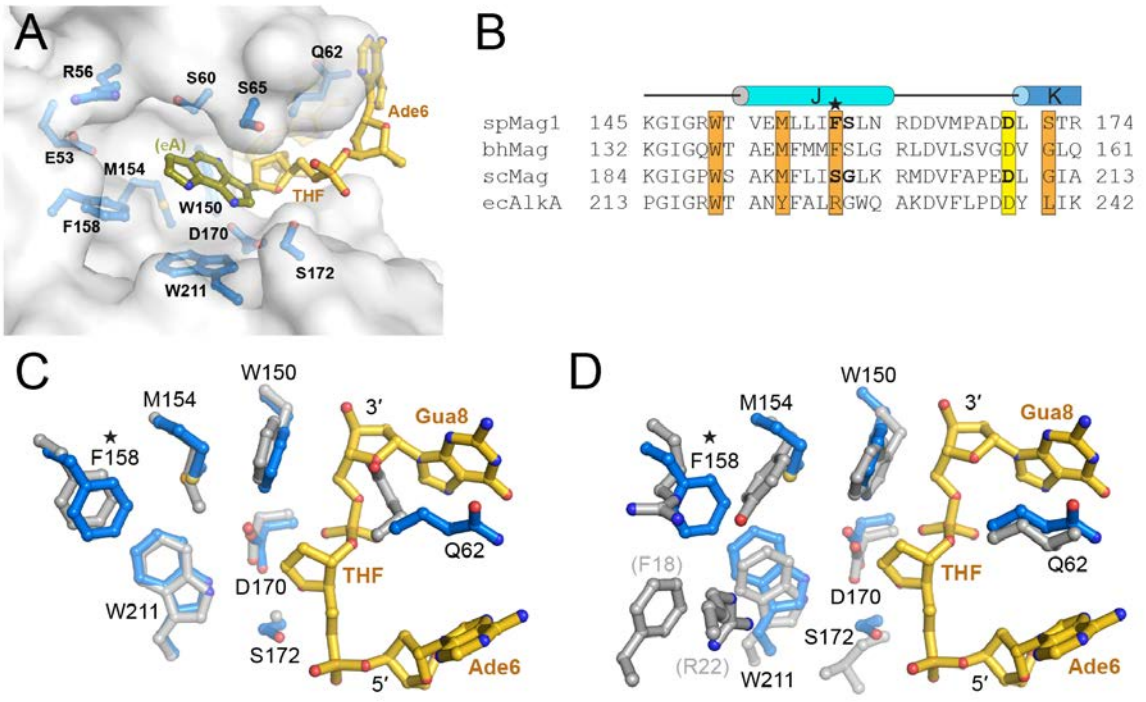


Fig. 2.6 SpMag1 nucleobase binding pocket. (A) The base binding pocket is shown as a transparent van der Waals surface. Protein side-chains likely to contact an extrahelical base are blue and the THF abasic site and the two flanking nucleotides are shown in gold. The εA base (olive) is modeled within the pocket from coordinates of εA-DNA bound to human AAG (PDB ID 1F4R). (B) Sequence alignment of spMag1, bhMag and ecAlkA structures with scMag sequence. Residues predicted to contact the substrate base are highlighted orange, and the conserved HhH aspartic acid is yellow. The nonconserved Phe158 is marked with a star. (C,D) Superposition of spMag1 (blue side chains, gold DNA) onto bhMag (C) and *E. coli* AlkA (D) (grey).

.

This is a fortuitous lattice contact likely irrelevant to spMag1 function given that spMag1 does not have a binding preference for 5′-overhangs (Table 4) and that this opening is normally occluded by the β-sheet domain in AlkA and presumably scMag. Nonetheless, it illustrates that the low activity toward εA is not a result of steric exclusion from the spMag1 active site.

There is remarkable agreement between the base binding residues in the Mag enzymes and AlkA (Fig. 2.6). Phe158 in spMag1 is the only non-conserved residue

predicted to contact the extrahelical nucleobase and is the most significant difference between spMag1 and scMag sequences (Fig. 2.6B) and the AlkA active site (Fig. 2.6D). We therefore tested the contribution of Phe158 and the spatially adjacent Ser159 to εA excision activity by swapping the corresponding residues between spMag1 and scMag. Neither spMag1 F158S,S159G nor scMag S197F,G198S double mutant affected the εA or 7mG excision activity relative to wild-type (Table 5, Fig. 2.6D). Thus, the substrate specificity differences between spMag1 and scMag cannot be explained by the differences in residues contacting the extrahelical base.

Table 4. DNA binding by spMag1

	K_d (μM)		**DNA sequences**
	Wild-type	H64S D170N	
THF	0.8 ± 0.1	0.8 ± 0.04	5′ TGACTACTACATG**X**TTGCCTACCAT 3′*ACTGATGATGTACCAACGGATGGTA
εA	0.6 ± 0.1	0.8 ± 0.1	5′ TGACTACTACATG**X**TTGCCTACCAT 3′*ACTGATGATGTACCAACGGATGGTA
Gua	1.3 ± 0.3		5′ TGACTACTACATG**G**TTGCCTACCAT 3′*ACTGATGATGTACCAACGGATGGTA
Gua, 5′-oh	0.9 ± 0.1		5′TTGACTACTACATG**G**TTGCCTACCA 3′*ACTGATGATGTACCAACGGATGGTA

Dissociation constants (K_d) were determined by fluorescence anisotropy changes upon adding protein to 6-carboxyfluorescein(*)-DNA as described in the Supplementary Methods. Values shown are averages ± standard deviations from three independent measurements.

In addition to steric exclusion, greater specificity toward 3mA and 7mG lesions may be influenced by the relatively weak catalytic power of some alkylpurine DNA glycosylases, since these bases have destabilized *N*-glycosidic bonds as a result of their

formal positive charges and therefore require minimal rate enhancement for removal over their spontaneous rate of depurination (Stivers and Jiang 2003; Rubinson, Gowda et al. 2010). Most monofunctional DNA glycosylases contain a catalytically essential, conserved aspartate at the mouth of the nucleobase binding pocket (Labahn, Scharer et al. 1996).

Whereas substitution of scMag Asp209 with asparagine reduced activity toward both substrates to less than 1% of the wild-type enzyme, spMag1 D170N retained 20% and 5% activity toward εA and 7mG, respectively (Table 5). Interestingly, this residual activity in the spMag1 aspartate mutant was also observed in 3mA-specific MagIII (Eichman, O'Rourke et al. 2003). The weaker catalytic potential of Asp170 in spMag1 could potentially be influenced by a polar interaction with Ser172, which is a glycine in scMag and bhMag (Fig. 2.6B-C).

Substitution of spMag1 Ser172 to glycine decreased εA activity 10-fold, whereas the corresponding Gly→Ser substitution in scMag did not have a significant effect (Table 5). Taken together, this data suggests that spMag1 and scMag may have subtle mechanistic differences in catalytic potential that could affect their ability to excise more stable alkylpurines.

Table 5. Base excision activity of wild-type and mutant Mag orthologs

	εA		7mG	
	k_{cat} (x10^{-5} sec^{-1})	Relative activity	k_{cat} (x10^{-3} sec^{-1})	Relative activity
spMag1				
WT	5.3 ± 0.6	1.0	2.6 ± 0.5	1.0
Q62A	0.6 ± 0.05	0.1	0.4 ± 0.09	0.2
L63A	0.01 ± 0.002	0.003	0.02 ± 0.003	0.01
H64S	21.1 ± 3.2	4.0	2.1 ± 0.4	0.8
FS158SG	5.7 ± 1.1	1.1	3.8 ± 0.6	1.5
D170N	1.1 ± 0.2	0.2	0.09 ± 0.02	0.04
S172G	0.7 ± 0.2	0.1		
scMag				
WT	12.8 ± 1.7	1.0	3.6 ± 0.7	1.0
S97H	2.0 ± 0.3	0.2	1.8 ± 0.4	0.5
SG197FS	14.2 ± 2.1	1.1	3.5 ± 0.7	1.0
D209N	0.1 ± 0.01	0.005	0.01 ± 0.002	0.003
G211S	8.3 ± 0.9	0.6		
bhMag				
WT	9.3 ± 0.4	1.0	3.4 ± 0.7	1.0
S53H	4.5 ± 0.7	0.5	1.7 ± 0.3	0.5

Single-turnover rate constants (k_{cat}) for excision of εA and 7mG opposite cytosine from a 25mer oligonucleotide were measured under saturating enzyme concentrations at pH 6.0 (εA) or 7.5 (7mG), 150 mM ionic strength, and 25°C. Values represent the average from three independent measurements ± standard deviations. Non-enzymatic rate constants for spontaneous depurination (k_{non}) under the same conditions were 7.2 x 10^{-8} sec^{-1} (εA) and 1.6 x 10^{-6} sec^{-1} (7mG).

2.2.4 Minor groove interactions are important for substrate specificity

Outside of the base binding pocket, side chains that intercalate into the DNA base stack serve two functions: to interrogate the DNA duplex prior to base flipping and to

stabilize the extrahelical base after flipping, and are thus essential for glycosylase activity (Stivers 2004; Banerjee, Santos et al. 2006; Qi, Spong et al. 2009). In the HhH glycosylases, these interrogating residues are located at the tip of the αC-αD loop (Fig. 2.7A-B). In spMag1, the extruded damaged strand and large kink in the duplex is stabilized by Gln62, which plugs the gap in the damaged strand, and Leu63, which wedges itself between the bases opposite the lesion (Fig. 2.7C). Both of these residues are conserved in scMag and bhMag (Fig. 2.7A). Substitution of spMag1 Gln62 and Leu63 with alanine resulted in an 80-90% and <99% decrease, respectively, in base excision activity for both εA and 7mG (Table 5), consistent with their importance to base excision.

SpMag1 contains an additional minor groove interaction adjacent to the plug and wedge residues that is not conserved in scMag or AlkA (Fig. 2.7A). The imidazole ring of His64 is positioned to form a hydrogen bond with either the N3 nitrogen of the adenine immediately 5′ to the lesion or the N3 nitrogen of Gua19 on the opposite strand, depending on the histidine conformer (Fig. 2.7C). This residue is not conserved among the other Mag enzymes or AlkA, which all have a serine in the same position (Fig. 2.7A). In an attempt to alter Mag specificity for εA, we changed spMag1 His64 to serine, and scMag Ser97 and bhMag Ser53 to histidine and measured their activities toward both εA and 7mG (Table 5, Fig. 2.7). Interestingly, the spMag1 H64S mutation *increased* the εA excision activity 3-fold relative to the wild-type, bringing the activity up to a level similar to scMag, but did not affect 7mG activity (Fig. 2.7D). In contrast, scMag S97H and bhMag S53H *decreased* their εA excision activities down to spMag1 levels (Fig. 2.7D), and had only marginal effect (< 2-fold) on 7mG excision activity (Table 5). We therefore

conclude that the minor groove interaction at this position in spMag1 plays a significant role in defining the substrate preference among the Mag enzymes.

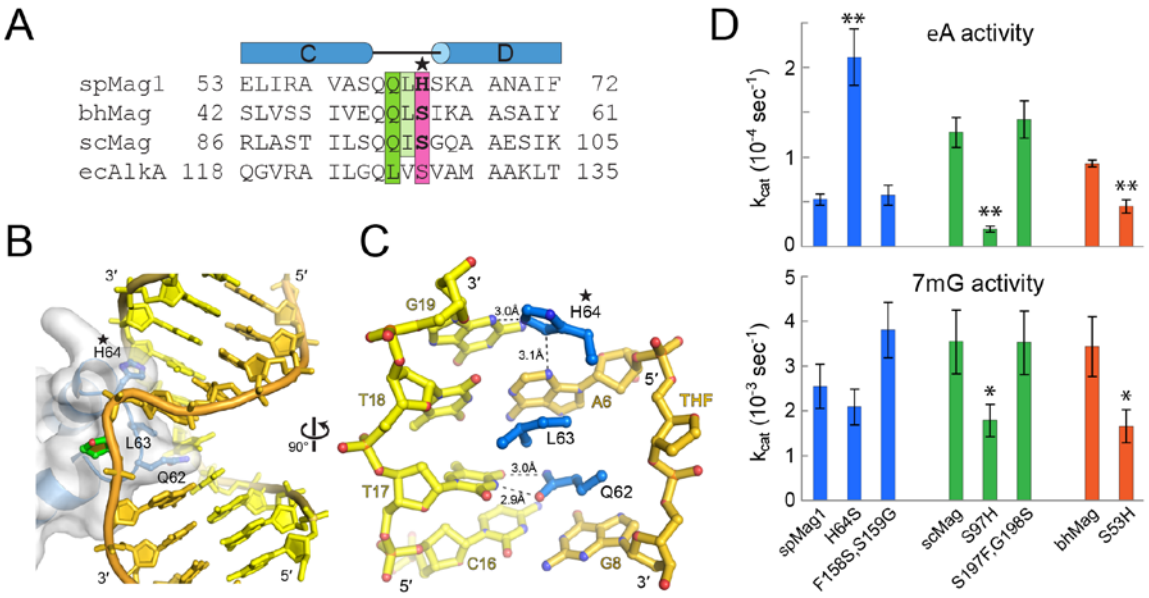


Fig. 2.7 DNA interrogation by spMag1. (A) Structure-based sequence alignment of the αC-αD loops from Mag and AlkA enzymes. Plug/wedge residues are highlighted green/light green, and the novel His64 contact in spMag1 is magenta. (B) The spMag1 αC-αD loop is shown in blue within a transparent molecular surface and intercalating residues rendered as sticks. The damaged DNA strand is colored gold and THF green. (C) Close-up of spMag1-DNA contacts at the lesion. (D) Single-turnover rates (k_{cat}) of εA (top) and 7mG (bottom) excision are plotted for wild-type and mutant spMag1 (blue), scMag (green), and bhMag (orange). Values are shown in Table 5 and raw data is shown in Fig. A5. Error bars represent the standard deviation (n=3). Asterisks denote p<0.05 (*) and p<0.002 (**).

Mutation of the minor groove intercalating residues in other glycosylases has been shown to abrogate base excision activity in a number of glycosylases (Jiang, Kwon et al. 2001; Vallur, Feller et al. 2002; Eichman, O'Rourke et al. 2003; Maiti, Morgan et al. 2009). Recent work by Verdine and colleagues has illustrated that these residues in MutM and AlkA make intimate contacts with undamaged DNA and thus likely act as sensors to distinguish damaged versus undamaged DNA (Banerjee, Santos et al. 2006;

Qi, Spong et al. 2009; Bowman, Lee et al. 2010). The effect of spMag1 His64 extends these results by demonstrating that probe residues are also capable of discriminating between particular types of damage. One possible mechanism for this is that the side chain at this position senses a local perturbation in the duplex prior to base flipping. *N*3-substituted purines would be identified directly in the minor groove, while etheno adducts and *N*7-substitutions could be sensed by a perturbation in base pair structure or stability. The general loss of activity from the S→H substitutions in scMag and bhMag, but not spMag1, suggest that there may be subtly different modes of detection of 7mG and εA lesions. Our crystal structure, which represents the product of the reaction, does not rule out the possibility that the enzyme-substrate complexes may differ between spMag1 and scMag or AlkA, or that His64 may act as an inhibitor by reducing the scanning rate as a result of the hydrogen bonds with the minor groove. Nonetheless, the structure and supporting biochemistry of base excision highlights the importance of residues outside of the base binding pocket in the lesion recognition process.

The enzymological differences between the yeast Mag enzymes certainly play an important role in alkylation resistance in cells since protein expression has been shown to complement the alkylation sensitivity of *tag alka E. coli* with different levels of effectiveness (Chen, Derfler et al. 1990; Chen and Samson 1991; Saparbaev, Kleibl et al. 1995; Memisoglu and Samson 1996; Memisoglu and Samson 2000; Alseth, Osman et al. 2005; Lingaraju, Kartalou et al. 2008). The reduced dependency of *S. pombe* to spMag1 in alkylation repair in cells can also be partially explained by specific cellular responses to alkylation damage in *S. pombe* and *S. cerevisiae* aside from glycosylase activity (Memisoglu and Samson 2000). In addition to BER, nucleotide excision and

recombination repair play significant roles in safeguarding *S. pombe* from alkylation damage (Memisoglu and Samson 2000; Alseth, Osman et al. 2005; Kanamitsu, Tanihigashi et al. 2007). Furthermore, *S. pombe* encodes a second Mag ortholog, Mag2, that lacks detectable glycosylase activity (Alseth, Osman et al. 2005) despite strong sequence similarity to Mag1, even in the functionally important residues described here. Thus, although the present work provides some biochemical insight into alkylation repair, the various other ways in which Mag proteins contribute to the alkylation response in yeast remains to be determined.

2.3 Experimental Procedures

2.3.1 Protein purification

The spMag1, scMag, and bhMag genes were PCR amplified from *S. pombe*, *S. cerevisiae*, and *B. halodurans* (ATCC BAA-125D-5) genomic DNA and cloned into expression vector pBG100 (Vanderbilt University Center for Structural Biology) that produces an N-terminal His_6-tagged protein. Wild-type and mutant yeast proteins were overexpressed in *E. coli* C41 (spMag1) or BL21 (scMag) cells for 4 hr at 25° C in LB media. BhMag was overexpressed in *E. coli* HMS174 (wild-type) or BL21 (S53H mutant) for 16 hr at 16° C. Cells were lysed in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 10% glycerol and proteins isolated using Ni-NTA (Qiagen) affinity chromatography. Following cleavage of the His_6 tag, proteins were purified by heparin and gel filtration chromatography in 20mM Tris-HCl (pH 7.5), 150 mM NaCl, 2mM DTT and 0.05 mM EDTA. Mutant protein constructs were generated using a Quik-Change kit (Stratagene), purified the same as wild-type, and their structural integrity verified using circular dichroism spectroscopy (Fig. A6). SeMet-spMag1 was overexpressed in C41 cells for 16

hr at 16° C in minimal media supplemented with 70 mg/L selenomethionine (Acros Organics) under conditions that suppress normal methionine biosynthesis (Van Duyne, Standaert et al. 1993) and was purified the same as wild-type, except that 5 mM methionine and 5 mM DTT were added to all purification buffers after the Ni-NTA step.

2.3.2 X-ray crystallography

Protein-DNA complexes were assembled by incubating 0.20 mM SpMag1 with 0.24 mM DNA (d(TGTCCA(THF)GTCT)/d(AAGACTTGGAC) at 4° C for 20 min. Crystals were obtained by vapor diffusion at 21°C against reservoir solution containing 100 mM MES (pH 6.5), 20% (w/v) PEG 8K and 2.4% (v/v) glycerol and were flash frozen in mother liquor containing 15% (v/v) glycerol prior to data collection. X-ray diffraction data (Table 2) were collected at the Advanced Photon Source (Argonne, IL) beamlines 21-ID (native) and 22-BM (SeMet) and processed using HKL 2000 (Otwinowski and Minor 1997). SAD data were collected at the selenium absorption peak. Positions of 10 Se atoms were identified and phases calculated using the program SHARP (Vonrhein, Blanc et al. 2007). A crystallographic model corresponding to amino acids 16-221 and nucleotides 1-22 for each of 2 protein/DNA complexes in the asymmetric unit was built into 2.8 Å Se-SAD electron density maps using Coot (Emsley and Cowtan 2004). The two crystallographically distinct protein/DNA complexes were virtually identical with an r.m.s. deviation of 0.33 Å for main chain atoms.

The crystallographic model was refined against native diffraction data extending to 2.2 Å using SAD phase combination and a maximum likelihood target as implemented in PHENIX (Adams, Grosse-Kunstleve et al. 2002). Improvements to the model were made by manual inspection of σ_A-weighted $2mF_o$-F_c and mF_o- DF_c electron density maps.

Translation/libration/screw-rotation (TLS) refinement was used to model anisotropic motion of each protein/DNA complex, and individual anisotropic B-factors derived from the refined TLS parameters were held fixed during subsequent rounds of refinement. The final model was validated using PROCHECK (Laskowski, Rullmannn et al. 1996) and deposited in the Protein Data Bank under accession number 3S6I.

2.3.3 Enzymatic activity

Base excision activities were measured by following the alkaline cleavage of the abasic DNA product of alkylbase excision from a 25-mer oligonucleotide duplex containing a centrally positioned εA•C or 7mG•C base pair. Oligonucleotides were purchased from Integrated DNA Technologies (IDT, USA), and 7mG-DNA was prepared enzymatically as previously described (Asaeda, Ide et al. 2000; Rubinson, Metz et al. 2008). Nucleotide sequences used were 5′-^{32}P-d(GACCA CTACA CC**X**TT TCCTA ACAAC) annealed to 5′-d(GTTGT TAGGA AACGG TGTAG TGGTC). Reaction mixtures (75 µL) contained 10 µM enzyme, 100 nM radiolabeled DNA duplex, 100 mM KCl, 2 mM DTT, and either 50mM sodium acetate (pH 6.0) for εA-DNA and HEPES (pH 7.5) for 7mG-DNA. We verified that under these conditions, the enzyme concentration is saturating. Reactions were initiated by addition of enzyme and incubated at 25 °C. As a control to verify that the reaction rates were not affected by changed in the protein as a result of the long reaction times, we performed control reactions in which the protein was pre-incubated for 4 hours under the reaction conditions prior to initiating the enzymatic reaction (Fig. A7). Aliquots (8 µL) were stopped at various time points by addition of 0.2 M NaOH and heated at 70 °C for 2 min. The cleaved 12-mer product and unreacted 25-mer substrate oligonucleotides were separated by 15% polyacrylamide/7M

urea gel electrophoresis and quantified by autoradiography. Fraction product (F_P) at each time point was calculated by $F_P = I_P/(I_S+I_P)$, where I_S and I_P are the integrated intensities of substrate and product bands. Rate constants (k_{cat}) were determined from single-exponential fits to the data from three separate experiments. Rate differences between wild-type and mutant proteins were judged to be significant based on *p*-values derived paired t-test analysis.

2.3.4 DNA binding

DNA binding was measured by the change in fluorescence anisotropy as spMag1 was added to oligonucleotide duplexes containing a centrally located THF or Gua in one strand and a 6-carboxyfluorescein on the 3'-end of the other (Table 4). Increasing concentrations of protein (0-30 μM) were added to a 50 nM DNA in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT, and 0.01 mM EDTA. Polarized fluorescence intensities using excitation and emission wavelengths of 485 and 538 nm were measured at 25° C using a SpectraMax M5 microplate reader (Molecular Devices). Dissociation constants were derived by fitting the data using the equation, $A = A_{max}[protein]/(K_d+[protein])$, in which A is the anisotropy value at a given protein concentration and A_{max} is the anisotropy value at maximal binding.

Chapter 3

NON-PRODUCTIVE DNA DAMAGE BINDING BY DNA GLYCOSYLASE-LIKE PROTEIN MAG2 FROM *SCHIZOSACCHAROMYCES POMBE**

Abstract

S. pombe contains two paralogous proteins, Mag1 and Mag2, related to the helix-hairpin-helix (HhH) superfamily of alkylpurine DNA glycosylases from yeast and bacteria. Phylogenetic analysis of related proteins from four *Schizosaccharomyces* and other fungal species shows that the Mag1/Mag2 duplication is unique to the genus *Schizosaccharomyces* and most likely occurred in its ancestor. Mag1 excises *N*3- and *N*7-alkylguanines and 1,N^6-ethenoadenine from DNA, whereas Mag2 has been reported to have no detectible alkylpurine base excision activity despite high sequence and active site similarity to Mag1. To understand this discrepancy, we determined the crystal structure of Mag2 bound to abasic DNA and compared it to our previously determined Mag1-DNA structure. In contrast to Mag1, Mag2 does not flip the abasic moiety into the active site or stabilize the DNA strand 5′ to the lesion, suggesting that it is incapable of forming a catalytically competent protein-DNA complex.

*The work presented in this chapter was published in **Adhikary, S.**, M. C. Cato, et al. (2012). "Non-productive DNA damage binding by DNA glycosylase-like protein Mag2 from Schizosaccharomyces pombe." DNA Repair (Amst) http://dx.doi.org/10.1016/j.dnarep.2012.12.001

Subtle differences in Mag1 and Mag2 interactions with the DNA duplex illustrate how Mag2 can stall at damage sites without fully engaging the lesion. We tested our structural predictions by mutational analysis of base excision and found a single amino acid responsible at least in part for Mag2's lack of activity. Substitution of Mag2 Asp56, which caps the helix at the base of the DNA intercalation loop, with the corresponding serine residue in Mag1 endows Mag2 with εA excision activity comparable to Mag1. This work provides novel insight into the chemical and physical determinants by which the HhH glycosylases engage DNA in a catalytically productive manner.

3.1 Introduction

All organisms are equipped with a number of DNA glycosylases that initiate base excision repair (BER) of potentially harmful alkylated, oxidized, and deaminated nucleobases from DNA (Dalhus, Helle et al. 2007). Upon locating an aberrant nucleotide, DNA glycosylases catalyze hydrolysis of the *N*-glycosidic bond to produce an abasic site, which is subsequently processed by an apurinic/apyrimidinic (AP) endonuclease, DNA polymerase, and DNA ligase activities to fully repair the damage (Dalhus, Helle et al. 2007). In order to break the *N*-glycosidic bond, DNA glycosylases trap the damaged nucleotide in an extrahelical conformation in which the nucleobase is pulled out of the DNA base stack and tucked into the active site of the enzyme. The extruded DNA is stabilized by several side chains that intercalate into the base stack at the damage site. The importance of these stabilizing interactions to catalysis is underscored by the fact that mutation of the intercalating residues renders the glycosylase inactive (Jiang, Kwon et al. 2001; Vallur, Feller et al. 2002; Eichman, O'Rourke et al. 2003; Livingston, Kundu et al. 2005; Maiti, Morgan et al. 2009), and recent work illustrates how they may play an

active role in detecting or even discriminating against certain types of damage (Banerjee, Santos et al. 2006; Qi, Spong et al. 2009; Bowman, Lee et al. 2010; Chapter II).

