7124–7136 Nucleic Acids Research, 2009, Vol. 37, No. 21 Published online 28 September 2009
doi:10.1093/nar/gkp774

Bioinformatics and functional analysis define four distinct groups of AlkB DNA-dioxygenases in bacteria

Erwin van den Born[1], Anders Bekkelund[1], Marivi N. Moen[2], Marina V. Omelchenko[3], Arne Klungland[2,4] and Pål Ø. Falnes[1,2,*]

[1]Department of Molecular Biosciences, University of Oslo, PO Box 1041 Blindern, 0316 Oslo, Norway, [2]Centre for Molecular Biology and Neuroscience, Institute of Medical Microbiology, Oslo University Hospital and University of Oslo, N-0027 Oslo, Norway, [3]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA and [4]Institute of Basic Medical Sciences, University of Oslo, PO Box 1018 Blindern, N-0315 Oslo, Norway

Received June 18, 2009; Revised and Accepted September 3, 2009

ABSTRACT

The iron(II)- and 2-oxoglutarate (2OG)-dependent dioxygenase AlkB from *Escherichia coli* (EcAlkB) repairs alkylation damage in DNA by direct reversal. EcAlkB substrates include methylated bases, such as 1-methyladenine (m^1A) and 3-methylcytosine (m^3C), as well as certain bulkier lesions, for example the exocyclic adduct 1,N^6-ethenoadenine (εA). EcAlkB is the only bacterial AlkB protein characterized to date, and we here present an extensive bioinformatics and functional analysis of bacterial AlkB proteins. Based on sequence phylogeny, we show that these proteins can be subdivided into four groups: denoted 1A, 1B, 2A and 2B; each characterized by the presence of specific conserved amino acid residues in the putative nucleotide-recognizing domain. A scattered distribution of AlkB proteins from the four different groups across the bacterial kingdom indicates a substantial degree of horizontal transfer of AlkB genes. DNA repair activity was associated with all tested recombinant AlkB proteins. Notably, both a group 2B protein from *Xanthomonas campestris* and a group 2A protein from *Rhizobium etli* repaired etheno adducts, but had negligible activity on methylated bases. Our data indicate that the majority, if not all, of the bacterial AlkB proteins are DNA repair enzymes, and that some of these proteins do not primarily target methylated bases.

INTRODUCTION

Alkylating agents can introduce several different types of deleterious lesions into the nucleic acids DNA and RNA. Such agents are abundant in the environment, e.g. in the form of methyl halides (1,2). Alkylating agents are also generated intracellularly as a result of normal cellular metabolism, and *S*-adenosylmethionine, which is the methyl donor in numerous enzyme-catalyzed methylation reactions, also induces a low level of aberrant methylations (3). Endogenous methylating agents may also arise through cellular nitrosation reactions (4). To protect their genomes against the harmful effects of alkylating agents, living organisms possess several enzymes dedicated to remove alkylation damage from DNA (5,6).

Alkylation repair enzymes were first discovered in the bacterium *Escherichia coli* (*E. coli*), where three different repair mechanisms have been identified (6). The expression of repair proteins is up-regulated as an adaptive response to alkylation damage, governed by the alkylation sensitive transcription factor Ada. Each of the mechanisms is responsible for repairing a specific subset of the lesions introduced by alkylation. One such mechanism involves the *E. coli* AlkB protein (EcAlkB), which directly reverses alkylations at the N1-position of purines and the N3-position of pyrimidines, for example, 1-methyladenine (m^1A) and 3-methylcytosine (m^3C); the first EcAlkB substrates identified (7,8). EcAlkB belongs to the superfamily of iron(II)- and 2-oxoglutarate (2OG) dependent dioxygenases (9). As for other members of this superfamily, EcAlkB requires ferrous iron as cofactor and 2-oxoglutarate as cosubstrate, which is decarboxylated, leading to formation of succinate and CO_2.

*To whom correspondence should be addressed. Tel: +47 22854840; Fax: +47 22854443; Email: pal.falnes@imbv.uio.no

EcAlkB uses molecular oxygen to oxidize a deleterious methyl group, and the resulting hydroxymethyl moiety is spontaneously released as formaldehyde, leading to the regeneration of the non-damaged base (7,8). The set of EcAlkB substrates has later been shown, in addition to m^1A and m^3C, to include the structurally analogous, but less abundant lesions 1-methylguanine and 3-methylthymine (10–12). EcAlkB can also repair bulkier adducts such as ethyl and propyl groups (13,14), as well as exocyclic etheno and ethano adducts (15–17).

Multicellular organisms generally possess several different AlkB homologues (ALKBH), and initial bioinformatics studies identified eight such proteins in mammals, denoted ALKBH1-8 (initially named ABH1-8) (5,9,18). In addition, it was recently found that the more distantly related obesity-associated protein FTO also is a functional ALKBH (19). ALKBH2, ALKBH3 and FTO have been shown to possess a repair activity similar to that of EcAlkB, while two somewhat conflicting reports have implicated ALKBH1 both in epigenetic gene regulation and in repair (20,21). The function of the remaining five proteins remains unknown, but there are indications that they may participate in processes other than DNA/RNA repair (22,23). Interestingly, both EcAlkB and ALKBH3 also display activity on methyl lesions in RNA substrates (24,25), and it has been shown that AlkB- or ALKBH3-mediated repair of tRNA and mRNA is accompanied by functional recovery of these molecules (26). Also, the genomes of some plant-infecting single-stranded RNA (ssRNA) viruses encode AlkB homologues capable of removing methylation damage from RNA, strongly indicating that some members of the AlkB family are true RNA repair enzymes (27).

In the present work, we have performed an extensive sequence analysis of bacterial AlkB proteins, and found that they can be subdivided into four distinct groups: 1A, 1B, 2A and 2B. A rather scattered distribution of AlkB proteins from the four resulting groups across the bacterial kingdom suggested a high degree of horizontal transfer of bacterial AlkB-encoding genes. We have functionally characterized nine bacterial AlkB proteins, representing all four groups. These proteins were investigated for their ability to repair alkylated nucleic acids in *E. coli*, and recombinant enzymes were also tested for *in vitro* repair activity. Finally, we discuss the likely functions of the AlkB proteins from the four defined groups in light of the results obtained by bioinformatics and experimental analysis.

MATERIALS AND METHODS

Sequence analysis and construction of phylogenetic trees

The selected set of bacterial AlkB proteins was identified in the GenBank non-redundant (NR) protein sequence database (NCBI, NIH) using the PSI-BLAST program (28). Multiple sequence alignment was constructed using the MUSCLE program (29). Maximum likelihood tree was generated using the ProtML program of the MOLPHY package (30) by optimizing the least-squares tree with local rearrangements (Jones-Taylor-Thornton evolutionary model (31) with adjustment for observed amino acid frequencies). The reliability of the internal tree branches was estimated with the RELL bootstrap method (32) using the ProtML program.

The protein sequence alignments in Figure 2 were constructed using the online version of the MAFFT program (33), saved in .pir format, converted to .msf format in Jalview (34), and the desired appearance of the alignment was achieved in GeneDoc (http://www.genedoc.us/).

When investigating the presence of AlkB proteins in completely sequenced bacteria, protein sequence searches were complemented by TBLASTN searches (where genomes were translated in all six possible reading frames) (35), due to the incomplete annotation of many proteomes.

To generate a 16S ribosomal RNA-based phylogenetic tree, an alignment of the relevant sequences was downloaded from the 'Ribosomal database project II' (36). A phylogenetic tree based on the alignment was constructed by using the 'Calculate tree—Average distance using identity' function in Jalview. The desired appearance of the tree was obtained in FigTree v. 1.1.2 (http://tree.bio.ed.ac.uk).

Plasmid construction and protein purification

Genes encoding AlkB were amplified by polymerase chain reaction (PCR) from bacterial genomic DNA obtained from the American Tissue Culture Collection (via LGC Standards, Boras, Sweden), except in the case of *Mycobacterium tuberculosis* and *Rhizobium etli* where genomic DNA were kind gifts from T. Tønjum (Oslo University Hospital, Norway) and Dr V. González (Universidad Nacional Autónoma de México, Cuernavaca, Mexico), respectively. The PCR products were subsequently cloned into the NdeI and BamHI sites in the plasmid pJB658 (37), which was used in bacterial reactivation assays. For expression and purification of N-terminally 6xHis-tagged recombinant protein in *E. coli*, expression plasmids were constructed by transferring the AlkB-encoding NdeI-BamHI fragments from the pJB658-derived plasmids to pET-28a(+) (Novagen, Darmstadt, Germany).

Recombinant EcAlkB was expressed and purified essentially as previously described (24), whereas the other recombinant bacterial AlkB proteins were obtained by a single affinity purification step (27).

Phage reactivation assay

The experiments were performed essentially as previously described (24). Plasmid pJB658 carrying a gene encoding a bacterial AlkB protein was transfected into the F-pilus-expressing, AlkB-deficient *E. coli* strain, HK82/F′. Expression of recombinant protein was induced by the addition of 2 mM toluic acid (Fluka/Sigma-Aldrich, St. Louis, MO) when the bacterial cultures had reached $A_{600} = 0.1$, after which they were further grown until $A_{600} = 0.8$.

