Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure–function studies of an unusual 3-methyladenine DNA glycosylase II (AlkA) from *Deinococcus radiodurans*

Elin Moe,[a] David R. Hall,[b]‡ Ingar Leiros,[a] Vivi Talstad Monsen,[c] Joanna Timmins[b]§ and Sean McSweeney[b]*

[a]The Norwegian Structural Biology Centre (NorStruct), Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway, [b]The Structural Biology Group, The European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble CEDEX 9, France, and [c]Department of Cancer Research and Molecular Medicine, The Norwegian University of Science and Technology, N-7006 Trondheim, Norway

‡ Present address: Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, England.
§ Present address: Institut de Biologie Structurale J.P. Ebel, 41, rue Jules Horowitz, F-38027 Grenoble Cedex, France.

Correspondence e-mail: seanmcs@esrf.fr

Received 24 November 2011
Accepted 4 March 2012

PDB References: SeMet DR_2584, 2yg8; native DR_2584, 2yg9.

3-Methyladenine DNA glycosylase II (AlkA) is a DNA-repair enzyme that removes alkylated bases in DNA *via* the base-excision repair (BER) pathway. The enzyme belongs to the helix–hairpin–helix (HhH) superfamily of DNA glycosylases and possesses broad substrate specificity. In the genome of *Deinococcus radiodurans*, two genes encoding putative AlkA have been identified (Dr_2074 and Dr_2584). Dr_2074 is a homologue of human AlkA (MPG or AAG) and Dr_2584 is a homologue of bacterial AlkAs. Here, the three-dimensional structure of Dr_2584 (DrAlkA2) is presented and compared with the previously determined structure of *Escherichia coli* AlkA (EcAlkA). The results show that the enzyme consists of two helical-bundle domains separated by a wide DNA-binding cleft and contains an HhH motif. Overall, the protein fold is similar to the two helical-bundle domains of EcAlkA, while the third N-terminal mixed α/β domain observed in EcAlkA is absent. Substrate-specificity analyses show that DrAlkA2, like EcAlkA, is able to remove both 3-methyladenine (3meA) and 7-methylguanine (7meG) from DNA; however, the enzyme possesses no activity towards 1,N^6-ethenoadenine (εA) and hypoxanthine (Hx). In addition, it shows activity towards the AlkB dioxygenase substrates 3-methylcytosine (3meC) and 1-methyladenine (1meA). Thus, the enzyme seems to preferentially repair methylated bases with weakened N-glycosidic bonds; this is an unusual specificity for a bacterial AlkA protein and is probably dictated by a combination of the wide DNA-binding cleft and a highly accessible specificity pocket.

1. Introduction

3-Methyladenine DNA glycosylases are DNA-repair enzymes that remove alkylated bases from DNA as part of the base-excision repair (BER) pathway. Two enzymes with this repair function have been identified in *Escherichia coli*. 3-Methyladenine DNA glycosylase I (Tag) is a constitutively expressed enzyme with specificity solely for 3-methyladenine damage, while 3-methyladenine DNA glycosylase II (AlkA) is induced upon exposure to alkylating agents and possesses a broad substrate specificity (Thomas *et al.*, 1982). In addition to 3-methyladenine (3meA), AlkA from various prokaryotic and eukaryotic organisms has been shown to remove 7-methyladenine (7meA), 3-methylguanine (3meG), 7-methylguanine (7meG), 1,N^6-ethenoadenine (εA) and hypoxanthine (Hx), amongst others (Saparbaev *et al.*, 1995; Saparbaev & Laval, 1994; Thomas *et al.*, 1982).

Three-dimensional structures are known for three bacterial AlkAs: *E. coli* AlkA (EcAlkA) alone and in complex with DNA (Hollis *et al.*, 2000; Labahn *et al.*, 1996; Yamagata *et al.*, 1996), *Helicobacter pylori* AlkA (HpMagIII); Eichman *et al.*, 2003) and *Bacillus halodurans* AlkA (BhAlkA; PDB entry

2h56; Joint Center for Structural Genomics, unpublished work). EcAlkA consists of three domains. The N-terminal domain (domain I) is a mixed α/β structure with a five-stranded antiparallel β-sheet flanked by two α-helices. The second domain (domain II) is a bundle of seven α-helices (αD–αJ) and the third domain (domain III) consists of a three-helix bundle (αK–αM) with one additional α-helix (αC) in the connective segment between domains II and III. Domains II and III are separated by a deep nonpolar cleft with a conserved catalytic aspartate (Asp238) pointing into it (Labahn *et al.*, 1996). The cocrystal structure of EcAlkA with DNA revealed that this is the DNA-binding cleft, that Asp238 allows the stabilization of a carbocation intermediate in order to facilitate the cleavage of the glycosidic bond and that Leu125 in the so-called minor-groove intercalating loop (αD–αE loop) fills in the gap in the DNA resulting from the flipping out of the damaged base (Hollis *et al.*, 2000). The same study showed that Trp272 stabilizes the flipped-out DNA base by aromatic ring-stacking interactions.

HpMagIII consists of two helical domains (similar to domains II and III of EcAlkA) separated by a DNA-binding cleft (Eichman *et al.*, 2003). However, its substrate specificity is different, with a strong preference for 3meA, a property that it shares with *E. coli* Tag (EcTag; O'Rourke *et al.*, 2000). Although HpMagIII possesses a similar fold to EcTag, the specificity for 3meA is determined by aromatic ring stacking in the specificity pocket, which excludes binding of 7-methylguanine (Eichman *et al.*, 2003), rather than the lack of a catalytic aspartate as is the case for EcTag (Drohat *et al.*, 2002).

Recently, the structure of *B. halodurans* AlkA (BcAlkA) has been described at 2.5 Å resolution (PDB entry 2h56; Joint Center for Structural Genomics), but no paper describing its structure or activity has been published to date. The structure of the AlkA from the archaeal *A. fulgidus* has recently been published and this structure closely resembles the three-domain arrangement of EcAlkA but possesses broader substrate specificity (Leiros *et al.*, 2007). In addition to the classic AlkA substrates such as 3meA and 7meG, it also removes the typical substrates of the oxidative demethylase AlkB, which directly reverses 1-methyladenine (1meA) and 3-methylcytosine (3meC) lesions (Begley & Samson, 2003). *A. fulgidus* does not contain the gene encoding AlkB dioxygenase in its genome and it is suggested that the activity of *A. fulgidus* AlkA towards 1meA and 3meC represents a different repair mechanism for these forms of damage in the third domain of life.

AlkA proteins are members of the helix–hairpin–helix (HhH) superfamily of DNA glycosylases, which includes the endonuclease III (Nth) and MutY DNA glycosylases, 8-oxoguanine DNA glycosylases I and II (OggI and Ogg2) and *N*-methylpurine DNA glycosylase (MpgII; Denver *et al.*, 2003). The HhH motif consists of two α-helices connected by a hairpin loop with the signature (Leu/Phe)-Pro-Gly-(Val/Ile)-Gly. In EcAlkA the HhH motif consists of helices αI–αJ linked by a hairpin loop with the signature sequence 212-FPGIG-216. The cocrystal structure of EcAlkA with DNA has shown that this motif anchors the DNA to the protein through hydrogen bonding and metal binding (Hollis *et al.*, 2000).

