

An unprecedented nucleic acid capture mechanism for excision of DNA damage

Emily H. Rubinson¹, A. S. Prakasha Gowda², Thomas E. Spratt², Barry Gold³ & Brandt F. Eichman¹

DNA glycosylases that remove alkylated and deaminated purine nucleobases are essential DNA repair enzymes that protect the genome, and at the same time confound cancer alkylation therapy, by excising cytotoxic N3-methyladenine bases formed by DNA-targeting anticancer compounds. The basis for glycosylase specificity towards N3- and N7-alkylpurines is believed to result from intrinsic instability of the modified bases and not from direct enzyme functional group chemistry. Here we present crystal structures of the recently discovered *Bacillus cereus* AlkD glycosylase in complex with DNAs containing alkylated, mismatched and abasic nucleotides. Unlike other glycosylases, AlkD captures the extrahelical lesion in a solvent-exposed orientation, providing an illustration for how hydrolysis of N3- and N7-alkylated bases may be facilitated by increased lifetime out of the DNA helix. The structures and supporting biochemical analysis of base flipping and catalysis reveal how the HEAT repeats of AlkD distort the DNA backbone to detect non-Watson-Crick base pairs without duplex intercalation.

Alkylation of DNA by endogenous methyl donors, environmental toxins and chemotherapeutic agents produces a diverse spectrum of cytotoxic and mutagenic lesions, including N3-methyladenine (3mA), N7-methylguanine (7mG) and 1,N⁶-ethenoadenine (εA), that threaten the survival of all organisms^{1–5}. 3mA is highly toxic owing to its inhibition of DNA polymerases during replication^{6,7}, and production of such lesions is the rationale behind the use of alkylating agents in chemotherapy. N7-substituted guanines are the most prevalent alkylation lesions and display a wide range of toxic and mutagenic biological properties⁸. By virtue of their positive charges at physiological pH, 3mA and 7mG are especially susceptible to spontaneous depurination, which generates abasic sites in DNA that can ultimately lead to single- and double-strand breaks.

DNA glycosylases initiate base excision repair of N3- and N7-methylpurines from the genome by catalysing hydrolysis of the N-glycosidic bond (Fig. 1a, b). Despite their structural diversity, all

DNA glycosylases studied so far use a common base-flipping mechanism to access damaged DNA and orient the substrate for catalysis by rotating the target nucleotide 180° around the phosphoribose backbone into a complementarily shaped active site pocket^{9,10}. The resulting distortion to the DNA is stabilized by an intercalating side-chain ‘plug’ that fills the void created by the extrahelical nucleotide. Glycosylases typically excise their target nucleobases by using a carboxylate side chain as a general base to activate a water nucleophile or to stabilize the carbocation transition state during base dissociation¹¹. Mutation of this residue, however, does not abolish catalytic activity in all cases, leading to a model in which conformational strain in the DNA arising from extensive binding energy helps to drive the reaction forward^{12,13}. The lack of a residue capable of performing general base catalysis in 3mA-specific DNA glycosylases (for example, *Escherichia coli* TAG)^{14–16} is consistent with the idea that excision of positively charged 3mA and 7mG does not require the same level of catalytic assistance as more stable ethenoadducts, although direct evidence for this has not been reported.

AlkC and AlkD proteins, recently discovered in *Bacillus cereus* and subsequently identified in all three kingdoms of life (Supplementary Fig. 1), have emerged as a unique DNA glycosylase superfamily specific for N3- and N7-alkylpurines^{17,18}. AlkD accelerates the rate of 7mG depurination¹⁹, prompting us to investigate the mechanism by which AlkD excises destabilized alkylated bases. Here we present crystal structures of *B. cereus* AlkD in complex with DNA damage resembling the substrate and product of the glycosylase reaction. These structures, together with supporting biochemistry of base-flipping and 7mG depurination activities, demonstrate how AlkD uses an unprecedented strategy to trap non-canonical base pairs that allows for specific hydrolysis of destabilized N-glycosidic bonds without direct chemical attack from the enzyme.

A new architecture for binding nucleic acids

We previously determined the crystal structure of *B. cereus* AlkD and identified residues important for DNA binding and catalysis¹⁹. AlkD

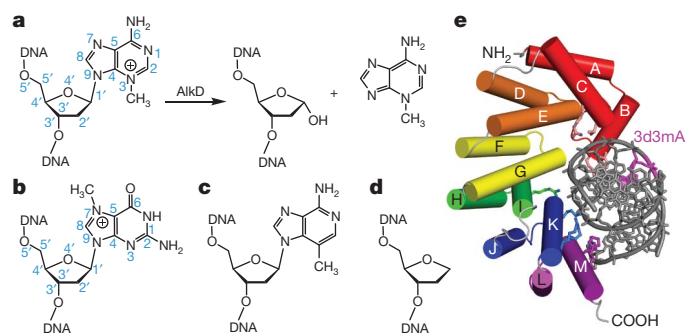


Figure 1 | Base excision repair of alkylated DNA by AlkD. **a**, AlkD catalyses the hydrolysis of the N-glycosidic bond to liberate an abasic site and free nucleobase. The enzyme is specific for positively charged N3-methyladenine (**a**) and N7-methylguanine (**b**). **c**, **d**, Structures of 3-deaza-3-methyladenosine (**c**) and tetrahydrofuran (**d**) used to trap AlkD in complex with alkylated and abasic DNA. **e**, Crystal structure of AlkD bound to 3d3mA-DNA. Each of the six HEAT repeats is coloured red-to-violet. The DNA is coloured silver with the 3d3mA nucleotide coloured magenta.

¹Department of Biological Sciences and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37232, USA. ²Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA. ³Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA.

is comprised entirely of HEAT repeats—tandem α -helical pairs that generate extended, non-enzymatic scaffolds that typically mediate protein, but not nucleic acid, interactions within their concave surfaces^{20–22}. AlkD's concave surface contains highly conserved residues important for 7mG excision and DNA binding activities and protection against bacterial sensitivity to alkylating agents^{17–19}.

