doi:10.1016/j.jmb.2008.05.078 J. Mol. Biol. (2008) 381, 13–23

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

A New Protein Architecture for Processing Alkylation Damaged DNA: The Crystal Structure of DNA Glycosylase AlkD

Emily H. Rubinson, Audrey H. Metz, Jami O'Quin and Brandt F. Eichman*

Department of Biological Sciences and Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA

Received 17 April 2008; received in revised form 28 May 2008; accepted 29 May 2008
Available online
5 June 2008

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

Keywords: DNA repair; 3-methyladenine; alkylpurine; DNA glycosylase; HEAT repeat

Edited by R. Huber

Introduction

The integrity of DNA is constantly challenged by chemical attack from endogenous metabolites and environmental agents. Chemical modification of DNA nucleobases by alkylation, oxidation, deamination, or hydrolysis produces mutagenic or cytotoxic lesions that can result in heritable disease, cancer, and cell death.[1] To safeguard against DNA damage, all organisms have evolved DNA repair mechanisms to eliminate modified bases from the genome. Base excision repair is the principal pathway by which small modifications to DNA nucleobases are removed in both prokaryotes and eukaryotes. DNA glycosylases initiate the base excision repair pathway by catalyzing the hydrolysis of the C1′-N glycosidic bond to liberate the modified base from the phosphoribose backbone. The resulting abasic site is further processed by AP endonuclease, phosphodiesterase, DNA polymerase, and DNA ligase functions to restore the DNA to an undamaged state.

DNA glycosylases are specific for a particular type of base damage. The mechanism by which these enzymes initially locate their target substrates is believed to proceed by a processive sliding search for destabilized base pairs along DNA.[2–5] DNA glycosylases read out the identity of the lesion by flipping the damaged nucleotide out of the DNA helix and into a nucleobase-binding pocket on the protein, the chemical and physical properties of which are complementary to the correct substrate. All glycosylases examined to date utilize a similar strategy for binding DNA and base flipping despite their structural diversity. There are five structural glycosylase superfamilies[6], represented by: (i) bacteriophage T4 pyrimidine dimer-specific DNA glycosylase, EndoV;[7] (ii) uracil DNA glycosylase, UDG;[8] (iii) bacterial 8-oxoguanine DNA glycosylase, MutM/FPG;[9,10] (iv) human alkyladenine DNA glycosylase, AAG/MNPG;[11] and (v) the helix-hairpin-helix (HhH) superfamily.[12,13] A broad range of substrate specificities exist within the HhH superfamily, which

*Corresponding author. E-mail address: brandt.eichman@vanderbilt.edu.

Abbreviations used: SAD, single-wavelength anomalous dispersion; HhH, helix-hairpin-helix.

includes *Escherichia coli* endonuclease III (EndoIII)[14], *E. coli* adenine DNA glycosylase (MutY)[15], human 8-oxoguanine DNA glycosylase (hOGG1)[9], and several alkylpurine-specific DNA glycosylases.[16–18]

DNA alkylation damage from endogenous methyl donors, environmental toxins, or from chemotherapeutic agents produces a chemically diverse spectrum of alkylated nucleobases.[19–21] DNA glycosylases remove N^3- and N^7-substituted alkylpurines, as well as 1,N^6-ethenoadenine (εA). The variation in substrate specificities of alkylpurine DNA glycosylases from different organisms has been useful in understanding the mechanisms of base selection and removal. *E. coli* 3-methyladenine (3mA) DNA glycosylase I (TAG)[22], and *Helcobacter pylori* methyladenine DNA glycosylase (MagIII)[23,24], are highly specific for 3mA. *Thermatoga maritima* methylpurine DNA glycosylase II (MpgII)[23] excises 7-methylguanine (7mG) in addition to 3mA. At the other extreme, *E. coli* 3mA DNA glycosylase II (AlkA), *Saccharomyces cerevisiae* MAG and human AAG have a broad, yet well-defined specificity toward 3mA, 7mG, εA, hypoxanthine, and various oxidized purines.[25–31]

Several alkylpurine DNA glycosylases have been identified recently in the genomic sequences of Gram-positive bacteria and lower eukaryotes *Entamoeba histolytica* and *Dictyostelium discoideum*.[32,33] *Bacillus cereus* AlkC and AlkD excise 3mA and 7mG, but not εA, and thus have an intermediate specificity similar to MpgII.[23,32] These enzymes share no sequence homology with any other glycosylase, and they have been predicted to represent a new structural class of glycosylase enzymes.[33] We present here the crystal structure of *B. cereus* AlkD at 2.1 Å resolution together with mutational analyses of 7mG excision and DNA binding. AlkD is composed of a tandem array of helical repeats reminiscent of HEAT motifs, which are known to facilitate protein–protein interactions and have not yet been associated with DNA binding or catalytic activity. A comprehensive structural analysis of the variant HEAT repeats in AlkD helps to explain how this protein interaction motif has been adapted to bind DNA. Importantly, the complementarity between AlkD's concave surface and the *B*-DNA backbone in our docking model raises the possibility that the enzyme uses a different strategy to engage DNA and access the damaged base.

Results

The structure of AlkD

The crystal structure of AlkD from *B. cereus* was determined using experimental phases from a single-wavelength anomalous dispersion (SAD) experiment from a crystal derivatized with platinum tetrachloroplatinate(II) (Table 1). A crystallographic model that consists of one AlkD molecule (residues 1–225) in the asymmetric unit was built into 2.0 Å Pt-SAD electron density (Fig. 1a). To avoid the possibility of structural artefacts resulting from platinum binding to the protein, the original SAD model was refined against native diffraction data extending to 2.1 Å resolution, resulting in a crystallographic residual (R_{cryst}) of 18.9% and an R_{free} value of 22.9% (Table 1).

Table 1. Data collection, phasing and refinement statistics

	Native	K_2PtCl_4
A. *Data collection*		
Space group	$P4_3$	$P4_3$
Cell dimensions		
a (Å)	77.9	78.8
b (Å)	77.9	78.8
c (Å)	55.1	55.3
α = β = γ (°)	90.0	90.0
Wavelength (Å)	1.0000	1.0718
Resolution (Å)	50–2.08 (2.15–2.08)	50–1.85 (1.92–1.85)
R_{sym} or R_{merge}	0.046 (0.531)	0.052 (0.321)
$I/\sigma I$	26.9 (3.4)	50.1 (3.1)
Completeness (%)	99.1 (95.3)	97.4 (84.3)
Redundancy	7.1 (6.2)	7.2 (4.8)
B. *Refinement*		
Resolution (Å)	50–2.10 (2.21–2.10)	
No. reflections	18356 (2605)	
R_{work}	0.188 (0.283)	
R_{free}	0.227 (0.313)	
No. atoms		
Protein	1893	
Ion/ligand	0	
Water	81	
B-factor		
Protein (Å^2)	52.7	
Water (Å^2)	55.2	
r.m.s.d. from ideal		
Bond lengths (Å)	0.018	
Bond angles (°)	1.571	

Data in parentheses refer to the highest resolution shells.

