Chapter 3

Structural Studies of Alkylpurine DNA Glycosylases

Emily H. Rubinson, Suraj Adhikary, and Brandt F. Eichman*

Department of Biological Sciences and Center For Structural Biology, Vanderbilt University, Nashville, TN 37232
***brandt.eichman@vanderbilt.edu**

Alkylation of DNA bases produces a broad spectrum of cytotoxic and mutagenic lesions that are removed from the genome by alkylpurine DNA glycosylases. These DNA repair enzymes exist in eukaryotes, archaea, and bacteria, and have varied, well-defined specificities for particular alkylpurine nucleobases. Crystal structures of these enzymes in complex with DNA and alkylated bases have illuminated some of the chemical determinants for selection of damage amidst a vast background of normal DNA. However, only now are we beginning to understand the basis for alkylpurine specificity. Here, we review the structures of alkylpurine DNA glycosylases determined to date. Comparison of these structures in the context of functional data provides insight into the mechanisms of alkylpurine selection and excision.

Introduction

Humans are exposed to alkylating agents from various environmental sources, including industrial processes, cigarette smoke, diet, and chemotherapy. These agents, in addition to endogenous methyl donors, chemically modify the nucleobases of DNA to produce a variety of cytotoxic and mutagenic lesions that disrupt DNA replication and thus lead to heritable diseases and cancer (reviewed in (*1*)). To maintain genomic integrity amidst the constant threat of DNA alkylation, all organisms have devised multiple DNA repair strategies to eliminate the damage. Bases methylated at exocyclic substituents (e.g., O^6-methylguanine) are directly demethylated by DNA methyltransferases, whereas ring-substituted

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

1-methyladenine (1mA) and 3-methylcytosine (3mC) are specifically repaired through oxidative deamination by DNA dioxygenase (reviewed in (*2*)). The majority of alkylated bases, however, are eliminated from the genome by the base excision repair (BER) pathway (reviewed in (*3*, *4*)). DNA glycosylases initiate BER by locating the damaged base and catalyzing the hydrolysis of the C1′-N glycosylic bond that links the base to the phosphoribose backbone. The resulting abasic site is further processed by apurinic (AP) endonuclease, DNA polymerase and DNA ligase, acting sequentially to restore the DNA to an undamaged state.

DNA glycosylases that remove alkylation damage have been characterized from eukaryotes, archaea, and bacteria. These include mammalian alkyladenine DNA glycosylase (AAG) (*5*, *6*), yeast methyladenine DNA glycosylase (*S. cerevisiae* MAG and *S. pombe* MagI) (*7–9*), *E. coli* 3-methyladenine (3mA) DNA glycosylase I (TAG) and II (AlkA) (*10*, *11*), *Thermotoga maritima* methylpurine DNA glycosylase II (MpgII) (*12*), *Helicobacter pylori* 3mA DNA glycosylase (MagIII) (*13*), and most recently *Bacillus cereus* AlkC and AlkD (*14*). Whereas most DNA glycosylases are specific for a single modification, alkylpurine DNA glycosylases can recognize a chemically diverse set of lesions (Figure 1), 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 (*15–17*). TAG and MagIII are highly specific for 3mA and 3mG (*13*, *18*), MpgII and AlkC/D are selective for positively charged lesions 3mA and 7mG (*12*, *14*), and AlkA and AAG can excise these lesions as well as other alkylated and modified bases, including εA and hypoxanthine (Hx) (*19–21*).

Alkylpurine DNA glycosylases can be classified into three distinct superfamilies based on their three-dimensional structures (Figure 2). The first is defined by the mixed α/β globular fold of AAG (*22*), which bears no structural resemblance to any other protein in the Protein Data Bank (PDB). A second, and by far the most common structural class, is the helix-hairpin-helix (HhH) superfamily of glycosylases and includes AlkA, TAG, MagIII, and MpgII (*12*, *23–25*). 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 (*26–29*). *S. cerevisiae* MAG and *S. pombe* Mag1 likely adopt the HhH fold based on sequence similarity to AlkA (*7*, *8*). A third alkylpurine DNA glycosylase architecture was identified recently from the AlkC and AlkD proteins from *Bacillus cereus*. AlkD forms a C-shaped α-helical fold from repeating HEAT motifs, and AlkC is expected to adopt a similar fold (*30*, *31*).

Although structurally divergent, AAG and HhH glycosylases have evolved a conserved base-flipping mechanism for gaining access to damaged nucleobases in DNA (reviewed in (*32*)). Base flipping is common among DNA processing enzymes, and allows the protein to correctly identify and orient the substrate for catalysis. Recognition of the substrate base is believed to proceed in two stages—processive interrogation of the DNA duplex through non-specific, electrostatic interactions, followed by base flipping of the target base into the active site of the enzyme (*33*, *34*). The active sites of AAG and HhH glycosylases consist of a concave pocket lined with aromatic side chains that base-stack

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

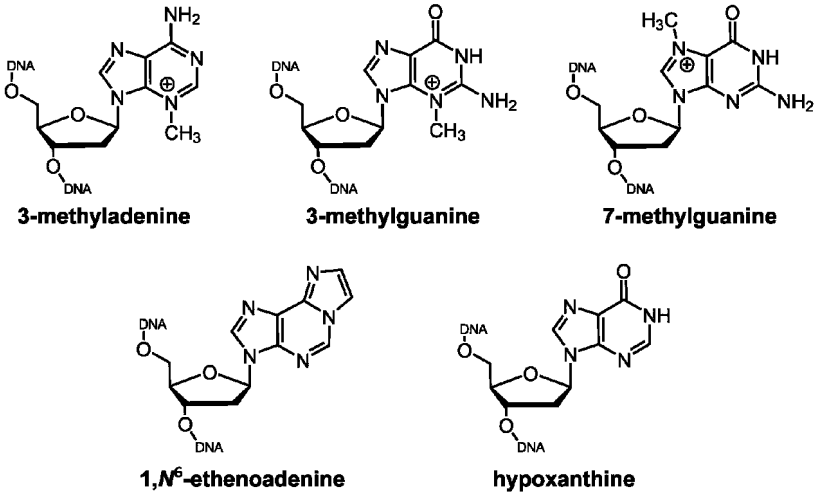

Figure 1. Nucleobases excised by alkylpurine DNA glycosylases.

with the flipped alkylpurine nucleobase, and most contain an ionizable side chain essential for catalysis. In order to stabilize the extrahelical nucleobase conformation, these glycosylases fill the gap left in the DNA by intercalating a set of side chains into the helical base stack. Although it remains to be determined if the HEAT glycosylases flip damaged bases, AlkD also contains an aromatic, electron-rich cleft but lacks identifiable intercalating residues typical of the other alkylpurine glycosylases.