DNA glycosylases specialized for alkylation damage are found in all organisms. The mammalian AAG enzyme is structurally unique and removes a variety of modified purines, including *N*3-methyladenine (3mA), *N*7-methylguanine (7mG), 1,N^6-ethenoadenine (εA), as well as hypoxanthine, a deamination product of adenine (Gallagher and Brent 1982; Singer, Antoccia et al. 1992; O'Connor 1993; Engelward, Weeda et al. 1997; Hang, Singer et al. 1997; O'Brien and Ellenberger 2004). The bacterial (TAG, AlkA) and yeast (Mag, Mag1) enzymes all belong to the helix-hairpin-helix (HhH) structural superfamily and exhibit a range of substrate specificities (Dalhus, Helle et al. 2007). TAG is constitutively active and highly specific for 3mA and 3mG, whereas AlkA is induced as part of the adaptive response to alkylation exposure and shares AAG's broad substrate preference (Samson and Cairns 1977; Dalhus, Helle et al. 2007). Similar to AlkA, *Saccharomyces cerevisiae* Mag is inducible, lacks specificity for any particular alkylpurine, and exhibits a strong mutator phenotype (Chen, Derfler et al. 1990; Chen and Samson 1991; Saparbaev, Kleibl et al. 1995; Lingaraju, Kartalou et al. 2008). The genetically tractable fission yeast, *Schizosaccharomyces pombe*, has provided a convenient model system to investigate the importance of individual components to BER (Memisoglu and Samson 2000). *S. pombe* contains one alkylpurine glycosylase, Mag1, which removes 3mA, 7mG, and εA at lower rates than Mag, partially explaining why Mag1 is not as critical to alkylation resistance as Mag and AlkA (Memisoglu and Samson 1996; Memisoglu and Samson 2000; Alseth, Osman et al. 2005; Lingaraju, Kartalou et al. 2008; Chapter II). The crystal structure of Mag1 in complex with abasic

DNA revealed a unique protein-DNA contact that inhibited activity toward εA, partially explaining the difference in activity with Mag (Chapter II). *S. pombe* contains a related protein, Mag2, which shares 41% sequence identity and 69% overall similarity to Mag1, but lacks glycosylase activity (Alseth, Osman et al. 2005). Genetic studies with *S. pombe* strains *mag1*Δ and *mag2*Δ showed no increase in alkylation sensitivity, whereas *nth1*Δ (AP lyase), *apn2*Δ (AP endonuclease) and *rad2*Δ (flap endonuclease) are all sensitive to MMS (Alseth, Osman et al. 2005). Interestingly, strains harboring double deletions in *nth1*Δ *mag1*Δ, *nth1*Δ *mag2*Δ, *apn2*Δ *mag1*Δ, and *rad2*Δ *mag1*Δ restored MMS resistance (Memisoglu and Samson 2000; Alseth, Osman et al. 2005; Kanamitsu, Tanihigashi et al. 2007), indicating that Mag1 is involved in BER and that Mag2 may play a role in Nth1-mediated DNA repair, although alkylation damage may not be the preferred substrate of either enzyme. In addition, the studies also showed that exogenous expression of either Mag1 or Mag2 in a *rad16*Δ/*mag1*Δ/*mag2*Δ triple mutant restores MMS resistance lost as a result of the deletions (Kanamitsu, Tanihigashi et al. 2007). However, the lack of glycosylase activity of Mag2 *in vitro* has made it difficult to understand its role and the need for two highly homologous enzymes in *S. pombe*.

In order to understand the lack of base excision activity by Mag2, we carried out an evolutionary analysis of Mag-related genes in yeasts and related fungi and determined the crystal structure of Mag2 in complex with DNA containing an abasic site. We find that duplication of Mag1 and Mag2 is specific to the *Schizosaccharomyces* clade, suggesting that the two proteins diverged at least 200 million years ago (Rhind, Chen et al. 2011), and that *S. pombe* Mag2 experienced an accelerated rate of evolution relative to Mag1. Interestingly, the Mag2 structure shows that unlike Mag1, the abasic site is not

extruded in the active site and the DNA 5′ to the lesion is disordered, suggesting that the lack of activity is a consequence of the inability of Mag2 to engage the substrate. Comparison of Mag1 and Mag2 DNA complexes illuminates structural features necessary for DNA damage recognition by this family of DNA glycosylases. We show by mutagenesis that inhibition of Mag2 base excision activity can be explained by a negatively-charged, helix-capping interaction at the base of the structural motif that normally stabilizes the extrahelical base, which further underscores the importance of DNA intercalating residues to base excision activity.

3.2 Results

3.2.1 Evolutionary analysis reveals that Mag1 and Mag2 were duplicated in the ancestor of the *Schizosaccharomyces* clade and preserved

BLAST analysis of Mag-related proteins across fungi finds a single high scoring homolog of Mag in most fungal genomes. An evolutionary analysis reveals that the duplication of Mag1 and Mag2 is specific to the *Schizosaccharomyces* clade with both paralogs retained in all 4 *Schizosaccharomyces* species with sequenced genomes (*S. pombe*, *S. octosporus*, *S. cryophilus*, and *S. japonicus*). As seen in Fig. 3.1, after duplication, the topology of the tree matches the established phylogenetic relationships of the species. Given that the divergence of the 4 *Schizosaccharomyces* species is estimated to have taken place approximately 200 million years ago (Rhind, Chen et al. 2011), the long-term preservation of both paralogs in all four species suggests that purifying selection is maintaining both copies. The phylogenetic tree also reveals that the lineage leading to *S. pombe* Mag2 experienced an accelerated rate of amino acid substitutions

relative to Mag1, which suggests that it has diverged further from the ancestral function of Mag than Mag1.

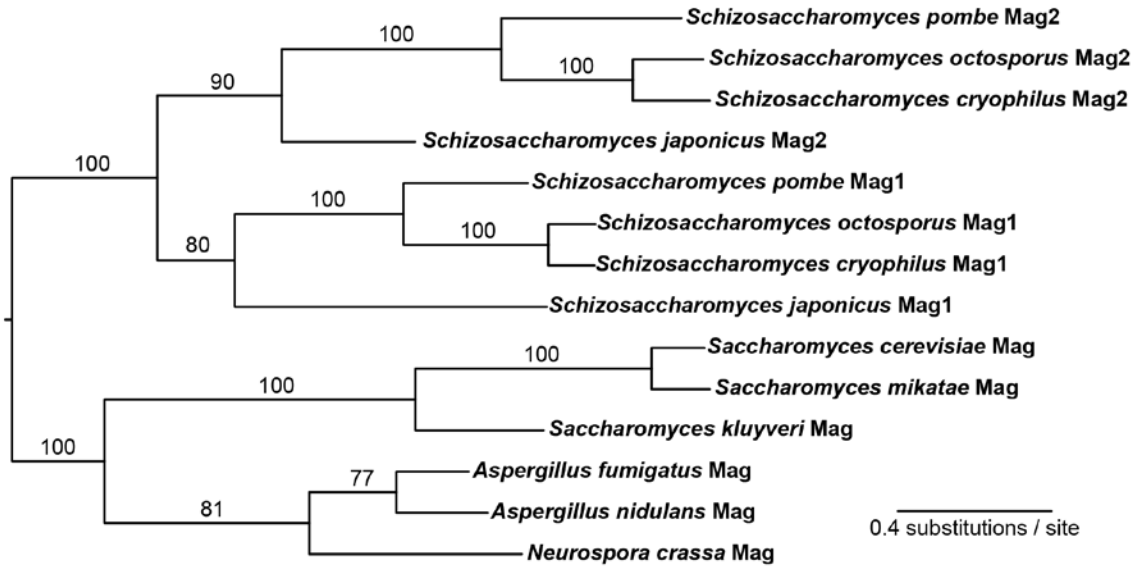


Fig. 3.1 Phylogenetic history of the Mag family of proteins from *Schizosaccharomyces* and representative fungal relatives. Numbers above branches represent bootstrap support and branch lengths represent protein evolutionary rate in units of amino acid substitutions per site. The Mag phylogeny shows that the duplication of Mag1 and Mag2 is specific and unique to the *Schizosaccharomyces* clade and that both paralogs have been retained in all 4 *Schizosaccharomyces* species.

3.2.2 Mag2 is not an alkylpurine DNA glycosylase

The reported lack of base excision activity by Mag2 is surprising given the sequence conservation to Mag1 in residues that we had previously identified to comprise the Mag1 active site. We therefore tested εA excision activity of Mag2 under conditions that support Mag1 εA activity. Using a standard assay involving alkaline hydrolysis of abasic sites generated by glycosylase action, we were unable to detect significant excision of εA at enzyme concentrations as high as 10 μM and reaction times of 48 hours (Fig. 3.5). Since all other Mag orthologs tested show at least low levels of activity for εA (Chapter

II), our results are consistent with Mag2 being deficient as an alkylpurine DNA glycosylase.

3.2.3 Structure of the Mag2-DNA complex

To glean structural insight into the lack of base excision activity, we crystallized Mag2 in complex with DNA containing a centrally located tetrahydrofuran (THF) abasic site analog on one strand. We used this same strategy previously to determine the structure of a Mag1/THF-DNA complex since THF closely mimics the structural properties of natural abasic sites but is refractory to hydrolysis (Pereira de Jesus, Serre et al. 2005).

In cases where both substrate and product complexes have been crystallized, the conformation of the abasic-DNA is remarkably similar to that of the substrate, consistent with the fact that glycosylases typically display high affinities for the abasic-DNA product (Norman, Bruner et al. 2001) (Tchou, Michaels et al. 1993; Castaing, Fourrey et al. 1999; Lingaraju, Kartalou et al. 2008; Imamura, Wallace et al. 2009). The Mag2-DNA crystal structure was determined by molecular replacement using the Mag1 protein as a search model. A crystallographic model with one Mag2-DNA complex in the asymmetric unit was refined against 1.9 Å native diffraction data to a crystallographic residual of 19.2% (R_{free} = 23.5%) (Table 6).

Table 6. Crystallographic data collection and refinement statistics

Data collection*	
Wavelength (Å)	0.9787
Space group	$P3_221$
Cell dimensions	
a, *b*, *c* (Å)	54.9, 54.9, 153.46
α, β, γ (°)	90, 90, 120
Resolution (Å)	29.91 – 1.90 (1.97-1.90)
R_{sym}	0.07 (0.39)
I / σ_I	30.3 (7.9)
Completeness (%)	99.7 (98.8)
Redundancy	10.6 (9.7)
Refinement	
Resolution (Å)	1.90
No. reflections	22187
R_{work} / R_{free}	0.192/0.234
No. of atoms	
Protein	1617
DNA	440
Solvent	114
B-factors	
Protein	28.9
DNA	93.3
Solvent	47.5
Wilson B-factor	24.9
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.117

*Values in parentheses refer to the highest resolution shell.

The structure confirmed that Mag2 adopts the HhH architecture common to yeast Mag/Mag1 and bacterial TAG/AlkA. Mag2 consists of two α-helical subdomains—one spanning helices C-J that contains the HhH DNA binding motif (helices I-J) and a second containing the N- and C-termini (helices A-B and K-M) (Fig. 3.2A). The HhH domain anchors the protein to the DNA through three similar helix-coil-helix interactions—helices CD, which contains the DNA interrogation loop, FG, and IJ, which forms the HhH motif (Fig. 3.2).

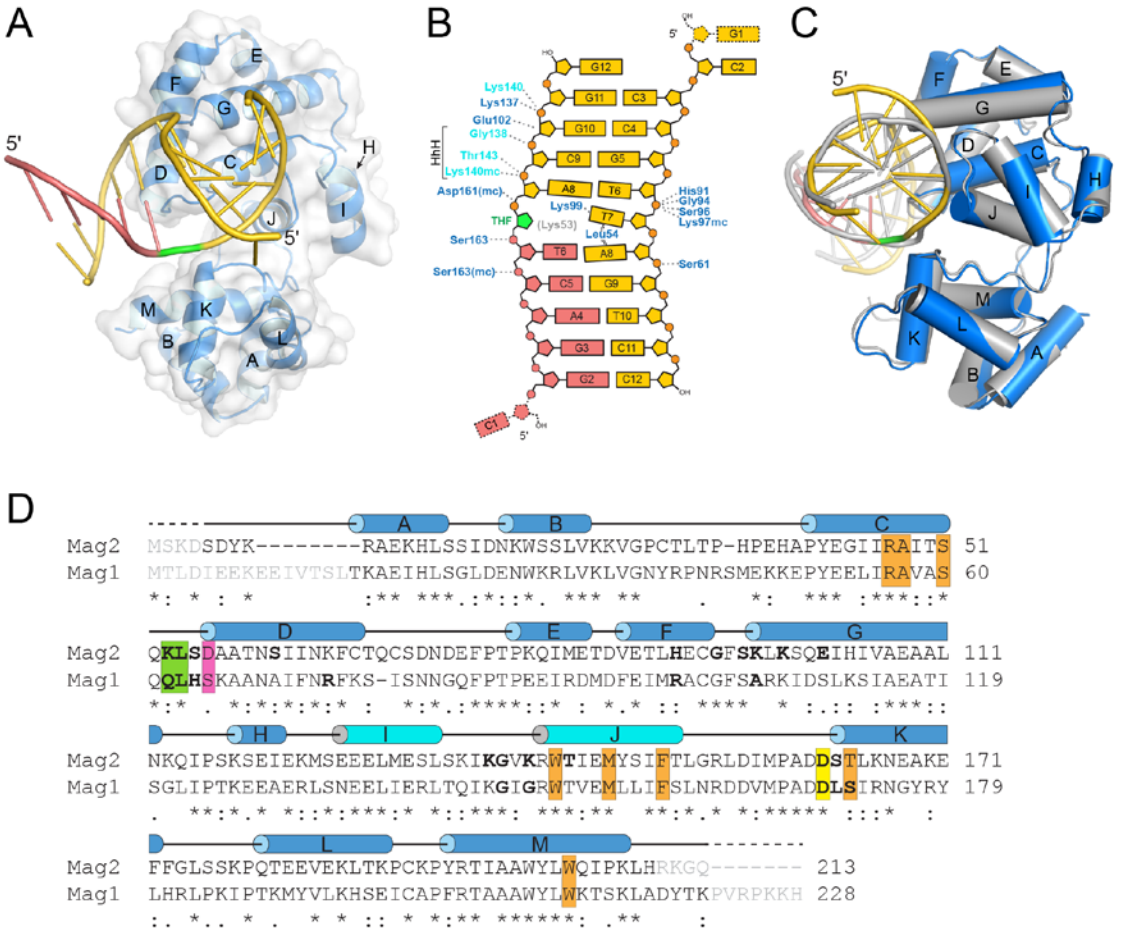


Fig. 3.2 Crystal structure of *S. pombe* Mag2 bound to abasic-DNA. (A,B) Mag2 (blue ribbons, white solvent accessible surface) in complex with DNA (gold) containing a THF abasic site analog (green). The disordered stretch of DNA is colored salmon. (B) Schematic of Mag2-DNA interactions, colored as in panel A with HhH residues in cyan. Dotted lines represent either hydrogen bonds or van der Waals interactions. *mc*, main chain. (C) Superposition of Mag2/THF-DNA (colored as in panel A) onto *S. pombe* Mag1/THF-DNA (PDB ID 3S6I, silver). (D) Structure based sequence alignment of Mag1 and Mag2, generated with Swiss PDB Viewer (Guex and Peitsch 1997). Secondary structure elements are shown above the alignment. The HhH loop is represented as light blue cylinders. Nucleobase binding pocket residues are highlighted orange, intercalating plug and wedge residues are green, and the catalytic aspartate found in HhH glycosylases is yellow. Mag2 Asp56, responsible for diminished base excision activity, is highlighted magenta.

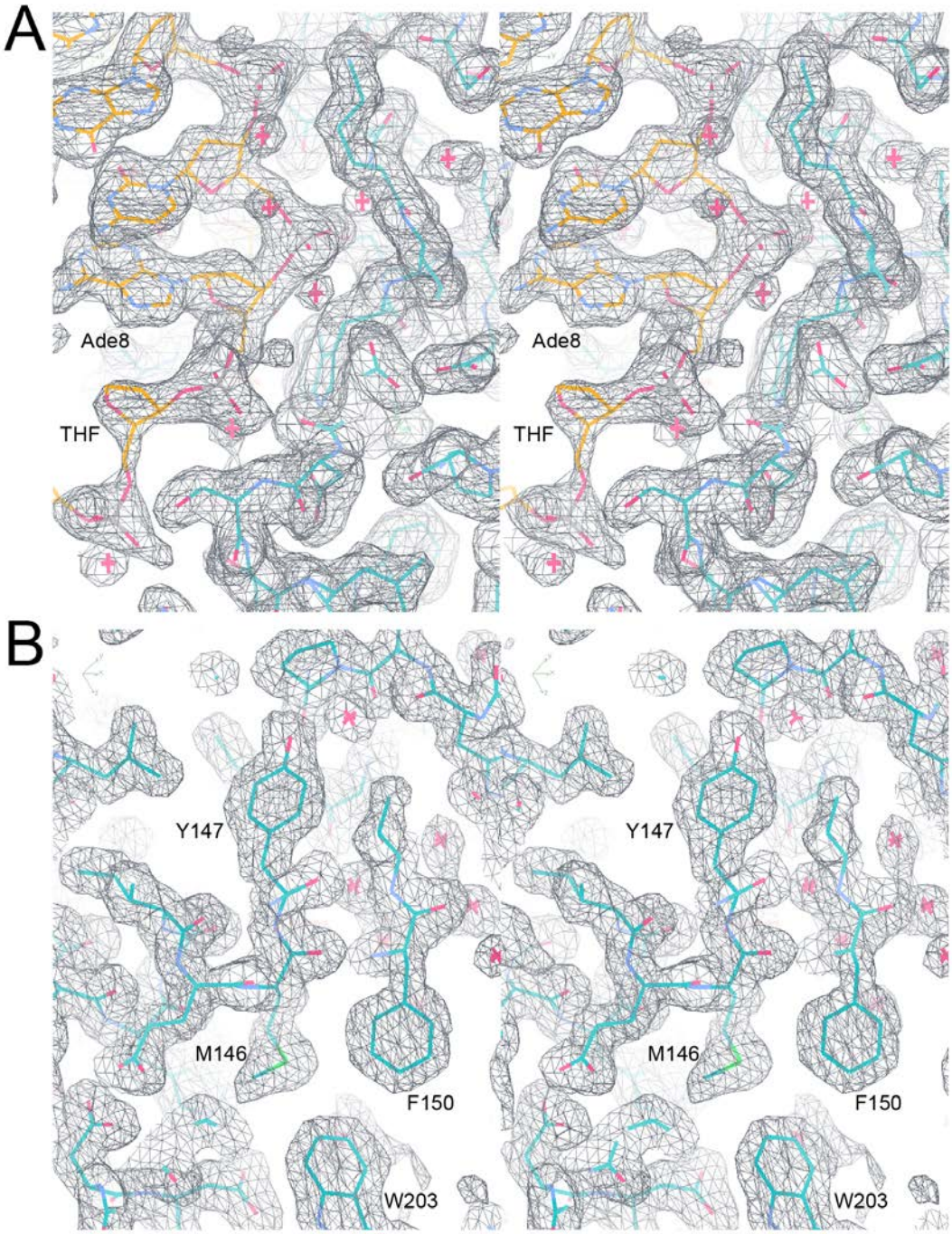


Fig. 3.3 Composite (2Fo-Fc) annealed omit electron density maps contoured at 1σ. Images are stereoviews of the region around the THF abasic site (A) and the apparent nucleobase binding pocket (B). Protein and DNA carbon atoms are colored cyan and gold, respectively.

The protein binds both strands of the DNA primarily through electrostatic interactions with the phosphate backbone; largely with the majority of protein-DNA contacts formed between the HhH motif and the 3′ half of the damaged strand (Fig. 3.2B). As a consequence, the abasic site is directed across the deep cavity at the interface between the two subdomains, which normally forms the active site/nucleobase binding pocket (Fig. 3.2A). There is a 70° bend in the DNA toward the major groove as a result of Leu54 at the tip of the helix C-D loop wedging itself into the duplex on the undamaged strand. This general mode of DNA binding is the same as that seen in the Mag1/THF-DNA structure. The structures of Mag1 and Mag2 are virtually identical, with an r.m.s.d. of 1.30 Å for all protein backbone atoms (Fig. 3.2C). The positions of side chains forming the nucleobase binding pocket in Mag1 are all conserved in Mag2, as expected from the sequence identity in these residues (Fig. 3.2).

3.2.4. A partially bound enzyme-DNA complex

In the structure of Mag1 in complex with THF-DNA, the THF moiety was flipped toward the nucleobase binding pocket, and the gap in the duplex stabilized by the Gln62 side chain. Surprisingly, in Mag2 the THF remains stacked into the duplex (Figs. 3.3 and 3.4A,B), despite the fact that the residues in the nucleobase binding pocket are identical between the two proteins. The electron density clearly supports this un-flipped conformation, with no discernible density in the location of the THF in the Mag1 structure (Fig. 3.3). Consequently, the side chain for Lys53, which corresponds to the intercalating Gln62 plug residue in Mag1, is disordered (Fig. 3.4A). In addition, the entire strand to the 5′-side of the lesion is disordered, as evident from the significantly higher B-factors for

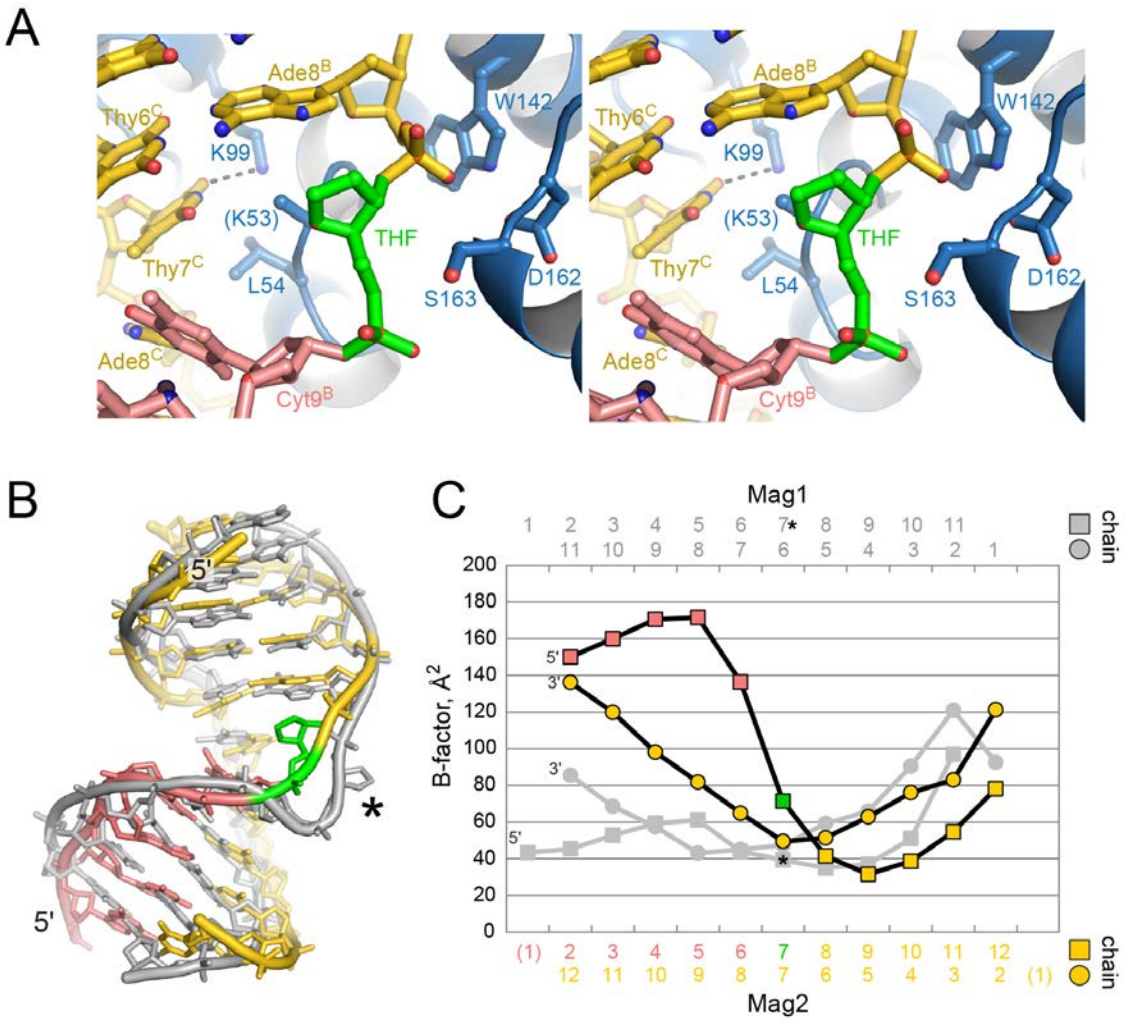


Fig. 3.4 Mag2 does not flip abasic sites. (A) Stereoview of the region around the THF abasic site in Mag2. Protein and DNA are colored by carbon atom as in Figure 2. (B) Superposition of DNA bound to Mag1 (grey) and Mag2 (gold/green/salmon). The THF moiety is flipped out of the duplex in the Mag1 structure (denoted with an asterisk) but remains stacked in the duplex in the Mag2 structure (green). (C) Average B-factor per nucleotide between Mag2 (gold/green/salmon) and Mag1 (grey) complexes. The THF-containing and unmodified strands are shown as squares and circles respectively.

those nucleotides (Fig. 3.4C). Interestingly, the 5′-side of the damaged strands in other HhH structures is generally well-ordered despite any significant protein-DNA contacts in that region of the duplex. We can rule out crystal packing as the reason for the disorder in the Mag2 DNA because the ends of the DNA are base stacked against a symmetry mate,

and the B-factors for the stacking residue G2 are lower than its neighbors (Fig. 3.4C). Thus, unlike other DNA glycosylases, Mag2 does not flip an abasic site toward the conserved nucleobase binding pocket and does not fully engage the DNA duplex on one side of the lesion.

3.2.5 A structural basis for the lack of glycosylase activity in Mag2

The conservation of the nucleobase binding pocket and bulky intercalating residues (Lys53, Leu54) would seemingly all favor a flipped conformation of the DNA. Why, then, does Mag2 not engage the abasic site? We reasoned that the answer lies in the subtle differences at the protein-lesion interfaces of Mag1 and Mag2. Helix D at the base of the intercalating loop normally forms a favorable dipole interaction with the DNA in which its N-terminal end is directed into the minor groove around the lesion (Fig. 3.5A).

In Mag2, this dipole is capped with an aspartate residue (Asp56) at the N-terminal end of helix D, whereas this position is occupied by serine or lysine in Mag1 orthologs and glycine in Mag (Figs. 3.2D and 3.7). We hypothesized that the neutralized dipole in Mag2 would weaken the interaction between the intercalating CD loop and the DNA, which would consequently impair the ability of the protein to stabilize an extrahelical nucleotide. We tested this by substituting Asp56 with serine and measuring εA excision activity. As seen in Figures 3.5B-C, the Asp56Ser mutation endows Mag2 with the ability to excise εA at levels similar to Mag1. This result strongly suggests that in addition to the direct protein-intercalating residues, a helix dipole in this vicinity of the DNA damage site is important for stabilizing the protein-DNA complex for catalysis.

