

Introduction to sequence similarity searches and sequence alignment

**MBV-INF4410/9410
Monday 21 November 2016**

**Torbjørn Rognes
Department of Informatics, University of Oslo
& Department of Microbiology, Oslo University Hospital
torognes@ifi.uio.no**

Overview of the presentation

PART 1

- An example showing how useful bioinformatics can be
- Searching sequence databases
- A walk-through of the BLAST search service

PART 2

- Alignments, sequence similarity and homology
- Significance of matches: What is a good match?
- How does BLAST work?

PART 3

- Iterative searching with a family of proteins (PSI-BLAST)

PART 4:

- Multiple sequence alignments

One example of how useful bioinformatics can be

- The protein AlkB was discovered in *E.coli* in 1984.
- It was known that it protected the bacterium when subjected to DNA-alkylation agents.
- No enzymatic activity was found.
- Perhaps some co-factors were missing?
- In 2001, a bioinformatics paper was published that shed light on the problem. Many similar sequences were found using advanced sequence similarity searches ...

<http://genomebiology.com/2001/2/3/research/0007.1>

Research

The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases

L Aravind and Eugene V Koonin

Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

Correspondence: L Aravind. E-mail: aravind@ncbi.nlm.nih.gov

Example...

Alignment showing conserved amino acids among many sequences

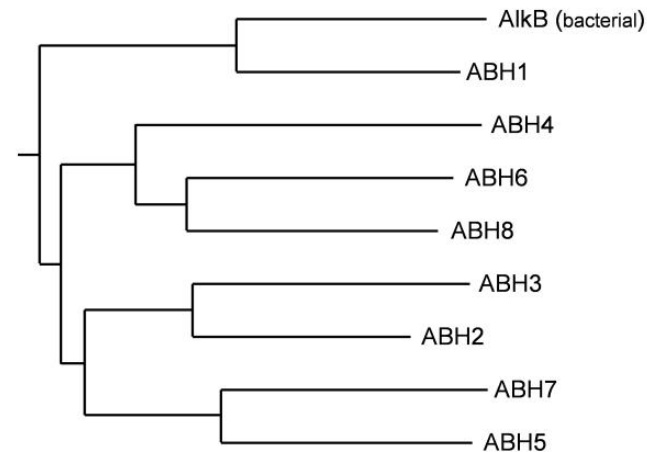
[illegible]

Example...

- By comparing *E.coli* AlkB to other sequences in the database it was found that AlkB had some features in common with more well-known enzymes
- Based on these similarities the following was suggested regarding AlkB:
 - That AlkB is a dioxygenase
 - That the enzyme is Iron(II) dependent
 - That the enzyme is 2-oxo-glutarate dependent
 - That AlkB repairs alkylated bases through a form of oxidation
 - That the enzyme could demethylate RNA as well (not just DNA)
 - That there were eukaryotic counterparts of the protein
- All of this was later verified in the lab and resulted in three publications in *Nature*.

Example...

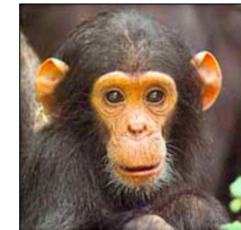
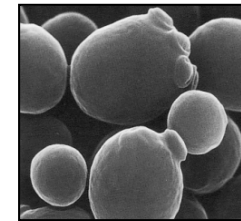
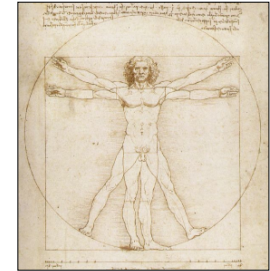
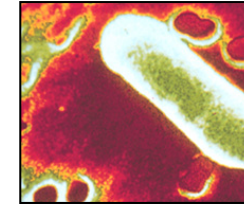
- By further sequence analysis 3 AlkB-like sequences were found in humans:
 - ALKBH1
 - ALKBH2
 - ALKBH3
- And by even more advanced analysis another 5 homologs were found in humans:
 - ALKBH4
 - ALKBH5
 - ALKBH6
 - ALKBH7
 - ALKBH8



- The function of these 8 enzymes are now being studied in detail. Some of them may be related to human diseases.

Genomes are a huge source of information

- More than 7000 "completely" sequenced genomes available – an enormous source of information. Many thousands of other genomes in progress*
- More than 1 000 000 000 000 basepairs in GenBank (incl. WGS) (2015)*
- Database sizes are growing exponentially
– doubling in about 18 months since 1982
- Searching sequence databases for a similar sequence is fundamental in many types of analyses in bioinformatics
- Searching a sequence database with a new amino acid or nucleotide sequence allow us to find out more about:
 - Gene function
 - Conserved and probably important residues
 - 3D structure of a protein
 - Distribution of the gene among species
 - Gene structure
 - Chromosomal localisation
- Save time in the lab!
- Database searching is highly compute intensive and is probably the task consuming the largest amount of computing time within bioinformatics.



* Sources: genomesonline.org & NCBI (<ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt>)

Searching sequence databases

- Goal: Identify which sequences in a database are significantly similar to a given DNA, RNA or protein sequence.
- How: The query sequence is compared (aligned) with each of the database sequences, and the amount of similarity is determined for each database sequence.

Example:

Query sequence:

acgatcgattagcca

Database sequences:

Identical (trivial):

acgatcgattagcca

Very similar (easy):

acga**c**cgat**g**agcca

Similar (moderate):

a**t**ga**c**ggat**g**agc**g**a

Very diverged (hard):


a**t**ga**c**g**g**gat**g**agc**g**a

Firefox

BLAST: Basic Local Alignment Search Tool

blast.ncbi.nlm.nih.gov/Blast.cgi

Google

BLAST®

Basic Local Alignment Search Tool

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

My NCBI
[\[Sign In\]](#) [\[Register\]](#)

► NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

New

DELTA-BLAST, a more sensitive protein-protein search

Go

BLAST Assembled RefSeq Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

<input type="checkbox"/> Human	<input type="checkbox"/> Oryza sativa	<input type="checkbox"/> Gallus gallus
<input type="checkbox"/> Mouse	<input type="checkbox"/> Bos taurus	<input type="checkbox"/> Pan troglodytes
<input type="checkbox"/> Rat	<input type="checkbox"/> Danio rerio	<input type="checkbox"/> Microbes
<input type="checkbox"/> Arabidopsis thaliana	<input type="checkbox"/> Drosophila melanogaster	<input type="checkbox"/> Apis mellifera