To investigate the mechanisms by which this novel enzyme binds DNA and catalyses base excision, we determined crystal structures of *B. cereus* AlkD in complex with DNAs resembling the substrate and product of 3mA excision (Fig. 1a). Trapping an alkylpurine DNA glycosylase onto a 3mA-containing substrate has presented a formidable challenge owing to the inherent instability of the *N*-glycosidic bond. To overcome this obstacle, we crystallized AlkD in complex with DNA containing 3-deaza-3-methyladenine (3d3mA), a structural 3mA mimetic in which the N3 nitrogen is replaced with carbon (Fig. 1c). The 3d3mA base is refractory to spontaneous depurination or excision by AlkD or human alkyladenine DNA glycosylase (AAG)⁷, presumably because the 3d3mA purine ring lacks the formal positive charge associated with 3mA. Importantly, the N3→C3 substitution does not affect duplex stability (Supplementary Information)²³. We also crystallized AlkD in complex with DNA containing a tetrahydrofuran (THF) moiety (Fig. 1d), which resembles the abasic site product. The AlkD–3d3mA–DNA and AlkD–THF–DNA structures were determined by molecular replacement and refined to 1.6 Å ($R/R_{\text{free}} = 15.9\%/18.3\%$) and 1.75 Å ($R/R_{\text{free}} = 18.5\%/22.5\%$), respectively (Supplementary Table 1 and Supplementary Fig. 2).

Both 3d3mA and THF complexes show the same general mode of nucleic acid binding despite their unique DNA sequences and crystal packing arrangements (Fig. 2). The DNA is positioned along AlkD's concave surface, which is lined with positively charged residues from the carboxy-terminal α -helix of each HEAT repeat (Fig. 1e and Supplementary Fig. 3). The C-shaped protein wraps halfway around the DNA helix with a footprint of ~ 10 bp. The contact surface is dominated by electrostatic interactions between side chains at the protein mid-region and the phosphoribose backbone of the DNA strand opposite the lesion. In contrast, contacts to the lesioned strand

are limited to base pairs further removed from the lesion and the protein termini (Fig. 2). The DNA axes are bent 30° away from AlkD's amino terminus as a result of helix α B (the only non-HEAT repeat in AlkD) projecting into the minor groove (Fig. 2). A 2 Å shift in helix α B is the only noticeable movement in the protein upon DNA binding (Supplementary Fig. 4).

A novel lesion capture mechanism

The most striking feature of the AlkD–DNA complexes is that both 3d3mA and THF reside on the face of the DNA duplex not in contact with the protein, whereas the base opposite the lesion is nestled into a cleft on the protein's concave surface (Figs 2 and 3). The 3d3mA•T unpredictably forms a highly sheared base pair in which 3d3mA remains stacked between T6 and A8, whereas the opposite thymine (T18) is displaced into the minor groove with no hydrogen bonds to 3d3mA (Fig. 3a). There are no protein contacts to the T18 base. Rather, it is held in this position by distortion of the T18/A19 backbone as a result of a hydrogen bond network among Asp 113–Arg 148 and Arg190. The protein–DNA interface is further strengthened by van der Waals interactions between tryptophans 109 and 187 and the phosphoribose backbone flanking the damaged base pair.

In the product complex, the abasic site is rotated $\sim 90^\circ$ around the phosphoribose backbone into the major groove, and is fully solvent exposed (Figs 2b and 3b). Interestingly, the opposite thymine is slipped completely out of the base stack and into the minor groove of the DNA, and is rotated ($\chi = 58^\circ$) so that the plane of the pyrimidine ring is virtually parallel with the helical axis. Unlike other DNA glycosylases, there is no intercalating side chain plugging the gap left by the flipped base. As a consequence, the duplex has collapsed to maintain base stacking interactions. Guanine G4, immediately 5' to the THF, is now stacked with cytosine C16 on the opposite strand (Fig. 3b). Importantly, the DNA backbone is highly distorted as a result of the large slide (4.4 Å) and twist (58°) between G4•C18 and G6•C16 base pairs (Figs 2b and 3b and Supplementary Fig. 12). A hydrogen bond between Tyr 27 at the C-terminal end of helix α B and the base 3' to the tipped thymine is the only specific AlkD–nucleobase contact (Fig. 3).

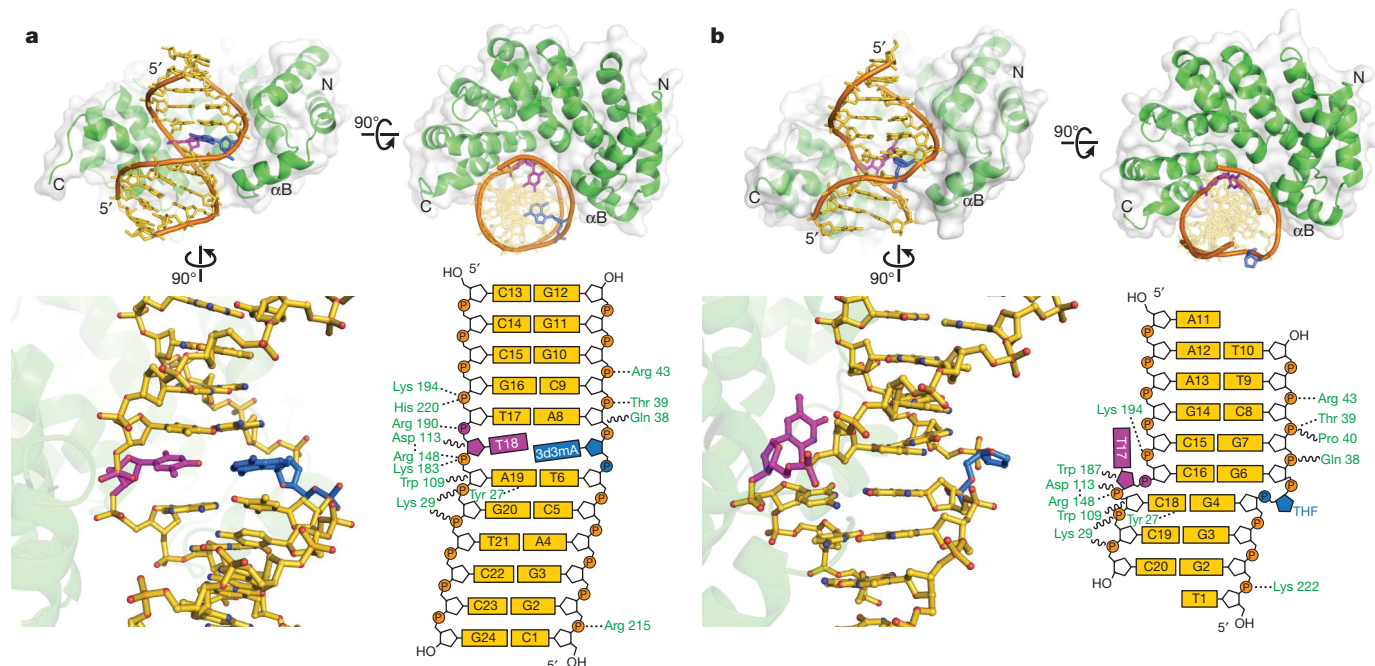


Figure 2 | Crystal structures of AlkD in complex with 3d3mA–DNA (a) and THF–DNA (b). The top of each panel shows orthogonal views of the AlkD protein (green) wrapping around the DNA duplex (gold). The modified 3d3mA and THF nucleotides are coloured blue, and the opposing thymine is

magenta. At the bottom, a side view of the atomic model and corresponding schematic illustrates the interactions between the modified base pairs and the protein. Dashed lines represent hydrogen bonds and wavy lines represent van der Waals interactions.

