Journal of Structural Biology 183 (2013) 66–75

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

A new family of proteins related to the HEAT-like repeat DNA glycosylases with affinity for branched DNA structures ☆

Paul H. Backe [a,b,1], Roger Simm [c,1], Jon K. Laerdahl [a], Bjørn Dalhus [a,b], Annette Fagerlund [c], Ole A. Økstad [c], Torbjørn Rognes [a,d], Ingrun Alseth [a], Anne-Brit Kolstø [c], Magnar Bjørås [a,b,*]

[a] Department of Microbiology, Oslo University Hospital and University of Oslo, P.O. Box 4950 Nydalen, 0424 Oslo, Norway
[b] Department of Medical Biochemistry, Oslo University Hospital and University of Oslo, P.O. Box 4950 Nydalen, 0424 Oslo, Norway
[c] Department of Pharmaceutical Biosciences, School of Pharmacy, University of Oslo, 0316 Oslo, Norway
[d] Department of Informatics, University of Oslo, P.O. Box 1080 Blindern, 0316 Oslo, Norway

ARTICLE INFO

Article history:
Received 21 November 2012
Received in revised form 9 April 2013
Accepted 16 April 2013
Available online 25 April 2013

Keywords:
DNA damage
Base excision repair
DNA glycosylases
AlkC
AlkD
Holliday junctions

ABSTRACT

The recently discovered HEAT-like repeat (HLR) DNA glycosylase superfamily is widely distributed in all domains of life. The present bioinformatics and phylogenetic analysis shows that HLR DNA glycosylase superfamily members in the genus *Bacillus* form three subfamilies: AlkC, AlkD and AlkF/AlkG. The crystal structure of AlkF shows structural similarity with the DNA glycosylases AlkC and AlkD, however neither AlkF nor AlkG display any DNA glycosylase activity. Instead, both proteins have affinity to branched DNA structures such as three-way and Holliday junctions. A unique β-hairpin in the AlkF/AlkG subfamily is most likely inserted into the DNA major groove, and could be a structural determinant regulating DNA substrate affinity. We conclude that AlkF and AlkG represent a new family of HLR proteins with affinity for branched DNA structures.

1. Introduction

Genomic DNA is constantly modified by chemical agents and physical processes of endogenous or exogenous origin, including reactive metabolites, environmental toxins and ionizing radiation (Lindahl, 1993). To withstand the harmful effects of these sources of potential mutagenesis or genomic instability, several DNA repair pathways have evolved (Friedberg et al., 2006). Base excision repair (BER) is the main cellular pathway for correcting non-bulky DNA base lesions. The BER pathway is initiated by a DNA glycosylase that recognizes and excises the damaged base by cleavage of the *N*-glycosylic bond between the 2′-deoxyribose and the base. The DNA glycosylases can be divided into five different structural superfamilies (Berti and McCann, 2006; Stivers and Jiang, 2003; Dalhus et al., 2009). Recently, *Bacillus cereus* AlkC and AlkD were shown to belong to the fifth structural superfamily of DNA glycosylases (Alseth et al., 2006; Dalhus et al., 2007). The enzymes within this superfamily are built entirely from α-helices in a solenoid-like-superhelix fold similar to ARM- and HEAT-repeat-containing proteins and has been termed the HEAT-like repeat (HLR) DNA glycosylase superfamily (Dalhus et al., 2007, 2009). The enzymes within this family was shown to display specific activities towards N3- and N7-alkylpurines (Alseth et al., 2006). So far, *B. cereus* AlkD is the only HLR DNA glycosylase with a thoroughly investigated 3D structure, first determined by homology modeling (Dalhus et al., 2007), and later confirmed experimentally by X-ray crystallography (Rubinson et al., 2008). The structure of AlkD has also been solved in complex with DNA, demonstrating a novel lesion capture mechanism, apparently not involving base flipping (Rubinson et al., 2010).

Here we present a clustering analysis suggesting that the newly discovered HLR DNA glycosylase superfamily can be categorized into three subfamilies: AlkC, AlkD and AlkF/AlkG. Unexpectedly, biochemical and genetic analysis demonstrates that AlkF and AlkG possess no DNA glycosylase activity. Instead these proteins bind branched DNA, such as Holliday junctions (HJ) and 3-way junctions (3WJ). A crystal structure of AlkF without DNA and a computer model of the protein in complex with DNA, suggest a model for DNA binding which is unique for the AlkF/AlkG family.

* Corresponding author at: Department of Microbiology, Oslo University Hospital and University of Oslo, P.O. Box 4950 Nydalen, 0424 Oslo, Norway. Fax: +47 23074061.
E-mail address: magnar.bjoras@rr-research.no (M. Bjørås).
[1] These authors contributed equally to this work.

1047-8477/$ - see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2013.04.007

2. Material and methods

2.1. PCR amplification, cloning and site-directed mutagenesis

The genes encoding proteins AlkF (BC3264, GenBank accession NP_833004) and AlkG (BC2926, NP_832674) were amplified by the polymerase chain reaction (PCR) from *B. cereus* type strain ATCC 14579 genomic DNA using standard protocols. Forward and reverse primers were 5′-ctagctagcatggattttaaaacagttatgc-3′ and 5′-cgcggatcctttaacaccttacatacttacgc-3′, respectively, for AlkF (restriction sites NheI and BamHI are underlined) and 5′-atatccatggtccatatgatgttacttgaagaagtaatgc-3′ and 5′-cgcggatcctttatttcactttctttttttcagc-3′, respectively, for AlkG (restriction sites NcoI and BamHI are underlined). After amplification, the PCR products were digested with NheI/NcoI and BamHI and cloned into the pET28b vector (Novagen) in-frame with an N-terminal hexahistidine tag separated from the inserted coding regions by a thrombin protease cleavage peptide to give pET28b-AlkF and pET28b-AlkG. Site-directed mutagenesis was performed according to the QuikChange mutagenesis protocol (Stratagene). The Arg203Ala, Lys206Ala and Lys207Ala mutations were simultaneously introduced by use of a single oligonucleotide and a complementary oligo, to give the AlkF mutant (AlkF mut): 5′-gtatagttgaagtcaaa**gcg**gataat**gcggcg**agcagt ttgttaaatg-3′. The mutant constructs were verified by sequencing.

2.2. Protein expression and purification

The pET28b-AlkF/G vectors were transformed into the *E. coli* BL21CodonPlus(DE3)-RIL strain (Stratagene) for protein expression. Cells were grown in LB-medium supplemented with 50 μg/ml kanamycin at 37 °C until an OD_{600} of 0.7 was reached and expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. After an additional 3 h incubation at 37 °C, cells were collected by centrifugation at 6000*g* for 30 min and stored at −20 °C. The thawed cell pellet was resuspended in 10 mM Tris–HCl pH 8.0, 300 mM NaCl, 10 mM β-mercaptoethanol (β-ME) and 10 mM imidazole buffer, using 15 ml buffer per liter of cell culture. Cells were sonicated for 3 × 30 s at 4 °C, followed by centrifugation at 27000*g*. The supernatant was loaded onto a Ni-NTA column pre-equilibrated with the sonication buffer. Bound AlkF, AlkF mut or AlkG was released from the resin using 10 mM Tris–HCl pH 8.0, 300 mM NaCl, 10 mM β-ME and 300 mM imidazole buffer. Both AlkF, AlkF mut and AlkG proteins were separated from their His-tags by thrombin cleavage with concomitant dialysis against 10 mM Tris–HCl pH 8.0, 50 mM NaCl and 10 mM β-ME (buffer A). The AlkF protein was further purified by use of a 6 ml Resource S cation exchange column, with a linear NaCl gradient in buffer A from 50 mM to 1 M NaCl. Fractions rich in AlkF protein were pooled and dialyzed against buffer A. The purification of AlkF, AlkF mut and AlkG was completed by a size exclusion polishing step using a Superdex75 gel filtration column with buffer A as the running buffer. Fractions with pure protein were pooled and concentrated to 18.5 mg/ml (AlkF), 24.0 mg/ml (AlkF mut) and 12.5 mg/ml (AlkG).