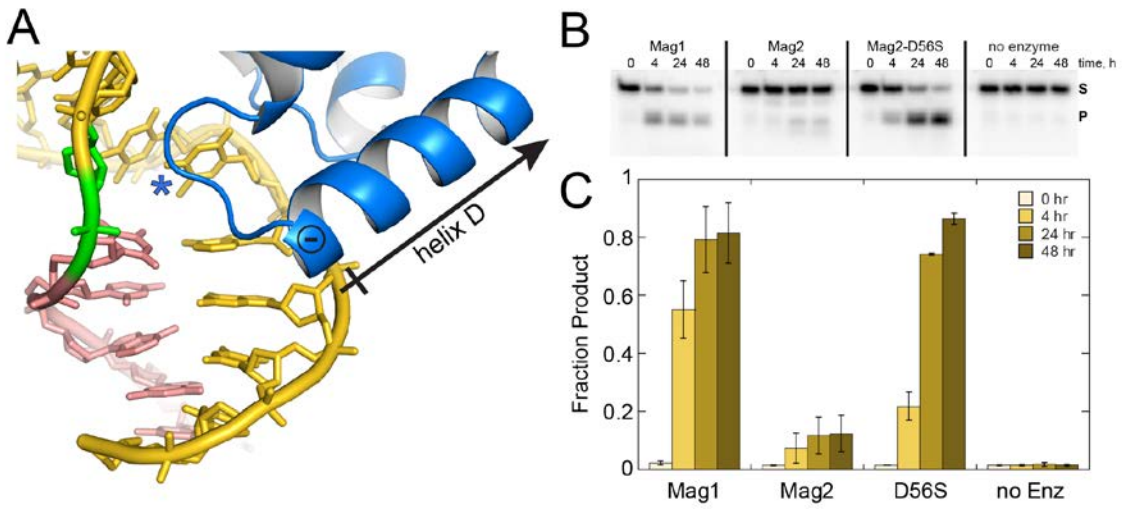


Fig. 3.5 Mag2 has a capped helix dipole that inhibits base excision activity. (A) Close-up view of the DNA interrogating loop. The arrow denotes the direction of the dipole of helix D, with the positive end toward the DNA. The position of Asp56 is marked with a negative sign (-), and the position of the intercalating plug residue, Lys53, is marked with an asterisk (*). (B) Representative εA excision activity for Mag1, Mag2, and Mag2-Asp56Ser mutant. The image shows the electrophoretic separation of the 25mer εA -DNA substrate (S) from 12mer product (P) generated from alkaline cleavage of abasic sites produced by the glycosylase. Reaction times in hours are shown at the top. (C) Quantitation of εA activity data, shown as the average of three independent experiments.

Further inspection of the structure offers several additional explanations for the disengaged DNA and lack of εA excision activity of Mag2. In Mag1, the backbone of the flipped THF is extruded toward the active site and stabilized by a van der Waals contact from the side chain of Leu171 (Fig. 3.6A). At this same position in Mag2 is a serine residue (Ser163), which would not provide the steric bulk to divert the backbone in the extruded position. In addition to steric effects, the overall surface charge likely contributes to the disengaged DNA complex in Mag2. Whereas the DNA binding surface of Mag1 is highly positively charged, Mag2 shows patches of negative charge (Fig. 3.6B). Most notably, the N/C-terminal domain that normally cradles the DNA 5′ to the

lesion is strongly negative in Mag2. Thus, although many of the specific protein-DNA contact points are similar between the two proteins, the overall electrostatic surface landscape may contribute to the instability of the protein-DNA complex.

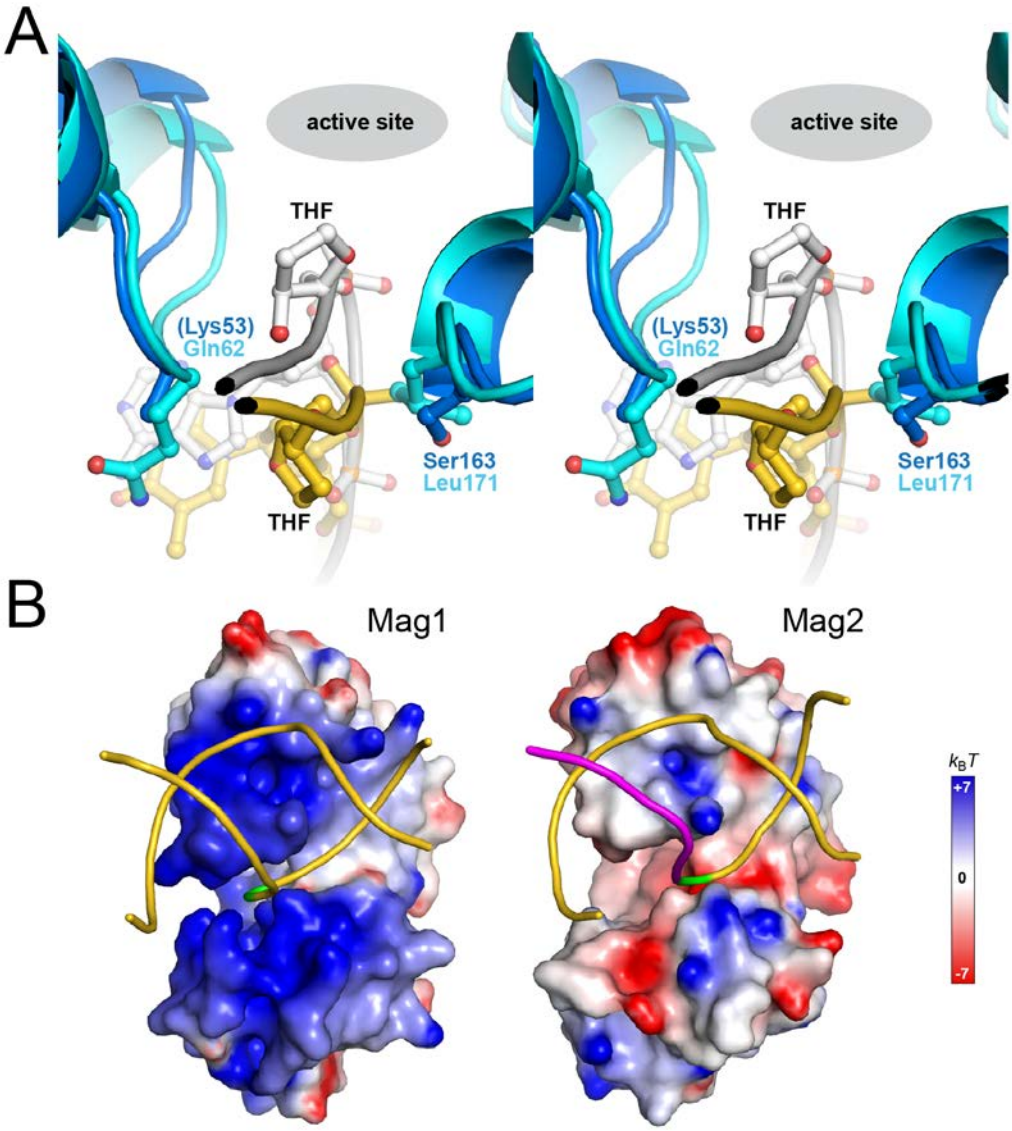


Fig. 3.6 Additional structural differences between Mag1 and Mag2 that likely impact DNA binding and base flipping. (A) Protein-DNA interactions around the abasic site. The Mag2 complex is colored blue (protein) and gold (DNA), and the Mag1 complex is cyan and silver. The phosphate backbone and main-chain protein atoms are traced with a cartoon. Only the damaged DNA strand is shown for clarity. The active site is at the top of the figure, and the outside of the protein (solvent) is at the bottom. (B) Solvent-accessible surfaces colored according to electrostatic potential (red, negative; blue, positive; −7 to +7 k_BT) calculated using DelPhi (Rocchia, Sridharan et al. 2002). The DNA is colored as in Fig. 3.2.

3.3 Discussion

3.3.1 New insight into how glycosylases stabilize the DNA complex

The high sequence and structural similarity between Mag1 and Mag2, and now the availability of both structures bound to THF-DNA offers a unique glimpse into the chemical and physical requirements for base excision activity by the HhH glycosylases, which include AlkA, TAG, Ogg1, MutY, Endo III, and the 5mC glycosylases DEMETER/ROS1 (Dalhus, Helle et al. 2007). Structural and biochemical studies to date have shown that glycosylase residues that intercalate the base stack are critical for interrogating the duplex in search of damage and stabilizing the extrahelical nucleotide and in a variety of protein architectures (Stivers 2004; Dalhus, Helle et al. 2007; Stivers 2008). To establish a fully docked enzyme-substrate complex, a side chain plug fills the void created in the duplex from the flipped nucleotide, and a bulky wedge residue stacks against the bases on the opposite strand to stabilize the sharp kink in the duplex. The present work reveals that in addition to the direct steric effects, there exist more subtle, electronic interactions with the duplex that are required for the protein to fully engage the damage, and that in this case can explain the discrepancy in Mag1 and Mag2 activities. The base excision activity observed in the Mag2 Asp56Ser mutant provides new insight into the ensemble of interactions necessary for the enzyme to fully engage damaged DNA. To our knowledge, this is the most dramatic gain of function for any DNA glycosylase. We previously showed that Mag1 contains an inhibitory histidine residue (His64) that resides on the plug/wedge loop and that interacts directly with nucleobases adjacent to the lesion. Removal of this interaction through a His64Ser substitution increased Mag1 εA activity 4-fold (Chapter II).

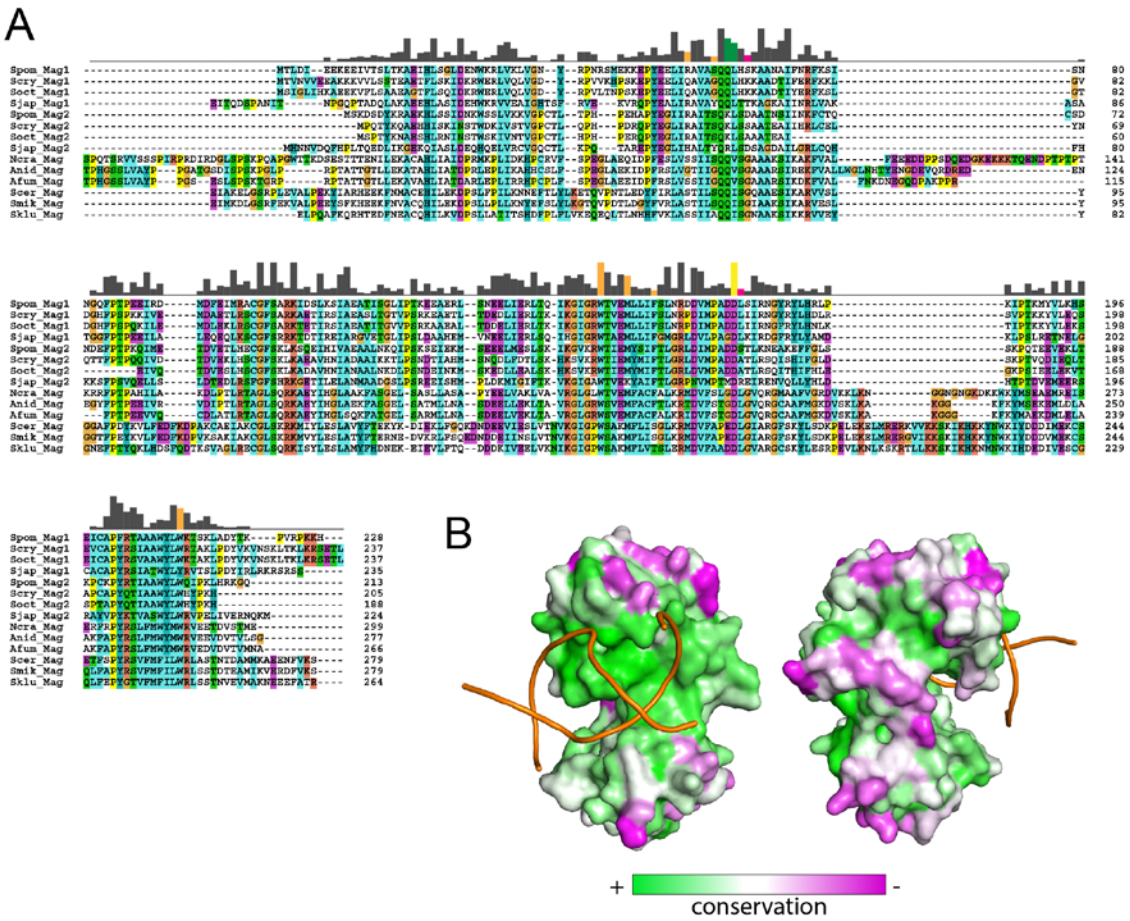


Fig. 3.7 Conservation among the Mag protein family from yeasts. (A) Sequence alignment of Mag orthologs from various organisms. Protein residues are highlighted according to side chain chemistry. Conservation is quantified by the bar graph above the alignment, and are colored according to their position in the structure (orange, base binding pocket; yellow, conserved catalytic HhH aspartate; green, plug/wedge DNA intercalating residues; magenta, putative Mag2 inhibitory residues). Spon, Schizosaccharomyces pombe; Scry, Schizosaccharomyces cryophilus; Soct, Schizosaccharomyces octosporus; Sjap, Schizosaccharomyces japonicus; Ncra, Neurospora crassa; Anid, Aspergillus nidulans; Afum, Aspergillus fumigatus; Scer, Saccharomyces cerevisiae; Smik, Saccharomyces mikatae; Sklu, Saccharomyces kluyveri. (B) Sequence conservation mapped onto the structure of Mag2. The views differ by 180° rotation around a vertical axis. The figure was prepared using the ConSeq (http://conseq.tau.ac.il/) and ConSurf (http://consurf.tau.ac.il/) servers and PyMOL.

We speculated that the inhibition by His64 could be due to a perturbation in enzyme-substrate complex or to a slowed search for damage from a greater number of protein-DNA interactions. In contrast, the Mag2 Asp56 interaction is more subtle and the inhibition of activity is most likely due to a weakened dipole-phosphate interaction or a greater electrostatic repulsion between aspartate and the phosphate backbone, either of which would lead to a sub-optimal interaction between helix D and the minor groove and an altered position of the intercalating loop that harbors the plug/wedge residues. Indeed, the Lys53 plug in Mag2 is disordered, even though its Cα is aligned with the gap in the DNA at the THF site. We note that all three helix-coil-helix interactions that anchor the protein to the DNA—helices CD, FG, and IJ—typically contain a positive charge at the N-terminal end of the second helix, which would strengthen the dipole interaction with the DNA (Fig. 3.5). More work will be required to determine if interactions involving helix D in particular play a more direct role in enzyme-substrate engagement aside from merely stabilizing the intercalation loop.

The un-flipped THF in Mag2 is reminiscent of TAG in complex with THF-DNA and free 3mA, which showed static disorder of the abasic site that interconverted between partially flipped and un-flipped conformations. Contrary to the Mag2-DNA complex, however, the DNA strand upstream of the abasic site in the TAG structure was highly ordered (Fig. 3.8). This implied that the stabilized DNA is a requisite for TAG activity, and it was speculated that the protein uses significant binding energy from protein-DNA interactions within the 3mA binding pocket to aid in breakage of the relatively weak 3mA *N*-glycosidic bond through steric strain (Metz, Hollis et al. 2007). Mag2, on the other hand, is apparently missing those key elements to help stabilize the DNA upstream of the

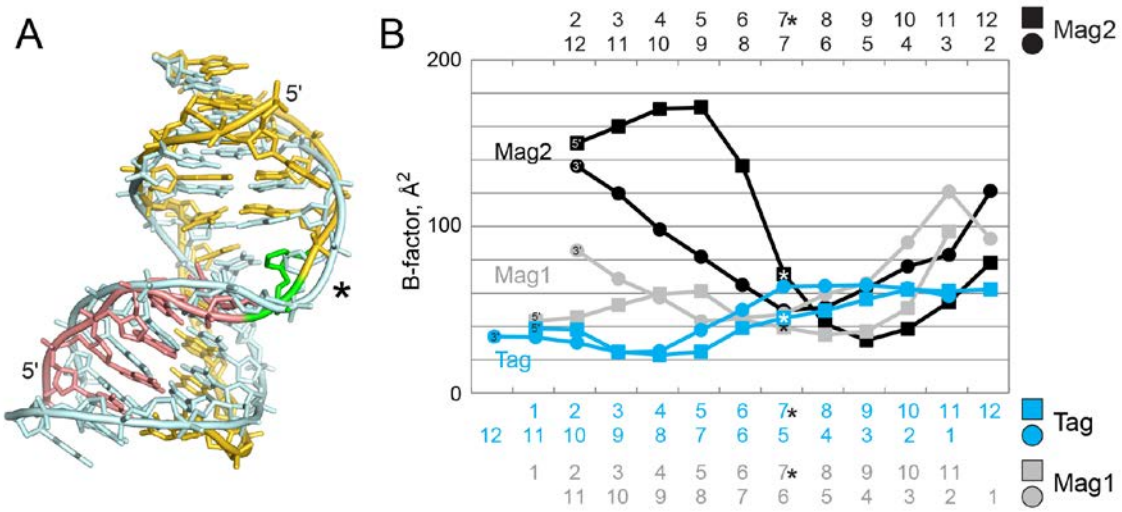


Fig. 3.8 Comparison of unflipped DNA bound to Mag2 and *E. coli* TAG. (A) THF-DNA from TAG (PDB ID 2OFI, silver) superimposed onto Mag2 (green THF, magenta disordered DNA). The position of the THF is marked with an asterisk (*). (B) Average B-factors for TAG (blue), Mag2 (black) and Mag1 (grey) THF complexes. In all cases, the THF moiety is in position #7 on the top strand (squares).

lesion, precluding glycosylase activity. The structural basis for the disordered 5′ strand in Mag2 (and the highly ordered 5′ strand in TAG and Mag1) is unclear given the lack of protein contacts. Nonetheless, the similarity in positions of the un-flipped THF in TAG and Mag2 suggest that the structure of the upstream DNA is not a directly coupled to the conformation of the abasic site. Furthermore, we speculate that unlike TAG, the nucleobase binding pockets in the Mag enzymes play only a minor role in stabilizing the complex, given the conservation of Mag1 and Mag2 pockets and the growing number of structures with abasic sites rotated toward the pocket even in the absence of damaged base.

3.3.2 Repair of base damage in *S. pombe*.

Most fungi have a single copy of Mag, with only the *Schizosaccharomyces* clade preserving two paralogous copies. The persistence of the Mag1 and Mag2 proteins in all

Schizosaccharomyces species suggests that both are functional, consistent with genetic analysis showing Mag1 and Mag2 play a role in *nth1*-mediated BER (Kanamitsu, Tanihigashi et al. 2007). The relatively longer evolutionary branch leading to *S. pombe* Mag2 relative to Mag1 suggests that the duplicates might have at least partially subfunctionalized, with Mag1 generally retaining a greater proportion of Mag's ancestral function and, thus, having more functionally in common with *S. cerevisiae* Mag. A potential alternative interpretation to a purely subfunctionalization model is that Mag2 may have gained additional functionality after its duplication in the *Schizosaccharomyces* ancestor and subsequent sequence divergence. Intriguingly, the trajectory of functional divergence between Mag1 and Mag2 after duplication may not be simple. For example, the relatively short terminal branch of Mag2 in *S. japonicus* suggests that this protein might have been more functionally constrained, and may have perhaps maintained a greater proportion of the ancestral function of Mag, than the Mag2 copies found in the three other *Schizosaccharomyces* species. Consequently, we can infer that Mag2's functional divergence from Mag1 occurred over a long period of evolutionary time.

Mag2's divergence may be indicative of a specialized DNA repair program in *S. pombe* compared to other fungal species and bacteria. Indeed, BER in *S. pombe* exhibits distinct features despite the conservation in protein components with *E. coli*. Firstly, the organism possesses only one bifunctional DNA glycosylase (Nth1) and no functional homologs of Fpg or NEIL enzymes (Kanamitsu and Ikeda 2010; Nilsen, Forstrom et al. 2012). Secondly, it was recently reported that unlike other organisms, repair of AP sites in *S. pombe* can be completed by a polynucleotide kinase Tdp1, independent of AP endonuclease (apn2) (Nilsen, Forstrom et al. 2012). Studies in *S. pombe* also have

revealed that sensitivity to alkylating agents do not rely solely on the glycosylase, suggesting crosstalk between various repair pathways (Memisoglu and Samson 2000; Alseth, Osman et al. 2005; Kanamitsu, Tanihigashi et al. 2007). Our evolutionary analysis leaves open the possibility that Mag2 does indeed have a function that has not yet been identified. The Mag2-DNA structure argues against a glycolytic role, but it may be that we have yet to uncover the true substrate for Mag2. It is also possible that the *in vivo* activity is modulated by a particular binding partner, as evident in other glycosylases (Campalans, Marsin et al. 2005; Fitzgerald and Drohat 2008; Baldwin and O'Brien 2009; Mutamba, Svilar et al. 2011). Although the true role of Mag2 remains to be determined, the work presented here highlights some key, previously hidden, DNA interactions necessary for stabilizing the glycosylase-DNA complex.

3.4 Experimental Procedures

3.4.1 Evolutionary analysis

The amino acid sequences of *S. cerevisiae* Mag and *S. pombe* Mag1 and Mag2 were used to query a fungal BLAST database containing more than 200 fungal genomes, as previously described (Slot and Rokas 2011). A preliminary neighbor joining tree constructed using the full sequences of the BLAST hits revealed that the Mag1/2 paralogs only occur in the *Schizosaccharomyces* clade, and that only a single Mag homolog exists in most other fungi. To accurately infer the evolutionary history of the Mag1/Mag2 duplication, an automated alignment was constructed from Mag1/Mag2 homologs in representative fungal species using MAFFT v6.952b (Katoh and Toh 2008) with settings recommended for high accuracy (maxiterate=1000 and localpair). The Mag1/Mag2 phylogeny was constructed from the alignment using the maximum likelihood optimality

criterion as implemented in RAxML 7.2.6 (Stamatakis 2006). The phylogenetic analysis was performed using the PROTGAMMALGF model, which had the best fit to the alignment data according to the ProteinModelSelection.pl test, packaged with RAxML. Robustness in phylogeny inference was assessed with 1000 bootstrap replicates and plotted on the maximum likelihood phylogeny.

3.4.2 Protein purification

The Mag2 gene was PCR amplified from *S. pombe* genomic DNA and cloned into pBG100 (Vanderbilt Center for Structural Biology) to produce an N-terminal His_6-tagged protein. Protein was overexpressed in *E. coli* BL21 cells for 16 hrs at 16° C upon addition of 0.5 mM IPTG. Cells were lysed in 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 10% glycerol and proteins isolated using Ni-NTA (Qiagen) affinity chromatography. Following cleavage of the His_6 tag by Rhinovirus 3C (PreScission) protease overnight, Mag2 was purified by anion-exchange (HiTrap Q-HP, GE Healthcare) chromatography in Buffer A (20mM Tris-HCl (pH 8.0), 1mM DTT, 0.05 mM EDTA) using a 0-1M NaCl linear gradient, followed by gel filtration (Superdex 200, GE Healthcare) in Buffer A/150 mM NaCl. Mutant proteins were generated using a Quik-Change kit (Stratagene) and purified the same as wild-type.

3.4.3 X-ray crystallography

Protein-DNA complexes were assembled by incubating 0.30 mM Mag2 with 0.36 mM DNA (d(CGGACT(THF)ACGGG)/d(GCCCGTTAGTCC) at 4° C for 20 min. Crystals were grown by vapor diffusion at 21°C against reservoir solution containing 50 mM NaHEPES (pH 7.5), 50% PEG 200, 200 mM KCl and 25 mM $MgSO_4$. Crystals were

flash frozen in mother liquor containing 60% PEG 200 prior to data collection. X-ray diffraction data (Supplementary Information) to 1.9 Å were collected at the 21-ID-F beamline at the Advanced Photon Source (Argonne, IL) and processed using HKL2000 (Otwinowski and Minor 1997). Phases were determined by molecular replacement using the protein structure of *S. pombe* Mag1 from PDB ID 3S6I as a search model. Simulated annealing and restrained coordinate refinement (Brunger, Adams et al. 1998) improved the quality of the maps and allowed visualization of the DNA, with the exception of nucleotides (C1-T6) on the damaged strand the 5′-overhang (G1) on the opposite strand. DNA atoms were built using COOT (Emsley and Cowtan 2004) and the model refined against 1.9 Å diffraction data using a maximum likelihood target in PHENIX (Adams, Grosse-Kunstleve et al. 2002). At this stage nucleotides G2-G3 and phosphates for A4-T6 were visible in 2Fo-Fc and Fo-Fc electron density maps, which allowed construction of the remaining nucleotides on the 5′-end of the damaged strand. Iterative coordinate and B-factor refinement and model building were carried out in PHENIX and COOT. Anisotropic motion of different protein/DNA fragments were modeled using translation/libration/screw-rotation (TLS) refinement. Individual anisotropic B-factors derived from the refined TLS parameters were held fixed during subsequent rounds of refinement. The final model, consisting of amino acids 5-209 and nucleotides 2-12 and one protein/DNA complex in the asymmetric unit, was validated using PROCHECK (Laskowski, Rullmannn et al. 1996). 94.1% and 5.9% of protein residues reside in allowed and additional allowed regions of the Ramachandran plot, respectively, and no residues reside in generously allowed or disallowed regions. The N9-C1′-O4′ (G2) and C3′-C4′-O4′ (G3) bond angles for nucleotides G2 and G3 on the disordered portion of the

damaged strand deviated more than 6 times the standard deviation from ideal values. Nucleic acid parameters were calculated using CURVES (Seeberg, Eide et al. 1995). The final model was deposited in the Protein Data Bank under accession number 4HSB.

3.4.4 Base excision activity

Base excision activity was measured as described for Mag1 (Chapter II). Unmodified and εA-containing oligonucleotides were purchased from Integrated DNA Technologies. The sequence tested was d(GACCACTACACC(εA)TTTCCTAACAAC)/ d(GTTGTTAGGAAAT<u>T</u>GGTGTAGTGGTC). Reactions were carried out at 25° C, 120 mM ionic strength, and pH 6.0. Reactions were terminated at various times with 0.2 M NaOH and heated at 70 °C for 2 min. Substrate and NaOH-cleaved product DNAs were separated on 15% polyacrylamide/7M urea gel and quantified by autoradiography. Experiments were performed in triplicate.

Chapter 4

CRYSTAL STRUCTURE OF ALKYLPURINE DNA GLYCOSYLASE (MAG) FROM *SACCHAROMYCES CEREVISIAE*.

Abstract

DNA glycosylases act to locate, recognize and excise damaged DNA bases that could either be mutagenic or cytotoxic if left unrepaired. *S. cerevisiae* encodes one inducible alkylpurine DNA glycosylase – scMag - which catalyzes the excision of multiple modified nucleobases including 3-methyladenine (3mA), 7-methylguanine (7mG) and 1, N^6-ethenoadenine (εA). Here we present the crystal structure of scMag, which confirms it to be a member of the helix-hairpin-helix (HhH) superfamily of DNA glycosylases. Contrary to a previous proposal, scMag lacks the TATA-box-binding-protein-like α/β domain seen in *E. coli* alkA. Structural and mutational analysis reveals that the additional ~30 amino acids (compared to its *S. pombe* homolog, spMag1) in the scMag N-terminus are important for catalytic efficiency and that the catalytic aspartate may have different role in scMag compared to spMag1. In addition, we also present crystal structure spMag1-His64Ser mutant bound to an abasic site containing DNA. Structural comparison of scMag with spMag1 and the new mutant provides an explanation for the importance of the minor groove Ser97 residue to εA specificity.