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a nucleotide database using a nucleotide query <i>Algorithms: blastn, megablast, discontinuous megablast</i>
protein blast	Search protein database using a protein query <i>Algorithms: blastp, psi-blast, phi-blast, delta-blast</i>
blastx	Search protein database using a translated nucleotide query
tblastn	Search translated nucleotide database using a protein query
tblastx	Search translated nucleotide database using a translated nucleotide query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- ☐ Make specific primers with [Primer-BLAST](#)
- ☐ Search [trace archives](#)
- ☐ Find [conserved domains](#) in your sequence (cds)
- ☐ Find sequences with similar [conserved domain architecture](#) (cdart)
- ☐ Search sequences that have [gene expression profiles](#) (GEO)
- ☐ Search sequences that have [BLAST](#)

News

[Improved BLASTX statistics](#)

BLASTX now uses composition based statistics (CBS).
Wed, 01 Aug 2012 17:00:00 EST

[More BLAST news...](#)

Tip of the Day

[How to Search Custom Databases in Web-Blast Using Entrez Queries.](#)

A powerful feature of the BLAST Web interface is the ability to limit BLAST searches to a subset of any database using a standard Entrez query.

[More tips...](#)

Search program variants

Query	Database	Comparisons	FASTA	BLAST	Description
Nucleotide	Nucleotide	Nucleotide (2)	fasta (fastn)	blastn	Compares directly both strands (forward and reverse complement) of the nucleotide query sequence to the nucleotide sequences in the database.
Amino acid	Amino acid	Amino acid (1)	fasta (fastp)	blastp	Compares the amino acid query sequence with the amino acid sequences in the database.
Amino acid	Nucleotide	Amino acid (6)	tfasta, tfastx, tfasty	tblastn	Translates the database nucleotide sequences into all six frames and compares the resulting amino acid sequences with the amino acid query sequences. tfasty allows intra-codon substitutions and frameshifts.
Nucleotide	Amino acid	Amino acid (6)	fastx, fasty	blastx	Translates the nucleotide query sequence into all six frames and compares the resulting amino acid sequences with the amino acid sequences in the database. fasty allows intra-codon substitutions and frameshifts.
Nucleotide	Nucleotide	Amino acid (36)	-	tblastx	Translates both the query nucleotide sequence and the database nucleotide sequences into all six frames and compares the resulting amino acid sequences with each other.

Firefox

Protein BLAST: search protein databases ...

blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LII

Google

BLAST®

Basic Local Alignment Search Tool

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

My NCBI
[\[Sign In\]](#) [\[Register\]](#)

[NCBI/BLAST/blastp suite](#) **Standard Protein BLAST**

[blastn](#) [blastp](#) [blastx](#) [tblastn](#) [tblastx](#)

BLASTP programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

MLDLFADAEFWQEP LAAGAVILRRFAFNAEQLIRDINDVASQSPFRQMTVPGGYTMSVA
MTNCGHLGWTTTHRQGYLYSPIDPQTINKFWPAMPQSFHNLQRAATAAGYPDFQPDACLIN
RYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNDPLKRLLEHGDVVVWGG
ESRLFYHGIQPLKAGFHPLTIDCRYNLTFRQAGKKE

[Clear](#)

Query subrange

From

To

Or, upload file

Browse...

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

UniProtKB/Swiss-Prot(swissprot)

Organism

Optional

Enter organism name or id—completions will be suggested

☐ Exclude

+

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude

Optional

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query

Optional

Enter an Entrez query to limit search

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)
☐ PSI-BLAST (Position-Specific Iterated BLAST)
☐ PHI-BLAST (Pattern Hit Initiated BLAST)
☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
Choose a BLAST algorithm

BLAST

Search database UniProtKB/Swiss-Prot(swissprot) using Blastp (protein-protein BLAST)

BLAST databases (protein)

nr:	All non-redundant GenBank CDS translations + RefSeq Proteins + PDB + UniProtKB/SwissProt + PIR + PRF
refseq:	RefSeq protein sequences from NCBI's Reference Sequence Project.
swissprot:	The SWISSPROT part of UniProt Knowledge Base (UniProtKB)
pat:	Patented protein sequences
pdb:	Sequences of proteins in the Protein Data Bank (PDB) containing the 3-dimensional structure of proteins
env_nr:	Protein sequences from metagenomic projects and environmental samples.

BLAST databases (nucleotides)

nr:	All GenBank + RefSeq Nucleotides + EMBL + DDBJ + PDB sequences (excluding HTGS0,1,2, EST, GSS, STS, PAT, WGS). No longer "non-redundant".
refseq_rna:	RNA entries from NCBI's Reference Sequence project
refseq_genomic:	Genomic entries from NCBI's Reference Sequence project
chromosome:	A database with complete genomes and chromosomes from the NCBI Reference Sequence project..
est:	Database of GenBank + EMBL + DDBJ sequences from EST Divisions
gss:	Genome Survey Sequence, includes single-pass genomic data, exon-trapped sequences, and Alu PCR sequences.
htgs:	Unfinished High Throughput Genomic Sequences: phases 0, 1 and 2 (finished, phase 3 HTG sequences are in nr)
pat:	Nucleotides from the Patent division of GenBank.
pdb:	Sequences derived from the 3-dimensional structure from Protein Data Bank (PDB)
alu_repeats:	Human ALU repeat elements
dbsts:	Database of GenBank+EMBL+DDBJ sequences from STS Divisions .
wgs:	A database for whole genome shotgun sequence entries
tsa:	Transcriptome shotgun assembly
16S:	16S ribosomal RNA from Bacteria and Archaea

Firefox

Protein BLAST: search protein databases ...

blast.ncbi.nlm.nih.gov/Blast.cgi

Google

BLAST

Search database UniProtKB/Swiss-Prot(swissprot) using Blastp (protein-protein BLAST)

☐ Show results in a new window

Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

General Parameters

Max target sequences

100

Select the maximum number of aligned sequences to display

Short queries

☒ Automatically adjust parameters for short input sequences

Expect threshold

10

Word size

3

Max matches in a query range

0

Scoring Parameters

Matrix

BLOSUM62

Gap Costs

Existence: 11 Extension: 1

Compositional adjustments

Conditional compositional score matrix adjustment

Filters and Masking

Filter

☐ Low complexity regions

Mask

☐ Mask for lookup table only

☐ Mask lower case letters

BLAST

Search database UniProtKB/Swiss-Prot(swissprot) using Blastp (protein-protein BLAST)

☐ Show results in a new window

BLAST is a registered trademark of the National Library of Medicine.