Despite progress in the field, the mechanisms by which DNA glycosylases select for a particular alkyl modification are not well understood. The importance of substrate specificity is underscored by the fact that these enzymes must locate very subtle modifications among a vast excess of normal base pairs. The diversity in their structural features despite overlapping functions presents an opportunity to understand the physical and chemical determinants of DNA alkylation damage recognition and removal. In this review, we compare the alkylpurine DNA glycosylase structures determined to date, and discuss the structural implications on enzyme specificity and catalysis.

Human Alkyladenine DNA Glycosylase AAG

AAG excises a broad range of alkylpurines, including 3mA and 7mG, and has a selective preference for neutral εA and Hx (*35*). The crystal structure of AAG in complex with DNA containing an abasic pyrrolidine transition-state analog showed that AAG is a single domain protein with a mixed α/β structure and a positively charged DNA binding surface (*22*, *36*). The protein crystallized lacked the N-terminal 79 amino acids and thus the presence of a second domain is unknown. The DNA is bent at the damage site by ~22°, with B-form helical arms swung away from the protein. The pyrrolidine is rotated out of the DNA duplex and into a cavity on the protein surface. Tyr162 on the tip of a β-hairpin plugs the gap in the DNA left by the flipped pyrrolidine, and presumably stabilizes the

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

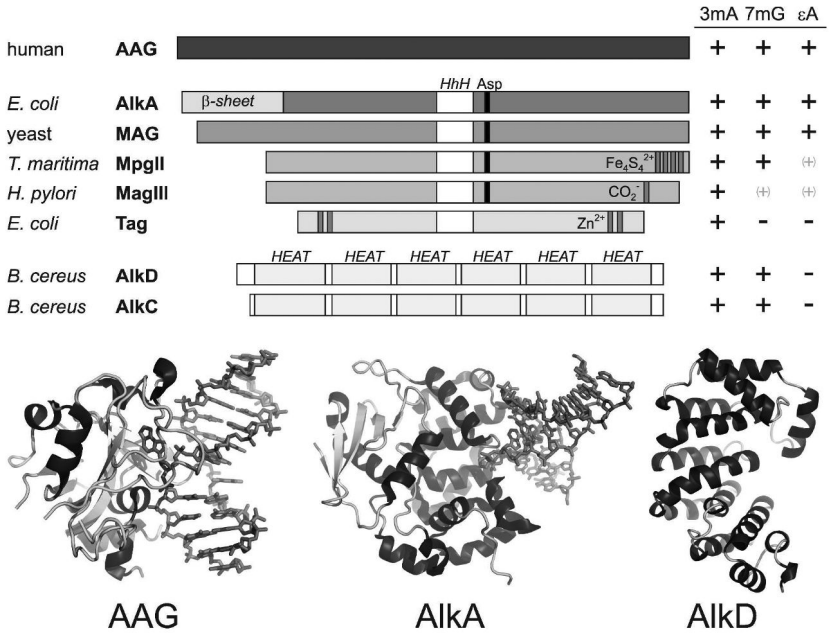

Figure 2. Structural superfamilies of alkylpurine DNA glycosylases. The three families are defined by the structural folds of human AAG, the helix-hairpin-helix (HhH) superfamily typified by AlkA, and HEAT-repeat proteins AlkC and AlkD. Substrate specificities are shown to the right of the schematic, and the crystal structures of representative proteins from each superfamily are shown at the bottom and shaded according to secondary structure. DNA is shown as sticks.

distorted conformation of the extrahelical DNA. The pyrrolidine binding pocket is lined with aromatic and polar residues. A subsequent crystal structure of AAG bound to εA-containing DNA showed the flipped εA base to be stacked between two tyrosine residues (Tyr127 and Tyr159) and His136 inside the active site cavity (Figure 3A) (*37*).

The AAG/DNA complexes have been important for directing biochemical studies aimed at understanding the molecular basis for AAG's substrate specificity and catalytic mechanism. Discrimination against normal purines is most likely due to their proper base pairing in the DNA duplex and to unfavorable interactions with exocyclic N6 and N2 amino groups inside the active site (*35*). For example, His136 donates a hydrogen bond to N6 of εA, whereas adenine cannot accept a hydrogen bond at this position. Furthermore, guanine is likely to be excluded on the basis of a steric clash between its exocyclic N2 amino group, which is absent in εA, Hx, and adenine, and the side chain of Asn169. In support of this, mutation of Asn169 gives AAG enhanced activity toward guanine (*38*). It has been suggested that AAG removes charged alkylpurine lesions because of their inherent instability and not through a structural recognition of the methyl group *per se*. Indeed, AAG's rate enhancement for excision of 3mA is one and three orders of magnitude less than that of εA and Hx, respectively (*35*). Regarding catalysis, an ordered water molecule sits adjacent to the N-glycosylic bond and is hydrogen bonded to the

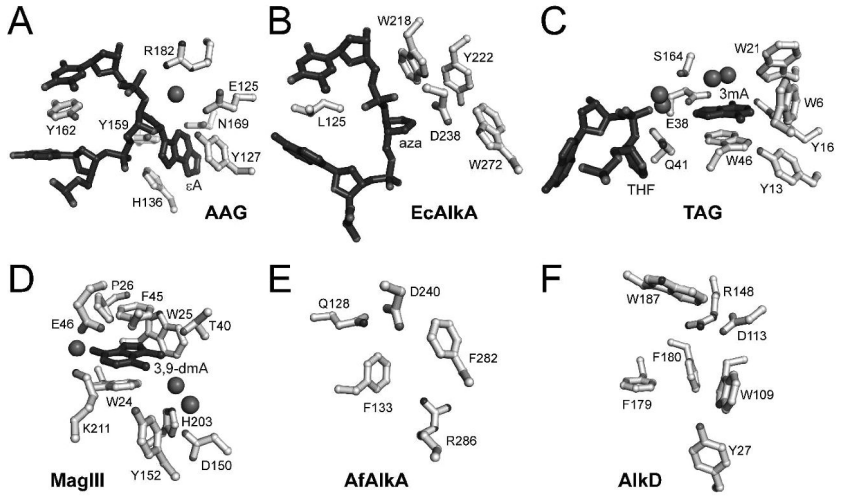

Figure 3. Active sites of alkylpurine DNA glycosylases. Protein and nucleic acid atoms are shaded black and grey, respectively. A. Human AAG in complex with εA-DNA, PDB 1ewn (37). B. E. coli AlkA bound to 1-azaribose-DNA, PDB 1diz (38). C. E. coli TAG/THF-DNA/3mA complex, PDB 2ofi (39). D. H. pylori MagIII bound to 3,9dma, PDB 1pu7 (25). E. A fulgidus AlkA, PDB 2jhj (40). F. B. cereus AlkD, PDB 3bvs (30).

