Molecular cloning and functional analysis of a human cDNA encoding an *Escherichia coli* AlkB homolog, a protein involved in DNA alkylation damage repair

Ying-Fei Wei*, Kenneth C. Carter, Rui-Ping Wang and Brenda K. Shell

Department of Molecular Biology, Human Genome Sciences Inc., 9620 Medical Center Drive, Suite 300, Rockville, MD 20850-3338, USA

Received October 11, 1995; Revised and Accepted January 19, 1996 EMBL accession no. X91992

ABSTRACT

The *Escherichia coli* AlkB protein is involved in protecting cells against mutation and cell death induced specifically by SN2-type alkylating agents such as methyl methanesulfonate (MMS). A human cDNA encoding a polypeptide homologous to *E.coli* AlkB was discovered by searching a database of expressed sequence tags (ESTs) derived from high throughput cDNA sequencing. The full-length human AlkB homolog (*hABH*) cDNA clone contains a 924 bp open reading frame encoding a 34 kDa protein which is 52% similar and 23% identical to *E.coli* AlkB. The hABH gene, which maps to chromosome 14q24, was ubiquitously expressed in 16 human tissues examined. When hABH was expressed in *E.coli* alkB mutant cells partial rescue of the cells from MMS-induced cell death occurred. Under the conditions used expression of hABH in skin fibroblasts was not regulated by treatment with MMS. Our findings show that the AlkB protein is structurally and functionally conserved from bacteria to human, but its regulation may have diverged during evolution.

INTRODUCTION

Alkylating agents from both exogenous and endogenous sources contribute to alkylation on more than a dozen types of DNA lesions, some of which can lead to mutation and cell death. In mammals, mutations caused by DNA alkylation can also lead to neoplasia. Based on their cytotoxicity, alkylating agents are employed as antiviral drugs and are used in chemotherapeutic treatment of cancers (1,2).

DNA alkylation damage repair is best studied in *Escherichia coli*, where two defense mechanisms are responsible for protecting the accuracy of the genetic information against attack by alkylating agents. One mechanism, which is constitutive, depends on expression of the *ogt* gene, encoding an *O*6-methylguanine (*O*6MeG) methyltransferase (MTase) (3,4), and the *tag* gene, encoding a 3-methyladenine (*3MeA*) DNA glycosylase (5,6). The second mechanism is induced upon exposure to a sublethal dose of alkylating agent and is called the adaptive response (7–9). Four genes are involved in this system: *ada, alkB, aidB* and *alkB*.

The *ada* gene encodes a DNA MTase which removes the methyl group from *O*6MeG and *O*4-methylthymine (*O*4MeT) to its active site Cys321 residue. Since *O*6MeG and *O*4MeT mispair during DNA replication, repair by the Ogt and Ada MTases protects *E.coli* from alkylation-induced mutation (8,10). In addition, Ada MTase transfers a methyl group from methylphosphotriester DNA lesions to a second cysteine active site (Cys69), upon which the Ada protein is transformed into a transcriptional activator for the four genes of the adaptive response, including itself. *alkA*, the second Ada-regulated gene, like *tag*, encodes a DNA glycosylase which repairs *3MeA, 3*-methylguanine, *O*2-methylthymine and *O*2-methylcytosine (11,12). The *alkA* and *Tag* glycosylases specifically protect cells from alkylation-induced cell death, since *3MeA* is a lethal lesion which blocks DNA replication (13,14). Eukaryotic homologs of *O*6MeG MTases have been cloned in yeast (15), mouse (16) and human (17,18) and *3MeA* glycosylase homologs have been cloned in yeast (19–21), rat (22), mouse (23), human (24–26) and *Arabidopsis thaliana* (27). Although some of these eukaryotic enzymes are inducible by alkylating agents, there is no conclusive evidence suggesting that eukaryotes have the same adaptive response mechanism to alkylating agents as *E.coli* (28,29). Unlike the above genes, little is known about the function of the *aidB* product in DNA alkylation damage repair.