DNA phage M13 or RNA phage MS2 was treated for 30 min at 30°C with different concentrations of methyl

methanesulphonate (MMS; Sigma-Aldrich) or chloroacetaldehyde (CAA; Sigma-Aldrich) to introduce methyl or etheno adducts, respectively. One hundred microliters of various dilutions of the treated phage was mixed with 300 µl of induced bacteria and 3 ml LB top agar, and plated onto LB plates that were placed overnight at 37°C. Phage survival was scored by counting the resulting plaques. All treatments and dilutions of the phages were performed in M9 minimum salt medium.

Bacterial survival assay

Cultures of *E. coli* strain HK82/F′ transformed with pJB658-derived plasmids encoding bacterial AlkB proteins were induced with toluic acid as described above and grown until $A_{600} = 0.5$. Bacteria in 500 µl of the culture were pelleted, resuspended in 250 µl of M9 minimum salt medium, and then mixed with an equal volume of M9 minimum salt medium containing the appropriate MMS concentration, followed by 30 min incubation at 30°C. Treated bacteria were diluted 10 000-fold and plated onto LB plates and incubated overnight at 37°C.

Assay for oxidative demethylation of [^{3}H]methylated ssDNA oligonucleotides

The [^{3}H]methylated ssDNA oligonucleotide (sequence: TAAAATAATAAATTAAA) was prepared by methylation with *N*-[^{3}H]methyl-*N*-nitrosourea as described previously (7). [^{3}H]methylated ssDNA was incubated with 100 pmol of recombinant, bacterial AlkB for 30 min at 37°C in a 50 µl reaction mixture containing 50 mM HEPES–KOH pH 7.5, 2 mM ascorbic acid, 100 µM 2-oxoglutarate, 50 µg/ml BSA and 40 µM $FeSO_4$. The released ethanol-soluble radioactivity was measured by scintillation counting as previously described (7).

In vitro repair of oligonucleotides containing site-specific lesions

DNA oligonucleotides (sequence: TAGACATTGCCATT CTCGATAGGATCCGGTCAAACCTAGACGAATTC CG) containing $\overline{m^1A}$, m^3C or 1,N^6-ethenoadenine (εa) at a specific position (underlined) were synthesized by ChemGenes, Wilmington, MA (m^1A and m^3C) and Midland Certified Reagent Company Inc., Midland, TX (εA). Corresponding 5′-[^{32}P]-labelled double-stranded DNA (dsDNA) substrates were generated as previously described (38,39). Repair reactions were performed essentially as described (39). Briefly, 100 pmol of bacterial AlkB was incubated with 16.6 nM m^1A, m^3C or (εa-containing 5′-[^{32}P]-labelled DNA in 50 mM Tris–HCl pH 8.0, 2 mM ascorbic acid, 100 µM 2-oxoglutarate, and 40 µM $FeSO_4$ in a final volume of 50 µl at 37°C for 30 min. Reactions were terminated by the addition of 18 mM EDTA, 0.4% SDS and 0.4 mg/ml proteinase K, and incubated at 37°C for 30 min. The DNA was precipitated with ethanol, dissolved in H_2O, and then subjected to digestion with 20 U of DpnII for 1 h at 37°C. The reaction products were separated by 20% denaturing PAGE, and visualized by phosphor imaging (Molecular Dynamics, Sunnyvale, USA).

RESULTS

Sequence analysis of bacterial AlkB proteins

We have previously generated a limited sequence phylogeny of bacterial AlkB proteins, based on ~60 protein sequences retrieved by PSI-BLAST searches, and divided these proteins into two major groups, denoted 1A and 1B, as well as the minor, heterogeneous group 2 (40). Since then, the number of sequenced bacterial genomes has increased considerably, reflected by the ~360 protein sequences retrieved through a new PSI-BLAST search. The phylogenetic tree based on an updated sequence alignment of bacterial AlkB proteins indicated that group 2 could be more appropriately subdivided into groups 2A and 2B (Figure 1). Group 2A represents an outgroup, while group 2B appears to be related to group 1B.

The recently published three-dimensional structures of EcAlkB in complex with various substrates have identified key residues involved in substrate recognition, and showed that the 216 aa protein consists of the following three domains: an N-terminal extension (aa 1–45) followed by a 'nucleotide-recognition lid' (NRL; aa 46–89), and a C-terminal dioxygenase domain, 'the catalytic core' (aa 90–216) (41,42). The dioxygenase domain is found in all members of the AlkB family of proteins and several conserved residues/motifs within this domain define the AlkB family of proteins (18). The NRL domain of EcAlkB is involved in binding the nucleic acid substrate, and appears to be unique to the AlkB proteins repairing DNA or RNA (42). The bacterial AlkBs are quite uniform in size (typically from 190 to 220 aa), and generally consist of a conserved dioxygenase domain and a less conserved N-terminal region (Supplementary Figure S2). We reasoned that this N-terminal part may contain key determinants of substrate specificity. To reveal possible similarities in this region, iterative PSI-BLAST searches were performed with the N-terminal part of different AlkB proteins as the initial query. It was found that group 1B members were retrieved in searches where this region from group 2B protein was the initial query, and vice versa. Accordingly, groups 1B and 2B proteins shared several conserved amino acids in the NRL domain (Figure 2A). When similar searches were performed with proteins from groups 1A and 2A, only proteins from the same group were retrieved. Although the iterative BLAST searches did not reveal sequence similarity between the N-terminal parts of the 1A and 1B proteins, they still appeared to share some conserved key residues in their NRL domain (Figure 2A), including residues of EcAlkB (Trp69 and Tyr76) that have been implicated in substrate binding (42). Notably, the regions of group 1B proteins displaying homology to group 1A or 2B proteins were non-overlapping and adjacent (Figure 2A). The observation that the NRL domain of group 1B proteins displayed sequence homology to both 1A and 2B proteins suggested to us that proteins from these three groups may recognize similar substrates.

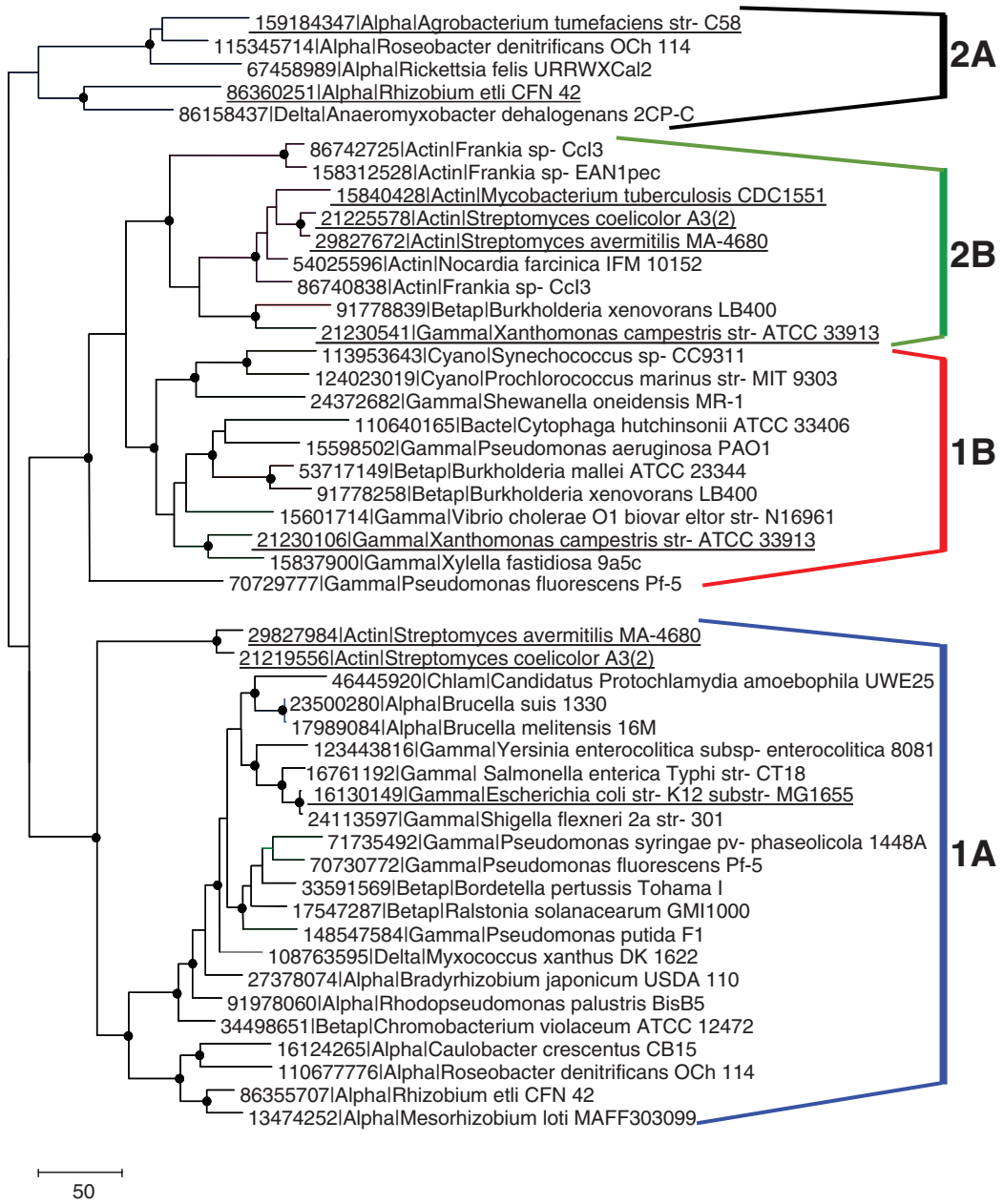

Figure 1. Phylogenetic tree of selected bacterial AlkB proteins. Dots indicate tree nodes with bootstrap support ≥70%. Numbers represent GenBank Identifier (gi) numbers. Proteins included in the experimental part of the present work have been underlined, and the four identified subgroups are indicated. Actin, Actinobacteria; Alpha, Alphaproteobacteria; Bacte, Bacteroidetes; Betap, Betaproteobacteria; Chlam, Chlamydiae; Cyano, Cyanobacteria; Delta, Deltaproteobacteria; Gamma, Gammaproteobacteria. The scale bar represents the number of substitutions per 100 positions. The alignment on which the tree was based is found in Supplementary Figure S1, and a more comprehensive alignment of bacterial AlkB proteins is shown in Supplementary Figure S2.