2.3. Crystallization, data collection, structure determination and refinement of AlkF

AlkF was crystallized by the hanging drop vapor diffusion method. A 1.0 μl protein droplet was mixed with 1.0 μl of precipitant solution containing 20% PEG 3350 and 0.15 M ammonium chloride, and equilibrated against the precipitant reservoir at room temperature. Crystals grew within a few days. The crystals were briefly soaked in 20% PEG 3350, 0.15 M ammonium chloride and 10% PEG 400 before being flash-frozen in liquid nitrogen and then exposed to X-rays. Crystallographic data to 1.58 Å resolution were collected at the BW7A Macromolecular Crystallography Beamline at the DESY synchrotron in Hamburg, Germany. The data was processed and scaled with Denzo and Scalepack within the HKL2000 package (Otwinowski and Minor, 1997). Data collection statistics are summarized in Table 1. The crystal belongs to the primitive orthorhombic space group $P2_12_12_1$, with unit cell parameters a = 52.92, b = 87.16 and c = 112.60 Å. The structure was determined by molecular replacement using MolRep/CCP4 (CCP4, 1994) with atomic coordinates taken from PDB deposition 1T06. The structure was refined to 1.6 Å resolution with Refmac5/CCP4 (Murshudov et al., 1997) interspersed with model building in Coot (Emsley and Cowtan, 2004). 5% of randomly distributed reflections were flagged for cross-validation during all steps of the refinement and not included in the model refinement. Improvement of the model was confirmed by the steady decrease in both R_{work} and R_{free}. The final model has an R_{work} and R_{free} of 0.165 and 0.205 respectively. The final model contains two protein chains, each with 235 residues, as well as 662 solvent water molecules with full occupancy. The refinement statistics are summarized in Table 1. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (Accession code: 3ZBO).

Table 1
Crystallographic data collection and refinement statistics.

Data collection and crystal statistics	
Beam line	BW7A – DESY
Wavelength (Å)	1.0000
Temperature (K)	100
Space group	$P2_12_12_1$
Cell dimensions (Å)	a = 52.92, b = 87.16, c = 112.60
Mosaicity (°)	0.60
Oscillation range (°)	0.25
Resolution (Å)	50.00–1.58 (1.64–1.58)*
I/σI	25.97 (3.76)
Total no. of reflections	281308
Unique reflections	72160
Average redundancy	3.90 (3.3)
Completeness (%)	99.6 (98.5)
R-sym†	0.050 (0.31)
Refinement statistics	
No. of reflections in refinement	68454
No. of reflections in test set	3629
R_{work}/R_{free} (%)	16.5/20.5
Number of water molecules	662
Number of nonhydrogen atoms	4529
Mean overall atomic B-factor (Å^2)	21.55
Rms deviation from ideal bond length (Å)	0.027
Rms deviation from ideal bond angles (°)	2.263
Ramachandran plot statistics (%)	
Residues in	
Most favourable region (%)	93.7
Additional allowed region (%)	5.8
Generous allowed region (%)	0.2
Disallowed region (%)	0.2

* Numbers in parenthesis correspond to the highest resolution bin (1.64–1.58 Å).
† $R_{sym} = \sum_{hkl}\sum_i |I(hkl)_i - < I(hkl) > | / \sum_{hkl}\sum_i < I(hkl)_i >$.

2.4. DNA substrates

HPLC purified oligonucleotides were purchased from Eurofins MWG Operon. The synthetic HJ (X26), was made by combining the oligonucleotides 1–4 in Table 2 as previously described (Constantinou et al., 2001). The 3WJ was made by combining the oligonucleotides 5–7 in Table 2 as previously described (Witte et al., 2008). The double-stranded (ds) linear non-damaged DNA was made by combining the oligonucleotides 8 and 9 in Table 2. Radioactive labelled oligonucleotides were annealed to their respective

Table 2
Oligonucleotides used in DNA substrates.

The synthetic Holliday junction (X26) was made by combining the oligonucleotides 1-4 as previously described Constantinou et al. (2001)
1: 5′-CCGCTACCAGTGATCACCAATGGATTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC-3′
2: 5′-TGGGTGAACCTGCAGGTGGGCAAAGATGTCCTAGCAATCCATTGTCTATGACGTCAAGCT-3′
3: 5′-GAGCTTGACGTCATAGACAATGGATTGCTAGGACATCTTTGCCGTCTTGTCAATATCGGC-3′
4: 5′-TGCCGATATTGACAAGACGGCAAAGATGTCCTAGCAATCCATTGGTGATCACTGGTAGCGG-3′
The 3-way junction was made by combining the oligonucleotides 5-7 previously described in Witte et al. (2008)
5: 5′ GGATACGTAACAACGCTTATGCATCGCCGCCGCTACATCCCTGAGCTGAC 3’
6: 5′ TGTGTTCGATCTCGATCAGAATGACGATGCATAAGCGTTGTTACGTATCC 3’
7: 5′ GTCAGCTCAGGGATGTAGCGGCGGAGTCATTCTGATCGAGATCGAACACA 3’,
The ds linear undamaged DNA was made by combining the oligonucleotides 8-9
8: 5′-TACGATCACGGATGATGGGAGTCGACGCAGTGGCCGG 3’
9: 5′-CCGGCCACTGCGTCGACTCCCATCATCCGTGATCGTA 3’

complementary strands by heating the solution to 95 °C for 5 min and slowly cooling to room temperature. In all substrates, one of the strands was labeled at the 5′ end prior to the annealing using [γ-^{32}P]ATP and T4 polynucleotide kinase. Labeled substrates were purified by non-denaturing 10% PAGE.

2.5. Electrophoretic mobility shift assay (EMSA)

To investigate the DNA binding properties of recombinant AlkF, AlkF mut and AlkG we used HJ, 3WJ and ds linear DNA. Purified protein was mixed with ^{32}P-labelled substrate DNA (10 fmol) in reaction buffer (70 mM MOPS (pH 7.5), 1 mM DDT, 1 mM EDTA and 5% glycerol) to a final volume of 10 μl on ice. After 30 min incubation on ice, protein-DNA complexes were analyzed by non-denaturing PAGE using 8% gels in low ionic strength buffer (taurin) on ice. Gels were analyzed by PhosphorImager (Typhoon 9419, GE Healthcare).

2.6. Formamidopyrimidine (faPy) and alkyl base (MNU) DNA glycosylase assays

All enzyme activities were assayed in reaction buffer as above. Calf thymus DNA containing alkylated bases or faPy was prepared by treatment with *N*-[^{3}H]-methyl-*N*-nitrosourea ([^{3}H]-MNU) (1.5 Ci mmol^{-1}) as described (Alseth et al., 1999; Bjelland et al., 1993). Briefly, 2.5 μg alkylated or faPy containing DNA (6000 dpm μg^{-1}), was incubated for 30 min at 37 °C with purified protein (20 pmol) in a total volume of 50 μl. DNA was precipitated with ethanol and the radioactivity in the supernatant was quantified in a liquid scintillation counter (Tri-Carb 2900TR, Paxkard). Purified human 8-oxoguanine DNA glycosylase (hOGG1) and human alkyladenine DNA glycosylase (AAG) were used as positive controls in the faPy- and MNU-assay, respectively.

2.7. Assay for enzyme cleavage of DNA oligonucleotides containing base lesions

Double-stranded DNA oligonucleotides containing single base damages including 7,8-dihydro-8-oxoguanine, 5-hydroxycytosine, uracil, ethenoadenine and inosin were generated as previously described (Alseth et al., 1999). Briefly, the 5′ end of the damage-containing oligonucleotides were labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP (3000 Ci mmol^{-1}), and annealed to complementary strands. The labelled DNA substrates (10 fmol) and proteins were incubated at 37 °C for 30 min in a total volume of 10 μl of reaction buffer (same as above). The resulting abasic sites after base removal were cleaved by adding 100 mM NaOH and continuing incubation for 20 min at 70 °C. The reactions were stopped by adding 100 mM HCl and 10 μl DNA loading buffer (90% formamide, 5 mM EDTA, 0.01% bromophenol blue), followed by heat denaturation at 80 °C for 3 min and separation of reaction products on 7 M urea-20% polyacrylamide gels. The radiolabelled fragments were visualized using a PhosphorImager (Typhoon 9419).

