

PROTEIN DATABASES

Protein sequence database

i-Prot

Universal **Protein** resource, a central repository of **protein** data created by combining the Swiss-Prot, TrEMBL and PIR-databases.

i-ProtKB

UniProt knowledge base is the central access point for extensive annotated protein information, including function, classification, and cross-reference.

UniProtKB/SWISS-PROT is a manually annotated universal protein database which is always reviewed.

UniProtKB/TrEMBL is an automatically annotated database and is not reviewed.

i-Prot Reference clusters (UniRef)

Provide clustered sets of sequences from the UniProtKB to obtain complete coverage of sequence space at several resolutions



EMBL-EBI

EMBL-EBI



SIB Swiss Institute of Bioinformatics



Swiss Institute of Bioinformatics

Georgian State University



Types of annotations provided

- ✓ Description of a protein's function, domain structure, conflicts between literature references and variants.
- ✓ Provides a high level of integration with other bio-molecular databases.

<https://www.ebi.ac.uk/uniprot/index>

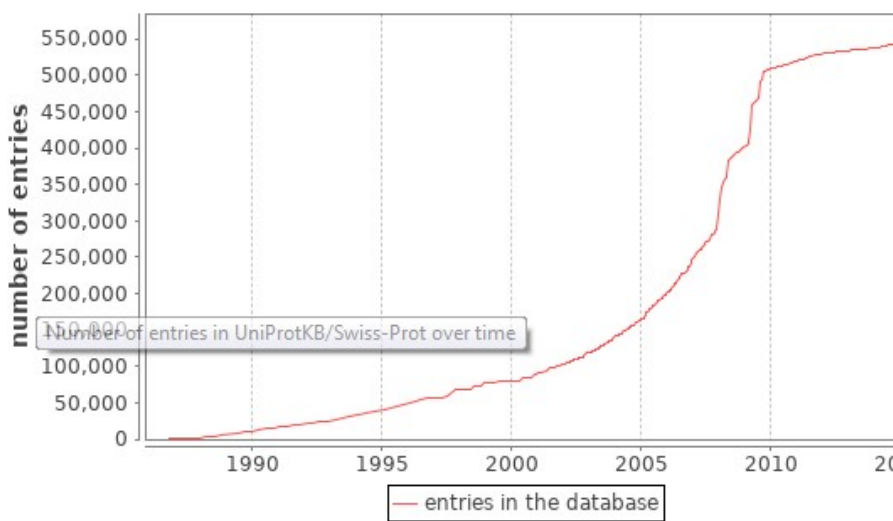
UniprotKB/SWISS-PROT

UniProtKB/Swiss-Prot is the manually annotated and reviewed section of the UniProt Knowledgebase (UniProtKB).

It is a high quality annotated and non-redundant protein sequence database, which brings together experimental results, computed features and scientific conclusions.

Since 2002, it is maintained by the UniProt consortium

Number of entries in UniProtKB/Swiss-Prot over time



Some annotation statistics

General Annotation (comments)

	Annotations	Entries
Allergenic properties	872	872
Alternative products	25,352	25,352
Biophysicochemical properties	8,756	8,756
Biotechnological use	1,069	1,053
Catalytic activity	294,640	240,300
Caution	13,239	12,970
Cofactor	221,158	0
Developmental stage	12,461	12,460
Involvement in disease	7,267	4,886
Mutation phenotype	15,321	15,318
Domain	50,415	43,220

Uni-ProtKB/SWISS-PROT file format

AROE_BUCA5 Reviewed; 273 AA.
B8D9R6;

28-JUL-2009, integrated into UniProtKB/Swiss-Prot.

03-MAR-2009, sequence version 1.

08-MAY-2019, entry version 66.

