# A Quick Guide to UniProtKB Swiss-Prot & TrEMBL

**UniProt**, the **Universal Protein Resource**, is produced by the UniProt Consortium, formed by the Swiss Institute of Bioinformatics (SIB), the European Bioinformatics Institute (EBI) and the Protein Information Resource (PIR). UniProt provides a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.The centrepiece of the UniProt databases is the UniProt knowledge base (**UniProtKB**), which comprises 2 sections: manually annotated UniProtKB/Swiss-Prot and automatically annotated UniProtKB/TrEMBL. Taken together, these 2 sections give access to all publicly available protein sequences.

**UniProtKB/Swiss-Prot** is a high quality manually annotated (reviewed) and non-redundant protein sequence database, which brings together experimental results and computed features. Although Swiss-Prot provides annotated entries for all species, it focuses on the annotation of proteins from model organisms of distinct taxonomic groups to ensure the presence of high quality annotation for representative members of all protein families. Protein families and groups of proteins are continuously reviewed to keep up with current scientific findings.

**UniProtKB/TrEMBL** is a computer-annotated (unreviewed) supplement to Swiss-Prot**,** which strives to gather all protein sequences that are not yet represented in Swiss-Prot.
The protein sequences are derived from the translation of coding sequences (CDS) submitted to the public nucleic acid databases (EMBL/GenBank/DDBJ) or from other sequence resources, such as Ensembl. Automated annotation of the highest currently available quality is integrated to TrEMBL entries. The usual Swiss-Prot annotation pipeline involves the manual annotation of TrEMBL entries, their integration into Swiss-Prot, with their original accession number, and subsequent deletion from TrEMBL.

Each Swiss-Prot entry contains information about one or more protein sequence(s) derived from one gene in one species. Different sections of the entry report specific biological information (Usermanual: http://www.uniprot.org/manual/).

**Entry information** Each entry is associated with a stable unique and citable identifier: the primary **accession number**. The entry name is unique, but not stable. Additional information on the entry history is provided.

| | |
|---|---|
| Entry name | AHSA1_HUMAN |
| Accession | Primary (citable) accession number: **O95433**<br>Secondary accession number(s): Q96IL6, Q9P060 |
| Entry history | Integrated into UniProtKB/Swiss-Prot: February 21, 2001<br>Last sequence update: May 1, 1999<br>Last modified: June 16, 2009<br>This is version 85 of the entry and version 1 of the sequence. [Complete history] |
| Entry status | Reviewed (UniProtKB/Swiss-Prot) |

**Names and origin** Protein name, synonyms and abbreviations are indicated, as well as gene and locus names, and taxonomy information.

| | |
|---|---|
| Protein names | **Activator of 90 kDa heat shock protein ATPase homolog 1**<br>*Also known as:*<br>AHA1<br>p38 |
| Gene names | Name: **AHSA1**<br>Synonyms: C14orf3<br>ORF Names: HSPC322 |
| Organism | **Homo sapiens (Human)** |
| Taxonomic identifier | 9606 [NCBI] |
| Taxonomic lineage | Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo |

**Evidence on protein existence** is provided in the 'Protein attributes' section.

**References** This section lists publications used to annotate the entry. The type of data retrieved from a cited article is specified (see token 'Cited for: ').

[6] **"p38: a novel protein that associates with the vesicular stomatitis virus glycoprotein."**
Sevier C.S., Machamer C.E.
Biochem. Biophys. Res. Commun. 287:574-582(2001) [PubMed: 11554768] [Abstract] [Article from publisher]
Cited for: INTERACTION WITH VSV G, SUBCELLULAR LOCATION, TISSUE SPECIFICITY.

**General annotation** Biological information and its source is stored in this subsection. Qualifiers (e.g. 'By similarity') are used in the absence of direct experimental evidence.

| | |
|---|---|
| Function | Cochaperone that stimulates HSP90 ATPase activity (By similarity). May affect a step in the endoplasmic reticulum to Golgi trafficking. |
| Subunit structure | Interacts with HSPCA/HSP90 and with the cytoplasmic tail of the vesicular stomatitis virus glycoprotein (VSV G). Interacts with GCH1. (Ref.6) (Ref.7) (Ref.8) (Ref.9) |
| Subcellular location | Cytoplasm › cytosol. Endoplasmic reticulum. Note: May transiently interact with the endoplasmic reticulum. (Ref.6) |
| Tissue specificity | Expressed in numerous tissues, including brain, heart, skeletal muscle and kidney and, at lower levels, liver and placenta. (Ref.6) |
| Induction | By heat shock and treatment with the HSP90 inhibitor 17-demethoxygeldanamycin (17AAG). (Ref.7) |
| Sequence similarities | Belongs to the AHA1 family. |