The precise biochemical function of the fourth member of the adaptive response, AlkB, is not clear. However, genetic studies indicate its importance in protecting cells from alkylating agent-induced DNA damage. *Escherichia coli* *alkB* mutant cells are extremely sensitive to MMS-induced mutation and cell death (30), suggesting that the AlkB pathway is extremely effective in defending against alkylation toxicity. At least two lines of evidence indicate that AlkB acts in repairing rather than preventing DNA alkylation damage: MMS-treated λ phage survive better in wild-type cells than in *alkB−* cells (30) and wild-type and *alkB* mutant *E.coli* genomes are alkylated to roughly the same extent when exposed to dimethyl sulfate (31). *E.coli* the AlkB repair pathway is probably independent of the AlkB glycosylase repair pathway, based on the following observations: (i) the sensitivity of an *alkB alkA* double mutant to

* To whom correspondence should be addressed
MMS and N-methyl-N’-nitrosoguanidine (MNNG) is at the level predicted from the additive sensitivity of each of the single mutants (32); (ii) purified E.coli AlkB does not reveal any glycosylase activity (33); (iii) wild-type alkB cannot complement an alkB mutant (31). Although the 27 kDa E.coli AlkB protein has been purified to homogeneity and a number of studies have been attempted to define its function, its mechanism of action remains unknown (33,34). Expression of E.coli AlkB in alkB-silenced and alkB-silent sensitive human cell lines rescues the cells from MMS-induced cell death (31), indicating that E.coli AlkB may function independently in mammalian cells and that mammalian cells might have a similar AlkB repair pathway. To date, no eukaryotic AlkB homolog has been identified. Identification of the eukaryotic AlkB homolog would represent an important step towards understanding the function of AlkB and may help elucidate a new pathway for eukaryotic alkylation damage repair. In addition, a number of DNA repair genes are involved in human neoplasia, e.g. the XP excision repair genes involved in xeroderma pigmentosum and Cockayne’s syndrome (35) and mismatch repair genes involved in hereditary non-polyposis colon cancer (36–40). Identification of new human DNA repair genes may lead to the discovery and identification of new disease genes.

In order to identify potential human AlkB homologs we searched a human cDNA database generated by high throughput cDNA sequencing and analysis using the expressed sequence tag (EST) method (41). One EST, derived from a human synovial sarcoma cDNA library, was found to be homologous to the C-terminal half of E.coli AlkB. The full-length hABH protein shares significant sequence homology with E.coli AlkB and is able to partially rescue an E.coli alkB mutant from MMS-induced cell death, suggesting that the AlkB protein is structurally and functionally conserved during evolution, like many other DNA repair enzymes.

**MATERIALS AND METHODS**

**Cells, plasmid and chemicals**

Escherichia coli strains HK81 (thr-1 leu-6 proA2 his-4 argE3 thi-1 lacY1 gall2 ara-14 syl-5 mtt-1 tex-33 stra31 supE44 nalA) and HK82 (as HK81 but alkB22), kindly provided by Dr Leona Samson (Harvard School of Public Health, Boston, MA), were used as the hosts of wild-type and mutant alkB respectively. Human skin fibroblast cell line CCD-965SK was purchased from the American Type Culture Collection. Plasmid vector pQE-9 was purchased from Qiagen. MMS was purchased from Aldrich Chemical Co. Inc. The cDNA libraries were constructed using the Stratagene Uni-ZAP XR cDNA kit.

**Sequence homology search against the human EST database**

A database containing ~500 000 human ESTs has been generated through the combined efforts of The Institute for Genomic Research and Human Genome Sciences Inc. using high throughput automated DNA sequence analysis of randomly selected human cDNA clones (41,42). Sequence homology comparisons of each EST were performed against the GenBank database using the blastn and tblastn algorithms (43). ESTs having homology with previously identified sequences (P ≤ 0.01) were given a tentative name based on the name of the sequence to which it was homologous. A specific homology search using the known E.coli AlkB amino acid sequence against this human EST database revealed one EST having >34% homology.

**Screening for full-length hABH**

The partial hABH cDNA (~1.3 kb in length) identified within the human EST database was radiolabeled and hybridized to 20 nylon membranes containing 600 000 individual plaques from a human synovial sarcoma Uni-ZAP II cDNA library. Hybridization and washing were under the conditions described by Maniatis et al. (44). Three positive clones each ~0.6 kb longer than the original partial cDNA clone, were identified.