RecName: Full=Shikimate dehydrogenase (NADP(+)) {ECO:0000255|HAMAP-Rule:MF_00222};

Short=SDH {ECO:0000255|HAMAP-Rule:MF_00222};

EC=1.1.1.25 {ECO:0000255|HAMAP-Rule:MF_00222};

Name=aroE {ECO:0000255|HAMAP-Rule:MF_00222};

OrderedLocusNames=BUAP5A_486;

Buchnera aphidicola subsp. Acyrthosiphon pisum (strain 5A).

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;

Erwiniaceae; Buchnera.

NCBI_TaxID=563178;

[1]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

STRAIN=5A;

PubMed=19150844; DOI=10.1126/science.1167140;

Moran N.A., McLaughlin H.J., Sorek R.;

"The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria.";

Science 323:379-382(2009).

-!- FUNCTION: Involved in the biosynthesis of the chorismate, which leads to the biosynthesis of aromatic amino acids. Catalyzes the reversible NADPH linked reduction of 3-dehydroshikimate (DHSA) to yield shikimate (SA). {ECO:0000255|HAMAP-Rule:MF_00222}.

-!- CATALYTIC ACTIVITY:

Reaction=NADP(+) + shikimate = 3-dehydroshikimate + H(+) + NADPH;

Xref=Rhea:RHEA:17737, ChEBI:CHEBI:15378, ChEBI:CHEBI:16630,

ChEBI:CHEBI:36208, ChEBI:CHEBI:57783, ChEBI:CHEBI:58349;

EC=1.1.1.25; Evidence={ECO:0000255|HAMAP-Rule:MF_00222};

-!- PATHWAY: Metabolic intermediate biosynthesis; chorismate biosynthesis; chorismate from D-erythrose 4-phosphate and phosphoenolpyruvate: step 4/7. {ECO:0000255|HAMAP-Rule:MF_00222}.

-!- SUBUNIT: Homodimer. {ECO:0000255|HAMAP-Rule:MF_00222}.

-!- SIMILARITY: Belongs to the shikimate dehydrogenase family. {ECO:0000255|HAMAP-Rule:MF_00222}.