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 S_N2 catalytic mechanism (*39*).

Helix-Hairpin-Helix Superfamily

The HhH glycosylases contain two α-helical subdomains separated by an active site cleft that accommodates the flipped substrate nucleobase (Figure 4). One of these domains (helices αD-αJ) is highly conserved and contains the HhH motif (αI-αJ), a DNA binding platform utilized by hundreds of repair proteins (*40*). The HhH anchors the protein to the DNA through electrostatic interactions between main-chain atoms from the hairpin region and the phosphoribose backbone. The HhH domain also contributes a bulky group (typically a Leu, Asn, or Gln side chain) that plugs the gap in the DNA left by the flipped-out nucleotide, and a second side chain (Phe, Tyr, Leu, or Pro) that wedges between the bases opposite the flipped out nucleotide (Figure 4). Both plug and wedge residues are important for stabilizing the bent conformation of the DNA, and the wedge residue has been implicated in probing the DNA helix during the search process (*34*). The second domain, formed from the N- and C-termini (the N/C domain, helices αB-αC and αK-αM), is more varied in structure and often contains additional structural elements, including a zinc binding motif (TAG), a carbamylated lysine (MagIII), and an iron-sulfur cluster (MpgII) (Figure 4). The

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

precise role of the N/C domain is not clear, but it is suspected that these elements help fold this domain in order to form the active site cleft.

The shape and chemical features of the active site cleft play a large role in defining the substrate specificity of these enzymes (*25*). Like AAG, HhH alkylpurine glycosylases contain aromatic, electron-rich nucleobase binding pockets that stack against ring-substituted purines. Methylated or protonated purines have enhanced π-orbital overlap between the modified base and the aromatic side chains (*42*), suggesting that the methyl group may be sensed by enhanced base stacking interactions in addition to, or in lieu of, a direct van der Waals interaction with the methyl group itself. With the exception of TAG, the HhH glycosylases contain a conserved, catalytically essential aspartic acid residue at the mouth of the active site.

E. coli AlkA

Crystal structures of *E. coli* AlkA revealed that the HhH architecture, first observed in *E. coli* Endo III (*26*), is also present in the alkylpurine DNA glycosylases (*23*, *43*). AlkA lacks the iron-sulfur cluster present in EndoIII and MutY, and instead contains an amino-terminal β-sheet domain that has no identified function but presumably stabilizes the overall fold. Crystal structures of AlkA in complex with DNA containing a 1-azaribose abasic site illuminated how DNA glycosylases utilize the HhH motif to anchor the protein to the DNA (*44*). Although the HhH does not directly participate in lesion recognition, it contributes most of the polar interactions between AlkA and the DNA. The DNA is highly distorted with a ~60° bend and widened minor groove around the site of the lesion. The 1-azaribose is rotated 180° around the phosphoribose backbone and points into a shallow cleft formed by several aromatic side chains (Figure 3B). Leu125 plugs the gap left by the flipped nucleotide in a manner similar to AAG Tyr162, and the hairpin between helices αG and αH wedge into the DNA strand opposite the lesion (Figures 3, 4, and 5).

AlkA's nucleobase binding surface is a shallow cleft that can accommodate a variety of alkylpurines. This open architecture helps explain AlkA's broad specificity. In addition, the substrate methylpurine base presumably stacks against the Trp272 indole ring, enhancing the preference of AlkA for positively charged bases. In the AlkA/DNA complex, rotation of the 1-azaribose into the active site places the N1′ nitrogen directly adjacent to the carboxylate group of the catalytic aspartate (Asp238), leaving no room for a water nucleophile necessary for an S_N2 catalytic mechanism (Figure 3B). This close proximity between the abasic site and Asp238 has led to the suggestion that AlkA utilizes an S_N1-type mechanism, whereby the ionized carboxylate stabilizes the carbocation intermediate formed on the ribose ring during nucleobase hydrolysis (*44*).

E. coli TAG

The 3mA-specific TAG enzyme is a divergent member of the HhH superfamily (*24*). Despite a conserved HhH domain, TAG lacks the conserved