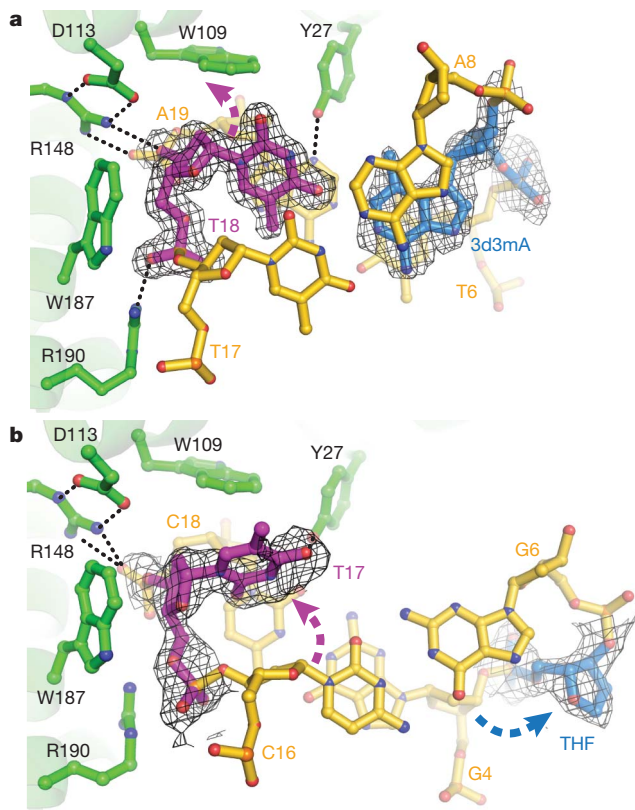


Figure 3 | Recognition of DNA damage by AlkD. **a**, 3d3mA-DNA (substrate) complex. **b**, THF-DNA (product) complex. Composite omit electron density (contoured to 1σ) for the modified base pairs is superimposed against the crystallographic models. Dashed arrows denote displacement of THF and opposing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

Thus, AlkD stabilizes the distortions in both substrate and product DNA—a sheared 3d3mA•T base pair and a single base THF•T bubble—through interactions with the phosphoribose backbone of the non-lesioned strand.

The solvent-exposed capture of DNA damage in the AlkD-DNA structures is both unexpected and unprecedented for a DNA glycosylase, and raises the possibility that either AlkD uses a different mechanism to catalyse base excision or that the crystal structures represent nonspecific, catalytically incompetent protein-DNA complexes. Indeed, the aromatic region at the centre of the concave cleft loosely resembles nucleobase binding pockets of other alkylpurine DNA glycosylases^{18,19}. However, several important differences argue against a traditional lesion binding pocket in AlkD. First, AlkD lacks the plug residue universally used by DNA glycosylases to prevent the flipped substrate base from re-entering the DNA base stack. Second, an extrahelical nucleobase would be sterically prohibited from full 180° rotation into this shallow cleft (Fig. 3). Third, high concentrations of free nucleobases do not inhibit base excision activity by AlkD as observed in other alkylpurine glycosylases (Supplementary Fig. 5)¹⁵. Fourth, the electrostatic interaction between Asp 113 and Arg 148 reduces the likelihood that Asp 113 acts as a general base in catalysis. Fifth, mutation of a putative base binding cleft directly adjacent to the catalytic Asp 113 and Arg 148 did not affect 7mG excision activity (Supplementary Fig. 6). Finally, whereas alkylpurine DNA glycosylases normally exhibit enhanced excision activity for mismatched alkylbases, presumably because of their greater propensity to base flip^{15,24,25}, AlkD does not discriminate against the base opposite the lesion (Supplementary Table 2 and Supplementary Fig. 7).

To determine the orientation of DNA relative to the central cleft during catalysis, we measured the rate of 7mG excision opposite a

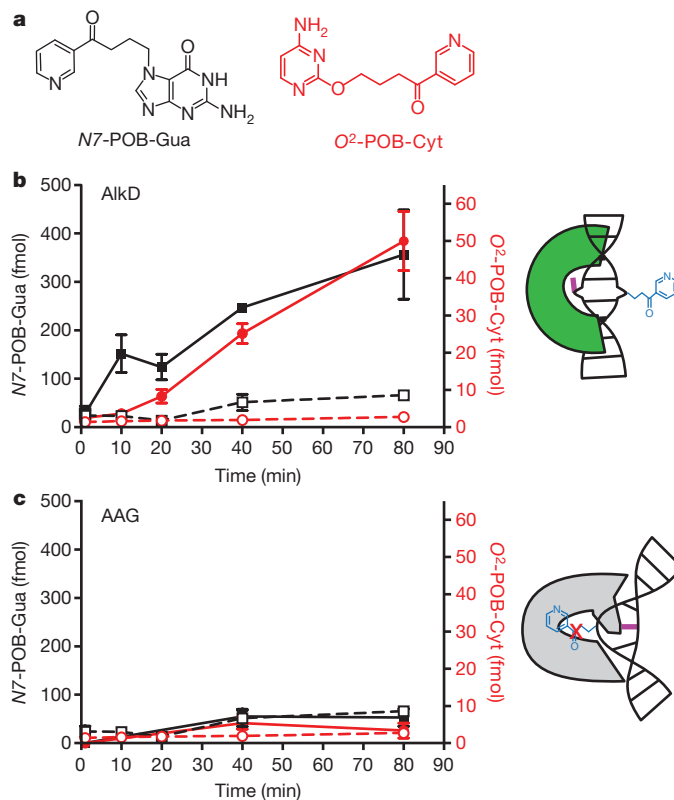


Figure 4 | Excision of *N*7- and *O*²-pyridyloxobutyl (POB) base adducts by AlkD. **a**, Chemical structures of *N*7-POB-guanine and *O*²-POB-cytosine. **b**, **c**, Time courses for the release of *N*7-POB-Gua (black squares) and *O*²-POB-Cyt (red circles) in the presence (filled symbols, solid lines) and absence (open symbols, dashed lines) of *B. cereus* AlkD (**b**) or human AAG (**c**). Error bars represent the standard deviation from three independent measurements.

bulky nucleotide. A pyrene nucleotide wedge across from uracil has been shown to enhance base excision by uracil DNA glycosylase (UDG) and rescue the loss of activity of UDG mutants that lack the Leu 191 plug side chain²⁶. In contrast, placing pyrene across from 7mG reduced AlkD's activity tenfold relative to a 7mG•C pair (Supplementary Fig. 7). Superposition of the pyrene onto the opposite thymine in the AlkD-DNA crystal structures showed that this bulky group would be hindered from rotating into this tipped position. Thus, the consistency between the crystal structures and partial inhibition of 7mG activity by an opposing pyrene argue strongly that the crystal structures represent a catalytically competent orientation of DNA.