**RNA blot analysis**

Human total RNA isolation was modified from the RNAzol B procedure of Biotex Laboratories Inc. Briefly, 2 g tissue were mixed with 40 ml RNAzol B solution and homogenized on a Polytron PT3000 (Brinkmann) at 22 000 r.p.m. for 2–3 min. Then 5 ml chloroform was added to the homogenate and the mixture was vortexed for 15 s and incubated on ice for 5 min. The suspension was centrifuged at 12 000 g at 4°C for 15 min and the upper phase homogenate was mixed with an equal volume of isopropanol and incubated at 4°C for 5 min to precipitate RNA. The samples were centrifuged again at 12 000 g for 15 min and the pellet was resuspended in 2 ml DEPC-treated water and extracted twice with an equal volume of phenol/chloroform (1:1 v/v, equilibrated with DEPC-water). The RNA was ethanol precipitated and resuspended in water. An aliquot of 15 mg total RNA was electrophoresed in a 1% agarose–formaldehyde gel (44), transferred to a nylon membrane and hybridized with the 1.9 kb hABH cDNA fragment. The blot was washed three times with 0.1× SSC and 0.1% SDS at 50°C for 1 h and subjected to autoradiography.

**Construction of pQE9hABH for expression of functional hABH protein**

The 1 kb open reading frame (ORF) coding for the hABH protein was amplified by the polymerase chain reaction (PCR) using the oligonucleotides 5′-gcgcggagtCATGTGCTCTTCATCGATGTTGCCAGAAACC, including a SalI site (underlined) followed by the first 20 nt of the C-terminal complementary sequence. The 1 kb PCR product was gel purified and digested with SalI and HindIII restriction enzymes and ligated to pQE-9 vector which had been digested with the same enzymes. The ligation mixture was transformed into XL1-Blue cells to verify the construct by restriction digestion and the pQE9hABH DNA was transformed into E.coli alkB mutant strain HK82 for functional complementation testing. The hABH expressed in this way was fused with a histidine tag at the N-terminus of the protein (Fig. 1).

**Survival curve**

Functional complementation of hABH in E.coli alkB mutant cells was measured by the cell survival rate on MMS plates. The wild-type and mutant E.coli cells harboring pRSValkB (E.coli alkB; see 31), pQE-9 vector or pQE9hABH were grown in LB/ampicillin medium to log phase. Cells were diluted with M9 medium and plated on LB/ampicillin plates containing 0, 0.02, 0.035 or 0.05% MMS and the plates incubated at 37°C overnight.
digoxigenin–dUTP (Boehringer Mannheim) and fluorescence in situ hybridization was carried out as detailed in Johnson containing a portion of the (Stratagene) using standard procedures (44). A genomic clone (45). Individual metaphase chromosome spreads were counter stained with DAPI and gene signals, were recorded using a triple bandpass filter set (Chroma Technology Inc., Brattleburo, VT) in combination with a charged couple device camera (Photometrics Inc., Tucson, AZ) and variable excitation wavelength filters (46). The DAPI and gene signals, which had 34% identity and 59% similarity at the amino acid level from a human synovial sarcoma cDNA library was identified using the blastn and tblastn algorithms (43). A 249 bp EST derived with a human cDNA database containing ∼500 000 human ESTs (47) and the homology, with 34% identity. The nucleotide sequence of this cDNA clone reveals a 1953 bp insert containing a 924 bp ORF extending from bp 200 to 1124 encoding 307 amino acid residues fused with 14 amino acids from the vector, including six histidines.

The cell survival rate at each dose was calculated by the number of colonies at that dose divided by the number of colonies on non-MMS plates.

Chromosomal mapping

Genomic clones of hABH were isolated from an IFIX-2 library (Stratagene) using standard procedures (44). A genomic clone containing a portion of the hABH gene was nick-translated using digoxigenin–dUTP (Boehringer Mannheim) and fluorescence in situ hybridization was carried out as detailed in Johnson et al. (45). Individual metaphase chromosome spreads were counter stained with DAPI and color digital images, containing both the DAPI and gene signals, were recorded using a triple bandpass filter set (Chroma Technology Inc., Brattleburo, VT) in combination with a charged couple device camera (Photometrics Inc., Tucson, AZ) and variable excitation wavelength filters (46). The same chromosome spreads were then stained using conventional g-banding procedures and images were again recorded. Chromosomes from the DAPI/gene images of the same spreads were aligned with the g-band images using the ISEE software package (Inovision Corp., Durham, NC), allowing assignment of precise g-band positions.