**Cross-references** This section provides links to numerous specialized databases.

| | |
|---|---|
| **Sequence Databases** | |
| EMBL | AJ243310 mRNA. Translation: CAB45684.1. |
| UniGene | Hs.204041 |
| **3D Structure Databases** | |
| PDB | *Structure determined by NMR spectroscopy:* 1X53. Chain A maps to 204-335. |
| **2D Gel Databases** | |
| REPRODUCTION-2DPAGE | O95433. HUMAN. |

**Keywords** are terms from a controlled vocabulary list, which summarises the features of the entry.

| | |
|---|---|
| Cellular component | Endoplasmic reticulum |
| Molecular function | Chaperone |
| Technical term | 3D-structure |

Relevant **Gene Ontology (GO)** terms are assigned based on experimental data from the literature.

**Sequence annotation** Over 30 feature keys (e.g. 'Modified residue') describe the sequence at the single residue level. Qualifiers ('Potential', 'Probable' and 'By similarity') indicate the computer-prediction of the feature or the existence of indirect experimental evidence. The sources of data are indicated (e.g. Ref.30).

**Regions**

| | | | | |
|---|---|---|---|---|
| DNA binding | 421 – 486 | 66 | Nuclear receptor | |
| Zinc finger | 421 – 441 | 21 | NR C4-type | |
| Zinc finger | 457 – 481 | 25 | NR C4-type | |
| Region | 1 – 420 | 420 | Modulating | |
| Region | 487 – 527 | 41 | Hinge | |
| Region | 528 – 777 | 250 | Steroid-binding | |
| Compositional bias | 399 – 418 | 20 | Glu/Pro/Ser/Thr-rich (PEST region) | |

**Amino acid modifications**

| | | | |
|---|---|---|---|
| Modified residue | 8 | 1 | Phosphothreonine (Ref.30) |
| Modified residue | 45 | 1 | Phosphoserine (Ref.30) |
| Modified residue | 113 | 1 | Phosphoserine (By similarity) |
| Modified residue | 134 | 1 | Phosphoserine (Ref.31) |
| Modified residue | 141 | 1 | Phosphoserine (By similarity) |
| Modified residue | 203 | 1 | Phosphoserine (Ref.24) |
| Modified residue | 211 | 1 | Phosphoserine (Ref.24) |
| Modified residue | 226 | 1 | Phosphoserine (Ref.31) |
| Modified residue | 234 | 1 | Phosphoserine (Ref.31) |
| Modified residue | 267 | 1 | Phosphoserine (Ref.31) |
| Cross-link | 277 | | Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in SUMO) |
| Cross-link | 293 | | Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in SUMO) |
| Cross-link | 419 | | Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin) (Probable) |

**Natural variations**

| | | | |
|---|---|---|---|
| Alternative sequence | 1 – 26 | 26 | Missing in isoform Alpha-B and isoform Beta-B. |
| Alternative sequence | 451 | 1 | G → GR in isoform Alpha-2 and isoform Beta-2. |

**Sequence** The protein sequence displayed in the entry is the most prevalent and/or the most similar to orthologous sequences found in other species. When the genomic sequence is completed, we generally display the protein sequence derived from genome translation. Sequence discrepancies are documented. The molecular weight is automatically calculated, independently of any experimental result. Direct links to protein sequence analysis tools (i.e. PeptideCutter) and Blast are provided.

| Sequence | Length | Mass (Da) | |
|---|---|---|---|
| O95433-1 [UniParc].<br>Last modified May 1, 1999. Version 1. | 338 | 38,274 | Blast |

```
        10         20         30         40         50         60
MAKWGEGDPR WIVEERADAT NVNNWHWTER DASNWSTDKL KTLFLAVQVQ NEEGKCEVTE

        70         80         90        100        110        120
VSKLDGEASI NNRKGKLIFF YEWSVKLNWT GTSKSGVQYK GHVEIPNLSD ENSVDEVEIS
```