An analysis of the genome (Makarova *et al.*, 2001) of the extreme radiation-resistant Gram-positive bacterium *Deinococcus radiodurans* (Mattimore & Battista, 1996) revealed that DNA glycosylases are among the expanded protein families in this organism. To date, 12 genes encoding DNA glycosylases have been identified in the genome of this organism: five uracil DNA glycosylases, three endonuclease IIIs, two AlkAs, MutY and MutM (Makarova *et al.*, 2001). The role of the DNA glycosylases in the radiation resistance of *D. radiodurans* remains unclear, but it has been shown previously that both the uracil-DNA *N*-glycosylase (UNG) and the mismatch-specific uracil-DNA glycosylase (MUG) possess structural modifications compared with other UNG and MUG proteins and that these modifications optimize the DNA-repair efficiency and repertoire of this organism (Leiros *et al.*, 2005; Moe *et al.*, 2006). Here, we present the three-dimensional structure of one of the two annotated AlkAs from *D. radiodurans* (DR_2584, hereafter called DrAlkA2). The results show that DrAlkA2 consists of two helical domains (*A* and *B*) separated by a wide DNA-binding cleft, and thus has an overall fold similar to domains II and III of EcAlkA, whilst the third N-terminal mixed α/β domain I is absent. Activity measurements reveal that DrAlkA2 possesses specificity for the typical AlkA substrates 3meA and 7meG but surprisingly not for εA and Hx. Instead, these experiments revealed that DrAlkA2 recognizes and removes the AlkB substrates 1meA and 3meC from DNA. This is the first reported bacterial AlkA protein with specificity for substrates characteristic of AlkB and we propose that this unusual specificity arises from a combination of a wide DNA-binding cleft and a highly accessible substrate-binding pocket.

2. Materials and methods

2.1. Cloning

The gene encoding 3-methyladenine DNA glycosylase II (AlkA) from *D. radiodurans* (Dr_2584) was inserted into the pDEST14 expression vector using Gateway technology following the manufacturer's instructions (Invitrogen). Two different constructs were made, one with nucleotides encoding a hexahistidine tag immediately upstream of the gene (DrAlkA2HIS) and one with both a hexahistidine tag and a TEV cleavage site (DrAlkA2HISTEV). The primers used for amplification of the gene were as follows (Sigma–Genosys): FDrAlkA2HIS, 5′-**CAT CAC CAT CAC CAT CAC** CGC GGC GGC GCC CTG CAC CAG-3′; FDrAlkA2HISTEV, 5′-CAT CAC CAT CAC CAT CAC **GAA AAC CTG TAT TTC CAG GGA GCA** CGC GGC GGC GCC CTG CAC CAG-3′; RDrAlkA2, 5′-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC **TCA** GAG CGG CGC CCC GCC CGC CTG C-3′; FDRHISTAG, 5′-G GGG ACA AGT TTG TAC AAA AAA GCA GGC TTC GAA GAT AGA ACC ATG **CAT CAC CAT CAC CAT CAC**-3′. The gene was first amplified using the primers FDrAlkA2HIS in order to engineer the N-terminal

Table 1
Data-collection statistics for the selenomethionine-derivatized and native proteins.

Values in parentheses are for the outer resolution shell.

Data set	Se *K*-edge peak	Native
Unit-cell parameters (Å, °)	a = 108.5, b = 51.3, c = 90.5, β = 99.8	a = 108.3, b = 51.4, c = 90.0, β = 99.7
Beamline	ID29, ESRF	ID14-1, ESRF
Detector	ADSC Q315r	ADSC Q210
Wavelength (Å)	0.97949	0.934
Resolution (Å)	49.75–2.00 (2.11–2.00)	42.68–1.95 (2.06–1.95)
$R_{\rm merge}$	0.100 (0.321)	0.14 (0.48)
$R_{\rm anom}$	0.073 (0.163)	—
$\langle I/\sigma(I)\rangle$	14.8 (5.1)	8.8 (3.0)
Multiplicity	6.3 (6.4)	4.1 (4.2)
Completeness (%)	100 (100)	99.8 (99.8)
Anomalous completeness (%)	99.9 (100)	—

hexahistidine tag (bold) or FDrAlkA2HISTEV in order to engineer the N-terminal hexahistidine tag and the TEV cleavage site (bold) and RDrAlkA2. The final gene products (DrAlkA2HIS and DrAlkA2HISTEV) were amplified using the FDRHISTAG (containing the *att*B1 site and nucleotides encoding the His tag) and the RDrAlkA2 primers, and used in a BP reaction along with the pDONR201 vector and in an LR reaction along with the pDEST14 vector. The sequence of the clones was confirmed by DNA sequencing (ABI 377 DNA sequencer, GE Healthcare).

2.2. Expression and purification

The *E. coli* BL21 (DE3) Codon Plus expression strain (Invitrogen) was transformed with the pDEST14 vector containing the genes encoding DrAlkA2HIS and DrAlkA2-HISTEV. Expression was induced for 16 h after the addition of 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) at 293 K in 1 l LB medium containing 100 µg ml^{-1} ampicillin and chloramphenicol. Soluble fractions of the lysed cells were loaded onto a 5 ml nickel-loaded HisTrap HP column (GE Healthcare) equilibrated with 50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl. The proteins were eluted from the column using a gradient of 0–100% 500 m*M* imidazole in equilibration buffer. The DrAlkA2HIS protein was then concentrated to approximately 10 mg ml^{-1} and stored at 277 K, whilst the DrAlkA2HISTEV protein was mixed with 1:10(*w*:*w*) TEV protease followed by incubation at 277 K overnight for cleavage of the His tag. The TEV protease was produced using the pRK793 vector (Kapust *et al.*, 2001) in *E. coli* BL21 (DE3) pLysS cells and purified using a HisTrap column according to the protocol described above. The TEV protease and the His tag were removed from the DrAlkA2 protein after cleavage by HisTrap purification using the protocol described above. Fractions of the flowthrough containing DrAlkA2 without His tag were pooled, concentrated to approximately 2.5 mg ml^{-1} and stored at 277 K.

The preparation of selenomethionine-substituted protein followed the method of Van Duyne *et al.* (1993). *E. coli* BL21 (DE3) Codon Plus cells (Invitrogen) were grown in M9 medium until they reached an absorbance at 600 nm (A_{600}) of 0.4. At this point, 100 mg each of L-lysine, L-phenylalanine and L-threonine, 50 mg each of L-isoleucine and L-leucine and 60 mg DL-selenomethionine (SeMet) were added to 1 l culture and the cells were grown at 310 K for 15 min before inducing expression with 0.5 m*M* IPTG overnight at 293 K. The cells were harvested and disrupted and the protein was purified, concentrated and stored as described above.

2.3. Determination of protein concentration and purity

Protein concentrations were determined with Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-Rad, Hercules, California, USA) based on the Bradford dye-binding procedure (Bradford, 1976) using bovine serum albumin (BSA) as a standard. Protein purity was determined using SDS–PAGE (NuPAGE) with 4–12% bis-Tris gels run in MES buffer. The gel was stained using Simply Blue Safe Stain according to the protocol supplied by the manufacturer (Invitrogen).

2.4. Crystallization and data collection

Crystallization of DrAlkA2 with the hexahistidine tag was performed by the hanging-drop method. Crystals were grown by mixing 1 µl drops of 10 mg ml^{-1} protein solution with a solution consisting of 1 *M* $LiCl_2$, 0.1 *M* MES pH 6.0, 10% PEG 6000 (Grid Screen PEG/LiCl solution B3, Hampton Research, California, USA). The drops were equilibrated at 291 K and platelet crystals appeared after 1 d. Crystals suitable for data collection were obtained after more than one week. The crystals were transferred into a cryosolution consisting of 1 *M* $LiCl_2$, 0.1 *M* MES pH 6.0, 10% PEG 6000 and 20% glycerol prior to direct flash-cooling using a nitrogen cold stream (Oxford Instruments) operating at a temperature of 100 K.

High-quality data were collected to 1.95 Å resolution from native protein crystals on beamline ID14-1 at the ESRF, Grenoble, but structure determination by molecular replacement using EcAlkA as a search model was unsuccessful. A crystal of SeMet-incorporated protein was subjected to a fluorescence scan around the Se *K* edge on beamline ID29 at the ESRF, Grenoble and the data were analysed with *CHOOCH* (Evans & Pettifer, 2001) to give a peak wavelength of 0.97949 Å. A single data set was then collected to a resolution of 2.0 Å from this crystal (Table 1) at the Se-peak wavelength. The crystal belonged to space group *C*2 and the data were indexed and integrated with *MOSFLM* (Leslie, 1992) and were reduced with *SCALA* (Winn *et al.*, 2011). The Matthews coefficient (Matthews, 1968) was determined to be 2.45 Å^3 Da^{-1} (corresponding to a solvent content of 49.8%) with two molecules in the asymmetric unit. Data from the native crystals are also included in Table 1.