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback


NCBI | NLM | NIH | DHHS

Firefox

NCBI Blast:Protein Sequence (216 letters)

blast.ncbi.nlm.nih.gov/Blast.cgi

Google

BLAST®

Basic Local Alignment Search Tool

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

My NCBI
[Sign In](#) [Register](#)

[NCBI/ BLAST/ blastp suite/ Formatting Results - 1ZKJE2S701N](#) [\[Formatting options\]](#)

Job Title: Protein Sequence (216 letters)

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq.

Superfamilies

1

25

50

75

100

125

150

175

200

216

20G-FeII_Oxy superfamily

Request ID	1ZKJE2S701N
Status	Searching
Submitted at	Mon Aug 6 10:01:50 2012
Current time	Mon Aug 6 10:01:53 2012
Time since submission	00:00:02

This page will be automatically updated in 7 seconds

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback

BLAST is a registered trademark of the National Library of Medicine.

NCBI | NLM | NIH | DHHS

Firefox

NCBI Blast:Protein Sequence (216 letters)

blast.ncbi.nlm.nih.gov/Blast.cgi

Google

BLAST®

Basic Local Alignment Search Tool

HomeRecent ResultsSaved StrategiesHelp

My NCBI[Sign In][Register]

NCBI/ BLAST/ blastp suite/ Formatting Results - 1ZKJE2S701N

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)

Protein Sequence (216 letters)

Query ID	Id 94622	Database Name	swissprot
Description	unnamed protein product	Description	Non-redundant SwissProt sequences
Molecule type	amino acid	Program	BLASTP 2.2.26+ Citation
Query Length	216		

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)

Graphic Summary

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. Superfamilies 20G-FeII_Oxy superfamily

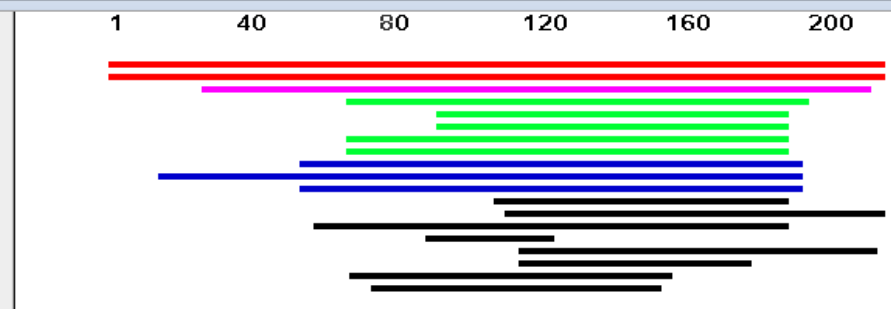
Distribution of 19 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

Color key for alignment scores

<40	40-50	50-80	80-200	>=200
-----	-------	-------	--------	-------

Descriptions



Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
P05050.1	RecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB; Alt	451	451	100%	2e-161	100%	S
P37462.2	RecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB; Alt	366	366	100%	8e-128	80%	
P0CAT7.1	RecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB hom	144	144	86%	6e-41	42%	
Q9SA98.2	RecName: Full=Alpha-ketoglutarate-dependent dioxygenase alkB; Alt	76.3	76.3	59%	3e-15	37%	GM
P0CB42.1	RecName: Full=Alkylated DNA repair protein alkB homolog 1; AltName:	72.8	72.8	45%	7e-14	37%	GM
Q13686.2	RecName: Full=Alkylated DNA repair protein alkB homolog 1; AltName:	70.9	70.9	45%	4e-13	34%	GM
Q60066.2	RecName: Full=Alpha-ketoglutarate-dependent dioxygenase abh1; Al	68.6	68.6	56%	1e-12	29%	G
Q54N08.1	RecName: Full=Alpha-ketoglutarate-dependent dioxygenase alkB; Alt	68.6	68.6	56%	2e-12	29%	
Q80Y20.1	RecName: Full=Alkylated DNA repair protein alkB homolog 8; AltName:	44.3	44.3	64%	4e-04	28%	GM
Q07G10.2	RecName: Full=Alkylated DNA repair protein alkB homolog 8; AltName:	42.7	42.7	82%	0.001	24%	G
A1A4L5.1	RecName: Full=Alkylated DNA repair protein alkB homolog 8; AltName:	40.8	40.8	64%	0.005	27%	GM
Q96BT7.2	RecName: Full=Alkylated DNA repair protein alkB homolog 8; AltName:	38.1	38.1	37%	0.036	36%	S GM
Q5UR03.1	RecName: Full=Uncharacterized protein L905	37.0	37.0	49%	0.058	27%	G
Q95K79.1	RecName: Full=Alkylated DNA repair protein alkB homolog 8; AltName:	37.7	37.7	61%	0.059	29%	
Q6DDW2.1	RecName: Full=UPF0577 protein KIAA1324-like homolog; AltName: Ful	33.5	33.5	16%	1.1	33%	G
B1PS76.1	RecName: Full=RNA-directed RNA polymerase; Includes: RecName: Fu	33.5	33.5	46%	1.5	30%	
Q67704.1	RecName: Full=RNA replication protein; Includes: RecName: Full=RNA	33.1	33.1	30%	1.9	36%	
Q8N661.2	RecName: Full=Lysoplasmalogenase; AltName: Full=Transmembrane p	32.0	32.0	41%	2.6	31%	GM
Q64962.1	RecName: Full=RNA replication protein; Includes: RecName: Full=RNA	31.2	31.2	37%	8.5	26%	