In a converse experiment, we tested the ability of AlkD to excise bulky pyridyloxobutyl (POB) base adducts (Fig. 4a), which arise in DNA upon exposure to cigarette-smoke-derived nitrosamine carcinogens²⁷. The expectation was that AlkD should excise POB bases from DNA, whereas the tightly constrained nucleobase binding pocket of human AAG would discriminate against bulky alkyl adducts²⁸. Indeed, AlkD liberated positively charged *N*7-POB-Gua and *O*²-POB-Cyt adducts from DNA, whereas neither of these modified bases was detected after treatment with AAG or in a mock reaction containing no enzyme (Fig. 4b, c). Neutral adducts *O*⁶-POB-Gua and *O*²-POB-Thy present in the DNA were not detected in the supernatant upon reaction with AlkD, consistent with the specificity of AlkD for positively charged lesions. This result indicates that AlkD need not flip the substrate base into an active site cavity to excise *N*3- or *N*7-alkylpurines from DNA.

AlkD traps and restructures destabilized base pairs

Recent work suggests that DNA glycosylases and oxidative demethylases detect damage by using side chains to probe for free energy

differences between normal and modified base pairs^{29–32}. The lack of lesion-specific and DNA intercalating interactions in the AlkD–DNA complexes implies that AlkD detects damage solely on the basis of DNA duplex destabilization resulting from altered stacking or pairing of non-canonical base pairs. In support of this, we crystallized the protein in complex with DNA containing a G•T mismatch (Fig. 5a and Supplementary Table 1), for which AlkD has no activity, but were unable to trap the protein onto the same oligonucleotide containing a G•C or A•T base pair at this same position. The resulting 1.5 Å AlkD–G•T–DNA structure is virtually identical to the 3d3mA•T complex (Supplementary Table 1 and Supplementary Fig. 4). The similarity in these structures, together with thermodynamic differences between modified and unmodified nucleobases, indicates that AlkD detects these energetic differences as opposed to specifically recognizing the N3- or N7-methyl groups (see Supplementary Information)^{23,33–36}.

Comparison of the G•T mismatch bound by AlkD and in the context of DNA alone provides a basis for DNA damage recognition by AlkD (Fig. 5). In DNA, G•T wobble mismatches form two Watson–Crick hydrogen bonds and are well stacked within the duplex³⁷ (Supplementary Fig. 10). AlkD restructures the G•T wobble so that the two bases protrude into opposite DNA grooves, disrupting base stacking and leaving only a single hydrogen bond between guanine N² and thymine O⁴ (Fig. 5a). Superposition of a canonical G•T wobble onto the AlkD structure revealed that the protein stabilizes this conformation by inducing a specific distortion to the DNA backbone to alleviate steric clashes (Fig. 5b) and to create optimal hydrogen bonding and van der Waals interactions at the DNA capture site (Supplementary Fig. 10). Thus, the enzyme detects non-Watson–Crick base pairs by resculpting the DNA backbone to create an optimized protein–DNA binding surface. In both 3d3mA•T and G•T complexes, specific protein–DNA contacts are mediated by Arg 148–Asp 113 and Arg 190. Substitution of any of these highly conserved residues reduces single-turnover rates of 7mG excision by an order of magnitude (Fig. 5c), highlighting the importance of these interactions to catalysis.

Base excision by solvent exposure

The specific structure of the DNA trapped in the AlkD complexes provides a rationale for the enzyme's specificity towards bases with a high propensity for depurination. We propose that the lesion capture mechanism facilitates base hydrolysis by increasing the lifetime that the N-glycosidic bond is exposed to solvent, consistent with spontaneous depurination rates of 7mG in different DNA secondary structural contexts (see Supplementary Information). However, the 100-fold rate enhancement of 7mG hydrolysis from duplex DNA by AlkD cannot be explained on the basis of solvent exposure alone. Close inspection of the highly distorted DNA backbone in the flipped abasic structure revealed

that the deoxyribose ring is positioned directly above a neighbouring phosphate and that several water molecules bridge this phosphate and the extrahelical deoxyribose C1' carbon (Supplementary Fig. 12a), raising the possibility that the phosphate groups participate in catalysis. DNA-mediated water positioning to facilitate hydrolysis is a plausible catalytic mechanism given the lack of a requirement for a general base in these probably highly dissociative reactions. Alternatively, electrostatic stabilization of an oxocarbenium intermediate by nearby phosphates, which has been reported for uracil DNA glycosylase^{38,39}, offers a second possible mechanism for DNA-stimulated catalysis.

Discussion

AlkD represents a novel glycosylase found in bacteria, archaea, plants and eukaryotes (Supplementary Fig. 1)¹⁷. To our knowledge, most if not all of these organisms contain at least one other alkylpurine DNA glycosylase, raising the question as to why an alternative mechanism has evolved to eliminate genomic alkylation damage. The redundancy of alkylation repair may provide enhanced protection to organisms faced with an onslaught of methylating agents. Alternatively, AlkD may be a general DNA binding protein that coincidentally accelerates hydrolysis of unstable N-glycosidic bonds, or, as speculated below, may have a supporting role in general lesion detection.

AlkD's activity towards bulky POB-DNA adducts normally associated with nucleotide excision repair^{40,41} may be indicative of a more generalized function of AlkD in genome maintenance. AlkD's lesion capture strategy is reminiscent of Rad4/XPC, which recognizes cyclopyrimidine dimers by binding to the opposing nucleotides⁴². Exposure of the lesion away from the protein has the biological advantage of damage accessibility by the rest of the nucleotide excision repair machinery. The AlkD–product complex may provide a platform for recruitment of a protein against the extrahelical abasic site, as seen in human APE1–DNA complexes⁴³. It is intriguing to speculate that AlkD may participate in alternative repair pathways by virtue of its ability to expose DNA damage. Indeed, non-enzymatic alkyltransferase-like proteins were recently found to trigger nucleotide excision repair of O⁶-alkylguanines by inducing a specific protein–DNA complex, as illustrated by the crystal structure of ATL bound to DNA containing O⁶-POB-dG⁴⁴.