RESULTS

Discovery of a human EST which is homologous to E.coli AlkB

As most of the DNA repair proteins are conserved during evolution, we reasoned that human AlkB might be homologous to E.coli AlkB. To identify a potential human AlkB homolog the amino acid sequence of E.coli AlkB (216 residues) was compared with a human cDNA database containing ∼500 000 human ESTs using the blastn and blastx algorithms (43). A 249 bp EST derived from a human synovial sarcoma cDNA library was identified which had 34% identity and 59% similarity at the amino acid level to a region within the C-terminal portion of E.coli AlkB (data not shown). This EST clone was discovered as part of a joint collaboration between scientists at The Institute for Genomic Research and Human Genome Sciences. The cDNA clone from which this EST was derived had a 1.3 kb cDNA insert. There was no similarity between the two genes at the nucleotide level.

Sequencing of the 1.3 kb cDNA insert revealed that it contained an ORF truncated at the 5′-end. The translated polypeptide sequence from this ORF shares homology with the C-terminal region of E.coli AlkB. We therefore searched for the full-length cDNA of this gene by screening the human Uni-ZAP XR II synovial sarcoma cDNA library from which the initial EST was obtained. As described in Materials and Methods, 600 000 plaques were screened and three positive clones, each ∼600 bp longer at the 5′-end than the original clone, were obtained. These three new clones were identical based on their restriction enzyme digestion pattern and 5′- and 3′-end sequence information. The nucleotide sequence of this cDNA clone reveals a 1953 bp insert containing a 924 bp ORF extending from bp 200 to 1124 and a poly(A) tail at the 3′-end (Fig. 2). The polypeptide encoded by this ORF (Fig. 2) showed significant homology to E.coli AlkB (Fig. 3). Therefore, this clone probably contains the full-length hABH cDNA. There are two putative translation initiation codons at the beginning of this ORF, one at bp 200 (immediately after an in-frame upstream stop codon) and one at bp 224. The ATG at bp 224 is likely to be the actual translation initiation site, since it lies in an optimal translation start consensus (47) and the homology to AlkB starts immediately thereafter (Fig. 3), whereas the ATG at position 200 is preceded by an unfavorable sequence (47) and there is no sequence homology to AlkB in the region from bp 200 to 230 (Fig. 3). Assuming that translation of hABH starts at the ATG at bp 224 and ends at the first in-frame stop codon at bp 1124, hABH contains 299 amino acid residues encoding a 34.040 kDa protein. This agrees with the protein size of in vitro translated hABH (data not shown). Two consensus AA TAAA polyadenylation signals (48) located at bp 1868 and 1912, which are 61 and 19 bp respectively upstream of the poly(A) tail, were observed. The amino acid sequences of hABH and E.coli AlkB are conserved throughout the coding region, with 52% similarity and 23% identity (Fig. 3). A stretch of 104 amino acids within the middle of the protein sequence (boxed in Fig. 3) reveals the best homology, with 34% identity. The nucleotide sequence of hABH cDNA has been deposited in GenBank with the accession no. X91992.
Figure 3. Sequence homology between hABH and E. coli AlkB (eAlkB). The deduced amino acid sequence from the entire 924 bp ORF of hAlkB was compared with the E. coli AlkB sequence using the Bestfit program (56). For the two protein sequences the straight lines indicate identical amino acids and dotted lines indicate similar amino acids (double dots are more similar than single dots). The human hABH sequence does not match E. coli AlkB until residue 11 of hABH, which agrees with our prediction that the ATG at residue 9 may be the actual translation initiation site. The two proteins share 52% similarity and 23% identity. The most conserved region between the two protein (34% identity) is boxed.

The hABH gene is ubiquitously expressed in normal human tissues

To determine the expression pattern of human AlkB in different tissues we carried out an RNA blot analysis on 16 normal adult human tissues, including brain, kidney, small intestine, testis, pancreas, prostate, heart, liver, lung, thymus, spleen, placenta, colon, ovary, leukocyte and muscle. As shown in Figure 4, hABH RNA is ubiquitously present in all these tissues as a 2.1 kb message, suggesting that hABH is probably a housekeeping gene that plays a fundamental role in most human tissues.