2.8. Construction of *alkF*$^-$ and *alkG*$^-$ knock-out mutants in *B. cereus*

Markerless single mutants of *alkF*$^-$ and *alkG*$^-$ as well as an *alkF*$^-$ *alkG*$^-$ double mutant were constructed in *B. cereus* ATCC 14579 basically as previously described (Arnaud et al., 2004; Janes and Stibitz, 2006). In short, the method involves cloning of a gene-replacement fragment containing the fused up- and downstream regions flanking the gene to be knocked out into a suicide vector. The gene-replacement fragment for *alkF* was cloned into a pMAD vector, modified to contain a I-SceI recognition site, whereas the gene-replacement fragment for *alkG* was cloned into pBKJ236, which also contains a I-SceI site. The plasmids pMAD (Arnaud et al., 2004) and pBKJ236 (Janes and Stibitz, 2006) are shuttle vectors able to replicate in *E. coli* as well as in *B. cereus* at a permissive temperature. The presence of the I-SceI recognition site enables forced recombination following insertion of the plasmids into the chromosome and introduction of a plasmid encoding I-SceI into that mutant. In addition, the pMAD vector contains the *bgaB* gene under control of a constitutive promoter that expresses a thermostable β-galactosidase. This enables blue/white screening of mutants carrying the plasmid which facilitates the identification of clones that has lost the plasmid following recombination. After transformation, the gene-replacement fragments of the vectors were recombined into the chromosome of *B. cereus*. Since autonomous replication of pBKJ236 and pMAD is restricted at 37 °C, selection of recombinants was done on LB plates containing erythromycin (5 μg/ml) at 37 °C. Recombinants were controlled by PCR using primers annealing to the chromosome 5′ and 3′ of the gene-replacement fragment. Correct recombinants were clean-streaked twice on LB plates supplemented with erythromycin (5 μg/ml). Next, recombinants were spread on LB agar plates and grown overnight at 37 °C. Resulting colonies were screened for erythromycin sensitivity on LB agar plates as well as LB agar plates containing erythromycin (5 μg/ml). Bacteria that had undergone the second recombination step were identified as erythromycin sensitive colonies. Screening for markerless mutants was done by PCR amplification using primers annealing to the chromosome up- and downstream of the knocked out genes. Deletions mutations were confirmed by sequencing.

2.9. RNA isolation

B. cereus were harvested by diluting the cultures with an equal volume of ice-cold methanol followed by centrifugation at 4000*g* for 5 min at 4 °C. Cells were disrupted by using a Precellys® 24 tis-

sue homogenizer. RNA isolation was performed using the RNeasy Mini Kit (Qiagen, Germany) according to the manufacturer's protocol, including the optional on-column DNAse treatment. After elution, the RNA was treated with Turbo DNAse (Applied Biosystems, USA) according to the manufacturer's manual followed by a second round of purification using the RNeasy Mini Kit. The concentration and purity of RNA was determined by UV-spectrophotometry and the integrity of the RNA was controlled by agarose gel electrophoresis.

2.10. Quantitative RT-PCR

Complementary strand DNA was synthesized from 1 μg of RNA using the Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. The cDNA was diluted 1:5 (1:2500 for the reference gene 16S). 3 μl of diluted cDNA was mixed with primers (0.5 μM) as well as qPCR&GO™ LC480 green master mix (MPBiomedicals, USA) to a final volume of 15 μl. Primers were designed to give an amplicon of approximately 100 bp. Quantitative PCR was performed with a Roche Lightcycler 480 (Roche Diagnostics GmbH, Germany). The cycling conditions were 95 °C for 1.5 min followed by 45 cycles with the following three steps: 95 °C for 10 s, 58 °C for 10 s and 72 °C for 10 s. The crossing point (Ct) values were determined by the 2nd derivative maximum of two technical replicates per biological replicate. Results were calculated by the ΔΔCt approximation. Expression ratios are averages of at least 2 biological replicates using 16S as the reference gene for normalization.

2.11. Sporulation assay

Bacteria were grown at 30 °C in LB medium overnight. A preculture was started by diluting the ON-culture 1:100 in fresh LB medium. The preculture was grown at 30 °C until OD_{600} reached 1. Bacteria were diluted to an OD_{600} of 0.01 in 10 ml modified G-medium (MGM) in 100 ml Erlenmeyer flasks. The cultures were supplemented with nalidixic acid (NAL) at 0, 10 and 50 μg/ml, and were grown for 24 h at 30 °C. Samples were taken at regular intervals for analysis of sporulation efficiency. At each time point two samples (A and B) were taken from each culture. Sample A was incubated at 70 °C for 20 min to kill all vegetative cells, leaving only the heat-resistant spores intact. In the meantime sample B was diluted in ice-cold water and plated on LB-agar plates. Similarly, following heat treatment, sample A was diluted in ice-cold water and plated on LB-agar plates. Sample B gives a measure of the total number of viable bacteria (cells and spores) present in the sample. Plates were incubated for 10–12 h at 30 °C before colonies were counted. The number of colony forming units (CFU) in the spore sample A was compared to the number of CFU in the viable bacteria sample B. The resulting ratio designates the sporulation efficiency at a specific condition. Each experiment was carried out in two biological replicates, with two technical replicates for each biological replicate. The experiments were repeated twice for the control sample (0 μg/ml NAL) and at least seven times for the NAL (10 or 50 μg/ml) stressed cells.

2.12. Sensitivity to genotoxic stress

Single knock out mutants of *alkF*$^-$ and *alkG*$^-$ as well as the *alkF*$^-$ *alkG*$^-$ double mutant in *B. cereus* were analyzed for genotoxic stress using methyl methanesulfonate (MMS), NAL, cisplatin (CP), hydrogen peroxide, gamma (γ) and ultraviolet (UV) radiation. Wild-type and mutant cells were grown to reach the exponential phase in LB-medium at 30 °C. The cells were diluted serially in MQ water and spotted onto LB plates containing MMS (3.0 mM), NAL (3.0 μM) or CP (1.3 μM) and incubated at 30 °C overnight. For hydrogen peroxide, the cells were exposed for 60 min at 600 μM concentration before spotting the cells to the plate. For γ and UV radiation exposure, 250 Gy and 30 J respectively, the cells were radiated after spotting the cells onto the LB plates. All experiments were performed in at least three replicates.

2.13. Sequence analysis

In order to identify homologs of the AlkC and AlkD DNA glycosylases that have been characterized earlier (Alseth et al., 2006), the AlkC and AlkD sequences from *B. cereus* strain ATCC 14579 (accessions NP_832800 and NP_834586, respectively) were used as queries for PSI-BLAST searches (Altschul et al., 1997) in the NCBI Reference Sequence (RefSeq) protein database (Pruitt et al., 2012) as of September 2012. PSI-BLAST was run with default parameters and was terminated after five iterations. Both searches gave approximately 1700 hits (expect values below 0.01) with more than 95% overlap between the sets. Very short (<180 residues) and long (>400) sequences were discarded in order to avoid erroneously annotated protein fragments or fusion genes resulting in a total of 1597 sequences. From this set, 244 proteins belonging to the *Bacillus* genus, 98 separate strains, were selected for further analysis.

An all-vs-all sequence comparison was performed using the SWIPE optimal local sequence alignment search tool (Rognes, 2011). Proteins aligned with expect values not greater than 0.0001 were considered significantly similar. The 244 sequences were subsequently clustered using a simple single-linkage clustering algorithm that allows only one sequence from each organism in the same cluster, in order to place paralogs in separate clusters. This procedure resulted in three separate networks with seven clusters. Cytoscape (Smoot et al., 2011) was used to draw protein network graphs using the force-directed layout with the alignment score as a parameter.