Description

References

Comments

DR EMBL; CP001161; ACL30837.1; -; Genomic_DNA.
 DR RefSeq; WP_009874444.1; NC_011833.1.
 DR EnsemblBacteria; ACL30837; ACL30837; BUAP5A_486.
 DR KEGG; bap:BUAP5A_486; -.
 DR HOGENOM; HOG000237876; -.
 DR KO; K00014; -.
 DR OMA; FGNP1KH; -.
 DR OrthoDB; 1054867at2; -.
 DR BioCyc; BAPH563178:G1GCU-505-MONOMER; -.
 DR UniPathway; UPA00053; UER00087.
 DR GO; GO:0050661; F:NADP binding; IEA:InterPro.
 DR GO; GO:0004764; F:shikimate 3-dehydrogenase (NADP+) activity; IEA:UniProtKB-UniRule.
 DR GO; GO:0009073; P:aromatic amino acid family biosynthetic process; IEA:UniProtKB-UniRule.
 DR GO; GO:0009423; P:chorismate biosynthetic process; IEA:UniProtKB-UniPathway.
 DR GO; GO:0019632; P:shikimate metabolic process; IEA:InterPro.
 DR HAMAP; MF_00222; Shikimate_DH_AroE; 1.
 DR InterPro; IPR036291; NAD(P)-bd_dom_sf.
 DR InterPro; IPR041121; SDH_C.
 DR InterPro; IPR011342; Shikimate_DH.
 DR InterPro; IPR013708; Shikimate_DH-bd_N.
 DR InterPro; IPR022893; Shikimate_DH_fam.
 DR InterPro; IPR006151; Shikm_DH/Glu-tRNA_Rdtase.
 DR Pfam; PF18317; SDH_C; 1.
 DR Pfam; PF01488; Shikimate_DH; 1.
 DR Pfam; PF08501; Shikimate_dh_N; 1.
 DR SUPFAM; SSF51735; SSF51735; 1.
 DR TIGRFAMs; TIGR00507; aroE; 1.
 PE 3: Inferred from homology;
 KW Amino-acid biosynthesis; Aromatic amino acid biosynthesis; NADP;
 KW Oxidoreductase.
 FT CHAIN 1 273 Shikimate dehydrogenase (NADP(+)).
 FT /FTId=PRO_1000124877.
 FT NP_BIND 131 135 NADP. {ECO:0000255|HAMAP-Rule:MF_00222}.
 FT REGION 19 21 Shikimate binding. {ECO:0000255|HAMAP-
 FT Rule:MF_00222}.
 FT ACT_SITE 70 70 Proton acceptor. {ECO:0000255|HAMAP-
 FT Rule:MF_00222}.
 FT BINDING 66 66 Shikimate. {ECO:0000255|HAMAP-
 FT Rule:MF_00222}.
 FT BINDING 91 91 Shikimate. {ECO:0000255|HAMAP-
 FT Rule:MF_00222}.
 FT BINDING 107 107 Shikimate. {ECO:0000255|HAMAP-
 FT Rule:MF_00222}.
 FT BINDING 218 218 NADP; via carbonyl oxygen.
 FT {ECO:0000255|HAMAP-Rule:MF_00222}.
 FT BINDING 220 220 Shikimate. {ECO:0000255|HAMAP-
 FT Rule:MF_00222}.
 FT BINDING 242 242 NADP; via carbonyl oxygen.
 FT {ECO:0000255|HAMAP-Rule:MF_00222}.
 SQ SEQUENCE 273 AA; 31002 MW; 4EAA9019E77213F9 CRC64;
 MFKCKNFNYA VFGNPINHSK SPEIHSLSFK QTGISHFYKS CNVPLNSLYA VLQDFFKKDG
 RGANITAPFK QEAYFFCNKL TKRAEVAQSV NTLKKIDNCN ILGONTDGIG LLSDLIRLNF
 IKKNYSILII GAGGAARGVL PLLSYGCSI CIFNRTVLNA EKLVLQFHKY GNINVFNTNS
 LHVKSFDLII NATSHFIQDK DNFIPFSCVS SKTCFYDMNY QTDNTFFFDW SRKTGSNFFS
 NGIGMLVFQA AHSFFLWHNV LPEIDYIDL LNK

Database cross reference

Feature table

Sequence

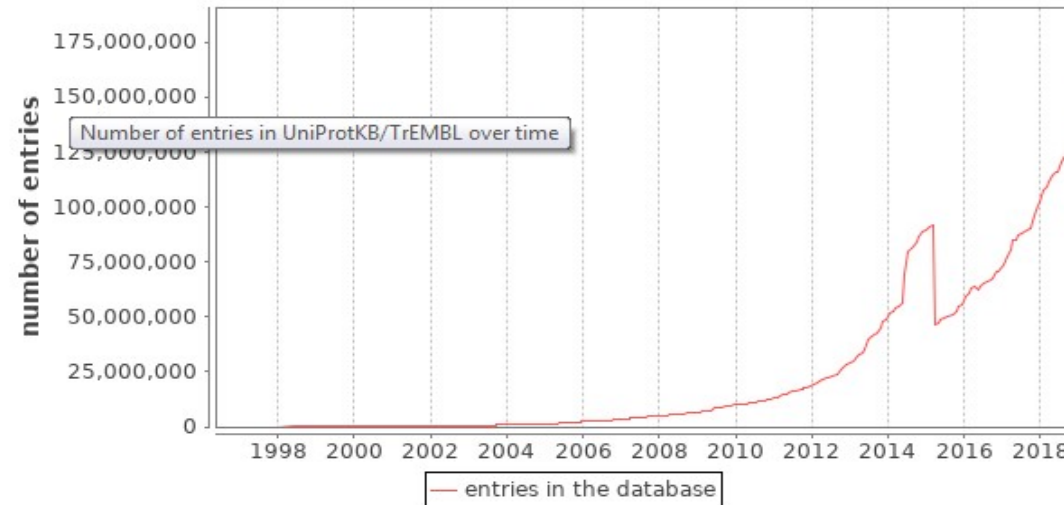
UniprotKB/TrEMBL

TrEMBL is a computer-annotated protein sequence database complementing the UniprotKB/SWISS-PROT.