Alignments

Firefox

NCBI Blast: Protein Sequence (216 letters) +

blast.ncbi.nlm.nih.gov/Blast.cgi

Google

Alignments

☐ Select All [Get selected sequences](#) [Distance tree of results](#) [Multiple alignment](#)

> [sp|P05050.1|ALKB_ECOLI](#) S RecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB;
AltName: Full=Alkylated DNA repair protein AlkB
Length=216

Score = 451 bits (1161), Expect = 2e-161, Method: Compositional matrix adjust.
Identities = 216/216 (100%), Positives = 216/216 (100%), Gaps = 0/216 (0%)

Query	1	MLDLFADAEFWQEPLAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTPGGYMSVA	60
Sbjct	1	MLDLFADAEFWQEPLAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTPGGYMSVA	60
Query	61	MTNCGHLGWITHRQGYLYSPIDPQTNKPWPAMPQSFHNLQRAATAAGYPDFQPDACLIN	120
Sbjct	61	MTNCGHLGWITHRQGYLYSPIDPQTNKPWPAMPQSFHNLQRAATAAGYPDFQPDACLIN	120
Query	121	RYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNPLKRLLEHGDVVVWGG	180
Sbjct	121	RYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNPLKRLLEHGDVVVWGG	180
Query	181	ESRLFYHGIQPLKAGFHPLTIDCRYNLTFRQAGKKE	216
Sbjct	181	ESRLFYHGIQPLKAGFHPLTIDCRYNLTFRQAGKKE	216

> [sp|P37462.2|ALKB_SALTY](#) RecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB;
AltName: Full=Alkylated DNA repair protein AlkB
Length=216

Score = 366 bits (940), Expect = 8e-128, Method: Compositional matrix adjust.
Identities = 172/216 (80%), Positives = 193/216 (89%), Gaps = 0/216 (0%)

Query	1	MLDLFADAEFWQEPLAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTPGGYMSVA	60
Sbjct	1	MLDLFADAEFWQEPLAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTPGGYMSVA	60
Query	61	MTNCGHLGWITHRQGYLYSPIDPQTNKPWPAMPQSFHNLQRAATAAGYPDFQPDACLIN	120
Sbjct	61	MTNCGHLGWITHRQGYLYSPIDPQTNKPWPAMPQSFHNLQRAATAAGYPDFQPDACLIN	120
Query	121	RYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNPLKRLLEHGDVVVWGG	180
Sbjct	121	RYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNPLKRLLEHGDVVVWGG	180
Query	181	ESRLFYHGIQPLKAGFHPLTIDCRYNLTFRQAGKKE	216
Sbjct	181	ESRLFYHGIQPLKAGFHPLTIDCRYNLTFRQAGKKE	216

> [sp|P0CAT7.1|ALKB_CAUCR](#) RecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB
homolog
[sp|B8GWW6.2|ALKB_CAUCN](#) RecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB
homolog
Length=220

Score = 144 bits (364), Expect = 6e-41, Method: Compositional matrix adjust.
Identities = 79/187 (42%), Positives = 107/187 (57%), Gaps = 5/187 (3%)

Query	27	FNAAEQLIRDINDVASQSPFRQMVTPGGYMSVAMTNCGLGWITHRQGYLYSPIDPQTN	86
-------	----	--	----

Firefox ▾

NCBI Blast: Protein Sequence (216 letters) +

blast.ncbi.nlm.nih.gov/Blast.cgi

Google

> [sp|Q13686.2|ALKB1_HUMAN](#) **GM** RecName: Full=Alkylated DNA repair protein alkB homolog 1; AltName: Full=Alpha-ketoglutarate-dependent dioxygenase ABH1; AltName: Full=DNA lyase ABH1 Length=389

GENE ID: 8846 [ALKB1](#) | alkB, alkylation repair homolog 1 (E. coli) [Homo sapiens] ([Over 10 PubMed links](#))

Score = 70.9 bits (172), Expect = 4e-13, Method: Compositional matrix adjust.
Identities = 33/98 (34%), Positives = 54/98 (55%), Gaps = 0/98 (0%)

Query 92 MPQSFHNLCQRAATAAGYPDFQPDACLINRYAPGAKLSLHQQDKDEPDLRAPIVSVSLGLP 151
P L ++ A A G+ DF+ +A ++N Y + L +H D+ E D P++S S G
Sbjct 192 FPSDLGFLSEQVAAACGFEDFRAEAGILNYYRLDSTLGIHVDRSELDHRSKPLLSFSFGQS 251

Query 152 AIFQFGGLKRNPLKRLLEHGDVVVWGGESRLFYHGI 189
AIF GGL+R++ + + GD+++ G SRL H +
Sbjct 252 AIFLLGGLQRDEAPTAMFMHSGDIMIMSGFSRLLNHAV 289

> [sp|O60066.2|ALKBH_SCHPO](#) **G** RecName: Full=Alpha-ketoglutarate-dependent dioxygenase abh1; AltName: Full=Alkylated DNA repair protein alkB homolog Length=297

GENE ID: 2539935 [SPBC13G1.04c](#) | alpha-ketoglutarate-dependent dioxygenase [Schizosaccharomyces pombe 972h-] ([10 or fewer PubMed links](#))

Score = 68.6 bits (166), Expect = 1e-12, Method: Compositional matrix adjust.
Identities = 36/124 (29%), Positives = 63/124 (51%), Gaps = 1/124 (1%)

Query 67 LGWTHRQGYLYSPIDPQINKPWPAMPQSFHNLCQRAAT-AAGYPDFQPDACLINRYAPG 125
L W T + Y ++ + P P+ + ++ + + ++ +A ++N Y+PG
Sbjct 135 LRWVTILGEQYDWTITKEYPDPSKSPGFPKDLGDFVEKVVKESTDFLHWKAEAAIVNFYSPG 194

Query 126 AKLSLHQQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNPLKRLLEHGDVVVWGGESRLF 185
LS H D+ E DL P++S+S+GL I+ G R++ L L GDVV+ G SR
Sbjct 195 DTLSAHIDSEEDLTPLISLSMGLDCIYLIGTESRSEKPSALRLHSGDVVIMTGTSRKA 254