3. Results

3.1. Sequence searching reveals two new members of the HEAT-like repeat DNA glycosylase superfamily in B. cereus, AlkF and AlkG

Using PSI-BLAST, 244 homologs (see Supplementary Table S1) of the AlkC and AlkD proteins within the *Bacillus* genus were identified. All pairs of these protein sequences were aligned to each other and then clustered into seven clusters containing no more than one protein from each strain. A protein similarity network graph of the proteins was generated (Fig. 1). The AlkC and AlkD homologs identified in *Bacillus* form three separate networks, *i.e.* they belong to three protein families. The first network contains AlkC, the second network contains AlkD, and the third network contains two previously unstudied proteins from the *B. cereus* ATCC 14579 type strain. These new proteins are termed AlkF and AlkG in order to highlight their evolutionary relationship with AlkC and AlkD, and because they are built from HEAT-like repeats (see below), recently suggested to be denoted ALK repeats (Rubinson and Eichman, 2012). The previously described *B. cereus* AlkE (Alseth et al., 2006) is evolutionary unrelated to the HEAT-like repeat superfamily.

Some *Bacillus* strains have an additional AlkF-like protein (Fig. 1, cyan), while other strains have one or two additional AlkC-like proteins (Fig. 1, light blue and orange). The *B. cereus* ATCC 14579 type strain has proteins belonging to four of the clusters, while five of the *Bacillus* strains examined had proteins in five different clusters. No strains examined had more than five. A multiple sequence alignment of *B. cereus* ATCC 14579 AlkF and AlkG and a number of homologs is shown in Supplementary Fig. S1. AlkF and AlkG

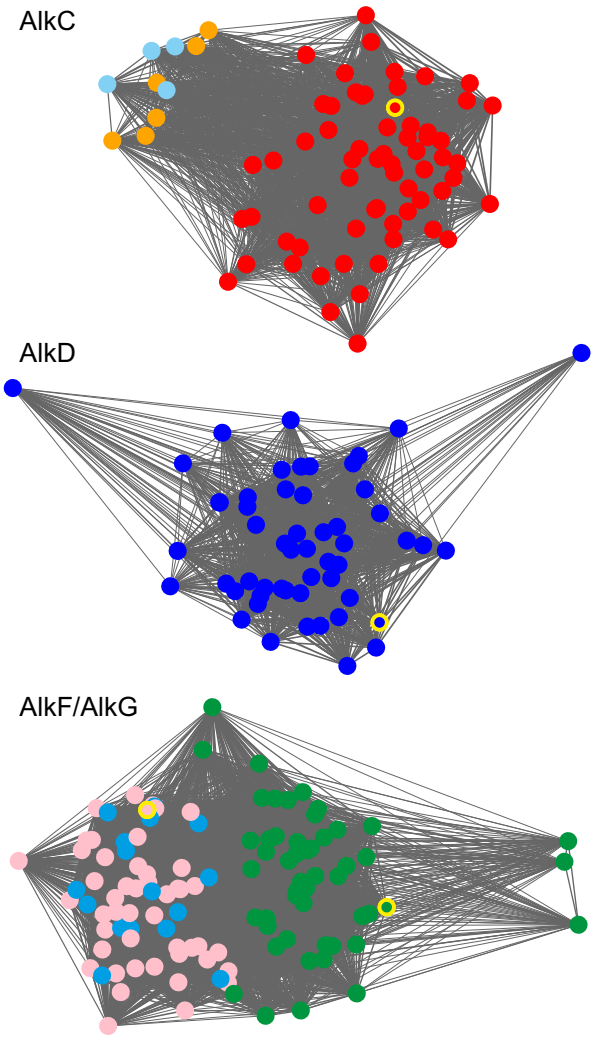

Fig.1. Network graph of the AlkD protein and homologs within the *Bacillus* genus. The nodes represent protein sequences, and significant similarities (expect values $<10^{-4}$) between them are indicated by lines between the nodes. Nodes are coloured according to which cluster they belong to. The widths of the lines are proportional to the alignment score for each protein pair. Proteins that are more similar to each other will generally be located closer to each other than less similar proteins. Also, sets of proteins that have many connections to each other will be located closer to each other than other proteins. The original AlkC and AlkD proteins from *Bacillus cereus* characterized earlier (Alseth et al., 2006), as well as the AlkF and AlkG proteins characterized in this work, are indicated with a yellow border. The proteins form three separate networks, each corresponding to a protein family, where each network is tightly connected. The network at the top contains the AlkC protein (YP_002339053) (red) that is most similar to the originally characterized AlkC protein. This network also contains light blue and orange nodes (paralogs) that are located at some distance from the red nodes. The middle network contains the AlkD protein (ZP_03237343) (blue) that is most similar to the first characterized AlkD protein. All nodes in this network belong to the same cluster. The bottom network contains the AlkF (NP_833004) (pink) and AlkG (NP_832674) (green) proteins. The pink and green nodes are clearly separated in the network. There are also some cyan nodes located in between the pink nodes. Four of the *Bacillus* strains studied had proteins belonging to all of these three clusters (pink, green, cyan). A list of strains and accession numbers for the proteins included in this figure is provided in Supplementary Table S1.

have a sequence identity of approximately 36% and are clearly in the same protein family. Pairwise sequence identity between AlkF/AlkG, AlkC, and AlkD are well below 20% and sequences belonging to the three separate families cannot be reliably aligned based on sequences alone (see discussion below).

In total, more than 1500 homologs of the AlkC and AlkD proteins were identified by sequence searching. As previously shown (Rubinson et al., 2010) proteins belonging to the AlkD family are found in all three domains of life. Members of the AlkC and AlkF/AlkG families, however, appear to be rare, if at all present, in Archaea and Eukaryota. Within the bacterial domain of life, all three families are wide-spread in Firmicutes, common in Actinobacteria and Bacteroidetes/Chlorobi, but rare or completely missing in other phyla such as the Proteobacteria, Chlamydiae and Cyanobacteria.

3.2. AlkF and AlkG exhibit no DNA glycosylase activity

Since both AlkF and AlkG are remote homologs of the AlkC and AlkD DNA glycosylases, we examined the abilities of recombinant AlkF and AlkG to remove alkylated bases from MNU-treated DNA, or faPy from faPy-containing DNA. The relative amounts of alkylated methylpurines present in the MNU-treated DNA substrate are 65% 7-methylguanine, 10% 3-methyladenine and 0.7% 3-methylguanine (Bjelland et al., 1993). These experiments showed that neither AlkF nor AlkG are able to remove alkylated DNA bases or formapyrimidines (Fig. 2A and B). Further, AlkF and AlkG showed no DNA glycosylase activity towards other types of DNA base lesions such as other oxidative DNA base lesions, ethenoadenine or deaminated DNA bases (data not shown). It thus appears that *B. cereus* AlkF and AlkG are structural, but not functional homologs of the AlkC and AlkD HEAT-like repeat DNA glycosylases.

3.3. AlkF and AlkG bind Holliday and 3-way junctions DNA with higher affinity than linear DNA

We used EMSA to study the DNA binding capability of purified recombinant AlkF and AlkG. Branched DNA such as HJ and 3WJ, and ds linear undamaged DNA were used as substrates. We observed very weak binding of both AlkF and AlkG to linear duplex DNA (Fig. 2C), even at the highest concentration. In EMSAs with HJ and 3WJ DNA as substrates, we observed progressively slower migrating species with increasing concentrations of protein, indicating that the proteins form multimeric complexes with DNA. These results demonstrate that both AlkF and AlkG bind to HJ and 3WJ with higher affinity as compared to ds linear DNA. It thus appears that AlkF and AlkG bind preferentially to branched DNA structures. Moreover, the triple mutant protein AlkF mut, in which the three positive residues Arg203, Lys206 and Lys207 within the β-hairpin have been mutated to alanine, displays significantly weaker binding towards the HJ DNA compared to wild type AlkF (Fig. 2D).