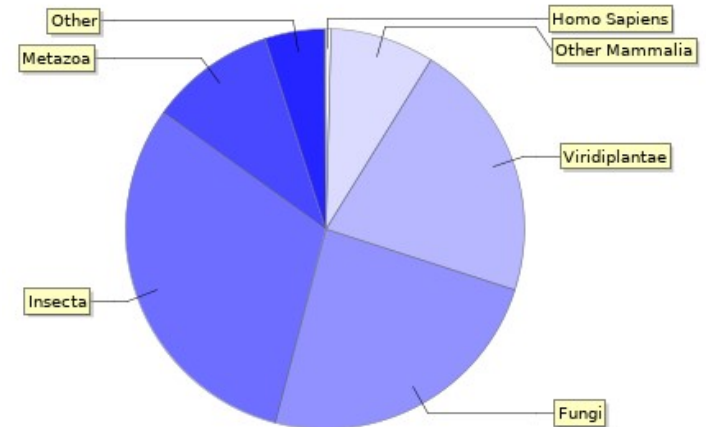
UniProtKB/TrEMBL contains the translations of all coding sequences ([CDS](#)) present in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases and also protein sequences extracted from the literature or submitted to UniProtKB/Swiss-Prot.

The database is enriched with automated classification and annotation.

Number of entries in UniProtKB/TrEMBL over time



TrEMBL entries in Eukaryota



PIR-Protein Information Resources

<https://proteininformationresource.org>

Integrated public [bioinformatics](#) resource to support [genomic](#) and [proteomic](#) research, and scientific studies. It contains protein sequences databases.

A division of National Biomedical Research Foundation (NBRF) in U.S.

One can search for entries or do sequence similarity search at PIR site.

located at Georgetown University Medical Center (GUMC)

The screenshot displays the PIR Protein Information Resource website. At the top, the PIR logo is shown with the text 'A UniProt Consortium Member' and 'Protein Information Resource'. Below this is a navigation bar with links: 'About PIR', 'Resources', 'Search/Analysis', 'Download', and 'Support'. The main heading reads 'INTEGRATED PROTEIN INFORMATICS RESOURCE FOR GENOMIC, PROTEOMIC AND SYSTEMS BIOLOGY RESEARCH'. Below this, a banner for UniProt states: 'The Universal Protein Resource (UniProt) provides the scientific community with a single, centralized, authoritative resource for protein sequences and functional information.' It also lists 'UniProtKB | UniRef | UniParc' and 'Current release: 2019_10'. The page is divided into three main sections: 1. PRO Protein Ontology: Includes 'Representation of protein objects with descriptions and relationships', 'Browse PRO', and 'Annotate with RACE-PRO'. It features a 'Sample PRO report' link. 2. iPTMnet Integrated Protein PTM Resource: Includes 'PTM enzyme-substrate-site relations', 'Visual analysis of PTM networks, cross-talks, and conservation', and 'Batch Retrieval'. It features a 'Sample iPTMnet report' link. 3. iProLINK Literature Information & Knowledge: Includes 'Access to text mining tools and annotated corpora', 'RLIMS-P extraction of kinase, substrate and site', and 'miRtex extraction of miRNA/target information'. It features a 'Sample RLIMS-P report' link. Each section has a corresponding icon and a list of features.

Informations that could be retrieved from Uni-Prot

- Function of the protein
- Post-translational modification
- Domains and sites.
- Secondary structure.
- Quaternary structure.
- Similarities to other proteins;
- Diseases associated with deficiencies in the protein
- Sequence conflicts, variants, etc.