4.1 Introduction

Constant assault on DNA from endogenous and environmental agents produces covalent nucleobase adducts (Lindahl 1993). If not dealt with properly, these adducts can either be cytotoxic or mutagenic (Friedberg, Walker et al. 2006). There exist multiple pathways by which cells deal with DNA damage and restore genomic stability. Base Excision Repair (BER) is the primary pathway used to remove single base lesions. This series of highly conserved enzymatic reactions is initiated by a class of enzymes known as DNA glycosylases. DNA glycosylases are responsible for locating, selecting and excising specific DNA s and the efficacy of the entire BER pathway depends on the efficiency and specificity of these enzymes (Brooks, Adhikary et al. 2012). Action of DNA glycosylases leaves behind an abasic site which is then further acted on by an AP-endonuclease resulting in a nick in the DNA phosphodiester backbone 5' of the excised base. The nick is processed by a deoxyribosephosphatase followed by gap-filling and sealing activities of a DNA polymerase and DNA ligase, respectively (Berti and McCann 2006).

Alkylation induced DNA damage is commonly repaired by a class of DNA glycosylases known as alkylpurine DNA glycosylases. Members of this class are found in all domains of life. All alkylpurine DNA glycosylases with the exception of the human enzyme share a common structural fold and DNA binding motif known as helix-hairpin-helix (HhH) motif (Denver, Swenson et al. 2003; Brooks, Adhikary et al. 2012). This common structural fold and motif is also found in enzymes involved in repair of oxidative DNA damage (endonuclease III (endoIII), 8-oxoguanine glycosylase (OGG1) and MutY) (Brooks, Adhikary et al. 2012). The HhH motif is used to non-specifically

bind the DNA and flip the target nucleotide 180° out of the duplex and into a nucleobase pocket where the glycosidic bond cleavage takes place. The flipped-out conformation is stabilized by one or more protein residues that act as a 'plug' to fill the void left by the flipped-out base and a second residue that 'wedges' between the bases across the lesion. Recent work from our lab as well as others has shown that these residues and others in the region could be important in substrate specificity as well as overall catalytic efficiency of the enzyme (Chapter II ; (Parikh, Mol et al. 1998; Jiang, Kwon et al. 2001; Vallur, Feller et al. 2002; Livingston, Kundu et al. 2005; Maiti, Morgan et al. 2009)).

Unlike *E. coli* (alkA and TAG) and *S. pombe* (spMag1 and spMag2), the *S. cerevisiae* genome encodes only one alkylpurine DNA glycosylase – scMag. It has been reported previously that scMag excises a wide array of damaged bases including 3-methyladenine, 3-methylguanine (3mA), 7-methylguanine (7mG), 1, N^6-ethenoadenine (εA), 7-(2-chloroethyl)guanine (7-CEG), 7-(2-hydroxyethyl)guanine (7-HEG), hypoxanthine (Hx) (Wyatt, Allan et al. 1999). This wide range of substrate preference is similar to that seen in the alkA (Wyatt, Allan et al. 1999). Despite their extensive sequence similarity, DNA glycosylases from yeast differ in various ways. The *S. pombe* enzymes lack the extended N-terminus of scMag and biochemically, their substrate preference is much more restricted compared to scMag (Chapter II). In fact, spMag2 has been reported to lack any glycosylase activity at all (Chapter III and (Alseth, Osman et al. 2005; Dalhus, Nilsen et al. 2012).

Our laboratory recently determined crystal structures of spMag1 bound to DNA containing a THF abasic analog (Chapter II). There were only two notable differences between scMag and spMag1 inside the active site. In an attempt to swap εA excision

activity between spMag1 and scMag we exchanged these residues (spMag1 158F159S→SG and scMag S197G198→FS double mutants) but found that it did not affect their relative εA excision rates. This observation provided evidence that the chemical environment of the binding pocket is not responsible for the lower εA excision activity of Mag1 (Chapter II). We also noticed a protein-DNA contact (His64) in the minor groove interrogation loop of spMag1 which had never before been seen in alkylpurine DNA glycosylases. His64 is substituted with a serine in scMag and AlkA orthologs (Rubinson, Adhikary et al. 2010; Moe, Hall et al. 2012, Chapter II). Surprisingly, swapping histidine and serine between spMag1 and scMag resulted in a dramatic increase in εA excision rate in spMag1 and a corresponding decrease in that of scMag, whereas the 7mG excision rates in both enzymes remained the same (Chapter II). These results illustrated that contacts to the minor groove may be important for damage detection and/or stabilizing a specific enzyme-substrate complex for catalysis. However this conclusion was based on sequence alignment and high resolution structures of scMag and spMag1-His64Ser mutant would be necessary to verify the roles of the active site and damage interrogating residues.

Here we present a crystal structure of scMag at 1.9 Å resolution together with biochemical data that help explain a basis for the functional differences in yeast alkylpurine DNA glycosylases. Our studies show that scMag adopts the canonical HhH fold seen in alkA, spMag1 and other related enzymes, but with an insertion not evident in other family members. Mutational/deletion analysis of scMag showed that the N-terminal amino acids (1-33) are important for overall protein stability. We also present a crystal structure of spMag1-His64Ser mutant that displayed an increased εA excision activity

compared to wild-type spMag1. This structure shows that the His64Ser mutation positions the Ser residue in the position of Ser97 present in scMag. In addition it also confirms that increased activity seen in the spMag1-His64Ser mutant is due to the differences in minor groove interactions and not gross changes in the overall protein fold.

4.2 Results

4.2.1 Crystal Structure of scMag

Limited proteolysis of scMag revealed two major products (Fig. 4.1A). Mass spectrometry on the fragments showed them to result from loss of the ~21 and 41 N-terminal residues (data not shown). Crystallization attempts with these N-terminal deletion mutants failed to produce diffraction quality crystals. We then decided to use *in situ* proteolysis which has been previously reported to aid in crystallization of recalcritant proteins (Dong, Xu et al. 2007). Crystals of selenomethionine-substituted scMag appeared overnight when chymotrypsin (1/1000 w/w) was added to crystal setups. An atomic model of scMag was built into electron density maps obtained from SAD X-ray diffraction data collected at the Se-absorption peak wavelength. The atomic model was refined to crystallographic residual of 17.2 % (R_{work}, 20 % R_{free}) (Table 7). We confirmed that scMag was active under the *in situ* proteolytic conditions, since scMag shows base excision activity against εA in presence of 1:1000 molar ratio of chymotrypsin and incubated overnight (Fig. A7).

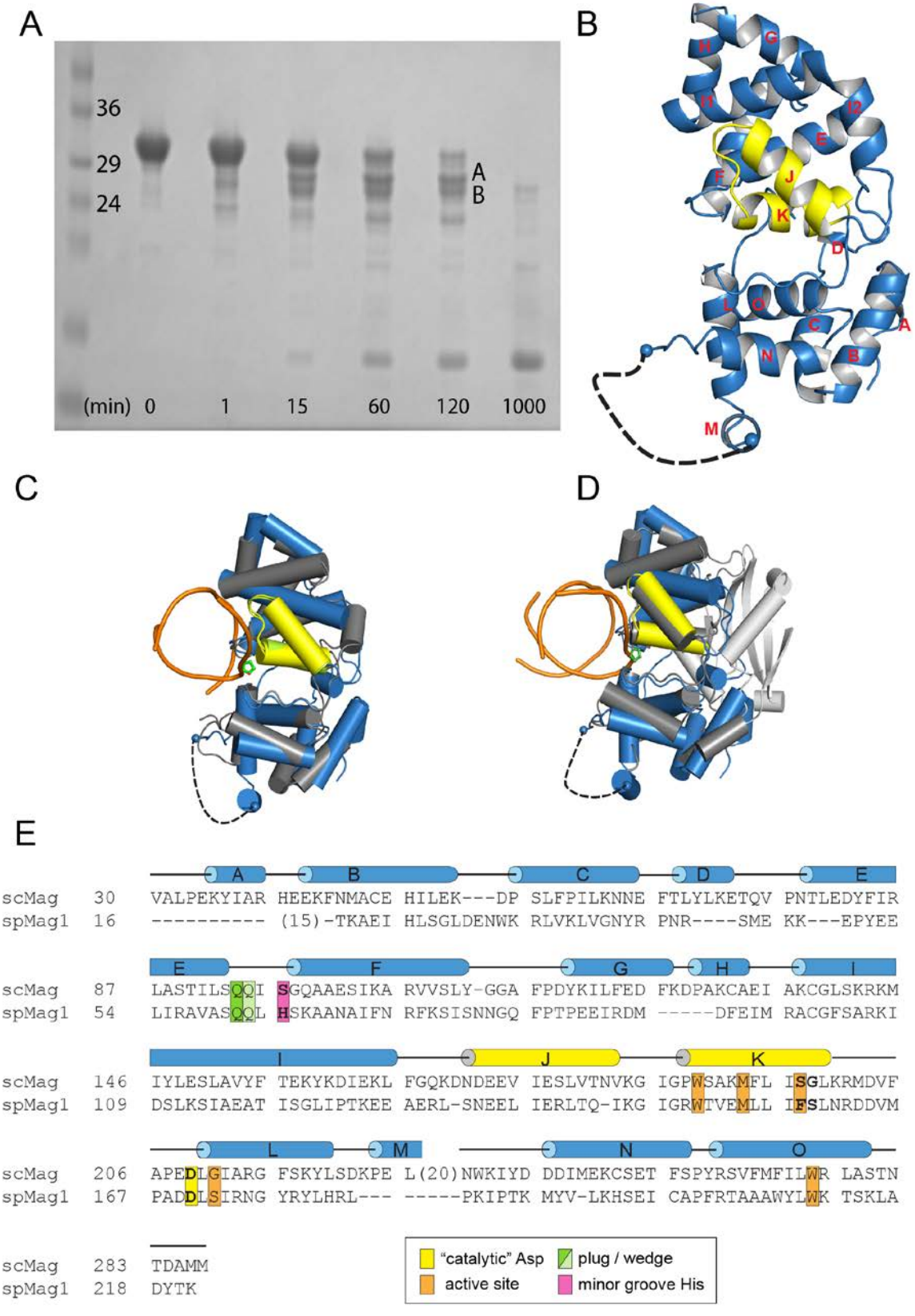

A
36
29
24
A
B
(min) 0 1 15 60 120 1000
B
C
D
E
scMag 30 VALPEKYIAR HEEKFNMACE HILEK---DP SLFPILKNNE FTLYLKETQV PNTLEDYFIR
spMag1 16 ---------- (15)-TKAEI HLSGLDENWK RLVKLVGNYR PNR----SME KK---EPYEE
scMag 87 LASTILSQQI SGQAAESIKA RVVSLY-GGA FPDYKILFED FKDPAKCAEI AKCGLSKRKM
spMag1 54 LIRAVASQQL HSKAANAIFN RFKSISNNGQ FPTPEEIRDM -----DFEIM RACGFSARKI
scMag 146 IYLESLAVYF TEKYKDIEKL FGQKDNDEEV IESLVTNVKG IGPWSAKMFL ISGLKRMDVF
spMag1 109 DSLKSIAEAT ISGLIPTKEE AERL-SNEEL IERLTQ-IKG IGRWTVEMLL IFSLNRDDVM
scMag 206 APEDLGIARG FSKYLSDKPE L(20)NWKIYD DDIMEKCSET FSPYRSVFMFILWR LASTN
spMag1 167 PADDLSIRNG YRYLHRL--- -----PKIPTK MYV-LKHSEI CAPFRTAAAWYLWK TSKLA
scMag 283 TDAMM
spMag1 218 DYTK
"catalytic" Asp
plug / wedge
active site
minor groove His

Fig. 4.1 Crystal structure of scMag. (A) SDS-PAGE of scMag subjected to limited proteolysis showing the two major products (A and B) of chymotrypsin treatment for various time points (in minutes) at 1:1000 protein: protease concentration. (B) Structure of scMag is presented is ribbon diagram. The helix-hairpin-helix (HhH) motif is shown in yellow. (C) Overlay of scMag (blue) with spMag1 (PDB ID 3S6I, grey). DNA from the spMag1 structure is colored orange. (D) Overlay of scMag with alkA (grey) with DNA from the *E. coli* alkA structure shown in orange (PDB ID 1DIZ). The extended N-terminal amino acids from alkA form a TATA-box-binding protein-like domain are colored silver. (E) Structure-based sequence alignment of scMag and spMag1. The schematic of the secondary structure elements from scMag are shown above the sequence. Residues lining the extrahelical base binding pocket are highlighted orange, and conserved aspartic acid residues important for catalysis in HhH glycosylases are highlighted yellow. DNA intercalating plug and wedge residues are highlighted green and light green, respectively, and the novel minor-groove interrogating histidine residue observed in the spMag1 structure is highlighted pink.

The refined structure of proteolyzed scMag revealed a predominantly helical protein comprised of 15 α-helices (A-O) (Fig. 4.1B and Table 7). Consistent with other related DNA glycosylases, the structure is divided into two lobes with helices A-D and L-O comprising the catalytic residues and helices E-K containing the HhH motif and the minor-groove interrogating residues (Labahn, Scharer et al. 1996; Chapter II-III) (Fig. 4.1B-E). The overall structure shows remarkable similarity with both spMag1 (Fig. 4.1 C) and alkA (Fig. 4.1D). The HhH motif is perfectly positioned to interact with the DNA as shown by the superimposition with spMag1 and alkA (Fig. 4.1 C,D). Probable effects of proteolysis can be observed at the termini where the first 30 and the last 10 amino acids are not visible. In addition, a 20 amino acid insertion is also not observable in the electron density maps (Figs. 4.1E, 4.2). Interestingly, this region also features a crystal contact between two protein molecules in the asymmetric unit. The loop formed by helices E-F juts into the space where the missing insertion would be expected for the symmetry-mate. This contact would be improbable with an intact loop.

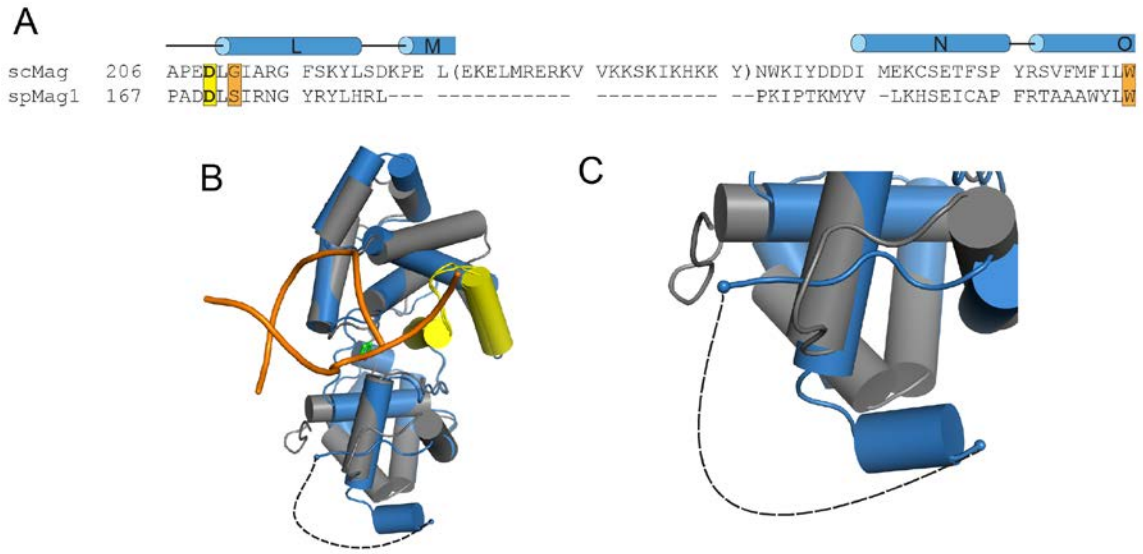


Fig. 4.2 *In situ* proteolysis of scMag. (A) Structure-based sequence alignment between scMag and spMag1 of the missing region (shown in parentheses). Color schemes are same as Fig. 4.1 E. (B) Overlay of scMag structure (blue) on spMag1 (gray and orange DNA) with the missing internal amino acids marked with a dotted line. (C) Close-up of the missing region.

4.2.2 N-terminal extension in alkA and scMag

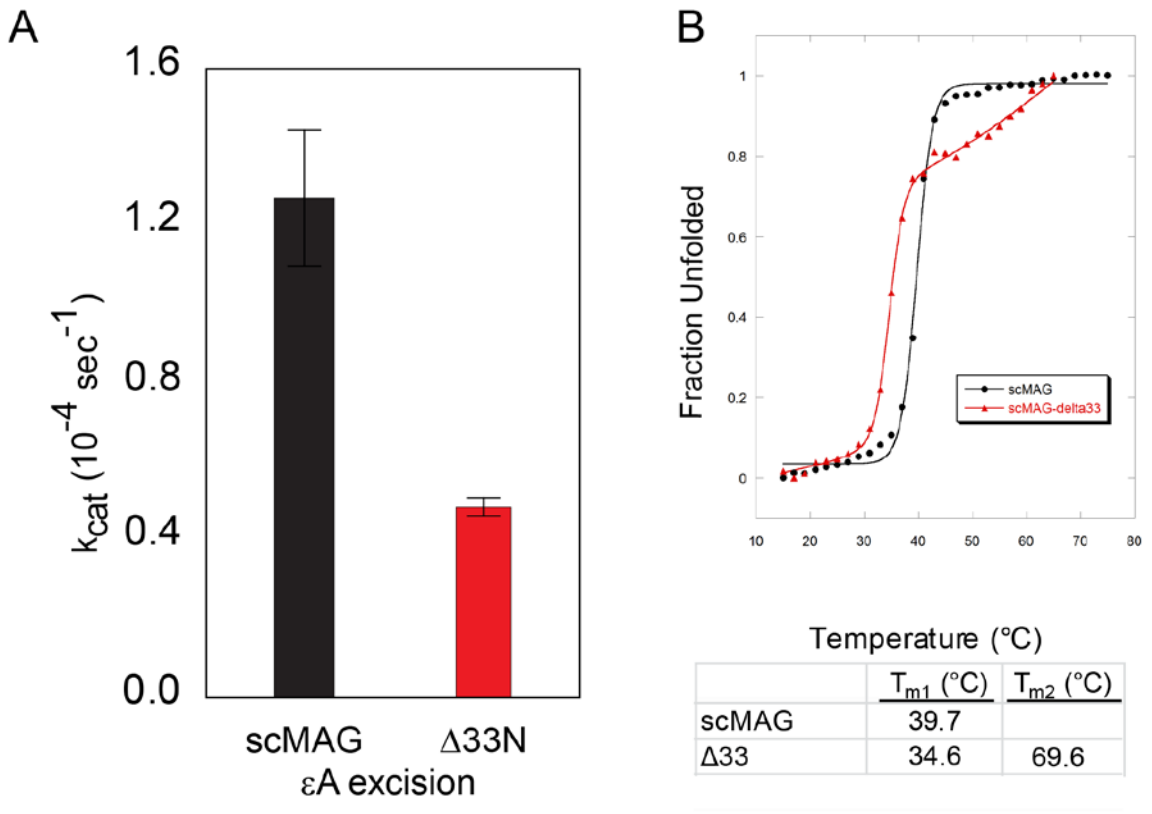
Among 3-methyladenine DNA glycosylases, alkA and scMag have an extended N-terminal region with no attributed function (Figs. 4.1E and (Labahn, Scharer et al. 1996)). In *E. coli* alkA, the N-terminal 89 amino acids form a mixed α-β structure reminiscent of the tandem repeats of the TATA-box binding protein (Labahn, Scharer et al. 1996). However, this similarity is thought to be a consequence of preferred topological arrangement of the helices and β-sheets rather than a functional connection (Labahn, Scharer et al. 1996). Interestingly, most of the N-terminal extension (AA's 1-29) in scMag is not visible in the current crystal structure (Fig. 4.1B).

Table 7. Crystallographic data collection and refinement statistics

	scMag	spMag1-His64Ser
Data collection*		
Wavelength (Å)	0.9787	0.9787
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	34.5, 131.6, 56.5	53.09 62.75 97.84
α, β, γ (°)	90, 107.7, 90	90.00 91.01 90.00
Resolution (Å)	49.8-1.9 (1.97-1.90)	30-2.75 (2.95-2.85)
R_{sym}	0.08 (0.34)	0.129 (0.717)
I / σ_I	4.7 (3.8)	10.1 (1.5)
Completeness (%)	93.1 (75.3)	96.2 (97.0)
Redundancy	6.2 (6.3)	5.2 (4.9)
Refinement		
Resolution (Å)	1.90	2.85
No. reflections	35558	16182
R_{work} / R_{free}	0.17/0.2	0.25/0.3
No. of atoms		
Protein	3925	3184
DNA	0	845
Solvent	106	12
R.m.s. deviations		
Bond lengths (Å)	0.008	0.01
Bond angles (°)	1.030	1.452

*Values in parentheses refer to the highest resolution shell.

Originally thought be similar to the N-terminal extension of alkA, this region in scMag is predicted to be predominantly disordered with some helical elements (Lingaraju, Kartalou et al. 2008). We generated an N-terminal deletion mutant (Δ33) to test the role of this extension in base excision activity. Compared to the wild-type protein, the excision rate for this mutant was more than 3-fold lower for εA (Fig. 4.3A) Although additional structural and biochemical data will be needed to precisely understand the role of these residues, test of structural integrity of the deletion mutant suggests that the decrease in activity is likely due to the loss of overall stability of the protein (Fig. 4.3B).



	T_{m1} (°C)	T_{m2} (°C)
scMAG	39.7	
Δ33	34.6	69.6

Fig. 4.3 N-terminal extension in scMag. (A) εA excision activity of wild-type and Δ33 mutant scMag proteins. (B) Thermal denaturation of wild-type and Δ33 mutant scMag. Protein unfolding was monitored using CD spectroscopy, measuring the change in molar ellipticity at 222 nm as a function of temperature.

4.2.3 Comparison of the active sites of scMag and spMag1

Our previous mutational analyses within the scMag active site were based on a sequence alignment with AlkA since the scMag structure was not yet available (Chapter II). The availability of scMag structure allowed us to compare the active site environment between scMag and spMag1. As predicted, the structural alignment shows that the active site environment is very similar to that seen for spMag1 with only noticeable difference being scMag residues Ser197, Gly198 and Gly211. SpMag1 has Phe158, Ser159 and 172Ser in the corresponding positions (Figs. 4.1E and 4.4).

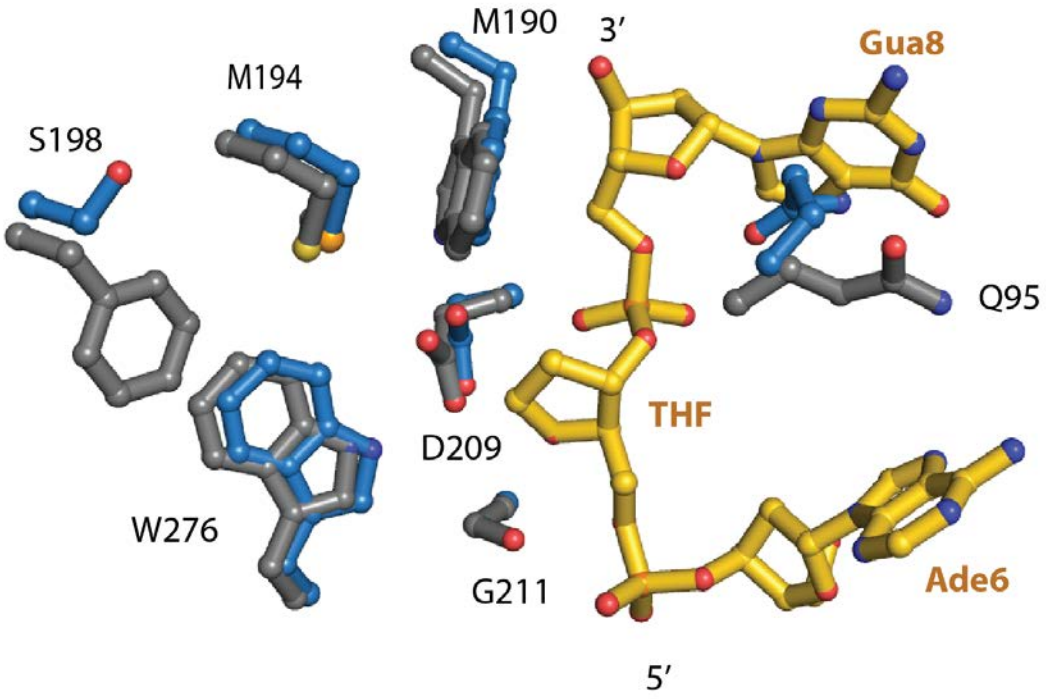


Fig. 4.4 ScMag nucleobase binding pocket. Overlay of the base binding pockets for scMag (blue sticks) with that of spMag1 (gray) with the DNA from the spMag1 structure shown as yellow sticks. Side chain information is from the scMag structure.