Query 186 YHGI 189
+H +
Sbjct 255 FHAV 258

> [sp|Q54N08.1|ALKB_DICDI](#) RecName: Full=Alpha-ketoglutarate-dependent dioxygenase alkB; AltName: Full=Alkylated DNA repair protein alkB Length=393

Score = 68.6 bits (166), Expect = 2e-12, Method: Compositional matrix adjust.
Identities = 36/123 (29%), Positives = 60/123 (49%), Gaps = 1/123 (1%)

Query 67 LGWTHRQGYLYSPIDPQINKPWPAMPQSFHNLCQRAATAAGYPDFQPDACLINRYAPGA 126
L W+T Y ++P + + + P L Q+ A A + + +A +N Y+ +
Sbjct 215 LAWSTLGYQYQWTP-RLYSEEFYEEFPDDLQELVQKIAIAIKFDPYVAEAAIVNFYSEDS 273

Query 127 KLSLHQQDKDEPDLRAPIVSVSLGLPAIFQFGGLKRNPLKRLLEHGDVVVWGGESRLFY 186
+ H D E ++ PI+S+S G A+F G R+ L + GD+V+ GG SR Y
Sbjct 274 IMGGHLDDAEQEMEKPIISISFGSTAVFLMGAETRDIAFPVPLFIRSGDIVIMGGRSRYCY 333

Query 187 HGI 189
HG+
Sbjct 334 HGV 336

> [sp|Q80Y20.1|ALKB8_MOUSE](#) **GM** RecName: Full=Alkylated DNA repair protein alkB homolog 8; AltName: Full=Probable alpha-ketoglutarate-dependent dioxygenase ABH8; AltName: Full=Scadenosul-L-methionine-dependent tRNA

Firefox

RecName: Full=Alkylated DNA repair pr... +

www.ncbi.nlm.nih.gov/protein/12643239?report=genbank&log\$=proalign&blast_rank=6&RID=1ZKJE2S701N

NCBI Resources How To My NCBI Sign In

Protein Protein Search

Limits Advanced Help

Display Settings: ☒ GenPept Send to: ☐

Change region shown

Customize view

Analyze this sequence ☐

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Articles about the ALKBH1 gene ☐

Homology modeling and function prediction of hABH1, involving in repair o [Interdiscip Sci. 2011]

Human AlkB homologue 1 (ABH1) exhibits DNA lyase activity at abasic [DNA Repair (Amst). 2010]

Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-n [J Biol Chem. 2008]

See all...

Identical proteins for Q13686.2 ☐

alkB, alkylation repair homolog 1 (E. [EAW81300]

alkylated DNA repair protein alkB ho [NP_006011]

ABH [Homo sapiens] [AAF01478]

See all...

Reference sequence information ☐

RefSeq protein

See the reference protein sequence for alkylated DNA repair protein alkB homolog 1

RecName: Full=Alkylated DNA repair protein alkB homolog 1; AltName: Full=Alpha-ketoglutarate-dependent dioxygenase ABH1; AltName: Full=DNA lyase ABH1

UniProtKB/Swiss-Prot: Q13686.2

FASTA Graphics

Go to: ☐

LOCUS ALKB1_HUMAN 389 aa linear PRI 13-JUN-2012

DEFINITION RecName: Full=Alkylated DNA repair protein alkB homolog 1; AltName: Full=Alpha-ketoglutarate-dependent dioxygenase ABH1; AltName: Full=DNA lyase ABH1.

ACCESSION Q13686

VERSION Q13686.2 GI:12643239

DBSOURCE UniProtKB: locus ALKB1_HUMAN, accession [Q13686](#); class: standard.

extra accessions: Q8TAU1, Q9ULA7

created: Dec 1, 2000.

sequence updated: Dec 1, 2000.

annotation updated: Jun 13, 2012.

xrefs: [X91992.1](#), [CAA63047.1](#), [AC008044.4](#), [AAF01478.1](#), [BC025787.1](#), [AAH25787.1](#), [S64736](#), [NP_006011.2](#)

xrefs (non-sequence databases): IPI: IPI00014482, UniGene: [Hs.94542](#), ProteinModelPortal: Q13686, STRING: Q13686, DMD: 12643239, PRIDE: Q13686, DNASU: 8846, Ensembl: ENSN00000216489, Ensembl: ENSP00000216489, Ensembl: ENSG00000100601, GeneID: [8846](#), KEGG: hsa:8846, UCSC: uc001xuc.1, CTD: 8846, GeneCards: GC14M078138, H-InvDB: [HIX0011855](#), HGNC: [17911](#), MIM: [605345](#), neXtProt: NX_Q13686, PharmGKB: PA134906996, eggNOG: COG3145, GeneTree: ENSGT00390000004599, HOGENOM: HOG000033905, HOVERGEN: HBG050487, InParanoid: Q13686, KO: K10765, OMA: HYNWDSK, OrthoDB: EOG4868CJ, PhylomeDB: Q13686, NextBio: 33208, ArrayExpress: Q13686, Bgee: Q13686, CleanEx: HS_ALKBH1, Genevestigator: Q13686, GermOnline: ENSG00000100601, GO: [0005739](#), GO: [0005634](#), GO: [0003906](#), GO: [0008198](#), GO: [0070579](#), GO: [0016702](#), GO: [0006307](#), GO: [0080111](#), GO: [0070989](#), GO: [0042245](#), InterPro: [IPR004574](#), InterPro: [IPR005123](#), Pfam: PF13532, TIGRFAMs: TIGR00568, PROSITE: PS51471

KEYWORDS Complete proteome; Dioxygenase; DNA damage; DNA repair; Iron; Lyase; Metal-binding; Mitochondrion; Multifunctional enzyme;

Firefox

RecName: Full=Alkylated DNA repair pr... +

www.ncbi.nlm.nih.gov/protein/12643239?report=fasta

NCBI Resources How To My NCBI Sign In

Protein Protein Search Limits Advanced Help

Display Settings: ☒ FASTA Send to: ☐

RecName: Full=Alkylated DNA repair protein alkB homolog 1; AltName: Full=Alpha-ketoglutarate-dependent dioxygenase ABH1; AltName: Full=DNA lyase ABH1