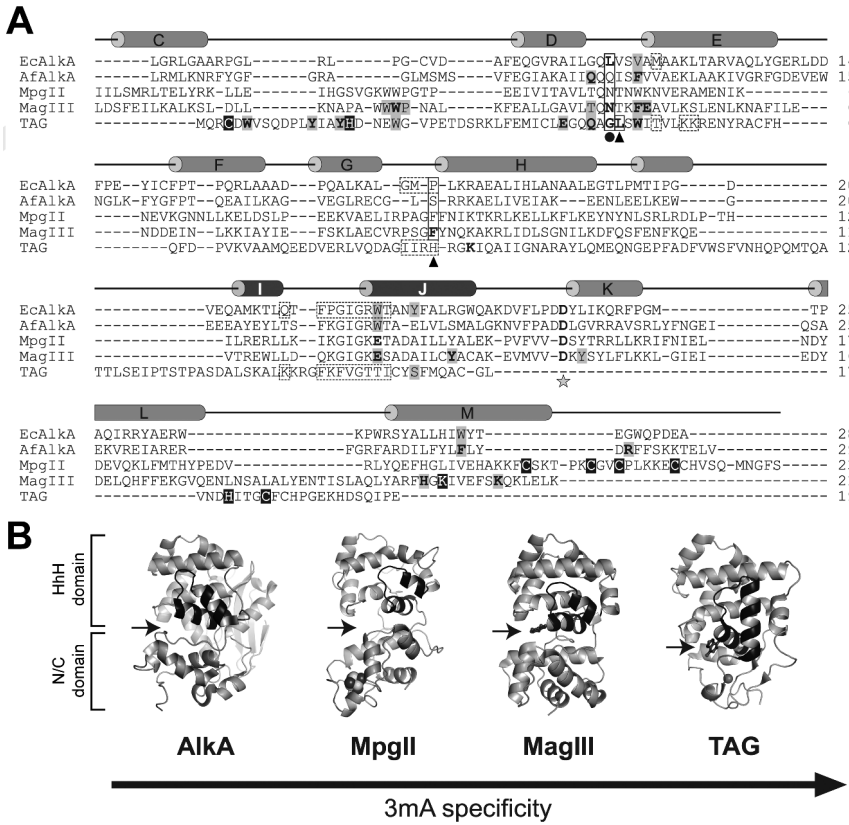

Figure 4. The helix-hairpin-helix superfamily. A. Structure based sequence alignment of E. coli AlkA, A. fulgidus AlkA, B. halodurans Mag, H. pylori MagIII, and S. typhimurium TAG is shown with secondary structure from AlkA. The MpgII sequence was aligned with MagIII. Residues that contact DNA in protein/DNA complexes of AlkA and TAG are highlighted with dashed boxes, and intercalating plug and wedge residues are boxed solid and marked beneath the sequences with a circle and triangle, respectively. Alkylpurine binding pocket residues are highlighted grey, and the catalytic aspartate is labeled with a star. Residues that coordinate ions in TAG (Zn^{2+}), MagIII (carbamylated lysine), and MpgII (iron-sulfur cluster) are shaded black. B. Crystal structures are the same as those in Figure 3 and include B. halodurans Mag, PDB 2h56 (41). The MpgII model was constructed using atomic coordinates from MagIII and MutY (1MUY, (27)) as described in the text. HhH motifs are shaded black, and the substrate base binding pockets are marked with an arrow.

catalytic aspartate present in all other HhH glycosylases, and the sequence and structure of the HhH motif itself is noticeably different (Figure 4). In addition, the N/C domain is devoid of any significant α-helical structure but rather contains a novel zinc binding motif that helps "snap" the N- and C-termini together (*45*). NMR and base perturbation studies revealed that *E. coli* TAG binds 3mA inside a deep pocket that is sterically constrained to exclude 7mG and εA bases (*46*).

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

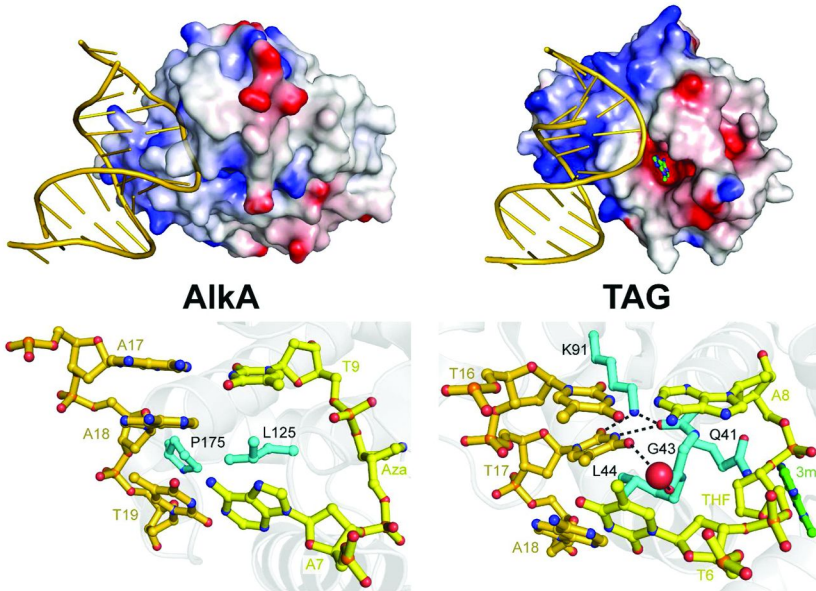

Figure 5. Comparison of TAG and AlkA DNA complexes. The overall structures of AlkA bound to 1-azaribose-containing DNA (left) and of TAG bound to THF-DNA and 3mA (right) are shown at the top, with protein rendered as an electrostatic potential surface (red, negative; blue, positive), DNA as gold cartoon, and 3mA nucleobase ball-and-stick (green carbons). At the bottom is a close-up view of the plug-and-wedge intercalation of the DNA duplexes by the proteins. A bridging water molecule in TAG is depicted as a red sphere. (see color insert)

The recent crystal structure of a TAG/DNA/3mA ternary complex provided insight into how TAG achieves is high selectivity for 3mA (*47*). *S. typhimurium* TAG, which shares 82% sequence identity (91% similarity) to the *E. coli* protein, was crystallized in the presence of free 3mA base and DNA containing a tetrahydrofuran (THF) abasic site (Figures 3C and 5). As in the AlkA/DNA complex, the HhH hairpin contributes most of the electrostatic interactions to the DNA backbone immediately 3′ to the lesion. For the first time in a DNA glycosylase structure, however, the THF moiety is not fully flipped into the active site and does not form any polar interactions with the protein. Instead, the abasic site was observed in the electron density map to interconvert between a stacked position normally found in B-DNA and one in which the ribose is partially rotated ~90° into the minor groove. The DNA is bent by ~65° as a consequence of the intercalating plug and wedge interactions, which in TAG are provided by a single hairpin loop between helices αD and αE. The main-chain of Gly43 plugs the abasic gap and the adjacent Leu44 side chain wedges between the bases across from the lesion (Figures 4 and 5). Despite the kink in the DNA, the helix remains essentially B-form as a result of the lack of specific interactions to the abasic site or to the DNA duplex on the 5′ side of the lesion. Thus, the DNA in the

TAG product complex is less distorted than in the AlkA transition state complex (Figure 5).

In the TAG/DNA crystal structure, the 3mA base resides 8 Å away from the THF moiety and deep inside the active site pocket. The 3mA ring is stacked between Trp46 and several ordered water molecules and is constrained on the sides by hydrogen bonds to Glu38 and Tyr16 and by van der Waals contacts to Trp6 (Figure 3C). Substitution of Trp46 to alanine reduced the rate of 3mA excision 10-fold with respect to wild-type TAG, further highlighting the importance of base stacking on alkylpurine glycosylase activity. The hydrogen bonds between the Glu38 carboxylate and the N6 amino and N7 imino nitrogens of 3mA suggest that 7mG is sterically excluded from the TAG active site (*46*, *47*). Interestingly, mutation of Glu38 to alanine, which should relax this constraint, did not provide TAG the ability to cleave 7mG from DNA. Rather, by two orders of magnitude with respect to the wild-type enzyme, suggesting that Glu38 is important for catalysis (*47*). A catalytic mechanism for 3mA excision has been proposed in which TAG provides a high-affinity base binding pocket that induces strain in the pre-catalytic TAG/DNA-3mA ground state complex (*46*). Release of this strain upon base hydrolysis is illustrated in the crystal structure of the TAG/DNA product complex by the large distance between 3mA and the abasic ribose and by the relatively small distortion to the conformation of the DNA helix as compared to the AlkA transition state complex (Figure 5). TAG's high specificity for 3mA, therefore, may be a result of the intrinsic instability of this lesion and the lack of a general acid or base to drive catalysis, rather than a mere steric exclusion of other bases from the active site.