4.2.4 Role of minor groove interrogating residue in εA excision

We previously showed a His64ser mutation in spMag1, which mimics the Ser97 residue in scMag, increased the εA excision activity of spMag1 relative to wild-type spMag1 (Chapter II). However, whether the mutation had an unforeseen structural impact on the protein was unclear. To ascertain the structural impact of His→Ser substitution at this position we determined a crystal structure of the Mag1-His64Ser mutant bound to DNA containing an abasic site (Fig. 4.5A). The structure confirms that the mutation does not exert global changes in the overall fold as the protein (wild-type and mutant) atoms align with an

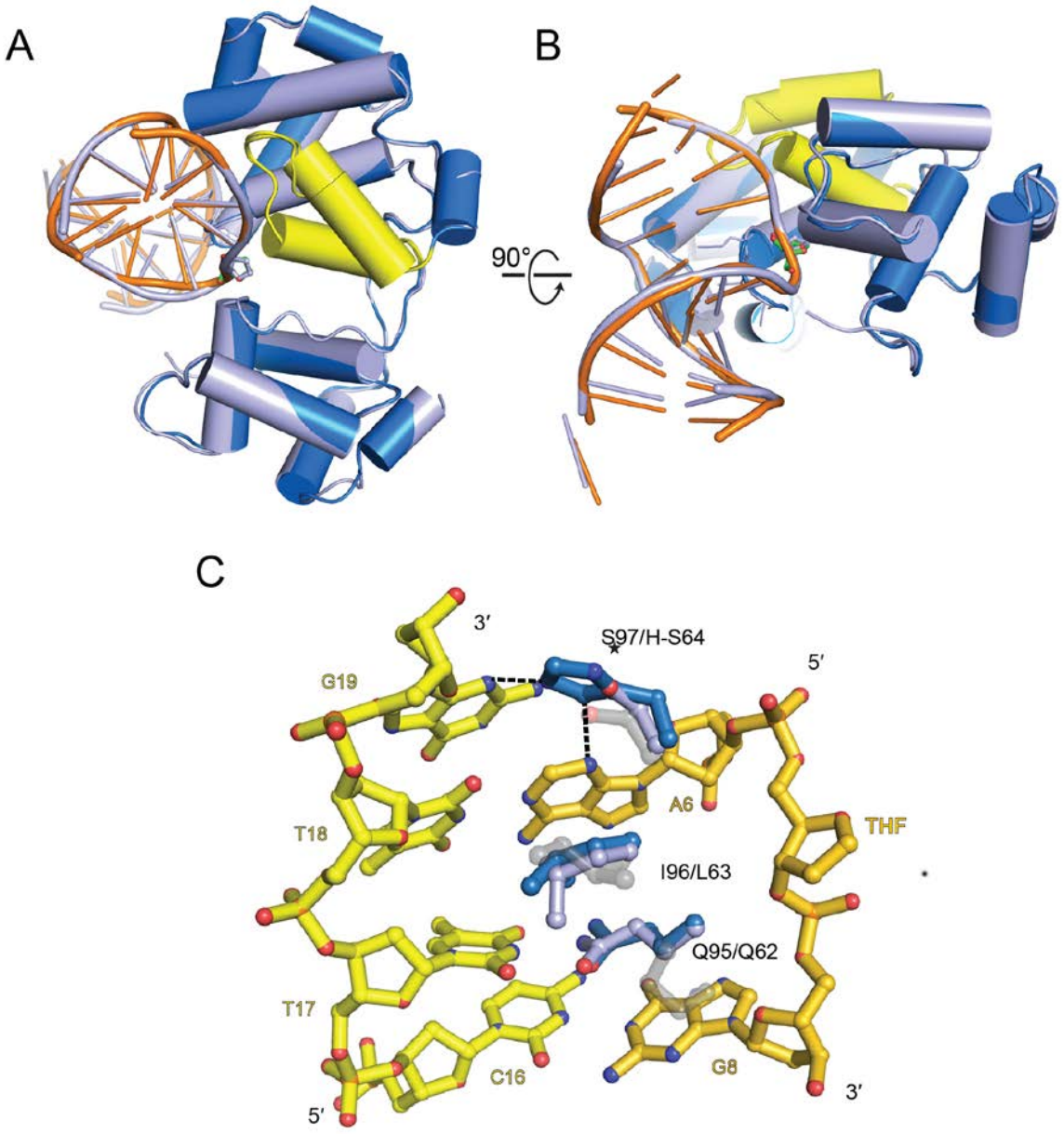


Fig. 4.5 Crystal structure of Mag1-His64Ser. (A) Overlay of Mag1-His64Ser-DNA mutant (light blue) on Mag1-wt (sky blue with orange DNA, 3S6I). (B) 90° rotation around the Y-axis to show the superimposition of DNAs between the two structures. (C) Close-up of the minor-groove intercalating residues. Amino acids from Mag1-His64Ser (light blue), Mag1-wt (sky blue) and Mag (gray) are superimposed with DNA from the Mag1-wt structure.

average r.m.s.d. of < 1 Å. The overall DNA structure also lines up almost identically to one observed in the wild-type protein/DNA structure and the position of the abasic site

remains unchanged (Fig. 4.5B). The structure shows that the serine substitution abolishes the two hydrogen bonded contacts made by His64 to Gua19 and Ade6 (Fig. 4.5C). It is worth noting that Ser64 in the spMag1-His64Ser structure aligns comparably with Ser97 observed in the scMag structure (Fig. 4.5C). This finding supports the inhibitory role of His64 either as steric hindrance or a damage sensor (Chapter II).

4.3 Discussion

In addition to completing our knowledge of the structures of yeast 3-methyladenine DNA glycosylases, the crystal structure of scMag provides an opportunity to test the determinants of structural and biochemical differences between it and other DNA glycosylases. Although further structural work, i.e., substrate/product bound structure, will be necessary to fully understand the basis of substrate specificity, this work suggests that the N-terminal extension seen in scMag is important for structural integrity and is not involved directly in catalysis. The study also adds to emerging research in the field that determinants of substrate specificity may lie outside the base binding pocket and that base recognition could take place before it is flipped into the active site.

In addition to the HhH and glycosylase domains, *E. coli* alkA also contains an N-terminal domain that shares structural homology with TATA-box binding protein (Labahn, Scharer et al. 1996). This domain is not involved in DNA binding or glycosylase activity and has not been observed in other alkylpurine DNA glycosylases (Labahn, Scharer et al. 1996). Interestingly, sequence alignment between spMag1 andscMag shows extra 35 amino acids on the N-terminus of scMag. This region has previously been suggested to be similar to the TATA-box-binding protein-like domain of alkA (Lingaraju, Kartalou et al. 2008). The structural and biochemical characterization of

scMag presented here suggests that the extra amino acids in scMag perform a different function than the domain seen in alkA. The domain seen in alkA is composed of 89 amino acids compared to 35 present in scMag. More importantly deleting these amino acids has a significant effect on the glycosylase activity of scMag which is not the case for alkA.

In situ proteolysis can be used to remove disordered and/or solvent exposed portions of the target protein, thereby increasing the favorable interactions between molecules to aid in crystal nucleation (Dong, Xu et al. 2007). The crystal structure of scMag supports this notion, as the entire DNA interacting portion is fully intact and ~81% of the protein residues are observed. Action of the protease removes the first 29 and last 9 residues. This was expected, as secondary structure prediction softwares and limited proteolysis experiments had shown these regions to be disordered and/or susceptible to proteolysis (Fig. 4.1A and (Slabinski, Jaroszewski et al. 2007)). There was however, an unexpected stretch of missing amino acids (230-247) between helices M and N. This insertion is conspicuously missing in other related DNA glycosylases (Fig. A2) and is predicted to form an α-helix. This is the most obvious difference between the structures of spMag1 and scMag, and it remains to be determined if this region has any biological/biochemical function. Our efforts at generating a truncated form of scMag with these residues deleted did not yield soluble protein (data not shown), indicating the structural importance of this region. Further study will be required to parse out the exact role of these missing residues in scMag.

With structural and biochemical data available on alkA and spMag1, we previously had tested comparable residues from scMag to characterize their role in

enzyme activity. The first residue we compared was the highly conserved aspartic acid residue. In most of the related glycosylases, including alkA, mutating this residue results in complete loss of base excision activity (Labahn, Scharer et al. 1996; Eichman, O'Rourke et al. 2003; Chapter II). Unsurprisingly, the scMag Asp209Asn mutation resulted in almost complete loss of activity against 7mG and εA (Chapter II). This result confirmed the role of this residue in hydrolysis of the glycosidic bond. However, the comparable mutation in spMag1 (Asp168Asn) had a significantly less severe effect (Chapter II). The reduced role of the 'catalytic' residue was also noted for the MagIII DNA glycosylase from *Helicobacter pylori* (Eichman, O'Rourke et al. 2003). Moreover, the highly specific TAG enzyme from *E. coli* has no comparable catalytic residue at all (Metz, Hollis et al. 2007). The comparison of spMag1 and scMag combined with existing structural and biochemical work suggest that the glycosylases that have wider substrate preference are much more reliant on the catalytic power of the general base residue in the base binding pocket.

4.4 Experimental Procedure

4.4.1 Cloning and Protein Purification

MAG from *S. cerevisiae* was amplified from genomic DNA (ATCC: 204511) using primers d(TGACGTGGATCCATGAAACTAAAAAG) and d(GCCGCGAATTCTTAGGATTTCACGAAA). The amplified PCR product was cloned into pBG100 expression plasmid which produces an N-terminal cleavable His_6-tagged protein. *E. coli* BL21 cells harboring the expression plasmid were grown to an OD_{600} ~ 0.5 at 37°C and protein production was induced after equilibrating the cultures at 16°C for 30 minutes by adding 0.05 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Cells

were harvested in 50 mM Tris.HCl pH 7.5, 500 mM NaCl and 10 % glycerol and lysed using three passes through an Emulsiflex cell homogenizer set at 25000 psi. The protein was purified to homogeneity using Ni-NTA affinity followed by ion exchange (HiTrap Q, GE) and size exclusion (Sephadex 200, GE Healthcare) chromatography. Selenomethionine-substituted scMag was expressed using metabolic inhibition as described previously (Chapter II). All scMag mutant proteins were purified similar to the wildtype protein. Mutant Mag1 from *Schizosaccharomyces pombe* was cloned and purified as described previously (Chapter II).

4.4.2 *In situ* proteolysis-aided crystallization.

Diffraction quality crystals of scMag were obtained using *in situ* proteolysis as described previously with some modifications (Dong, Xu et al. 2007). 1:1000 (w/w) chymotrypsin or trypsin was added to 10 mg/mL selenomethionine-substituted scMag immediately before setting up crystallization trials. Highest quality crystals were seen after ~20 hours in 2.5-3.0 M $(NH_4)_2SO_4$ and 0.1M HEPES pH 7.5. Crystals were cryoprotected in 3.2 M $(NH_4)_2SO_4$ and 0.1M HEPES pH 7.5 before being flash-frozen.

4.4.3 Data Collection and Refinement

X-ray diffraction data was collected at the Advanced Photon Source (Argonne, IL) LS-CAT beamline 21-ID-D. Single Anomalous Dispersion (SAD) data were collected at the selenium absorption peak. Electron density for the protein was calculated using the positions of the Se atoms using the PHENIX suite AutoSol feature (Adams, Afonine et al. 2007). An initial crystallographic model corresponding to amino acids 35-283 from two copies of the protein in the asymmetric unit was built by the AutoSol program. This model was further improved by manual inspection of σ_A-weighted 2 mF_o-

F_c and mF_o- DF_c electron density maps using Coot (Emsley and Cowtan 2004) and refined using a maximum likelihood target as implemented in PHENIX (Adams, Afonine et al. 2007) to a crystallographic residual of 17.5 % (R_{free} = 20%). The final model was validated using PROCHECK (Laskowski, Rullmannn et al. 1996)

4.4.4 Enzymatic Activity

Base excision activities were measured by following the alkaline cleavage of the abasic DNA product of alkylbase excision from a 25-mer oligonucleotide duplex containing a centrally positioned εA•C or 7mG•C base pair. Oligonucleotides were purchased from Integrated DNA Technologies (IDT, USA), and 7mG-DNA was prepared enzymatically as previously described (Asaeda, Ide et al. 2000; Rubinson, Metz et al. 2008). Nucleotide sequences used were 5′-^{32}P-d(GACCA CTACA CC**X**TT TCCTA ACAAC) annealed to 5′-d(GTTGT TAGGA AACGG TGTAG TGGTC). Reaction mixtures (75 µL) contained 10 µM enzyme, 100 nM FAM-labeled DNA duplex, 100 mM KCl, 2 mM DTT, 1mM EDTA, and 50mM sodium acetate (pH 6.0). Under these conditions, the enzyme concentration is saturating. Reactions were initiated by addition of enzyme and incubated at 25 °C. Aliquots (8 µL) were stopped at various time points by addition of 0.2 M NaOH and heated at 70 °C for 2 min. The cleaved 12-mer product and unreacted 25-mer substrate oligonucleotides were separated by 15% polyacrylamide/7M urea gel electrophoresis and quantified using Typhoon scanner (GE Lifesciences). Fraction product (F_P) at each time point was calculated by $F_P = I_P/(I_S+I_P)$, where I_S and I_P are the integrated intensities of substrate and product bands. Rate constants (k_{cat}) were determined from single-exponential fits to the data from three separate experiments.

4.4.5 Thermal Denaturation

Protein unfolding was monitored using CD spectroscopy, measuring the change in molar ellipticity at 222 nm as a function of temperature as described in chapter II. Molar ellipticity values were normalized to reflect the fraction of unfolded protein and plotted against temperature. The melting temperatures were determined by fitting the data using the equation $f_u = [(A_f-A_u)/(1+e^{-(T-Tm)/w})]$, in which f_u is the fraction of unfolded protein, A_f and A_u are the CD values for folded and unfolded protein respectively, T_m is the melting temperature, and *w* is the cooperativity coefficient for the transition. For curves showing bi-phasic transitions, melting temperatures for each transition were determined by fits to the following equation: $f_u = [(\Delta A_1)/(1+e^{-(T-Tm1)/w1})] + [(\Delta A_2)/(1+e^{-(T-Tm2)/w2})]$, in which ΔA refers to the change in CD signal for a particular transition.

Chapter 5

DISCUSSION AND FUTURE DIRECTIONS

DNA is subject to modification from cellular metabolites and industrial and environmental agents. Given the dire consequences of DNA damage, all organisms possess a multitude of repair pathways to maintain genomic stability. The majority of single nucleobase adducts are repaired by the base excision repair pathway (BER). Chapters II-IV detailed structural and biochemical characterization of three enzymes that initiate BER.

There is a relative dearth of structural and biochemical information on DNA glycosylases from *S. cerevisiae* and *S. pombe*. In addition to providing three new high resolution crystal structures, my work on yeast alkylpurine DNA glycosylases has provided novel insights into the mechanism through which the enzymes search for damaged bases, substrate specificity (Chapter II and IV) and fundamental requirements for productive DNA binding and catalytic activity (Chapter III). In this chapter, I will summarize the findings in a broader context and discuss potential follow-up studies.

5.1 Substrate specificity in DNA Glycosylases

Crystal structures of DNA glycosylases in complex with DNA containing substrate or substrate-mimics show base binding pockets that are sculpted to accommodate a particular nucleobase and exclude others. Based on these studies, the base binding pockets have been proposed to confer a strict substrate preference (Bruner, Norman et al. 2000; Parikh, Walcher et al. 2000; Eichman, O'Rourke et al. 2003; Metz,

Hollis et al. 2007). Conversely, an open base binding pocket has been interpreted as the source of relaxed substrate specificity (Hollis, Ichikawa et al. 2000; Lau, Wyatt et al. 2000). Taken together, these studies have contributed to the notion that the active site environment solely dictates substrate specificity in DNA glycosylases (Rubinson, Adhikary et al. 2010).

The discrepancy in εA excision observed in two homologous yeast alkylpurine glycosylases, spMag1 and scMag, presented an opportunity to test this hypothesis. Interestingly, crystal structures of spMag1 and scMag showed virtually identical base binding pockets that can both easily accommodate the flipped εA base (Chapters II, IV). Moreover efforts at switching substrate preference by mutating active site residues proved unsuccessful. More surprising was the discovery that a single residue (spMag1-His64) in the minor-groove interrogating loop, a region that also contains the plug/wedge residues, was able to dictate preference for εA. This is, to my knowledge, the first evidence of a gain-of-function mutation away from the base binding pocket in alkylpurine DNA glycosylase. The structural and biochemical comparisons of spMag1 and scMag presented in this thesis have shown that residues away from the base binding pocket can be important in dictating preference for damaged nucleobase. Moreover, they illustrate that DNA glycosylases can in fact 'scan' the duplex for damage and select a substrate even before flipping it into the active site. Section 5.2 will address how the role of His64 fits with its proposed role as a damage sensor.

5.1.1 Future directions – substrate complex and *in vivo* studies

High resolution structures of DNA glycosylases in complex with substrate DNA will be very important to fully understand the basis of substrate selection and to design

potential inhibitors. As of now there is only one crystal structure of an alkylpurine glycosylase-substrate complex, i.e., AAG in complex with εA-containing DNA (Lau, Wyatt et al. 2000). Although the structure showed how the enzyme sculpts the DNA to flip the base out of the helix and into a snug base binding pocket, the only damage specific contact with εA was a hydrogen-bond interaction between the amide nitrogen on His136 and N^6 of εA. Thus, the preference of the enzyme for εA and hypoxanthine over other damaged base was not entirely clear from the structure (O'Brien and Ellenberger 2004).

The availability of three new alkylpurine glycosylase crystal structures now allows us to design future crystallization studies to capture substrate bound and late reaction intermediates. Catalytically inactivating mutation of the conserved aspartate or nonhyrdolyzable substrates can be used for this purpose (Lee, Bowman et al. 2008; Rubinson, Gowda et al. 2010). DNA bound structures of spMag1 and spMag2 also can be used to design crosslinking experiments to obtain stable protein-DNA complexes of advanced intermediates suitable for crystallization studies. This method has served very well for a number of other DNA binding proteins (Banerjee, Yang et al. 2005; Banerjee, Santos et al. 2006; Radom, Banerjee et al. 2007; Yang, Yi et al. 2008; Qi, Spong et al. 2010).

Despite being homologous, spMag1 and scMag have been reported to have markedly different biochemical properties (See Introduction section 1.5.5.5). With the structural information at hand, we can begin to introduce substitution and deletion mutants into *S. pombe/S. cerevisiae* cells for *in vivo* functional analysis. For example, the

effect of deleting the additional internal helix in scMag, the spMag2 gain-of-function mutant (Asp56Ser) can be investigated *in vivo*.

5.2 Duplex interrogation and damage location

One of the main outstanding questions in the field pertains to how DNA glycosylases locate and and maintain exquisite specificity for a certain DNA lesion in ~70,000-fold overabundance of normal DNA. The currently accepted model is that of an enzyme using a combination of 3D diffusion and short-distance sliding to translocate on DNA (Zharkov and Grollman 2005; Friedman and Stivers 2010). This model is consistent with the inhibitory role of His64 seen in spMag1, where substituting the residue with serine increased the εA excision activity of the enzyme. His64 is located immediately C-terminal to the plug and wedge residues. The effect of helix intercalating residues had been reported previously but mutating them showed predominantly inhibitory effects on catalysis (Parikh, Mol et al. 1998; Jiang, Kwon et al. 2001; Vallur, Feller et al. 2002; Eichman, O'Rourke et al. 2003; Livingston, Kundu et al. 2005; Maiti, Morgan et al. 2009). Thus the increased preference for εA shown by spMag1-His64Ser mutatnt represents, for the first time, the role of minor groove intercalating residue in imparting substrate selectivity.

This discovery presents an opportunity to probe the role of these DNA interrogating residues in base flipping and formation of a catalytically competent complex. Tryptophan fluorescence and fluorescently active nucleotides have been used to understand the rates of flipping in human alkylpurine glycosylase AAG and can be modified to investigate the role of His64 as steric inhibitor and/or minor-groove scanner (Wolfe and O'Brien 2009; Hendershot, Wolfe et al. 2011). If His64 indeed functions as a

steric hindrance, the wild-type spMag1 should show a slower rate of εA-flipping compared to scMag and spMag1-His64Ser mutant.

5.3 Curious case of Mag2

In Chapter II, I surmised that lack of glycosylase activity reported for spMag2 could be partially explained by the inability of the protein to form a catalytically competent complex with DNA. In addition, a single mutation (Asp56Ser) was able to impart the protein with detectable εA excision activity. This work has further highlighted the importance of protein-DNA contacts in the minor-groove intercalating loop in the activity of DNA glycosylases.

Concurrent to the publication of the article corresponding to Chapter III, Dalhus *et al.* published structures of spMag2 bound to abasic site-containing DNA (Dalhus, Nilsen et al. 2012). The study confirms that expression of spMag2 is induced upon DNA damage, but the protein is not an active DNA glycosylase and does not flip the abasic site. However, unlike the structure reported in Chapter III, the region of DNA 5′ to the THF moiety is fully ordered. The authors also contend that Mag2 sculpts the THF containing DNA to induce a 'wider' minor groove, reminiscent of DNA conformation seen in the MutSβ-DNA structure (a protein involved in mismatch repair, see section 1.2.3). This observation coupled with electrophoretic mobility shift assay (EMSA) data showing that presence of spMag2 increases binding of MutSβ to DNA, was used to suggest that spMag2 plays the role of shuttling abasic sites to the MMR pathway. This is an intriguing idea and as the authors mention, has been observed in alkyltransferase-like proteins (ATLs) where the protein binds O^6-methylguanine lesions and presents it to the NER machinery (Tubbs, Latypov et al. 2009; Tubbs and Tainer 2010). It must be noted

that given the extremely transient nature of the interaction between the protein and the abasic site compared to ATLs, further study will be required to verify the role of spMag2 in damage protection.

The ability of the spMag2-Asp56Ser mutant to excise εA provides an excellent opportunity to test the current hypothesis of catalysis in DNA glycosylases. As reported in Chapter III and (Dalhus, Nilsen et al. 2012), spMag2 does not flip the abasic site toward the active site. Therefore if base hydrolysis requires base flipping, the Asp56Ser mutant has acquired this ability. Crystal structures of spMag2 bound to εA containing DNA would help answer this question and better understand the fundamental requirement(s) of being a catalytically active DNA glycosylase.

5.4 Phylogenetic analysis of yeast alkylpurine DNA glycosylases

In an attempt to understand the lack of base excision activity seen in spMag2, we performed a phylogenetic analysis of Mag from *S. cerevisiae* and spMag1 and spMag2 from *S. pombe* (Chapter III). One of the surprising findings of the study was that despite their orthologous functions, scMag and spMag1 have experienced different rates of amino acid substitution. Mag has evolved at a rate faster than that observed for spMag1. This observation suggests that spMag1 has retained more of the ancestral function compared to scMag. Whereas scMag has much more robust activity against a wide spectrum of alkylated bases, spMag1 is limited to excising relatively labile and positively charged lesions (Wyatt, Allan et al. 1999). It is therefore interesting to speculate that Mag has evolved its stronger glycosylase activity via selective pressure from being the sole enzyme responsible for removing nucleobase adducts produced by alkylating agents in *S. cerevisiae*. Furthermore, repair of alkylation damage in *S. pombe* may involve crosstalk

between multiple pathways. The latter argument also is supported by genetic studies showing crosstalk between rad13 (NER) and rhp51 (recombination repair, RR) (Memisoglu and Samson 2000).

Schizosaccharomyces is also the only clade with two copies of *mag*-related genes. The second copy, *mag2*, lacks excision activity but has been conserved with an accelerated rate of substitution for over 200 million years (Chapter III and (Alseth, Osman et al. 2005)). Taken together, these observations suggest not a glycolytic but possibly an undiscovered function for spMag2. Given the protective role proposed by Dalhus et al., it would be interesting to test whether rates of evolution of proteins involved in MMR, especially MutSβ, or other pathways cluster with that seen for spMag2. Results could identify a pathway that may have coevolved to accommodate the novel function of spMag2.

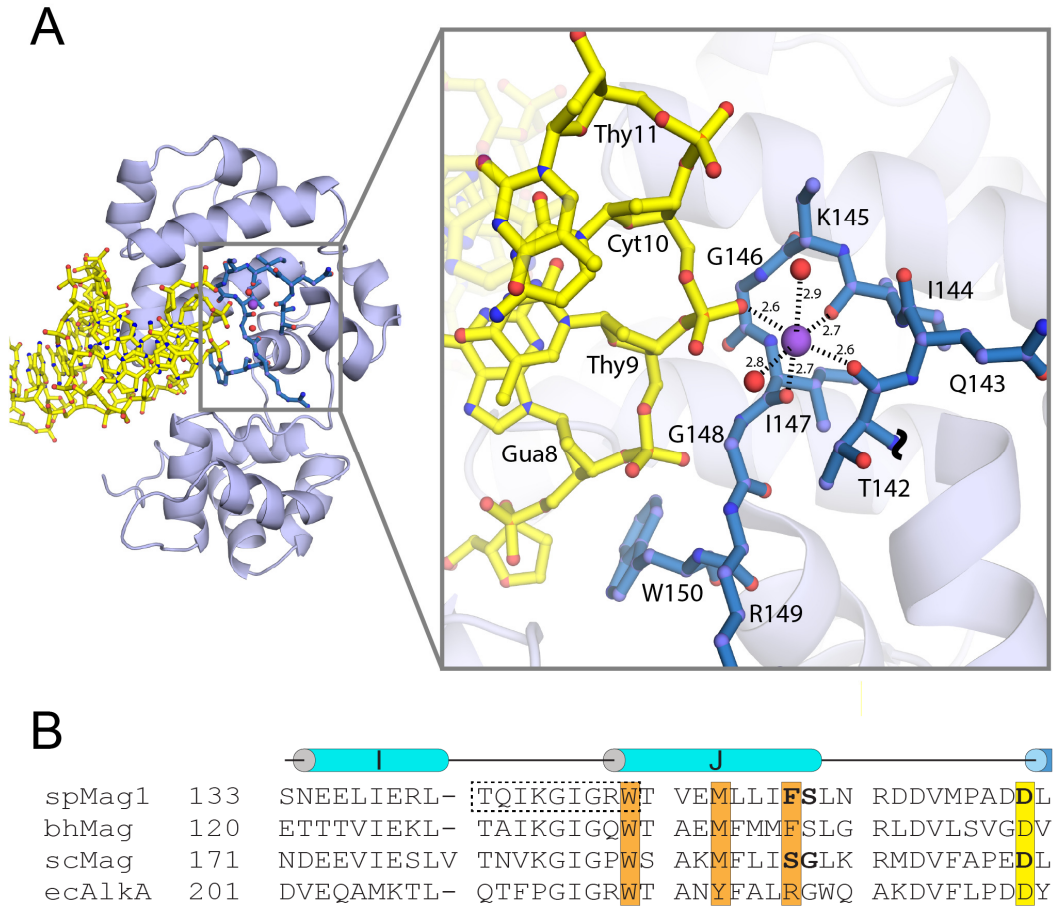


Fig. A1 Helix-hairpin-helix mediated protein-DNA contact in spMag1. (A) Contacts between the HhH motif (helices αI-αJ) and the DNA (yellow) backbone are mediated by the hairpin loop (blue sticks), and are largely conserved among AlkA, EndoIII and hOgg1 protein-DNA structures (Kuo, McRee et al. 1992; Bruner, Norman et al. 2000; Hollis, Ichikawa et al. 2000). The hairpin turn ligates a metal ion, which we have modeled as a sodium ion (purple sphere) on the basis of octahedral coordination and bonds lengths, shown as dashed lines and distances in Angstroms. The sodium ion is coordinated by main-chain carbonyl oxygen atoms of Thr142, Ile144 and Ile147, the O1P phosphate oxygen of Cyt10, and two water molecules. In addition to metal ion mediated contacts, the Cyt10 phosphate is anchored to the protein through a hydrogen bond from main-chain Gly146 nitrogen. The adjacent phosphate group from Thy9, immediately 3′ to the THF abasic site, is hydrogen bonded to Gly148 nitrogen and the side-chain O_γ oxygen from Thr151. (B) Structure based sequence alignment between Mag proteins with AlkA. Hairpin atoms are enclosed by a dashed box, and base binding pocket and catalytic aspartate residues are highlighted orange and yellow, respectively.

