Chapter 2

UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View

Emmanuel Boutet, Damien Lieberherr, Michael Tognolli, Michel Schneider, Parit Bansal, Alan J. Bridge, Sylvain Poux, Lydie Bougueleret, and Ioannis Xenarios

Abstract

The Universal Protein Resource (UniProt, http://www.uniprot.org) consortium is an initiative of the SIB Swiss Institute of Bioinformatics (SIB), the European Bioinformatics Institute (EBI) and the Protein Information Resource (PIR) to provide the scientific community with a central resource for protein sequences and functional information. The UniProt consortium maintains the UniProt KnowledgeBase (UniProtKB), updated every 4 weeks, and several supplementary databases including the UniProt Reference Clusters (UniRef) and the UniProt Archive (UniParc).

The Swiss-Prot section of the UniProt KnowledgeBase (UniProtKB/Swiss-Prot) contains publicly available expertly manually annotated protein sequences obtained from a broad spectrum of organisms. Plant protein entries are produced in the frame of the Plant Proteome Annotation Program (PPAP), with an emphasis on characterized proteins of *Arabidopsis thaliana* and *Oryza sativa*. High level annotations provided by UniProtKB/Swiss-Prot are widely used to predict annotation of newly available proteins through automatic pipelines.

The purpose of this chapter is to present a guided tour of a UniProtKB/Swiss-Prot entry. We will also present some of the tools and databases that are linked to each entry.

Key words Swiss-Prot, TrEMBL, UniProt, Protein database, Amino-acid sequence, Manual annotation

1 Introduction

In late 2002 the SIB Swiss Institute of Bioinformatics (SIB), the European Bioinformatics Institute (EBI) and the Protein Information Resource (PIR) (*see* Note 1) joined forces by creating the Universal Protein Resource (UniProt) consortium [1]. The aim of this consortium is to provide high quality protein databases that are freely accessible to the scientific community.

The centerpiece of UniProt is the UniProt Knowledgebase (UniProtKB, http://www.uniprot.org), a comprehensive and

DOI 10.1007/978-1-4939-3167-5_2, © Springer Science+Business Media New York 2016

David Edwards (ed.), Plant Bioinformatics: Methods and Protocols, Methods in Molecular Biology, vol. 1374,

annotated protein sequence knowledgebase, which consists of two sections: UniProtKB/Swiss-Prot, containing manually expertly annotated entries, and UniProtKB/TrEMBL, containing computer translation and annotation of CoDing Sequences (CDS) extracted from the European Molecular Biology Laboratory nucleotide sequence database (EMBL) [2, 3] as well as sequences and annotation imported from Ensembl (http://www.ensembl. org), EnsemblGenomes (http://ensemblgenomes.org) including EnsemblPlants (http://plants.ensembl.org), and in the future, from RefSeq (http://www.ncbi.nlm.nih.gov/refseq/). Taking advantage of the expertly curated UniProtKB/Swiss-Prot section, automatic annotation procedures based on well described proteins are created and maintained to improve the annotation of related proteins in the UniProtKB/TrEMBL section. UniProtKB entries contain information curated by biologists or produced by annotation rules, and provide users with cross-links to about 140 external databases and give access to additional information or tools. UniProtKB/Swiss-Prot contributes actively to the "Gene Ontology" (GO, 10) annotation effort of proteins by manually assigning GO terms during the annotation process.

UniProtKB/Swiss-Prot is characterized by extended expert annotation (sequence properties, corresponding literature, etc.), minimal redundancy (separate entries for the same gene product in a given species and same cultivar/isolate are merged into a single protein entry), integration with other databases (cross-links to other life science databases including sequence-related databases as well as specialized data collections) and documentation (large number of index files and specialized documentation files) (*see* **Note 2**).

UniProtKB/TrEMBL, a computer-annotated database, mainly consists of translations of all coding sequences (CDS) proposed by the submitters to the EMBL/GenBank/DDBJ nucleotide databases, which are not integrated into UniProtKB/Swiss-Prot, and by proteomes imported from Ensembl and EnsemblPlants. Some additional protein sequences are also extracted from the literature or directly submitted to UniProtKB. In addition to the preliminary information given by the submitters, UniProtKB/TrEMBL entries are processed according to automatic annotation procedures such as: (i) transfer of general annotation, domains and functional sites from well-characterized UniProtKB/Swiss-Prot entries belonging to protein family groups defined by InterPro [4], (ii) removal of redundancy by merging identical full-length sequences from the same organism, (iii) attribution of evidence to identify the source of individual data items (*see* Note 3).

In addition to UniProtKB, the UniProt consortium maintains several other protein databases, including:

 The UniProt Archive (UniParc), which stores and maps all publicly available protein sequences from numerous databases, including UniProtKB, RefSeq, Patent offices, etc. (obsolete data excluded from UniProtKB are also present in UniParc) The UniProt Reference Clusters (UniRef), which consists of clusters of sequences sharing 100 % identity for UniRef100, 90 % for UniRef90 and 50 % for UniRef50 (*see* Note 4). These databases are based on both UniProtKB and UniParc.

The Swiss-Prot group has initiated the Plant Proteome Annotation Program (PPAP) in 2001 [5] (http://www.uniprot. org/program/plants/). The current priority of this program is to annotate the proteomes of *Arabidopsis thaliana* and *Oryza sativa*, but without neglecting to annotate the proteins from other plant species. Our goals are the annotation of characterized plant specific and plant family proteins according to the Swiss-Prot standards [3]. At the beginning of March 2014 (UniProt release 2014_02), 34,824 plant sequence entries are present in UniProtKB/Swiss-Prot. Among them 12,665 are from *A. thaliana* and 3130 from *O. sativa*. In UniProtKB/Swiss-Prot, more than 1976 different plant species are present with at least one annotated protein (up-todate statistics are available at http://www.uniprot.org/statistics/, http://web.expasy.org/docs/relnotes/relstat.html and http:// www.uniprot.org/program/plants/statistics).

To cope with the large and growing amount of sequenced genomes, UniProt assigns unique proteome identifiers giving the possibility to select proteins of a given organism. A subset of well-studied or biomedically and biotechnologically interesting organisms, selected to provide broad coverage of the tree of life, are manually defined as standard for a particular user community, and their proteome are "Reference proteomes" (*see* **Note 5**).

2 Materials

UniProtKB is hosted by uniprot.org (*see* Note 6). This chapter will always refer to the UniProtKB interface format used by the uniprot.org server (http://www.uniprot.org/), and will focus on UniProtKB/Swiss-Prot entries. The database is updated every four weeks. It is possible to download a local version of UniProtKB (*see* Notes 7 and 8).

2.1 UniProtKB Entries The main distribution format of UniProtKB is a custom text-based format. Entries are represented by lines beginning with a two-letter code that identifies the type of data contained in the line. Each line follows a strictly defined format and the lines themselves are organized in such a way as to be easily legible to human users and simple to parse by computer programs (http://www.expasy.org/sprot/userman.html#entrystruc). However, UniProtKB proteins are also available in the more modern and structured XML/RDF format for computational use (http://www.uniprot.org/docs/uniprot.xsd).

- 2.1.2 Web View When accessing UniProtKB entries from the uniprot.org server, the default format is topic-wise organized in a user-friendly format when compared to the text-based format (*see* Fig. 1). The general elements of an entry in the uniprot.org view format are (from top to bottom): (i) UniProt header and search tool, (ii) UniProt tools (BLAST, alignment, mapping/retrieval in batch), (iii) general help, contact and basket tools, (iv) the header of the UniProtKB entry, (v) tools applicable to the current UniProtKB entry, (vi) current UniProtKB entry centric comment, feedback and external data tools, (vii) UniProtKB entry's section navigation bar organized by topics, (viii) the content of the current UniProtKB entry, (ix) details about the history of the current UniProtKB entry.
- 2.1.3 Content of an Entry In most cases, each entry corresponds to a protein sequence encoded by a single gene locus (*see* Note 9). However, a few protein entries contain different coding loci merged into a single record when these loci are highly similar (e.g., histones, ubiquitins). References to residue positions within a sequence are made using sequential numbering starting with 1 at the N-terminal position. Displayed sequences correspond to the precursor forms of proteins, before posttranslational modifications and processing.

2.2 Tools and Databases Linked to UniProtKB The uniprot.org website provides dedicated tools designed to exploit both protein sequences (BLAST, [6], alignments, database identifier mapping tool) and functional annotations (friendly but advanced search tool). SIB has developed the Expert Protein Analysis System proteomic server (ExPASy), which is another entry point to UniProtKB [7–9]. On http://www.expasy.org/, tools are available to deal with several aspects of protein analysis, including BLAST search, proteomics and sequence analysis, and take into account all splice variants as annotated in UniProtKB (*see* Note 10). Results obtained by these tools or links from other specific databases points to the corresponding UniProtKB entries.

3 Methods

3.1 Introduction The main goal of UniProt is to provide a central resource for protein sequences and functional annotation. Together with UniProtKB/TrEMBL, UniProtKB/Swiss-Prot contains all known proteins, without species restriction. Currently the plant protein entries represent about 20 % of eukaryotes proteins and 7 % of the total content of UniProtKB/Swiss-Prot and our main effort is focused on the annotation of plant specific proteins characterized in literature from *Arabidopsis thaliana* and *Oryza sativa*. Any new genome fully sequenced, deposited in the public nucleotide database (EMBL/GenBank/DDBJ) and for which a gene prediction



| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|----------------------------|-----------------------------|---------------|---------------------------|------------------|-----------------------|
| Metal binding ⁱ | 391 | 1 | Zinc; catalytic | | |
| Metal binding ⁱ | 393 | 1 | Zinc; catalytic | | |
| Binding site ⁱ | 393 | 1 | Substrate | | |
| Metal binding ⁱ | 659 | 1 | Zinc; catalytic | | |
| Binding site ¹ | 662 | 1 | Substrate | | |
| Active site ¹ | 681 | 1 | Proton acceptor | | |
| Metal binding ⁱ | 736 | 1 | Zinc; catalytic | | |
| egions | n - 141 | · · · · · · · | | 2 million in the | Feature |
| Feature key | Position | (s) Leng | th Description | Graphical view | identifier |
| Nucleotide bind | ling ⁱ 289 - 296 | 8 | ATP # Reviewed Prediction | | |

Fig. 1 Header of a UniProtKB entry in the uniprot.org display format; partial view (http://www.uniprot.org/uniprot/080452) has been performed will be processed automatically. The predicted set of proteins is added to the UniProtKB/TrEMBL section as soon as the data is publicly available.

One of the great strengths of the UniProt Knowledgebase is the extensive integration and interconnectivity of numerous tools and external databases. The knowledgebase is cross-linked to about 140 other databases while most of the tools are adapted to allow analysis of all spliced isoforms described in the entry.

The UniProt Knowledgebase is constantly evolving and all recent modifications are detailed at http://www.uniprot.org/help/?query=*&fil=section:news while the forthcoming modifications are listed in http://www.uniprot.org/changes.

To further improve the quality of our annotation, we encourage users to submit comments and update requests (http://www.uniprot. org/update?entry= *primary accession number* accessible by the buttons and links present in each UniProtKB entries, *see* Fig. 1 iii and vi).

3.2 Accessing and Analyzing UniProtKB Entries Quick and advanced text search (*see* Fig. 1 i) can be accessed directly from the UniProt home page (http://www.uniprot. org) (*see* Note 11). The advanced text search is designed to help users in writing complex queries by restricting terms to specific fields of the database (*see* Fig. 2 i), organized in the same topics of entry's sections. "Intelligent" filters are suggested to restrict the query with most likely terms (*see* Fig. 2 ii). Proteins of interest can be stored in the "basket" by checking boxes (*see* Figs. 2 iii and 3 i) and clicking on the button "Add to basket" for later comparison or download. When accessing the basket (*see* Fig. 3 ii), previously selected entries are listed and different actions are available: "Align", "BLAST", and "Download" (*see* Fig. 3 iii). The result table can be customized to fit user's requirement (*see* Fig. 4). A drag and drop



Fig. 2 Text search result; partial view. Partial view of the result of a text search made on UniProtKB with "amp deaminase" as query

| | | | | | | (ii) | 🏛 Basket 🖪 🗸 |
|-----------|----------|---------|---|---------------------------------------|--|-------------|----------------|
| | | Ur | niProtKB (3) | UniRef (0) | UniParc (0) | (max | 400 entries) 🗙 |
| 🗶 Columns | * BLAST | Entry | Entry nam | ne | Organism | Rei | move |
| Entry | Entry na | P23109 | AMPD1_H | UMAN | Homo sapiens (Hum | an) | ā |
| Q01432 | AMPD3_I | Q01433 | AMPD2_H | UMAN | Homo sapiens (Hum | an) | Ô |
| (i) | | Q01432 | AMPD3_H | UMAN | Homo sapiens (Hum | an) | □ |
| Q01433 | AMPD2_I | ≡ Align | * BLAST | ± Download | (iii) | Clea | r Full View |
| | | | deaminase i | Format: | Text | G X | |
| ₽23109 | AMPD1_H | UMAN 🏓 | AMP deam 3.5.4.6) (AM deaminase is (Myoadenyla deaminase) | nose I (ce np soform M) ate | FASTA (canonical & isof Tab-delimited Text Excel GFF | orm) (H | luman) |
| Q9DBT5 | AMPD2_M | OUSE 🏂 | AMP deami 3.5.4.6) (AM deaminase is | n ase 2 (EC 1P soform L) | | musculus (1 | Mouse) |

Fig. 3 The UniProt basket. View of the UniProt basket containing three UniProtKB protein entries (e.g., P23109, Q01433, and Q01432)



Fig. 4 The UniProt customization interface. View of the UniProt customization tool



Fig. 5 The UniProt alignment tool. View of the UniProt protein alignment tool

tool makes it possible to change column order (*see* Fig. 4 i). A search engine is available to select for a favorite topic to display in the result table (*see* Fig. 4 ii). Each entry section can also be browsed in details (*see* Fig. 4 iii). When downloading selected entries in "tab-delimited" format, the columns of the output file are the same as the personalized display (*see* Fig. 2 iv). UniProt web services follow the representational state transfer (REST) architectural style to help sharing or storing favorite requests; this also permits easy programmatic access (*see* http://www.uniprot.org/faq/28).