AlkD adopts a single α-helical domain. Of the 14 helices (αA-αN), 12 pair in an antiparallel fashion to form six tandemly repeated α-α motifs (αA/αC, αD/αE, αF/αG, αH/αI, αJ/αK, and αL/αM) (Figs. 1b and 3). The helical repeats are stacked into a superhelical solenoid in which helices B, C, E, G, I, K and M form a concave surface with an aromatic cleft at its center (Fig. 1b). Residues within this cleft were shown previously to impair the base excision activity of AlkD.[33] The concave surface of the protein is strikingly electropositive (Fig. 1c), a feature that is distinct from other helical repeat proteins and one that likely facilitates binding to DNA (discussed below). A structural homology search of the PDB using the DALI server[34] indicated that AlkD is most similar in tertiary structure to *Enterococcus faecalis* EF3068 (PDB ID *2B6C*) and *B. cereus* BC3264 (PDB ID *1T06*), two hypothetical proteins of unknown function determined from the Midwest Center for Structural Genomics (Supplementary Data Fig. S1, Table S1). EF3068 shares 35% sequence identity with AlkD (Supplementary Data Fig. S2) and was used in an earlier study as the basis for a homology model of the *Bacillus* enzyme.[33] The structures of AlkD and EF3068 are similar in overall fold (r.m.s.d. of 1.38 Å for all backbone atoms), with the most notable differences in helices αA/αC at the N-terminus (Supplementary Data Fig. S1).

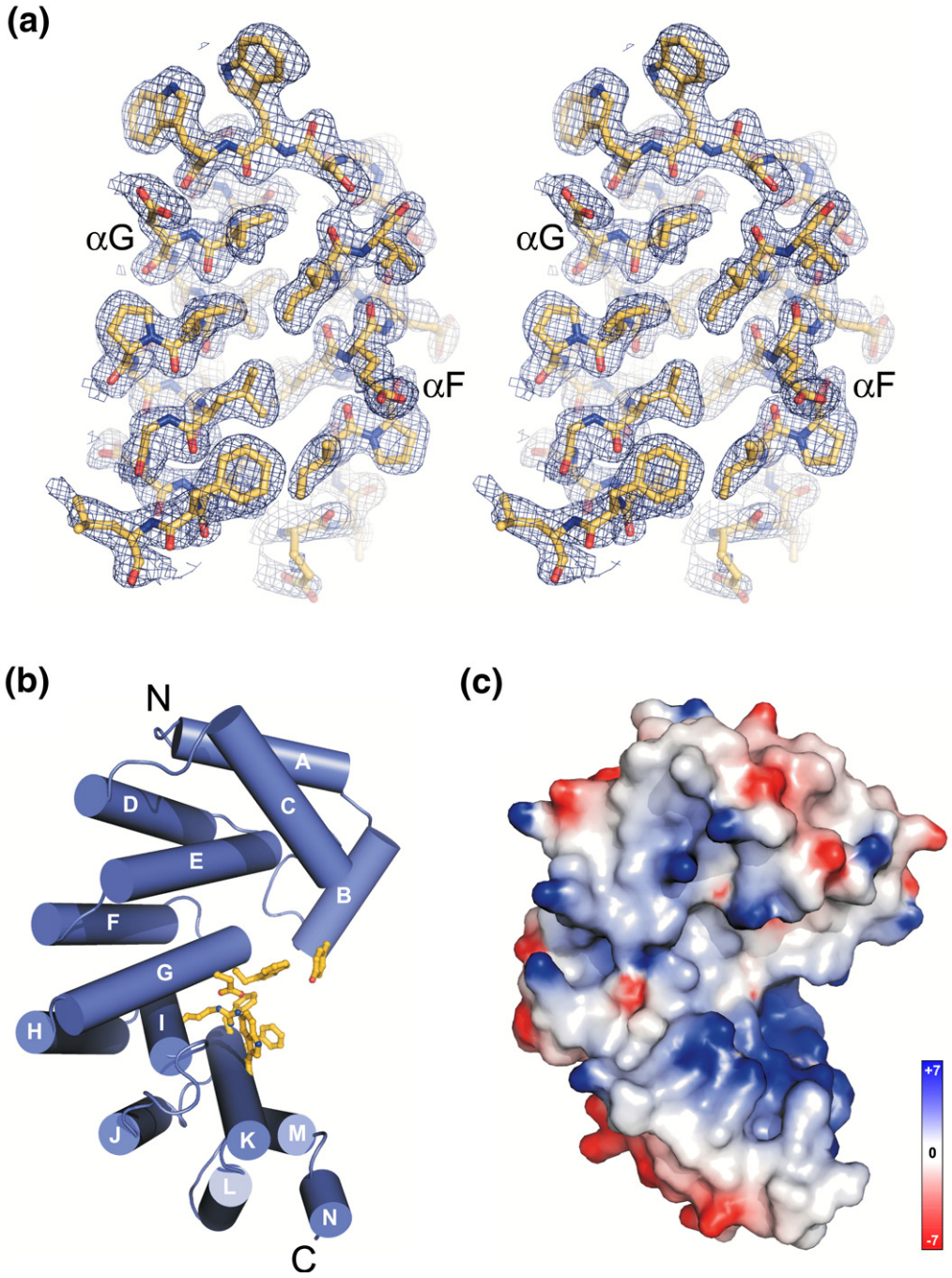

Fig. 1. Structure of *B. cereus* AlkD. (a) Stereo-view of helices αG and αF from the refined crystallographic model superimposed on a 2.0-Å Pt-SAD experimental electron density map (contoured at 1σ). (b) AlkD forms a solenoid structure with a left-handed superhelical twist. Residues forming an aromatic cleft on the concave surface are shown as yellow sticks. (c) Solvent-accessible surface representation of AlkD in the same orientation as (b) colored according to electrostatic potential (red, negative; blue, positive; −7 to +7 k_BT). Potentials in Figs. 1 and 5 were calculated with the program DelPhi.[64] All molecular images were rendered using PyMOL (http://www.pymol.org).

Putative active site

One striking feature of the AlkD structure is the apparent similarity in the chemical environment of the aromatic cleft to the nucleobase-binding pockets of other alkylpurine DNA glycosylases. AAG, AlkA, TAG, and MagIII all feature electron-rich, aromatic pockets that accommodate the extrahelical alkylpurine base through both shape complementarity and π-π stacking interactions.[18,35–38] With the exception of TAG, which excises exclusively 3mA and 3mG, the alkylpurine binding pockets also feature a conserved acidic residue that is essential for base excision activity.