3.4. *alkF*$^{-}$ *alkG*$^{-}$ double mutant displays very low sensitivity to genotoxic stress

To investigate a possible role of *alkF* and *alkG* in response to genotoxic stress, single and double mutants were constructed in *B. cereus* and examined for sensitivity towards MMS, NAL, CP, hydrogen peroxide, and γ and UV radiation by survival assays. The single and double mutants displayed no sensitivity to genotoxic stress except for a very weak sensitivity towards MMS and NAL (Fig. 3), indicating that *alkF* and *alkG* are not essential for DNA repair.

3.5. The structure of AlkF resembles the HEAT-like repeat AlkD DNA glycosylase

The 3D structure of AlkF was determined by X-ray crystallography to 1.58 Å resolution (Fig. 4A). The overall structure of AlkF is similar to that of AlkD (rigid superimposition with Rapido (Mosca et al., 2008) gives a RMSD of 4.5 Å of 172 residues aligned), adopting a typical HEAT-like repeat fold that comprises 13 α-helices

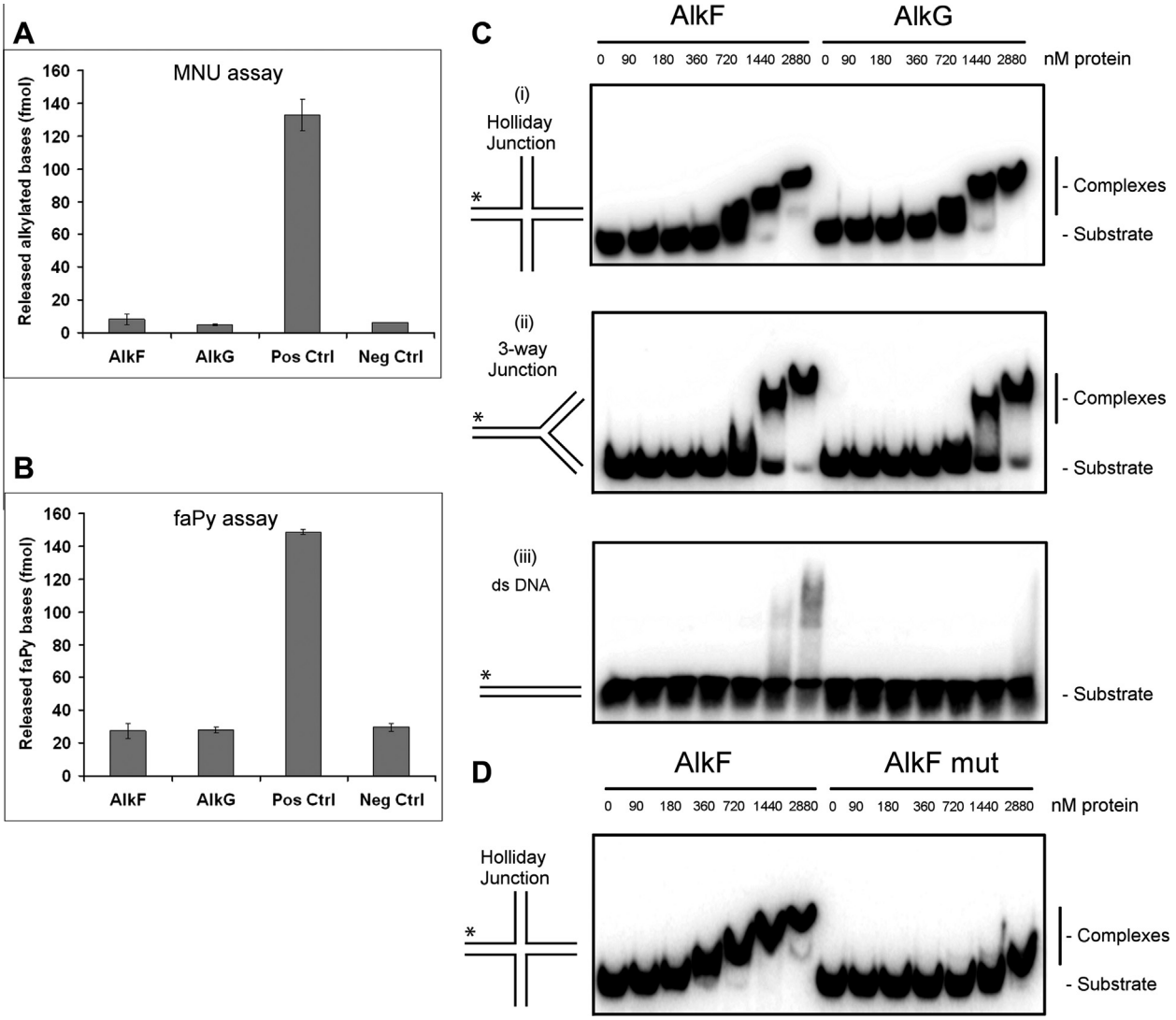

Fig.2. Enzymatic activities and DNA binding of AlkF and AlkG. Alkylbase (A) and faPy (B) DNA glycosylase activity of recombinant AlkF and AlkG. 2.5 μg alkylated or faPy containing DNA was incubated for 30 min at 37 °C with purified protein (20 pmol) in a total volume of 50 μl. The DNA was ethanol precipitated and the supernatant subjected to scintillation counting. Purified human OGG1 and human AAG were used as positive controls (Pos Ctrl) in the faPy- and alkylbase DNA glycosylase assay, respectively. Dilution buffer (25 mM HEPES pH 7.9, 15% glycerol, 1 mM EDTA, 1 mM DTT and 0.1 mg of BSA/ml) was used as negative control (Neg Ctrl). A three-parallel experiment was performed for both substrates. (C) Structure specificity of DNA binding by AlkF and AlkG. Panels i–iii are gel assays showing binding of AlkF and AlkG to the structures depicted. Binding reactions contained 1 nM of the ^{32}P-labelled substrate and AlkF and AlkG as indicated. The asterisks indicate ^{32}P label at the 5′-end DNA. (D) Binding of AlkF and AlkF mut to HJ DNA. Binding reactions contained 1 nM of the ^{32}P-labelled substrate and AlkF and AlkG as indicated. The asterisks indicate ^{32}P label at the 5′-end DNA.

(αA-αM) with one concave and one convex surface (Fig. 4A and B). The dense packing of the α-helices results in a scoop-shaped overall structure containing a wide groove on the concave side of the protein. Mapping the electrostatic potential onto the molecular surface of AlkF, reveals that the molecule acts as a dipole, with the groove/concave side of the protein being positively charged and the convex side mainly negatively charged, making the protein overall neutral (Fig. 4D). As for AlkD, 12 of the α-helices pair in an antiparallel way to form six tandemly repeated α-α-motifs (αA/αC, αD/αE, αF/αG, αH/αI, αJ/αK and αL/αM). However, the C-terminal region comprising the helices αL and αM, does have some characteristic differences compared to the AlkD structure. A 16 residue long loop, forming a two-stranded antiparallel β-sheet (β-hairpin), is inserted between helix αL and αM (Fig. 4A–C). This β-hairpin seems to be to a certain extent flexible as it displays two different conformations in the two molecules in the asymmetric unit. Maximum movement/distance measured for the β-hairpin when the two molecules are superimposed is about 8.6 Å, see Fig. 4A. Interestingly, this β-hairpin partly covers the shallow cleft at the center of the concave surface corresponding to the active site of AlkD (Fig. 4A).

The alignment of AlkD (PDB structure 3BVS (Rubinson et al., 2008)) and AlkF (this study) by the Dali structural alignment program (Holm and Rosenstrom, 2010) (Supplementary Fig. S2A) clearly demonstrates that the two proteins are in the same structural superfamily (Z-score = 13), but in two separate protein families. The Dali alignment shows that the sequence identity between full length AlkF and AlkD is approximately 12% (Supplementary Fig. S2B).