- 2. An alignment tool based on Clustal Omega (https://www.ebi. ac.uk/Tools/msa/clustalo/) is available at http://www.uniprot.org/align/ (see Fig. 5). The alignment output (see Fig. 6) is interactive and gives the possibility to highlight in different colors sequence features (see Fig. 6 ii) annotated in UniProtKB as well as amino acid properties by selecting properties of interest (see Fig. 6 i). When more than two protein sequences are aligned, an alignment tree is also available.
- 3. BLAST is available at http://www.uniprot.org/blast/ (see Fig. 7). Standard parameters can be modified, default settings being: UniProtKB for the data set, 10 for the E-threshold, Matrix auto, no low complexity filtering and gap allowed (see Note 12). The BLAST output (see Fig. 8) gives, on the top, a list of sequences classified by level of similarity to the query, displayed in a graphical view of the query sequence with a similarity-dependent color gradient, and linked to the corresponding UniProtKB entries (see Fig. 8 i). A mechanism to allow the user to toggle between similarity based graphics and e-value based graphics will be soon available. All splice variants

| Display ALLownext ALLOWNEXT ALLOWNEXT ALLOWNEXT Pailog TREE Coll 433 AMPD_HUBAN 1 RESULT INFO Coll 432 AMPD_HUBAN 1 Highlight Coll 432 AMPD_HUBAN 24 Coll 432 AMPD_HUBAN 24 Coll 432 AMPD_HUBAN 24 Coll 432 Coll 432 <thcoll 432<="" th=""> <thcol 432<="" th=""> <thc< th=""><th>Align</th><th></th><th></th><th></th><th></th><th></th></thc<></thcol></thcoll> | Align | | | | | |
|---|--------------------|---------|-----------------|------|--|-----|
| ALIGNMENT ALIGNMENT ALIGNMENT ALIGNMENT REE P23109 ANPD1 HUHAN 1 Q01432 ANPD2 HUHAN 1 HNNROQCLFRLESRCFLAGSLPLAGAGREKGLDVAEPOPSRCRSDSPAVAAVVPARASTPS 60 Q01432 ANPD2 HUHAN 1 HNNROQCLFRLESRCFLAGSLPLAGAGREKGLDVAEFOPSRCRSDSPAVAAVVPARASTPS 60 Q01432 ANPD1 HUHAN 260 DEPKPLPYPNLDEYDAUVVRALLTAGEVKTYTHEPLKFLSSKFOMHULINERDELKEL 319 Annotation Q01432 ANPD2 HUHAN 241 CSSPLEVPPLOEFVANUVRALTINGETKYCHEPULYSKFOMHULINENEEKEL 300 Matural variant Q01432 ANPD3 HUHAN 240 SKNLPPVD FYVLEKVDTH HAA CSKNLPPVD FYVLEKVDTH HAA 300 Natural variant Chain P23109 ANPD3 HUHAN 300 KSNPPHOPTVD FVDRVDVTH HAA WOHABCPTVD FVDRVDVTH HAA 300 Natural variant P23109 ANPD3 HUHAN 300 KSNPPHOPTVD FVDRVDVTH HAA 300 Region Q01433 ANPD3 HUHAN 300 KSNPPHOPTVD FVDRVDVAHADINT HEPDKYNAKVPTOKSELDLYLKTNTNIKGEYFATI 399 Geguence conflict Q01432 ANPD3 HUHAN 400 IKEVGADUVEAKVOTHERDKYNDVGASELRDUYKKSENVUNVKKSELRDUYKKTNTIKGEYFYATI 499 | Display 🗛 | ± Downl | oad | | | |
| TREE P23109 AMPD1 HUMAN I INNRGQGLFFLESSCTLHGSLPLGGGREKGLDVAEPOPSRCRDSPAVAAVPAMASYPS 60 C01432 AMPD2 HUMAN I INNRGQGLFFLESSCTLHGSLPLGGGREKGLDVAEPOPSRCRDSPAVAAVPAMASYPS 60 Highlight Q01432 AMPD2 HUMAN I INNRGQGLFFLESSCTLHGSLPLGGVADVNLALLINGOPVKTYTHRRLKFLSSSFGVMUQULNEHBELKEL 319 Annotation I INNGGGULFFLESSCTLHGSLPYPDLGEVVADVNLALLINGOPVKTYTHRRLKFLESSFSUEHLINEHSEKKEL 300 Sequence Q01432 AMPD2 HUMAN 320 KNNPHRDFYNCKKVDTHIHAL INNCKLLSF IKKSNGLDADEVVYSTKEKNLTLKELFA 379 Chain Q01432 AMPD2 HUMAN 320 KNNPHRDFYNCKKVDTHIHAL INNCKLLSF IKKSNGLDADEVVYSTKEKNLTLKVEFA 379 Chain Q01432 AMPD2 HUMAN 300 KNNPHRDFYNCKNVDVAEPDEPTNVEKNVOQAEELDI VKTVNVGQKCTLEVYFA 461 Sequence Q01432 AMPD2 HUMAN 361 GLMBPTYNVKVOQAEELDI VKTVNVQQKETLESKTV 461 Chain P23109 AMPD1 HUMAN 360 KLWPRD7TNIKKVDTBTHAA 700 700 FKVGADVVAEKELESTVVKYKKSTLLEVF 461 701432 4190 400 700 700 700 700 700 700 700 700 | ALIGNMENT | Aligni | ment | | | |
| RESULT INFO CO1433 AMPD2_HUMAN I INRREQCIF LESSCT.HGSL.PLGAGREKCLEV/AEPOPSRCREDSPAYAAVPAMASYPS GO Highlight CO1433 AMPD3_HUMAN I INRREQCIF LESSCT.HGSL.PFGAGREKCLEV/AEPOPSRCREDSPAYAAVPAMASYPS GO Highlight P23109 AMPD3_HUMAN 260 DEPKPLPYPNLDTFLDDINFLALIAQOPVKTYTHERLKFLSSKFQVHQMLNEKNELALA 400 Annotation I Exercise Co1433 AMPD3_HUMAN 241 COCHAPTINE AMPD3_HUMAN 241 2420 241 241 2420 241 241 241 241 241 241 241 241 241 241 241 241 <th>TREE</th> <th>P23109</th> <th>AMPD1 HUMAN</th> <th>1</th> <th>MNVRIFYS</th> <th>8</th> | TREE | P23109 | AMPD1 HUMAN | 1 | MNVRIFYS | 8 |
| Image: Result INFO Q01432 AMPD3_HUNAN Image: Result INFO Q01432 AMPD3_HUNAN Image: Result INFO Q01432 AMPD1_HUNAN 260 DEPKPLPYPNLDTFLDDMNFLLALIAQGPVKTYTHRRLKFLSSKFQHVNLLMENKELAAQ 401 Annotation Q01432 AMPD3_HUNAN 320 CSEVSLPYPDLGEYVADUWLALIINGPIKSTCYRRLQVLSSKFQHVNLLMENKELAAQ 401 Alternative Q01432 AMPD3_HUNAN 320 CSEVSLPYPDLGEYVADUWLALLINGPIKSTCYRRLKSYQIDADRVVYSTKEKNLLKELA 379 Alternative Q01432 AMPD3_HUNAN 320 CSEVSLPYPNCRKVDTH IMA_GENCKKLLFIKKSYQIDADRVVYSTKEKNLLKELK 379 Chain P23109 AMPD1_HUNAN 360 KINPPRDFYINKRVDTH IMA_GENCKKLLFIKKITYCTEPDEV3EKRORSTILRQVFD 461 Chain P23109 AMPD1_HUNAN 360 KIKHPYDLTVDSLDVHAGRQTFQRFDKYNKNYVGASELRDLYLKTDNYINGEYFATI 439 Goldasi AMPD3_HUNAN 462 SINLTATDSUVTLDVBLDVHAGRQTFVRFTKRDKYNKYVGASELRDLYLKTDNYINGEYFATI 439 Goldasi AMPD3_HUNAN 460 SINLTATDSUVTLDVBLDVHAGRQTFVRFTKRVKNYVGASELRDLYLKTDNYINGEYFATI 439 Goldasi AMPD3_HUNAN 461 SINLTATDSUVTLDVBLDVHAGRGTFNRTKNYVGASELRDLYLKTDNYINGEYFATI 439 Goldasis AMPD3_HUNAN 462 | 2 | Q01433 | AMPD2_HUMAN | 1 | MRNRGQGLFRLRSRCFLHQSLPLGAGRRKGLDVAEPGPSRCRSDSPAVAAVVPAMASYPS | 60 |
| Highlight P23109 AMPD1_HURAN 260 DEPKPLPYPNLDTFLDDINFLLALIAQOPVKTYTHRRLKFLSSKFQUHONLNEHDELKEL 319 Annotation 01433 AMPD2_HURAN 342 CSEVSLPYPDLQEFVADUXUNALININGPIKSFCYRRLQVLSSKFQUHOLLNEHKELAAQ 401 Matation 1 P23109 AMPD1_HURAN 320 CSEVSLPYPDLQEFVADUXUNALININGPIKSFCYRLQVLSSKFQUHOLLNEHKELAAQ 401 Matation 1 P23109 AMPD1_HURAN 320 KNNPHROFTIKKKYDTFLEAK-INCLESKFANKELLEFY IKARKHELEFIVHCQCRCOTLEFVER 300 Calternative 201433 AMPD2_HURAN 301 KNNPHROFTIKKKYDTFLEAK-INCLEFYKRHKELEFY IKARKHELEFYKARKHETYKARKHETYKARKHETYKARKHETYKARKHEKEFYKARKHETYKARKHETYKARKARKARKARKARKARKARKARKARKARKARKARKARK | RESULT INFO | Q01432 | AMPD3_HUMAN | : | | 0 |
| Importing Q01433 AMPD2_HUMAN 241 CSEVELPYPDLGEFVADVULALLINGFIKSFCYRLQYLSSKFOMHULLNERKELAAQ 401 Annotation Q01432 AMPD3_HUMAN 241 CSEVELPYPDLGEFVADVULALLINGFIKSFCYRLQYLSSKFOMHULLNERKELAAQ 401 Matal binding P23109 AMPD3_HUMAN 241 CSEVELPYPDLGEFVADVULALLINGFIKSFCYRLQYLSSKFOMHULLNERKELAAQ 401 Matal binding P23109 AMPD3_HUMAN 420 KNPHRPFTNCKKVDTEIHA OKHLLPFIKKSYQIDADRVVYSTKEKNILTKELFA 379 Galass AMPD3_HUMAN 402 KKVPHRPFTNKKVDTEIHA OKHLLPFIKKSYQIDADRVVYSTKEKNILTKEVFE 441 Chain Q01432 AMPD3_HUMAN 360 KKVPHRPFTNKKVDTEIHA OKHLLPFIKKTVTPCASELEDLYLKTDNYINGKYFAII 439 Region Q01432 AMPD3_HUMAN 462 SMILTAYDLSVDTLDVIADRADTFHRPKYNKYPVGASELEDLYLKTDNYINGKYFAII 521 Binding site Q01432 AMPD3_HUMAN 440 IKEVGADLVEAKYQHAEPRLSIYGSRDEUWRLAWAWHINUSPNKRUIQVPRIPUPY 499 Sequence conflict Q01432 AMPD3_HUMAN 522 IKEVKJADLESKYQJSEPRUSIYCKARAVAHKWLBSPNKRUIQVPRIPUPY 551 Similarity Q01432 AMPD3_HUMAN 522 IKEVASADLESKYQJSEPRUSIYLYWY INKWUMERLENKSPVKNRWLGPRSKKSSS <td>Highlight</td> <td>P23109</td> <td>AMPD1_HUMAN</td> <td>260</td> <td>${\tt DEPKPLPYPNLDTFLDDMNFLL&LI&QGPVKTYTHRRLKFLSSKFQVHQMLNEMDELKEL$</td> <td>319</td> | Highlight | P23109 | AMPD1_HUMAN | 260 | ${\tt DEPKPLPYPNLDTFLDDMNFLL&LI&QGPVKTYTHRRLKFLSSKFQVHQMLNEMDELKEL$ | 319 |
| Annotation P23109 AMPD1_HUMAN 320 KNNPHRDFYNCRKVDTHIHAC_IDCKHLERFIKKSYQIDADRVVYSTKEKNLTLKELFA 379 Atternative Q01433 AMPD2_HUMAN 300 KNVPHRDFYNLRKVDTHIHAC_IDCKHLERFIKKSYQIDADRVVYSTKEKNLTLKELFA 379 Sequence Q01432 AMPD1_HUMAN 300 KNVPHRDFYNLRKVDTHIHAC_IDCKHLERFIKKANKRHLEEIVHVQGREGTLREVFF 461 Matral variant P23109 AMPD1_HUMAN 380 KLKMPYDLTVDSLDVHAGRGTFORFDKFNDKYNPVGASELRDLYLKTDNVINCEYFATI 499 Chain P23109 AMPD1_HUMAN 360 KLKMPYDLTVDSLDVHAGRGTFORFDKFNDKYNPVGASELRDLYLKTDNVINCEYFATI 499 Region Q01432 AMPD1_HUMAN 361 GLHMDPYDLTVDSLDVHAGRGTFORFDKFNDKYNPVGASELRDLYLKTDNVINCEYFARM 420 Binding site Sinut TAYDLSEVTLVABROTTHERPEKTRAKKARVALENAVAMHKVIDSPNVRUGGVETPARM 420 1 ************************************ | Fighight | Q01433 | AMPD2_HUMAN | 342 | CSEVELPYPDLQEFVADVNVLMALIINGPIKSFCYRRLQYLSSKFQMHVLLNEMKELAAQ | 401 |