Sequence comparison of bacterial and eukaryotic AlkB proteins

ALKBH2 and ALKBH3, which are functional vertebrate homologues of EcAlkB, display the highest sequence similarity to bacterial AlkB proteins from group 1B (40). Indeed, ALKBH2, ALKBH3 and group 1B bacterial AlkBs share numerous conserved residues in the region corresponding to the NRL domain of EcAlkB (Figure 2B). Most of these residues are also shared between group 1B proteins and members of either group 1A or 2B (Figure 2; compare panels A and B), further supporting our hypothesis that proteins from these three groups may act on similar nucleic acid substrates. Notably, the NRL residues that in a recently published crystal structure of human ALKBH2 were shown to be in close contact with a m^1A-containing dsDNA substrate, i.e. Arg110, Tyr122, Phe124 and Ser125 (41), are all among these conserved, shared residues (Figure 2B). Moreover, Phe102 in ALKBH2, which was shown to be

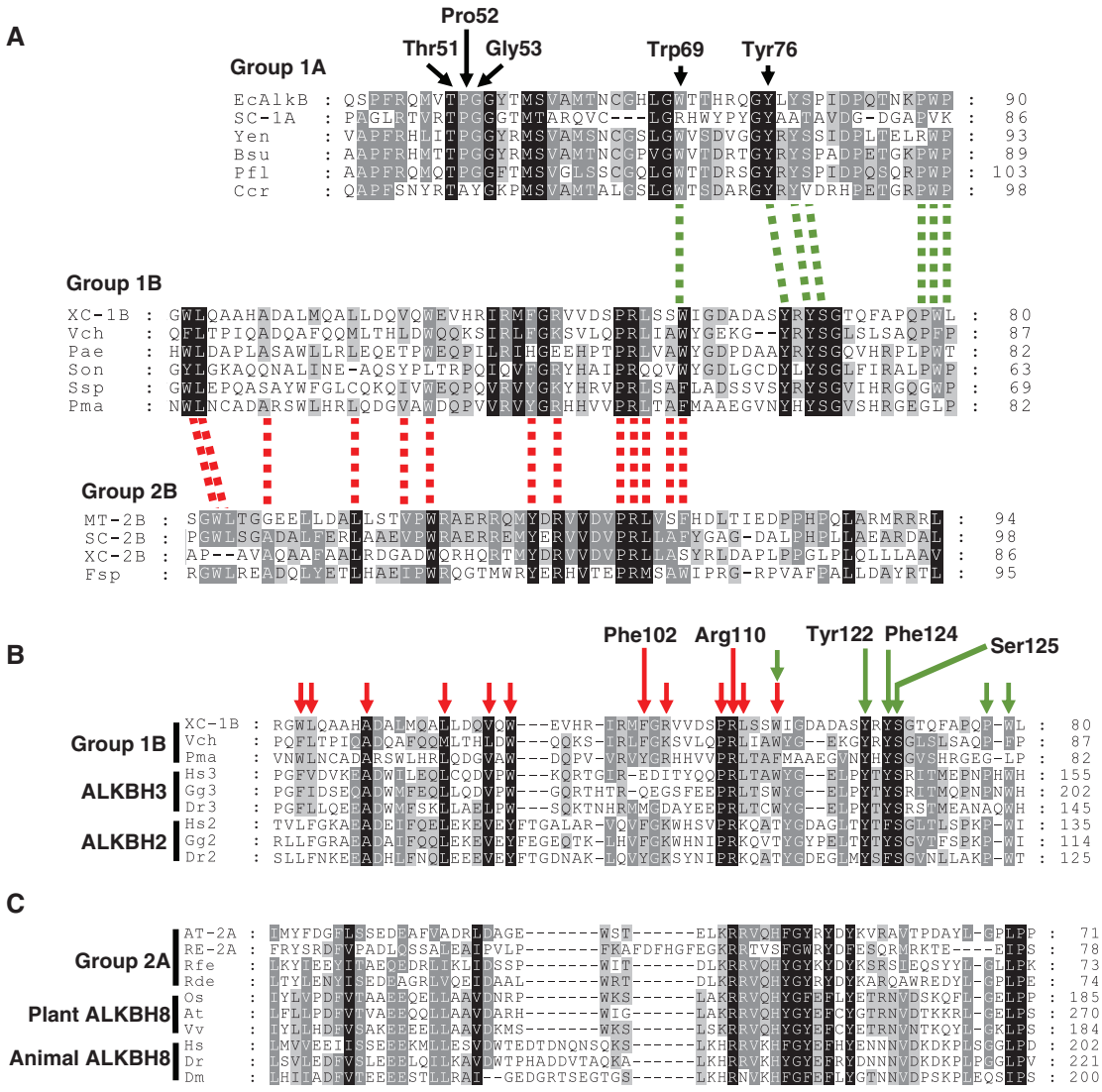

Figure 2. Sequence alignments of the region encompassing the putative nucleotide-recognition lid (NRL) of various AlkB proteins. (**A**) Alignments of the NRL regions of AlkB of selected members of group 1A, 1B or 2B. Thr51, Pro52, Gly53, Trp69 and Tyr76, which in the three-dimensional structure of EcAlkB were shown to interact with the DNA substrate, are indicated. Dashed lines indicate partially or fully conserved residues that are shared between group 1B members and those from groups 1A (green) or 2B (red). (**B**) Alignment of the NRL region of group 1B bacterial proteins and vertebrate ALKBH2 and ALKBH3. Arrows indicate residues that are conserved between group 1B bacterial AlkB proteins and ALKBH2 or ALKBH3, as well as between group 1B proteins and proteins from groups 2B (red) or 1A (green). Indicated are the residues Phe102, Arg110, Tyr122, Phe124 and Ser125, which in the three-dimensional structure of human ALKBH2 were shown to interact with the DNA substrate. (**C**) Alignment of the NRL region of group 2A bacterial proteins and ALKBH8 from multicellular eukaryotes. Yen, *Yersinia enterocolitis*; Bsu, *Brucella suis*; Pfl, *Pseudomonas fluorescens*; Ccr, *Caulobacter crescentus*; Vch, *Vibrio cholerae*; Pae, *Pseudomonas aeruginosa*; Son, *Shewannella oneidensis*; Ssp, *Synechococcus sp. CC9311*; Pma, *Prochlorococcus marinus MIT 9303*; Fsp, Frankia sp. CcI3; Hs, *Homo sapiens*; Gg, *Gallus gallus*; Dr, *Danio rerio*; Dm, *Drosophila melanogaster*; Rfe, *Rickettsia felis*; Rde, *Roseobacter denitrificans*; Os, *Oryza sativa*; At, *Arabidopsis thaliana*; Vv, *Vitis vinifera*. Proteins included in the experimental part of this work are indicated by the abbreviations introduced in Table 2 (SC-1A, XC-1B, etc.).

important for flipping out the damaged base from the dsDNA structure (41), is conserved as an aromatic residue in ALKBH2 and group 1B proteins, but not in ALKBH3 proteins, which are inactive on dsDNA substrates (Figure 2B), suggesting that group 1B proteins may also be active on dsDNA.

Among bacterial AlkB proteins, those from group 2A appeared as a small outgroup (Figure 1), whose N-terminal part displayed little sequence similarity to proteins belonging to the other three groups. However, putative NRL domain of group 2A proteins displayed substantial homology to a corresponding region in ALKBH8 proteins (Figure 2C), which are found in most multicellular eukaryotes and probably have a function other than DNA repair (see Discussion section).

Distribution of AlkB proteins across bacterial species

Many, but not all bacteria possess an AlkB protein, and an iterative PSI-BLAST search of 547 completely sequenced and annotated bacterial genomes using EcAlkB as query returned 215 bacterial AlkB proteins

Table 1. Distribution of bacterial AlkB proteins in completely sequenced genomes

No. of AlkBs	Group(s)	No. of bacterial sp.	Bacterial genera/species (examples)
1	1A	65	*Escherichia, Salmonella, Bradyrhizobium, Rhodopseudomonas*
1	1B	61	*Shewanella, Burkholderia, Synechococcus, Xylella*
1	2A	6	*Agrobacterium, Rickettsia*
1	2B	12	*Mycobacterium, Nocardia*
1	All	144	
2	1A	1	*Ralstonia metallidurans*
2	1A + 1B	2	*Marinobacter aquaeolei, Pseudomonas fluorescens*
2	1A + 2A	5	*Sinorhizobium, Rhizobium, Roseobacter*
2	1A + 2B	3	*Streptomyces, Rhodococcus*
2	1B	1	*Acaryochloris marina*
2	1B + 2A	0	–
2	1B + 2B	4	*Xanthomonas, Burkholderia xenovorans*
2	2A	0	–
2	2A + 2B	0	–
2	2B	2	*Frankia*
2	All	18	
3	1A(2) + 1B	1	*Burkholderia sp. 383*
3	1B	1	*Flavobacterium johnsoniae*
3	All	2	
Any	All	164	

(as of February 2008), of which 186 remained after removal of the redundancy corresponding to different strains of the same bacterial species. Of these, 144 proteins were found in bacteria possessing a single AlkB protein, whereas the rest were found in species having two (18 species; 36 proteins) or three (2 species; 6 proteins) AlkB proteins (Table 1). In our previous bioinformatics study of bacterial AlkB proteins, we noted that several of the group 2 proteins were found in bacteria possessing two AlkB proteins, in most cases one from group 1 and one from group 2 (40). This still holds partially true with the considerably higher number of genomes analyzed here; of 34 group 2 proteins, 12 are found in bacteria that also have a group 1 AlkB protein. However, no particular combination of two AlkBs is particularly frequent, and seven of the ten possible combinations (1A + 1A, 1A + 1B, 1A + 2A, etc.) are actually found (Table 1).