The AlkD–DNA structures illustrate how HEAT repeats engage nucleic acids and, to our knowledge, provide the first structural example of a HEAT motif with enzymatic activity. Comparison with nuclear import factor importin β , which uses HEAT repeats to bind a highly charged region of importin α ²¹ and Ran GTPase²², demonstrates that the concave surface of the HEAT domain is a generalized macromolecular binding platform. HEAT repeats have been identified in chromatin-remodelling factors, including condensins, cohesins and some SWI2/SNF2 proteins⁴⁵, as well as DNA-damage-response protein

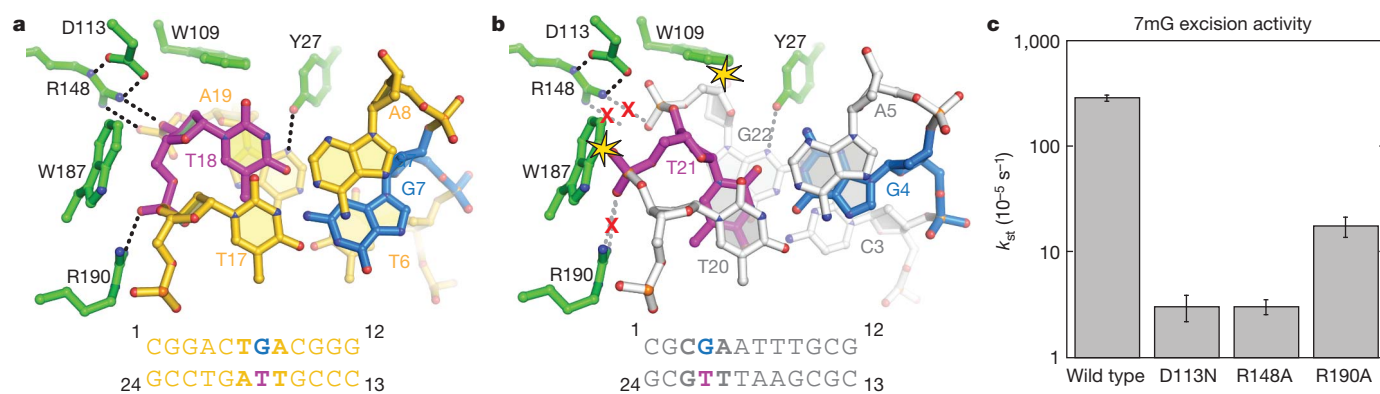


Figure 5 | Remodelling of a G•T wobble base pair by AlkD. **a**, AlkD–G•T–DNA complex viewed down the helical axis. AlkD is in green. **b**, The structure of a G•T wobble base pair in DNA alone (Protein Data Bank ID 113D) is superimposed onto the AlkD–G•T complex. Steric clashes between the protein and DNA are highlighted by yellow stars, and disrupted hydrogen bonds

(dashed lines) are shown by a red X. **c**, Relative single-turnover rates (k_{cat}) of 7mG excision from a 25mer oligonucleotide duplex by wild-type AlkD and the indicated AlkD mutants. Wild-type, D113N and R148A data are from ref. 19. Error bars represent the standard deviation from three independent measurements.

kinases ATM, ATR and DNA-PK⁴⁶. Recently, HEAT domains were visualized by electron microscopy and crystal structures of the catalytic subunit of DNA-PK^{47,48}, raising the possibility that other structurally uncharacterized DNA processing enzymes use HEAT domains to bind DNA in a manner similar to AlkD.

METHODS SUMMARY

Preparation of 3-deaza-3-methyladenine. The 3-methyl-3-deazaadenine phosphoramidite was prepared as previously described⁴⁹. The 3-methyl-3-deazaadenine-modified deoxynucleotide oligomers were synthesized at the University of Pittsburgh DNA core facility, purified by reverse-phase HPLC, desalted on Sephadex G20 and analysed by MALDI-TOF-MS. All other oligonucleotides were synthesized by Integrated DNA Technologies.

AlkD-DNA crystal structure determination. Wild-type and mutant *B. cereus* AlkD proteins were purified as described previously¹⁹. AlkD-DNA complexes were assembled using a 1:1.2 molar ratio of AlkD:DNA. AlkD-THF-DNA crystals were grown at 16 °C by vapour diffusion against a reservoir containing 0.1 M Bis-Tris pH 6.5, 0.1 mM NaCl and 9% PEG 3350, and were flash frozen in a 30% glycerol/reservoir solution. AlkD-3d3mA-DNA and AlkD-G•T-DNA crystals were grown at 21 °C from reservoir solutions containing 85 mM NaAcetate pH 4.6, 170 mM ammonium acetate, 25.5% PEG 4000, and 15% glycerol, and were flash frozen directly from the mother liquor. X-ray data (Supplementary Table 1) were collected at the Advanced Photon Source (21-ID-D, LS-CAT). All structures were determined by molecular replacement using the unliganded AlkD structure (Protein Data Bank ID 3BVS) as a search model.

Biochemical assays. Base excision and DNA binding activity assays were performed as previously described¹⁹. Kinetic data were analysed by standard single-turnover techniques. Activity towards POB-nucleobases was measured by incubation of NNK-treated genomic DNA with AlkD or AAG, followed by DNA precipitation and mass-spectrometric detection of N7-POB-Gua and O²-POB-Cyt in the supernatant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 November 2009; accepted 17 August 2010.

Published online 3 October 2010.