2.5. Structure determination and refinement

The single-wavelength anomalous dispersion data were phased using *SHARP* (de La Fortelle & Bricogne, 1997), which found and refined four out of six possible Se sites for a dimer in the asymmetric unit (phasing power = 1.121, $R_{\rm Cullis}$ = 0.804). After solvent flattening with twofold averaging in *DM*, *ARP*/*wARP* (Perrakis *et al.*, 1999) built 398 residues (390 with side chain assigned) out of 413 in the final model.

Table 2
Statistics for the final models.

	SeMet DrAlkA2	Native DrAlkA2
Resolution (Å)	49.75–2.00	42.68–1.95
Protein residues	413	413
Protein atoms	3085	3122
Waters	613	612
Ions	4 Cl^-	2 Ni^{2+}, 4 Cl^-
Ligands	1 MES, 1 glycerol	1 glycerol
R_{work} (%)/No. of observations	15.8/30038	17.0/32168
R_{free} (%)/No. of observations	20.3/1693	23.1/1802
Average isotropic thermal parameters (Å^2)		
Molecule *A* overall	10.9	11.8
Molecule *B* overall	10.0	11.5
Main chain	9.4	11.0
Side chain	11.6	12.4
Solvent	23.4	23.0
Residues in dual conformation	Glu185*A*, Gln219*A*	Asp28*A*, Arg117*A*, Glu185*A*, Arg199*A*, Gln219*A*, Arg117*B*, Arg194*B*
Residues truncated to alanine	His135*B*	His135*B*
R.m.s.d. bond lengths (Å)	0.016	0.016
R.m.s.d. angles (°)	1.43	1.48
PDB code	2yg8	2yg9

Manual model building using *Coot* (Emsley & Cowtan, 2004) was interspersed with crystallographic refinement using *REFMAC*5 (Murshudov *et al.*, 2011) and gave rise to a final model with an *R* factor of 15.8% (R_{free} = 20.3%). Details of the model and refinement can be found in Table 2.

The crystal structure of the native protein was later determined by molecular replacement using the SeMet structure as a search model (Table 2). All the structural illustrations were generated by using *PyMOL*.

2.6. Specificity measurements

The activity and specificity measurements were performed on DrAlkA2 without the hexahistidine tag using either [^{3}H]-methylated DNA substrates containing a variety of alkylated nucleotides or fluorescence-labelled dsDNA oligonucleotides with specific methylation damage. The [^{3}H]-methylated DNA substrate was prepared as described previously (Bjelland *et al.*, 1993) using *N*-[^{3}H]-methyl-*N*-nitrosourea (MNU; 0.5 mCi; GE Healthcare) and 200 µg poly(dA–dT) and poly(dG–dC) (Sigma–Aldrich). The assay was performed by incubating increasing amounts of DrAlkA2 (0, 0.02, 0.27 and 2.63 µ*M* final concentration) and 3-methyladenine DNA glycosylase I (Tag; 0, 0.22, 2.12 and 4.43 µ*M* final concentration) from *Vibrio cholerae* (from our laboratory, unpublished work) and 5 µl substrate (∼5000 cpm in total) in a buffer consisting of 10 m*M* HEPES–KOH pH 7.4, 100 m*M* KCl, 1 m*M* EDTA, 0.1 m*M* DTT in a total volume of 20 µl at 310 K for 10 min. The nucleic acids were precipitated by addition of NaCl to 1 *M* and of two volumes of ice-cold 96% ethanol and incubation at 153 K for 10 min. The released [^{3}H]-methylated nucleotides were separated from the nucleic acids by centrifugation at 13 000 rev min^{-1} in a microcentrifuge for 10 min at 277 K. The amount of released methylated nucleotides was quantified by subjecting 50 µl of the supernatant to scintillation counting.

The specificity of DrAlkA2 and mouse 3-methyladenine DNA glycosylase (mAag; Nordic Biosite AB) towards specific methylation damages was measured using 30 bp dsDNA oligonucleotide substrates containing εA·T and Hx·T mismatches labelled with HEX in the 5′ end (Sigma–Aldrich) and a 49 bp dsDNA oligonucleotide containing 3meC·G and 1meA·T labelled with FAM in the 5′ end (MedProbe). The dsDNA oligonucleotide labelled with HEX had the sequence 5′-ATATACCGCGG*X*GGCCGATCAAGCTTATT-3′, where *X* was either εA or Hx and the complementary strand had T as a partner to the *X*. The oligonucleotide labelled with FAM had the sequence 5′-TAGACATTGCCATTCTCGATAGG**AT**C-CGGTCAAACCTAGACGAATTCCG-3′, where the A and C in bold were located in the cleavage site of the *Dpn*II restriction enzyme and were either 1meA or 3meC, with T or G as a partner in the complementary strand (Ringvoll *et al.*, 2006). Assays were performed by mixing increasing amounts of enzyme (0–100 pmol) with 1 pmol DNA in 20 m*M* Tris–HCl pH 8.0, 0.1 mg ml^{-1} BSA, 1 m*M* EDTA, 1 m*M* DTT in a 20 µl reaction mixture for 1 h at 310 K. The reaction was stopped and the resulting abasic site was cleaved off by adding NaOH to a final concentration of 100 m*M* and incubating at 372 K for 10 min; the mixture was then analysed on a denaturing 15% TBE–urea PAGE. The products were visualized using a Typhoon scanner (GE Healthcare).

2.7. HPLC-MS/MS analysis

DNA from the DNA glycosylase assays using the FAM-labelled substrates (containing 3meC and 1meA) was hydrolyzed to nucleosides by nuclease P1, snake-venom phosphodiesterase and alkaline phosphatase (Crain, 1990). This was followed by the addition of three volumes of methanol and centrifugation at 16 000*g* for 30 min. The supernatants were dried under vacuum and the resulting residues were dissolved in 50 µl 0.1% formic acid/5% methanol (*v*/*v*) in water for 3meC and 1meA analysis by HPLC-MS/MS. A fraction of each sample was diluted to 1:25 000 for quantitation of the corresponding deoxycytosine (dC) and deoxyadenosine (dA). The HPLC apparatus consisted of an HP1100 system (Agilent Technologies) with binary pump, autosampler and solvent degasser. Gradient chromatographic separation of 3meC and 1meA analytes was performed on a Zorbax SB-C18 2.1 × 150 mm i.d. (3.5 µm) reverse-phase column equipped with an Eclipse XDB-C8 2.1 × 12.5 mm i.d. (5 µm) guard column (both from Agilent Technologies) at ambient temperature and a flow rate of 0.2 ml min^{-1}. The mobile phase consisted of *A* (0.1% formic acid in water) and *B* (0.1% formic acid in methanol) in a gradient elution starting with 95% *A*/5% *B* for 3.5 min, followed by a 7 min linear gradient of 5–50% *B* and finally 25 min re-equilibration with the initial mobile-phase conditions. Chromatographic analysis of dC and dA was performed under isocratic conditions on a Supelcosil LC-C18-S 4.6 × 150 mm i.d. (5 µm) reverse-phase column (Sigma–Aldrich). This was maintained at ambient temperature with a mobile phase of 80% *A*/20% *B* at a flow rate of 1 ml min^{-1}. Online

mass-spectrometric detection was performed using an Applied Biosystems/MDS Sciex 4000 hybrid triple-quadrupole/linear ion-trap mass spectrometer (Applied Biosystems Sciex) with TurboIonSpray probe operating in positive electrospray ionization mode. The analytes were monitored in the multiple-reaction monitoring mode using the mass transitions 242–126 for 1me(dC) and 266–150 for 1me(dA).