| Mateal Dindinx P23109 AMPD1 HUMAN 320 KNNPHRDFYNCRKVDTHIHAG OKHLLRFIKKSYQIDADRVVYSTKEKNLTLKELFA 379 Alternative Q01433 AMPD2 HUMAN 402 KKVPHRDFYNIRKVDTHIHAG OKHLLRFIKKSYQIDADRVVYSTKEKNLTLKELFA 379 sequence Q01432 AMPD1 HUMAN 301 KKVPHRDFYNIRKVDTHIHAG OKHLLRFIKKTYCPTPRTVAERKORKTILEQVP 461 Sequence Q01432 AMPD1 HUMAN 380 KLKHPYDLTVDSLDVHAGRGTFORFDKYNKYNPCASELRDLYLKTDWYINGSKYFAH 439 Region Q01432 AMPD3 HUMAN 361 Smint ATVLSVOTLDVHAGRGTFORFDKYNSKYMPCASELRDLYLKTEWINGSKYFAH 420 Binding site Q01432 AMPD3 HUMAN 361 IKEVGADLVEAKYQHAEPRLSIYGRSPDEWSKLSSWFVCNR HECPNHTWHIQVPRIYDF 499 Sequence conflict Q01432 AMPD3 HUMAN 440 IKEVGADLVEAKYQHAEPRLSIYGRSPDEWSKLSWFVCNR HECPNHTWHIQVPRIYDF 499 Active site Q01432 AMPD3 HUMAN 421 VEVARELESKYGMAELRSIYGRSPDEWSKLARWHERVBSNVFWUVQVPRLPVY 551 Amino acid P23109 AMPD1 HUMAN 560 RSKNFLPHFCKMLENIFLPFFATUPRASHELHLIFLYDGPSVDDESKHSDHFPSK 559 properties Q01432 AMPD3 HUMAN 560 SSKNFLPHYCKMLEHLHIFLPFFATUPRASHELHLIFLYDGPSVDDESKHSDHFPSK < | Annotation (j) | Q01432 | ANPD3_HOHAN | 241 | QEPSEPTPDETTIDDESTINDESTINGTION (ESSTSLEEDERSET KEL . ****:*:: *:. ::*** :** *:::****::*.***.:* | 300 |
| Alternative Q01433 AHPD2_HUHAN 402 KKVPHRDFYNIRKVDTELERASI DKHLLRFIKRAMKRHLEFIVHVEQGREGTLREVFF 461 sequence Q01432 AHPD3_HUHAN 301 KKVPHRDFYNIRKVDTELERASI DKHLLRFIKRAMKRHLEFIVHVEQGREGTLREVFF 461 Natural variant P23109 AMPD1_HUMAN 380 KLKMHPYDLTVDSLDVHAGRGTPGRFDKFNDKYNPVGASELRDLVLKTDNYINGEYFATI 439 Region Q01433 AMPD3_HUMAN 361 KLKMHPYDLTVDSLDVHAGRGTPGRFDKFNDKYNPVGASELRDLVLKTDNYINGEYFATI 439 Binding site ************************************ | Metal binding | P23109 | AMPD1 HUMAN | 320 | KNNPHRDFYNCRKVDTH I HAK ON OKHLLRFIKKSYOIDADRVVYSTKEKNLTLKELFA | 379 |
| sequence Q01432 AMPD3_HUMAN 301 KSNPHRPFYMVRVDTHIRAA | Alternative | Q01433 | AMPD2 HUMAN | 402 | KKVPHRDFYNIRKVDTHIHASS HNDKHLLRFIKRAMKRHLEEIVHVEQGREQTLREVFE | 461 |
| Natural variant | sequence | Q01432 | AMPD3_HUMAN | 301 | KSNPHRDFYNVRKVDTHIHA COOKHLLRFIKHTYQTEPDRTVAEKRGRKITLRQVFD | 360 |
| ChainP23109AMPD1 HUMAN380KLKMHPYDLTVDSLDVHAGRQTFQRFDKFNDKYNPVGASELRULYLKTDNYINGEYFATI439RegionQ01433AMPD2 HUMAN462SMNLTAYDLSVDLDVHADRNTFHRFDKFNAKYNPIGESULRE IFIKTDNRVSGKYFAHI420Binding siteCLHMPD7LTVDSLDVHADRNTFHRFDKFNAKYNPIGESULRE IFIKTDNRVSGKYFAHI420Modified residueP23109AMPD1 HUMAN440Gequence conflictQ01432AMPD2 HUMAN522Q01432AMPD3 HUMAN522IKEVMSDLEESKYQNAELLSIYGRSPDEWSLLSWFVCNRHCHCPNHTWHIQVPRIPTDF499Active siteQ01432AMPD3 HUMAN521VKEVARLEESKYQNSEPLSIYGRSPDEWSLLARWAVMHRVHSPNKRWIQVPRIPTDF480rift if ifitigrift if ifitigPropertiesQ01432AMPD2 HUMAN500Q01432AMPD2 HUMAN500RSKNFLPHFGKHLENIFMPVFEATINPQADPELSVFLKHITGFDSVDDESKHSGHMFSSK559RTKGQLANFCORMLENIFLPLFRATINPQDHELLHFLKYTOFDSVDDESKHSGHMFSSK540SimilarityQ01432AMPD2 HUMAN560PositiveQ01433AMPD2 HUMAN540SplPEdWULENPFYYLYYMYNNINVLNSLRKERGINTFLFRPHCGEAGALTHLMTAF619PositiveQ01433AMPD2 HUMAN541SplPEdWUENENPFYYLYYMYNNINVLNSLRKERGINTFLFRPHCGEAGALTHLMTAF619AromaticQ01433AMPD2 HUMAN541TinyP23109AMPD1 HUMAN540Smallrift if | Natural variant | | | | | |
| Region Q01433 AMPD2_HUMAN 462 SINULTAYDLSYUPLDYHADRNTFHEFDKFNAKYNPIGESULREIFIKTDNRVSGKYFAHI 521 Binding site | Chain | P23109 | AMPD1_HUMAN | 380 | KLKMHPYDLTVDSLDVHAGRQTFQRFDKFNDKYNPVGASELRDLYLKTDNYINGEYFATI | 439 |
| Binding site 001432 XHPD3_HUMAN 361 GLHMDPTDLTUSLUVHACKQIPHEP DKFNSKINPVGASELKULTLKTENTLGEDETFARK 420 Binding site Modified residue 23109 ANPD1_HUMAN 440 IKEVASUCALVEAKYQHAEPRLSIYGRSPEEWSKLSSWFVCNRIHCENNTUMICOVPRITVUF 499 Sequence conflict Q01433 AMPD2_HUMAN 421 IKEVASULEESKYQNAELRLSIYGRSPEEWSKLSSWFVCNRIHCENNTUMICOVPRITVUF 499 Active site Q01432 AMPD3_HUMAN 421 IKEVASULEESKYQNAELRLSIYGRSPEEWSKLARWAVMHRVHSPNNRWLVQVPRIFDVF 480 mino acid P23109 AMPD1_HUMAN 500 RSKNFLPHFGKMLENIFIPVFEATINPQAPELSVFLKHITGFDSVDESKHSGHHFSSK 559 properties Q01432 AMPD3_HUMAN 500 RSKNFLPHFGKMLENIFIPVFEATINPQAPELSVFLKHITGFDSVDESKHSGHHFSSK 559 Similarity Q01432 AMPD3_HUMAN 500 RSKNFLPHFGKMLENIFIPVFFATINPQDHRELHFLFLFVFSVDESKHSDHHFSDK 540 Similarity Q01432 AMPD3_HUMAN 560 SPKPQEUTLEKNPSYTYTYAYNYNANIMVLNSLRKERGMNTFIFFRHCGEAGALTHLMTAF 619 Negative Q01433 AMPD2_HUMAN 540 SPKPQEUTLEKNPSYTYTYAYNYNANIMVLNSLRKERGMNTFIFFRHCGEAGALTHLMTAF 619 Arbhatic Co1433 AMPD2_HUMAN 540 SPKPQEWTESQN | Region | Q01433 | AMPD2_HUMAN | 462 | SMNLTAYDLSVDTLDVHADRNTFHRFDKFNAKYNPIGESVLREIFIKTDNRVSGKYFAHI | 521 |
| Modified residue P23109 ANPD1 HUMAN 440 IKEVGADLVEAKYQHAEPRLSIYGRSPDEWSKLSWFVCNRHCPNHTWHIQVPRIYDVF 499 Sequence conflict Q01432 ANPD2 HUMAN 522 IKEVMSDLEESKYQNAELALSIYGRSPDEWSKLSWFVCNRHCPNTWRUQVPRIPDVF 499 Active site Q01432 ANPD3 HUMAN 421 IKEVMSDLEESKYQNSEPRLSIYGRSPDEWSKLSWFVCNRHCPNTWRUQVPRIPDVF 480 Amino acid P23109 AMPD1 HUMAN 500 RSKVFLPHFGKHLENIFUPVFEATINPQADPELSVFLKHITGFDSVDDESKHSGHNFSSK 559 properties Q01432 ANPD3 HUMAN 500 RSKVFLPHFGKHLENIFUPLFEATINPQADPELSVFLKHITGFDSVDDESKHSGHNFSSK 559 Similarity Q01432 ANPD3 HUMAN 481 RSKKLPNFGKHLENIFUPLFEATINPQADPELSVFLKHITGFDSVDDESKHSGHNFSSK 540 Negative Q01433 AMPD2 HUMAN 560 SPKPC@UTLEKNPSYTYYAYMYMAINMULNSLKEERGNTFLFRPHCGEAGALTHLMTAF 619 Positive Q01432 ANPD3 HUMAN 541 SPNPDVUTSEQUPYSYTYYAYMYMAINMULINERGEGSTFLFFRPHCGEAGALTHLWSAF 701 Aliphatic Timy P23109 ANPD1 HUMAN 620 MIADDISHGLNLKKSPVLQYLFFLAQIPIANSPLSNNSLFLEYAKNPFLDFULGKLHISL 679 Aromatic Q01432 ANPD3 HUMAN 620 MIADDISHGLNLKKSPVLQYLYLAQIPIANSP | Binding site | QU1432 | AMPD3_HUMAN | 361 | GLHMDPYDLTVDSLDVHAGRQIFHRFDKFNSKYNPVGASELRDLYLKTENYLGGEYFARM | 420 |
| Definition 223109 ANPD1 HUMAN 440 IKEVGADUCEALVYCHAEPRLSIYGRSPDEUBKLSSWFVCRRIECPNNTHULQVPRIPDVF 499 Sequence conflict Q01432 AMPD2 HUMAN 522 IKEVGADUCEALVYCHAEPRLSIYGRSPDEUBKLARWAVMHKVHSPNVRULQVPRIPDVF 480 Active site Q01432 AMPD3 HUMAN 421 VKEVGARLEESKYQNAELELSIYGRSPDEUBKLARWAVMHKVHSPNVRULQVPRIPDVF 480 Amino acid P23109 ANPD1 HUMAN 500 PSKNFLPHFGKHLENIFNPVFEATINPQAPELSVFLKHITGFDSVDDESKHSGHPTSSK 559 properties Q01432 AMPD3 HUMAN 500 PSKNFLPHFGKHLENIFNPVFEATINPQAPELSVFLKHITGFDSVDDESKHSGHPTSSK 559 Similarity Q01432 AMPD3 HUMAN 540 PSKNFLPHFGKHLENIFNPVFEATINPQDHELLFLFLWVTGFDSVDDESKHSGHPTSSK 540 Negative Q01432 AMPD2 HUMAN 560 SPKPQEUTLEKNPSYTYYAYYMYNANIMVLNSLRKERGMNTFLFRPEGCEAGALTHLMTAF 619 Positive Q01432 AMPD3 HUMAN 540 SPKPQEUTLEKNPSYTYYAYYMYNANIMVLNSLRKERGLSTFLFRPEGCEAGALTHLMTAF 619 Aromatic Q01432 AMPD3 HUMAN 540 SPKPQEUTLEKNSPVLYYYYYNYYNYNYNUKUNUNSLSFLSTSTHVSAF 660 Tiny P23109 AMPD1 HUMAN 620 MIADDISHGLLKKSPVLQYLYFLAQIDIAMSPLSTHSUSSTFLSTHKNEKLHKUSSCH | Modified residue | | | | | |
| Active site Q01433 AMPD2_RUMAN S22 IKEV MSUCLEXESTQUEELEXED TORSEDUELARMAVMENTENTENT NEW LOUVERLEDUT S81 Active site Q01432 AMPD3_HUMAN 421 VKEVARELEESKYQVSEPENSITYGRSPEEVENLAWAVMENTENTENT 111111111111111111111111111111111111 | | P23109 | AMPD1_HUMAN | 440 | IKEVGADLVEAKYQHAEPRLSIYGRSPDEWSKLSSWFVCNRIHCPNMTWMIQVPRIYDVF | 499 |
| Active site Q01402 AMPD1_HUMAN SOO RSKNFLPHFGKMLENIFMPVFEATINPQADPELSVFLKHITGFDSVDDESKHSGHMFSSK SSS properties Q01433 AMPD2_HUMAN SOO RSKNFLPHFGKMLENIFHPVFEATINPQADPELSVFLKHITGFDSVDDESKHSGHMFSSK SSS Similarity Q01432 AMPD3_HUMAN SOO RSKNFLPHFGKMLENIFLPLFEATUHPASHPELMLFLEHUGFDSVDDESKHSGHMFSSK SS9 Hydrophobic Positive Q01433 AMPD2_HUMAN S60 SPKPQEUTLEKNPSTTYTAYMYNANIMVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAF 619 Positive Q01433 AMPD2_HUMAN S60 SPKPQEUTLEKNPSTTYTAYMYNANIMVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAF 619 Positive Q01433 AMPD2_HUMAN S60 SPKPQEUTLEKNPSTYTYAYMYNANIMVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAF 619 Jaliphatic Timy P23109 AMPD1_HUMAN S60 STMPDVUTSEQNPYSYVLYYTRAMIMVLINNLRERCISTFLFRPHCGEAGSITHLVSAF 600 Aromatic Q01433 AMPD2_HUMAN 620 MIADDISHGLNLKKSPVLQYLFFLAQIPIAMSPLSNNSLFLSYMNPFLDFLKGKLWSL 660 Charged Q01432 AMPD1_HUMAN 620 MIADDISHGLNLKKSPVLQYLFFLAQIPIAMSPLSNNSLFLSYMPLPEYLKKGLWSL 660 Polar Q01433 AMPD2_HUMAN 610 LTANISHGLLEKAPVLQYLYLAQI | E sequence connict | 001433 | AMPD2 HUMAN | 421 | IKEVNSDLEESKIQNAELKLSIIGKSKDE0DKLAK0AVNHKVHSPNVK0LVQVPKLFDVI IKEVNSDLEESKIQNAELKLSIIGKSKDE0DKLAK0AVNHKVHSPNVK0LVQVPKLFDVI | 480 |