S. cerevisiae MAG and *S. pombe* Mag1

S. cerevisiae MAG and *S. pombe* Mag1 share 42% and 47% overall sequence similarity to *E. coli* AlkA. Despite the similarity, MAG and Mag1 are less versatile than AlkA in their ability to excise a wide range of substrates. 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 (*19*, *48–51*). In addition, whereas MAG protects yeast cells against the toxic effects of alkylating agents and restores MMS resistance to *E. coli tag alka* mutants, *S. pombe mag1* mutants are only moderately sensitive to methylation damage (*7–9*, *52*). Interestingly, Mag1 expression is not induced by exposure to alkylating agents to the same extent as AlkA and MAG (*9*). These and other reports suggest that MAG and Mag1 play different roles in protection of yeast against alkylation damage than do the bacterial glycosylases.

The structures of MAG or Mag1 have not been determined. However, a search of MAG/Mag1 sequences against structures in the Protein Databank revealed an unpublished structure of a Mag ortholog from *Bacillus halodurans* (BhMag) determined by the Joint Center for Structural Genomics (PDB ID 2h56, (*41*)). This protein is a clear member of the HhH superfamily of DNA glycosylases (Figure 4), and shares 27% sequence identity and 65% overall similarity with the yeast proteins, although it has not been functionally characterized. Structural

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

alignment of BhMag and AlkA shows a strong conservation in active site residues (Figure 4). The most notable difference in the base binding cleft is the presence of a methionine, which is invariant among MAG/Mag1 sequences, in place of the Tyr222 in AlkA (Figure 6A). It remains to be determined if this substitution accounts for the substrate specificity difference between AlkA and MAG/Mag1 proteins.

The asymmetric unit of the BhMag crystals contained three independent copies of the protein that adopted one of two distinct conformations. Superposition of the HhH domains of the two different conformations showed an approximate 30° rotation of the N/C-domains with respect to one another (Figure 6B). This domain displacement moves the N/C-domain into a position superimposable with AlkA, and would thus presumably allow the protein to interact more favorably with DNA. To our knowledge, this is the first observation of movement of the HhH and N/C domains in HhH glycosylases with respect to one another, and suggests that Mag orthologs might change conformation upon binding DNA. The lack of movement of the N/C-domain upon AlkA binding DNA is likely due to the presence of the unique β-sheet subdomain packed against the back-side of both HhH and N/C-domains. In the case of TAG, movement of the N/C domain was not expected since it did not contact the DNA. The effect of DNA on Mag protein structure awaits further investigation. Both MAG and Mag1 will be critical to our understanding of alkylation damage specificity by the highly divergent HhH family of DNA glycosylases.

H. pylori MagIII

MagIII and MpgII are two related prokaryotic alkylpurine glycosylases identified by their sequence similarity to EndoIII (*12*, *13*). The crystal structure of MagIII provided additional insight into 3mA excision and specificity (*25*). As predicted, MagIII's HhH domain structure is most similar to the iron-sulfur containing EndoIII and MutY glycosylases. Instead of a metal center, the N/C domain of MagIII contains a carbamylated lysine (Lys205) that forms extensive electrostatic interactions and likely stabilizes the protein fold (Figure 4). The interface of HhH and N/C domains forms a deep, electronegative nucleobase binding pocket lined with aromatic residues and perfectly shaped to provide a snug fit for 3mA. Structures of MagIII bound to positively charged 3,9-dimethyladenine (3,9-dmA) and neutral εA bases showed that nucleobases stack between Phe45 and Trp24 at the faces and between Trp25, Pro26, and Lys211 around the edges (Figure 3D). Other than van der Waals and π-stacking interactions, there are no specific contacts to the adenine rings like those observed in TAG. Superposition of 7mG onto the 3,9-dmA and εA rings shows that 7mG is sterically excluded from the MagIII pocket, but that the guanine ring would be accommodated by an outward rotation the Lys211 side chain. Importantly, the MagIII structures show that specific protein-nucleobase hydrogen bonds are not necessary for 3mA specificity (*25*).

In contrast to TAG, MagIII is able to excise mispaired 7mG•T and εA•C from DNA (*25*), albeit at a much lower level than AlkA and AAG (*35*, *53*). MagIII's

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

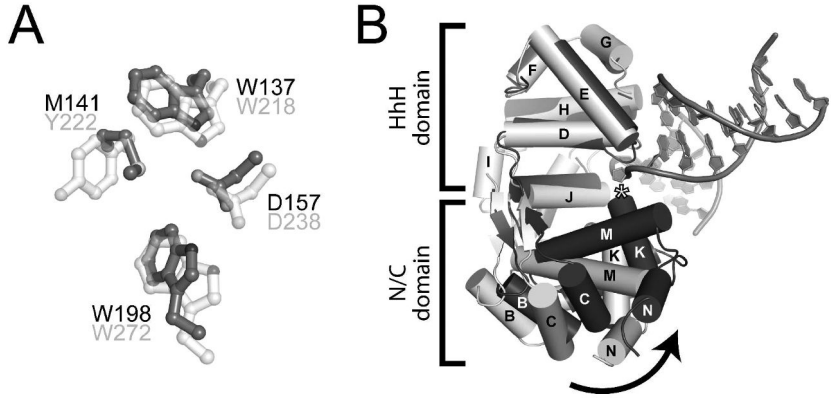

Figure 6. The crystal structure of B. halodurans Mag. A. Superposition of BhMag (black) and AlkA (grey) shows putative nucleobase binding residues are identical except for one substitution. B. Superposition of the HhH domains of the two distinct conformations of BhMag (black, grey). DNA from the AlkA/DNA crystal structure was docked onto the structure by structural alignment of the BhMag and AlkA proteins. The position of the catalytic aspartic acid is highlighted with an asterisk. The 30° rotation of the N/C-domain toward the active site of the protein places it in position to favorably interact with DNA.

weak activity for εA is likely provided by the presence of the conserved catalytic aspartate (Asp150), because mutation of this residue reduces 7mG and εA excision below measureable levels. Interestingly, the D150N mutant exhibits only a 20-fold reduction in the rate of 3mA excision. This residual 3mA activity in the aspartate mutant is further evidence that little catalytic assistance is required for hydrolysis of the labile 3mA glycosylic bond.