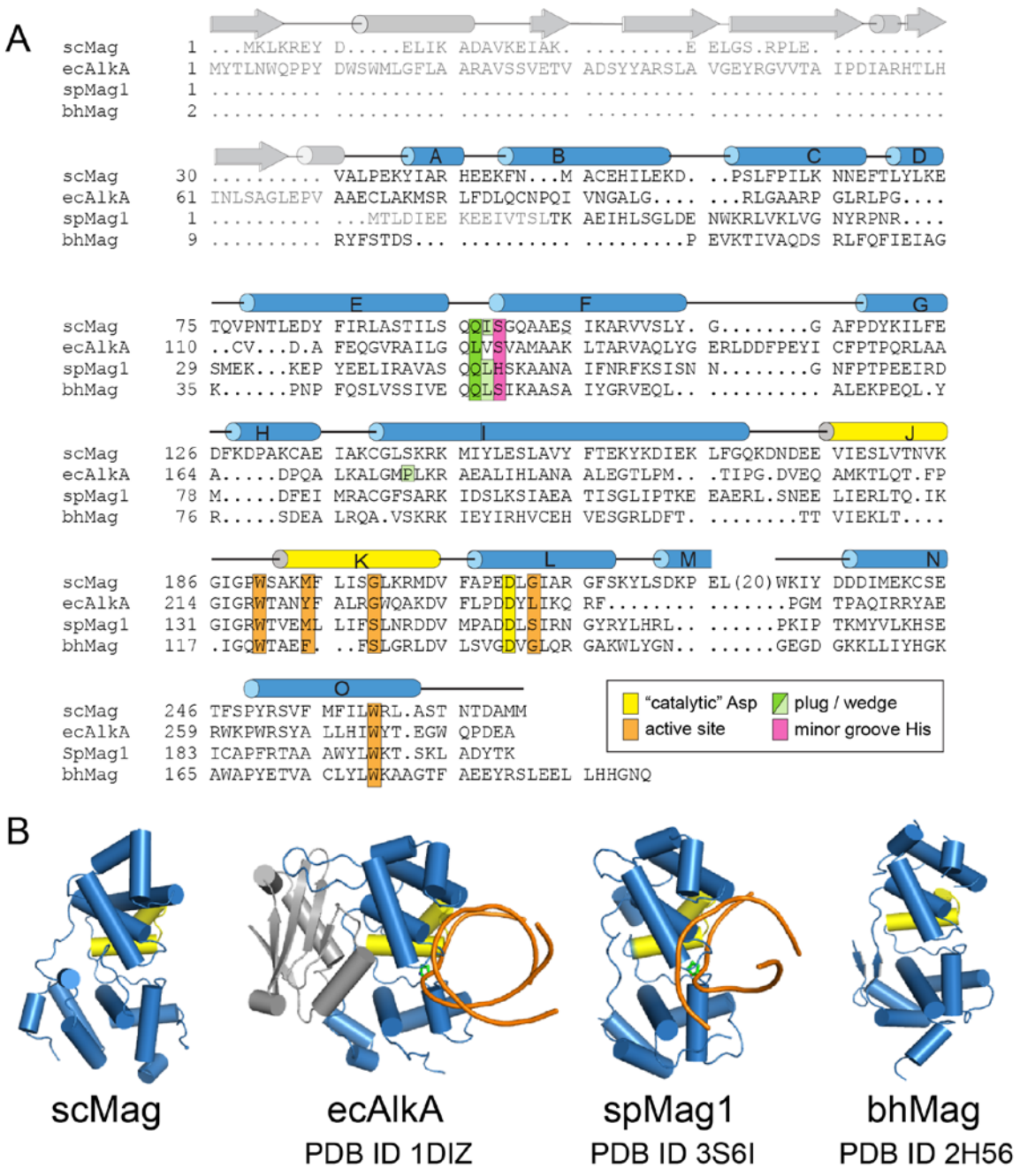


Fig. A2 Comparison of Mag orthologs. (A) Structure-based sequence alignment of 3-methyladenine DNA glycosylases from *Saccharomyces cerevisiae* (scMag), *Escherichia coli* (ecAlkA), *Schizosaccharomyces pombe* (spMag1) and *Bacillus halodurans* (bhMag). The alignment of spMag1, bhMag and ecAlkA was generated by structural superposition using PDB ID 3S6I, 2H56, and 1DIZ. Structural alignment was performed using STRAP (Gille, Lorenzen et al. 2003). Secondary structure of scMag (blue, cyan) and ecAlkA (grey) is shown schematically above the alignment. Except for the α/β domain of ecAlkA, amino acids not observed in the crystal structures are shown in lower opacity. Residues lining the extrahelical base binding pocket are highlighted orange, and conserved aspartic

acid residues important for catalysis in HhH glycosylases are highlighted yellow. DNA intercalating plug and wedge residues are highlighted green and light green, respectively, and the novel minor-groove interrogating histidine residue observed in spMag1 structure is highlighted pink. (B) Crystallographic models of scMag, ecAlkA (Hollis, Ichikawa et al. 2000), spMag1 and bhMag (PDB 2006) are shown in the same orientation. HhH motifs are shown in yellow and AlkA residues 1-89 are grey. DNA is shown as an orange backbone trace.

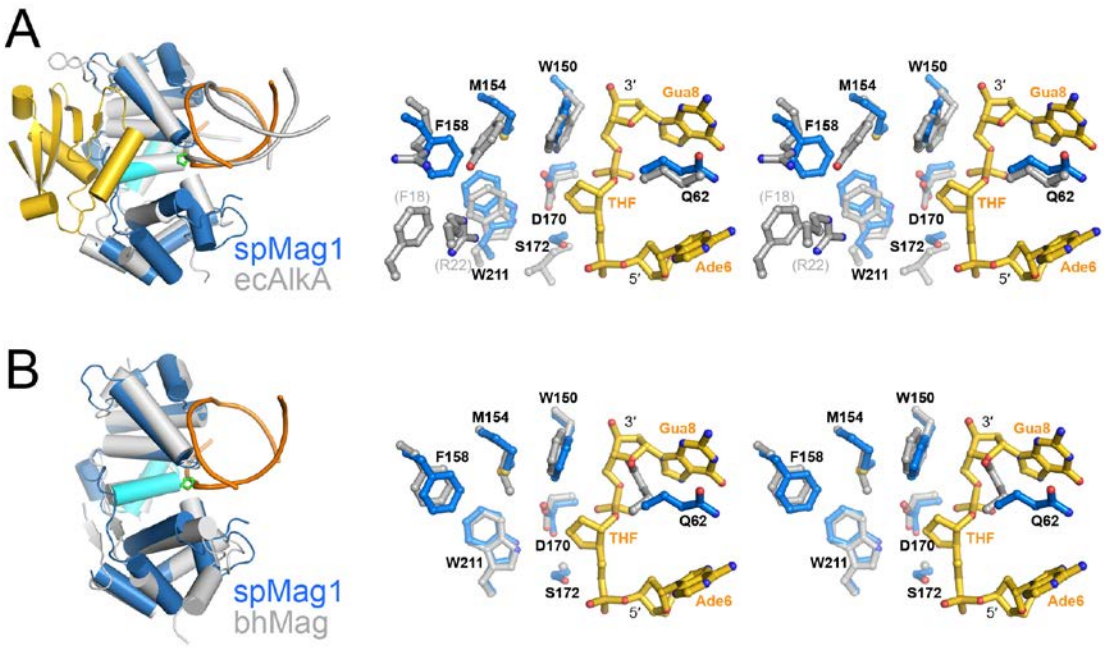


Fig. A3 Structural alignment of spMag1 with ecAlkA (A) and bhMag (B). In the superpositions at left, the structures are in the same orientation as in Fig. A2, with SpMag colored blue and AlkA/bhMag colored grey. AlkA residues 1-89 are colored gold. At the right are stereo views of the nucleobase binding pockets, in which spMag1 is blue with gold DNA and AlkA/bhMag are grey. Structures were taken from PDB ID 1DIZ (ecAlkA) and 2H56 (bhMag).

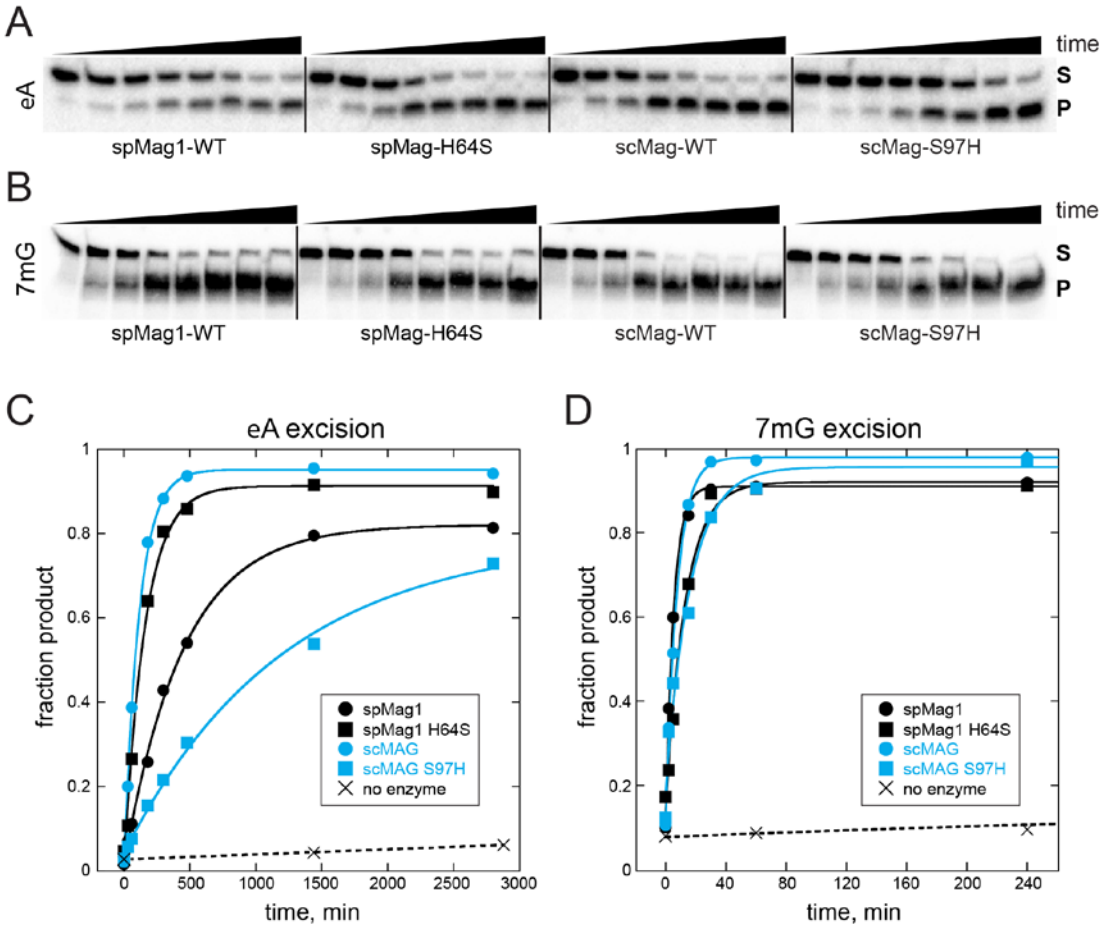
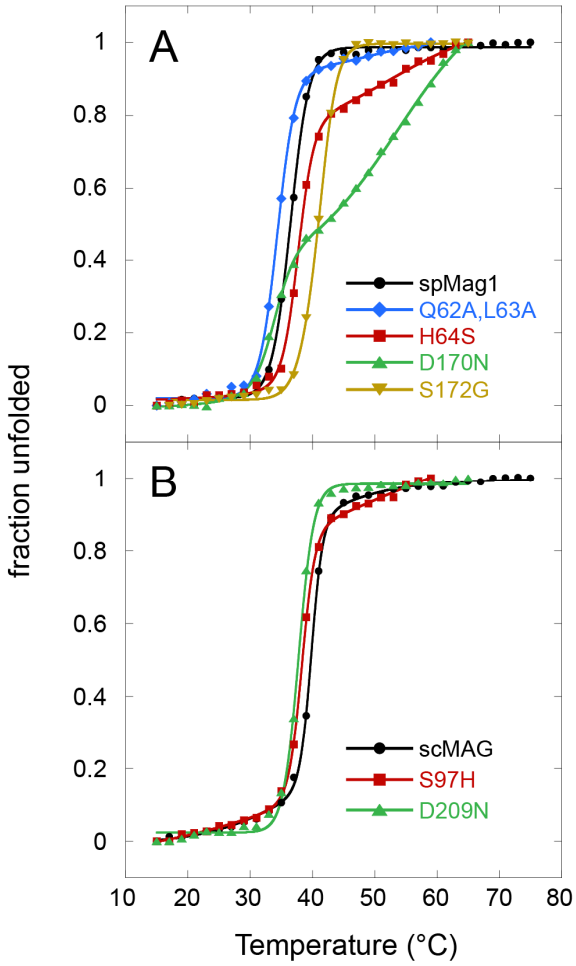


Fig. A4 Excision of εA and 7mG by Mag1 H64S and Mag S97H. (A) Denaturing polyacrylamide gels showing the disappearance of radiolabeled εA-containing 25-mer DNA substrate (S) and appearance of alkaline-cleaved 12-mer abasic-DNA product (P) as a function of time after addition of wild-type (WT) or mutant enzymes. For each enzyme, lanes correspond to 0, 30, 60, 180, 300, 480, 1440, 2880 min reaction times. (B) Same as (A) but using a 7mG-DNA substrate. For each enzyme, lanes correspond to 0, 2, 5, 15, 30, 60, 240, 1440 min reaction times. (C,D) Quantification of the data shown in (A) and (B), respectively. Black curves correspond to spMag1 and blue curves correspond to scMag. Circles denote wild-type and squares denote mutant proteins. Non-enzymatic εA and 7mG depurination is shown as crosses and a dotted line curve fit.



C

	T_{m1} (°C)	T_{m2} (°C)
spMag1	36.5	
Q62A,L63A	34.6	
H64S	37.7	52.4
D170N	33.9	53.7
S172G	40.9	
scMAG	39.7	
S97H	38.3	
D209N	37.7	

Fig. A5 Thermal denaturation of wild-type and mutant spMag1 (A) and scMag (B) proteins. Protein unfolding was monitored using CD spectroscopy, measuring the change in molar ellipticity at 222 nm as a function of temperature. (A,B) Molar ellipticity values were normalized to reflect the fraction of unfolded protein and plotted against temperature. The melting temperatures were determined by fitting the data using the equation $f_u = [(A_f\text{-}A_u)/(1+e^{-(T\text{-}Tm)/w})]$, in which f_u is the fraction of unfolded protein, A_f and A_u are the CD values for folded and unfolded protein respectively, T_m is the melting temperature, and w is the cooperativity coefficient for the transition. For curves showing bi-phasic transitions, melting temperatures for each transition were determined by fits to the following equation: $f_u = [(\Delta A_1)/(1+e^{-(T\text{-}Tm1)/w1})] + [(\Delta A_2)/(1+e^{-(T\text{-}Tm2)/w2})]$, in which ΔA refers to the change in CD signal for a particular transition.

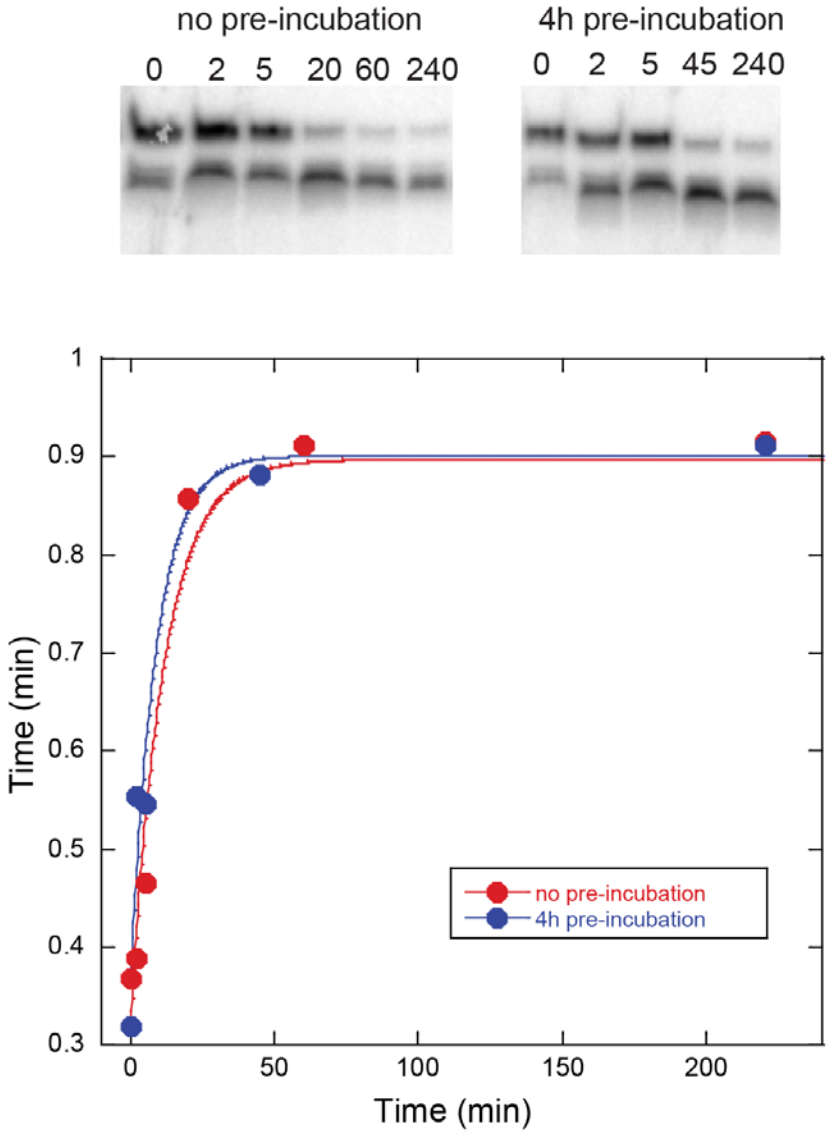


Fig. A6 Protein stability control for spMag1 7mG activity. Denaturing polyacrylamide gels showing the disappearance of radiolabeled 7mG-containing 25-mer DNA substrate and appearance of alkaline-cleaved 12-mer abasic-DNA product as a function of time after addition of spMag1 that had been either pre-incubated for 4h in the reaction buffer or not pre-incubated. The plot at the bottom shows the quantification of the data shown in the gel strips.

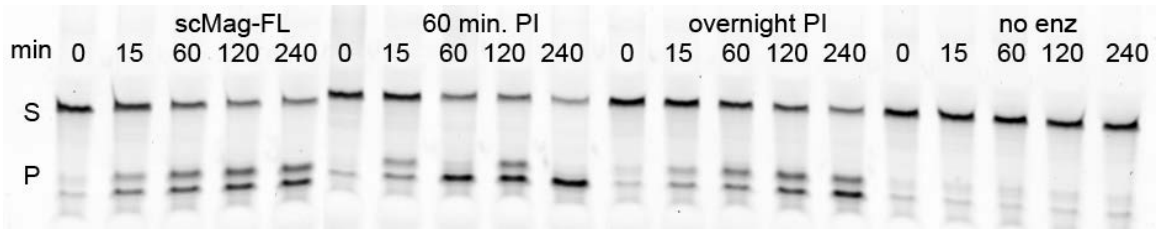


Fig. A7 Effect of chymotrypsin on base excision by scMag. Excision of εA from a 25-mer oligonucleotide (S) resulting in a 12-mer product (P) after NaOH treatment, separated on a denaturing PAGE. Base hydrolysis by scMag - lanes 1-5 : untreated protein, 6-10 : after being treated with 1:1000 molar equivalent of chymotrypsin for 60 minutes, 11-15 : after being treated with 1:1000 molar equivalent chymotrypsin overnight (16 hours) molar equivalent of chymotrypsin, 16-20 : no enzyme (scMag) control.

Table 8. DNA glycosylases specific for oxidized, alkylated, mismatched, uracil, and 5-methylcytosine bases

	Eukary-otes	Archa-ea	Prokaryot-es	Protein Fold	Substrates	PDB entries
Oxidation						
	OGG1	OGG	Ogg	HhH	8oxoG•C, FapyG, FapyA	1KO9 (hOGG1) 1EBM (K249Q/8oxoG-DNA) 1FN7 (THF-DNA) 1HU0, 1LWV, 1LWW, 1LWY (NaBH4-trapped DNA complex) 1M3H (D268E/nicked-DNA) 1M3Q (D268E/abasic-DNA/8-aminoG) 1N39 (D268E/THF-DNA) 1N3A (D268Q/THF-DNA) 1N3C (D268N/THF-DNA) 1YQK (N149C/G•C-DNA XL) 1YQL (N149C,K249Q/7-deaza-8-azaguanine-DNA XL) 1YQM (N149C,K249Q/7-deazaguanine-DNA XL) 1YQR (N149C,K249Q/8oxoG•C-DNA XL) 2I5W (N149C/8oxoG•G-DNA XL) 2NOB (N149C,K249Q,H270A/8oxoG•C-DNA XL) 2NOE (G42A,K249Q/8oxoG•C-DNA) 2NOF (N149C,Q315F/8oxoG•C/DNA XL) 2NOH (K249Q,Q315A/8oxoG•C-DNA) 2NOI (G42A,N149C,K249Q/8oxoG•C-DNA XL) 2NOL (K249Q,S292C/8oxoG•C-DNA) 2NOZ (S292C,Q315F/8oxoG•C-DNA) 2XHI (K249C,C253K,D268N/8oxoG•C-DNA) 3KTU (2′F-8oxoG-DNA)
		OGG2		HhH	8oxoG (paired with any base)	3FHF (MjOGG) 3KNT (MjOGG K129G/8oxoG•C-DNA) 3FHG (SsOGG)
		AGOG		HhH	8oxoG (ssDNA, dsDNA)	1XQO 1XQP (free 8oxoG)
			MutM/Fpg	H2TH	8oxoG, FaPy, 7mFapyG, Sp, Gh, Tg, Ug, DHT, DHU, 5-OHU, 5-OHC, FU, urea, oxazolone, oxaluric acid, oxidized εA derivatives, sulfur mustard guanine N7-adduct, ring-opened oxidized aminofluorene guanine C8-adduct, 5-hydroxy-5-methylhydantoin, 3-[(aminocarbonyl)amino]-	1EES TtMutM 2F5Q, 2F5S (E3Q GsMutM/8-oxoG•C-DNA XL) 1R2Y (E3Q GsMutM/8oxoG•C-DNA) 1R2Z (E3Q GsMutM/DHU-DNA) 1LIT (GsMutM/reduced abasic site-DNA) 1LIZ (GsMutM/NaBH4-trapped DNA complex) 1L2B (GsMutM/nicked DNA complex) 1L2C (GsMutM/HPD•T-DNA)

					(2R)-hydroxy-2-methylpropanoic acid	1L2D (GsMutM/HPD•G-DNA) 2F5N, 2F5P (Q166C GsMutM/A•T-DNA XL) 2F5O (Q166C GsMutM/G•C-DNA XL) 2F5Q, 2F5S (E3Q,Q166C GsMutM/8oxoG•C-DNA XL) 3JR4, 3JR5 (N174C GsMutM/G•C-DNA XL) 3GP1 (Q166C,V222P GsMutM/8oxoG•C-DNA XL) 3GPP (Q166C,T224P GsMutM/8oxoG•C-DNA XL) 3GPU, 3GQ3, 3GO8 (Δ220-235,Q166C GsMutM/8oxoG•C-DNA XL) 3GPX (Δ220-235,Q166C GsMutM/G•C-DNA XL) 3GPY (Q166C GsMutM/8oxoG•C-DNA XL) 3GQ4 (GsMutM/8oxoG•C-DNA XL) 3GQ5 (Q166C,T224P GsMutM/G•C-DNA XL) 3SAR, 3SAU (Q166C GsMutM/A•T-DNA XL) 3SAS, 3SAT (Q166C,R112A GsMutM/G•C-DNA XL) 3SAV (A149S,Q166C, GsMutM/G•C-DNA XL) 3SAW (GsMutM/G•C-DNA XL) 3SBJ (Q166C,V222P GsMutM/A•T-DNA XL) 1XC8, 1TDZ (LlMutM/FapyG•C-DNA) 1PJI, 1NNJ (LlMutM/PDI•C-DNA) 1PJJ, 1PM5 (LlMutM/THF•C-DNA) 1KFV(P1G LlMutM/PDI•C-DNA) 3C58 (LlMutM/7bFapyG•C-DNA) 2XZF (LlMutM-HC-DNA) 2XZU (LlMutM-HC-DNA XL)
	NTH1	EndoIII	Nth/EndoIII	HhH/FeS[2]	Tg, Ug, DHU, 5-OHU, 5-OHC, urea	2ABK (EcEndoIII) 1ORN, 1ORP (GsEndoIII/NaBH4-trapped DNA complexes) 1P59 (GsEndoIII/THF-DNA)
	NEIL1		Nei/Endo VIII	H2TH	Tg, DHT, DHU, 5-OHU, 5-OHC, 5fU, 5hmU, FapyG, FapyA, urea, 8oxoA, Gh, Sp, Ia; (Nei only: Ug, 8oxoG, 7mFapyG, 5,6dhC, 5-OHT)	1Q39 (EcEndoVIII) 1Q3B, 1Q3C (EcEndoVIII R252 and E2A mutants) 1K3W, 1K3X (EcEndoVIII/NaBH4-trapped DNA complexes) 2EA0, 2OPF, 2OQ4 (EcEndoVIII/PED-DNA) 1TDH (NEIL1) 3A45 (MvNei1) 3A46 (MvNei1/THF-DNA)
	NEIL2			H2TH	Gh/Ia, 5-OHU, FapyG	(Ref. Hazra, Kow et al. 2002; Katafuchi, Nakano et al. 2004)
	NEIL3			H2TH	Sp, Gh, FapyG, FapyA	(Ref. Hazra, Kow et al. 2002; Liu, Bandaru et al. 2010)
Alkylation						