| Amino acid P23109 AMPD1_HUMAN 500 RSKNFLPHFGKMLENIFMPVFEATINPQADPELSVFLKHITGFDSVDDESKHSGHMFSSK 559 properties Q01433 AMPD2_HUMAN 500 RSKNFLPHFGKMLENIFHPVFEATINPQADPELSVFLKHITGFDSVDDESKHSGHMFSSK 559 Similarity Q01432 AMPD3_HUMAN 441 RSKRLPHFGKMLENIFHPVFEATINPQADPELSVFLKHITGFDSVDDESKHSGHMFSSK 540 Hydrophobic P23109 AMPD1_HUMAN 560 SPKPQEVTLEKNPSYTYYAYNYANIHVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAF 619 Positive Q01433 AMPD2_HUMAN 560 SPKPQEVTLEKNPSYTYYAYNYANIHVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAF 619 Aliphatic Tix * . ********************************* | - Active site | 201102 | All Do _ Hollan | 10 1 | 1*** 1* *!*** !* ******** !** !*! * ! .!!!.**! *!!!*!! | 100 |
| The second se | Amino acid | D22100 | ANDDA WIINAM | 500 | DOUNDI DUPOUNI ENITENDUPELITINDALDEI QUEI UUTTAPDQUDDEQUUQAUMPQQU | FEO |
| properties Q01432 AMPD3_HUMAN 481 RSKKLLPNFGKMLENIFLPLKATINPQDHRELHLFLKVVTGFDSVDDESKHSDHMFSDK 540 Similarity Similarity Pagative Paga | nronerties | Q01433 | AMPD2 HUMAN | 582 | RTKGQLÅNFQEMLEN IFLPLFEATVHPÅSHPELHLFLEHVDGFDSVDDESKHSOHHFSSK | 641 |
| Hydrophobic P23109 AMPD1_HUMAN 560 SPKPQEWTLEKNPSYTYYAYMYMANIMVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAF 619 Positive Q01432 AMPD2_HUMAN 642 SPLPEAUVEEDNPPYAYULYTYANMALINLRRQRGFHTFVLRPHCGEAGALTHLMTAF 619 Positive Q01432 AMPD3_HUMAN 642 SPLPEAUVEEDNPPYAYULYTYANMALINLRRQRGFHTFVLRPHCGEAGALTHLMTAF 619 Aliphatic Tiny P23109 AMPD1_HUMAN 620 MIADDISHGLNLKKSPVLQYLFFLAQIPIAMSPLSNNSLFLEYAKNPFLOFLOKUNISLS 679 Aromatic Q01433 AMPD2_HUMAN 620 MIADDISHGLNLKKSPVLQYLFFLAQIPIAMSPLSNNSLFLEYKNPFLDFLYRGLMYSL 660 Charged Q01433 AMPD1_HUMAN 660 STDDPHOFHFTKEPLMEEYAIAAQVFKLSTCDHCEVARNSVLOGSISHEEKVKFLGDNYL 739 Polar Q01433 AMPD2_HUMAN 660 STDDPHOFHFTKEPLMEEYAIAAQVFKLSTCDHCEVARNSVLOGSISHEEKVKFLGDNYL 739 Big Q01433 AMPD2_HUMAN 661 STDDPHOFHFTKEPLMEEYAIAAQVFKLSTCDHCEVARNSVLOGSISHEGKNYKFLGONYT 621 Big Q01433 AMPD2_HUMAN 661 STDDPHOFHFTKEPLMEEYAIAAQVFKLSTCDHCEVARNSVLOSGISHEEKVKFLGONYT 730 Big Q01433 AMPD2_HUMAN 661 STDDPHOFHFTKEPLMEEYAIAAQVFKLSTCDHCEVARNSVLOSGISHEEKVKFLGONYT | Cimilarity | Q01432 | AMPD3_HUMAN | 481 | RSKKLLPNFGKMLENIFLPLFKATINPQDHRELHLFLKYVTGFDSVDDESKHSDHMFSDK | 540 |
| Negative P23109 ANPD1_HUMAN 560 SPKPQEWTLEKNPSYTYYAYNYANIMVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAF 619 Positive Q01433 ANPD2_HUMAN 642 SPLPEAUVEEDNPYAYULYTTFANMAMLINLRRQRGFHTFVLRPHCGEAGALTHLMTAF 619 Aliphatic Ininy P23109 ANPD1_HUMAN 540 SPLPEAUVEEDNPYAYULYTTFANMAMLINLRRQRGFHTFVLRPHCGEAGALTHLWSAF 701 Aliphatic Ininy P23109 ANPD1_HUMAN 520 MIADDISHGLNLKKSPVLQYLFFLAQIPIANSPLSNNSLFLEYAKNPFLOFLGKGLMISL 679 Aromatic Q01433 ANPD2_HUMAN 702 MLARNISHGLLKKAPVLQYLYTLAUGIANSPLSNNSLFLSYHNPFLPEVLSRGLMYSL 761 Charged Q01433 ANPD2_HUMAN 601 LTANNISHGLLKKAPVLQYLYTLAUGIANSPLSNNSLFLSYNNENFLEVKKELGUNYL 660 Polar Q01433 ANPD2_HUMAN 680 STDDPHOFHFTKEPLMEEYSIAAQVFKLSCDMCEUARNSVLWSCFSHKVKFLGDNYL 739 Big Q01433 ANPD2_HUMAN 661 STDDPHOFHTYKEPLMEEYSIAAQVFKLSTCDMCEUARNSVLWSGFSHKVKSHULGPNYT 621 Big Q01433 ANPD2_HUMAN 661 STDDPHOFHTYKEPLMEEYSIAAQVFKLSTCDMCEUARNSVLWGFSHKVKSHULGPNYT 739 Serine Threonine Serine Threonine STDPPMOFHYTKEPLMEEYSIAAQVFKLSTCDMCEUARNSVLWGSGSHGEKQKFLGQNYT 72 | | | | | *:* * .* :*******:*:*:*: . ** :**::: ******** | |
| Degative Q01433 AMPD2_HUMAN 642 SPLPEAWVEEDNPYAYULYTYFAMMAHLMLRRQROFHTFVLRPECGEAGFIHHLVSAF 701 Positive Q01432 AMPD3_HUMAN 541 SPNPDVWTSEQNPPYSYLLYTMANIHULNNLRRERGLSTFLFRPHCGEAGSITHLVSAF 600 Aliphatic *** *: *. ***************************** | | P23109 | AMPD1 HUMAN | 560 | SPKPQEUTLEKNPSYTYYAYYMYANIMVLNSLRKERGMNTFLFRPHCGEAGALTHLMTAF | 619 |
| Positive Q01432 AMPD3_HUMAN S41 SPNPDVUTSEQNPPYSYLYTMYANIHVLANLRRERCLSTFLFRPHCCEAGSITHLVSAF 600 Aliphatic Tiny P23109 AMPD1_HUMAN 620 MIADDISHGLINLKKSPVLQYLFFLAQIPIAMSPLSNNSLFLEYAKNPFLDFLQKQLMISL 679 Aromatic Q01432 AMPD3_HUMAN 600 LIADNISHGLLEKKSPVLQYLFFLAQIPIAMSPLSNNSLFLEYAKNPFLDFLQKGLMISL 679 Charged Q01432 AMPD3_HUMAN 601 LIADNISHGLLEKKSPVLQYLYLAQIPIAMSPLSNNSLFLEYSKNPLREFLHKGLHVSL 660 Small STDDPMOFHFTKEPLMEEYSIAAQVFKLSTCDMCEVARNSVLQCGISHEEKVKFLGDNYL 739 739 Big Q01433 AMPD3_HUMAN 661 STDDPMOFHFTKEPLMEEYSIAAQVFKLSTCDMCEVARNSVLGGSHSHEKVKSHGUMYM 821 Big Q01433 AMPD3_HUMAN 661 STDDPMOFHFTKEPLMEEYSIAAQVFKLSTCDMCEVARNSVLGGSHSHEKVKSHGUMYM 821 Serine Threonine 651 STDDPMOFHTKEPLMEEYSIAAQVFKLSTCDMCEVARNSVLGGSHSHEKVKSHGUMYM 720 | I Negative | Q01433 | AMPD2_HUMAN | 642 | SPLPEAWVEEDNPPYAYYLYYTFANMAMLNHLRRQRGFHTFVLRPHCGEAGPIHHLVSAF | 701 |
| Aliphatic Tiny P23109 AMPD1 HUMAN 620 MIADDISHGLNLKKSPVLQYLFFLAQIPIAMSPLSNNSLFLEYAKNPFLDFLQKGLMISL 679 Aromatic Q01433 AMPD2 HUMAN 702 MLAENISHGLLEKKSPVLQYLYTLAQIFIAMSPLSNNSLFLEYAKNPFLDFLQKGLMISL 679 Charged Q01432 AMPD3 HUMAN 601 LTADNISHGLLEKKSPVLQYLYTLAQIFIAMSPLSNNSLFLEYSKNPLREFLHKGLHVSL 660 Small Polar P23109 AMPD1 HUMAN 680 STDDPMQFHFTKEPLMEEYSIAAQVFKLSTCDMCEVARNSVLQCGISHEEKVKFLGDNYL 739 Big Q01432 AMPD3 HUMAN 661 STDDPMQFHFTKEPLMEEYSIAAQVFKLSTCDMCELARNSVLMSFSISHKVSHMUGPNYT 821 Serine Threonine Stripterstaatestaat | | Q01432 | AMPD3_HUMAN | 541 | SPNPDVUTSEONPPYSYYLYYMYANINVLNNLRRERGLSTFLFRPHCGEAGSITHLVSAF | 600 |
| Tiny P23109 AMPD1 HUMAN 620 MIADDISHGLNLKKSPULQYLFFLAQIPIANSPLSNNSLFLEYAKNPFLDFLQKGLNISL 679 Aromatic Q01433 AMPD2 HUMAN 702 MLAENISHGLLLKKSPULQYLYLAQIGIANSLSLSNNSLFLEYAKNPFLDFLQKGLNISL 660 Charged Q01433 AMPD3 HUMAN 660 LTADNISHGLLLKKSPULQYLYLAQIGIANSLSFLEYKNPLPEPTLSKGLNYSL 660 Small Polar Q01433 AMPD2 HUMAN 660 STDDPMOFHFTKEPLMEEYSIAAQVFKLSTCDMCEVARNSVLQCGISHEEKVKFLGDNYL 739 Big Q01433 AMPD2 HUMAN 661 STDDPMOFHFTKEPLMEEYSIAAQVFKLSTCDMCEVARNSVLQSGISHEEKVKFLGDNYT 621 Serine Threonine Striptic Stri | Aliphatic | | | | | |
| Aromatic Q01433 AMPD2 HUMAN 702 HLAENISHGLLEKAPVLQYLYYLAQIGIAMSPLSYNSLFLSYHNPLPEYLSRGLMYSL 761 Charged Q01432 AMPD3_HUMAN 601 LTADNISHGLLEKAPVLQYLYYLAQIFIAMSPLSYNSLFLSYHNPLPEYLSRGLMYSL 761 Charged | 🗆 Tiny | P23109 | AMPD1_HUMAN | 620 | MIADDISHGLNLKKSPVLQYLFFLAQIPIAMSPLSNNSLFLEYAKNPFLDFLQKGLMISL | 679 |
| Charged Q01432 AMP03_BORAN GOI FileNUISHOLLKASPUQUEITILAQUPIAASLESINSELEISANPERELEISANP | 🗆 Aromatic | Q01433 | AMPD2_HUMAN | 702 | MLAENISHGLLLRKAPVLQYLYYLAQIGIAMSPLSNNSLFLSYHRNPLPEYLSRGLNVSL | 761 |
| Small P23109 AMPD1 HUMAN 680 STDPMOFHFTKEPLMEEYAIAAQVFKLSTCDMCEVARNSVLQCGISHEEKVKKFLGDNYL 739 Polar Q01433 AMPD2_HUMAN 680 STDPMOFHFTKEPLMEEYAIAAQVFKLSTCDMCEVARNSVLQCGISHEEKVKKFLGDNYL 739 Big Q01432 AMPD3_HUMAN 660 STDPMOFHFTKEPLMEEYAIAAQVFKLSTCDMCEVARNSVLMSGFSHKVKSHULGPNYL 739 Serine Threonine Serine Threonine Stripper State Stat | 🗆 Charged | 001432 | ANPDS_HOHAN | 001 | : *::***** *:*:******::**** ********** | 000 |
| Polar P23109 AMPD1_HUMAN 680 STDDPMOFHFTKEPLMEEYAIAAQVFKLSTCDMCEVARNSVLQCGISHEEKVKFLGDNVL 739 Q01433 AMPD2_HUMAN 762 STDDPLQFHFTKEPLMEEYSIATQVWKLSSCDMCELARNSVLQSGISHQEKKVKSHULGPNYL 821 Big Q01432 AMPD3_HUMAN 661 STDDPLQFHFTKEPLMEEYSIATQVWKLSSCDMCELARNSVLQSGISHQEKKVKSHULGPNYT 821 Serine Threonine StdopLopHofHYTKEALMEEYAIAAQVFKLSTCDLCEIARNSVLQSGISHQEKKVKSHUGPNYT 720 | 🗆 Small | | | | | |
| Big CO1432 AMPDG_HUMAN 661 STEDPLCF HI KREALETS LA LOVELSS DALCLARMSVLAS OF SHKVKSHULGRNYT 821 Serine Threenine Serine Threenine State | 🗆 Polar | P23109 | AMPD1_HUMAN | 680 | STDDPMQFHFTKEPLMEEYAIAAQVFKLSTCDMCEVARNSVLQCGISHEEKVKFLGDNYL | 739 |
| Serine Threonine | Big | 001433 | AMPD3 HUMAN | 661 | STDDPMOFHYTKEALMEEYAIAAQVUKLSTCDLCEIARNSVLOSGLSHOEKOKFLGONYY | 720 |
| | Serine Threonine | | | | ****** | |