AlkD contains a cluster of aromatic residues (Trp109, Trp187, Phe179, Phe180, and Tyr27) that

form a shallow cleft at the center of the concave surface (Fig. 2a). At the back of this cleft, Asp113 and Arg148 form an electrostatic interaction and were observed in the crystal structure to hydrogen bond to several well-ordered water molecules (Fig. 2a). The relative positions of these side chains are similar to the aromatic and catalytic Asp238 residues inside the shallow nucleobase-binding pocket of AlkA (Fig. 2b).[16,37] Asp113, Arg148 and Trp109 are invariant in sequence among AlkD homologs (Supplementary Data Fig. S2) and were identified in an earlier study to be important for release of methylated bases from *N*-methyl-*N*-nitrosourea-treated genomic DNA and for growth of *E. coli tag alkA* mutant cells in the presence of methylmethane sulfonate.[33] Because wild-type AlkD excises several alkylated nucleobases produced from *N*-methyl-*N*-nitrosourea or treatment with methylmethane sulfonate,[32] we tested the effect of AlkD mutants on the specific excision of 7mG from a defined oligonucleotide substrate. As shown in Fig. 2c and d, Asp113Asn and Arg148Ala each resulted in a dramatic (100-fold) decrease in the single-turnover

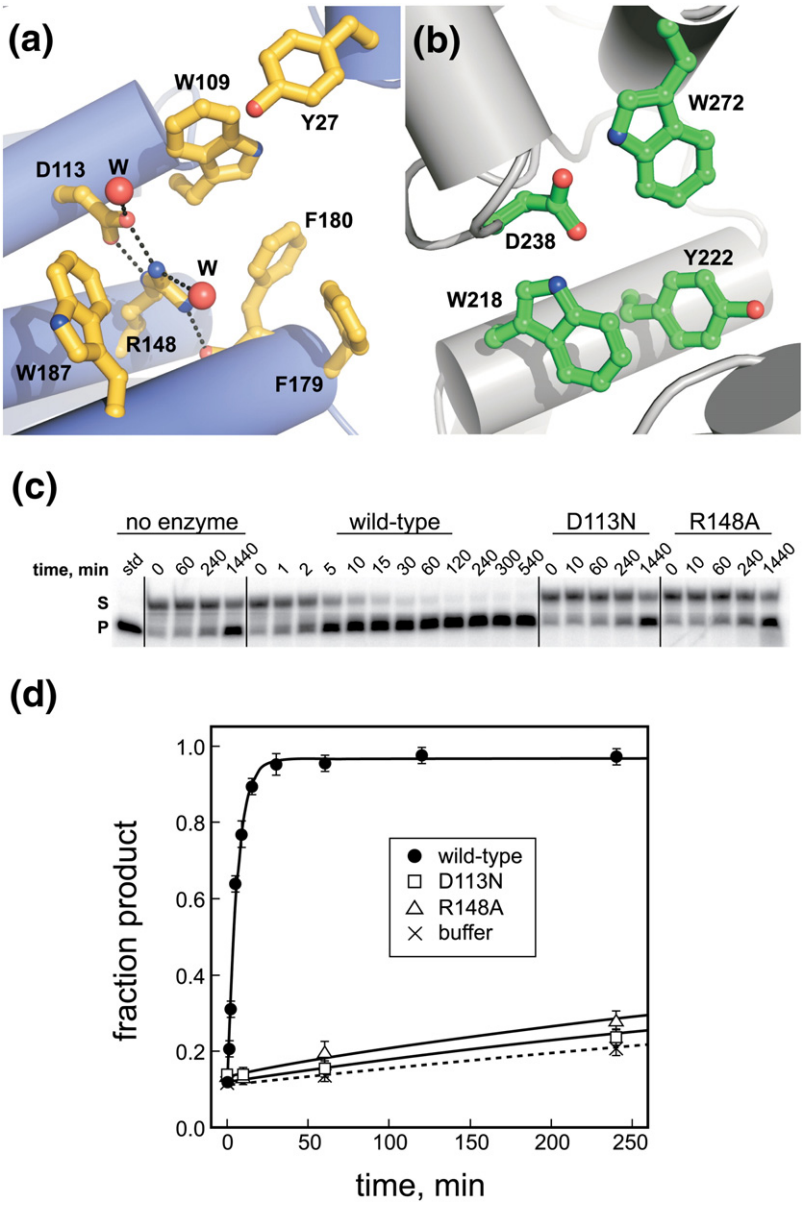

Fig. 2. The putative active site of AlkD. (a) Close-up of the concave cleft shows aromatic and charged residues implicated in nucleobase and DNA binding. Water molecules are shown as red spheres and hydrogen bonds are depicted as black dotted lines. (b) The electron-rich active site of AlkA, showing the conserved Asp238 essential for base excision. (c) 7-Methylguanine (7mG) excision by AlkD. Denaturing polyacrylamide gel showing the disappearance of 7mG-DNA substrate (S) and appearance of alkaline-cleaved abasic-DNA product (P) as a result of reaction with no enzyme, wild-type AlkD, Asp113Asn, and Arg148Ala mutants for increasing amounts of time. (d) Quantification of the data in (c) showing the inactivity of the Asp113Asn (squares) and Arg148Ala (triangles) mutants as compared to wild-type AlkD (filled circles). Non-enzymatic 7mG hydrolysis is shown as crosses and a dotted line curve fit. Error bars represent the standard deviation from the average of three independent experiments.

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

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

[a] First-order single-turnover rate constants for 7mG excision from a 25mer oligonucleotide duplex containing a 7mG·C base pair. Values represent the averages and standard deviations from three experiments.
[b] Rate enhancements are based on the fold increase above a non-enzymatic control ($k_{non}=0.7\times10^{-3}$ min^{-1}).

rate of 7mG excision relative to the wild-type enzyme (Table 2), indicating that Asp113 and Arg148 have a specific role in 7mG excision by AlkD. These results show that AlkD provides catalytic assistance to liberate 7mG from DNA, and implicate the aromatic cleft as the active site of the enzyme.

The variant HEAT motif

The list of structurally homologous proteins from the DALI analysis was exclusive to proteins that contain tandem helical repeats (Supplementary Data Table S1). Apart from EF3068 and BC3264, AlkD is most similar in structure to the PR65/A subunit of protein phosphatase 2A (PP2A)[39,40], and to the SCF ubiquitin ligase regulatory subunit Cand1[41], both of which contain Huntington/Elongation/A subunit/Target of rapamycin (HEAT) repeats.[42] HEAT motifs are ~45-amino acid sequences arranged as two antiparallel α-helices, which are packed together by a conserved hydrophobic interface and are tandemly repeated to form superhelical α-structures.[42] The α-α pairs stack in a parallel arrangement to form a solenoid in which the C-terminal helix of each repeat lines the inside groove of the superhelix. The solenoid structure is typically formed from no less than 14 HEAT motifs and provides a molecular scaffold that facilitates protein–protein interactions. Out of 393 structurally similar proteins identified from the PDB search, only three were eukaryotic nucleic acid-binding proteins, including transcription factor Rcd-1 (*Z*-score=5.9)[43] and the RNA-binding proteins Pumilio1 (*Z*-score=4.8)[44] and Ro autoantigen (*Z*-score=3.5).[45] Ro contains HEAT repeats that are quite divergent from the canonical motif, while Rcd-1 and Pumilio-1 are comprised entirely of armadillo and pumilio repeats, respectively.[42,46] Thus, AlkD represents the first example of a DNA-binding HEAT protein.