UniProtKB/Swiss-Prot: Q13686.2

[GenPept](#) [Graphics](#)

```
>gi|12643239|sp|Q13686.2|ALKB1_HUMAN RecName: Full=Alkylated DNA repair protein
alkB homolog 1; AltName: Full=Alpha-ketoglutarate-dependent dioxygenase ABH1;
AltName: Full=DNA lyase ABH1
MGKMAAAGSVATLATEPGEDAFRKLFRFYRQSRPGTADLEGVIDFSAHAARGKGPQAQKVIKSQLNVS
SVSEQNAYRAGLQPVSKWQAYGLKGYPGFIFIPNPFPGYQWHWVKQCLKLYSQKPNVCNLDKHSKEET
QDLWEQSKEFLRYKEATKRRPRSLEKLWVTVGYHYNWDSKKYSADHYTFPPSDLGFLSEQVAAACGFE
DFRAEAGILNYYRLDSTLGIHVDRSELDHSPKLLSFSFGQSAIFLLGGLQRDEAPTAMFMHSGDIMIMSG
FSRLLNHAVPRVLPNPEGEGLPHCLEAPLPAVLPRDSMVPEPCSMEDWQVCASYLKTARVNMTVRQVLATD
QNFPLEPIEDEKRDISTEGFCHLDDQNSEVKRARINPDS
```

Analyze this sequence

- [Run BLAST](#)
- [Identify Conserved Domains](#)
- [Highlight Sequence Features](#)
- [Find in this Sequence](#)

Articles about the ALKBH1 gene

- [Homology modeling and function prediction of hABH1, involving in repair o \[Interdiscip Sci. 2011\]](#)
- [Human AlkB homologue 1 \(ABH1\) exhibits DNA lyase activity at abasic \[DNA Repair \(Amst\). 2010\]](#)
- [Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-n \[J Biol Chem. 2008\]](#)

[See all...](#)

Identical proteins for Q13686.2

- [alkB, alkylation repair homolog 1 \(E. \[EAW81300\]](#)
- [alkylated DNA repair protein alkB ho \[NP_006011\]](#)
- [ABH \[Homo sapiens\] \[AAF01478\]](#)

[See all...](#)

Reference sequence information

[RefSeq protein](#)

See the reference protein sequence for alkylated DNA repair protein alkB homolog 1 (NP_006011.2).

Firefox

RecName: Full=Alkylated DNA repair pr... +

www.ncbi.nlm.nih.gov/protein/12643239?report=genbank&log\$=proalign&blast_rank=6&RID=1ZKJE2S701N

PROSITE:PS51471

KEYWORDS Complete proteome; Dioxxygenase; DNA damage; DNA repair; Iron; Lyase; Metal-binding; Mitochondrion; Multifunctional enzyme; Nucleus; Oxidoreductase; Polymorphism; Reference proteome; RNA repair.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (residues 1 to 389)

AUTHORS Wei,Y.F., Carter,K.C., Wang,R.P. and Shell,B.K.

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

JOURNAL Nucleic Acids Res. 24 (5), 931-937 (1996)

PUBMED [8600462](#)

REMARK NUCLEOTIDE SEQUENCE [MRNA].;
TISSUE=Synovial sarcoma

REFERENCE 2 (residues 1 to 389)

AUTHORS Heilig,R., Eckenberg,R., Petit,J.L., Fonknechten,N., Da Silva,C., Cattolico,L., Levy,M., Barbe,V., de Berardinis,V., Ureta-Vidal,A., Pelletier,E., Vico,V., Anthouard,V., Rowen,L., Madan,A., Qin,S., Sun,H., Du,H., Pepin,K., Artiguenave,F., Robert,C., Cruaud,C., Bruls,T., Jaillon,O., Friedlander,L., Samson,G., Brottier,P., Cure,S., Segurens,B., Aniere,F., Samain,S., Crespeau,H., Abbasi,N., Aiach,N., Boscus,D., Dickhoff,R., Dors,M., Dubois,I., Friedman,C., Gouyvenoux,M., James,R., Madan,A., Mairey-Estrada,B., Mangenot,S., Martins,N., Menard,M., Oztas,S., Ratcliffe,A., Shaffer,T., Trask,B., Vacherie,B., Bellemere,C., Belser,C., Besnard-Gonnet,M., Bartol-Mavel,D., Boutard,M., Briez-Silla,S., Combette,S., Dufosse-Laurent,V., Ferron,C., Lechaplais,C., Louesse,C., Muselet,D., Magdelenat,G., Pateau,E., Petit,E., Sirvain-Trukniewicz,P., Trybou,A., Vega-Czarny,N., Bataille,E., Bluet,E., Bordelais,I., Dubois,M., Dumont,C., Guerin,T., Haffray,S., Hammadi,R., Muanga,J., Pellouin,V., Robert,D., Wunderle,E., Gauguier,G., Roy,A., Sainte-Marthe,L., Verdier,J., Verdier-Discala,C., Hillier,L., Fulton,L., McPherson,J., Matsuda,F., Wilson,R., Scarpelli,C., Gyapay,G., Wincker,P., Saurin,W., Quetier,F., Waterston,R., Hood,L. and Weissenbach,J.

TITLE The DNA sequence and analysis of human chromosome 14

JOURNAL Nature 421 (6923), 601-607 (2003)

PUBMED [12508121](#)

REMARK NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

REFERENCE 3 (residues 1 to 389)

AUTHORS Gerhard,D.S., Wagner,L., Feingold,E.A., Shenmen,C.M., Grouse,L.H., Schuler,G., Klein,S.L., Old,S., Rasooly,R., Good,P., Guyer,M., Peck,A.M., Derge,J.G., Lipman,D., Collins,F.S., Jang,W., Sherry,S., Feolo,M., Misquitta,L., Lee,E., Rotmistrovsky,K., Greenhut,S.F.,

More about the ALKBH1 gene

See the reference protein sequence for alkylated DNA repair protein alkB homolog 1 (NP_006011.2).