	AAG			AAG	3mA, 7mG, εA, Hx, A, G	1EWN (εA-DNA) 1BNK (pyrrolidine-DNA) 3QI5 (εC-DNA)
	MAG, Mag1			HhH	3mA, 3mG, 7mG, 7-CEG, 7-HEG, εA, Hx, G	3S6I (SpMag1/THF-DNA)
		AfAlkA		HhH	3mA, 7mG, εA, 1mA, 3mC	2JHN (AfAlkA)
		MpgII		HhH	3mA, 7mG	(Ref. Begley, Haas et al. 1999)
			AlkA	HhH	3mA, 3mG, 7mG, 7-CEG, 7-HEG, 7-EG, O^2-mT, O^2-mC, εA, Hx, A, G, T, C, Xa	1MPG (EcAlkA) 1DIZ (EcAlkA/1-azaribose-DNA) 1PVS (EcAlkA/Hx base) 3OGD, 3OH6, 3OH9 (EcAlkA/undamaged DNA XL) 2H56 (BhAlkA) 2YG9 (DrAlkA)
			MagIII	HhH	3mA, mispaired 7mG	1PU6 1PU7 (MagIII/3,9-dimethylA) 1PU8 (MagIII/εA)
			TAG	HhH^1	3mA, 3mG	2OFK (StTAG) 2OFI (StTAG/THF-DNA/3mA) 1NKU, 1LMZ (EcTAG NMR) 1P7M (EcTAG/3mA NMR) 4AIA (SaTAG)
			AlkC, AlkD	ALK/HEAT	3mA, 3mG, 7mG, 7-POB-G, O^2-POB-C	3BVS (AlkD) 3JX7 (AlkD/3d3mA-DNA) 3JXY (AlkD/G•T-DNA) 3JXZ (AlkD/THF•T-DNA) 3JY1 (AlkD/THF•C-DNA)
Adenine						
	MUTYH	MutY	MutY	HhH/FeS^2	A•8oxoG, A•G,	1MUY, 1KG2, 1KG3 (EcMutY CD) 1RRQ (GsMutY/A•8oxoG-DNA XL) 1RRS, IVRL (GsMutY/HPD•8oxoG-DNA XL/adenine) 3N5N (HsMUTYH) 1WEF (K20A EcMutY CD) 1WEG, 1KG4 (K142A EcMutY CD) 1WEI (K20A EcMutY CD/adenine) 1KG5 (K142Q EcMutY CD) 1KG6 (K142R EcMutY CD) 1KG7 (E161A EcMutY CD) 1KQJ (C199H EcMutY CD) 1MUD (D138N EcMutY CD/adenine) 1MUN (D138N EcMutY CD)
U/T/5mC						
	UDG		Ung	UDG-1	U•G	1AKZ (HsUDG) 1SSP (HsUDG/U-DNA) 1LAU, 1UDG (HSV1 UDG) 1UDH (HSV1 UDG/uracil) 1EUG, 2EUG, 3EUG, 5EUG (EcUng/U, EcUng/glycerol)
	SMUG			UDG-3	U (ssDNA), U•G, U•A, 5hmU, 5-OHU, 5fU	1OE4 (XlSMUG/THF-DNA) 1OE5 (XlSMUG/THF-DNA/U) 1OE6 (XlSMUG/THF-DNA/5hmU)
	TDG		MUG	UDG-2	T•G, U•G, U•A, 5fC, 5caC, 5FU•G, 5FU•A, 5BrU•G, 5BrU•A, 5hmU•G 5-OHU•G, Tg•G, εC•G, εC•A, Hx•G, 8hmεC, εG Xa	2D07 (HsTDG/SUMO3) 1WYW (HsTDG/SUMO1) 2RBA (HsTDG/THF-DNA) 3UO7 (HsTDG/5caC-DNA) 3UFJ (HsTDG/dU analog) 1MWJ (EcMUG/U-DNA)

						1MTL, 1MWI (EcMUG/AP-DNA) 1MUG (EcMUG)
		UDG		UDG-4	U (ssDNA), U•G	1UI0 (TtUDG)
	MBD4			HhH	T•G, U•G, 5FU•G, εC, 5mC	1NGN (MmMBD4 CD) 3IHO (HsMBD4 CD) 4DK9 (HsMBD4/THF-DNA) 4EVV (MmMBD4/T•G-DNA) 4EW0 (MmMBD4/5hmU•G-DNA) 4EW4 (MmMBD4/AP-DNA)
		MIG		HhH/FeS[2]	T•G	1KEA
	DME, ROS1, DML2, DML3			HhH/FeS[2]	5mC, T•G	Plants only; (Ref. Choi, Gehring et al. 2002; Gong, Morales-Ruiz et al. 2002; Penterman, Zilberman et al. 2007; Ortega-Galisteo, Morales-Ruiz et al. 2008)

[1] TAG adopts the HhH architecture, but lacks the conserved catalytic aspartate and lysine residues present in mono- and bifunctional HhH glycosylases

[2] EndoIII, MutY, MIG, and DME/ROS incorporate Fe_4S_4-type iron sulfur clusters (FeS) into their HhH architecture

Abbreviations: AP, abasic site; THF, tetrahydrofuran; HPD, 1-hydroxypentane-3,4-diol; PDI, 3-hydroxypropyl; PED, pentane-3,4-diol; HC; hydantoin carbanucleoside; 8oxoG, 8-oxo-7,8-dihydroguanine; FapyG, 2,6-diamino-4-hydroxy-5-formamidopyrimidine; FapyA, 4,6-diamino-5-formamidopyrimidine; 7mFapyG, *N*7-methylFapyG; 7bFapyG; *N*7-benzylFapyG;Tg, thymine glycol; DHT, dihydrothymine; DHU, dihydrouracil; 5-OHC, 5-hydroxycytosine; 5-OHT, 5-hydroxythymine; 5,6dhC, 5,6-dihydroxycytosine; 5-OHU, 5-hydroxyuracil; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine; 5hmU, 5-hydroxymethyluracil; 5fU, 5-formyluracil; 5FU, 5-fluorouracil; 5BrU, 5-bromouracil; Gh, guanidinohydantoin; Ia, iminoallantion; Sp, spiroiminodihydantoin; 3mA, *N*3-methyladenine; 3mG, *N*3-methylguanine, 7mG, *N*7-methylguanine; 7-CEG, 7-(2-chloroethyl)guanine; 7-HEG, 7-(2-hydroxyethyl)guanine; 7-POB-G, *N*7- pyridyloxobutylguanine; O^2-POB-C, O^2-pyridyloxobutylcytosine; εA, 1,N^6-ethenoadenine; εG, 1,N^2-ethenoguanine; εC, 3,N^4-ethenocytosine; 8hmεC, 8-(hydroxymethyl)-3,N^4-ethenocytosine; 3d3mA, 3-deaza-*N*3-methyladenine; Hx, hypoxanthine; Xa, xanthine; ssDNA, single-stranded DNA; XL, covalent cross-linked protein-DNA; CD, catalytic domain; O^2-mT, *O2*-methylthymine;O^2-mC, *O2*-methylcytosine

Organisms: Hs, *Homo Sapeiens*; Mm, *Mus musculus*; Xl, *Xenopus laevis*; Mv, mimivirus; HSV1, Herpes simplex virus 1; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Mt, *Methanothermobacter thermautotrophicus*; Af, *Archaeoglobus fulgidus*; Ss, *Sulfolobus solfataricus*; Tt, *Thermus thermophilus*; Gs, *Geobacillus stearothermophilus*; Mj, *Methanocaldococcus jannaschii*; Ec, *Escherichia coli*; Ll, *Lactobacillus lacti*; St, *Salmonella* typhi; Sa, *Staphylococcus aureus*; Bh, *Bacillus halodurans*; Dr, *Deinococcus radiodurans*

FOOTNOTES

Work on DNA repair enzymes in the Eichman lab is funded by the National Institutes of Health (R01 ES019625) and the American Cancer Society (RSG-07-063-01-GMC). SpMag1 plasmid was a generous gift from Tom Hollis. I would also like to thank the Advanced Photon Source LS-CAT and SER-CAT beamline staff for their assistance with X-ray diffraction experiments. Use of the APS 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).

REFERENCES

Aamodt, R. M., P. O. Falnes, et al. (2004). "The Bacillus subtilis counterpart of the mammalian 3-methyladenine DNA glycosylase has hypoxanthine and 1,N6-ethenoadenine as preferred substrates." The Journal of biological chemistry **279**(14): 13601-13606.

Aburatani, H., Y. Hippo, et al. (1997). "Cloning and characterization of mammalian 8-hydroxyguanine-specific DNA glycosylase/apurinic, apyrimidinic lyase, a functional mutM homologue." Cancer research **57**(11): 2151-2156.

Adams, P. D., P. V. Afonine, et al. (2007). Automated structure determination with Phenix. Evolving Methods for Macromolecular Crystallography. R. J. Read and J. L. Sussman. Dordrecht, Springer**:** 101-109.

Adams, P. D., R. W. Grosse-Kunstleve, et al. (2002). "PHENIX: building new software for automated crystallographic structure determination." Acta Crystallogr D Biol Crystallogr **58**(Pt 11): 1948-1954.

Adhikary, S., M. C. Cato, et al. (2012). "Non-productive DNA damage binding by DNA glycosylase-like protein Mag2 from Schizosaccharomyces pombe." DNA Repair (Amst).

Adhikary, S. and B. F. Eichman (2011). "Analysis of substrate specificity of Schizosaccharomyces pombe Mag1 alkylpurine DNA glycosylase." EMBO Rep **12**(12): 1286-1292.

Al-Tassan, N., N. H. Chmiel, et al. (2002). "Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors." Nature genetics **30**(2): 227-232.

Aller, P., M. A. Rould, et al. (2007). "A structural rationale for stalling of a replicative DNA polymerase at the most common oxidative thymine lesion, thymine glycol." Proc Natl Acad Sci U S A **104**(3): 814-818.

Alseth, I., F. Osman, et al. (2005). "Biochemical characterization and DNA repair pathway interactions of Mag1-mediated base excision repair in Schizosaccharomyces pombe." Nucleic Acids Res **33**(3): 1123-1131.

Alseth, I., T. Rognes, et al. (2006). "A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from Bacillus cereus, AlkC and AlkD." Mol Microbiol **59**(5): 1602-1609.

Arai, K., K. Morishita, et al. (1997). "Cloning of a human homolog of the yeast OGG1 gene that is involved in the repair of oxidative DNA damage." Oncogene **14**(23): 2857-2861.

Asaeda, A., H. Ide, et al. (2000). "Substrate specificity of human methylpurine DNA N-glycosylase." Biochemistry **39**(8): 1959-1965.

Bailly, V., W. G. Verly, et al. (1989). "Mechanism of DNA strand nicking at apurinic/apyrimidinic sites by Escherichia coli [formamidopyrimidine]DNA glycosylase." The Biochemical journal **262**(2): 581-589.

Baldwin, M. R. and P. J. O'Brien (2009). "Human AP endonuclease 1 stimulates multiple-turnover base excision by alkyladenine DNA glycosylase." Biochemistry **48**(25): 6022-6033.

Banerjee, A., W. L. Santos, et al. (2006). "Structure of a DNA glycosylase searching for lesions." Science **311**(5764): 1153-1157.

Banerjee, A. and G. L. Verdine (2006). "A nucleobase lesion remodels the interaction of its normal neighbor in a DNA glycosylase complex." Proc Natl Acad Sci U S A **103**(41): 15020-15025.

Banerjee, A., W. Yang, et al. (2005). "Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA." Nature **434**(7033): 612-618.

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

Barrett, T. E., R. Savva, et al. (1998). "Structure of a DNA base-excision product resembling a cisplatin inter-strand adduct." Nature structural biology **5**(8): 697-701.

Barrett, T. E., R. Savva, et al. (1998). "Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions." Cell **92**(1): 117-129.

Barrett, T. E., O. D. Scharer, et al. (1999). "Crystal structure of a thwarted mismatch glycosylase DNA repair complex." EMBO J **18**(23): 6599-6609.

Begley, T. J., B. J. Haas, et al. (1999). "A new member of the endonuclease III family of DNA repair enzymes that removes methylated purines from DNA." Curr Biol **9**(12): 653-656.

Berdal, K. G., M. Bjørås, et al. (1990). "Cloning and expression in Escherichia coli of a gene for an alkylbase DNA glycosylase from Saccharomyces cerevisiae; a homologue to the bacterial alkA gene." EMBO J **9**(13): 4563-4568.

Berdal, K. G., R. F. Johansen, et al. (1998). "Release of normal bases from intact DNA by a native DNA repair enzyme." Embo J **17**(2): 363-367.

Berti, P. J. and J. A. McCann (2006). "Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer." Chem Rev **106**(2): 506-555.

Birkeland, N. K., H. Anensen, et al. (2002). "Methylpurine DNA glycosylase of the hyperthermophilic archaeon Archaeoglobus fulgidus." Biochemistry **41**(42): 12697-12705.

Bjelland, S., N. K. Birkeland, et al. (1994). "DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the AlkA enzyme in *Escherichia coli*." J Biol Chem **269**(48): 30489-30495.

Bjelland, S., M. Bjoras, et al. (1993). "Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of *Escherichia coli*." Nucleic Acids Res **21**(9): 2045-2049.

Bjørås, M., A. Klungland, et al. (1995). "Purification and properties of the alkylation repair DNA glycosylase encoded the MAG gene from Saccharomyces cerevisiae." Biochemistry **34**(14): 4577-4582.

Bjørås, M., L. Luna, et al. (1997). "Opposite base-dependent reactions of a human base excision repair enzyme on DNA containing 7,8-dihydro-8-oxoguanine and abasic sites." The EMBO journal **16**(20): 6314-6322.

Bjørås, M., E. Seeberg, et al. (2002). "Reciprocal "flipping" underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase." J Mol Biol **317**(2): 171-177.

Blainey, P. C., G. Luo, et al. (2009). "Nonspecifically bound proteins spin while diffusing along DNA." Nat Struct Mol Biol **16**(12): 1224-1229.

Blainey, P. C., A. M. van Oijen, et al. (2006). "A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA." Proc Natl Acad Sci U S A **103**(15): 5752-5757.

Boorstein, R. J. and G. W. Teebor (1988). "Mutagenicity of 5-hydroxymethyl-2'-deoxyuridine to Chinese hamster cells." Cancer research **48**(19): 5466-5470.

Bowman, B. R., S. Lee, et al. (2008). "Structure of the E. coli DNA glycosylase AlkA bound to the ends of duplex DNA: a system for the structure determination of lesion-containing DNA." Structure **16**(8): 1166-1174.

Bowman, B. R., S. Lee, et al. (2010). "Structure of Escherichia coli AlkA in complex with undamaged DNA." J Biol Chem **285**(46): 35783-35791.

Brent, T. P. (1979). "Partial purification and characterization of a human 3-methyladenine-DNA glycosylase." Biochemistry **18**(5): 911-916.

Brieba, L. G., B. F. Eichman, et al. (2004). "Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase." EMBO J **23**(17): 3452-3461.

Brinkmeyer, M. K., M. A. Pope, et al. (2012). "Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays." Chemistry & biology **19**(2): 276-286.

Bronner, C. E., D. L. Welker, et al. (1992). "Mutations affecting sensitivity of the cellular slime mold Dictyostelium discoideum to DNA-damaging agents." Mutat Res **274**(3): 187-200.

Brooks, S. C., S. Adhikary, et al. (2012). "Recent advances in the structural mechanisms of DNA glycosylases." Biochim Biophys Acta **1834**(1): 247-271.

Bruner, S. D., D. P. Norman, et al. (2000). "Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA." Nature **403**(6772): 859-866.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-921.

Burdett, V., C. Baitinger, et al. (2001). "In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair." Proc Natl Acad Sci U S A **98**(12): 6765-6770.

Burrows, C. J. and J. G. Muller (1998). "Oxidative Nucleobase Modifications Leading to Strand Scission." Chemical reviews **98**(3): 1109-1152.

Burrows, C. J., J. G. Muller, et al. (2002). "Structure and potential mutagenicity of new hydantoin products from guanosine and 8-oxo-7,8-dihydroguanine oxidation by transition metals." Environmental health perspectives **110 Suppl 5**: 713-717.

Campalans, A., S. Marsin, et al. (2005). "XRCC1 interactions with multiple DNA glycosylases: a model for its recruitment to base excision repair." DNA Repair (Amst) **4**(7): 826-835.

Cao, C., Y. L. Jiang, et al. (2006). "The catalytic power of uracil DNA glycosylase in the opening of thymine base pairs." J Am Chem Soc **128**(40): 13034-13035.

Cao, C., Y. L. Jiang, et al. (2004). "Dynamic opening of DNA during the enzymatic search for a damaged base." Nat Struct Mol Biol **11**(12): 1230-1236.
Cao, C., K. Kwon, et al. (2003). "Solution structure and base perturbation studies reveal a novel mode of alkylated base recognition by 3-methyladenine DNA glycosylase I." J Biol Chem **278**(48): 48012-48020.
Castaing, B., J. L. Fourrey, et al. (1999). "AP site structural determinants for Fpg specific recognition." Nucleic Acids Res **27**(2): 608-615.
Castaing, B., A. Geiger, et al. (1993). "Cleavage and binding of a DNA fragment containing a single 8-oxoguanine by wild type and mutant FPG proteins." Nucleic Acids Res **21**(12): 2899-2905.
Ceska, T. A., J. R. Sayers, et al. (1996). "A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease." Nature **382**(6586): 90-93.
Chang, D. Y. and A. L. Lu (2005). "Interaction of checkpoint proteins Hus1/Rad1/Rad9 with DNA base excision repair enzyme MutY homolog in fission yeast, Schizosaccharomyces pombe." J Biol Chem **280**(1): 408-417.
Chang, P. W., A. Madabushi, et al. (2009). "Insights into the role of Val45 and Gln182 of Escherichia coli MutY in DNA substrate binding and specificity." BMC biochemistry **10**: 19.
Chen, J., B. Derfler, et al. (1990). "Saccharomyces cerevisiae 3-methyladenine DNA glycosylase has homology to the AlkA glycosylase of E. coli and is induced in response to DNA alkylation damage." Embo J **9**(13): 4569-4575.
Chen, J. and L. Samson (1991). "Induction of S.cerevisiae MAG 3-methyladenine DNA glycosylase transcript levels in response to DNA damage." Nucleic Acids Res **19**(23): 6427-6432.
Cheng, K. C., D. S. Cahill, et al. (1992). "8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions." The Journal of biological chemistry **267**(1): 166-172.
Chmiel, N. H., A. L. Livingston, et al. (2003). "Insight into the functional consequences of inherited variants of the hMYH adenine glycosylase associated with colorectal cancer: complementation assays with hMYH variants and pre-steady-state kinetics of the corresponding mutated E.coli enzymes." J Mol Biol **327**(2): 431-443.
Choi, Y., M. Gehring, et al. (2002). "DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis." Cell **110**(1): 33-42.
Cleaver, J. E. (1968). "Defective repair replication of DNA in xeroderma pigmentosum." Nature **218**(5142): 652-656.
Cleaver, J. E. (1989). "DNA repair in man." Birth Defects Orig Artic Ser **25**(2): 61-82.
Connor, E. E. and M. D. Wyatt (2002). "Active-site clashes prevent the human 3-methyladenine DNA glycosylase from improperly removing bases." Chem Biol **9**(9): 1033-1041.
Coste, F., M. Ober, et al. (2004). "Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase." J Biol Chem **279**(42): 44074-44083.
Coulondre, C., J. H. Miller, et al. (1978). "Molecular basis of base substitution hotspots in Escherichia coli." Nature **274**(5673): 775-780.

Crenshaw, C. M., K. Nam, et al. (2012). "Enforced presentation of an extrahelical guanine to the lesion-recognition pocket of the human 8-oxoguanine DNA glycosylase, hOGG1." J Biol Chem.

Crick, F. H. (1958). "On protein synthesis." Symp Soc Exp Biol **12**: 138-163.

Dalhus, B., M. Forsbring, et al. (2011). "Separation-of-function mutants unravel the dual-reaction mode of human 8-oxoguanine DNA glycosylase." Structure **19**(1): 117-127.

Dalhus, B., I. H. Helle, et al. (2007). "Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats." Nucleic Acids Res **35**(7): 2451-2459.

Dalhus, B., J. K. Laerdahl, et al. (2009). "DNA base repair--recognition and initiation of catalysis." FEMS microbiology reviews **33**(6): 1044-1078.

Dalhus, B., L. Nilsen, et al. (2012). "Sculpting of DNA at Abasic Sites by DNA Glycosylase Homolog Mag2." Structure.

David, S. S., V. L. O'Shea, et al. (2007). "Base-excision repair of oxidative DNA damage." Nature **447**(7147): 941-950.

Delaney, S., W. L. Neeley, et al. (2007). "The substrate specificity of MutY for hyperoxidized guanine lesions in vivo." Biochemistry **46**(5): 1448-1455.

Denver, D. R., S. L. Swenson, et al. (2003). "An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases." Mol Biol Evol **20**(10): 1603-1611.

Dong, A., X. Xu, et al. (2007). "In situ proteolysis for protein crystallization and structure determination." Nat Methods **4**(12): 1019-1021.

Drohat, A. C., J. Jagadeesh, et al. (1999). "Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase." Biochemistry **38**(37): 11866-11875.

Drohat, A. C., K. Kwon, et al. (2002). "3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member." Nat Struct Biol **9**(9): 659-664.

Duclos, S., P. Aller, et al. (2012). "Structural and biochemical studies of a plant formamidopyrimidine-DNA glycosylase reveal why eukaryotic Fpg glycosylases do not excise 8-oxoguanine." DNA Repair (Amst)(In press).

Duncan, B. K. and J. H. Miller (1980). "Mutagenic deamination of cytosine residues in DNA." Nature **287**(5782): 560-561.

Duncan, J., L. Hamilton, et al. (1976). "Enzymatic degradation of uracil-containing DNA. II. Evidence for N-glycosidase and nuclease activities in unfractionated extracts of Bacillus subtilis." J Virol **19**(2): 338-345.

Dzantiev, L., N. Constantin, et al. (2004). "A defined human system that supports bidirectional mismatch-provoked excision." Mol Cell **15**(1): 31-41.

Eichman, B. F., E. J. O'Rourke, et al. (2003). "Crystal structures of 3-methyladenine DNA glycosylase MagIII and the recognition of alkylated bases." Embo J **22**(19): 4898-4909.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.

Engelward, B. P., G. Weeda, et al. (1997). "Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase." Proc Natl Acad Sci U S A **94**(24): 13087-13092.

Evans, M. D., M. Dizdaroglu, et al. (2004). "Oxidative DNA damage and disease: induction, repair and significance." Mutat Res **567**(1): 1-61.

Faucher, F., S. Doublié, et al. (2012). "8-Oxoguanine DNA Glycosylases: One Lesion, Three Subfamilies." International Journal of Molecular Sciences **13**(6): 6711-6729.

Faucher, F., S. Duclos, et al. (2009). "Crystal structures of two archaeal 8-oxoguanine DNA glycosylases provide structural insight into guanine/8-oxoguanine distinction." Structure **17**(5): 703-712.

Faucher, F., S. M. Robey-Bond, et al. (2009). "Structural characterization of Clostridium acetobutylicum 8-oxoguanine DNA glycosylase in its apo form and in complex with 8-oxodeoxyguanosine." J Mol Biol **387**(3): 669-679.

Faucher, F., S. S. Wallace, et al. (2009). "Structural basis for the lack of opposite base specificity of Clostridium acetobutylicum 8-oxoguanine DNA glycosylase." DNA Repair (Amst) **8**(11): 1283-1289.

Fitzgerald, M. E. and A. C. Drohat (2008). "Coordinating the initial steps of base excision repair. Apurinic/apyrimidinic endonuclease 1 actively stimulates thymine DNA glycosylase by disrupting the product complex." The Journal of biological chemistry **283**(47): 32680-32690.

Friedberg, E. C., A. Aguilera, et al. (2006). "DNA repair: from molecular mechanism to human disease." DNA Repair (Amst) **5**(8): 986-996.

Friedberg, E. C., G. C. Walker, et al. (2006). DNA Repair and Mutagenesis. Washimgton, D.C., ASM Press.

Friedman, J. I., A. Majumdar, et al. (2009). "Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage." Nucleic Acids Res **37**(11): 3493-3500.

Friedman, J. I. and J. T. Stivers (2010). "Detection of damaged DNA bases by DNA glycosylase enzymes." Biochemistry **49**(24): 4957-4967.

Fromme, J. C., A. Banerjee, et al. (2004). "Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase." Nature **427**(6975): 652-656.

Fromme, J. C., A. Banerjee, et al. (2004). "DNA glycosylase recognition and catalysis." Curr Opin Struct Biol **14**(1): 43-49.

Fromme, J. C., S. D. Bruner, et al. (2003). "Product-assisted catalysis in base-excision DNA repair." Nat Struct Biol **10**(3): 204-211.

Fromme, J. C. and G. L. Verdine (2002). "Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM." Nat Struct Biol **9**(7): 544-552.

Fromme, J. C. and G. L. Verdine (2003). "DNA lesion recognition by the bacterial repair enzyme MutM." J Biol Chem **278**(51): 51543-51548.

Fromme, J. C. and G. L. Verdine (2004). "Base excision repair." Adv Protein Chem **69**: 1-41.

Fu, D., J. A. Calvo, et al. (2012). "Balancing repair and tolerance of DNA damage caused by alkylating agents." Nat Rev Cancer **12**(2): 104-120.

Gallagher, P. E. and T. P. Brent (1982). "Partial purification and characterization of 3-methyladenine-DNA glycosylase from human placenta." Biochemistry **21**(25): 6404-6409.

Gallinari, P. and J. Jiricny (1996). "A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase." Nature **383**(6602): 735-738.

Gilboa, R., D. O. Zharkov, et al. (2002). "Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA." J Biol Chem **277**(22): 19811-19816.

Gille, C., S. Lorenzen, et al. (2003). "KISS for STRAP: user extensions for a protein alignment editor." Bioinformatics **19**(18): 2489-2491.

Gogos, A. and N. D. Clarke (1999). "Characterization of an 8-oxoguanine DNA glycosylase from Methanococcus jannaschii." The Journal of biological chemistry **274**(43): 30447-30450.

Gong, Z., T. Morales-Ruiz, et al. (2002). "ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase." Cell **111**(6): 803-814.

Gowers, D. M. and S. E. Halford (2003). "Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling." EMBO J **22**(6): 1410-1418.

Guan, Y., R. C. Manuel, et al. (1998). "MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily." Nat Struct Biol **5**(12): 1058-1064.

Guarne, A. (2012). "The functions of MutL in mismatch repair: the power of multitasking." Prog Mol Biol Transl Sci **110**: 41-70.

Guex, N. and M. C. Peitsch (1997). "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling." Electrophoresis **18**(15): 2714-2723.

Halford, S. E. (2001). "Hopping, jumping and looping by restriction enzymes." Biochem Soc Trans **29**(Pt 4): 363-374.

Halford, S. E. and M. D. Szczelkun (2002). "How to get from A to B: strategies for analysing protein motion on DNA." Eur Biophys J **31**(4): 257-267.

Hang, B., B. Singer, et al. (1997). "Targeted deletion of alkylpurine-DNA-N-glycosylase in mice eliminates repair of 1,N6-ethenoadenine and hypoxanthine but not of 3,N4-ethenocytosine or 8-oxoguanine." Proc Natl Acad Sci U S A **94**(24): 12869-12874.

Haushalter, K. A., M. W. Todd Stukenberg, et al. (1999). "Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors." Current biology : CB **9**(4): 174-185.

Hazra, T. K., Y. W. Kow, et al. (2002). "Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions." J Biol Chem **277**(34): 30417-30420.

Hendershot, J. M., A. E. Wolfe, et al. (2011). "Substitution of active site tyrosines with tryptophan alters the free energy for nucleotide flipping by human alkyladenine DNA glycosylase." Biochemistry **50**(11): 1864-1874.

Henderson, P. T., J. C. Delaney, et al. (2003). "The hydantoin lesions formed from oxidation of 7,8-dihydro-8-oxoguanine are potent sources of replication errors in vivo." Biochemistry **42**(31): 9257-9262.

Hill, J. W., T. K. Hazra, et al. (2001). "Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair." Nucleic Acids Res **29**(2): 430-438.

Hollis, T., Y. Ichikawa, et al. (2000). "DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, Escherichia coli AlkA." Embo J **19**(4): 758-766.

Hollis, T., A. Lau, et al. (2000). "Structural studies of human alkyladenine glycosylase and E. coli 3-methyladenine glycosylase." Mutat Res **460**(3-4): 201-210.

Holt, S., T. Y. Yen, et al. (1998). "Detection of 1,N6-ethenoadenine in rat urine after chloroethylene oxide exposure." Carcinogenesis **19**(10): 1763-1769.