To verify if the assignment of bacterial AlkB proteins to four different groups also holds true in a natural microbial community, we analyzed the AlkB sequences from the NCBI environmental sample database, mainly consisting of marine samples from the Sargasso Sea (43). A PSI-BLAST search retrieved 143 protein sequences, and after reducing redundancy (so that no sequence was >95% identical to any other sequence) and removing sequences of apparent eukaryotic origin, 86 sequences remained. These sequences showed a distribution between the four groups (1A:1B:2A:2B = 23:50:6:7) which was not dramatically different from that found when studying completely sequenced bacteria (Table 1; 1A:1B:2A:2B = 79:73:11:23).

To relate the distribution of bacterial AlkBs to bacterial phylogeny, a phylogenetic tree of selected bacterial species was constructed based on 16S ribosomal RNA sequences, and the distribution of AlkB proteins across the four groups has been indicated on this tree (Figure 3). As previously reported (40), bacteria from certain phyla and subdivisions, such as firmicutes, spirochaetes and ε-proteobacteria appear to lack AlkB proteins altogether. Members of some bacterial phyla tend to possess AlkBs from certain groups; actinobacteria usually have group 2B AlkBs, most cyanobacteria possess AlkBs from group 1B, whereas α-proteobacteria usually contain an AlkB protein from group 1A, and in some cases also one from group 2A. On the other hand, a more heterogeneous distribution of AlkB proteins is found within the β- and γ-subdivisions of proteobacteria, and there are several examples that closely related bacteria possess AlkB proteins from different groups. For example, in the genus *Pseudomonas,* species exist that have AlkBs from group 1A (*P. putida and P. syringae*), group 1B (*P. aeruginosa*), as well as a combination of the two (*P. fluorescens*). Also, a group 1A AlkB is found in *Yersinia enterocolitica*, while no AlkB encoding gene is present in *Yersinia pestis* (Figure 3).

In vivo repair of methylated ssDNA

For the experimental studies, AlkB proteins were selected from three bacterial species that have two such proteins and where genomic DNA was available through the ATCC, i.e. *Xanthomonas campestris* (1B and 2B), and *Streptomyces coelicolor* and *Streptomyces avermitilis* (both 1A and 2B). The six respective proteins were designated accordingly; XC-1B, XC-2B, SC-1A, SC-2B, SA-1A and SA-2B (Table 2). In addition, we included a group 2A protein from *Agrobacterium tumefaciens*, AT-2A, a group 2A protein from *Rhizobium etli*, RE-2A (*R. etli* also has a group 1A protein) and a group 2B protein from *Mycobacterium tuberculosis*, MT-2B (Table 2).

To study the *in vivo* repair activity of the various bacterial AlkB proteins, AlkB-deficient (AlkB$^-$) *E. coli* was transformed with expression plasmids encoding bacterial AlkBs, and the effect of exposing bacterial or bacteriophage genomes to alkylating agents was assessed. When the ssDNA bacteriophage M13 is exposed to methylmethanesulfonate (MMS), substantial amounts of the AlkB substrates m^1A and m^3C are introduced, and a dramatically lower number of progeny phage is obtained with AlkB$^-$ *E. coli* relative to wild-type *E. coli*, due to inefficient repair of replication blocking lesions (44). This repair deficiency can be complemented by expressing a functional AlkB protein (13,24), and we found that all groups 1A and 1B proteins, (SA-1A, SC-1A and XC-1B) as well as most of the group 2B proteins (MT-2B, SA-2B, SC-2B) displayed a complementing ability similar to, or slightly lower than that of EcAlkB (Figure 4A). No complementation was observed for RE-2A and XC-2B. The AT-2A protein was inactive in this, as well as in subsequent, complementation assays (data not shown), and since AT-2A also was insoluble when expressed in *E. coli*, the negative results on this protein have been omitted from figures and tables.

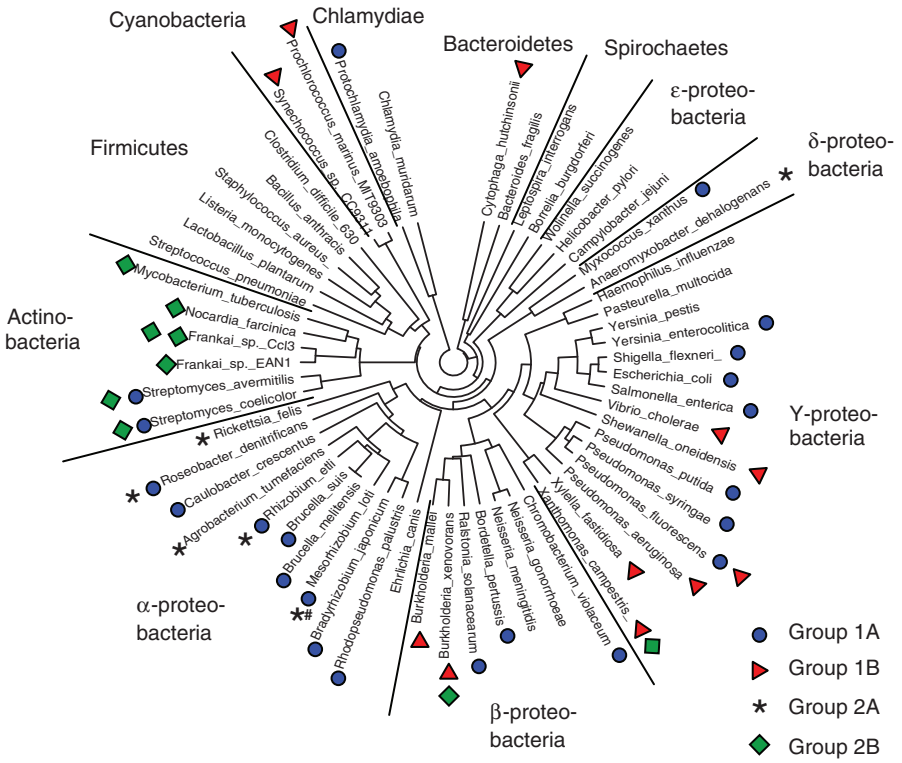

Figure 3. Distribution of AlkB proteins among bacterial species. A phylogenetic tree of selected bacterial species, including all species represented in Figure 1, was generated based on 16S ribosomal RNA sequences. The distribution of AlkB proteins from the four groups has been indicated. Additionally, AlkB homologs are found in the following lineages: Acidobacteria (*Solibacter usitatus*), Planctomycetes (*Gemmata obscuriglobus*), Spirochaetes (*Leptospira biflexa*), GNS (green non-sulfur bacteria) bacteria (*Dehalococcoides sp-VS*), Chlamydiae/Verrucomicrobia group (*Verrucomicrobium spinosum*). #This protein is only annotated as a partial AlkB sequence (92 aa) in the NCBI protein database, and has therefore not been included in Figure 1.

Table 2. Bacterial AlkB proteins used in present study

Abbreviation	Origin	Accession	Gi number	Size (aa)
AT-2A[a]	Agrobacterium tumefaciens	NP_353535	15887854	209
MT-2B	*Mycobacterium tuberculosis*	NP_335465	15840428	205
RE-2A	*Rhizobium etli*	YP_472140	86360251	189
SA-1A	*Streptomyces avermitilis*	NP_822618	29827984	222
SA-2B	*Streptomyces avermitilis*	NP_822306	29827672	208
SC-1A	*Streptomyces coelicolor*	NP_625335	21219556	216
SC-2B	*Streptomyces coelicolor*	NP_631357	21225578	210
XC-1B	*Xanthomonas campestris*	NP_636023	21230106	201
XC-2B	*Xanthomonas campestris*	NP_636458	21230541	211
EcAlkB	*Escherichia coli*	NP_754639	26248599	216

[a]When this study was initiated the AT protein was annotated as indicated, but this protein has now been replaced by NP_353535.2 [gi:159184347], representing translation initiation 42 nucleotides downstream, resulting in a protein of 195 aa.

These experiments indicate that most 1A/1B proteins, as well as several 2B proteins, exhibit a repair function similar to that of EcAlkB.