In contrast to the regular, repeating α-α structures found in archetypical HEAT proteins, the AlkD repeats exhibit greater structural variation despite the conservation of hydrophobic residues that stabilize the interface between helices (Fig. 3). The most obvious outlier is αA/αC, which has the opposite handedness from the other repeats, as well as an inserted helix (αB) that contributes considerable surface area to the concave face of the protein (Fig. 3a). The r.m.s.d. among the five remaining repeats αD/αE-αL/αM in AlkD is 2.48 Å, as compared with 1.76 Å among five consecutive HEAT repeats in PP2A.[47] The AlkD helices are, on average, one turn shorter and, with the exception of αF/αG, do not contain the intrahelical kink characteristic of HEAT repeats (Fig. 3b). Thus, the helical repeats that comprise AlkD represent a structural variant of the canonical HEAT motif and thus give rise to a unique architecture.

The most striking feature of the AlkD variant HEAT repeats is the abundance of basic residues along the C-terminal helix (Fig. 3c). The positions

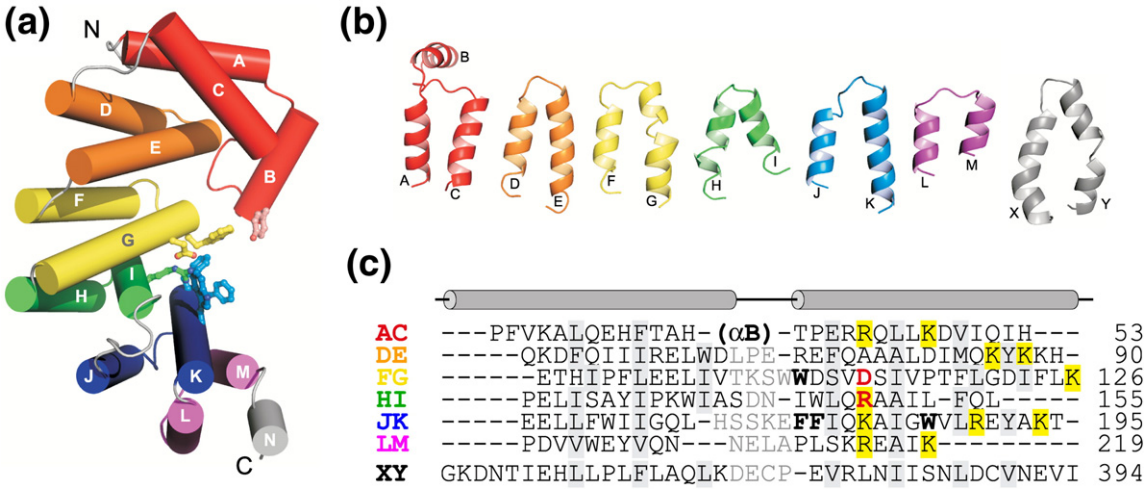

Fig. 3. AlkD is composed of six variant HEAT motifs. (a) Cylinder representation of AlkD with HEAT-like tandem repeats colored independently. Residues lining the putative active site are shown as ball-and-stick. (b) Structures of individual AlkD α-α pairs (AC, DE, FG, HI, JK, LM) are compared to canonical HEAT repeat 10 (XY) from protein phosphatase 2A PR65/A subunit, PDB ID 1B3U.[47] (c) Structure-based sequence alignment of HEAT motifs in (b) with AlkD residues lining the aromatic cleft in boldface and Asp113/Arg148 colored red. Positively charged residues contributing to the electropositive concave surface are highlighted in yellow. Interdigitating residues important for the relative orientation of paired α-helices are shaded gray.

and spacing of these basic residues are conserved among the repeats and, consequently, the concave surface of the protein is highly electropositive (Fig. 1c). This positive charge along the concave surface is conserved among AlkD orthologs (Supplementary Data Figs. S1 and S2), but is not observed in other helical repeat proteins (Supplementary Data Fig. S3), which faithfully utilize this surface to bind polypeptides.[40,41,48,49] Thus, it seems that the AlkD variant HEAT repeats have evolved this additional electrostatic feature as a means to stabilize the negatively charged DNA backbone at the molecular binding interface.

A DNA-binding model

Despite their structural divergence, the five existing superfamilies of DNA glycosylases use a common mode of binding DNA and gaining access to the lesion via a base-flipping mechanism (Fig. 4).[6,50] Structures of these enzymes in complex with DNA show that they all use a positively charged surface that helps anchor the DNA against the enzyme. The DNA duplex is highly kinked, allowing extrusion of the damaged base into a sterically constrained active site pocket that lies adjacent to the positive surface (Fig. 4). The distorted conformation of the DNA is stabilized by a pair of side chains that intercalate into the DNA helix at the site of the lesion.

The solenoid structure of AlkD is distinctly different from the other DNA glycosylases (Fig. 4). AlkD is the only glycosylase to form a linear, cylindrical groove along the entire width of the protein. Consequently, the environment around the putative active site is much less constrained than other DNA glycosylase architectures. To investigate how such a unique glycosylase architecture might engage DNA, we constructed a model of DNA docked onto the structure of AlkD (Fig. 5). The lack of structural similarity between AlkD and the other glycosylases precluded us from superimposing previous glycosylase–DNA complexes onto AlkD. Instead, we took advantage of the fact that the dimensions across the concave opening are ~24 Å, which can easily accommodate a *B*-DNA duplex. Canonical *B*-DNA (PDB ID *1BNA*) was docked manually onto the concave surface with the minor groove of the DNA positioned toward the putative active site, consistent with structures of other glycosylase–DNA complexes. Importantly, the linear DNA model formed favorable van der Waals and electrostatic interactions along the entire groove of the protein with no steric collision (Fig. 5).

The DNA docking model was used as a guide for mutational and chemical modification experiments designed to probe the nature of DNA binding to AlkD. The DNA backbone in our docking model was in close proximity to the catalytically important