Fig. 6 Protein alignment result. Partial view of the protein alignment result made on UniProtKB between P23109, Q01433 and Q01432 protein entries

annotated in UniProtKB are considered during the BLAST (their UniProt accessions are followed by "-n" where "n" is a digit for Swiss-Prot alternative splicing products) (*see* Fig. 8 i). On the lower part of the output BLAST result, a detailed list of the matched proteins is displayed, with a graphical view of the best alignment for each hit represented in a graphical view with the color code described previously, and linked to all corresponding local alignments between the query and the hit sequences (*see* Fig. 8 ii). All options available for text search result are applicable to this list (*see* Fig. 4).

4. Database Entries can be downloaded in batch. Several sets of protein sequences are proposed for download at http://www.uniprot.org/downloads. Entries present in the basket can be retrieved in different formats (*see* Fig. 3 iv). A dedicated tool to convert and download a list of proteins is available at http://

Add sequence or UniProt identifier



Fig. 7 The UniProtBLAST tool. View of the UniProt BLAST tool

| -ifter by | < Ed | it and resubmit | | | | | | |
|--|-------------|---|---|---|--|--------|----|--|
| Reviewed (5) Swiss-Prot | | Order b | y Score | | | | | _ |
| Unreviewed (245) TREMBL | | Color s | ale: Identity (%) 💌 1 | 00 80 | 60 | 40 | 20 | 0 |
| With 3D structure (1) | | 08280 | 4 ELF3_ARATH | Protein EARLY FLOWE | RING 3 | | | 100.0% |
| Complete Proteomes | | ROFUN | ROFUN8_9BRAS | Uncharacterized prote | ein 💻 | | | 84.0% |
| 107) | | V4NU7 | 1 V4NU71_THESL | Uncharacterized prote | ein 📖 | - | | 73.0% |
| raanisms | | D7LC2 | D7LC25_ARALL | Putative uncharacteria | zed protein | | | 92.0% |
| Moat (10) | | 00000 | 4-2 EUF3_ARATH | Isoform 2 of Protein E | AKLY | | | 100.0% |
| OLCA (13) | | M4CU8 | M4CU80 BRARP | Uncharacterized prote | ein 📕 | | | 68.0% |
| SOYEN (13) | | M4EZP | M4EZP1_BRARP | Uncharacterized prote | ein 🔳 | | | 74.0% |
| SOLTLI (11) | | F5B4H | F584H0_BRACM | EARLY FLOWERING 3 | | | | 84.0% |
| OPBA (10) | | F6H38 | F6H384_VITVI | Putative uncharacteria | zed protein 💻 | | | 55.0% |
| ther ereanieme | | B9RRV | 1 B9RRW1_RICCO | Putative uncharacteria | zed protein 💻 | | | 54.0% |
| AUTEL OF GOLFISTINS | | | | | | | | |
| | | More | | | | | | |
| Go | | More | | | | | | |
| Go Map To | €.Co | Iumns 🔧 BLA | at 👘 Align 🛃 Do | wnload 💼 Add to bask | | | 1 | to 250 of 250 Show |
| Go Map To UniProtKB | ₽ Co | Iumns 🔧 BLA | ST = Align 🛃 Do | wnload) 曾 Add to bask | at | | 1 | to 250 of 250 Show |
| Go Map To JniProtkB JniRef | | Iumns 🔧 BLA | II Alian de Do | wnload 🖆 Add to bask EARLY FLOWERING 3 - V | et view alignment | | 1 | to 250 of 250 Show |
| Go Map To JniProtKB JniRef JniParc | | Iumns % BLA Entry 4 082604 E | ST = Align 2 Do lignment overview LF3_ARATH - Protein | early flowering 3 - V | et View alignment | | 1 | to 250 of 250 Show 1070 E-value: 0.0 Score: 3,669 |
| Go Map To JniProtKB JniRef JniParc Fiew by | ≥ co | Iumns & BLA Entry 4 082804 E | Alion Do Ignment overview LF3_ARATH - Protein | wnload 前 Add to bask | et View alignment | | 1 | to 250 of 250 Show Info E-value: 0.0 Score: 3,669 Ident.: 100.0% |
| Go Map To JniProtKB JniRef JniParc View by | | Iumns & BLA | Ignment overview IGNMENT - Protein | Add to basic | et View alignment | | 1 | to 250 of 250 Show Info E-value: 0.0 Score: 3,669 Ident.: 100.0% |
| Go App To JniProtKB JniParc Fiew by asconomy | | Entry 4 082804 E ROFUN8 F | III Alian Do IIIgnment overview LF3_ARATH - Protein | EARLY FLOWERING 3 - V | et View alignment ew alignment | _ | 1 | to 250 of 250 Show Info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 |
| Go Map To JniProtKB JniParc fiew by faxonomy fext version | | More Entry A O82804 E ROFUN8 R | IT Alian Do Ilignment overview F3_ARATH - Protein OFUN8_9BRAS - Und | eARLY FLOWERING 3 - Vienaracterized protein - Vie | et View alignment ew alignment | | 1 | to 250 of 250 Show info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 Score: 1,766 |
| Go Map To IniProtKB IniParc Fiew by axonomy ext version Demo | | More Entry A O82804 E ROFUN8 R | III Alian Do IIIanment overview LF3_ARATH - Protein OFUN8_9BRAS - Und | eARLY FLOWERING 3 - V | et View alignment ew alignment | | 1 | to 250 of 250 Show Info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 Score: 1,786 Ident.: 84.0% |
| Go App To JniPatkB JniParc Aniew by awonomy fext version Demo | | More | III = Align do Do IIgnment overview LF3_ARATH - Protein OFUN8_9BRAS - Und MH (71, THESL - Lind | eAdd to bask | et View alignment | | 1 | to 250 of 250 Show Info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 Score: 1,786 Ident.: 84.0% E-value: 0.0 |
| Go Aap To JniProtKB JniParc Fiew by faxonomy fext version Demo | | More | T Alian Do Ilgarment overview LF3_ARATH - Protein DFUIN8_9BRAS - Und | wnload Add to besk EARLY FLOWERING 3 - V naracterized protein - Vie | et View alignment ew alignment ew alignment | | | to 250 of 250 Show Info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 Score: 1,786 Ident.: 84.0% E-value: 0.0 Score: 1,081 |
| Go Map To JniProtKB JniParc Aew by Faxonomy Fext version Demo | | More | II Alian Do IIanment overview F3_ARATH - Protein OFUIN8_9BRAS - Und ANU71_THESL - Und | eARLY FLOWERING 3 - V haracterized protein - Vie | et view alignment aw alignment ew alignment | | | to 250 of 250 Show info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 Score: 1,786 Ident.: 84.0% E-value: 0.0 Score: 1,081 Ident.: 73.0% |
| Go Map To JniProtKB JniParc Arew by faxonomy fext version Demo | | More Iumns ★ BLA Entry A 082804 E ROFUN8 R V4NU71 V (i) | III Alian Do IIIanment overview LF3_ARATH - Protein OFUN8_9BRAS - Und | eARLY FLOWERING 3 - V haracterized protein - Vie | et View alignment ew alignment ew alignment | | | to 250 of 250 Show Info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 Score: 1,786 Ident.: 84.0% E-value: 0.0 Score: 1,081 Ident.: 73.0% |
| Go Map To JniProtKB JniParc Fiew by fextonomy fext version Demo | | Nore | T Alian Do Ilgarment overview LF3_ARATH - Protein OFUIN8_9BRAS - Und ANU71_THESL - Und ANU71_THESL - Und | wnload Add to besk EARLY FLOWERING 3 - Vie haracterized protein - Vie | et View alignment ew alignment ew alignment tein - View alignm | sent. | | to 250 of 250 Show Info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 Score: 1,786 Ident.: 84.0% E-value: 0.0 Score: 1,081 Ident.: 73.0% E-value: 0.0 |
| Go Map To JniProtKB JniParc Arew by Faxonomy Fext version Demo | | Nore | T Alian Do Iighten overview F3_ARATH - Protein OFUN8_98RAS - Und ANU71_THESL - Und 7LC25_ARALL - Puta | eAdd to besk EARLY FLOWERING 3 - V haracterized protein - Vie haracterized protein - Vie | et View alignment ew alignment ew alignment tein - View alignm | ient : | | to 250 of 250 Show info E-value: 0.0 Score: 3,669 Ident.: 100.0% E-value: 0.0 Score: 1,786 Ident.: 73.0% E-value: 0.0 Score: 1,081 Ident.: 73.0% |

Fig. 8 BLAST result. Partial view of the result of the BLAST made on UniProtKB with 082804 entry as query

Upload Lists



Fig. 9 The UniProt downloading tool. View of the UniProt downloading tool

www.uniprot.org/uploadlists/ (*see* Fig. 9). The user provides a list of accessions in any of the supported formats (*see* http:// www.uniprot.org/help/uploadlists and Fig. 9 i) and can convert this list into any of the listed databases (*see* Fig. 9 ii). When the "from" database is "UniProtKB (AC/ID)" and the "to" database is "UniProt", the user can retrieve UniProtKB protein entries from a UniProtKB accession list.

5. UniProtKB entries are also present or cross-linked in several other biological databases and tools such as ExPASy (http://www.expasy.org/), the NCBI (http://www.ncbi.nlm.nih.gov/protein/) and TAIR (http://www.arabidopsis.org).

When accessing the UniProt website, some elements are always present at the top of the page: the UniProt logo to return to the home page, the search box (*see* Fig. 1-i), access to additional tools including BLAST, alignment and download, described elsewhere (*see* Fig. 1-ii), links to help (*see* **Note 13**), contact, and to the basket containing selected entries (*see* Fig. 1-iii).

3.3 The Web View of a UniProtKB Entry

3.3.1 UniProt Banner

34 Emmanuel Boutet et al.

| 3.3.2 Entry Header | The first block of each entry details (<i>see</i> Fig. 1-iv) accession numbers, status (<i>reviewed</i> for UniProtKB/Swiss-Prot and <i>unre- viewed</i> for UniProtKB/TrEMBL), as well as protein and gene names and synonyms. The primary accession number (AC, e.g., O80452) of an entry (<i>see</i> Note 14, documentation available at http://www.uniprot.org/manual/accession_numbers) is stable and provides a unique identifier which allows unambiguous cita- tion of the entry (<i>see</i> Notes 15 and 16). The entry name (ID, e.g., AMPD_ARATH) consists of up to 11 characters and takes the general form X_Y. Both X and Y represent mnemonic codes of up to 5 alphanumeric characters for both the protein name (X) and the species (Y) (documentation is available at http://www. uniprot.org/manual/entry_name). Entry names, corresponding to protein/gene name abbreviations, are subject to revision and therefore do not provide a stable means of identifying individual entries. Because entry names are prone to change, researchers who wish to cite entries in publications should always cite the pri- mary accession number. |
|---|---|
| 3.3.3 Analysis Tabs | Direct access to BLAST, alignment, and download, tools described in Subheading 3.2, is available from the protein entry view (<i>see</i> Fig. 1-v). Entry can also be stored in the basket. |
| 3.3.4 Contribution to Entry Annotation Tabs | Suggestions to update the content of the current entry can be sent via "comment" or "feedback" features (<i>see</i> Fig. 1-vi). |
| 3.3.5 Entry's Section Navigation Panel | The content of a protein entry is organized in 15 topics. To navi- gate and switch between topics, a display menu containing direct links to the different blocks of the entry is always visible on the left side of the screen (<i>see</i> Fig. 1-vii). Check boxes in this menu permit to hide/display the corresponding section. |
| 3.3.6 Entry Content View | In the main central area, the content of the current protein entry is displayed by thematic topics (<i>see</i> Fig. 1-viii). When a term is followed by "i" as exponent, this means that contextual information are available for this term. Most of the information in this section is extracted from the literature. Some information is also based on unproven empirical biological evidence, determined by computer prediction, or propagated from homologous members of the family (for details about annotation procedures, <i>see</i> http://www.uniprot.org/faq/45). In these cases, non-experimental qualifiers are added (<i>see</i> http://www.uniprot.org/manual/non_experimental_qualifiers); the qualifiers are: "Potential" for computer predicted, logical or conclusive evidence (<i>see</i> Note 17, represented on the website as "Reviewed prediction" in Swiss-Prot and as "Predicted" in TrEMBL), "Probable" for non-direct experimental evidence (<i>see</i> Note 18, |

represented on the website as "Inferred" in Swiss-Prot), and "By similarity" for experimental evidence in a close member of the family. Explanations of non-experimental qualifiers can be obtained by clicking on them in the entry.

Annotations are mainly distributed in four different types:

- 1. *General annotation*: provides general information about the protein, mostly biological knowledge, in different subsections (*see* http://www.uniprot.org/manual/general_annotation).
- 2. Sequence feature: information associated with specific residues of the current protein sequence (see http://www.uniprot.org/manual/sequence_annotation). Each sequence feature contains a "Feature key" (see Note 19), "Position(s)" indicates limits of the feature according to the amino acid residue positions of the displayed sequence (see Note 20), the "Length" of the feature is also given, a "Description" of the feature (see Note 21), a "Graphical view" to visualize the region in the consensus sequence, and, when available, "Feature identifier". UniProtKB/Swiss-Prot entries contain extensive annotation of all features that are predicted (and compatible with the protein function), experimentally proven, or determined by resolution of the protein structure.
- 3. *Cross-references*: used to point to information related to entries and found in data collections other than UniProtKB (*see* http://www.uniprot.org/help/cross_references_section).
- 4. Ontologiesand controlled vocabularies: a combination of controlled vocabularies and ontologies is used to summarize the functional implication of the current protein. The controlled vocabulary is developed by UniProtKB/Swiss-Prot (see http:// www.uniprot.org/keywords/), and GO terms (GO, [10]), a formal representation of terms that can be used to describe biological function, process and component, are developed and curated by the GO consortium (see http://www.uniprot. org/help/gene_ontology). Some keywords are derived from automatic annotation in UniProtKB/TrEMBL entries, but the vast majority is added manually in UniProtKB/Swiss-Prot entries. They describe the main characteristics of the protein.

The information contained in the entry is organized in a total of 15 topics, each accessible form the display panel. Depending of the information available in each entry, some sections might appear or not.

The 15 sections used in UniProtKB and their respective subsections are listed below:

1. "Function": (*see* Fig. 10 and http://www.uniprot.org/help/ function_section).

Contains information pertinent to biological knowledge of the protein function.