T. maritima MpgII

Unlike MagIII, MpgII shows robust activity toward 7mG, which is intriguing given the sequence similarity between MagIII and MpgII (Figure 4) (*12*). MpgII differs from MagIII in only two residues within the active site, Trp52 and Lys53 (Phe45 and Glu46 in MagIII, respectively). Additionally, the N/C domain of MpgII contains an iron-sulfur cluster which is absent in MagIII (Figure 4). To help understand the specificity difference between these two enzymes, we generated an MpgII model using the crystal structures of MagIII and MutY as templates (Figure 4). A chimeric MagIII-MutY template was constructed by superposition of their HhH domains followed by fusion of polypeptides from MutY residues 1-18, 171-177, 191-224 to MagIII residues 22-176 and 196-211. The MpgII sequence was threaded onto the template structure and energy minimized using the Swiss-PDBViewer and SWISS-MODEL (*54*). This model predicts that the MpgII active site is less spatially constrained than MagIII as a result of electrostatic repulsion between Lys53 and Lys204 (normally a Glu46-Lys211 salt bridge in MagIII). Substituting MagIII Glu46 with lysine to mimic the MpgII enzyme resulted in an 8-fold increase in 7mG•T activity, suggesting that steric

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

exclusion of 7mG partially accounts MagIII's low activity toward methylguanine bases (*25*).

MpgII is the only alkylpurine-specific DNA glycosylase that contains an iron-sulfur cluster. Iron-sulfur clusters play diverse enzymatic roles and are found in a variety of DNA processing enzymes (*55–60*), although their function in DNA repair enzymes remains unclear. The iron-sulfur clusters in MutY and EndoIII glycosylases are ~15 Å from both the active site and the bound DNA (*61*, *62*). These structural observations, together with biochemical studies on MutY and EndoIII, suggest that iron-sulfur clusters play purely structural roles in DNA glycosylases (*63*, *64*). It is intriguing to speculate that the iron-sulfur cluster contributes to MpgII's enhanced activity toward more stable εA lesions.

A. fulgidus AlkA

The major single-stranded DNA alkylation products 1mA and 3mC are repaired by oxidative deamination by AlkB (*65*, *66*), a DNA dioxygenase found in bacteria and mammals but with no known orthologs in archaea. 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 (*67–69*). The crystal structure of AfAlkA reveals a globular, three-domain architecture with a HhH domain similar to *E. coli* AlkA (EcAlkA) (*69*). However, the putative substrate binding pocket of AfAlkA is markedly different than that of EcAlkA (Figure 3E). The flipped substrate base is predicted to base stack between Phe133 and Phe282 in a manner similar to that observed in MagIII (Figure 3D). Substitution of either residue with alanine impairs εA and 1mA base excision, and the double mutant almost abolishes glycosylase activity, demonstrating the importance of base stacking to AfAlkA activity. Arg286 is predicted to help orient εA in the active site through hydrogen bonding, but would potentially repel the protonated amine groups of 1mA and 3mC (*69*). Mutation of Asp240, which is structurally equivalent to catalytic Asp238 in EcAlkA, results in a complete loss of function. AfAlkA represents the first reported glycosylase to have activity towards 1mA and 3mC, and further investigation will be required to determine the specificity .

HEAT Repeat Glycosylases AlkC and AlkD

Two alkylpurine DNA glycosylases, AlkC and AlkD, were identified in *Bacillus cereus* as functional complements to *E. coli* AlkA (*14*). The sequences of AlkC and AlkD are distantly related (Figure 7A) and are distinct from other known proteins. Both specifically excise positively charged bases and have no measureable activity toward εA or Hx. AlkC is highly specific for 3mA and 3mG, while AlkD also efficiently removes 7mG from DNA. The high resolution crystal structure of *Bacillus cereus* AlkD shows that the protein adopts a C-shaped globular fold composed exclusively of helical HEAT-like repeats (Figure 7B), and thus represents an unprecedented DNA glycosylase architecture (*30*). HEAT motifs are common protein binding domains that have been adapted by AlkD to

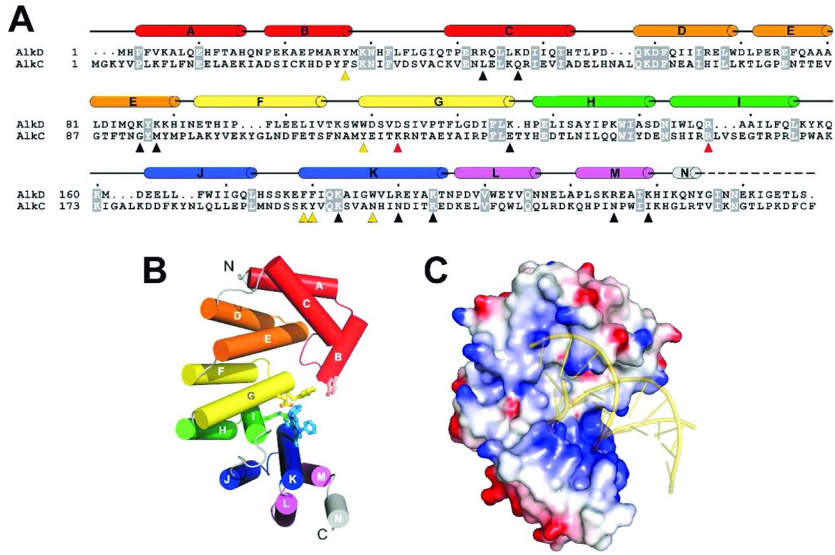

Figure 7. The AlkC/AlkD superfamily. A. Sequence alignment of B. cereus AlkC and AlkD. The secondary structure of AlkD is shown schematically at the top, and colored by heat repeat. Invariant residues are highlighted grey. Triangles denote AlkD residues identified from the crystal structure: putative active site (yellow), catalytic D113 and R148 (red), and positive charges lining the concave cleft (black). B. Crystal structure of AlkD, colored as in A. Putative active site side chains are shown as ball-and-stick. C. Electrostatic surface potential of AlkD (red, negative; blue, positive). DNA (transparent gold cartoon) modeled onto the surface highlights that the concave electropositive surface of AlkD is complementary in shape and charge to B-form DNA duplex. (see color insert)

bind DNA. The C-terminal α-helix of each HEAT repeat forms the inner, concave surface of the protein and contains lysine or arginine residues at conserved positions. Consequently, the concave surface of AlkD is positively charged and perfectly shaped to accommodate a DNA duplex (Figure 7C).

At the heart of AlkD's concave cleft is a cluster of aromatic and charged residues that resembles the active sites of other alkylpurine DNA glycosylases, implicating this region in catalysis (Figures 3F and 7B). Most notably, Asp113 and Arg148, which form an electrostatic bond at the rear of this shallow cleft, are essential for 7mG excision (*30*). Additionally, AlkD mutants with substitutions at Asp113, Arg148, Trp109, or Trp187 fail to complement the MMS sensitive phenotype of an *E. coli tag alkA* strain and decrease or abolish 3mA excision from a methylated genomic DNA substrate (*31*). Despite the evidence for an AlkD active site, the mechanism of base excision remains to be determined.