Hsu, G. W., M. Ober, et al. (2004). "Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase." Nature **431**(7005): 217-221.

Huffman, J. L., O. Sundheim, et al. (2005). "DNA base damage recognition and removal: new twists and grooves." Mutat Res **577**(1-2): 55-76.

Imamura, K., S. S. Wallace, et al. (2009). "Structural characterization of a viral NEIL1 ortholog unliganded and bound to abasic site-containing DNA." J Biol Chem **284**(38): 26174-26183.

Jiang, D., Z. Hatahet, et al. (1997). "Characterization of Escherichia coli endonuclease VIII." J Biol Chem **272**(51): 32230-32239.

Jiang, Y. L., K. Kwon, et al. (2001). "Turning On uracil-DNA glycosylase using a pyrene nucleotide switch." J Biol Chem **276**(45): 42347-42354.

Jones, A. S., M. K. Khan, et al. (1968). "Synthetic analogues of polynucleotides. 3. The synthesis and polymerisation of 2',3'-O-isopropylidene-5'-O-(3-vinylacryloyl)uridine." J Chem Soc Perkin 1 **12**: 1454-1457.

Joseph, N., V. Duppatla, et al. (2006). "Prokaryotic DNA mismatch repair." Prog Nucleic Acid Res Mol Biol **81**: 1-49.

Kanamitsu, K. and S. Ikeda (2010). "Early Steps in the DNA Base Excision Repair Pathway of a Fission Yeast Schizosaccharomyces pombe." J Nucleic Acids **2010**.

Kanamitsu, K., H. Tanihigashi, et al. (2007). "Involvement of 3-methyladenine DNA glycosylases Mag1p and Mag2p in base excision repair of methyl methanesulfonate-damaged DNA in the fission yeast Schizosaccharomyces pombe." Genes Genet Syst **82**(6): 489-494.

Karran, P. and T. Lindahl (1980). "Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus." Biochemistry **19**(26): 6005-6011.

Katafuchi, A., T. Nakano, et al. (2004). "Differential specificity of human and Escherichia coli endonuclease III and VIII homologues for oxidative base lesions." J Biol Chem **279**(14): 14464-14471.

Katoh, K. and H. Toh (2008). "Recent developments in the MAFFT multiple sequence alignment program." Brief Bioinform **9**(4): 286-298.

Kavli, B., O. Sundheim, et al. (2002). "hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup." J Biol Chem **277**(42): 39926-39936.

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

Klaunig, J. E. and L. M. Kamendulis (2004). "The role of oxidative stress in carcinogenesis." Annual review of pharmacology and toxicology **44**: 239-267.

Kreutzer, D. A. and J. M. Essigmann (1998). "Oxidized, deaminated cytosines are a source of C --> T transitions in vivo." Proceedings of the National Academy of Sciences of the United States of America **95**(7): 3578-3582.

Krokan, H. E., F. Drablos, et al. (2002). "Uracil in DNA--occurrence, consequences and repair." Oncogene **21**(58): 8935-8948.

Kryston, T. B., A. B. Georgiev, et al. (2011). "Role of oxidative stress and DNA damage in human carcinogenesis." Mutation research **711**(1-2): 193-201.

Kuo, C. F., D. E. McRee, et al. (1992). "Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III." Science **258**(5081): 434-440.

Kuper, J. and C. Kisker (2012). "Damage recognition in nucleotide excision DNA repair." Curr Opin Struct Biol **22**(1): 88-93.

Kuznetsov, N. A., V. V. Koval, et al. (2005). "Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA glycosylase." Nucleic Acids Res **33**(12): 3919-3931.

Kwon, K., C. Cao, et al. (2003). "A novel zinc snap motif conveys structural stability to 3-methyladenine DNA glycosylase I." J Biol Chem **278**(21): 19442-19446.

Labahn, J., O. D. Scharer, et al. (1996). "Structural basis for the excision repair of alkylation-damaged DNA." Cell **86**(2): 321-329.

Laskowski, R. A., J. A. Rullmannn, et al. (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-486.

Lau, A. Y., O. D. Scharer, et al. (1998). "Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision." Cell **95**(2): 249-258.

Lau, A. Y., M. D. Wyatt, et al. (2000). "Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG." Proc Natl Acad Sci U S A **97**(25): 13573-13578.

Lee, C. Y., J. C. Delaney, et al. (2009). "Recognition and processing of a new repertoire of DNA substrates by human 3-methyladenine DNA glycosylase (AAG)." Biochemistry **48**(9): 1850-1861.

Lee, S., B. R. Bowman, et al. (2008). "Synthesis and structure of duplex DNA containing the genotoxic nucleobase lesion N7-methylguanine." J Am Chem Soc **130**(35): 11570-11571.

Lee, S., C. T. Radom, et al. (2008). "Trapping and structural elucidation of a very advanced intermediate in the lesion-extrusion pathway of hOGG1." J Am Chem Soc **130**(25): 7784-7785.

Lee, S. and G. L. Verdine (2009). "Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase." Proceedings of the National Academy of Sciences of the United States of America **106**(44): 18497-18502.

Leipold, M. D., J. G. Muller, et al. (2000). "Removal of hydantoin products of 8-oxoguanine oxidation by the Escherichia coli DNA repair enzyme, FPG." Biochemistry **39**(48): 14984-14992.

Leiros, I., M. P. Nabong, et al. (2007). "Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA." EMBO J **26**(8): 2206-2217.
Li, G. M. (2003). "DNA mismatch repair and cancer." Front Biosci **8**: d997-1017.
Li, G. M. (2010). "Novel molecular insights into the mechanism of GO removal by MutM." Cell research **20**(2): 116-118.
Lindahl, T. (1974). "An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues." Proc Natl Acad Sci U S A **71**(9): 3649-3653.
Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." Nature **362**(6422): 709-715.
Lindahl, T. (2000). "Suppression of spontaneous mutagenesis in human cells by DNA base excision-repair." Mutation research **462**(2-3): 129-135.
Lindahl, T. and D. E. Barnes (2000). "Repair of endogenous DNA damage." Cold Spring Harb Symp Quant Biol **65**: 127-133.
Lindahl, T. and B. Nyberg (1972). "Rate of depurination of native deoxyribonucleic acid." Biochemistry **11**(19): 3610-3618.
Lingaraju, G. M., M. Kartalou, et al. (2008). "Substrate specificity and sequence-dependent activity of the Saccharomyces cerevisiae 3-methyladenine DNA glycosylase (Mag)." DNA Repair (Amst) **7**(6): 970-982.
Liu, J. and P. W. Doetsch (1998). "Escherichia coli RNA and DNA polymerase bypass of dihydrouracil: mutagenic potential via transcription and replication." Nucleic acids research **26**(7): 1707-1712.
Liu, M., V. Bandaru, et al. (2010). "The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo." Proceedings of the National Academy of Sciences of the United States of America **107**(11): 4925-4930.
Livingston, A. L., S. Kundu, et al. (2005). "Insight into the roles of tyrosine 82 and glycine 253 in the Escherichia coli adenine glycosylase MutY." Biochemistry **44**(43): 14179-14190.
Livingston, A. L., V. L. O'Shea, et al. (2008). "Unnatural substrates reveal the importance of 8-oxoguanine for in vivo mismatch repair by MutY." Nature chemical biology **4**(1): 51-58.
Lloyd, R. S., P. C. Hanawalt, et al. (1980). "Processive action of T4 endonuclease V on ultraviolet-irradiated DNA." Nucleic Acids Res **8**(21): 5113-5127.
Lu, A. L., H. Bai, et al. (2006). "MutY and MutY homologs (MYH) in genome maintenance." Frontiers in bioscience : a journal and virtual library **11**: 3062-3080.
Luncsford, P. J., D. Y. Chang, et al. (2010). "A structural hinge in eukaryotic MutY homologues mediates catalytic activity and Rad9-Rad1-Hus1 checkpoint complex interactions." J Mol Biol **403**(3): 351-370.
Luo, W., J. G. Muller, et al. (2001). "Characterization of hydantoin products from one-electron oxidation of 8-oxo-7,8-dihydroguanosine in a nucleoside model." Chemical Research in Toxicology **14**(7): 927-938.
Maiti, A., M. T. Morgan, et al. (2009). "Role of two strictly conserved residues in nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase." J Biol Chem **284**(52): 36680-36688.

Maiti, A., M. T. Morgan, et al. (2008). "Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition." Proceedings of the National Academy of Sciences of the United States of America **105**(26): 8890-8895.

Mansfield, C., S. M. Kerins, et al. (2003). "Characterisation of Archaeglobus fulgidus AlkA hypoxanthine DNA glycosylase activity." FEBS letters **540**(1-3): 171-175.

Manuel, R. C., K. Hitomi, et al. (2004). "Reaction intermediates in the catalytic mechanism of Escherichia coli MutY DNA glycosylase." J Biol Chem **279**(45): 46930-46939.

McCann, J. A. and P. J. Berti (2008). "Transition-state analysis of the DNA repair enzyme MutY." Journal of the American Chemical Society **130**(17): 5789-5797.

McCarthy, T. V., P. Karran, et al. (1984). "Inducible repair of O-alkylated DNA pyrimidines in *Escherichia coli*." Embo J **3**(3): 545-550.

McCullough, A. K., M. L. Dodson, et al. (1999). "Initiation of base excision repair: glycosylase mechanisms and structures." Annual review of biochemistry **68**: 255-285.

McKibbin, P. L., A. Kobori, et al. (2012). "Surprising repair activities of nonpolar analogs of 8-oxoG expose features of recognition and catalysis by base excision repair glycosylases." Journal of the American Chemical Society **134**(3): 1653-1661.

Memisoglu, A. and L. Samson (1996). "Cloning and characterization of a cDNA encoding a 3-methyladenine DNA glycosylase from the fission yeast Schizosaccharomyces pombe." Gene **177**(1-2): 229-235.

Memisoglu, A. and L. Samson (2000). "Base excision repair in yeast and mammals." Mutat Res **451**(1-2): 39-51.

Memisoglu, A. and L. Samson (2000). "Contribution of base excision repair, nucleotide excision repair, and DNA recombination to alkylation resistance of the fission yeast Schizosaccharomyces pombe." J Bacteriol **182**(8): 2104-2112.

Metz, A. H., T. Hollis, et al. (2007). "DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG)." Embo J **26**(9): 2411-2420.

Michaels, M. L. and J. H. Miller (1992). "The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine)." Journal of bacteriology **174**(20): 6321-6325.

Modrich, P. (1997). "Strand-specific mismatch repair in mammalian cells." J Biol Chem **272**(40): 24727-24730.

Moe, E., D. R. Hall, et al. (2012). "Structure-function studies of an unusual 3-methyladenine DNA glycosylase II (AlkA) from Deinococcus radiodurans." Acta crystallographica. Section D, Biological crystallography **68**(Pt 6): 703-712.

Mol, C. D., A. S. Arvai, et al. (2002). "Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases." J Mol Biol **315**(3): 373-384.

Mol, C. D., A. S. Arvai, et al. (1995). "Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis." Cell **80**(6): 869-878.

Mol, C. D., C. F. Kuo, et al. (1995). "Structure and function of the multifunctional DNA-repair enzyme exonuclease III." Nature **374**(6520): 381-386.

Morikawa, K., O. Matsumoto, et al. (1992). "X-ray structure of T4 endonuclease V: an excision repair enzyme specific for a pyrimidine dimer." Science **256**(5056): 523-526.

Morland, I., L. Luna, et al. (2005). "Product inhibition and magnesium modulate the dual reaction mode of hOgg1." DNA Repair (Amst) **4**(3): 381-387.

Mutamba, J. T., D. Svilar, et al. (2011). "XRCC1 and base excision repair balance in response to nitric oxide." DNA Repair (Amst) **10**(12): 1282-1293.

Nagashima, M., A. Sasaki, et al. (1997). "Presence of human cellular protein(s) that specifically binds and cleaves 8-hydroxyguanine containing DNA." Mutation research **383**(1): 49-59.

Nash, H. M., S. D. Bruner, et al. (1996). "Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily." Curr Biol **6**(8): 968-980.

Nash, H. M., R. Lu, et al. (1997). "The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOgg1: direct identification, ablation and chemical reconstitution." Chem Biol **4**(9): 693-702.

Neeley, W. L. and J. M. Essigmann (2006). "Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products." Chemical Research in Toxicology **19**(4): 491-505.

Nghiem, Y., M. Cabrera, et al. (1988). "The mutY gene: a mutator locus in Escherichia coli that generates G.C----T.A transversions." Proceedings of the National Academy of Sciences of the United States of America **85**(8): 2709-2713.

Nilsen, L., R. J. Forstrom, et al. (2012). "AP endonuclease independent repair of abasic sites in Schizosaccharomyces pombe." Nucleic Acids Res **40**(5): 2000-2009.

Norman, D. P., S. D. Bruner, et al. (2001). "Coupling of substrate recognition and catalysis by a human base-excision DNA repair protein." J Am Chem Soc **123**(2): 359-360.

Norman, D. P., S. J. Chung, et al. (2003). "Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase." Biochemistry **42**(6): 1564-1572.

O'Brien, P. J. (2006). "Catalytic promiscuity and the divergent evolution of DNA repair enzymes." Chemical reviews **106**(2): 720-752.

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

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

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

O'Connor, T. R. (1993). "Purification and characterization of human 3-methyladenine-DNA glycosylase." Nucleic Acids Res **21**(24): 5561-5569.

O'Rourke, E. J., C. Chevalier, et al. (2000). "A novel 3-methyladenine DNA glycosylase from helicobacter pylori defines a new class within the endonuclease III family of base excision repair glycosylases." J Biol Chem **275**(26): 20077-20083.

Ortega-Galisteo, A. P., T. Morales-Ruiz, et al. (2008). "Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks." Plant molecular biology **67**(6): 671-681.

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

Parikh, S. S., C. D. Mol, et al. (1998). "Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA." EMBO J **17**(17): 5214-5226.

Parikh, S. S., G. Walcher, et al. (2000). "Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects." Proc Natl Acad Sci U S A **97**(10): 5083-5088.

Parker, J. B., M. A. Bianchet, et al. (2007). "Enzymatic capture of an extrahelical thymine in the search for uracil in DNA." Nature **449**(7161): 433-437.

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

Penterman, J., D. Zilberman, et al. (2007). "DNA demethylation in the Arabidopsis genome." Proceedings of the National Academy of Sciences of the United States of America **104**(16): 6752-6757.

Pereira de Jesus, K., L. Serre, et al. (2005). "Structural insights into abasic site for Fpg specific binding and catalysis: comparative high-resolution crystallographic studies of Fpg bound to various models of abasic site analogues-containing DNA." Nucleic Acids Res **33**(18): 5936-5944.

Porecha, R. H. and J. T. Stivers (2008). "Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils." Proc Natl Acad Sci U S A **105**(31): 10791-10796.

Purmal, A. A., Y. W. Kow, et al. (1994). "Major oxidative products of cytosine, 5-hydroxycytosine and 5-hydroxyuracil, exhibit sequence context-dependent mispairing in vitro." Nucleic acids research **22**(1): 72-78.

Qi, Y., M. C. Spong, et al. (2009). "Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme." Nature **462**(7274): 762-766.

Qi, Y., M. C. Spong, et al. (2010). "Entrapment and structure of an extrahelical guanine attempting to enter the active site of a bacterial DNA glycosylase, MutM." The Journal of biological chemistry **285**(2): 1468-1478.

Radicella, J. P., C. Dherin, et al. (1997). "Cloning and characterization of hOGG1, a human homolog of the OGG1 gene of Saccharomyces cerevisiae." Proceedings of the National Academy of Sciences of the United States of America **94**(15): 8010-8015.

Radom, C. T., A. Banerjee, et al. (2007). "Structural characterization of human 8-oxoguanine DNA glycosylase variants bearing active site mutations." J Biol Chem **282**(12): 9182-9194.

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

Rhind, N., Z. Chen, et al. (2011). "Comparative functional genomics of the fission yeasts." Science **332**(6032): 930-936.

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

Robey-Bond, S. M., R. Barrantes-Reynolds, et al. (2008). "Clostridium acetobutylicum 8-oxoguanine DNA glycosylase (Ogg) differs from eukaryotic Oggs with respect to opposite base discrimination." Biochemistry **47**(29): 7626-7636.

Rocchia, W., S. Sridharan, et al. (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." J Comput Chem **23**(1): 128-137.

Roldan-Arjona, T., Y. F. Wei, et al. (1997). "Molecular cloning and functional expression of a human cDNA encoding the antimutator enzyme 8-hydroxyguanine-DNA glycosylase." Proceedings of the National Academy of Sciences of the United States of America **94**(15): 8016-8020.

Rosenquist, T. A., D. O. Zharkov, et al. (1997). "Cloning and characterization of a mammalian 8-oxoguanine DNA glycosylase." Proceedings of the National Academy of Sciences of the United States of America **94**(14): 7429-7434.

Rubinson, E. H., S. Adhikary, et al. (2010). Structural Studies of Alkylpurine DNA Glycosylases. ACS Symposium Series : Structural Biology of DNA Damage and Repair. M. P. Stone. Washingon, D.C., American Chemical Society. **1041:** 29-45.

Rubinson, E. H. and B. F. Eichman (2012). "Nucleic acid recognition by tandem helical repeats." Current opinion in structural biology **22**(1): 101-109.

Rubinson, E. H., A. S. Gowda, et al. (2010). "An unprecedented nucleic acid capture mechanism for excision of DNA damage." Nature **468**(7322): 406-411.

Rubinson, E. H., A. H. Metz, et al. (2008). "A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD." J Mol Biol **381**(1): 13-23.

Samson, L. and J. Cairns (1977). "A new pathway for DNA repair in Escherichia coli." Nature **267**(5608): 281-283.

Saparbaev, M., K. Kleibl, et al. (1995). "*Escherichia coli*, *Saccharomyces cerevisiae*, rat and human 3-methyladenine DNA glycosylases repair 1,N^6-ethenoadenine when present in DNA." Nucleic Acids Res **23**(18): 3750-3755.

Saparbaev, M., S. Langouet, et al. (2002). "1,N(2)-ethenoguanine, a mutagenic DNA adduct, is a primary substrate of Escherichia coli mismatch-specific uracil-DNA glycosylase and human alkylpurine-DNA-N-glycosylase." The Journal of biological chemistry **277**(30): 26987-26993.

Saparbaev, M. and J. Laval (1994). "Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases." Proc Natl Acad Sci U S A **91**(13): 5873-5877.

Sartori, A. A., G. M. Lingaraju, et al. (2004). "Pa-AGOG, the founding member of a new family of archaeal 8-oxoguanine DNA-glycosylases." Nucleic acids research **32**(22): 6531-6539.

Savva, R., K. McAuley-Hecht, et al. (1995). "The structural basis of specific base-excision repair by uracil-DNA glycosylase." Nature **373**(6514): 487-493.

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

Sedgwick, B. (2004). "Repairing DNA-methylation damage." Nat Rev Mol Cell Biol **5**(2): 148-157.
Seeberg, E., L. Eide, et al. (1995). "The base excision repair pathway." Trends Biochem Sci **20**(10): 391-397.
Serre, L., K. Pereira de Jesus, et al. (2002). "Crystal structure of the Lactococcus lactis formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA." EMBO J **21**(12): 2854-2865.
Shi, G., D. Y. Chang, et al. (2006). "Physical and functional interactions between MutY glycosylase homologue (MYH) and checkpoint proteins Rad9-Rad1-Hus1." The Biochemical journal **400**(1): 53-62.
Shikazono, N., C. Pearson, et al. (2006). "The roles of specific glycosylases in determining the mutagenic consequences of clustered DNA base damage." Nucleic acids research **34**(13): 3722-3730.
Shuker, D. E., E. Bailey, et al. (1987). "The determination of urinary 3-methyladenine in humans as a potential monitor of exposure to methylating agents." Carcinogenesis **8**(7): 959-962.
Shuker, D. E. and P. B. Farmer (1992). "Relevance of urinary DNA adducts as markers of carcinogen exposure." Chem Res Toxicol **5**(4): 450-460.
Singer, B., A. Antoccia, et al. (1992). "Both purified human 1,N6-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N6-ethenoadenine and 3-methyladenine." Proc Natl Acad Sci U S A **89**(20): 9386-9390.
Slabinski, L., L. Jaroszewski, et al. (2007). "XtalPred: a web server for prediction of protein crystallizability." Bioinformatics **23**(24): 3403-3405.
Slot, J. C. and A. Rokas (2011). "Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi." Curr Biol **21**(2): 134-139.
Slupphaug, G., C. D. Mol, et al. (1996). "A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA." Nature **384**(6604): 87-92.
Stamatakis, A. (2006). "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models." Bioinformatics **22**(21): 2688-2690.
Stivers, J. T. (2004). "Site-specific DNA damage recognition by enzyme-induced base flipping." Prog Nucleic Acid Res Mol Biol **77**: 37-65.
Stivers, J. T. (2008). "Extrahelical damaged base recognition by DNA glycosylase enzymes." Chemistry **14**(3): 786-793.
Stivers, J. T. and Y. L. Jiang (2003). "A mechanistic perspective on the chemistry of DNA repair glycosylases." Chem Rev **103**(7): 2729-2759.
Sugahara, M., T. Mikawa, et al. (2000). "Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8." Embo J **19**(15): 3857-3869.
Tchou, J., H. Kasai, et al. (1991). "8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity." Proc Natl Acad Sci U S A **88**(11): 4690-4694.
Tchou, J., M. L. Michaels, et al. (1993). "Function of the zinc finger in Escherichia coli Fpg protein." J Biol Chem **268**(35): 26738-26744.

Thomas, L., C. H. Yang, et al. (1982). "Two DNA glycosylases in Escherichia coli which release primarily 3-methyladenine." Biochemistry **21**(6): 1162-1169.
Tominaga, Y., Y. Ushijima, et al. (2004). "MUTYH prevents OGG1 or APEX1 from inappropriately processing its substrate or reaction product with its C-terminal domain." Nucleic Acids Res **32**(10): 3198-3211.
Truglio, J. J., E. Karakas, et al. (2006). "Structural basis for DNA recognition and processing by UvrB." Nat Struct Mol Biol **13**(4): 360-364.
Tubbs, J. L., V. Latypov, et al. (2009). "Flipping of alkylated DNA damage bridges base and nucleotide excision repair." Nature **459**(7248): 808-813.
Tubbs, J. L. and J. A. Tainer (2010). "Alkyltransferase-like proteins: molecular switches between DNA repair pathways." Cell Mol Life Sci **67**(22): 3749-3762.
Tudek, B. (2003). "Imidazole ring-opened DNA purines and their biological significance." J Biochem Mol Biol **36**(1): 12-19.
Valko, M., C. J. Rhodes, et al. (2006). "Free radicals, metals and antioxidants in oxidative stress-induced cancer." Chemico-biological interactions **160**(1): 1-40.
Vallur, A. C., J. A. Feller, et al. (2002). "Effects of hydrogen bonding within a damaged base pair on the activity of wild type and DNA-intercalating mutants of human alkyladenine DNA glycosylase." J Biol Chem **277**(35): 31673-31678.
van der Kemp, P. A., D. Thomas, et al. (1996). "Cloning and expression in Escherichia coli of the OGG1 gene of Saccharomyces cerevisiae, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine." Proc Natl Acad Sci U S A **93**(11): 5197-5202.
Van Duyne, G. D., R. F. Standaert, et al. (1993). "Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin." J Mol Biol **229**(1): 105-124.
van Loon, B., E. Markkanen, et al. (2010). "Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine." DNA repair **9**(6): 604-616.
Vasen, H. F., P. Watson, et al. (1999). "New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC." Gastroenterology **116**(6): 1453-1456.
Vassylyev, D. G., T. Kashiwagi, et al. (1995). "Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition." Cell **83**(5): 773-782.
Vidal, A. E., I. D. Hickson, et al. (2001). "Mechanism of stimulation of the DNA glycosylase activity of hOGG1 by the major human AP endonuclease: bypass of the AP lyase activity step." Nucleic Acids Res **29**(6): 1285-1292.
Vonrhein, C., E. Blanc, et al. (2007). "Automated structure solution with autoSHARP." Methods Mol Biol **364**: 215-230.
Wiederholt, C. J., M. O. Delaney, et al. (2003). "Repair of DNA containing Fapy.dG and its beta-C-nucleoside analogue by formamidopyrimidine DNA glycosylase and MutY." Biochemistry **42**(32): 9755-9760.
Wolfe, A. E. and P. J. O'Brien (2009). "Kinetic mechanism for the flipping and excision of 1,N(6)-ethenoadenine by human alkyladenine DNA glycosylase." Biochemistry **48**(48): 11357-11369.

Wu, P., C. Qiu, et al. (2003). "Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4." The Journal of biological chemistry **278**(7): 5285-5291.

Wyatt, M. D., J. M. Allan, et al. (1999). "3-methyladenine DNA glycosylases: structure, function, and biological importance." Bioessays **21**(8): 668-676.

Yamagata, Y., M. Kato, et al. (1996). "Three-dimensional structure of a DNA repair enzyme, 3-methyladenine DNA glycosylase II, from Escherichia coli." Cell **86**(2): 311-319.

Yang, C. G., C. Yi, et al. (2008). "Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA." Nature **452**(7190): 961-965.

Yoshida, M., K. Makino, et al. (1997). "Substrate and mispairing properties of 5-formyl-2'-deoxyuridine 5'-triphosphate assessed by in vitro DNA polymerase reactions." Nucleic acids research **25**(8): 1570-1577.

Zhang, Y., F. Yuan, et al. (2005). "Reconstitution of 5'-directed human mismatch repair in a purified system." Cell **122**(5): 693-705.

Zhao, B. and P. J. O'Brien (2011). "Kinetic mechanism for the excision of hypoxanthine by Escherichia coli AlkA and evidence for binding to DNA ends." Biochemistry **50**(20): 4350-4359.

Zharkov, D. O., G. Golan, et al. (2002). "Structural analysis of an Escherichia coli endonuclease VIII covalent reaction intermediate." EMBO J **21**(4): 789-800.

Zharkov, D. O. and A. P. Grollman (2005). "The DNA trackwalkers: principles of lesion search and recognition by DNA glycosylases." Mutat Res **577**(1-2): 24-54.

Zharkov, D. O., G. V. Mechetin, et al. (2010). "Uracil-DNA glycosylase: Structural, thermodynamic and kinetic aspects of lesion search and recognition." Mutat Res **685**(1-2): 11-20.

Zharkov, D. O., T. A. Rosenquist, et al. (2000). "Substrate specificity and reaction mechanism of murine 8-oxoguanine-DNA glycosylase." J Biol Chem **275**(37): 28607-28617.

Zharkov, D. O., G. Shoham, et al. (2003). "Structural characterization of the Fpg family of DNA glycosylases." DNA Repair (Amst) **2**(8): 839-862.

Zhu, X., X. Yan, et al. (2012). "A model for 3-methyladenine recognition by 3-methyladenine DNA glycosylase I (TAG) from Staphylococcus aureus." Acta crystallographica. Section F, Structural biology and crystallization communications **68**(Pt 6): 610-615.