- Friedberg, E. C. *et al.* DNA repair: from molecular mechanism to human disease. *DNA Repair (Amst.)* **5**, 986–996 (2006).
- Singer, B. & Grunberger, D. *Molecular Biology of Mutagens and Carcinogens: Intrinsic Properties of Nucleic Acids* (Plenum, 1983).
- Holt, S., Yen, T. Y., Sangai, R. & Swenberg, J. A. Detection of 1,N⁶-ethenoadenine in rat urine after chloroethylene oxide exposure. *Carcinogenesis* **19**, 1763–1769 (1998).
- Shuker, D. E., Bailey, E., Parry, A., Lamb, J. & Farmer, P. B. The determination of urinary 3-methyladenine in humans as a potential monitor of exposure to methylating agents. *Carcinogenesis* **8**, 959–962 (1987).
- Shuker, D. E. & Farmer, P. B. Relevance of urinary DNA adducts as markers of carcinogen exposure. *Chem. Res. Toxicol.* **5**, 450–460 (1992).
- Larson, K., Sahm, J., Shenkar, R. & Strauss, B. Methylation-induced blocks to *in vitro* DNA replication. *Mutat. Res.* **150**, 77–84 (1985).
- Plosky, B. S. *et al.* Eukaryotic Y-family polymerases bypass a 3-methyl-2'-deoxyadenosine analog *in vitro* and methyl methanesulfonate-induced DNA damage *in vivo*. *Nucleic Acids Res.* **36**, 2152–2162 (2008).
- Gates, K. S., Nooner, T. & Dutta, S. Biologically relevant chemical reactions of N7-alkylguanine residues in DNA. *Chem. Res. Toxicol.* **17**, 839–856 (2004).
- Stivers, J. T. Site-specific DNA damage recognition by enzyme-induced base flipping. *Prog. Nucleic Acid Res. Mol. Biol.* **77**, 37–65 (2004).
- Stivers, J. T. Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chemistry* **14**, 786–793 (2008).
- Stivers, J. T. & Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **103**, 2729–2760 (2003).
- Parikh, S. S. *et al.* 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 (2000).
- Mol, C. D., Arvai, A. S., Begley, T. J., Cunningham, R. P. & Tainer, J. A. Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases. *J. Mol. Biol.* **315**, 373–384 (2002).
- Drohat, A. C., Kwon, K., Krosky, D. J. & Stivers, J. T. 3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member. *Nature Struct. Biol.* **9**, 659–664 (2002).
- Eichman, B. F., O'Rourke, E. J., Radicella, J. P. & Ellenberger, T. Crystal structures of 3-methyladenine DNA glycosylase MagIII and the recognition of alkylated bases. *EMBO J.* **22**, 4898–4909 (2003).
- Metz, A. H., Hollis, T. & Eichman, B. F. DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG). *EMBO J.* **26**, 2411–2420 (2007).
- Alseth, I. *et al.* A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol. Microbiol.* **59**, 1602–1609 (2006).
- Dalhus, B. *et al.* Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats. *Nucleic Acids Res.* **35**, 2451–2459 (2007).
- Rubinson, E. H., Metz, A. H., O'Quin, J. & Eichman, B. F. A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD. *J. Mol. Biol.* **381**, 13–23 (2008).
- Andrade, M. A. & Bork, P. HEAT repeats in the Huntington's disease protein. *Nature Genet.* **11**, 115–116 (1995).
- Cingolani, G., Petosa, C., Weis, K. & Muller, C. W. Structure of importin-β bound to the IBB domain of importin-α. *Nature* **399**, 221–229 (1999).
- Vetter, I. R., Arndt, A., Kutay, U., Gorlich, D. & Wittlinghofer, A. Structural view of the Ran-Importin β interaction at 2.3 Å resolution. *Cell* **97**, 635–646 (1999).
- Ganguly, M., Wang, R.-W., Marky, L. A. & Gold, B. Thermodynamic characterization of DNA with 3-deazaadenine and 3-methyl-3-deazaadenine substitutions. *J. Phys. Chem. B* **114**, 7656–7661 (2010).
- O'Brien, P. J. & Ellenberger, T. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.* **279**, 9750–9757 (2004).
- O'Brien, P. J. & Ellenberger, T. The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J. Biol. Chem.* **279**, 26876–26884 (2004).
- Jiang, Y. L., Kwon, K. & Stivers, J. T. Turning on uracil-DNA glycosylase using a pyrene nucleotide switch. *J. Biol. Chem.* **276**, 42347–42354 (2001).
- Hecht, S. S. DNA adduct formation from tobacco-specific N-nitrosamines. *Mutat. Res.* **424**, 127–142 (1999).
- Lau, A. Y., Schärer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258 (1998).
- Banerjee, A., Santos, W. L. & Verdine, G. L. Structure of a DNA glycosylase searching for lesions. *Science* **311**, 1153–1157 (2006).
- Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612–618 (2005).
- Yang, C. G., Garcia, K. & He, C. Damage detection and base flipping in direct DNA alkylation repair. *ChemBioChem* **10**, 417–423 (2009).
- Yang, C. G. *et al.* Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961–965 (2008).
- Aboul-ela, F., Koh, D., Tinoco, I. Jr & Martin, F. H. Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A, C, G, T). *Nucleic Acids Res.* **13**, 4811–4824 (1985).
- Ezaz-Nikpay, K. & Verdine, G. L. Aberrantly methylated DNA: site-specific introduction of N-7-methyl-2'-deoxyguanosine into the Dickerson/Drew dodecamer. *J. Am. Chem. Soc.* **114**, 6562–6563 (1992).
- Ezaz-Nikpay, K. & Verdine, G. L. The effects of N7-methylguanine on duplex DNA structure. *Chem. Biol.* **1**, 235–240 (1994).
- Lee, S., Bowman, B. R., Ueno, Y., Wang, S. & Verdine, G. L. Synthesis and structure of duplex DNA containing the genotoxic nucleobase lesion N7-methylguanine. *J. Am. Chem. Soc.* **130**, 11570–11571 (2008).
- Hunter, W. N. *et al.* The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.* **262**, 9962–9970 (1987).
- Dinner, A. R., Blackburn, G. M. & Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **413**, 752–755 (2001).
- Jiang, Y. L., Ichikawa, Y., Song, F. & Stivers, J. T. Powering DNA repair through substrate electrostatic interactions. *Biochemistry* **42**, 1922–1929 (2003).
- Brown, P. J., Bedard, L. L. & Massey, T. E. Repair of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-induced DNA pyridyloxobutylations by nucleotide excision repair. *Cancer Lett.* **260**, 48–55 (2008).
- Li, L. *et al.* The influence of repair pathways on the cytotoxicity and mutagenicity induced by the pyridyloxobutylations pathway of tobacco-specific nitrosamines. *Chem. Res. Toxicol.* **22**, 1464–1472 (2009).
- Min, J. H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570–575 (2007).
- Mol, C. D., Izumi, T., Mitra, S. & Tainer, J. A. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature* **403**, 451–456 (2000).
- Tubbs, J. L. *et al.* Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* **459**, 808–813 (2009).
- Neuwald, A. F. & Hirano, T. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome Res.* **10**, 1445–1452 (2000).
- Perry, J. & Kleckner, N. The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell* **112**, 151–155 (2003).
- Williams, D. R., Lee, K. J., Shi, J., Chen, D. J. & Stewart, P. L. Cryo-EM structure of the DNA-dependent protein kinase catalytic subunit at subnanometer resolution reveals α helices and insight into DNA binding. *Structure* **16**, 468–477 (2008).
- Sibanda, B. L., Chirgadze, D. Y. & Blundell, T. L. Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature* **463**, 118–121 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Stivers for providing the pyrene phosphoramidite, Z. Warzak and LS-CAT beamline staff at the Advanced Photon Source (APS) for assistance with X-ray data collection, and T. Ellenberger, J. Stivers and P. O'Brien for

comments on the manuscript. Use of the APS was supported by the US Department of Energy Office of Basic Energy Sciences. Use of LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor. This research was supported by a grant from the American Cancer Society (to B.F.E.) and the NIH (RO1 CA29088 to B.G.). E.H.R. was supported in part by the Vanderbilt Training Program in Molecular Toxicology. Additional support for local crystallography facilities was provided by the Vanderbilt Center in Molecular Toxicology and the Vanderbilt-Ingram Cancer Center.