There are important structural differences which distinguish AlkD from other glycosylases and argue against a base-flipping mechanism. In all other glycosylases, the catalytic aspartate is never involved in a salt bridge and typically resides at the mouth of the active site, and is thus positioned in close proximity to the glycosylic bond of the flipped substrate base. A theoretical

In Structural Biology of DNA Damage and Repair; Stone, M.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2010.

model of DNA docked onto AlkD illustrates that the phosphate backbone of a B-DNA duplex would electrostatically contact the Asp113 side chain. In order to accommodate a flipped base, the DNA backbone would sit farther away from the protein and consequently would lose favorable electrostatic interactions (Figure 7C). In support of this, D113N and R148A mutants cause a two-fold increase and decrease in DNA binding affinity, respectively, over wild-type AlkD. In addition, AlkD binds with the same affinity to oligonucleotides containing either G, 7mG, THF, 1-azaribose, or pyrrolidine, suggesting that the protein does not specifically interact with the lesion (*30*). These observations and the unique protein architecture imply that AlkD utilizes a novel strategy to manipulate DNA in its search for alkylpurine bases. Structures of AlkD in complex with DNA should help resolve this issue.

Summary

The structures of alkylpurine DNA glycosylases define three of the six known protein folds of DNA glycosylases, and show that various protein architectures can be used to create a DNA binding platform suitable for nucleobase excision. In addition, these structures have provided insight into enzymatic selection and hydrolysis of alkylpurines, a diverse array of lesions created from exposure of DNA to alkylating agents. Despite their diversity, the alkylpurine DNA glycosylases utilize the same general strategy for DNA damage recognition as do other glycosylases, in which amino acids near the active site are able to sense an energetic difference between modified and unmodified base pairs. Nevertheless, the alkylpurine specific active sites seem to have evolved unique mechanisms for excision of either relatively unstable, positively charged bases (e.g., cytotoxic 3mA) in addition to more stable adducts like mutagenic εA.

Acknowledgments

The authors wish to thank Patrick O'Brien for critical reading of the manuscript. Work on DNA alkylation repair in the Eichman laboratory is supported by the American Cancer Society (RSG-07-063-01-GMC), the Vanderbilt-Ingram Cancer Center (IRG-58-009-47), the Sartain-Lanier Family Foundation, and the Vanderbilt Center in Molecular Toxicology (T32 ES07028).

References

1. Friedberg, E. C.; Walker, G. C.; Siede, W.; Wood, R. D.; Schultz, R. A.; Ellenberger, T. *DNA Repair and Mutagenesis*, 2nd ed.; ASM Press: Washington, D.C., 2006.
2. Sedgwick, B. *Nat. Rev.* **2004**, *5*, 148–157.
3. Fromme, J. C.; Verdine, G. L. *Adv. Protein Chem.* **2004**, *69*, 1–41.
4. Huffman, J. L.; Sundheim, O.; Tainer, J. A. *Mutat. Res.* **2005**, *577*, 55–76.
5. Karran, P.; Lindahl, T. *Biochemistry* **1980**, *19*, 6005–6011.
6. Brent, T. P. *Biochemistry* **1979**, *18*, 911–916.

7. Chen, J.; Derfler, B.; Samson, L. *EMBO J.* **1990**, *9*, 4569–4575.
8. Berdal, K. G.; Bjoras, M.; Bjelland, S.; Seeberg, E. *EMBO J.* **1990**, *9*, 4563–4568.
9. Memisoglu, A.; Samson, L. *Gene* **1996**, *177*, 229–235.
10. Riazuddin, S.; Lindahl, T. *Biochemistry* **1978**, *17*, 2110–2118.
11. Thomas, L.; Yang, C. H.; Goldthwait, D. A. *Biochemistry* **1982**, *21*, 1162–1169.
12. Begley, T. J.; Haas, B. J.; Noel, J.; Shekhtman, A.; Williams, W. A.; Cunningham, R. P. *Curr. Biol.* **1999**, *9*, 653–656.
13. O'Rourke, E. J.; Chevalier, C.; Boiteux, S.; Labigne, A.; Ielpi, L.; Radicella, J. P. *J. Biol. Chem.* **2000**, *275*, 20077–20083.
14. Alseth, I.; Rognes, T.; Lindback, T.; Solberg, I.; Robertsen, K.; Kristiansen, K. I.; Mainieri, D.; Lillehagen, L.; Kolsto, A. B.; Bjoras, M. *Mol. Microbiol.* **2006**, *59*, 1602–1609.
15. Holt, S.; Yen, T. Y.; Sangaiah, R.; Swenberg, J. A. *Carcinogenesis* **1998**, *19*, 1763–1769.
16. Shuker, D. E.; Bailey, E.; Parry, A.; Lamb, J.; Farmer, P. B. *Carcinogenesis* **1987**, *8*, 959–962.
17. Shuker, D. E.; Farmer, P. B. *Chem. Res. Toxicol.* **1992**, *5*, 450–460.
18. Bjelland, S.; Bjoras, M.; Seeberg, E. *Nucleic Acids Res.* **1993**, *21*, 2045–2049.
19. Saparbaev, M.; Kleibl, K.; Laval, J. *Nucleic Acids Res.* **1995**, *23*, 3750–3755.
20. McCarthy, T. V.; Karran, P.; Lindahl, T. *EMBO J.* **1984**, *3*, 545–550.
21. Bjelland, S.; Birkeland, N. K.; Benneche, T.; Volden, G.; Seeberg, E. *J. Biol. Chem.* **1994**, *269*, 30489–30495.
22. Lau, A. Y.; Schärer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, T. *Cell* **1998**, *95*, 249–258.
23. Labahn, J.; Schärer, O. D.; Long, A.; Ezaz-Nikpay, K.; Verdine, G. L.; Ellenberger, T. E. *Cell* **1996**, *86*, 321–329.
24. Drohat, A. C.; Kwon, K.; Krosky, D. J.; Stivers, J. T. *Nat. Struct. Biol.* **2002**, *9*, 659–664.
25. Eichman, B. F.; O'Rourke, E. J.; Radicella, J. P.; Ellenberger, T. *EMBO J.* **2003**, *22*, 4898–4909.
26. Kuo, C. F.; McRee, D. E.; Fisher, C. L.; O'Handley, S. F.; Cunningham, R. P.; Tainer, J. A. *Science* **1992**, *258*, 434–440.
27. Guan, Y.; Manuel, R. C.; Arvai, A. S.; Parikh, S. S.; Mol, C. D.; Miller, J. H.; Lloyd, S.; Tainer, J. A. *Nat. Struct. Biol.* **1998**, *5*, 1058–1064.
28. Bruner, S. D.; Norman, D. P.; Verdine, G. L. *Nature* **2000**, *403*, 859–866.
29. Mol, C. D.; Arvai, A. S.; Begley, T. J.; Cunningham, R. P.; Tainer, J. A. *J. Mol. Biol.* **2002**, *315*, 373–384.
30. Rubinson, E. H.; Metz, A. H.; O'Quin, J.; Eichman, B. F. *J. Mol. Biol.* **2008**, *381*, 13–23.
31. Dalhus, B.; Helle, I. H.; Backe, P. H.; Alseth, I.; Rognes, T.; Bjoras, M.; Laerdahl, J. K. *Nucleic Acids Res.* **2007**, *35*, 2451–2459.
32. Roberts, R. J.; Cheng, X. *Annu. Rev. Biochem.* **1998**, *67*, 181–198.
33. Stivers, J. T.; Jiang, Y. L. *Chem. Rev.* **2003**, *103*, 2729–2759.
34. Banerjee, A.; Santos, W. L.; Verdine, G. L. *Science* **2006**, *311*, 1153–1157.