Function⁴

AMP deaminase plays a critical role in energy metabolism. Essential for the transition from zygote to embryo. #1 Publication -Catalytic activity AMP + H2O = IMP + NH3. #1 Publication -Cofactor Binds 1 zinc ion per subunit. #1 Publication -Enzyme regulation Activated by ATP. Activated by sulfate ions (in vitro). Inhibited by phosphate ions. #1 Publication --Manual assertion based on experiment described in: Kinetics "Membrane association, mechanism of action, and structure of Arabidopsis embryonic factor 1 (FAC1)." K_M=6.7 mM for AMP (in the absence of ATP) *#***1** Publication *▼* Han B.W., Bingman C.A., Mahnke D.K., Bannen R.M., Bednarek S.Y., Sabina R.L., Phillips G.N. Jr. K_M=0.26 mM for AMP (in the presence of 1 mM ATP) J. Biol. Chem. 281:14939-14947(2006) [PubMed] [Europe PMC] [Abstract]

Cited for: X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS) OF 140-839 IN COMPLEX WITH COFORMYCIN 5'-PHOSPHATE;

-

V_{max}=17 µmol/min/mg enzyme (in the absence of ATP) V_{max}=375 µmol/min/mg enzyme (in the presence of 1 mM ATP)

Pathway

Purine metabolism; IMP biosynthesis via salvage pathway; IMP from AMP: step 1/1. Sites

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|----------------------------|-------------|--------|---------------------------------|----------------|-----------------------|
| Metal binding ⁱ | 391 | 1 | Zinc; catalytic | | |
| Metal binding ⁱ | 393 | 1 | Zinc; catalytic | | |
| Binding site ⁱ | 393 | 1 | Substrate | | |
| Metal binding ⁱ | 659 | 1 | Zinc; catalytic | | |
| Binding site ⁱ | 662 | 1 | Substrate | | |
| Active site ⁱ | 681 | 1 | Proton acceptor <i>Inferred</i> | | |
| Metal binding ⁱ | 736 | 1 | Zinc; catalytic | | |

Regions

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|---------------------------------|-------------|--------|---------------------------|----------------|-----------------------|
| Nucleotide binding ⁱ | 289 - 296 | 8 | ATP # Reviewed Prediction | | |

GO - Molecular function¹

AMP deaminase activity

Inferred from genetic interaction¹ Ref.1 Source: TAIR

```
ATP binding
Inferred from electronic annotation<sup>i</sup> Source: UniProtKB-KW
```

metal ion binding Inferred from electronic annotation¹ Source: UniProtKB-KW Complete GO annotation...

Keywords - Molecular functionⁱ Hydrolase Keywords - Biological processⁱ Nucleotide metabolism

Keywords - Ligandⁱ

ATP-binding, Metal-binding, Nucleotide-binding, Zinc

Enzyme and pathway databases

| SABIO-RK | 080452. |
|------------|---------------------|
| UniPathway | UPA00591; UER00663. |

Fig. 10 Function section of a UniProtKB entry. View of the "function" section of the UniProtKB protein 080452

The different subsections of the function section are:

- (a) General annotation dealing with function, catalytic activity, cofactor, enzyme regulation, biophysicochemical properties, and pathway
- (b) Sequence features describing active site, metal binding, binding site, site, calcium binding, zinc finger, and DNA binding with a graphical view
- (c) GO terms of the 'Molecular function' section
- (d) Keywords of 'Molecular function', 'Biological process', and 'Ligand' subsections
- (e) Cross-references that point to family, enzyme, and pathway databases
- 2. "Names & Taxonomy": (*see* Fig. 11 and http://www.uniprot. org/help/names_and_taxonomy_section).

This block describes protein names, gene names and taxonomy of the organism. The recommended protein name is given in the first row, followed by the alternative names used in the literature. In the case of an enzyme, the Enzyme Commission (EC) number is given as synonym. This EC number is an active link to the Enzyme database (http://www.expasy.org/ enzyme/) [11], which contains detailed information about enzyme activity and lists all UniProtKB/Swiss-Prot entries having the same EC number. The second row of this block describes the gene encoding the protein in the following order: gene name, synonyms, ordered locus name when applicable (see Note 22) and ORF names used by the genomic sequencing projects, when available. Following the gene description, the organism name, the NCBI taxonomy identifier, and the summarized taxonomic hierarchy are actively linked to the UniProt taxonomy browser (http://www.uniprot.org/ taxonomy/) which contains details on the organism and gives access to all UniProtKB entries of that organism (see Note 23).

 "Subcellular location": (*see* Fig. 12 and http://www.uniprot. org/help/subcellular_location_section).
 Contains information pertinent to biological knowledge of the protein localization and topology.

The different subsections of the subcellular location section are:

- (a) General annotation dealing with subcellular location
- (b) Sequence features describing transmembrane and topological domain with a graphical view
- (c) GO terms of the 'Cellular component' section
- (d) Keywords of the 'Cellular component' section
- 4. "Pathology & Biotech": (*see* Fig. 13 and http://www.uniprot. org/help/pathology_and_biotech_section).

Names & Taxonomy

| Protein names ⁱ | Recommended name: AMP deaminase • Short name: AtAM • EC: 3.5.4.6 Alternative name(s): • Protein EMBRYONIC | PD FACTOR 1 |
|--------------------------------------|--|---|
| Gene names ⁱ | Name: Synonyms: Ordered Locus Names: ORF Names: | AMPD FAC1 At2g38280 F16M14.21 |
| Organism ⁱ | Arabidopsis thaliana (M | ouse-ear cress) [Reference proteome] |
| Taxonomic identifier ⁱ | 3702 [NCBI] | |
| Taxonomic lineage ⁱ | Eukaryota > Viridiplanta eudicotyledons > core e | e > Streptophyta > Embryophyta > Tracheophyta > Spermatophyta > Magnoliophyta > udicotyledons > rosids > malvids > Brassicales > Brassicaceae > Camelineae > Arabidopsis |

Organism-specific databases

TAIR AT2G38280.

Fig. 11 Names & taxonomy section of a UniProtKB entry. View of the "names and taxonomy" section of the UniProtKB protein 080452

Subcellular location:

Membrane; Single-pass membrane protein. Microsome membrane <u>Note:</u> Might be associated with the inner mitochondrial membrane **#By similarity** . **#1 Publication Topology**

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|--|--|-------------|--------------------------------|----------------|-----------------------|
| Transmembrane ⁱ | 8 - 28 | 21 | Helical; # Reviewed Prediction | | |
| GO - Cellular comp cytosol Inferred from direct endoplasmic reticulu Inferred from electro | onent ⁱ assay ⁱ Ref.7 m onic annotation | PubMed 2110 | niProtKB-KW | | |
| integral to mitochon | drial outer men | nbrane | | | |
| Inferred from direct | assay ⁱ PubM | ed 21896887 | Source: TAIR | | |
| nucleus | | | | | |
| Inferred from direct Complete GO anno | assay ⁱ Ref.7 Si tation | ource: TAIR | | | |
| Keywords - Cellula Endoplasmic reticulu | r component ⁱ m, Membrane, | Microsome | | | |

Fig. 12 Subcellular location section of a UniProtKB entry. View of the "subcellular location" section of the UniProtKB protein 080452

Pathology & Biotech

Allergenic properties

Causes an allergic reaction in human. Common symptoms of mite allergy are bronchial asthma, allergic rhinitis and conjunctivitis. Binds to IgE in 80% of patients with house dust allergy.

Mutagenesis

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|-----------------------------|-------------------|--------|---------------------------------------|----------------|-----------------------|
| Mutagenesis ⁱ | 132 | 1 | $C \rightarrow A$: Loss of activity. | | |
| Mutagenesis ⁱ | 150 | 1 | N → E: Loss of N-glycosylation. | | |
| Keywords - Dise Allergen | ease ⁱ | | | | |
| Protein family/g | roup databas | es | | | |

| Allergome | 1232. Der p 1.0111. |
|-----------|---------------------|
| | 310. Der p 1. |

Fig. 13 Pathology & biotech section of a UniProtKB entry. View of the "pathology and biotech" section of the UniProtKB protein P08176

Contains information pertinent to biological knowledge of disease(s) and phenotype(s) associated with the deficiency of the protein.

The different subsections of the Pathology & Biotech section are:

- (a) General annotation dealing with involvement in disease, natural variant, allergenic properties, biotechnological use, toxic dose, and pharmaceutical use
- (b) Sequence features describing disruption phenotype and mutagenesis with a graphical view
- (c) Keywords of the 'Disease' section
- (d) Cross-references that point to organism-specific databases
- 5. "Post translational modification (PTMs) / Processing": (*see* Fig. 14 nd http://www.uniprot.org/help/ptm_processing_ section).

Contains information pertinent to biological knowledge of the protein posttranslational modifications.

The different subsections of the PTM / processing section are:

- (a) Sequence features describing initiator methionine, signal, pro-peptide, transit peptide, chain, peptide, modified residue, lipidation, glycosylation, disulfide bond, and crosslink with a graphical view
- (b) General annotation dealing with posttranslational modification

PTM / Processing^{*}

Molecule processing

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|------------------------------|-------------|--------|---|----------------|-----------------------|
| Transit peptide ⁱ | 1 - 55 | 55 | Chloroplast 🖋 Inferred | | |
| Chain ⁱ | 56 - 333 | 278 | Adenylate isopentenyltransferase 3, chloroplastic | | PRO_0000391072 |
| Propeptide ⁱ | 334 - 336 | 3 | Removed in mature form | | PRO_0000396781 |

Amino acid modifications

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|-------------------------------|-------------|--------|-----------------------|----------------|-----------------------|
| Modified residue ⁱ | 333 | 1 | Cysteine methyl ester | | |
| Lipidation ⁱ | 333 | 1 | S-farnesyl cysteine | | |

Post-translational modification

Farnesylated.

Keywords - PTMⁱ Lipoprotein, Methylation, Prenylation

| Proteomic databas | oteomic databases | | | | |
|-------------------|-------------------|--|--|--|--|
| PRIDE | Q93WC9. | | | | |

Fig. 14 PTM/processing section of a UniProtKB entry. View of the "PTM/processing" section of the UniProtKB protein Q93WC9

- (c) Keywords of the 'PTM' section
- (d) Cross-references that point to proteomics and PTM databases
- 6. "Expression": (*see* Fig. 15 and http://www.uniprot.org/help/ expression_section).

Contains information pertinent to biological knowledge of the protein expression.

The different subsections of the expression section are:

- (a) General annotation dealing with tissue specificity, developmental stage and induction
- (b) Keywords of the 'Developmental stage' section
- (c) Cross-references that point to gene expression databases
- 7. "Interaction": (*see* Fig. 16 and http://www.uniprot.org/help/ interaction_section).

| Expression: |
|--|
| Tissue specificity Expressed in seedlings, roots, leaves, flowers, pollen grains, pollen tubes and siliques, and at a lower level in stems. |
| Developmental stage Expressed in both male and female gametophytes, at the zygote stage, in the endosperm, and during early embryo development. |
| Observed in cotyledonary embryos and in the basal part of the embryo, but not in the suspensor or in mature embryos. Also expressed during somatic embryogenesis. #1 Publication - |
| Gene expression databases |

| Genevestigator |
|----------------|
|----------------|

Fig. 15 Expression section of a UniProtKB entry. View of the "expression" section of the UniProtKB protein 080452

| ubunit stru omodimer. I inary inter | ucture Interacts with AHK4 actions | 🛷 2 Publ | ications 👻 | | |
|---|--|----------|------------|-------------------------|-------|
| With | Entry | #E | хр. | IntAct | Notes |
| AHK4 | Q9C5U0 | 2 | | EBI-1807679,EBI-1100775 | |
| rotein-pro | tein interaction da | tabases | | | |
| | | IntAct | 080452. | 2 interactions. | |

Fig. 16 Interaction section of a UniProtKB entry. View of the "interaction" section of the UniProtKB protein 080452

Contains information pertinent to biological knowledge of the protein interactions.

The different subsections of the interaction section are:

- (a) General annotation dealing with subunit structure
- (b) Specific annotation describing binary interactions
- (c) Cross-references that point to protein–protein interaction databases
- 8. "Structure": (*see* Fig. 17 and http://www.uniprot.org/help/ structure_section).

Contains information pertinent to biological knowledge of the protein structure.

The different subsections of the structure section are:

- (a) Sequence features describing turn, beta strand and helix with a graphical view (when available)
- (b) Cross-references that point to 3D structure databases
- 9. "Family & Domains": (see Fig. 18 and http://www.uniprot. org/help/family_and_domains_section).

| Structure ⁴ | | | | | | |
|---|---------------|-----------------|------------------------|------------|----------------------|---------------|
| Secondary structure 1 Show details Legend Helix Turn Beta strand 3D structure databases | | | | | | 839 |
| Select the link destinations: PDBe RCSB PDB PDBj | Entry 2A3L | Method X-ray | Resolution (Å) 3.34 | Chain A | Positions 140-839 | PDBsum [»] |
| ProteinModelPortal | 080452. | | | | | |
| SMR | 080452. P | ositions 212- | 839. | | | |
| Miscellaneous databases | | | | | | |
| EvolutionaryTrace | 080452. | | | | | |

Fig. 17 Structure section of a UniProtKB entry. View of the "structure" section of the UniProtKB protein 080452

Contains information pertinent to biological knowledge of the protein family and domains

The different subsections of the Family & Domains section are:

- (a) Sequence features describing domain, repeat, compositional bias, region, coiled coil, motif, and domain with a graphical view with a graphical view
- (b) General annotation dealing with sequence similarities; a comment describing to which family the protein may belong may be included. It is linked to a UniProt query that lists all UniProtKB entries belonging to the same family (*see* **Note 24** and Fig. 18 i). In the case of transporter families, the transport classification (TC) number is present when available, and a cross-link to the transport classification database (http://www.tcdb.org) is also included.
- (c) Keywords of the 'Domain' section
- (d) Cross-references that point to phylogenomic and family and domain databases
- 10. "Sequence": (*see* Fig. 19 and http://www.uniprot.org/help/ sequences_section).

Contains general metadata determined for the given sequence, such as sequence length, molecular weight, and CRC64 checksum (64 bit Cyclic Redundancy Check value) [12] (*see* **Note 25**). Each subsection contains information pertinent to biological knowledge of the protein sequence. On the right side of all sequences, a quick access to the FASTA format (http://en.wikipedia.org/wiki/FASTA_format) of the sequence and to sequence/proteomic tools is present (*see* Fig. 19 i).