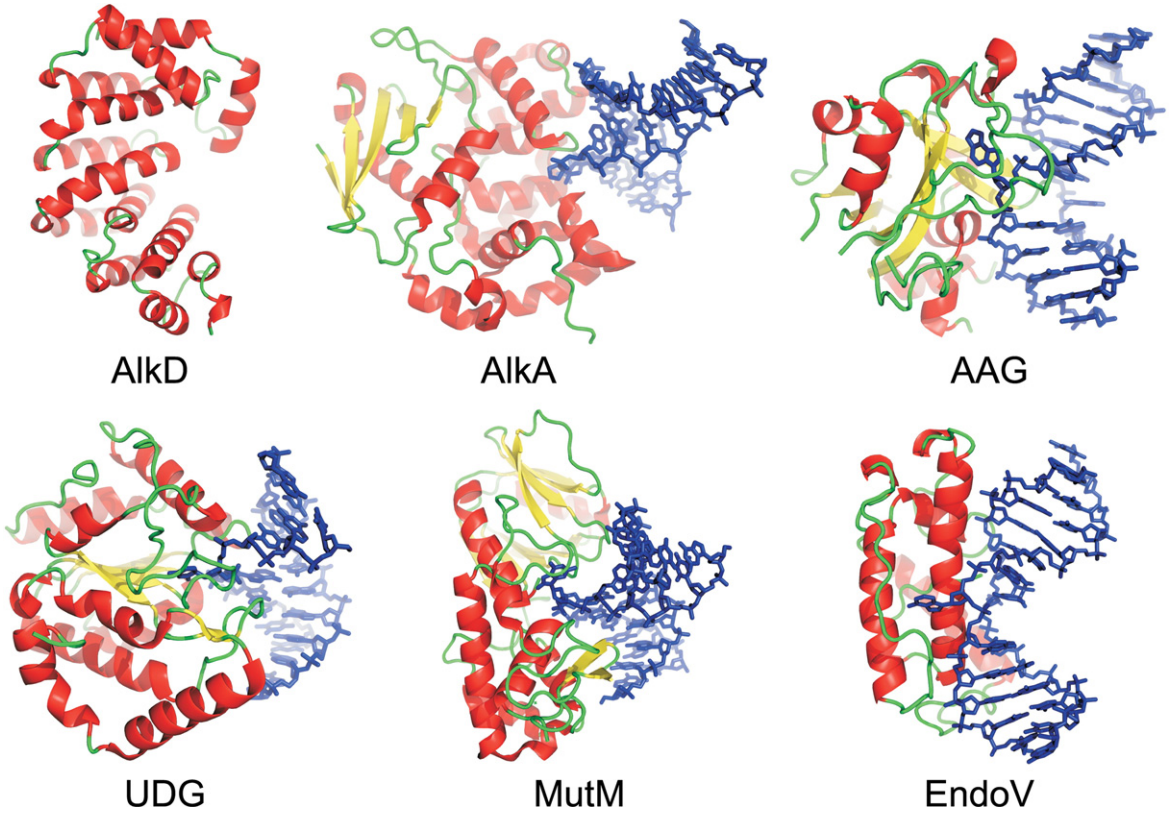

Fig. 4. The six structural superfamilies of DNA glycosylases. Representative crystal structures shown are: *B. cereus* AlkD; *E. coli* 3-methyladenine DNA glycosylase II, AlkA (1DIZ);[37] human methylpurine DNA glycosylase, AAG (1EWN);[36] human uracil-DNA glycosylase, UDG (1EMH);[65] *Bacillus stearothermophilus* 8-oxoguanine DNA glycosylase, MutM (1L1T);[66] and T4 pyrimidine dimer DNA glycosylase, EndoV (1VAS).[67] Proteins are colored according to secondary structure, and lesion-containing DNA is shown as blue sticks.

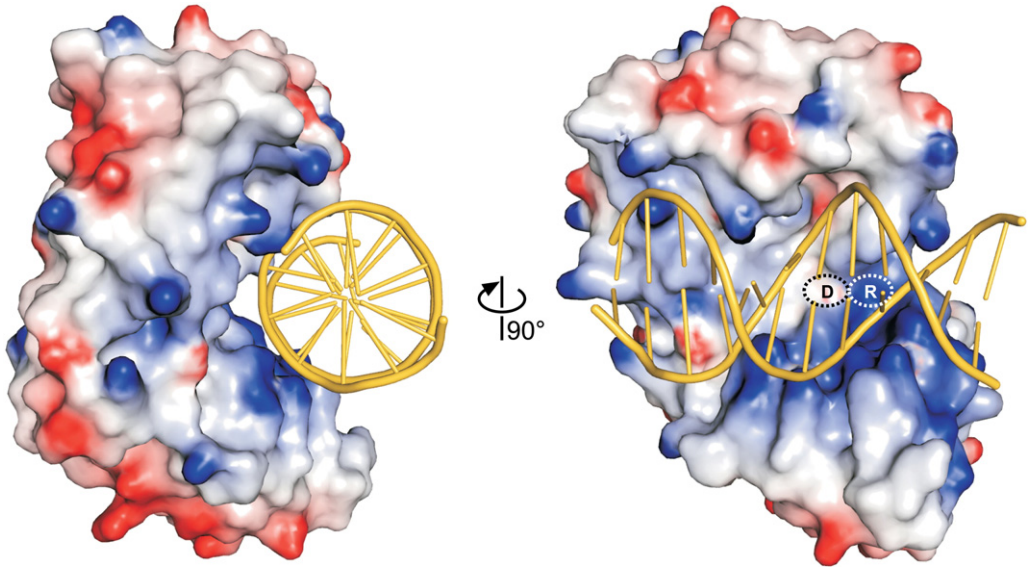

Fig. 5. Theoretical model of DNA bound to AlkD. *B*-DNA from PDB entry 1BNA docked manually onto the AlkD crystal structure shows that the DNA need not adopt a distorted conformation to access the concave surface and putative active site of the protein. The molecular surface of the protein is colored according to electrostatic potential (red, negative; blue, positive; −7 to +7 k_BT). Asp113 and Arg148 are labeled D and R, respectively.

Asp113 and Arg148 residues (Fig. 5). The effects of these residues on DNA binding were quantified using an *in vitro* fluorescence anisotropy assay (Table 3; Supplementary Data Fig. S4). Substitution of Asp113 with asparagine increased the affinity of AlkD for DNA containing either 7mG substrate or a tetrahydrofuran abasic product analog (Table 3). Likewise, an Arg148Ala mutation resulted in a modest but significant decrease in DNA binding. These results are consistent with an electrostatic interaction between Asp113–Arg148 and the DNA backbone. To further investigate this interaction, we measured binding of wild-type AlkD to positively charged 1-azaribose (C1′→N) and pyrrolidine (O4→N) abasic sites, which are potent inhibitors of many DNA glycosylases and were designed to mimic the positive charge of the proposed transition state for hydrolysis of the *N*-glycosylic bond.[37,51–53] The affinity of AlkA for pyrrolidine-DNA is three orders of magnitude greater than for tetrahydrofuran–DNA, and AlkA binds to both pyrrolidine and 1-azaribose ~15-fold tighter than to 7mG-DNA.[54] It was therefore surprising that wild-type AlkD bound with the same affinity to all DNAs tested (Table 3), suggesting that AlkD engages DNA in the active site differently than AlkA. Taken together, the docking model and mutational data suggest that DNA need not obtain a distorted, extrahelical conformation in order to bind to the protein active site surface.

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

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

Dissociation constants (K_d) for a 25mer oligonucleotide duplex containing the specified modification paired with cytosine were measured by fluorescence anisotropy as described in Experimental Procedures. THF, tetrahydrofuran; Pyr, pyrrolidine; Aza, 1-azaribose; n.d., not determined.

Discussion

The crystal structure of *B. cereus* alkylpurine DNA glycosylase AlkD reveals variant HEAT-like motifs that form an electropositive concave surface necessary to engage DNA. At the heart of the concave surface lies a shallow cleft of aromatic and charged residues, which we verified to be important for binding the DNA backbone and for 7mG excision. These catalytically essential residues and the unique DNA-binding HEAT repeat architecture were previously predicted from an AlkD homology model[33], which was constructed from an unpublished structure (EF3068) determined by the Midwest Structural Genomics Center as part of the Protein Structure Initiative.[55] In the present study, we extend these results by a structural and mutagenic analysis of nucleic acid binding by the variant HEAT motifs in AlkD.