In vivo repair of etheno lesions in ssDNA

Etheno (ε) lesions on DNA bases are characterized by a C = C bridge connecting the exocyclic and heterocyclic nitrogen atoms of adenine, cytosine, or guanine. Such adducts are both mutagenic and cytotoxic, and can arise from endogenous lipid peroxidation products, as well as from environmental carcinogens, such as vinyl chloride. Cytochrome P450 enzymes can convert vinyl chloride to the reactive secondary metabolite chloroacetaldehyde (CAA) (45), which is also frequently used as an experimental tool for introducing ε-adducts *in vitro*. The main lesions induced by treatment of DNA with CAA are 1,N^6-ethenoadenine (εA), 3,N^4-ethenocytosine (εC), 1,N^2-ethenoguanine (1,N^2-εG), and N^2,3-ethenoguanine (N^2,3-εG) with the following

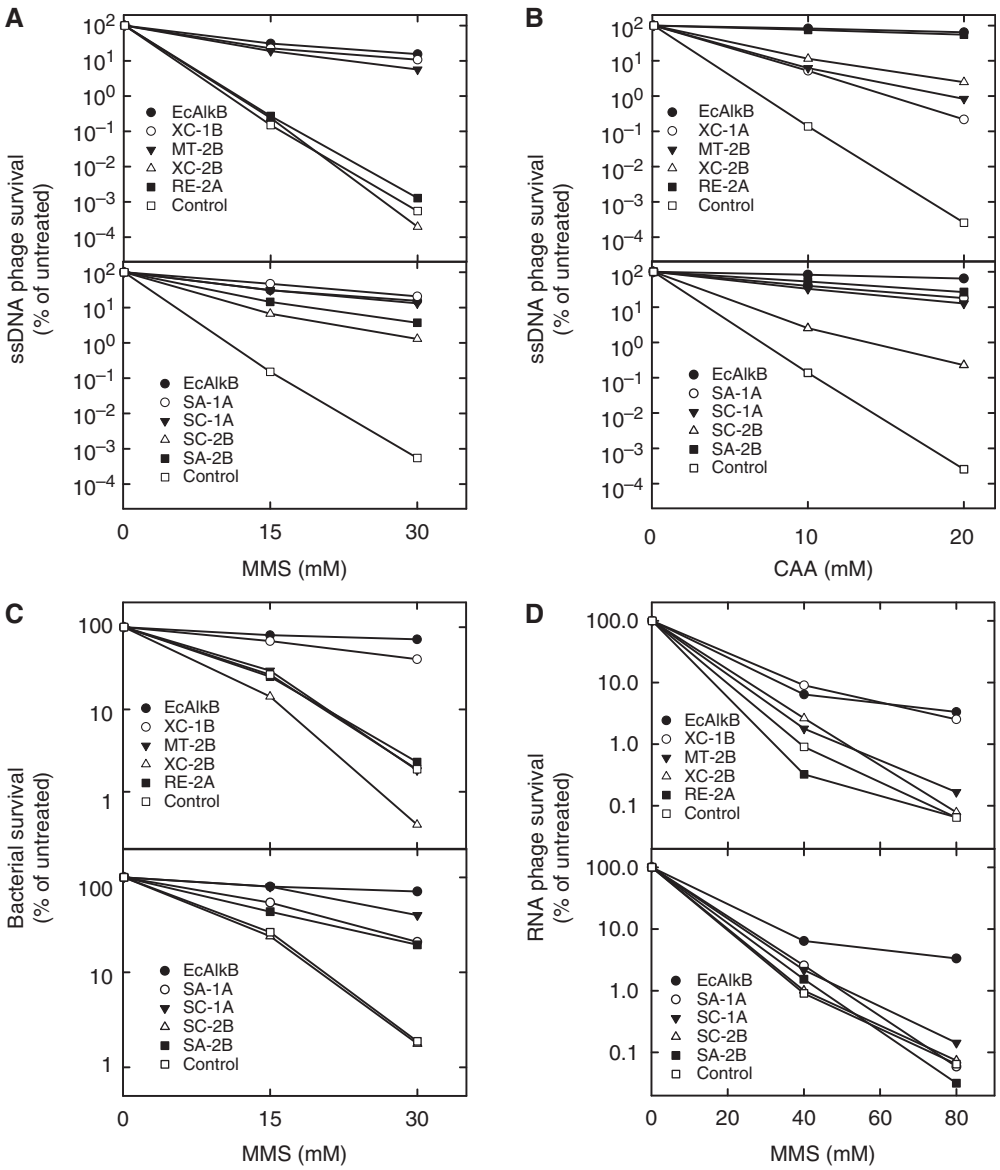

Figure 4. *In vivo* repair of DNA/RNA damage in *E. coli* by bacterial AlkB proteins. (**A**) Repair of methyl lesions in ssDNA. AlkB$^-$ *E. coli* expressing the indicated bacterial AlkB proteins were infected with ssDNA bacteriophage M13 that had been treated with the indicated concentrations of MMS. The results are presented as the percentage of plaque numbers obtained with untreated M13. 'Control' represents bacteria transformed with the empty expression plasmid. (**B**) Repair of etheno adducts in ssDNA. Similar experiment as in (A), except that M13 was treated with CAA. (**C**) Ability to complement the MMS-sensitive phenotype of AlkB$^-$ bacteria. AlkB$^-$ *E. coli* expressing the indicated bacterial AlkB proteins was treated with the indicated concentrations of MMS, and the bacterial survival was assessed. (**D**) Repair of methyl lesion in ssRNA. Similar experiment as in (A), except that ssDNA phage M13 was replaced by RNA phage MS2. For all these assays, similar results were obtained in three or more independent experiments.

efficiencies: $\varepsilon A > \varepsilon C > N^2,3\text{-}\varepsilon G > 1,N^2\text{-}\varepsilon G$ (46,47). It has recently been shown that EcAlkB, as well as the mammalian homologues ALKBH2 and ALKBH3, can remove etheno adducts in a reaction where the etheno moiety is released as glyoxal (15,17,39).

To study the ability of the various bacterial AlkB proteins to repair etheno adducts, CAA-treated ssDNA bacteriophage M13 was used to infect AlkB$^-$ *E. coli* expressing the different AlkB proteins and the number of resulting progeny phage was quantified. As expected, expression of EcAlkB did cause a dramatic increase in the number of progeny phage obtained from CAA-treated M13, and similar, albeit in some cases slightly weaker, effects were observed with all the tested proteins (Figure 4B). Notably, RE-2A and XC-2B, which were inactive on methylated M13, showed robust activity on the CAA-treated phage.

Complementation of the methylation sensitive phenotype of AlkB$^-$ *E. coli*

It was next tested whether the various bacterial AlkB proteins could complement the MMS-sensitive phenotype of AlkB$^-$ bacteria. This approach is similar to the ssDNA

phage reactivation experiments above, but lesions are likely to be present also in dsDNA. When AlkB$^-$ *E. coli* expressing different bacterial AlkBs was treated with MMS, all 1A/1B proteins (SA-1A, SC-1A and XC-1B) were found to increase the survival, but only one out of four 2B proteins and none of the 2A proteins had such an effect (Figure 4C). These results suggest that the 1A/1B proteins may be more important than the 2A/2B proteins for protecting bacteria against methylating agents.

In vivo repair of methylated ssRNA

It has been demonstrated that several AlkB proteins can repair lesions in RNA as well as in DNA (24,27). To study whether the different bacterial AlkBs were active on RNA, they were expressed in AlkB$^-$ *E. coli*, which was infected with MMS-treated RNA bacteriophage MS2. Expression of EcAlkB increased the number of progeny from the methylated MS2 phage (Figure 4D), in agreement with previous observations (24). A similar effect was observed for XC-1B, but not for any of the other bacterial AlkB proteins. This suggests that DNA rather than RNA is the preferred substrate for most bacterial AlkB proteins.

Demethylating activity of recombinant proteins on [^{3}H]methylated ssDNA

The nine bacterial AlkBs were expressed in *E. coli* and purified as hexahistidine-tagged proteins (Figure 5A). Judged by their ability to decarboxylate 2OG to succinate, they all appeared to be purified in an active state (Supplementary Figure S3). With the exception of EcAlkB (and possibly XC-1B), the decarboxylation of 2OG by bacterial AlkBs was not stimulated by the ribonucleoside 1-methyladenosine (Supplementary Figure S3), in agreement with the observed inability of these proteins to repair methylated RNA (Figure 4D).

For the next series of experiments, a [^{3}H]m^{1}A-containing AlkB-substrate was generated by treating an A-rich ssDNA oligonucleotide with the radiolabelled methylating agent *N*-[^{3}H]methyl-*N*-nitrosourea, as previously described (7). When this substrate was incubated with the bacterial AlkB proteins, a robust demethylating activity, although somewhat lower than that of EcAlkB, was observed for all proteins, except RE-2A and XC-2B, where only negligible activity was measured (Figure 5B). These results are in very good agreement with those obtained with MMS-treated M13 (Figure 4B), where m^{1}A also represents the major deleterious lesion (7).