3. Results and discussion

3.1. The overall structure of DrAlkA2

From the experimental data set, two molecules related by noncrystallographic twofold symmetry were modelled in the asymmetric unit and the refined model of each molecule showed that DrAlkA2 comprises a single polypeptide chain, in which amino-acid residues 19–225 of chain *A* and 20–225 of chain *B* were defined in electron density. No electron density was observed for the N-terminal residues. The overall root-mean-square (r.m.s.) deviation between molecules *A* and *B* was 0.339 Å and all further discussion will be based on monomer *A* of the model. The protein consists of two helical domains separated by a wide cleft (Fig. 1*a*). One domain (*A*) consists of two N-terminal helices αA and αB and three C-terminal helices αJ–αL, while the second domain (*B*) consists of seven helices αC–αI and includes the HhH domain (αH–hairpin loop–αI) and the minor-groove intercalating loop between helices αC and αD.

The crystal structure resembles domains II and III of the well studied *E. coli* AlkA (PDB entries 1mpg and 1diz; Hollis *et al.*, 2000; Labahn *et al.*, 1996), with an r.m.s. deviation of 2.5 Å for 163 aligned C^{α} atoms of the two domains (Fig. 1*b*), while domain I of EcAlkA is absent. DrAlkA2 can also be structurally aligned with the crystal structures of *B. halodurans* AlkA (PDB entry 2h56; Joint Center for Structural Genomics, unpublished work), *H. pylori* MagIII (PDB entry 1pu6; Eichman *et al.*, 2003), *A. fulgidus* AlkA (PDB entry 2jhn; Leiros *et al.*, 2007), *E. coli* 3-methyladenine DNA glycosylase I (Tag; PDB entry 1lmz; Drohat *et al.*, 2002), *E. coli* endonuclease III (PDB entry 2abk; Thayer *et al.*, 1995), *E. coli* MutY (PDB entry 1muy; Guan *et al.*, 1998) and *E. coli* 8-oxoguanine DNA glycosylase (MutM; PDB entry 1ee8; Sugahara *et al.*, 2000). DrAlkA2 can thus be described as a member of the HhH family of DNA glycosylases. In this work, we have focused on a comparative study of the DrAlkA2 structure with the EcAlkA and BhAlkA structures.

In the DrAlkA2 structure two ions were modelled into density as chloride ions. One chloride is coordinated by the main-chain amide N atoms of Gly155 and Thr158 close to the hairpin of the HhH motif. The other chloride ion is coordinated by the main-chain amide N atom of Ser112 and the amide group of Lys115 in the loop between αF and αG. They are believed to be an artefact of the crystallization conditions since they are not bound to any residues shown to be of significance for the biological activity of the protein.

The most striking structural difference between DrAlkA2 and EcAlkA is the absence of domain I in DrAlkA2 (Fig. 1). A sequence alignment between DrAlkA2 and EcAlkA shows that there is only 20% sequence identity between the two enzymes (Fig. 2), and there are pronounced differences in the length and amino-acid content of the N-terminal sequences of the enzymes which explain the absence of an N-terminal domain in DrAlkA2. It is not possible to observe the first 18 residues in the N-terminal domain in the electron-density map

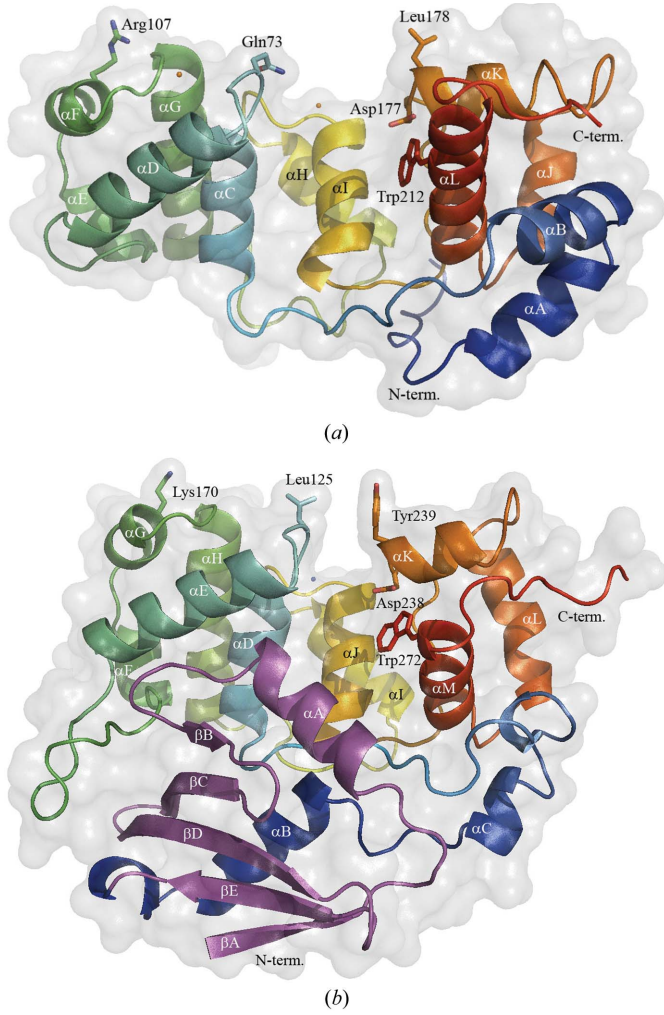

Figure 1
Overall structure and visualization of the protein surfaces of DrAlkA2 and EcAlkA (PDB entry 1diz without DNA). (*a*) DrAlkA2 consists of a two-domain α-helix structure separated by a wide DNA-binding cleft. Domain *A* consists of αA, αB and αJ–αL (coloured blue and orange/red). Domain *B* consists of αC–αG (coloured cyan/green). The two domains are separated by the DNA-binding cleft and the helix–hairpin–helix (HhH) motif (αH–αI; coloured yellow). Residues involved in catalysis (Asp177), base stacking (Trp212), DNA intercalation and stabilization (Gln73 and Leu178) and DNA interaction (Arg107) are indicated. Two chloride ions are indicated in orange. (*b*) EcAlkA consists of three domains. Domain I is a mixed α/β structure with a five-stranded antiparallel β-sheet flanked by two α-helices (αA, βA–βE and αB; coloured magenta). Domain II consists of a globular bundle of five α-helices (αD–αH) coloured cyan/green and domain III consists of a three-α-helix bundle (αK–αM) coloured red/orange. In addition, α-helix C (αC) is found in the connective segment between domains II and III (αI–αJ; coloured yellow), which are also separated by the DNA-binding cleft and the HhH motif. Residues involved in catalysis (Asp238), base stacking (Trp272), DNA intercalation and stabilization (Leu125 and Tyr239) and DNA interaction (Lys170) are indicated. One metal ion (sodium) is indicated in blue.

of DrAlkA2, which indicates that this part of the structure is highly flexible. The sequence identity between DrAlkA2 and BhAlkA is 41%, and from the sequence alignment (Fig. 2) it can be observed that the N-terminus of *B. halodurans* AlkA (BhAlkA) is short (as in DrAlkA2). It appears that domain I is also absent in this structure, which consists only of two helical domains (*A* and *B*) as in DrAlkA2.

3.2. The DNA-binding cleft

Another obvious difference between the DrAlkA2 and EcAlkA structures is the wider DNA-binding cleft of DrAlkA2 (Figs. 1*a* and 1*b*). The distance between the C^{α} main-chain atoms of Leu178 and Gln73 at the entrance of the cleft of DrAlkA2 is 15.3 Å, compared with 8.1 Å between the corresponding residues in EcAlkA (Leu125 and Tyr239). The wide cleft is possibly permitted by the absence of domain I, which interacts with domain II, in EcAlkA and places constraints upon the structure. In EcAlkA, interactions between the loop linking αA and βA in domain I of EcAlkA and helix αE in domain II and interactions between βC–βD and the loop between helices αE and αF seem to both push and constrain the αD–αE helices into a fixed position (Hollis *et al.*, 2000). Additional interactions can be observed between helices αC (domain I) and αD (domain II) and the helices in domain III of the EcAlkA structure.