3.6. Molecular modeling of AlkF in complex with DNA

Attempts to co-crystallize AlkF and AlkG with ds DNA and other DNA substrates have so far been unsuccessful. In order to predict important residues for the interaction between AlkF and DNA, we generated a theoretical docking model of a putative AlkF in complex with linear duplex DNA. Atomic coordinates were extracted from the structure of AlkD in complex with a 12-mer ds DNA con-

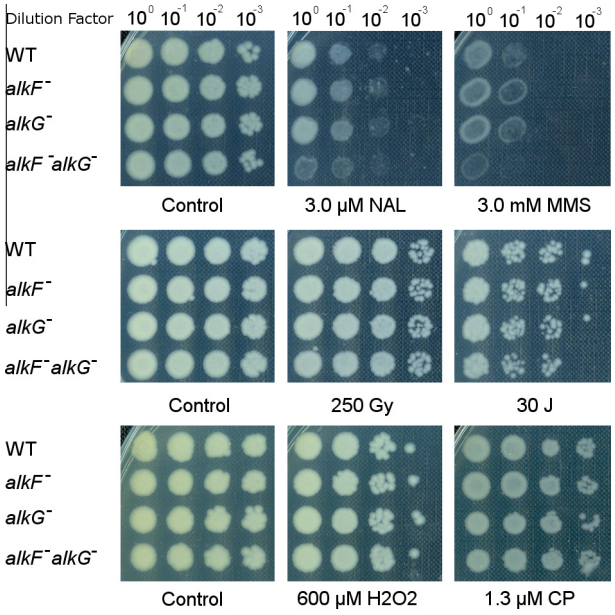

Fig.3. Characterization of sensitivity of *B. cereus* *alkF*⁻ and *alkG*⁻ single and double mutants to genotoxic stress. *B. cereus* wild type (WT) and mutants were analyzed for sensitivity against MMS, NAL, CP, hydrogen peroxide, and γ- and UV-radiation with doses as indicated. An aliquot of 10 µL serially diluted mid-log phase cultures of wild type, *alkF*⁻, *alkG*⁻ (single mutants) and *alkF*⁻ *alkG*⁻ (double mutant) were spotted onto LB plates and incubated at 30 °C overnight.

taining 3-deaza-3-methyladenine (3d3 mA), a structural 3 mA mimetic in which the N3 nitrogen is replaced with carbon (PDB structure 3JX7 (Rubinson et al., 2010)), and superimposed onto the three-dimensional structure of AlkF. The model (Fig. 5A) suggests that the DNA backbone interacts with conserved Lys and Arg residues along the edges of the wide AlkF groove in a similar fashion as AlkD. Furthermore, in the model, the β-hairpin which is only found in the AlkF/AlkG family of the HLR DNA glycosylase superfamily, is inserted into the major groove of the DNA, suggesting it may have a role as a sensor/anchoring point. Interestingly, within this loop there are four positively charged amino acids, of which three (Arg203, Lys206 and Lys207) are directed towards the DNA backbone (Fig. 5B).

3.7. Expression of alkF but not alkG is increased upon entry into stationary phase

Analysis of the growth of *B. cereus* ATCC 14579 and its isogenic *alkF*⁻ *alkG*⁻ double mutant showed that the double mutant grow similarly to the wild type in LB medium at 30 °C (Fig. 6A), indicating that AlkF and AlkG are not essential for maintaining normal growth in rich medium. To determine the gene expression level of *alkF* and *alkG* under optimal growth conditions, qRT-PCR analysis was carried out at regular intervals of growth in LB medium at 30 °C (Fig. 6B). This experiment indicates that *alkF*, but not *alkG* expression, is increasing upon entry into stationary phase of growth. Moreover, the absolute expression level of *alkG* in exponential phase is approximately 1% as compared to that of *alkF* (data not shown), suggesting that AlkF may have a specialized function not required under normal physiology.

3.8. alkF and alkG are not essential to the sporulation process

Two recent reports demonstrated that the DisA (DNA integrity scanning protein) controls a *Bacillus subtilis* sporulation checkpoint in response to DNA damage by binding branched DNA structures such as HJ (Bejerano-Sagie et al., 2006; Witte et al., 2008). In order to examine if *alkG* or *alkF* are involved in the sporulation process, experiments analyzing the sporulation efficiency of the *B. cereus* ATCC 14579 (wild type) and the isogenic *alkG*⁻ *alkF*⁻ double mutant were conducted. The bacteria were grown in MGM medium at 30 °C and samples were taken for analysis at regular intervals. The samples were analyzed by microscopic examination as well

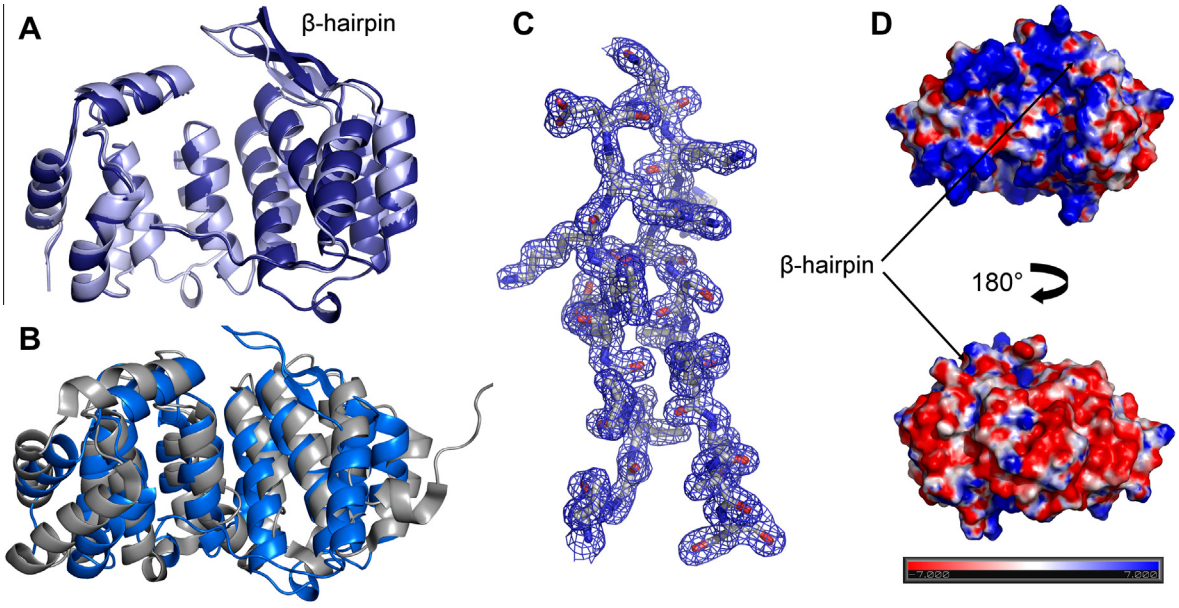

Fig.4. Structural characteristics of AlkF. (A) Cartoon representation of the overall fold and tertiary structure of *B. cereus* AlkF. The figure shows a structural superposition of the two molecules in the asymmetric unit. (B) Superposition of AlkF (blue) and AlkD (grey). Structural superposition was done using RAPIDO (Mosca et al., 2008). (C) sA-weighted composite omit map at 1.0 s calculated using the CCP4 software (CCP4, 1994) showing the β-hairpin between helix αL and αM. (D) Electrostatic potential (−7 to + 7 kT/e) of AlkF mapped onto the solvent-accessible protein surface (blue indicates positive regions; red indicates negative regions). Electrostatic potential was calculated using APBS (Baker et al., 2001). All panels were prepared with PyMOL <http://pymol.org>. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