The AlkD solenoid is a unique DNA glycosylase fold, and therefore represents a sixth structural superfamily for these DNA repair enzymes (Fig. 4). Perhaps the most unique feature of the AlkD glycosylase is its concave surface, which forms a groove perfectly complementary in shape and charge to a linear *B*-DNA duplex. Our DNA docking model shows an excellent fit between AlkD and an undistorted *B*-DNA molecule. As a comparison, we superimposed the kinked DNA from the AlkA/DNA

complex[37] onto our structure, which has been done for the theoretical homology model.[33] The protein surface occupied by the highly bent DNA was not as extensive as the undistorted *B*-DNA contact surface, suggesting that AlkD might bind DNA differently than other glycosylases. In order to stabilize extrahelical nucleobases and a large helical distortion in the DNA duplex, DNA glycosylases typically intercalate a pair of side chains into the DNA base stack. Interestingly, neither linear nor kinked DNA–AlkD models revealed any obvious candidate side chains that might penetrate the stacked bases in order to stabilize a flipped nucleotide. We cannot exclude the possibility that AlkD might adjust its conformation in order to engage a flipped-out nucleotide, although such a DNA-induced protein conformational change has not been observed in other glycosylases. Nevertheless, the open architecture and the apparent lack of DNA intercalating residues suggests strongly that AlkD interrogates damaged DNA differently than other glycosylases. Our model illustrates how AlkD might utilize extensive contacts with the minor groove of both DNA strands in order to sense a subtle distortion in 3mA·T and 7mG·C base pairs. This binding regime, although speculative, illustrates that this novel glycosylase architecture is well suited to scan DNA in a search for a lesion without distorting the conformation of the double helix.

The distinct substrate specificity of AlkD for only positively charged alkylpurines might be explained by the Asp113–Arg148 salt-bridge within the putative active site. Because it participates in an electrostatic interaction, Asp113 is not analogous to the conserved aspartate or glutamate residues found in other glycosylases (e.g., Asp238 in AlkA). The crystal structure of AlkA in complex with DNA containing 1-azaribose revealed a direct interaction between the extrahelical positively charged abasic site and Asp238.[37] The observation that neither 1-azaribose nor pyrrolidine enhances DNA binding by AlkD (Table 3) suggests that AlkA and AlkD use different mechanisms to catalyze nucleobase excision. Acidic residues within glycosylase active sites have been suggested to catalyze base hydrolysis by activating a water molecule or protein nucleophile for attack of the C1′ anomeric carbon[13,56], or by directly stabilizing a carbocation intermediate formed during cleavage of the *N*-glycosylic bond.[37] However, the formal positive charges of 3mA and 7mG nucleotides make these bases favorable to hydrolysis even in the absence of direct activation by a catalytic residue. Mutation of Arg148 reduced activity by two orders of magnitude, suggesting that Asp113 alone is not sufficient for catalysis, or that the salt-bridge stabilizes the local conformation of the active site. Indeed, our DNA docking model predicts that the precise orientations of the Asp113 and Arg148 side chains with respect to the lesion are critical. In this model, the phosphoribose backbones of both DNA strands are abutted against the Asp113–Arg148 pair, leaving no room for a flipped nucleotide (Fig. 5). This supports the scanning mechanism described above, and raises the possibility that this aromatic cleft is not an extrahelical nucleobase-binding pocket. In support of this, binding of free alkylpurine bases to AlkD could not be detected in solution by intrinsic tryptophan fluorescence quenching or 7mG excision inhibition assays, nor did we observe electron density corresponding to 3mA or 7mG bases that were soaked into AlkD crystals at tenfold molar excess of free base (data not shown). Based on the inherent instability of the 3mA and 7mG nucleotides and the proximity of Asp113–Arg148 to the DNA backbone in our model, AlkD should be able to gain access to the glycosylic bond for catalysis without flipping the base into the aromatic cleft.

AlkD is the first structural example of a HEAT protein that contains catalytic activity. Based on the adaptability of the HEAT motif to bind protein and DNA, other examples are likely. Of note is deoxyhypusine hydroxylase, an iron(II)-containing enzyme that catalyzes the final step in post-translational hypusinylation of eukaryotic initiation factor 5A (eIF5A). A theoretical model of deoxyhypusine hydroxylase predicts the protein to form a pair of HEAT domains similar in structure to AlkD.[57] Mutation of glutamate residues lining the concave surfaces of the protein impairs binding to the positively-charged DOH-eIF5A substrate.[58] The apparent similarity between catalytic residues within AlkD and deoxyhypusine hydroxylase suggests that HEAT proteins might have evolved a common feature to couple substrate binding and catalysis.

Materials and Methods

AlkD purification and crystallization

The AlkD gene was PCR amplified from *Bacillus cereus* genomic DNA (ATCC 14579) and cloned into a pET27 (Novagen) derived expression vector (pBG103, Vanderbilt Center for Structural Biology) that produces a cleavable N-terminal His_6-SUMO-fusion protein. *E. coli* HMS174 cells transformed with the AlkD/pBG103 plasmid were propagated in LB medium and protein was over-expressed for 3 h at 37°C upon addition of 0.5 mM IPTG. Cells were harvested in 50 mM Tris–HCl (pH 7.5), 500 mM NaCl, and 10% (v/v) glycerol and lysed with an Avestin Emulsifier C3 homogenizer operating at ~20000 psi (1 psi ≈6.9 kPa). AlkD-fusion protein was purified using Ni-NTA (Qiagen) affinity chromatography, followed by cleavage of the His_6-SUMO tag. AlkD was further purified by heparin affinity and gel-filtration chromatography to >99% homogeneity. Protein was concentrated to 12.5 mg/ml and stored in 20 mM bis-Tris–Propane, 100 mM NaCl, 0.1 mM EDTA. Mutant proteins were prepared by site-directed mutagenesis using a Quik-Change Kit (Stratagene) and purified in the same manner as wild-type AlkD. Structural integrity of mutant proteins was verified by circular dichroism spectroscopy.

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

X-ray data collection, phasing, and structure refinement

X-ray diffraction data (Table 1) were collected at beamlines 21-ID (native) and 22-ID (derivative) at the Advanced Photon Source (Argonne, IL) and processed using the HKL2000 package.[59] AlkD crystals belong to space group $P4_3$ and contain one molecule in the asymmetric unit.

Experimental X-ray phases were obtained from a SAD experiment using a crystal soaked with K_2PtCl_4. Diffraction data extending to 1.85 Å (Table 1) were collected at the energy corresponding to the platinum absorption peak (11.567 keV). Significant anomalous differences between $|F^+|$ and $|F^-|$ amplitudes extended to 2 Å. The program SHARP[60] was used to locate and refine the positions of two platinum sites and to carry out phase calculation and density modification. A continuous protein chain corresponding to residues 1–225 was built into the resulting 2.0 Å electron density map. Two non-native N-terminal residues (Val –1, Pro 0) resulting from cleavage of the His_6-SUMO tag were readily identified in the experimental electron density and were included in the final model, while the C-terminal 12 residues (226–237) were not observed.