Repair activity on DNA substrates containing site-specific lesions

The presence of alkyl lesions in dsDNA can prevent cleavage by certain restriction enzymes. Thus, AlkB-mediated repair of site-specific lesions in dsDNA oligonucleotides may be observed as the regeneration of a functional restriction site (38,48). To further explore the repair activity of the bacterial AlkB proteins, 5′-[^{32}P]-end-labelled oligonucleotides containing specific lesions at a DpnII-site (recognition sequence: GATC), were utilized. DpnII-mediated cleavage of the 49-nt oligonucleotide into a 22-nt end-labelled (and 27-nt unlabelled) fragment will

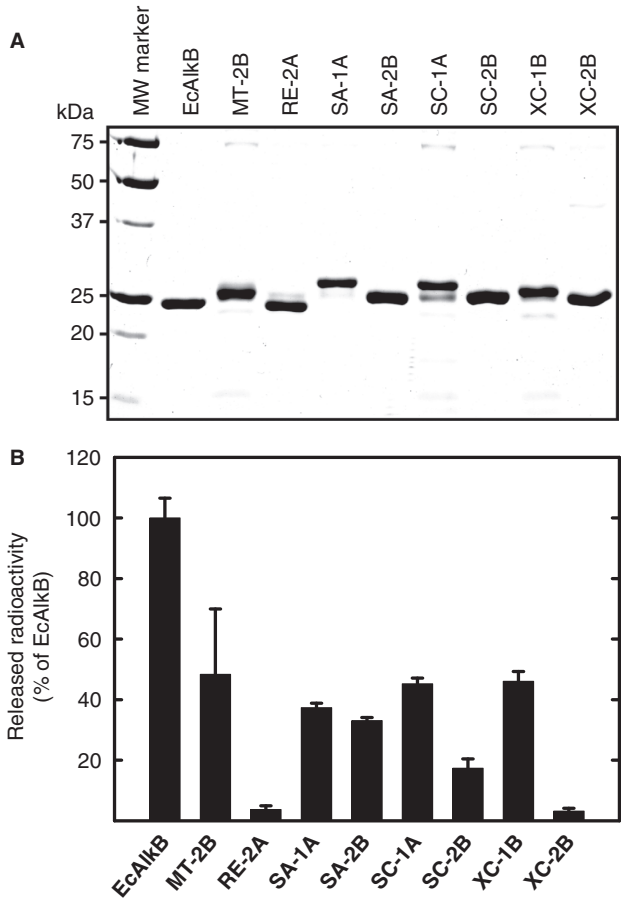

Figure 5. *In vitro* characterization of bacterial AlkB proteins. (**A**) Purified recombinant His-tagged AlkB proteins used in this study. Proteins were visualized by Coomassie staining of a 12% SDS–PAGE gel. (**B**) Activity of AlkB proteins on methylated ssDNA. [^{3}H]methylated oligonucleotides were incubated with 100 pmol of AlkB and the ethanol soluble radioactivity released was measured by scintillation counting. Error bars represent the range between duplicate measurements.

only occur if the lesion has been repaired (Figure 6A). When using a dsDNA oligonucleotide containing an m^{1}A lesion, we could detect repair activity above background levels in the case of EcAlkB, MT-2B and XC-1B (Figure 6B). Strong repair activity towards an m^{3}C-containing oligo was observed for the majority of the proteins, but weak or no activity was observed for RE-2A, SC-2B and XC-2B (Figure 6B). Two of the latter, RE-2A and XC-2B, showed strong repair activity towards εA (Figure 6B), in good agreement with the results of experiments with CAA-treated M13. Some proteins, such as SA-1A and SC-1A, which were able to efficiently reactivate CAA-treated M13 ssDNA *in vivo*, did only show background levels of repair of the εA-containing dsDNA substrate *in vitro*. To elucidate this apparent discrepancy, which could reflect different activities of the enzymes on ssDNA versus dsDNA, an *in vitro* repair experiment was performed on a single-stranded substrate prior to annealing to the complementary strand. The results were very similar to those obtained for εA-containing dsDNA, but a weak activity

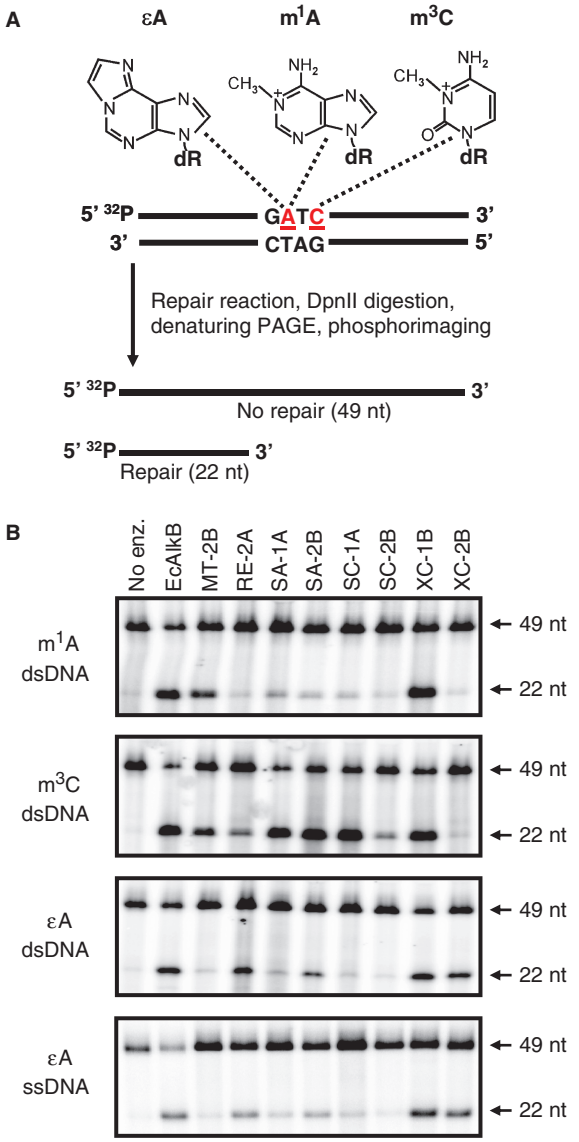

Figure 6. *In vitro* repair of oligonucleotides containing site-specific lesions. (**A**) Schematic representation of the experimental set-up. Double-stranded (ds) ^{32}P-end-labelled 49-nt DNA containing m^1A, m^3C, or εA in one strand of the DpnII restriction site GATC was treated with repair enzyme, subsequently digested with DpnII, and the products were separated by denaturing PAGE and visualized by phosporimaging. To study repair activity on single-stranded DNA, the lesion-containing ^{32}P-end-labelled 49-nt ssDNA substrate was subjected to repair before annealing to the complementary strand. (**B**) Repair activity of the different bacterial AlkB proteins on various dsDNA and ssDNA substrates. ^{32}P-end-labelled substrate was incubated with 100 pmol of the indicated proteins.

of SA-1A and SC-1A on εA-containing ssDNA could indeed be observed.

DISCUSSION

We have here performed a sequence analysis of bacterial AlkB proteins, showing that they can be subdivided into four distinct groups: 1A, 1B, 2A and 2B. We found that proteins from groups 1A and 2B shared several conserved amino acids with the 1B proteins in the region corresponding to the nucleotide recognition lid of EcAlkB, which suggested that these three groups of proteins may act on similar nucleic acid substrates.

Furthermore, we have undertaken a functional characterization of nine different bacterial AlkB proteins, with a particular focus on bacteria possessing two AlkBs. We studied both the *in vivo* repair activity of the proteins when expressed in *E. coli*, and the *in vitro* repair activity of the purified recombinant proteins (summarized in Table 3). In general, there is a good consistency within the dataset obtained for any given protein, and the data convincingly demonstrated that the groups 1A, 1B and 2B proteins all possess DNA repair activity. Also, one of the two tested group 2A proteins, RE-2A, displayed robust repair activity. Intriguingly, RE-2A and XC-2B, which are present in bacteria which also possess an AlkB from groups 1A or 1B, efficiently repaired etheno lesions in DNA, but displayed very low activity towards methyl adducts.

Group 1A

The group 1A proteins are distributed among a wide range of bacteria, including all the AlkB-containing subdivisions of proteobacteria (Figure 3). This group includes the founding member EcAlkB, and we have here studied two other members, SA-1A and SC-1A, which both come from the genus *Streptomyces*, and are 79% identical on the protein sequence level. In accordance with this, the two proteins displayed very similar properties and they displayed robust repair activity in most of the assays. These data strengthen the notion that group 1A consists of proteins with a function similar to that of the founding member EcAlkB.

Group 1B

The group 1B proteins are primarily found in the β- and γ-subdivisions of proteobacteria, as well as in cyanobacteria (Figure 3). These proteins are similar to the ALKBH2/ALKBH3 proteins, which are vertebrate proteins with a similar repair activity as EcAlkB. Accordingly, the only group 1B protein studied here, XC-1B, displayed repair activity in all assays. Our bioinformatics and experimental analysis, therefore, indicates that group 1B proteins are repair proteins with a function similar to that of EcAlkB.

Group 2B

Group 2B AlkB proteins are generally found in actinobacteria (Figure 3). However, they are also found in some, but not all, subspecies of the root-associated plant pathogens *Xanthomonas* (γ-proteobacteria) and *Burkholderia* (β-proteobacteria), which may have acquired their AlkBs by horizontal transfer from actinobacteria, most of which are soil bacteria (49). Protein sequence analysis revealed that the proteins in groups 1B and 2B shared several conserved residues in the NRL region (Figure 2A), implying also the group 2B proteins in DNA repair. Indeed, the investigated group 2B proteins, MT-2B, SA-2B, SC-2B and XC-2B, displayed

Table 3. Summary of experiments

Protein	*In vivo* experiments				*In vitro* experiments				
	ssDNA (MMS)	ssDNA (CAA)	Survival (MMS)	ssRNA (MMS)	[^{3}H]Me-ssDNA	dsDNA - m^{1}A	dsDNA - m^{3}C	dsDNA - εA	ssDNA - εA
SA-1A	++	++	+	–	+	–	++	–	+
SC-1A	++	++	++	–	+	–	++	–	+
XC-1B	++	+	++	++	+	++	++	++	++
RE-2A	–	++	–	–	–	–	+	++	++
MT-2B	++	+	–	–	+	++	++	–	–
SA-2B	++	++	+	–	+	–	++	+	+
SC-2B	++	+	–	–	+	–	+	–	–
XC-2B	–	++	–	–	–	–	–	++	++

++, activity comparable to, or slightly lower than, that of EcAlkB; +, substantially lower activity than EcAlkB; –, no detectable activity.

repair activity in most of the assays. However, the substrate specificity of these proteins was quite diverse, ranging from a protein (MT-2B) with high activity towards methyl lesions and low activity on etheno adducts, to the opposite (XC-2B).