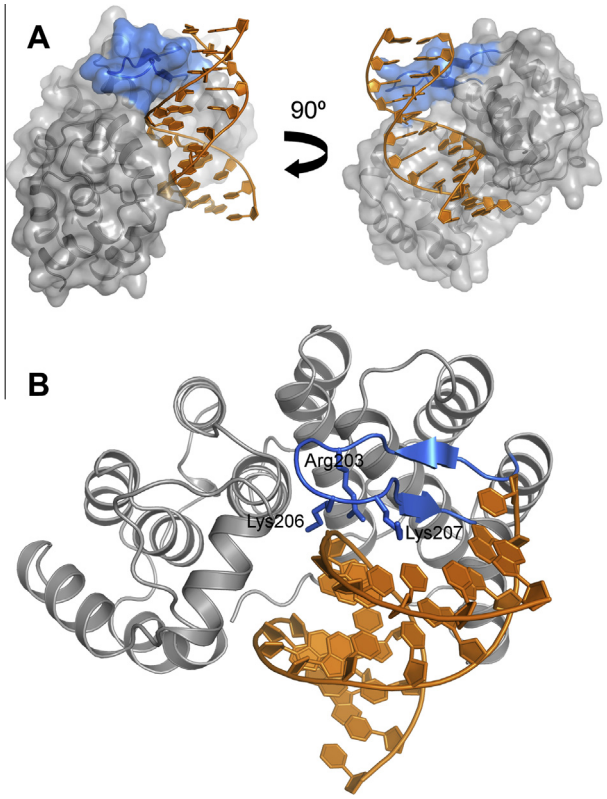

Fig.5. Proposed AlkF-DNA complex. (A) Theoretical model of AlkF in complex with double stranded DNA, illustrating possible, but weak, binding of DNA in the positively charged groove and how the β-hairpin (in blue) might be inserted in the major groove of the DNA double helix. (B) The theoretical model shows that three of the four positively charged amino acids (Arg203, Lys206 and Lys207) in the β-hairpin are directed towards the DNA backbone. DNA coordinates was extracted from PDB entry 3JX7 (Rubinson et al., 2010) and superimposed onto the three dimensional structure of AlkF using RAPIDO (Mosca et al., 2008).

as by comparing the number of colony forming units after heat treatment (70 °C for 20 min) with the total number of viable cells in a sample. The results showed no difference in sporulation between the wild type and the *alkG*⁻ *alkF*⁻ double mutant. Both strains start to sporulate after approximately 8 h (starting at OD_{600} of 0.01) as determined by the appearance of endospores and the occurrence of a heat resistant sub-population of bacteria. After 24 h incubation in the precense of NAL (10 μg/ml), the ratio of heat resistant to viable cells was 79.5% ± 18.6% and 76.8% ± 9.6% for the wild type and *alkF*⁻ *alkG*⁻ double mutant respectively. It should be noted that in the presence of NAL, the absolute number of heat resistant as well as viable cells at a given time point varied between experiments. This may be explained by the microscopic examination of the cultures, which revealed a deformed filamentous cell morphology/growth pattern. This phenotype was similar in both the wild type and the *alkF*⁻ *alkG*⁻ double mutant. It thus appears that *alkF* and *alkG* are not essential to sporulation.

4. Discussion

We have previously shown that AlkD and AlkC are single domain DNA glycosylases belonging to a new, fifth structural superfamily of DNA glycosylases (Alseth et al., 2006; Dalhus et al., 2007). It is generally accepted that the 3D structure is more conserved than sequence in distantly related proteins. Protein domains with significant sequence similarity, usually better than roughly 30% sequence identity, are classified as belonging to the same protein domain family. Protein domains that have very low or insignificant sequence similarity, but still clearly are evolutonary related based on 3D structure and functional features, are classified in the same protein domain superfamily. This protein domain classification scheme is for example employed in the most widely used domain classification hierarchies, SCOP (Andreeva et al., 2008) and CATH (Sillitoe et al., 2013), where a major fraction of the domain superfamilies comprises several families. AlkD and AlkC have a right-handed alpha-alpha superhelix fold related in structure to ARM/HEAT-repeat containing proteins (Andrade et al., 2001; Rubinson and Eichman, 2012). AlkD and AlkC belong in the same structural superfamily of proteins, but clearly in separate protein families due to very low sequence identity, well below 20%. AlkF and AlkG, here described for the first time, belong in a third unique protein family, with sequence identity between AlkD and AlkF at roughly 12%. This value is based on a structural alignment of a previously published AlkD 3D structure and the present AlkF structure (Supplementary Fig. S2B). At this low level of sequence identity, well below the "twilight zone" (Rost, 1999), it is not possible to make meaningful alignments of the protein sequences, or derive reliable phylogenies, based on sequences alone.

Mapping the sequence conservation of the AlkD family proteins onto the AlkD structure shows a concentrated patch of highly conserved residues with solvent exposed side-chains in one end of the wide DNA-binding cleft. We and others have previously shown that this is the active site region for the AlkD DNA glycosylases (Dalhus et al., 2007; Rubinson et al., 2008, 2010). Rubinson and

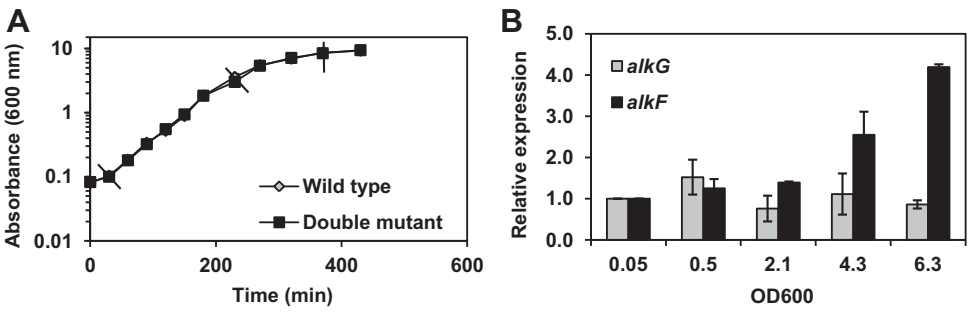

Fig.6. Bacterial growth and gene expression analysis. (A) Growth of *B. cereus* ATCC 14579 wild type (diamonds) and the isogenic *alkF*⁻ *alkG*⁻ double mutant (squares) in LB medium at 30 °C. Bacteria were diluted 1:100 from a logarithmically growing pre-culture ($OD_{600} \sim 1$) which resulted in a starting OD of approximately 0.01. Shown are averages and standard deviations from two independent experiments. The lines that are perpendicular to the growth curve indicate the boundaries of the logarithmic growth phase as well as the beginning of the stationary phase. (B) Gene expression analysis of *alkF and alkG* by qRT-PCR. Averages and standard deviations from three experiments are shown. The relative expression of *alkF* and *alkG* at different OD_{600} values is normalized to the expression level of each gene respectively, at OD_{600} of 0.05. The expression level of *alkG* in exponential phase is approximately 1% to that of *alkF*. 16S RNA was used as a control to normalize the data.

co-workers (Rubinson et al., 2010) listed 3 catalytic residues and 15 additional residues directly involved in DNA binding and other DNA interactions in AlkD (Supplementary Fig. S2B). Among these residues only Thr39 is conserved in AlkF (as residue Thr36). However, this residue is not conserved in either the AlkD or the AlkF/AlkG families (See Refs. (Rubinson et al., 2008, 2010) and Supplementary Fig. S1). It may be speculated that AlkF residues Lys40, Tyr92, Asp96, and Lys219 functionally correspond to the AlkD catalytic and DNA-binding residues Arg43, Trp109, Asp113, and His220 (Supplementary Fig. S2B). These residues are conserved in the AlkF/AlkG family, but are not invariant (Supplementary Fig. S1). In conclusion, few, if any, of the functionally important residues in AlkD are conserved in AlkF.

Unlike the AlkD family, there is no patch of highly conserved residues with solvent exposed side-chains on the surface of the AlkF/AlkG family proteins. Among the 25 invariant residues in the AlkF/AlkG family (Supplementary Fig. S1), only *B. cereus* AlkF Gly33 (Gly36 in AlkD), Ala68 (Ala80), Trp116 (Trp138), and Ser119 (Ser141) are conserved in *B. cereus* AlkD. Only Gly33/Gly36 is invariant in both the AlkF/AlkG and AlkD families (Rubinson et al., 2008, 2010), making this residue a candidate as the single fully conserved residue in this protein superfamily. More or less all 25 invariant AlkF/AlkG residues have buried side-chains and their conservation appears to be important mainly for maintaining the correct protein fold.