The model was refined against both native and derivative amplitudes using experimental Pt-SAD phases and a maximum likelihood target as implemented in REFMAC 5.4.[61] Improvements to the model were guided by manual inspection of σ-weighted $2mF_o-Df_c$ and mF_o-Df_c electron density maps, and were judged successful by a decrease in R_{free} during refinement. Translation/libration/screw-rotation (TLS) refinement in REFMAC was used to model anisotropic motion of the protein. Refinement of the Pt-derivative model resulted in R_{cryst} and R_{free} values of 20.3% and 21.4%, respectively, which are higher than expected at 1.85 Å resolution. Efforts to resolve regions in the model corresponding to a partially chlorinated Pt(II) ion and a PEG 4000 molecule inside the putative active site and a second Pt (II) ion bound to Met1 using both experimental and refined electron density were unsuccessful. In order to avoid any potential structural perturbation of the putative active site from Pt or PEG 4000, the native model was used in all structural analysis despite the lower resolution of the native data (Table 1). The r.m.s.d. between the native and derivative models is 0.72 Å for all atoms. The native AlkD model was validated using PROCHECK.[62]

Glycosylase activity assay

AlkD glycosylase activity was measured by alkaline cleavage of the abasic DNA product of 7mG excision from a 25mer oligonucleotide duplex containing a 7mG·C base pair. 7mG was incorporated into DNA duplexes enzymatically as described.[63] Briefly, the primer oligonucleotide 5′-GACCACTACACC was ^{32}P-labeled at the 5′-end, annealed to a threefold excess of the complementary strand (5′- GTTGTTAGGAAACGGTGTAGTGGTC), and extended using DNA polymerase I Klenow fragment (New England Biolabs) in the presence of deoxy-7-methylguanosine 5′-triphosphate (d7mGTP, Sigma), dCTP, dTTP, and dATP. In a 10 μl glycosylase reaction, 100 nM radiolabeled DNA duplex was incubated with 2 μM AlkD in 50 mM HEPES (pH 7.5), 100 mM KCl, 10 mM DTT, 2 mM EDTA. The reaction was stopped at various times by addition of NaOH to a final concentration of 0.2 M, and heated at 70 °C for 10 min. The 12mer product and remaining 25mer substrate DNA strands were separated by denaturing PAGE (15% polyacrylamide gel) in 7M urea and quantified by autoradiography.

DNA binding assay

DNA binding of wild-type and mutant AlkD was measured by the change in fluorescence anisotropy as protein was added to an oligonucleotide duplex that contained a site-specific modification (X) in the middle of one strand [d(GACCACTACACCXTTTCCTAACAAC)] and a 6-carboxyfluorescein on the 3′-end of the complementary strand [d(GTTGTTAGGAAACGGTGTAGTGGTC)-FAM]. Oligonucleotides containing abasic sites were chemically synthesized, and those containing 7mG were prepared as above but without the ^{32}P label. Increasing concentrations of protein (0–30 μM) were added to a 50 nM DNA in 20 mM Bis-tris propane (pH 6.5), 100 mM NaCl, 2 mM DTT, 0.1 mM EDTA. Polarized fluorescence intensities using excitation and emission wavelengths of 485 and 538 were measured at ambient temperature using a SpectraMax M5 microplate reader (Molecular Devices). Measurements were recorded within 2 min of adding wild-type AlkD to DNA to ensure that no more than 50% of the 7mG-DNA substrate was converted to an abasic DNA product. Dissociation constants were derived by fitting a two-state binding model to data from three independent experiments.

Protein Data Bank accession code

The native AlkD model has been deposited along with structure factors in the Protein Data Bank under accession code 3BVS.

Acknowledgements

The authors thank Zdzislaw Wawrzak for help with data collection, and the staff at the Life Sciences (LS-CAT) and Southeast Regional Collaborative Access Teams (SER-CAT) at the Advanced Photon Source (Argonne, IL). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). This work was funded by the American Cancer Society (RSG-07-063-01-GMC). E.H.R. was supported by the Vanderbilt Training Program in Environmental Toxicology (T32 ES07028).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.05.078

References

1. Friedberg, E. C., Aguilera, A., Gellert, M., Hanawalt, P. C., Hays, J. B., Lehmann, A. R. *et al.* (2006). DNA repair: from molecular mechanism to human disease. *DNA Repair (Amst)*, **5**, 986–996.

2. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. (2006). A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA*, **103**, 5752–5757.
3. Francis, A. W. & David, S. S. (2003). *Escherichia coli* MutY and Fpg utilize a processive mechanism for target location. *Biochemistry*, **42**, 801–810.
4. Banerjee, A., Santos, W. L. & Verdine, G. L. (2006). Structure of a DNA glycosylase searching for lesions. *Science*, **311**, 1153–1157.
5. Stivers, J. T. & Jiang, Y. L. (2003). A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **103**, 2729–2759.
6. Fromme, J. C., Banerjee, A. & Verdine, G. L. (2004). DNA glycosylase recognition and catalysis. *Curr. Opin. Struct. Biol.* **14**, 43–49.
7. Morikawa, K., Matsumoto, O., Tsujimoto, M., Katayanagi, K., Ariyoshi, M., Doi, T. *et al.* (1992). X-ray structure of T4 endonuclease V: an excision repair enzyme specific for a pyrimidine dimer. *Science*, **256**, 523–526.
8. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. & Tainer, J. A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*, **80**, 869–878.
9. Bruner, S. D., Norman, D. P. & Verdine, G. L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859–866.
10. Sugahara, M., Mikawa, T., Kumasaka, T., Yamamoto, M., Kato, R., Fukuyama, K. *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**, 3857–3869.
11. Lau, A. Y., Schärer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. (1998). Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell*, **95**, 249–258.
12. Doherty, A. J., Serpell, L. C. & Ponting, C. P. (1996). The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res.* **24**, 2488–2497.
13. Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Spooner, E. *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**, 968–980.
14. Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P. & Tainer, J. A. (1992). Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. *Science*, **258**, 434–440.
15. Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H. *et al.* (1998). MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nature Struct. Biol.* **5**, 1058–1064.
16. Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L. & Ellenberger, T. E. (1996). Structural basis for the excision repair of alkylation-damaged DNA. *Cell*, **86**, 321–329.
17. Drohat, A. C., Kwon, K., Krosky, D. J. & Stivers, J. T. (2002). 3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member. *Nature Struct. Biol.* **9**, 659–664.
18. Eichman, B. F., O'Rourke, E. J., Radicella, J. P. & Ellenberger, T. (2003). Crystal structures of 3-methyladenine DNA glycosylase MagIII and the recognition of alkylated bases. *EMBO J.* **22**, 4898–4909.
19. Rydberg, B. & Lindahl, T. (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *EMBO J.* **1**, 211–216.
20. Hecht, S. S. (1999). DNA adduct formation from tobacco-specific *N*-nitrosamines. *Mutat. Res.* **424**, 127–142.
21. Hurley, L. H. (2002). DNA and its associated processes as targets for cancer therapy. *Nature Rev. Cancer*, **2**, 188–200.
22. Bjelland, S., Bjørås, M. & Seeberg, E. (1993). Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of *Escherichia coli*. *Nucleic Acids Res.* **21**, 2045–2049.
23. Begley, T. J., Haas, B. J., Noel, J., Shekhtman, A., Williams, W. A. & Cunningham, R. P. (1999). A new member of the endonuclease III family of DNA repair enzymes that removes methylated purines from DNA. *Curr. Biol.* **9**, 653–656.
24. O'Rourke, E. J., Chevalier, C., Boiteux, S., Labigne, A., Ielpi, L. & Radicella, J. P. (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**, 20077–20083.
25. Bjelland, S., Birkeland, N. K., Benneche, T., Volden, G. & Seeberg, E. (1994). DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the AlkA enzyme in *Escherichia coli*. *J. Biol. Chem.* **269**, 30489–30495.
26. McCarthy, T. V., Karran, P. & Lindahl, T. (1984). Inducible repair of O-alkylated DNA pyrimidines in *Escherichia coli*. *EMBO J.* **3**, 545–550.
27. Saparbaev, M., Kleibl, K. & Laval, J. (1995). *Escherichia coli*, *Saccharomyces cerevisiae*, rat and human 3-methyladenine DNA glycosylases repair 1,N^6-ethenoadenine when present in DNA. *Nucleic Acids Res.* **23**, 3750–3755.
28. Bjørås, M., Klungland, A., Johansen, R. F. & Seeberg, E. (1995). Purification and properties of the alkylation repair DNA glycosylase encoded the MAG gene from Saccharomyces cerevisiae. *Biochemistry*, **34**, 4577–4582.
29. Singer, B., Antoccia, A., Basu, A. K., Dosanjh, M. K., Fraenkel-Conrat, H., Gallagher, P. E. *et al.* (1992). Both purified human 1,N^6-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N^6-ethenoadenine and 3-methyladenine. *Proc. Natl Acad. Sci. USA*, **89**, 9386–9390.
30. O'Connor, T. R. (1993). Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic Acids Res.* **21**, 5561–5569.
31. Saparbaev, M. & Laval, J. (1994). Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl Acad. Sci. USA*, **91**, 5873–5877.
32. Alseth, I., Rognes, T., Lindbäck, T., Solberg, I., Robertsen, K., Kristiansen, K. I. *et al.* (2006). A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol. Microbiol.* **59**, 1602–1609.
33. Dalhus, B., Helle, I. H., Backe, P. H., Alseth, I., Rognes, T., Bjørås, M. & Laerdahl, J. K. (2007). Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats. *Nucleic Acids Res.* **35**, 2451–2459.
34. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