Group 2A

The group 2A proteins are primarily found in the α-subdivision of proteobacteria, such as *Agrobacterium*, *Rickettsia* and *Rhizobium* species (Figure 3). Several of these bacteria are plant pathogens that invade their hosts, and exhibit substantial DNA exchange with the host cell. In addition, *Roseobacter denitrificans*, which also has a group 2A AlkB protein, is a marine bacterium which populates the surfaces of green seaweeds (50). Interestingly, group 2A AlkBs from several bacteria (*R. denitrificans, R. etli, Sinorhizobium meliloti, Sinorhizobium medicae*) reside on plasmids rather than on the chromosome, suggesting that they may have been acquired from the host genome.

The group 2A AlkBs share strong homology with the ALKBH8 proteins from plants and animals, which, by bioinformatics analysis, have been implicated in tRNA modification (22). This may suggest that group 2A proteins are involved in tRNA modification rather than in repair, in agreement with our inability to measure any repair activity for AT-2A, but in apparent disagreement with the observed activity of RE-2A on etheno adducts. However, compared to AT-2A, RE-2A is considerably more distantly related to the eukaryotic ALKBH8 proteins (Figure 2C), and this protein may, conceivably, have evolved from a tRNA modification enzyme into a repair protein. Unfortunately, due to its insolubility when expressed in *E. coli*, we were not able to investigate the enzymatic activity of AT-2A.

Distribution of bacterial AlkB proteins

Our phylogenetic analysis clearly showed that the phylogeny of bacterial AlkB protein sequences is very different from that of the bacterial 16S RNA sequences, indicating that a substantial degree of horizontal transfer of bacterial AlkB genes has occurred. In particular, several bacterial genera exist in which closely related species possess very different AlkB proteins.

The somewhat patchy distribution of AlkB proteins across bacterial species is similar to what is observed in the case of plant-infecting RNA viruses (51). It has been suggested that viruses exposed to a particularly high exogenous load of methylation damage may selectively have acquired AlkBs (51). Similarly, bacteria exposed to relatively high doses of methylating agents may profit from possessing the AlkB function, whereas bacteria experiencing less methylation damage may have disposed of this function, or never acquired it, resulting in an uneven distribution of AlkBs among bacterial species.

The lack of an AlkB function in many bacteria may also be explained by the fact that certain bacteria have translesion DNA polymerases that can efficiently bypass AlkB-substrates, such as m^{1}A and m^{3}C, or that these lesions are repaired through other cellular functions. Indeed, it was recently found that the DNA alkylbase glycosylase AlkA from the archaeon *Archaeoglobus fulgidus*, which was active on typical AlkA substrates such as 3-methyladenine (52), also acted on the AlkB substrates m^{1}A and m^{3}C (53). Generally, archaea do not possess AlkB proteins and it may be that all bacteria lacking AlkB compensate for this by possessing a DNA alkylbase glycosylase active on m^{1}A and m^{3}C. Another possible explanation for the apparent lack of AlkB-encoding sequences in some bacteria is that functional AlkB homologues may exist that are not readily identifiable by current bioinformatics approaches.

The O_2-dependent AlkB mechanism will not be applicable in obligate anaerobic bacteria (23). Thus, AlkB genes appear to be absent from typical anaerobes, e.g. *Clostridium, Fusobacterium, Desulfitobacterium, Bacteroides, Bifidobacterium*, *Propionobacterium* and *Thermotoga ssp*.

Function of bacterial AlkB proteins

In *E. coli*, the N-terminal part of the Ada protein (AdaA) is an alkylation sensitive transcription factor which mediates the up-regulation of AlkB and other alkylation repair proteins in an adaptive response to alkylation damage (54). Moreover, Ada and AlkB are in *E. coli* part of the same operon, being encoded by a common

transcript. To investigate the possible involvement of AdaA in regulating the expression of the various bacterial AlkB proteins, we first found, by BLAST searches, that the genomes of *M. tuberculosis, R. etli, S. avermitilis S. coelicolor,* and *X. campestris* all encoded an AdaA homologue (data not shown). However, *adaA* did not form an operon with *alkB* in these bacteria, but rather (in all cases except *R. etli*) with the genes encoding the alkylbase glycosylase AlkA and DNA alkyltransferase Ogt. Moreover, the *alkB* genes were generally not localized adjacent to genes implicated in alkylation repair (data not shown). Thus, the possible role of AdaA in regulating the expression of these AlkB proteins may best be addressed by studying the inducibility of the relevant genes in response to various alkylating agents.

The results of the present study indicate that most bacterial AlkBs have low activity on RNA, and that AlkB-mediated RNA repair is rare, if not absent, in bacteria. This is also supported by the observation that the activity of EcAlkB on RNA substrates is ~10-fold lower than on DNA (24,25). Bacterial mRNAs are turned over very rapidly and may not acquire substantial methylation damage during their life-time, thus obviating the need for repair. We rather favor the idea that AlkB-mediated RNA repair may be important in the case of RNAs with a considerably longer lifespan, such as the genomes of RNA viruses or long-lived mRNA species in higher eukaryotes (26,27).

Due to the mutagenicity and carcinogenicity of etheno adducts, their formation and repair has been subject of extensive studies (55). Several different repair mechanisms have been shown to target etheno adducts, but these also act on additional substrates, thus making it hard to address the significance of the observed activity. Indeed, EcAlkB and the mammalian homologue ALKBH2, have similar or stronger activity on m^1A than on etheno adducts, and the alkylbase DNA glycosylases AlkA (*E. coli*) and AAG (mammals) prefer 3-methyladenine over etheno lesions. Thus, two of the proteins studied here, RE-2A and XC-2B, represents the first alkylation repair enzymes which are active on etheno adducts, but with low or no activity on methylated bases.

CONCLUSIONS

Our experiments firmly demonstrate that most, if not all, bacterial AlkB proteins are DNA repair enzymes. Moreover, they indicate that some AlkB proteins preferentially target etheno adducts, and that the proteins within each group do not represent an identical repair function. Conceivably, AlkB proteins have the ability to evolve different specificities, governed by the type and amount of damaging agents encountered by the individual bacteria.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The [^{3}H]methylated ssDNA AT-oligonucleotide was a kind gift from Hanne Korvald and Ingrun Alseth. The plasmids pJB658-EcAlkB and pET28-EcAlkB were kindly provided by Hans Krokan. The authors are thankful to T. Tønjum and V. González for providing bacterial genomic DNA. We are grateful to Torbjørn Rognes and Ophélie Aussedat for assistance with bioinformatics analysis.

FUNDING

FRIBIOMOL and FUGE programs of the Research Council of Norway and the Intramural Research Program of the National Library of Medicine at the National Institutes of Health. Funding for open access charge: National Institutes of Health Intramural Funds.

Conflict of interest statement. None declared.

REFERENCES

1. Keppler,F., Eiden,R., Niedan,V., Pracht,J. and Scholer,H.F. (2000) Halocarbons produced by natural oxidation processes during degradation of organic matter. *Nature*, **403**, 298–301.
2. Redeker,K.R., Wang,N., Low,J.C., McMillan,A., Tyler,S.C. and Cicerone,R.J. (2000) Emissions of methyl halides and methane from rice paddies. *Science*, **290**, 966–969.
3. Rydberg,B. and Lindahl,T. (1982) Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *EMBO J.*, **1**, 211–216.
4. Taverna,P. and Sedgwick,B. (1996) Generation of an endogenous DNA-methylating agent by nitrosation in *Escherichia coli*. *J. Bacteriol.*, **178**, 5105–5111.
5. Drablos,F., Feyzi,E., Aas,P.A., Vaagbo,C.B., Kavli,B., Bratlie,M.S., Pena-Diaz,J., Otterlei,M., Slupphaug,G. and Krokan,H.E. (2004) Alkylation damage in DNA and RNA–repair mechanisms and medical significance. *DNA Repair (Amst)*, **3**, 1389–1407.
6. Sedgwick,B. (2004) Repairing DNA-methylation damage. *Nat. Rev. Mol. Cell Biol.*, **5**, 148–157.
7. Falnes,P.O., Johansen,R.F. and Seeberg,E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature*, **419**, 178–182.
8. Trewick,S.C., Henshaw,T.F., Hausinger,R.P., Lindahl,T. and Sedgwick,B. (2002) Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature*, **419**, 174–178.
9. Aravind,L. and Koonin,E.V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.*, **2**, RESEARCH0007.
10. Delaney,J.C. and Essigmann,J.M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **101**, 14051–14056.
11. Falnes,P.O. (2004) Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Res.*, **32**, 6260–6267.
12. Koivisto,P., Robins,P., Lindahl,T. and Sedgwick,B. (2004) Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J. Biol. Chem.*, **279**, 40470–40474.
13. Duncan,T., Trewick,S.C., Koivisto,P., Bates,P.A., Lindahl,T. and Sedgwick,B. (2002) Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl Acad. Sci. USA*, **99**, 16660–16665.
14. Koivisto,P., Duncan,T., Lindahl,T. and Sedgwick,B. (2003) Minimal methylated substrate and extended substrate range of *Escherichia coli* AlkB protein, a 1-methyladenine-DNA dioxygenase. *J. Biol. Chem.*, **278**, 44348–44354.