35. O'Brien, P. J.; Ellenberger, T. *J. Biol. Chem.* **2004**, *279*, 9750–9757.
36. Scharer, O. D.; Nash, H. M.; Jiricny, J.; Laval, J.; Verdine, G. L. *J. Biol. Chem.* **1998**, *273*, 8592–8597.
37. Lau, A. Y.; Wyatt, M. D.; Glassner, B. J.; Samson, L. D.; Ellenberger, T. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13573–13578.
38. Connor, E. E.; Wyatt, M. D. *Chem. Biol.* **2002**, *9*, 1033–1041.
39. O'Brien, P. J.; Ellenberger, T. *Biochemistry* **2003**, *42*, 12418–12429.
40. Doherty, A. J.; Serpell, L. C.; Ponting, C. P. *Nucleic Acids Res.* **1996**, *24*, 2488–2497.
41. Protein Data Bank, Joint Center for Structural Genomics, PDB ID 2H56, 2006.
42. Ishida, T.; Shibata, M.; Fujii, K.; Inoue, M. *Biochemistry* **1983**, *22*, 3571–3581.
43. Yamagata, Y.; Kato, M.; Odawara, K.; Tokuno, Y.; Nakashima, Y.; Matsushima, N.; Yasumura, K.; Tomita, K.; Ihara, K.; Fujii, Y.; Nakabeppu, Y.; Sekiguchi, M.; Fujii, S. *Cell* **1996**, *86*, 311–319.
44. Hollis, T.; Ichikawa, Y.; Ellenberger, T. *EMBO J.* **2000**, *19*, 758–766.
45. Kwon, K.; Cao, C.; Stivers, J. T. *J. Biol. Chem.* **2003**, *278*, 19442–19446.
46. Cao, C.; Kwon, K.; Jiang, Y. L.; Drohat, A. C.; Stivers, J. T. *J. Biol. Chem.* **2003**, *278*, 48012–48020.
47. Metz, A. H.; Hollis, T.; Eichman, B. F. *EMBO J.* **2007**, *26*, 2411–2420.
48. Saparbaev, M.; Laval, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5873–5877.
49. Bjoras, M.; Klungland, A.; Johansen, R. F.; Seeberg, E. *Biochemistry* **1995**, *34*, 4577–4582.
50. Berdal, K. G.; Johansen, R. F.; Seeberg, E. *EMBO J.* **1998**, *17*, 363–367.
51. Alseth, I.; Osman, F.; Korvald, H.; Tsaneva, I.; Whitby, M. C.; Seeberg, E.; Bjoras, M. *Nucleic Acids Res.* **2005**, *33*, 1123–1131.
52. Memisoglu, A.; Samson, L. *J. Bacteriol.* **2000**, *182*, 2104–2112.
53. O'Brien, P. J.; Ellenberger, T. *J. Biol. Chem.* **2004**, *279*, 26876–26884.
54. Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714–2723.
55. Flint, D. H.; Allen, R. M. *Chem. Rev.* **1996**, *96*, 2315–2334.
56. Cunningham, R. P.; Asahara, H.; Bank, J. F.; Scholes, C. P.; Salerno, J. C.; Surerus, K.; Munck, E.; McCracken, J.; Peisach, J.; Emptage, M. H. *Biochemistry* **1989**, *28*, 4450–4455.
57. Michaels, M. L.; Pham, L.; Nghiem, Y.; Cruz, C.; Miller, J. H. *Nucleic Acids Res.* **1990**, *18*, 3841–3845.
58. Rudolf, J.; Makrantoni, V.; Ingledew, W. J.; Stark, M. J.; White, M. F. *Mol. Cell* **2006**, *23*, 801–808.
59. Weiner, B. E.; Huang, H.; Dattilo, B. M.; Nilges, M. J.; Fanning, E.; Chazin, W. J. *J. Biol. Chem.* **2007**, *282*, 33444–33451.
60. Klinge, S.; Hirst, J.; Maman, J. D.; Krude, T.; Pellegrini, L. *Nat. Struct. Mol. Biol.* **2007**, *14*, 875–877.
61. Fromme, J. C.; Verdine, G. L. *EMBO J.* **2003**, *22*, 3461–3471.
62. Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L. *Nature* **2004**, *427*, 652–656.
63. Fu, W.; O'Handley, S.; Cunningham, R. P.; Johnson, M. K. *J. Biol. Chem.* **1992**, *267*, 16135–16137.

64. Golinelli, M. P.; Chmiel, N. H.; David, S. S. *Biochemistry* **1999**, *38*, 6997–7007.
65. Falnes, P. O.; Johansen, R. F.; Seeberg, E. *Nature* **2002**, *419*, 178–182.
66. Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. *Nature* **2002**, *419*, 174–178.
67. Birkeland, N. K.; Anensen, H.; Knaevelsrud, I.; Kristoffersen, W.; Bjoras, M.; Robb, F. T.; Klungland, A.; Bjelland, S. *Biochemistry* **2002**, *41*, 12697–12705.
68. Mansfield, C.; Kerins, S. M.; McCarthy, T. V. *FEBS Lett.* **2003**, *540*, 171–175.
69. Leiros, I.; Nabong, M. P.; Grosvik, K.; Ringvoll, J.; Haugland, G. T.; Uldal, L.; Reite, K.; Olsbu, I. K.; Knaevelsrud, I.; Moe, E.; Andersen, O. A.; Birkeland, N. K.; Ruoff, P.; Klungland, A.; Bjelland, S. *EMBO J.* **2007**, *26*, 2206–2217.