From the deposited structure of BhAlkA it is observed that there are three molecules in the asymmetric unit. A closer inspection of these molecules reveals that the distance between the cleft-edge residues (Gln51 and Val158) is 15 Å in molecule *A* and *B*, whilst in molecule *C* it is 7 Å (Figs. 3*a* and 3*b*). There is no indication that the different domain conformations observed are caused by crystal-packing interactions. It appears that BhAlkA is a flexible molecule and several conformations have been observed in the crystal forms. The high level of identity between DrAlkA2 and BhAlkA both at the structure and the sequence level (41%) suggests that DrAlkA2 may also possess such flexibility. These observations suggest that both DrAlkA and BhAlkA undergo structural changes upon substrate binding, which is also observed in

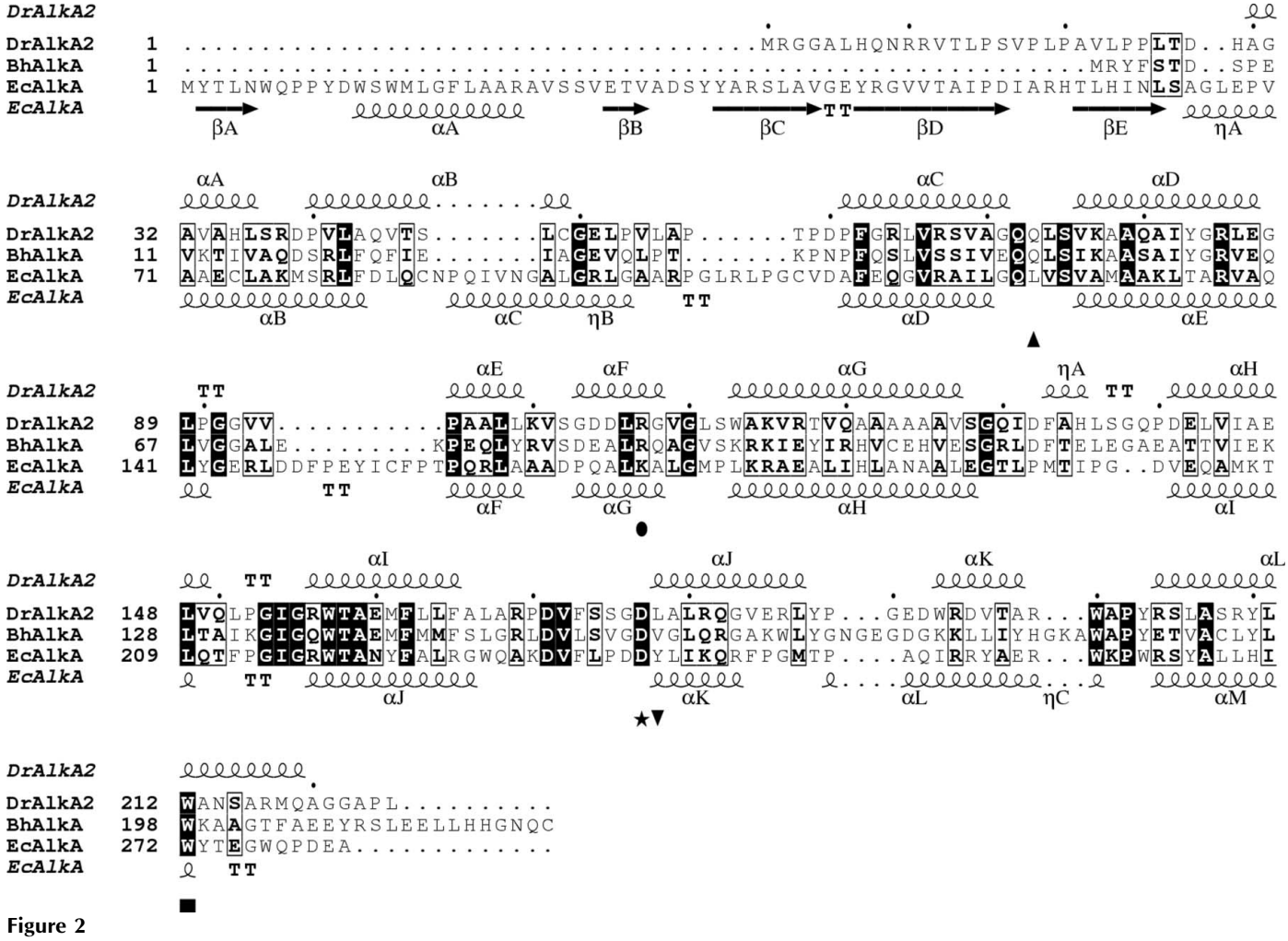

Figure 2
Sequence alignment of DrAlkA2, BhAlkA and EcAlkA. The secondary-structure elements of DrAlkA2 are indicated above the alignment and those of EcAlkA are indicated below. Identical residues for all sequences are indicated as white letters on a black background, while equivalent residues are shown as black letters on a white background. The catalytic residue (Asp177 in DrAlkA2) is marked with a star, the residues at the entrance of the specificity pocket (Leu178 and Gln73 in DrAlkA2) are marked with triangles, the substrate-stabilization residue (Trp212 in DrAlkA2) is marked with a square and the suggested DNA-interacting residue (Arg107) is marked with a circle. The HhH motif is identified as residues 140–165 in DrAlkA2.

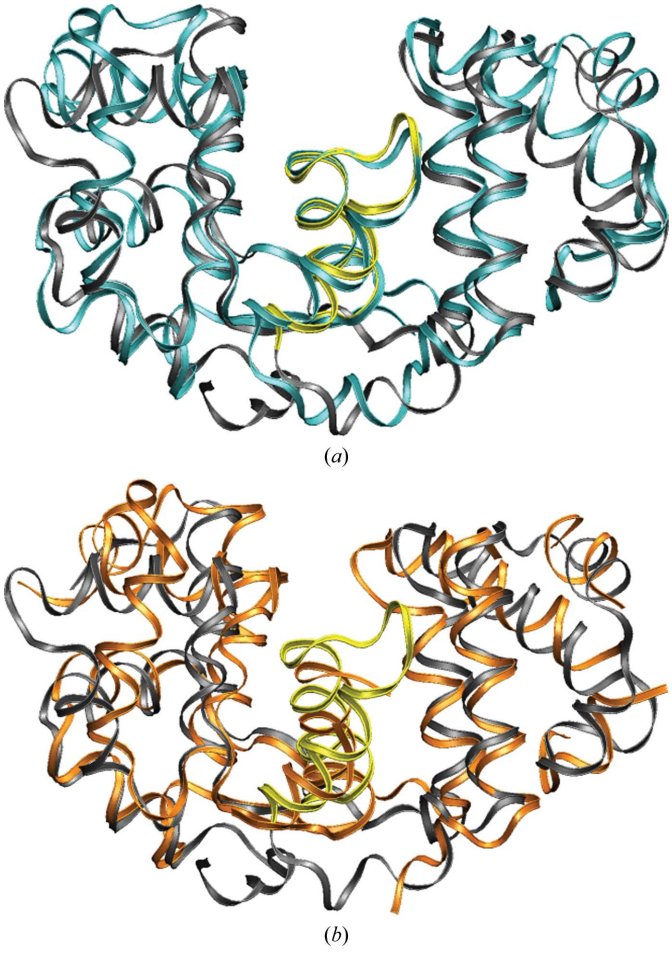

Figure 3
Structural comparison of DrAlkA2 with open and closed structures of BhAlkA. (*a*) Overlay of DrAlkA2 (grey) and molecule *A* (open structure) of BhAlkA (cyan). (*b*) Overlay of DrAlkA2 (grey) and molecule *C* (closed structure) of BhAlkA (orange). The HhH motif is coloured yellow for DrAlkA2.