Family & Domains:

| Region | | | | | |
|---|---|----------------------------|---|----------------|-----------------------|
| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
| Region ⁱ | 462 - 467 | 6 | Substrate binding | | |
| Region ⁱ | 737 - 740 | 4 | Substrate binding | | |
| Compositional bia | is | | | | |
| Feature key | Position(s |) Lengt | n Description | Graphical view | Feature identifier |
| Compositional bia | as ⁱ 86 - 92 | 7 | Poly-Gly | | |
| Compositional bia | as ⁱ 158 - 161 | 4 | Poly-Asp | | |
| Sequence similari Belongs to the ade Keywords - Doma Transmembrane, Tr Phylogenomic dat | ities mosine and AMF in ¹ ransmembrane H tabases |) deaminas nelix | es family. | | |
| | eg | gNOG CO |)G1816. | | |
| | HOG | ENOM HO | G000092200. | | |
| | InPar | anoid O | 0452. | | |
| | | ко ко | 1490. | | |
| | | OMA HE | VYSDN. | | |
| | Phylo | meDB OS | 80452. | | |
| | ProtCl | ustDB PL | N02768. | | |
| Family and domain | n databases | | | | |
| | Int | erPro IP IP IP [G | R006650. A/AMP_deam_AS. R001365. A/AMP_deaminase_ R006329. AMP_deaminase. raphical view] | dom. | |
| | PAN | THER PT | HR11359. PTHR11359. 1 hit. | | |
| | | Pfam PF | 00962. A_deaminase. 1 hit. raphical view] | | |
| | TIGF | FAMs TI | GR01429. AMP_deaminase. 1 | hit. | |
| | PRO | DSITE PS | 00485. A_DEAMINASE. 1 hit. raphical view] | | |

Fig. 18 Family & domains section of a UniProtKB entry. View of the "family and domains" section of the UniProtKB protein 080452

The different subsections of the sequence section are:

- (a) The sequence status, either complete or fragment(s)
- (b) Sequence processing when accurate; details about this processing are described in the "PTM/Processing" section
- (c) The canonical protein sequence
- (d) Alternative products with sequence and additional related information, when existing. The alternative products subsection describes the proteins which may be produced by

Sequence⁴

```
Sequence status<sup>i</sup>: Complete.
                                                                               Length:
                                                                                            839
080452-1 [UniParc] ± FASTA
                                                                               Mass (Da): 95,130
                                                                               Last modified:
  MEPNIYQLAL AALFGASFVA VSGFFMHFKA LNLVLERGKE RKENPDGDEP 50
                                                                                June 1, 2002. Version 2.
                                                                                             : 188F1F4A589A17DA<sup>1</sup>
                                                                               Checksum
  QNPTLVRRRS QVRRKVNDQY GRSPASLPDA TPFTDGGGGG GGDTGRSNGH 100
  VYVDEIPPGL PRLHTPSEGR ASVHGASSIR KTGSFVRPIS PKSPVASASA 150
                                                                               Blast
                                                                                             GO
                                                                               Blast
  EYSIAASVWK LSACDLCEIA RNSVYOSGFS HALKSHWIGK DYYKRGPDGN 800
                                                                                ProtParam
                                                                                                I
  DIHKTNVPHI RVEFRDTIWK EEMOOVYLGK AVISDEVVP
                                                             839
                                                                                Compute pl/MW
                                                                                ProtScale
                                                                                PeptideMass
 « Hide
                                                                               PeptideCutter
Sequence caution
The sequence BAD94943.1 differs from that shown. Reason: Intron retention.
Sequence databases
  Select the link destinations:
                                    AC003028 Genomic DNA. Translation: AAC27176.2.
    · EMBL
                                     CP002685 Genomic DNA. Translation: AEC09516.1.
    C GenBank
                                    CP002685 Genomic DNA, Translation: AEC09517.1.
    C DDBJ
                                    AY056301 mRNA. Translation: AAL07150.1.
                                     AY133852 mRNA. Translation: AAM91786.1.
                                     AK316943 mRNA. Translation: BAH19646.1.
                                     AK221552 mRNA, Translation: BAD94943.1, Sequence problems.
                                IPI IPI00546126.
                                PIR T01259.
                             RefSeq NP_565886.1. NM_129384.2.
                                     NP_850294.1. NM_179963.2.
                            UniGene At.12466.
Genome annotation databases
                      EnsemblPlants AT2G38280.1; AT2G38280.1; AT2G38280.
                                     AT2G38280.2; AT2G38280.2; AT2G38280.
                            GeneID 818408.
                              KEGG ath: AT2G38280.
```

Fig. 19 Sequence section of a UniProtKB entry. View of the "sequence" section of the UniProtKB protein 080452

alternative splicing or promoter usage. Modifications of the canonical sequence necessary to produce the alternative product sequence are described in the sequence features subsection (*see* Fig. 20).

- (e) General annotation dealing with sequence caution, caution, polymorphism, RNA editing and mass spectrometry
- (f) Sequence features describing natural variant, alternative sequence, sequence uncertainty, sequence conflict, nonadjacent residues, non-terminal residue, and non-standard residue with a graphical view
- (g) Keywords of the 'Coding sequence diversity' section
- (h) Cross-references that point to sequence, genome annotation databases and polymorphism databases

Sequences (2)*

| Sequence status ⁱ : Complete. | |
|---|--|
| This entry describes 2 isoforms i produced by alternative splicing. \square Alig | In |
| Isoform 1 (identifier: 082804-1) [UniParc] This isoform has been chosen as the 'canonical' sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry. | Length: 695 Mass (Da): 77,206 Last modified:November 1, 1998 - v1 Checksum: 607A0720ED381C08 ⁴ BLAST GO |
| Isoform 2 (identifier: O82804-2) [UniParc] | Length:339Mass (Da):37,760Checksum:4CBEAD87D3292DA6 ⁱ BLASTGO |
| MKRGKDEEKI LEPMFPRLHV NDADKGGPRA PPRNKMALYE QLSIPSQRFG 50 DHGTMNSRSN NTSTLVHPGP SSQPCGVERN LSVQHLDSSA ANQATEKFVS 100 QMSFMENVRS SAQHDQRKMV REEEDFAVFV YINSRRSQSH GRTKSGIEKE 150 KHTPMVAPSS HHSIRFQEVN QTGSKQNVCL ATCSKPEVRD QVKANARSGG 200 FVISLDVSVT EEIDLEKSAS SHDRVNDYNA SLRQESRNRL YRDGGKTRLK 250 DTDNGAESHL ATENHSQEGH GSPEDIDNDR EYSKSRACAS LQQINEEASD 300 DVSDDSMVDS ISSIDVSPDD VVGILGQKRF WRARKAIAK 339 | |

Note: No experimental confirmation available.

Sequence conflict

The sequence CAA72719.1 differs from that shown. Reason: Frameshift at positions 437, 472 and 485.

Cautionⁱ

Isoform-2 : No experimental confirmation available.

Alternative sequence

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier |
|----------------------|-------------|--------|---------------------------------|----------------|-----------------------|
| Alternative sequence | 339 | 1 | $N \rightarrow K$ in isoform 2. | | VSP_004042 |
| Alternative sequence | 340 - 695 | 356 | Missing in isoform 2. | | VSP_004043 |

Fig. 20 Sequence section of a UniProtKB entry containing alternative products. View of the "sequence" section of the UniProtKB protein 082804; only details concerning the alternative splicing are shown

11. "Cross-references": (*see* Fig. 21 and http://www.uniprot.org/ help/cross_references_section).

The cross-references section is divided into subsections organized by themes. This section links the protein to several other databases that contain information relevant to that protein. Many of these cross-links are automatically added to UniProtKB/TrEMBL entries, but some are manually created in UniProtKB/Swiss-Prot entries (*see* Note 19). Each row of this block corresponds to a single database, the name of which

Cross-references

| Sequence databases | |
|--|--|
| Select the link destinations: © EMBL © GenBank © DOBJ | AC0030289 Genomic DNA. Translation: AAC07176.2 . CP002085 Genomic DNA. Translation: AEC07516.1 . CP002085 Genomic DNA. Translation: AEC07517.1 . AY050201 mRNA. Translation: AAU07150.1 . AY139502 mRNA. Translation: AAU07156.1 . AX130549 mRNA. Translation: BA119646.1 . AX131649 mRNA. Translation: BA119646.1 . |
| | IP100546126. |
| U PIR | T01259. |
| RefSeq | NP_565886.1. NM_129384.2. NP_850294.1. NM_179963.2. |
| UniGene | At.12466. |

3D structure databases

| Select the link destinations: PDBe C RCSB PDB C PDBj | Entry 2A3L | Method X-ray | Resolution (Å) 3.34 | Chain A | Positions 140-839 | PDBsum [>] |
|---|---------------|-----------------|------------------------|------------|----------------------|----------------|
| ProteinModelPortal | 080452. | | | | | |
| SMR | 080452. P | ositions 212- | 339. | | | |
| ModBase | Search | | | | | |

Proteomic databases

| veoine databases | |
|------------------|---------|
| PaxOb | 080452. |
| PRIDE | 080452. |
| | |

Protocols and materials databases

StructuralBiologyKnowledgebase Search...

Genome annotation databases

| EnsemblPlants | AT2G38280.1; AT2G38280.1; AT2G38280. AT2G38280.2; AT2G38280.2; AT2G38280. |
|---------------|--|
| GeneID | 818408. |
| KEGG | ath: AT2G38280. |

Organism-specific databases

TAIR AT2G38280.

Phylogenomic databases

| eggNOG | COG1816. |
|-------------|--------------|
| HOGENOM | HOG00092200. |
| InParanoid | 080452. |
| KO | K01490. |
| OMA | HRVYSDN. |
| PhylomeOB | 080452. |
| ProtChastOR | PI N02769 |

Enzyme and pathway databases

| UniPathway | UPA00591; UER00563. |
|---------------------------|---------------------|
| SABIO-RK | 060452. |
| Miscellaneous databases | |
| EvolutionaryTrace | 080452. |
| Gene expression databases | |
| Genevestigator | 080452. |

| Ontologies | |
|-----------------------------|--|
| PRO | O80452. |
| Family and domain databases | |
| InterPro | IPR006650. A/AMP_deam_AS. IPR001365. A/AMP_deaminase_dom. IPR006329. AMP_deaminase. [Craphical view] |
| PANTHER | PTHR11359. PTHR11359. 1 hit. |
| Pfam | PF00962. A_deaminase. 1 hit. [Graphical view] |
| TIGRFAMs | TIGR01429. AMP_deaminase. 1 hit. |
| PROSITE | PSC0485. A_DEAMINASE. 1 hit. [Graphical view] |
| ProtoNet | Search |

Fig. 21 Cross-references section of a UniProtKB entry. View of the "cross-references" section of the UniProtKB protein 080452

Table 1 Plant-specific cross-references present in UniProtKB

| Database name and URL and goals | DR line format |
|---|---|
| GeneFarm [13] http://genoplante-info.infobiogen.fr/Genefarm/ Structural and functional annotation of <i>Arabidopsis thaliana</i> gene and protein families (<i>see</i> http://www.uniprot.org/database/DB-0032). | DR GeneFarm; GeneID; FamilyID. In UniProtKB/Swiss-Protonly |
| Gramene; a comparative mapping resource for grains [14] http://www.gramene.org/ Curated, open-source, Web-accessible data resource for comparative genome analysis in the grasses (<i>see</i> http://www. uniprot.org/database/DB-0039). | DR Gramene; UniProtKB_AC; In UniProtKB/Swiss-Protand UniProtKB/TrEMBL |
| MaizeGenetics/GenomicsDatabase(MaizeGDB) [15] http://www.maizegdb.org/ Central repository for public maize information (see http://www.uniprot.org/database/DB-0058). | DR MaizeDB; ProteinID; In UniProtKB/Swiss-Protonly |
| The Arabidopsis Information Resource (TAIR) [16] http://www.arabidopsis.org/index.jsp Searchable relational database on Arabidopsis thaliana, which includes many different molecular data types and provides a comprehensive resource for the scientific community (<i>see</i> http://www.uniprot.org/database/ DB-0102). | DR TAIR; Order_locus_name; In UniProtKB/Swiss-Protand UniProtKB/TrEMBL |

is indicated in the first column (*see* Fig. 21 i). A link to the relevant data in the cross-linked database is present in next columns. Plant specific databases that are currently cross-linked in UniProtKB entries are listed in Table 1. They have been chosen because of their content, their stability and their frequent updates. All of them give additional information about the protein and are linked back to UniProtKB.

The different subsections of the cross-references section are:

- (a) 2D gel databases
- (b) **3D structure databases**; Cross-references to the PDB database (http://www.rcsb.org/pdb/) are present when protein structures are available. PDB cross-links contain information about the crystallographic method, the number of chains, and the range of residues present in the structure.
- (c) Enzyme and pathway databases
- (d) Family and domain databases
- (e) Gene expression databases

- (f) Genomeannotationdatabases
- (g) Ontologies
- (h) Organism-specific databases
- (i) Phylogenomic databases
- (j) Polymorphism databases
- (k) Proteomic databases
- (l) Protein-protein interaction databases
- (m) Protein family/group databases
- (n) PTM databases
- (o) Sequence databases; Cross-references to the EMBL database (http://www.embl-heidelberg.de/) are displayed in the same order as the corresponding references associated with a sequence submission. EMBL crosslinks contain a nucleic acid sequence ID, a protein sequence ID and a molecule type to indicate the origin of the sequence (e.g., mRNA or Genomic_DNA) (see Note 26). When no coding sequence to translate the nucleic acid sequence into the protein sequence was provided by the submitters to the EMBL, the flag "No translation available" is present to replace the lacking protein sequence ID. When the sequence displayed in UniProt differs from the original EMBL sequence, a flag "Sequence problems" is added and the differences between the two sequences are summarized in the "Sequence" section.
- (p) Other
- 12. "Publications": (see Fig. 22 and http://www.uniprot.org/ help/publications_section). This block lists all references used for the annotation of the protein entry. The first references are usually associated with sequence submission, followed by references providing other information concerning the function and structure of the protein. Each reference is numbered and contains title, authors, and conventional citation information for the reference, including cross-links to PubMed and digital object identifier (DOI), thus allowing retrieval of the electronic version of the article. In addition, an indication of what information was extracted from the article, strain and tissues used is also mentioned when available. In the case of references associated with a sequence submission, the sequenced molecule type is mentioned and, if relevant, the corresponding isoform is indicated. Each author name is linked to a UniProtKB query that retrieves all entries where that author is cited.
- 13. "Entry information": (see Fig. 23a and http://www.uniprot. org/help/entry_information_section). In addition to the

Publications:

« Hide 'large scale' publications ± Download 1. "EMBRYONIC FACTOR 1 encodes an AMP deaminase and is essential for the zygote to embryo transition in Arabidonsis." Xu J., Zhang H.-Y., Xie C.-H., Xue H.-W., Dijkhuis P., Liu C.-M. Plant J. 42:743-756(2005) [PubMed] [Europe PMC] [Abstract] Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA], MUTAGENESIS OF ASP-598, FUNCTION, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE. Strain: cv. Landsberg erecta. 2. "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana." Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L., Moffat K.S., Cronin L.A. 🛲 Venter J.C. Nature 402:761-768(1999) [PubMed] [Europe PMC] [Abstract] Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA]. Strain: cv. Columbia. 3. The Arabidopsis Information Resource (TAIR) Submitted (APR-2011) to the EMBL/GenBank/DDBJ databases Cited for: GENOME REANNOTATION. Strain: cv. Columbia. ... 8. "Toward an interaction map of the two-component signaling pathway of Arabidopsis thaliana." Dortay H., Gruhn N., Pfeifer A., Schwerdtner M., Schmuelling T., Heyl A. J. Proteome Res. 7:3649-3660(2008) [PubMed] [Europe PMC] [Abstract] Cited for: INTERACTION WITH AHK4. 10. "Crystallization and preliminary X-ray crystallographic analysis of adenosine 5'-monophosphate deaminase (AMPD) from Arabidopsis thaliana in complex with coformycin 5'-phosphate." Han B.W., Bingman C.A., Mahnke D.K., Sabina R.L., Phillips G.N. Jr. Acta Crystallogr. F 61:740-742(2005) [PubMed] [Europe PMC] [Abstract] Cited for: CRYSTALLIZATION, X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS) OF 140-839 IN COMPLEX WITH COFORMYCIN 5'-PHOSPHATE AND ZINC IONS.