## UniProtKB/Swiss-Prot specificity and utility

Data integrated into Swiss-Prot, including the protein sequence and associated current knowledge, are **manually checked and continuously updated**.
- In order to have **minimal redundancy** and improve **sequence reliability**, all protein sequences encoded by the same gene are merged into a single record.
- A special emphasis is laid on the **annotation of biological events that generate protein diversity** and that cannot be predicted at the genomic level. Alternative products (alternative splicing, RNA editing…), post-translational modifications (PTMs) are extensively annotated. For additional information, see Boeckmann et al., C. R. Biol. 328:882-899 (2005).
- Swiss-Prot contains numerous cross-references, thus playing the role of a **central hub for biological data,** linking together relevant protein resources.

**Swiss-Prot is particularly suitable for similarity searches, protein identification (proteomics) and training of prediction software tools.**

## What you can find in UniProtKB/Swiss-Prot

Reliable **protein sequences** and description of **alternative protein products** (due to alternative splicing, alternative promoter usage, alternative initiation, RNA editing…); **protein and gene names** using standardised official nomenclature and synonyms used in the literature and other databases; **protein function**; enzyme-specific information (catalytic activity, cofactors, metabolic pathway, regulation); biologically relevant **domains** and sites; **PTMs**; **subcellular location(s)**; **tissue expression**; expression during embryonic development and /or cell differentiation; secondary and quaternary structure information; **polymorphisms**; similarities to other proteins; involvement in diseases; **cross-references** to numerous databases; controlled vocabularies in several (sub)sections; qualifiers for predicted or propagated data; documentation and FAQ files (http://www. uniprot.org/help/); etc.
See also our usermanual: http://www.uniprot.org/manual/

## What you can find through UniProtKB/Swiss-Prot

Specialized information beyond the scope of Swiss-Prot is made available via cross-references to biologically relevant resources, such as the EMBL/GenBank/DDBJ nucleotide sequence databases, 2D and 3D protein structure databases, various protein domain and family databases, PTM databases, species-specific data collections, variant and disease databases; a list of cross-referenced databases is available at http://www.uniprot.org/docs/dbxref: explicit links are provided in the Swiss-Prot flat file, additional implicit links are created 'on the fly' by the UniProt server.

## Interactive access to UniProtKB

You can access UniProtKB from http://www.uniprot.org/ through the 'Search' tool.

## How to obtain a local copy of UniProtKB

To download complete data sets in the original flat file format, fasta format, XML or RDF format, go to www.uniprot.org/downloads. UniProtKB is released every 4 weeks. Please check our current release notes.

## How to retrieve '**Complete proteomes**'?

Example (homo sapiens):
Query: organism:9606 AND keyword:complete proteome
See: http://www.uniprot.org/faq/15
A selection of complete proteomes (fasta format) is available at www.uniprot.org/downloads

## Submission of updates and new data

To submit **updates** and/or corrections to Swiss-Prot entries and for any enquiries you can either use the e-mail address help@uniprot.org or go at www.uniprot.org/contact

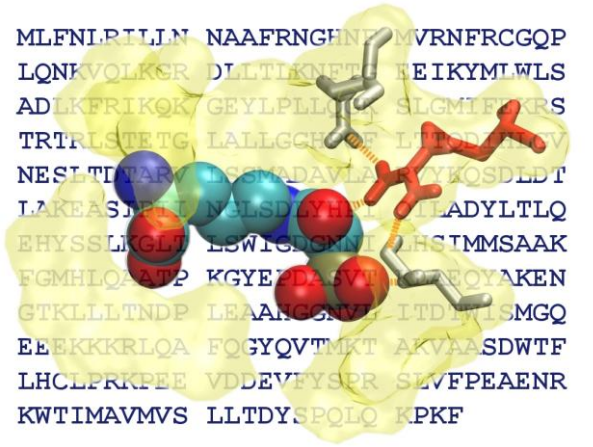
## How to cite UniProtKB

Please use the following reference:
The UniProt Consortium, **Ongoing and future developments at the Universal Protein Resource,** Nucleic Acids Res. 39: D214-D219 (2011).

Updated: September 2011

# A Quick Guide to UniProtKB Swiss-Prot & TrEMBL

**www.uniprot.org**



Contribute
Send feedback

**help@uniprot.org**