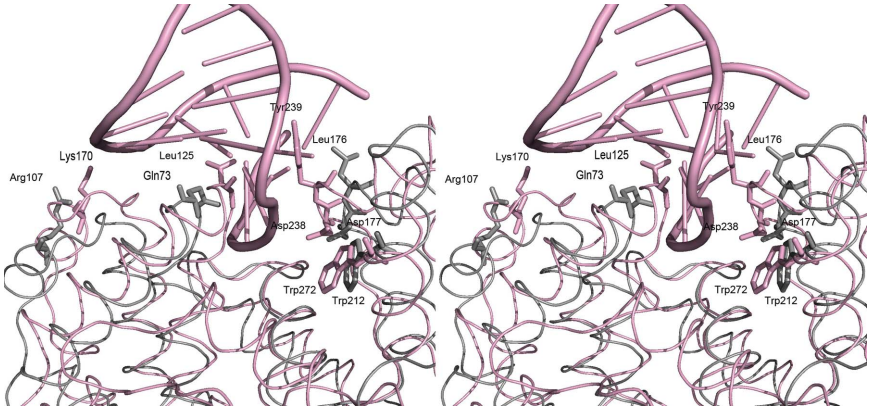

Figure 4
Stereoview of the positioning of important residues in the substrate-binding pockets and the DNA-interacting parts of DrAlkA2 (grey) and EcAlkA in complex with DNA (PDB entry 1diz, pink): the catalytic residues (Asp177 and Asp238), the suggested base-stacking residues (Trp212 and Trp272), the DNA-intercalating residues (Gln73 and Leu125), Tyr239 and Leu178 which are located at the entrance of the DNA-binding cleft and the DNA-interacting residues (Arg107 and Lys170).

other DNA glycosylases, for example uracil-DNA *N*-glycosylase and MutM (Koval *et al.*, 2004; Parikh *et al.*, 1998).

3.3. The catalytic and base-stabilization residues are conserved

The residues that are involved in catalysis and stabilization of the flipped-out damaged base are conserved between DrAlkA2 and EcAlkA. Sequence alignment of DrAlkA2 and EcAlkA shows that the catalytic Asp residue is conserved and is found in position 177 in DrAlkA2 (Fig. 2). The suggested stacking Trp residue is conserved and is found in position 212 in DrAlkA2. A closer examination of the substrate-binding pocket shows that even though the DNA-binding cleft is wider for DrAlkA2 than for EcAlkA, the side chains of both Asp177 and Trp212 are found in similar positions in EcAlkA (Asp238 and Trp272) and should thus be able to serve the same functions as in EcAlkA (Fig. 4).

3.4. DNA interaction and stabilization

The structure of EcAlkA in complex with DNA shows that a specific residue in the protein (Leu125) intercalates with the DNA and displaces the base targeted for removal (Hollis *et al.*, 2000). The corresponding residue in DrAlkA2 is Gln73. Comparison of the EcAlkA–DNA complex with a model of DrAlkA2 in complex with DNA shows that Gln73 in DrAlkA2 is in an equivalent position to Leu125 in EcAlkA (Fig. 4) and should thus be able to perform a similar action. In the DrAlkA2 structure, however, the side chain of Gln73 does not seem to point in the direction of the DNA modelled in the binding site. This orientation was also observed for Leu125 in the structure of EcAlkA without DNA, suggesting that this residue may move upon DNA binding. Glutamine residues serving as a DNA-intercalating residue have been found in the same position as Gln73 in DrAlkA2 in a number of protein–DNA complexes of other members of the HhH DNA glycosylase family [*e.g.* endonuclease III (Thayer *et al.*, 1995) and *E. coli* MutY (Guan *et al.*, 1998)], supporting its putative role in DNA stabilization in DrAlkA2 during repair of methylated DNA bases.

Lys170 is the only residue in EcAlkA which interacts with phosphates in DNA. In DrAlkA2 a positively charged residue (Arg107) is also found in this position (Figs. 2 and 4) and may therefore be involved in similar ionic interactions with the negatively charged backbone of the DNA.

The HhH signature sequence in DrAlkA2 is 151-LPGIG-156, while in EcAlkA it is 212-FPGIG-216, both of which correspond to the consensus HhH motif (Labahn *et al.*, 1996). In *E. coli* AlkA the HhH motif is found to anchor the DNA to the protein through hydrogen bonds from the main-chain amide N atoms of Gly214, Gly216 and

Thr219 and a metal-mediated interaction coordinated by the carbonyl O atoms of Gln210, Phe212 and Ile215 in the hairpin of the HhH motif (Hollis *et al.*, 2000). A superpositioning of the structure of the protein–DNA complex of EcAlkA and DrAlkA2 shows that the HhH motifs of the two structures are well aligned (Fig. 4) and suggest that the main-chain amide N atoms of Gly153, Gly155 and Thr158 in drAlkA2 (corresponding to Gly214, Gly216 and Thr219 in EcAlkA) are in a position to form hydrogen bonds to the DNA. In addition, metal-mediated interactions between the carbonyl O atoms of Val149, Leu151 and Ile154 in the HhH hairpin of DrAlkA2 (corresponding to Gly214, Gly216 and Thr219 in EcAlkA) and DNA would also be possible in order to anchor the DNA to the HhH motif.

3.5. Substrate specificity

The open accessible substrate-binding pocket of DrAlkA2 may affect the substrate specificity of DrAlkA2 compared with other AlkAs and several substrates were tested to identify the specificity of DrAlkA2. One of the substrates used was poly-dA·dT or poly-dG·dC DNA treated with *N*-[^{3}H]-methyl-*N*-nitrosourea (MNU). Treatment of high-molecular-weight oligomeric dsDNA with MNU is believed to give the following methylation damage: 67% 7meG, 9% 3meA, 6.3% O^6-G, 1.2% 7meA, 1.3% 1meA, 0.8% 3meG and 0.6% 3meC (Beranek *et al.*, 1980). Treating poly-dA·dT and poly-dG·dC DNA with MNU allowed us to identify specificity for either 3meA (using the dA·dT substrate) or 7meG (using the dG·dC substrate).

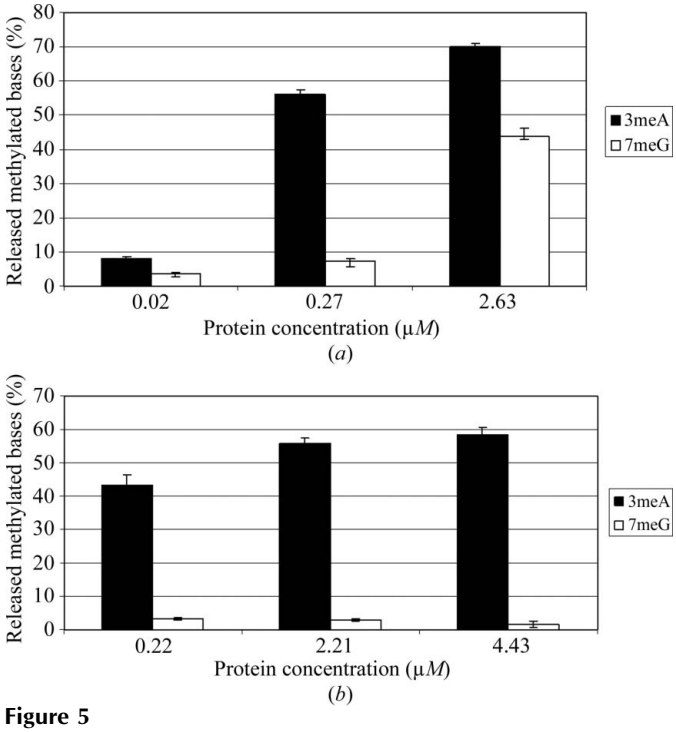

Figure 5
Specificity of three different concentrations of DrAlkA2 (*a*) and VcTag (*b*) towards MNU-treated poly-dA·dT (3meA) and poly-dG·dC (7meG) substrates. The activity is presented as the percentage of released methylated bases relative to the maximum cpm value (5000 cpm). DrAlkA2 possesses specificity for both 3meA and 7meG substrates, while VcTag only possesses specificity towards 3meA substrate, as expected.

3-Methyladenine DNA glycosylase I from *V. cholerae* (VcTag) was used as a control as it possess specificity only for 3meA in the dA·dT substrate. The results showed that DrAlkA2 possesses activity towards both 3meA and 7meG substrates, while VcTag possesses activity only for the dA·dT substrate, as expected (Fig. 5).