More about the ALKBH1 gene

This gene encodes a homolog to the E. coli alkB gene product. The E. coli alkB protein is part of the adaptive response mechanism of DNA alk... Also Known As: ABH, ABH1, ALKBH, alkB,...

Homologs of the ALKBH1 gene

The ALKBH1 gene is conserved in chimpanzee, dog, cow, mouse, rat, chicken, zebrafish, fruit fly, mosquito, and C.elegans.

LinkOut to external resources

Q13686 [Domain Mapping of Disease Mut...]

Transcript/Protein Information [PANTHER Classification System]

Pathway Commons [Pathway Commons]

ABH1 antibody [Abcam plc]

ABH1 antibody [Abcam plc]

ABH1 antibody [hABH1-151] [Abcam plc]

ABH1 antibody [Abcam plc]

protein and peptide [ExactAntigen/Labome]

antibody [ExactAntigen/Labome]

others [ExactAntigen/Labome]

siRNA and shRNA [ExactAntigen/Labome]

cDNA clone [ExactAntigen/Labome]

Firefox

RecName: Full=Alkylated DNA repair pr... +

www.ncbi.nlm.nih.gov/protein/12643239?report=genbank&log\$=proalign&blast_rank=6&RID=1ZKJE2S701N

Peck,A.M., Derge,J.G., Lipman,D., Collins,F.S., Jang,W., Sherry,S., Feolo,M., Misquitta,L., Lee,E., Rotmistrovsky,K., Greenhut,S.F., Schaefer,C.F., Buetow,K., Bonner,T.I., Haussler,D., Kent,J., Kiekhaus,M., Furey,T., Brent,M., Prange,C., Schreiber,K., Shapiro,N., Bhat,N.K., Hopkins,R.F., Hsie,F., Driscoll,T., Soares,M.B., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Piao,Y., Dudekula,D.B., Ko,M.S., Kawakami,K., Suzuki,Y., Sugano,S., Gruber,C.E., Smith,M.R., Simmons,B., Moore,T., Waterman,R., Johnson,S.L., Ruan,Y., Wei,C.L., Mathavan,S., Gunaratne,P.H., Wu,J., Garcia,A.M., Hulyk,S.W., Fuh,E., Yuan,Y., Sneed,A., Kowis,C., Hodgson,A., Muzny,D.M., McPherson,J., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madari,A., Young,A.C., Wetherby,K.D., Granite,S.J., Kwong,P.N., Brinkley,C.P., Pearson,R.L., Bouffard,G.G., Blakesly,R.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Griffith,M., Griffith,O.L., Krzywinski,M.I., Liao,N., Morin,R., Palmquist,D., Petrescu,A.S., Skalska,U., Smailus,D.E., Stott,J.M., Schnerch,A., Schein,J.E., Jones,S.J., Holt,R.A., Baross,A., Marra,M.A., Clifton,S., Makowski,K.A., Bosak,S. and Malek,J.

CONSRTM MGC Project Team

TITLE The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC)

JOURNAL Genome Res. 14 (10B), 2121-2127 (2004)

PUBMED 15489334

REMARK NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].; TISSUE=Pancreas

Erratum:[Genome Res. 2006 Jun;16(6):804. Morrin, Ryan [corrected to Morin, Ryan]]

REFERENCE 4 (residues 1 to 389)

AUTHORS Tsujikawa,K., Koike,K., Kitae,K., Shinkawa,A., Arima,H., Suzuki,T., Tsuchiya,M., Makino,Y., Furukawa,T., Konishi,N. and Yamamoto,H.

TITLE Expression and sub-cellular localization of human ABH family molecules

JOURNAL J. Cell. Mol. Med. 11 (5), 1105-1116 (2007)

PUBMED 17979886

REMARK SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.

REFERENCE 5 (residues 1 to 389)

AUTHORS 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., Slupphaug,G., Otterlei,M. and Krokan,H.E.

TITLE Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA

JOURNAL J. Biol. Chem. 283 (36), 25046-25056 (2008)

PUBMED 18603530

REMARK FUNCTION, COFACTOR, SUBCELLULAR LOCATION, MUTAGENESIS OF ILE-218; HIS-231; ASP-233; HIS-287; ARG-338 AND ARG-344, MASS SPECTROMETRY, AND TISSUE SPECIFICITY

cDNA clone [ExactAntigen/Labome]

Related information

BLink

Related Sequences

Identical Proteins

BioSystems

CDD Search Results

Conserved Domains (Concise)

Conserved Domains (Full)

Domain Relatives

Full text in PMC

Gene

Gene Genotype

GeneView in dbSNP

Map Viewer

OMIM

Protein (RefSeq)

PubMed

PubMed (Weighted)

Related Structures (List)

Related Structures (Summary)

Taxonomy

Recent activity

Turn Off Clear

RecName: Full=Alkylated DNA repair protein alkB homolog 1; AltName: Full=Alpha- Protein

UNKNOWN19899 BLAST

UNKNOWN10268 BLAST

UNKNOWN13424 BLAST

Firefox

RecName: Full=Alkylated DNA repair pr... +

www.ncbi.nlm.nih.gov/protein/12643239?report=genbank&log\$=protalign&blast_rank=6&RID=1ZKJE2S701N

Google

UNKNOWN13424 BLAST

UNKNOWN28903 BLAST

See more...

REMARK FUNCTION, COFACTOR, SUBCELLULAR LOCATION, MUTAGENESIS OF ILE-218; HIS-231; ASP-233; HIS-287; ARG-338 AND ARG-344, MASS SPECTROMETRY, AND TISSUE SPECIFICITY.

REFERENCE 6 (residues 1 to 389)

AUTHORS Muller,T.A., Meek,K. and Hausinger,R.P.