Author Contributions E.H.R. purified and crystallized AlkD, determined crystal structures and performed 7mG activity assays; B.G. synthesized 3d3mA

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

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 3JX7 (3d3mA•T), 3JXY (G•T), 3JXZ (THF•T) and 3JY1 (THF•C). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.F.E. (brandt.eichman@vanderbilt.edu).

METHODS

AlkD purification and crystallization. AlkD proteins were purified as described previously¹⁹. Briefly, *Bacillus cereus* AlkD was overexpressed as an N-terminal His₆-SUMO-AlkD fusion protein in *E. coli* HMS174 cells for 3 h at 37 °C. AlkD was isolated using Ni-NTA (Qiagen) affinity chromatography, followed by cleavage of the His₆-SUMO tag and further purification by heparin affinity and gel filtration chromatography. Protein was concentrated to 12.5 mg ml⁻¹ and stored in 20 mM Bis-Tris propane, 100 mM NaCl, 2 mM DTT and 0.1 mM EDTA. Site-directed mutagenesis of the wild-type AlkD vector was performed using a Quik-Change Kit (Stratagene). Mutant proteins were overexpressed and purified identically to wild-type AlkD, and their structures verified by circular dichroism spectroscopy.

AlkD-DNA complexes were assembled by incubating 0.45 mM protein and 0.54 mM oligonucleotide for 15 min at 4 °C. Oligonucleotide sequences used were d(TGGG(THF)GGCTT)/d(AAAGCCYCCC), in which Y = T or C, and d(CGGACTXACGGG)/d(CCCGTTTCCG), in which X = 3d3mA⁴⁹ or G. AlkD-THF-DNA crystals were grown at 16 °C by mixing 2 µl protein-DNA complex with 2 µl reservoir solution containing 0.1 M Bis-Tris pH 6.5, 0.1 mM NaCl, and 19% PEG 3350 and 2% glycerol. Crystals were soaked in 30% glycerol/reservoir solution for 1 min and flash frozen in a liquid nitrogen stream. Crystals of 3d3mA-DNA and G•T-DNA complexes were grown from reservoir solutions containing 85 mM NaAcetate pH 4.6, 170 mM ammonium acetate, 25.5% PEG 4000, and 15% glycerol at 21 °C, and were flash frozen in liquid nitrogen directly from this solution.

X-ray data collection, phasing and structure refinement. X-ray data (Supplementary Table 1) were collected at a wavelength of 0.97850 Å and 110 K at the Advanced Photon Source beamlines 21-ID-D and 21-ID-G (LS-CAT) and processed with HKL2000⁵⁰. Molecular replacement using unliganded AlkD (Protein Data Bank ID 3BVS) as a search model in Phaser⁵¹ gave a clear solution for each structure. After one round of simulated annealing refinement in CNS⁵², the entire DNA molecules could be discerned and were built into 2F_o - F_c and F_o - F_c electron density using XtalView⁵³. Atomic coordinates and B-factors for the protein-DNA models were refined in Phenix⁵⁴. TLS refinement with protein and each DNA chain defined as three separate TLS groups was carried out for each model except the GT-complex. Individual anisotropic B-factors were derived from the refined TLS parameters and held fixed during subsequent rounds of refinement, which significantly decreased the crystallographic residuals and improved the electron density maps. Instead of TLS refinement, individual anisotropic B-factors were explicitly refined for the G•T complex. Adjustments to the model, including addition of solvent molecules, using Coot⁵⁵ were guided by manual inspection of 2F_o - F_c and F_o - F_c electron density maps and were judged successful by a decrease in R_{free} during refinement.

Protein and DNA models were validated using PROCHECK⁵⁶ and DNA parameters were quantified using CURVES 5.2⁵⁷. All but one out of the total 223–231 protein residues resided in the most favoured (191–198 residues) or allowed (14–15 residues) regions of the Ramachandran plot. As in the unliganded structure¹⁹, Thr 54 in all four DNA complex structures remained in the disallowed region despite an excellent fit to 2F_o - F_c electron density maps.

Enzyme activity. Excision of 7mG by AlkD was measured by incubating the enzyme with a 25mer oligonucleotide containing a centrally located 7mG and following the appearance of abasic DNA product after alkaline cleavage. 7mG was enzymatically incorporated into DNA duplexes using the previously described method⁵⁸, in which an oligonucleotide primer (d(GACCACTACACC)) was ³²P-labelled at the 5' end, annealed to a threefold excess of the complementary strand (d(GTTGTTAGGAAACGGTGTAGTGGTC)) and extended using DNA polymerase I Klenow fragment (New England Biolabs) in the presence of 2'-deoxy-7-methylguanosine 5'-triphosphate (Sigma), dCTP, dTTP and dATP. To create 7mG mispairs, 100-fold excess of complementary strand with T, G, A or pyrene in place of C at position 13 was re-annealed to the 7mG containing oligonucleotide. Single-stranded 7mG containing strands were obtained by re-annealing to 100-fold excess of unlabelled lesion strand with G in place of 7mG (d(GACCACTACACCGTTTCCTAACAAC)).