Both EcAlkA and human Aag have been reported to possess specificity for 1,N^6-ethenoadenine (εA) and hypoxanthine (Hx) (Saparbaev *et al.*, 1995; Saparbaev & Laval, 1994). In order to test whether DrAlkA2 also recognizes these types of DNA damage, specificity assays were performed using fluorescence-labelled double-stranded DNA (dsDNA) substrates containing either εA or Hx. The results showed that

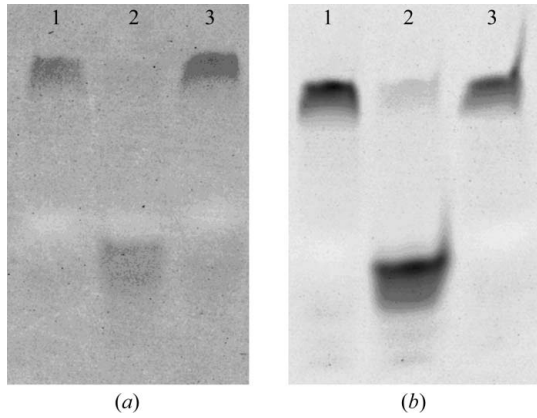

Figure 6
15% TBE–urea gel substrate-specificity analysis towards εA·T and Hx·T substrates. (*a*) Lane 1, εA·T; lane 2, εA·T and mAag; lane 3, εA·T and DrAlkA2. (*b*) Lane 1, Hx·T; lane 2, Hx·T and mAag; lane 3, Hx·T and DrAlkA2.

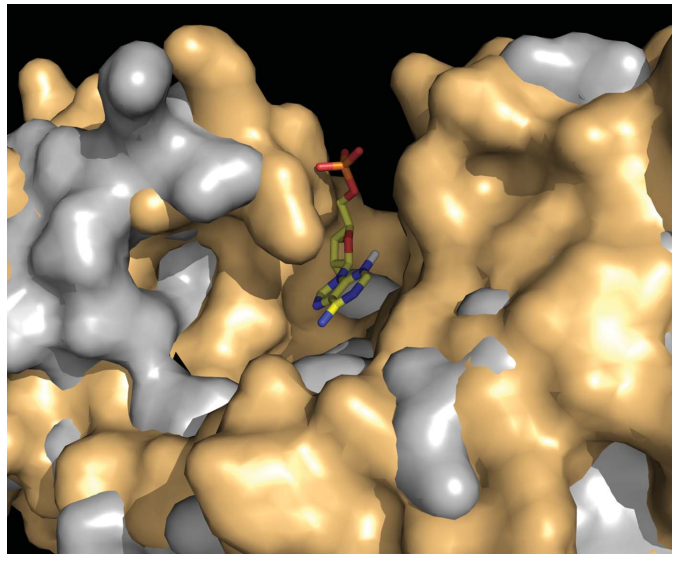

Figure 7
Illustration of the binding of 3-methyladenine in the substrate-binding pockets of DrAlkA2 and the closed conformation of BhAlkA. The surface of DrAlkA2 is visualized in grey, while the surface of the closed conformation of BhAlkA2 is shown in orange. The 3-methyladenosine molecule is modelled into the active site based on a superpositioning of DrAlkA2, the closed conformation of BhAlkA and EcAlkA in complex with DNA (PDB entry 1diz).

DrAlkA2 possesses no activity against the εA and Hx substrates (Fig. 6). Analysis of the DrAlkA2 structure reveals that the substrate-binding pocket is large enough to accommodate these bases (Fig. 7) and there are no structural hindrances that may block their placement in the pocket. However, it has been shown that some AlkA proteins more readily excise alkylated bases with unstable N-glycosidic bonds. Both εA and Hx are neutral alkyl adducts and have more stable N-glycosidic bonds than 3meA and 7meG, and probably require some stabilizing interactions in order to be cleaved off. A closer look into the structure of EcAlkA shows that an N-terminal arginine residue (Arg22) in αA is indeed in a position to stabilize the neutral εA and Hx and thus contribute to the catalysis of these bases. In the recently published structure of AlkA from the archaeon *A. fulgidus* (AfAlkA) a similar stabilizing interaction with neutral substrates by an arginine residue (Arg286) has also been illustrated. In DrAlkA2 there are no positively charged residues close to the catalytic pocket which can contribute to similar stabilizing interactions, which explains why it does not have any capacity for the removal of εA and Hx.

The recently reported structure of AfAlkA also shows that it possesses specificity for the AlkB dioxygenase substrates 1meA and 3meC (Leiros *et al.*, 2007). *A. fulgidus* does not contain the gene encoding AlkB dioxygenase in its genome and neither does *D. radiodurans*; thus, the capacity of DrAlkA2 to repair these methylated bases was investigated. An assay using fluorescence-labelled dsDNA substrates containing 1meA and 3meC and different enzyme concentrations was performed and the resulting product formation was analyzed both by denaturating gel and HPLC/MS/MS. The results of the gel analysis are illustrated in Fig. 8 and show that DrAlkA2 is able to convert substrate to product and that product formation increases with increasing enzyme concentration.

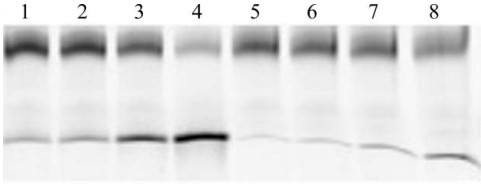

Figure 8
15% TBE–urea gel substrate-specificity analysis of DrAlkA2 using 3meC and 1meA substrates. Lane 1, 3meC·G; lane 2, 3meC·G and DrAlkA2 (1 pmol); lane 3, 3meC·G and DrAlkA2 (10 pmol); lane 4, 3meC·G and DrAlkA2 (100 pmol); lane 5, 1meA·T; lane 6, 1meA·T and DrAlkA2 (1 pmol); lane 7, 1meA·T and DrAlkA2 (10 pmol); lane 8, 1meA·T and DrAlkA2 (100 pmol).

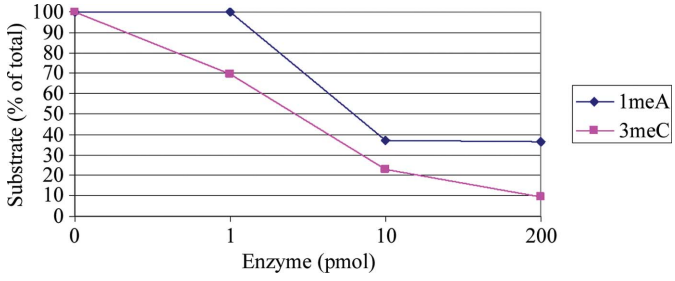

Figure 9
HPLC/MS/MS analysis of the substrate specificity of DrAlkA2 using 3meC and 1meA substrates. Only 10% of the 3meC bases were left in the substrate after incubation at 310 K for 60 min; thus, 90% of the 3meC bases have been removed from the substrate.

The results of the gel analysis were confirmed by HPLC/MS/MS analysis, which showed that only 10% of the 3meC bases were left in the substrate after incubation at 310 K for 60 min using 200 pmol enzyme; thus, 90% of the 3meC bases have been removed from the substrate (Fig. 9). The reduction of 1meA is lower and approximately 35% of the bases were left in the substrate after incubation with 200 pmol enzyme; thus, 65% of the bases were released from the substrate. The results correlate well with the results of the gel analysis and show that the amount of methylated base released from the substrate increases with increasing enzyme concentration.

The substrate-specificity analysis thus shows that DrAlkA2 possesses specificity towards both 3meA and 7meG like other AlkA proteins and can therefore be classified as a 3-methyladenine DNA glycosylase II. Surprisingly, the enzyme does not possess specificity towards the εA and Hx substrates and in this sense has narrower substrate specificity than EcAlkA. On the other hand, DrAlkA2 is able to remove the AlkB dioxygenase substrates 3meC and 1meA from DNA, a specificity that has never been observed previously among bacterial and mammalian AlkAs but that has been observed for an archaeal AlkA (AfAlkA; Leiros *et al.*, 2007).