15. Delaney,J.C., Smeester,L., Wong,C., Frick,L.E., Taghizadeh,K., Wishnok,J.S., Drennan,C.L., Samson,L.D. and Essigmann,J.M. (2005) AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nat. Struct. Mol. Biol.*, **12**, 855–860.
16. Frick,L.E., Delaney,J.C., Wong,C., Drennan,C.L. and Essigmann,J.M. (2007) Alleviation of 1,N6-ethanoadenine genotoxicity by the *Escherichia coli* adaptive response protein AlkB. *Proc. Natl Acad. Sci. USA*, **104**, 755–760.
17. Mishina,Y., Yang,C.G. and He,C. (2005) Direct repair of the exocyclic DNA adduct 1,N6-ethenoadenine by the DNA repair AlkB proteins. *J. Am. Chem. Soc.*, **127**, 14594–14595.
18. Kurowski,M.A., Bhagwat,A.S., Papaj,G. and Bujnicki,J.M. (2003) Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC. Genomics*, **4**, 48.
19. Gerken,T., Girard,C.A., Tung,Y.C., Webby,C.J., Saudek,V., Hewitson,K.S., Yeo,G.S., McDonough,M.A., Cunliffe,S., McNeill,L.A. *et al.* (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*, **318**, 1469–1472.
20. Pan,Z., Sikandar,S., Witherspoon,M., Dizon,D., Nguyen,T., Benirschke,K., Wiley,C., Vrana,P. and Lipkin,S.M. (2008) Impaired placental trophoblast lineage differentiation in Alkbh1(-/-) mice. *Dev. Dyn.*, **237**, 316–327.
21. Westbye,M.P., Feyzi,E., Aas,P.A., Vagbo,C.B., Talstad,V.A., Kavli,B., Hagen,L., Sundheim,O., Akbari,M., Liabakk,N.B. *et al.* (2008) Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J. Biol. Chem.*, **283**, 25046–25056.
22. Falnes,P.O., van den Born,E. and Meza,T.J. (2009) Demethylation of DNA and RNA by AlkB Proteins. In Grosjean,H. (ed.), *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Bioscience, Austin, USA.
23. Sedgwick,B., Bates,P.A., Paik,J., Jacobs,S.C. and Lindahl,T. (2007) Repair of alkylated DNA: recent advances. *DNA Repair (Amst)*, **6**, 429–442.
24. Aas,P.A., Otterlei,M., Falnes,P.O., Vagbo,C.B., Skorpen,F., Akbari,M., Sundheim,O., Bjoras,M., Slupphaug,G., Seeberg,E. *et al.* (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature*, **421**, 859–863.
25. Falnes,P.O., Bjoras,M., Aas,P.A., Sundheim,O. and Seeberg,E. (2004) Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res.*, **32**, 3456–3461.
26. Ougland,R., Zhang,C.M., Liiv,A., Johansen,R.F., Seeberg,E., Hou,Y.M., Remme,J. and Falnes,P.O. (2004) AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Mol. Cell*, **16**, 107–116.
27. van den Born,E., Omelchenko,M.V., Bekkelund,A., Leihne,V., Koonin,E.V., Dolja,V.V. and Falnes,P.O. (2008) Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic Acids Res.*, **36**, 5451–5461.
28. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
29. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
30. Hasegawa,M., Kishino,H. and Saitou,N. (1991) On the maximum likelihood method in molecular phylogenetics. *J. Mol. Evol.*, **32**, 443–445.
31. Jones,D.T., Taylor,W.R. and Thornton,J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.*, **8**, 275–282.
32. Waddell,P.J., Kishino,H. and Ota,R. (2002) Very fast algorithms for evaluating the stability of ML and Bayesian phylogenetic trees from sequence data. *Genome Inform.*, **13**, 82–92.
33. Katoh,K., Misawa,K., Kuma,K. and Miyata,T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059–3066.
34. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. *Bioinformatics.*, **20**, 426–427.
35. Gertz,E.M., Yu,Y.K., Agarwala,R., Schaffer,A.A. and Altschul,S.F. (2006) Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.*, **4**, 41.
36. Cole,J.R., Chai,B., Farris,R.J., Wang,Q., Kulam-Syed-Mohideen,A.S., McGarrell,D.M., Bandela,A.M., Cardenas,E., Garrity,G.M. and Tiedje,J.M. (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res.*, **35**, D169–D172.
37. Blatny,J.M., Brautaset,T., Winther-Larsen,H.C., Haugan,K. and Valla,S. (1997) Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Appl. Environ. Microbiol.*, **63**, 370–379.
38. Ringvoll,J., Nordstrand,L.M., Vagbo,C.B., Talstad,V., Reite,K., Aas,P.A., Lauritzen,K.H., Liabakk,N.B., Bjork,A., Doughty,R.W. *et al.* (2006) Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *EMBO J.*, **25**, 2189–2198.
39. Ringvoll,J., Moen,M.N., Nordstrand,L.M., Meira,L.B., Pang,B., Bekkelund,A., Dedon,P.C., Bjelland,S., Samson,L.D., Falnes,P.O. *et al.* (2008) AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer Res.*, **68**, 4142–4149.
40. Falnes,P.O. and Rognes,T. (2003) DNA repair by bacterial AlkB proteins. *Res. Microbiol.*, **154**, 531–538.
41. Yang,C.G., Yi,C., Duguid,E.M., Sullivan,C.T., Jian,X., Rice,P.A. and He,C. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature*, **452**, 961–965.
42. Yu,B., Edstrom,W.C., Benach,J., Hamuro,Y., Weber,P.C., Gibney,B.R. and Hunt,J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879–884.
43. Venter,J.C., Remington,K., Heidelberg,J.F., Halpern,A.L., Rusch,D., Eisen,J.A., Wu,D., Paulsen,I., Nelson,K.E., Nelson,W. *et al.* (2004) Environmental genome shotgun sequencing of the *Sargasso Sea. Science*, **304**, 66–74.
44. Dinglay,S., Trewick,S.C., Lindahl,T. and Sedgwick,B. (2000) Defective processing of methylated single-stranded DNA by *E. coli* AlkB mutants. *Genes Dev*, **14**, 2097–2105.
45. el Ghissassi,F., Barbin,A. and Bartsch,H. (1998) Metabolic activation of vinyl chloride by rat liver microsomes: low-dose kinetics and involvement of cytochrome P450 2E1. *Biochem. Pharmacol.*, **55**, 1445–1452.
46. Dosanjh,M.K., Chenna,A., Kim,E., Fraenkel-Conrat,H., Samson,L. and Singer,B. (1994) All four known cyclic adducts formed in DNA by the vinyl chloride metabolite chloroacetaldehyde are released by a human DNA glycosylase. *Proc. Natl Acad. Sci. USA*, **91**, 1024–1028.
47. Kim,M.Y., Zhou,X., Delaney,J.C., Taghizadeh,K., Dedon,P.C., Essigmann,J.M. and Wogan,G.N. (2007) AlkB influences the chloroacetaldehyde-induced mutation spectra and toxicity in the pSP189 supF shuttle vector. *Chem. Res. Toxicol.*, **20**, 1075–1083.
48. Lee,D.H., Jin,S.G., Cai,S., Chen,Y., Pfeifer,G.P. and O'Connor,T.R. (2005) Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. *J. Biol. Chem.*, **280**, 39448–39459.
49. Madigan,M.T., Martinko,J.M. and Parker,J. (2003) *Brock Biology of Microorganisms*. Prentice Hall, Upper Saddle River, NJ.
50. Shiba,T. (1992) The genus roseobacter. In Starr,M.P., Stolp,H., Trüpper,H.G., Balows,A. and Schelgel,H.G. (eds), *The Prokaryotes: a Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*. Springer, Berlin, pp. 202–206.
51. Martelli,G.P., Adams,M.J., Kreuze,J.F. and Dolja,V.V. (2007) Family Flexiviridae: a case study in virion and genome plasticity. *Annu. Rev. Phytopathol.*, **45**, 73–100.
52. Birkeland,N.K., Anensen,H., Knaevelsrud,I., Kristoffersen,W., Bjoras,M., Robb,F.T., Klungland,A. and Bjelland,S. (2002) Methylpurine DNA glycosylase of the hyperthermophilic archaeon Archaeoglobus fulgidus. *Biochemistry*, **41**, 12697–12705.
53. Leiros,I., Nabong,M.P., Grosvik,K., Ringvoll,J., Haugland,G.T., Uldal,L., Reite,K., Olsbu,I.K., Knaevelsrud,I., Moe,E. *et al.* (2007) Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA. *EMBO J.*, **26**, 2206–2217.
54. Sedgwick,B. and Lindahl,T. (2002) Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene*, **21**, 8886–8894.
55. Gros,L., Ishchenko,A.A. and Saparbaev,M. (2003) Enzymology of repair of etheno-adducts. *Mutat. Res.*, **531**, 219–229.