Conserved residues with solvent exposed side chains in the AlkF/AlkG family appears to be limited to Lys and Arg residues lining the cleft on the concave side of the proteins. Among these are *B. cereus* AlkF residues Lys19, Lys20, Lys40, Lys47, Arg166, Lys219, and residues 229–231 (Supplementary Figs. S1 and S2). These residues appear to be ideally located for interacting with negatively charged phosphate groups in bound DNA in a similar fashion as in AlkD. However, the lack of an obvious patch of highly conserved residues that could indicate an active site suggests that neither AlkF nor AlkG are DNA glycosylases. This is supported by the biochemical characterization of AlkF and AlkG, showing no glycosylase activity for a range of DNA lesions, including alkylated, oxidized and deaminated bases. This is in contrast to AlkC and AlkD in which the recombinant enzymes excised methylated bases from DNA substrates containing 7-methylguanine, 3-methyladenine, and 3-methylguanine lesions (Alseth et al., 2006; Rubinson et al., 2008, 2010). In addition, the *alkF*$^-$ *alkG*$^-$ double mutant bacteria displayed modest if any sensitivity to genotoxic stress induced by the DNA damaging agents MMS, NAL, CP and hydrogen peroxide, as well as γ- and UV-radiation.

AlkF and AlkG both bind preferentially to branched DNA. It is tempting to speculate whether the proteins may have a function in other processes than DNA repair, including replication, cell division, transcription or nucleoid compaction. Interestingly, the newly discovered checkpoint protein DisA recognizes branched DNA structures in prokaryotes (Bejerano-Sagie et al., 2006; Witte et al., 2008). DisA scans the genome of *B. subtilis* and induces a sporulation checkpoint in response to chromosomal damage. Although the *alkF*$^-$ and *alkG*$^-$ single and double mutants show no or very low sensitivity to genotoxic stress or apparent alterations in sporulation, it is possible that AlkF and AlkG operates as sensory proteins in signaling a DNA damage response by scanning/monitoring the genome for specific DNA structures or lesions. Our experimental structure of AlkF and the corresponding model of AlkF binding to DNA show a DNA–protein interface with several charged residues interacting with the DNA backbone, but no pocket for base recognition. In addition, our model of AlkF in complex with DNA indicates that the β-hairpin, containing positively charged amino acids and unique to the AlkF/AlkG subfamily, interacts with the major groove of DNA. Indeed, by mutating the three positive residues Arg203, Lys206 and Lys207 within the β-hairpin to alanine, we demonstrated a significantly weaker binding mutated AlkF to HJ DNA compared to wild type AlkF protein. Although this model does not reflect a complex with branched DNA like three-way or Holliday junction, the β-hairpin may have a role as a DNA sensor/anchoring point, possibly by binding to specific structures in DNA without the need for an active site for recognition of a particular sequence/base signature.

Funding

The work in the Bjørås laboratory was supported by the Research Council of Norway and the Norwegian Cancer Society. PHB, BD and MB were supported by the South-Eastern Norway Regional Health Authority (Grants No. 2009100, 2011040 and 2012085) for establishing the Regional Core Facility for Structural Biology and Bioinformatics. The work in the Kolstø-Økstad laboratory was supported by a FUGE II grant from the Research Council of Norway.

Acknowledgment

The authors acknowledge the technical support at the BW7A Macromolecular Crystallography Beamline at the DESY synchrotron in Hamburg.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2013.04.007.

References

Alseth, I., Eide, L., Pirovano, M., Rognes, T., Seeberg, E., Bjoras, M., 1999. The Saccharomyces cerevisiae homologues of endonuclease III from *Escherichia coli*, Ntg1 and Ntg2, are both required for efficient repair of spontaneous and induced oxidative DNA damage in yeast. Mol. Cell Biol. 19, 3779–3787.

Alseth, I., Rognes, T., Lindback, T., Solberg, I., Robertsen, K., Kristiansen, K.I., Mainieri, D., Lillehagen, L., Kolsto, A.B., Bjoras, M., 2006. A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. Mol. Microbiol. 59, 1602–1609.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W., Bork, P., 2001. Comparison of ARM and HEAT protein repeats. J. Mol. Biol. 309, 1–18.

Andreeva, A., Howorth, D., Chandonia, J.M., Brenner, S.E., Hubbard, T.J., Chothia, C., Murzin, A.G., 2008. Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res. 36, D419–D425.

Arnaud, M., Chastanet, A., Debarbouille, M., 2004. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. Appl. Environ. Microbiol. 70, 6887–6891.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA 98, 10037–10041.

Bejerano-Sagie, M., Oppenheimer-Shaanan, Y., Berlatzky, I., Rouvinski, A., Meyerovich, M., Ben-Yehuda, S., 2006. A checkpoint protein that scans the chromosome for damage at the start of sporulation in *Bacillus subtilis*. Cell 125, 679–690.

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

Bjelland, S., Bjoras, M., Seeberg, E., 1993. Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of *Escherichia coli*. Nucleic Acids Res. 21, 2045–2049.

CCP4, 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. 50, 760–763.

Constantinou, A., Davies, A.A., West, S.C., 2001. Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells. Cell 104, 259–268.

Dalhus, B., Helle, I.H., Backe, P.H., Alseth, I., Rognes, T., Bjoras, M., Laerdahl, J.K., 2007. Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats. Nucleic Acids Res. 35, 2451–2459.

Dalhus, B., Laerdahl, J.K., Backe, P.H., Bjoras, M., 2009. DNA base repair–recognition and initiation of catalysis. FEMS Microbiol. Rev. 33, 1044–1078.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.
Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., Ellenberger, T., 2006. DNA Repair and Mutagenesis. ASM Press, Washington, DC.
Holm, L., Rosenstrom, P., 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545–W549.
Janes, B.K., Stibitz, S., 2006. Routine markerless gene replacement in *Bacillus anthracis*. Infect. Immun. 74, 1949–1953.
Lindahl, T., 1993. Instability and decay of the primary structure of DNA. Nature 362, 709–715.
Mosca, R., Brannetti, B., Schneider, T.R., 2008. Alignment of protein structures in the presence of domain motions. BMC Bioinformatics. 9, 352.
Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240–255.
Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326.
Pruitt, K.D., Tatusova, T., Brown, G.R., Maglott, D.R., 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 40, D130–D135.
Rognes, T., 2011. Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. BMC Bioinformatics. 12, 221.
Rost, B., 1999. Twilight zone of protein sequence alignments. Protein Eng. 12, 85–94.
Rubinson, E.H., Eichman, B.F., 2012. Nucleic acid recognition by tandem helical repeats. Curr. Opin. Struct. Biol. 22, 101–109.
Rubinson, E.H., Metz, A.H., O'Quin, J., Eichman, B.F., 2008. A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD. J. Mol. Biol. 381, 13–23.
Rubinson, E.H., Gowda, A.S., Spratt, T.E., Gold, B., Eichman, B.F., 2010. An unprecedented nucleic acid capture mechanism for excision of DNA damage. Nature 468, 406–411.
Sillitoe, I., Cuff, A.L., Dessailly, B.H., Dawson, N.L., Furnham, N., Lee, D., Lees, J.G., Lewis, T.E., Studer, R.A., Rentzsch, R., Yeats, C., Thornton, J.M., Orengo, C.A., 2013. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic Acids Res. 41, D490–D498.
Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Ideker, T., 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27, 431–432.
Stivers, J.T., Jiang, Y.L., 2003. A mechanistic perspective on the chemistry of DNA repair glycosylases. Chem. Rev. 103, 2729–2759.
Witte, G., Hartung, S., Buttner, K., Hopfner, K.P., 2008. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. Mol. Cell 30, 167–178.