11. "Membrane association, mechanism of action, and structure of Arabidopsis embryonic factor 1 (FAC1)." Han B.W., Bingman C.A., Mahnke D.K., Bannen R.M., Bednarek S.Y., Sabina R.L., Phillips G.N. Jr. J. Biol. Chem. 281:14939-14947(2006) [PubMed] [Europe PMC] [Abstract] Cited for: X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS) OF 140-839 IN COMPLEX WITH COFORMYCIN 5'-PHOSPHATE; PHOSPHATE AND ZINC IONS, CATALYTIC ACTIVITY, SUBUNIT, COFACTOR, SUBCELLULAR LOCATION, ENZYME REGULATION, BIOPHYSICOCHEMICAL PROPERTIES.

Fig. 22 Publications section of a UniProtKB entry. View of the "publications" section of the UniProtKB protein 080452

primary accession number, a protein entry may contain one or more secondary accession numbers, which follow the primary accession number. These are usually accession numbers of UniProtKB/TrEMBL entries that have been merged into a single UniProtKB/Swiss-Prot entry. The history of the current protein entry give the date when the entry was first created, the date of last modification of the sequence and the date of last modification of annotation, respectively. The corresponding releases are also indicated. A quick access to this history is also available beneath the entry remote control (*see* Fig. 1 ix).

| 1000 CONTRACTOR | ime ¹ AMP | PD_ARATH | | | | | |
|---|--|--|----------------------------|------------|--------------|------|--|
| Accessio | on ⁱ Prim Sec | nary (citable) accession number: 080452 condary accession number(s): B9DFX9, Q56XX1, Q93ZR9 | | | | | |
| Entry his | story ⁱ Inte Uni | egrated into May ProtKB/Swiss-Prot: | 30, 2006 | | | | |
| Last Last This | t sequence update: June | 1, 2002 | | | | | |
| | Las | t modified: Octo | nodified: October 16, 2013 | | | | |
| | Thi | This is version 94 of the entry and version 2 of the sequence. [Complete history] | | | | | |
| Entry st | atus ⁱ Rev | iewed (UniProtKB/Swiss-Prot) | | | | | |
| Annotat | ion program Plan | it Protein Annotation Program | | | | | |
| Aisce | llaneous (b | 6 | | | | | |
| rabidop: rabidop: ATHWA\ | sis thaliana sis thaliana: entrie: / comments | s and gene names | | | | | |
| rabidops rabidops ATHWAY ndex of DB cross ndex of MILARI ndex of | sis thaliana sis thaliana: entrie: / comments metabolic and bios s-references Protein Data Bank TY comments protein domains ar | s and gene names ynthesis pathways (PDB) cross-references Id families | | | | | |
| rabidops rabidops ATHWAN ndex of DB cross ndex of IMILARI INILARI Simila | sis thaliana sis thaliana: entrie: / comments metabolic and bios s-references Protein Data Bank TY comments protein domains ar r proteinS ¹ | s and gene names ynthesis pathways (PDB) cross-references Id families | | | | | |
| rabidops rabidops ATHWAN ndex of DB cross DB cross IMILARI ndex of Simila | sis thaliana sis thaliana: entrie: / comments metabolic and bios s-references Protein Data Bank TY comments protein domains ar r proteins: 90% Identity | s and gene names ynthesis pathways (PDB) cross-references Id families | | | | | |
| rabidops rabidops ATHWAN ndex of DB cross ndex of IMILARI INILARI INILARI Comila | sis thaliana sis thaliana: entrie: comments metabolic and bios s-references Protein Data Bank TY comments protein domains ar r proteinS ¹ (90% Identity 51 | s and gene names ynthesis pathways (PDB) cross-references id families C 1% ldentity :) Organisms | Length | Cluster ID | Cluster name | Size | |

Fig. 23 Entry information, miscellaneous and similar proteins sections of a UniProtKB entry. View of the "information, miscellaneous and similar proteins" sections of the UniProtKB protein 080452

- 14. "Miscellaneous": (*see* Fig. 23b and http://www.uniprot.org/ help/miscellaneous_section). Links to relevant documents (*see* Note 2) and keywords of the 'Technical term' section are listed.
- 15. "Similar proteins": (*see* Fig. 23c and http://www.uniprot.org/ help/similar_proteins_section). This section provides links to UniRef100, UniRef90, and UniRef50, corresponding to protein sequences sharing 100 %, 90 %, or 50 % identity, respectively. UniRef are sequence clusters, used to speed up sequence similarity searches (*see* Note 4).

4 Notes

- 1. The SIB (Switzerland, Geneva), in collaboration with the EBI (UK, Hinxton) and PIR (USA, Georgetown University Medical Center and National Biomedical Research Foundation), develop the UniProt protein resource that contain a Protein knowledgebase (UniProtKB), Sequence clusters (UniRef), and a sequence archive (UniParc).
- 2. For more information, *see* http://www.uniprot.org/docs and http://www.expasy.org/sprot/userman.html. UniProt propose also demonstration videos on its YouTube channel: https://www.youtube.com/channel/UCkCR5RJZCZZoVTQzTYY92aw.
- 3. For more information, *see* http://www.uniprot.org/manual/ non_experimental_qualifiers.
- 4. The UniRef reference clusters combine closely related sequences into a single record on order to speed sequence similarity searches. The UniRef100 database combines identical sequences and subfragments of the UniProt Knowledgebase (from any species) and selected UniParc records into a single UniRef entry (http://www.uniprot. org/help/uniref). UniRef90 and UniRef50 yield a database size reduction of approximately 40 % and 65 %, respectively, providing for significantly faster sequence searches.
- UniProtKB proteomes are listed at http://www.uniprot.org/ taxonomy/complete-proteomes. Each protein of a reference organism has the keyword "Reference proteome" (see http:// www.uniprot.org/keywords/KW-1185).
- 6. UniProt is currently hosted by a unified UniProt website http://www.uniprot.org/.
- 7. Major releases usually introduce important format changes. They are distinguishable from other releases by a new primary number followed by ".0".
- 8. To download a local version of UniProtKB, use the web page ftp://ftp.uniprot.org/pub.
- 9. When a gene encodes different isoforms and/or when different protein sequences for the same gene of a given species (given cultivar/strain/isolate) are available, they are merged into a single UniProtKB entry (e.g., Jasmonic acid-amido synthetase JAR1, entry **Q9SKE2**).
- Other tools and databases developed by the EBI and PIR are available at http://www.ebi.ac.uk/services/ [17] and http:// pir.georgetown.edu/, respectively.
- 11. For users of the Mozilla Web browser (http://www.mozilla. org/), the biobar navigation bar, dedicated to search into various

biological databases, is available at https://addons.mozilla. org/en-US/firefox/addon/biobar/. An ExPASy navigation bar is available at http://expasybar.mozdev.org, it allows searches to be performed in several databases hosted by ExPASy.

- 12. A complete documentation about BLAST parameters is available on the UniProt website at this address: http://www.uniprot.org/help/sequence-searches.
- 13. Your feedback is highly important and allows us to continuously improve our knowledgebase according to your needs.
- 14. UniProtKB accessions (AC) contain six characters and respect one of these regular expressions [A-N,R-Z][0-9][A-Z] [A-Z,0-9][A-Z,0-9][0-9] or [O,P,Q][0-9][A-Z,0-9] [A-Z,0-9][A-Z,0-9][0-9] (e.g., O80452). To face the fast increasing amount of new protein entries, an additional accession format extended to 10 alphanumerical characters for entries integrated after all 6 characters accessions will be used, possibly in 2014. The format of this new format will be [A-N,R-Z][0-9][A-Z][A-Z,0-9][A-Z,0-9][0-9][A-Z] [A-Z,0-9][A-Z,0-9][0-9]. Both 6 and 10 characters accessions will coexist. All accessions are stable in time and should be used for UniProtKB protein citation.
- 15. It can also (but rarely) happen that the primary accession number becomes a secondary accession number (e.g., when an entry is split in two entries).
- 16. An accession number uniquely identifies an entry. If an entry is deleted, its AC will never be attributed to another entry.
- 17. A typical example is the annotation of N-glycosylation sites in the entries of non-cytoplasmic domains or proteins.
- 18. A typical example is the annotation of nuclear subcellular location in the entries of active transcription factors in eukaryotic organisms.
- 19. Exhaustive information about all cross-references present into UniProtKB (more than 140 in 2014) is available at http://www.uniprot.org/database/ and http://www.uniprot.org/docs/dbxref.
- 20. Amino-acid residue numbering begins at the N-terminus of the precursor protein (the displayed sequence).
- 21. The description of the feature may contain a non-experimental qualifier (*see* http://www.uniprot.org/manual/non_experimental_qualifiers).
- 22. In the case of *Arabidopsis thaliana* and *Oryza sativa* (and in other organisms following the same standards), we use the following nomenclature according to the standard defined for *A. thaliana*: [first letter of the genius name]-[first letter of the

species name]-[chromosome number]-[g, for gene]-[locus number] (e.g., At1g15690, Os03g16440).

- Currently, Oryza sativa has three different taxonomy identifiers in UniProtKB/TrEMBL: 39947 for japonica cultivars, 39946 for indica cultivars, and 4530 for unspecified rice cultivars. In UniProtKB/Swiss-Prot, when possible, cultivars are specified for each reference related to a sequence deposition.
- 24. The family classification is exclusively based on sequence similarities, not on functions.
- The algorithm to compute the CRC64 is described in the ISO 3309 standard [12].
- 26. Additional qualifiers may be present: ALT_SEQ, ALT_INIT, ALT_TERM, or ALT_FRAME. These are used in the case of discrepancies between the EMBL derived CDS and the displayed protein sequence. These may be due to gross differences in the predicted CDS sequence (arising from the failure to correctly predict all exons for a given gene for instance), incorrect selection of the initiating methionine, and termination of the sequence or frameshifts, respectively. For more details, see the documentation (http://www.uniprot.org/help/sequence_caution).

Acknowledgments

UniProt is mainly supported by the National Institutes of Health (NIH) grant 1 U41 HG006104. Additional support for the EBI's involvement in UniProt comes from the NIH grant 2P41 HG02273. Swiss-Prot activities at the SIB are supported by the Swiss Federal Government through The State Secretariat for Education, Research and Innovation SERI. PIR's UniProt activities are also supported by the NIH grants 5R01GM080646-07, 3R01GM080646-07S1, 5G08LM010720-03, and 8P20GM103446-12, and the National Science Foundation (NSF) grant DBI-1062520. We would like to thank all Swiss-Prot curators and developers for their contribution to the expert annotation of proteins and their critical reading of the manuscript.

References

- The UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res 42(Database issue):D191–D198
- Bairoch A, Boeckmann B, Ferro S, Gasteiger E (2004) Swiss-Prot: juggling between evolution and stability. Brief Bioinform 5:39–55
- Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S,

Schneider M (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31: 365–370

4. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Muellenet P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJ, Scheremetjew M, Tate J, Thimmajanarthanan M, Thomas PD, Wu CH, Yeats C, Yong SY (2012) InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res 40(Database issue): D306–D312

- Schneider M, Lane L, Boutet E, Lieberherr D, Tognolli M, Bougueleret L, Bairoch A (2009) The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. J Proteomics 72(3):567–573
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
- 7. Gattiker A, Gasteiger E, Bairoch A (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. Appl Bioinforma 1:107–108
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy Server. In: Walker JM (ed) The proteomics protocols handbook. Humana, Totowa, NJ, pp 571–607
- Dimmer EC, Huntley RP, Alam-Faruque Y, Sawford T, O'Donovan C, Martin MJ et al (2012) The UniProt-GO Annotation database in 2011. Nucleic Acids Res 40:D565–D570
- 11. Bairoch A (2000) The ENZYME database in 2000. Nucleic Acids Res 28:304–305
- Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1993) Numerical recipes in C,

2nd edn. Cambridge University Press, Cambridge, pp 896–902

- Aubourg S, Brunaud V, Bruyere C, Cock M, Cooke R, Cottet A, Couloux A, Dehais P, Deleage G, Duclert A, Echeverria M, Eschbach A, Falconet D, Filippi G, Gaspin C, Geourjon C, Grienenberger J-M, Houlne G, Jamet E, Lechauve F, Leleu O, Leroy P, Mache R, Meyer C, Nedjari H, Negrutiu I, Orsini V, Peyretaillade E, Pommier C, Raes J, Risler J-L, Riviere S, Rombauts S, Rouze P, Schneider M, Schwob P, Small I, Soumayet-Kampetenga G, Stankovski D, Toffano C, Tognolli M, Caboche M, Lecharny A (2005) GeneFarm, structural and functional annotation of Arabidopsis gene and protein families by a network of experts. Nucleic Acids Res 33:D641–D646
- Ware DH, Jaiswal P, Ni J, Yap IV, Pan X, Clark KY, Teytelman L, Schmidt SC, Zhao W, Chang K, Cartinhour S, Stein LD, McCouch SR (2002) Gramene, a tool for grass genomics. Plant Physiol 130:1606–1613
- Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V (2004) MaizeGDB, the community database for maize genetics and genomics. Nucleic Acids Res 32(Database issue): D393–D397
- 16. Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res 31:224–228
- 17. Harte N, Silventoinen V, Quevillon E, Robinson S, Kallio K, Fustero X, Patel P, Jokinen P, Lopez R (2004) European Bioinformatics Institute. Public web-based services from the European Bioinformatics Institute. Nucleic Acids Res 32(Web Server issue):W3–W9