TITLE Human AlkB homologue 1 (ABH1) exhibits DNA lyase activity at abasic sites

JOURNAL DNA Repair (Amst.) 9 (1), 58-65 (2010)

PUBMED [19959401](#)

REMARK COMMENT FUNCTION, MUTAGENESIS OF HIS-231; ASP-233 AND HIS-287, AND SUBUNIT. On Jul 17, 2007 this sequence version replaced gi:[2134723](#). [FUNCTION] Dioxygenase that repairs alkylated single-stranded DNA and RNA containing 3-methylcytosine by oxidative demethylation. Requires molecular oxygen, alpha-ketoglutarate and iron. May have a role in placental trophoblast lineage differentiation (By similarity). Has DNA lyase activity and introduces double-stranded breaks at abasic sites. Cleaves both single-stranded DNA and double-stranded DNA at abasic sites, with the greatest activity towards double-stranded DNA with two abasic sites. DNA lyase activity does not require alpha-ketoglutarate and iron. [CATALYTIC ACTIVITY] The C-O-P bond 3' to the apurinic or apyrimidinic site in DNA is broken by a beta-elimination reaction, leaving a 3'-terminal unsaturated sugar and a product with a terminal 5'-phosphate. [COFACTOR] Binds 1 Fe(2+) ion per subunit. [SUBUNIT] Monomer. Interacts with DNAJB6 (By similarity). [SUBCELLULAR LOCATION] Mitochondrion. Nucleus (By similarity). Note=Mainly localizes in euchromatin, largely excluded from heterochromatin and nucleoli (By similarity). [TISSUE SPECIFICITY] Ubiquitous. [SIMILARITY] Belongs to the alkB family. [SIMILARITY] Contains 1 Fe2OG dioxygenase domain. [CAUTION] Detected in cytoplasm and nucleus When expressed as fusion protein with an N-terminal tag, or when the first 26 N-terminal residues are removed (PubMed:17979886 and PubMed:18603530). The endogenous, unmodified protein localizes to mitochondria (PubMed:18603530). [SEQUENCE CAUTION] Sequence=CAA63047.1; Type=Frameshift; Positions=Several.

FEATURES Location/Qualifiers

source 1..389 /organism="Homo sapiens" /db_xref="taxon:[9606](#)"

gene 1..389 /gene="ALKBH1" /gene_synonym="ABH" /gene_synonym="ABH1" /gene_synonym="ALKBH"

Protein 1..389

Firefox

RecName: Full=Alkylated DNA repair pr... +

www.ncbi.nlm.nih.gov/protein/12643239?report=genpept

Google

Protein

1..389

/gene_synonym="ALKBH"

/gene="ALKBH1"

/gene_synonym="ABH"

/gene_synonym="ABH1"

/gene_synonym="ALKBH"

/product="Alkylated DNA repair protein alkB homolog 1"

/EC_number="1.14.11.-"

/EC_number="4.2.99.18"

/note="Alpha-ketoglutarate-dependent dioxygenase ABH1; DNA lyase ABH1"

/UniProtKB_evidence="Evidence at protein level"

Region

1..389

/gene="ALKBH1"

/gene_synonym="ABH"

/gene_synonym="ABH1"

/gene_synonym="ALKBH"

/region_name="Mature chain"

/experiment="experimental evidence, no additional details recorded"

/note="Alkylated DNA repair protein alkB homolog 1."

/FTId=PRO_0000066668."

Region

119..289

/gene="ALKBH1"

/gene_synonym="ABH"

/gene_synonym="ABH1"

/gene_synonym="ALKBH"

/region_name="2OG-FeII_Oxy"

/note="2OG-Fe(II) oxygenase superfamily; cl15773"

/db_xref="CDD:210173"

Region

133

/gene="ALKBH1"

/gene_synonym="ABH"

/gene_synonym="ABH1"

/gene_synonym="ALKBH"

/region_name="Conflict"

/experiment="experimental evidence, no additional details recorded"

/note="K -> T (in Ref. 1; CAA63047)."

Region

135

/gene="ALKBH1"

/gene_synonym="ABH"

/gene_synonym="ABH1"

/gene_synonym="ALKBH"

/region_name="Variant"

/experiment="experimental evidence, no additional details recorded"

/note="M -> I (in dbSNP:rs17825440). /FTId=VAR_048221."

Site

170

Firefox

RecName: Full=Alkylated DNA repair pr... +

www.ncbi.nlm.nih.gov/protein/12643239?report=genpept

Google

Site

338

/gene="ALKBH1"

/gene_synonym="ABH"

/gene_synonym="ABH1"

/gene_synonym="ALKBH"

/site_type="mutagenized"

/experiment="experimental evidence, no additional details recorded"

/note="R->A: Reduced Fe2OG dioxygenase activity."

Site

344

/gene="ALKBH1"

/gene_synonym="ABH"

/gene_synonym="ABH1"

/gene_synonym="ALKBH"

/site_type="mutagenized"

/experiment="experimental evidence, no additional details recorded"

/note="R->A: Reduced Fe2OG dioxygenase activity."

Region

388

/gene="ALKBH1"

/gene_synonym="ABH"

/gene_synonym="ABH1"

/gene_synonym="ALKBH"

/region_name="Conflict"

/experiment="experimental evidence, no additional details recorded"

/note="D -> H (in Ref. 3; AAH25787)."

ORIGIN

1 mgkmaaavgs vatlatepge dafrklfrfy rqsrrgtadl egvidfsaah aargkpggaq

61 kviksqnlvs svseqnayra glqpvskwqa yglkgypgfi fipnpflpgy qwhwvkqcl

121 lysqkpnvcn ldkhmskeet qdlweqskef lrykeatkrr prslleklrw vtvgyhynwd

181 skkysadhyt pfpsdlgfls eqvaaacgfe dfraeagiln yyrldstlgi hvdrsel dhs

241 kp11sfsgq saifllgglq rdeaptamfm hsgdimimg fsrllnhavg rvlpnpegeg

301 lphcleaplp avlprds mve pcsmedwqvc asylktarvn mtrqvlatd qnfplepied

361 ekrdistegf chlddqns ev krarinpds

//

You are here: NCBI > Proteins > Protein Database

Write to the Help Desk

GETTING STARTED

NCBI Education

NCBI Help Manual

NCBI Handbook

Training & Tutorials

RESOURCES

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

POPULAR

PubMed

Nucleotide

BLAST

PubMed Central

Gene

FEATURED

Genetic Testing Registry

PubMed Health

GenBank

Reference Sequences

Map Viewer

NCBI INFORMATION

About NCBI

Research at NCBI

NCBI Newsletter

NCBI FTP Site

NCBI on Facebook