35. Hollis, T., Lau, A. & Ellenberger, T. (2001). Crystallizing thoughts about DNA base excision repair. *Progr. Nucleic Acid Res. Mol. Biol.* **68**, 305–314.
36. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. (2000). Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase. *AAG. Proc. Natl Acad. Sci. USA*, **97**, 13573–13578.
37. Hollis, T., Ichikawa, Y. & Ellenberger, T. (2000). DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli* AlkA. *EMBO J.* **19**, 758–766.
38. Metz, A. H., Hollis, T. & Eichman, B. F. (2007). DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG). *EMBO J.* **26**, 2411–2420.
39. Magnusdottir, A., Stenmark, P., Flodin, S., Nyman, T., Kotenyova, T., Nilsson, P. *et al*. Crystal structure of the human Pp2A regulatory subunit, B56G. PDB ID 2JAK-A.
40. Xu, Y., Xing, Y., Chen, Y., Chao, Y., Lin, Z., Fan, E. *et al.* (2006). Structure of the protein phosphatase 2A holoenzyme. *Cell*, **127**, 1239–1251.
41. Goldenberg, S. J., Cascio, T. C., Shumway, S. D., Garbutt, K. C., Liu, J., Xiong, Y. & Zheng, N. (2004). Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. *Cell*, **119**, 517–528.
42. Andrade, M. A., Petosa, C., O'Donoghue, S. I., Muller, C. W. & Bork, P. (2001). Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* **309**, 1–18.
43. Garces, R. G., Gillon, W. & Pai, E. F. (2007). Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with in vitro nucleic acid binding properties. *Protein Sci.* **16**, 176–188.
44. Wang, X., Zamore, P. D. & Hall, T. M. (2001). Crystal structure of a Pumilio homology domain. *Mol. Cell.* **7**, 855–865.
45. Stein, A. J., Fuchs, G., Fu, C., Wolin, S. L. & Reinisch, K. M. (2005). Structural insights into RNA quality control: the Ro autoantigen binds misfolded RNAs via its central cavity. *Cell*, **121**, 529–539.
46. Macdonald, P. M. (1992). The Drosophila pumilio gene: an unusually long transcription unit and an unusual protein. *Development*, **114**, 221–232.
47. Groves, M. R., Hanlon, N., Turowski, P., Hemmings, B. A. & Barford, D. (1999). The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell*, **96**, 99–110.
48. Conti, E. & Kuriyan, J. (2000). Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. *Structure*, **8**, 329–338.
49. Matsuura, Y. & Stewart, M. (2004). Structural basis for the assembly of a nuclear export complex. *Nature*, **432**, 872–877.
50. Huffman, J. L., Sundheim, O. & Tainer, J. A. (2005). DNA base damage recognition and removal: new twists and grooves. *Mutat. Res.* **577**, 55–76.
51. Schärer, O. D., Ortholand, J. Y., Ganesan, A., Ezaznikpay, K. & Verdine, G. L. (1995). Specific binding of the DNA-repair enzyme alka to a pyrrolidine-based inhibitor. *J. Am. Chem. Soc.* **117**, 6623–6624.
52. Schärer, O. D., Nash, H. M., Jiricny, J., Laval, J. & Verdine, G. L. (1998). Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases. *J. Biol. Chem.* **273**, 8592–8597.
53. Makino, K. & Ichikawa, Y. (1998). Synthesis of a 2-deoxyribose type 1-*N*-iminosugar. *Tetrahedron Lett.* **39**, 8245–8248.
54. O'Brien, P. J. & Ellenberger, T. (2004). The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J. Biol. Chem.* **279**, 26876–26884.
55. Osipiuk, J., Hatzos, C., Moy, S., Collart, F., Joachimiak, A. X-ray structure of predicted DNA alkylation repair enzyme from Enterococcus faecalis. PDB ID 2B6C-A.
56. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P. & Tainer, J. A. (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.* **14**, 4108–4120.
57. Park, J. H., Aravind, L., Wolff, E. C., Kaevel, J., Kim, Y. S. & Park, M. H. (2006). Molecular cloning, expression, and structural prediction of deoxyhypusine hydroxylase: a HEAT-repeat-containing metalloenzyme. *Proc. Natl Acad. Sci. USA*, **103**, 51–56.
58. Park, M. H., Joe, Y. A. & Kang, K. R. (1998). Deoxyhypusine synthase activity is essential for cell viability in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 1677–1683.
59. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
60. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230.
61. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240–255.
62. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
63. Asaeda, A., Ide, H., Asagoshi, K., Matsuyama, S., Tano, K., Murakami, A. *et al.* (2000). Substrate specificity of human methylpurine DNA *N*-glycosylase. *Biochemistry*, **39**, 1959–1965.
64. Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A. & Honig, B. (2002). Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **23**, 128–137.
65. Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M. & Tainer, J. A. (2000). Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc Natl Acad. Sci. USA*, **97**, 5083–5088.
66. Fromme, J. C. & Verdine, G. L. (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature Struct. Biol.* **9**, 544–552.
67. Vassylyev, D. G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E. & Morikawa, K. (1995). Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell*, **83**, 773–782.