In a 10 µl glycosylase reaction, 100 nM [³²P]-DNA duplex was incubated with 0–20 µM AlkD in 50 mM HEPES pH 7.5, 100 mM KCl, 10 mM DTT and 2 mM EDTA. The reaction was quenched at various times by the addition of 0.2 N NaOH and heated at 70 °C for 2 min. Substrate 25mer and product 12mer DNA strands were separated by denaturing 20% polyacrylamide gel electrophoresis in 7 M urea and quantified by autoradiography. Kinetic data were analysed by standard single-turnover techniques⁵⁹, which have been extensively used for DNA glycosylases^{60–65}. Enzymatic rate constants (*k*) were obtained from single-exponential fits to the data ($f_p = 1 - e^{-kt}$, in which *f_p* is the fraction of product). For determination of the single-turnover rate constant, *k_{st}*, AlkD was at least fivefold in excess over the *K_{1/2}* for a particular labelled DNA substrate (for example, 5 µM for 7mG•C). For *K_{1/2}*

determinations, the 7mG excision assay was performed over a range of enzyme concentrations and *K_{1/2}* obtained by fitting the Michaelis–Menten plot with the equation, $k_{obs} = V_{max}[AlkD]/(K_{1/2} + [AlkD])$. We note that our *K_{1/2}* for maximal activity may differ from the *K_m* value for multiple turnover because the *K_m* can be affected by product release. Stoichiometric 7mG excision was performed in the presence of 10 µM unlabelled 25mer DNA duplex (*K_{1/2}* for this DNA was determined to be 0.9 ± 0.1 µM). Spontaneous rates of 7mG hydrolysis were determined using the sequence d(GACCACTACACC(7mG)ATTCCTTACAAC) that had been re-annealed to 100-fold excess complementary strand d(GTTGTAAGG AAT(C/T)GGTGTAGTGGTC).

POB adduct excision. DNA (catalogue no. D1501), alkaline phosphatase (P8361), esterase (E2884) micrococcal nuclease (N3755) and phosphodiesterase II (P9041) were purchased from Sigma. The tetra-deuterated standards were provided by S. S. Hecht. NNKOAc was synthesized by D. Desai.

For NNKOAc-damaged DNA, 5 ml DNA (2 mg ml⁻¹) dissolved in 100 mM sodium phosphate (pH 7.0), 1 mM EDTA and 50 mM NaCl was reacted with 1 mM NNKOAc and esterase (200 units) at 37 °C for 2 h. The reaction was diluted to 10 ml with H₂O and extracted with 10 ml CHCl₃/iso-amyl alcohol (24/1) to remove the protein and 10 ml ethyl acetate to remove any unreacted NNKOAc. The DNA was precipitated by the addition of 40 ml ethanol and washed twice with 70% ethanol. Residual amounts of ethanol were removed by rotary evaporation and the DNA was dissolved in H₂O, aliquoted and stored at -80 °C before use.

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

For HPLC-MS/MS, deuterated standards (100 fmol each of O²-POB-C-d₄ and N7-POB-G-d₄) were added to the ethanol supernatant and the solvent evaporated. The sample was dissolved in 50 µl methanol for MS analysis. The samples were analysed with a MDS/Sciex 4000 QTrap instrument with electrospray ionization (ESI) coupled to an Agilent 1100 HPLC system. Samples (20 µl) were loaded onto a column (Luna C18(2) 150 × 2 mm, 3 µm) which was eluted with 10 mM ammonium formate at 0.1 ml min⁻¹. The POB-DNA adducts, along with their deuterated standards, were monitored by selected reaction monitoring. The ion transitions were as follows N7-POB-Gua, *m/z* 299.1 [M + 1]⁺ to *m/z* 148.1 [POB]⁺; [pyridine-D₄]N7-POB-Gua, *m/z* 303.1 [M + 1]⁺ to *m/z* 152.1 ([pyridine-D₄]POB)⁺ and [Gua + H]⁺; O²-POB-Cyt, *m/z* 259.1 [M + 1]⁺ to *m/z* 148.1 [POB]⁺; [pyridine-D₄]O²-POB-Cyt, *m/z* 263.1 [M + 1]⁺ to *m/z* 152.1 ([pyridine-D₄]POB)⁺. Prior to HPLC-ESI-MS/MS analysis of the samples the MS parameters were optimized for each deuterated POB-DNA adduct standard. For analysis, the MS parameters were set as follows: curtain gas, 40 p.s.i.; ion spray voltage, 4 kV; source temperature, 650 °C; nebulizer gas (GS1), 70 p.s.i.; heater gas (GS2), 70 p.s.i.; and collision gas, 12 p.s.i. The fragmentation potentials were optimized for each ion. For 299.1 and 303.1: declustering potential (DP), 65 V; entrance potential (EP), 10 V; collision energy (CE), 20 V; collision cell exit potential (CXP), 12 V. For 259.1 and 263.1: DP, 40 V; EP, 8 V; CE, 15 V; and CXP, 6 V. The amount of each POB-DNA adduct was determined by comparing the MS peak area ratio of each adduct to its deuterated standard with a calibration curve. Calibration standards were prepared by spiking different amounts of each adduct with a constant amount of the corresponding internal standard in H₂O and then analysed by LC-MS/MS without undergoing the sample preparation procedure described above. The calibration curves were constructed by plotting concentration ratio versus MS peak area ratios of each adduct to its deuterated standard.

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

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

50. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
51. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
52. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
53. McRee, D. E. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
54. Adams, P. D. *et al.* in *Evolving Methods for Macromolecular Crystallography* (eds Read, R. J. & Sussman, J. L.) 101–109 (Springer, 2007).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
56. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
57. Lavery, R. & Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**, 63–91 (1988).
58. Asaeda, A. *et al.* Substrate specificity of human methylpurine DNA N-glycosylase. *Biochemistry* **39**, 1959–1965 (2000).
59. Jones, B. N., Quang-Dang, D. U., Oku, Y. & Gross, J. D. A kinetic assay to monitor RNA decapping under single-turnover conditions. *Methods Enzymol.* **448**, 23–40 (2008).
60. Baldwin, M. R. & O'Brien, P. J. Human AP endonuclease 1 stimulates multiple-turnover base excision by alkyladenine DNA glycosylase. *Biochemistry* **48**, 6022–6033 (2009).
61. Lyons, D. M. & O'Brien, P. J. Efficient recognition of an unpaired lesion by a DNA repair glycosylase. *J. Am. Chem. Soc.* **131**, 17742–17743 (2009).
62. Maher, R. L. & Bloom, L. B. Pre-steady-state kinetic characterization of the AP endonuclease activity of human AP endonuclease 1. *J. Biol. Chem.* **282**, 30577–30585 (2007).
63. Maher, R. L., Vallur, A. C., Feller, J. A. & Bloom, L. B. Slow base excision by human alkyladenine DNA glycosylase limits the rate of formation of AP sites and AP endonuclease 1 does not stimulate base excision. *DNA Repair (Amst.)* **6**, 71–81 (2007).
64. Maiti, A., Morgan, M. T. & Drohat, A. C. Role of two strictly conserved residues in nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase. *J. Biol. Chem.* **284**, 36680–36688 (2009).
65. Bennett, M. T. *et al.* Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **128**, 12510–12519 (2006).