Expression of hABH can partially protect E. coli alkB mutant cells against MMS-induced cell death

To examine whether hABH plays a similar role to E. coli AlkB in defending cells against alkylating agent-induced cell death we expressed an N-terminal tagged hABH in E. coli alkB mutant cells and tested for its ability to protect the E. coli alkB mutant from MMS-induced cell death. The bacterial expression construct pQE9hABH (Fig. 1) was prepared as described in Materials and Methods. As it is not certain which ATG is the actual translation initiation codon, we inserted the entire ORF from the first ATG at bp 200 to the first in-frame stop codon at bp 1124 into the pQE-9 vector. The hABH expressed from this plasmid contains 307 amino acid residues from its coding region fused with 14 amino acids from the vector, including six histidines at the N-terminus. The His tag will facilitate protein purification by nickel affinity column chromatography.

The plasmid pQE9hABH was transformed into E. coli alkB mutant strain HK82 and the MMS resistance of the transformants was examined by a plate assay (see Materials and Methods). As a control, plasmid vector pQE-9 was transformed into wild-type HK81 and alkB mutant HK82 cells. Six independent HK82/pQE9hABH transformants were tested, each of which showed an increased survival rate of ~10-fold (Fig. 5). Escherichia coli alkB can fully complement HK82 under the same conditions (Fig. 5). This result suggests that hABH plays a similar role to E. coli AlkB in protecting cells against alkylating agent-induced cell death. It also indicates that the AlkB alkylation damage repair pathway has been conserved during evolution.

Expression of the hABH gene is not altered by MMS

Expression of E. coli AlkB is inducible upon treatment with a low dose of alkylating agents as part of the adaptive response to
and bacterial AlkB proteins includes homology in their respective primary structures and an apparently strong functional conservation, indicated by the ability of the human gene to partially rescue bacterial alkB mutant cells. The two proteins share large areas of amino acid homology, particularly within a 102 residue stretch in the middle portion of each protein (boxed in Fig. 3). Although the precise biochemical role of AlkB remains unknown, it is reasonable to suggest that these highly conserved regions may represent core functional domains. In addition to the areas of highly conserved sequence, hAlkB is significantly larger and appears to contain five insertions of non-homologous domains (Fig. 3). These five insertions form two clusters; the first two insertions are clustered at the N-terminus of the 102 residue highly conserved region and the last three insertions are clustered at the C-terminus of this region. These regions, like the more conserved areas of the protein, do not share sequence homology with any other known genes or functional domains. It is possible that these insertions reflect increased complexity of DNA alkylation repair pathways in higher eukaryotes. For example, other proteins may participate in the repair process through interaction with these regions. At present no other higher eukaryotic AlkB homologs are known, so sequence comparisons cannot be done. A few putative yeast genes have been shown to be able to complement E.coli alkB mutant cells (55). However, the yeast genes do not share primary sequence homology with either bacterial or human AlkB.

The existence of a human counterpart to bacterial AlkB had been previously suggested by the observation that E.coli AlkB can function independently in humans to protect cells against DNA alkylation-induced cell death (31). Our data suggest that the human protein can function similarly in bacterial cells. Interestingly, hABH only partially rescued the E.coli alkB mutant phenotype (Fig. 5), which may be due either to inefficient expression of hABH in E.coli or to variations in the microenvironment of E.coli which are not optimal for hABH activity. Alternatively, the non-homologous regions within the larger human protein might function as regulatory elements, with maximal functioning of the human protein requiring interaction with cellular proteins not present in E.coli. However, because the precise biochemical function of AlkB is unknown and there is no assay system available, it is hard to identify the specific role of different regions within either the human or bacterial proteins. It should be noted that the E.coli alkB gene can fully complement E.coli alkB mutant cells (31). As we used His-tagged hAlkB for the complementation experiment, it would be interesting to see if His-tagged E.coli AlkB can still function fully or non-tagged hABH can rescue E.coli alkB better than tagged hABH. The lack of full complementation in our experiment may also be due to the extra eight amino acids at the N-terminus, assuming that native hABH starts at the second methionine (at the ATG at bp 224).