There is no gene encoding AlkB in *D. radiodurans* and the specificity of DrAlkA2 for 3meC and 1meA has probably evolved in order to repair these damages in the organism. The substrate-binding pockets of DrAlkA2 and BhAlkA2 (Fig. 7) shows that there are no steric or other structural hindrances in the pocket of DrAlkA2 that would interfere with the binding of 3meC and 1meA.

N1 of adenine and N3 of cytosine are shielded from alkylation in dsDNA owing to their involvement in Watson–Crick base pairing; thus, 1meA and 3meC are preferentially introduced into single-stranded DNA (ssDNA) by methylating agents. It has been shown that both bacterial AlkB and one of the human AlkB homologues (hABH3) have a preference for the repair of 1meA and 3meC in ssDNA. However, it has also been shown that human AlkB homologue 2 (hABH2) possesses specificity for the repair of 1meA and 3meC in dsDNA (Falnes *et al.*, 2004). It has further been shown that hABH2 locates to the replication foci, which suggests that it removes 1meA and 3meC from dsDNA close to the replication forks (Beranek, 1990). In this study we have used dsDNA substrates with 1meA and 3meC, and in future studies we will try to identify the cellular location of the enzyme and its expression level during normal and stress-related conditions in order to investigate the biological relevance of this enzyme for the radiation resistance of *D. radiodurans*.

4. Concluding remarks

Here, we present the three-dimensional structure and a substrate-specificity analysis of one of the 3-methyladenine

DNA glycosylases II from *D. radiodurans* (Dr_2584). The results show that the enzyme has an unusual AlkA structure and consists of two helical domains separated by a wide DNA-binding cleft and possesses an HhH motif, which confirms that it belongs to the HhH family of DNA glycosylases. Substrate-specificity analyses further show that it possesses specificity for the classical AlkA substrates 3meA and 7meG and, more surprisingly, for the AlkB dioxygenase substrates 1meA and 3meC, but not for εA or Hx. 3meA, 7meG, 3meC and 1meA all possess delocalized positive charges and weakened N-glycosidic bonds and are thus more prone to nucleophilic attack and catalysis compared with the neutral εA and Hx substrates. The wide DNA-binding cleft and the accessible substrate-binding pocket do not offer any residues (such as Arg22 in EcAlkA) for the stabilization of intermediate reaction products, which is necessary to ensure the efficient removal of the εA and Hx bases. This further suggests that the role of the N-terminal domain as observed in EcAlkA is to determine/modify the substrate specificity of the various AlkA proteins studied to date.

This is, to our knowledge, the first report of a bacterial alkylpurine DNA glycosylase which shows specificity towards the AlkB substrates and we suggest that this specificity may have developed as a result of the absence of a gene encoding an AlkB homologue in the genome of *D. radiodurans*. The same substrate specificity has recently been discovered for AlkA from the archaeon *A. fulgidus*, which also lacks a gene encoding AlkB in its genome.

DrAlkA2 is therefore another example of a *D. radiodurans* DNA glycosylase enzyme with an unusual activity that may contribute to improving the DNA-repair repertoire of the organism and its extreme radiation resistance.

This work was financed by the Functional Genomics programme (FUGE) of the Research Council of Norway (RCN) and the European Synchrotron Radiation Facility (ESRF) in Grenoble. The work was performed in the CIBB laboratory and on the beamlines of the Structural Biology Group at the ESRF and in the laboratory of the Molecular Biology Research Group at the Department of Cancer Research and Molecular Medicine at the Norwegian University of Science and Technology (NTNU). Professor Hans E. Krokan, Cathrine Vågbø and Per Arne Aas at NTNU are especially acknowledged for their help with the substrate-specificity experiments.

References

Begley, T. J. & Samson, L. D. (2003). *Trends Biochem. Sci.* **28**, 2–5.
Beranek, D. T. (1990). *Mutat. Res.* **231**, 11–30.
Beranek, D. T., Weis, C. C. & Swenson, D. H. (1980). *Carcinogenesis*, **1**, 595–606.
Bjelland, S., Bjørås, M. & Seeberg, E. (1993). *Nucleic Acids Res.* **21**, 2045–2049.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Crain, P. F. (1990). *Methods Enzymol.* **193**, 782–790.
Denver, D. R., Swenson, S. L. & Lynch, M. (2003). *Mol. Biol. Evol.* **20**, 1603–1611.
Drohat, A. C., Kwon, K., Krosky, D. J. & Stivers, J. T. (2002). *Nature Struct. Biol.* **9**, 659–664.
Eichman, B. F., O'Rourke, E. J., Radicella, J. P. & Ellenberger, T. (2003). *EMBO J.* **22**, 4898–4909.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
Falnes, P. Ø., Bjørås, M., Aas, P. A., Sundheim, O. & Seeberg, E. (2004). *Nucleic Acids Res.* **32**, 3456–3461.
Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H., Lloyd, S. & Tainer, J. A. (1998). *Nature Struct. Biol.* **5**, 1058–1064.
Hollis, T., Ichikawa, Y. & Ellenberger, T. (2000). *EMBO J.* **19**, 758–766.
Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). *Protein Eng.* **14**, 993–1000.
Koval, V. V., Kuznetsov, N. A., Zharkov, D. O., Ishchenko, A. A., Douglas, K. T., Nevinsky, G. A. & Fedorova, O. S. (2004). *Nucleic Acids Res.* **32**, 926–935.
Labahn, J., Schärer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L. & Ellenberger, T. E. (1996). *Cell*, **86**, 321–329.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Leiros, I., Moe, E., Smalås, A. O. & McSweeney, S. (2005). *Acta Cryst.* D**61**, 1049–1056.
Leiros, I., Nabong, M. P., Grøsvik, K., Ringvoll, J., Haugland, G. T., Uldal, L., Reite, K., Olsbu, I. K., Knaevelsrud, I., Moe, E., Andersen, O. A., Birkeland, N. K., Ruoff, P., Klungland, A. & Bjelland, S. (2007). *EMBO J.* **26**, 2206–2217.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V. & Daly, M. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 44–79.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 499–501.
Mattimore, V. & Battista, J. R. (1996). *J. Bacteriol.* **178**, 633–637.
Moe, E., Leiros, I., Smalås, A. O. & McSweeney, S. (2006). *J. Biol. Chem.* **281**, 569–577.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
O'Rourke, E. J., Chevalier, C., Boiteux, S., Labigne, A., Ielpi, L. & Radicella, J. P. (2000). *J. Biol. Chem.* **275**, 20077–20083.
Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). *EMBO J.* **17**, 5214–5226.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Ringvoll, J., Nordstrand, L. M., Vågbø, C. B., Talstad, V., Reite, K., Aas, P. A., Lauritzen, K. H., Liabakk, N. B., Bjørk, A., Doughty, R. W., Falnes, P. Ø, Krokan, H. E. & Klungland, A. (2006). *EMBO J.* **25**, 2189–2198.
Saparbaev, M., Kleibl, K. & Laval, J. (1995). *Nucleic Acids Res.* **23**, 3750–3755.
Saparbaev, M. & Laval, J. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 5873–5877.
Sugahara, M., Mikawa, T., Kumasaka, T., Yamamoto, M., Kato, R., Fukuyama, K., Inoue, Y. & Kuramitsu, S. (2000). *EMBO J.* **19**, 3857–3869.
Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P. & Tainer, J. A. (1995). *EMBO J.* **14**, 4108–4120.
Thomas, L., Yang, C. H. & Goldthwait, D. A. (1982). *Biochemistry*, **21**, 1162–1169.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yamagata, Y., Kato, M., Odawara, K., Tokuno, Y., Nakashima, Y., Matsushima, N., Yasumura, K., Tomita, K., Ihara, K., Fujii, Y., Nakabeppu, Y., Sekiguchi, M. & Fujii, S. (1996). *Cell*, **86**, 311–319.