Despite the structural and functional conservation between human and E.coli AlkB, the two proteins also differ in ways besides the inserted regions of non-homologous sequence seen in the human protein. First, in E.coli the alkB gene is in the same operon as the ada MTase gene, whereas in the human alkB and MGMT are located on different chromosomes (we have shown here that hABH is on chromosome 14 and Rydberg et al. have shown that human MGMT is on chromosome 10; see 18). Second, expression of E.coli alkB is regulated by Ada as part of the adaptive response to alkylating agents (7–9). In the human no similar regulatory system has been defined, although both
Figure 7. Fluorescence in situ hybridization mapping of hABH. A genomic clone containing a portion of the hABH gene was hybridized to normal human male chromosomes. (A) A chromosome spread from a single cell showing hybridization to the q arm of each chromosome 14 (arrows). (B) DAPI stained (blue) and aligned g-band (black and white) images of two individual chromosomes containing hABH signal (see Materials and Methods). Most gene signals appeared at the distal border of band q24. (C) Assignment of band position based on analysis of images from 36 individual chromosomes.

MGMT and ANPG may be slightly induced upon treatment with DNA alkylating agents in certain cell types (50, 51). Our results show that expression of hABH in skin fibroblasts is not altered by MMS under the conditions used. Hence, regulation of hABH expression, at least in this cell type and under this treatment, is regulated differently than in E.coli. Induction of human MGMT and ANPG by MMS and other DNA damaging agents has been reported in a number of human transformed cell lines, such as a hepatoma cell line and a glioblastoma cell line (50). It will be important to see if hAlkB is inducible in these cell lines.

Cloning AlkB and other homologs to bacterial DNA repair genes in various species is an important step towards elucidating DNA repair pathways in higher eukaryotes. However, cloning cross-species homologs with limited sequence homology can be problematical, particularly when the precise biochemical functions of the protein are unknown. The functional complementation approach has been successful in cloning a number of eukaryotic DNA repair gene homologs (see, for example, 15, 17, 19, 24), but when it was used to screen for a yeast alkB homolog none of several genes that showed an ability to partially complement the E.coli alkB mutant contained sequence homology with alkB (55). It will be important to determine whether these yeast genes are conserved in higher eukaryotes and whether they play a role in DNA alkylation repair. An approach commonly used to clone sequence-homologous genes in different species is to perform degenerate PCR based on the conserved sequence motifs among the known genes; this usually requires multiple genes to determine these motifs and is therefore unsuitable for alkB or the functional homologs defined in yeast. Using a large database of cDNA EST sequences to identify homologous genes, as was used here to clone halkB, provides a valuable and in some cases essential approach to identifying homologous genes. This relatively new approach has already led to identification of other DNA repair genes, including three DNA mismatch repair genes involved in human hereditary non-polyposis colorectal cancer (39, 40).
Finally, as was the case with the above mismatch repair genes, mutations in hABH may be involved in human neoplasia. Alterations in chromosome band 14q24, the site of the structural gene for hABH, have been observed in ≈90% of reported cytogenetic analyses of leiomysarcoma cases and are seen sporadically in several other cancers (53). Mutations in the neighboring band, 14q31, are only very rarely reported. In addition, one of the loci for autosomal dominant cerebellar ataxia is mapped on chromosome 14q24–31 (54). There is no evidence suggesting that these diseases are directly related to a deficiency in DNA damage repair, however, as DNA alklylation introduces mutations, it can cause defective proteins in general. Further studies are needed to define the potential relationship of hABH to any of these human diseases.

ACKNOWLEDGEMENTS

We would like to thank Drs Craig Rosen, Steve Ruben and Guo-Liang Yu for their helpful comments and stimulating discussions. Many thanks to Ms Julie Nelson for typing the references. The human tissues used in this study were supplied by the Cooperative Human Tissue Network, which is funded by the National Cancer Institute. The initial EST in this study was sequenced at The Institute for Genomic Research. Nucleotide sequencing of the hABH cDNA clone and related constructs was carried out at Human Genome Sciences Inc. sequencing core facility.

REFERENCES