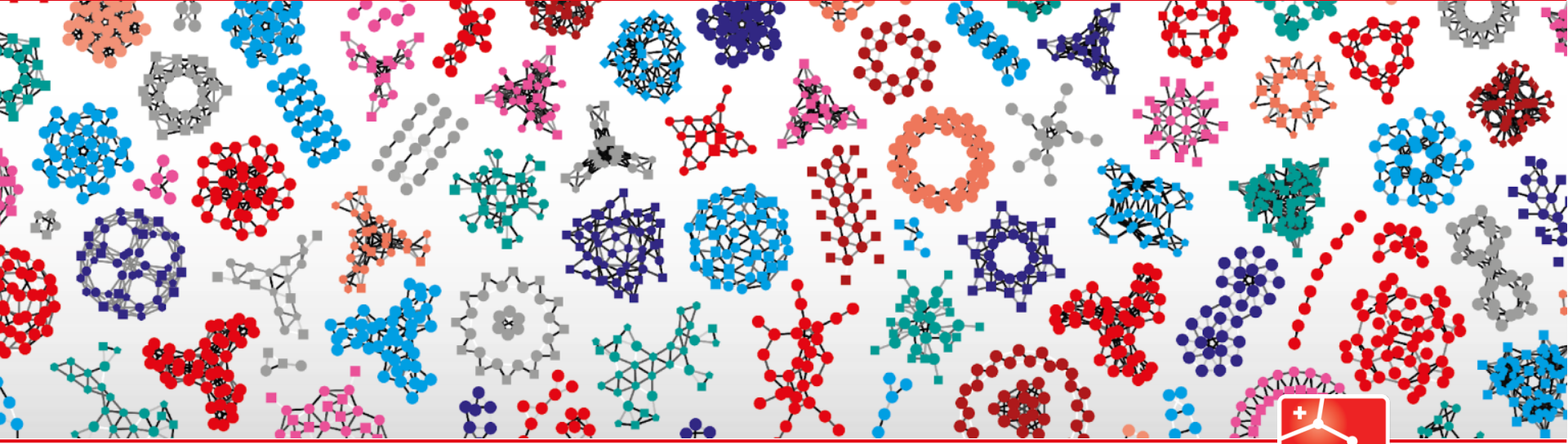


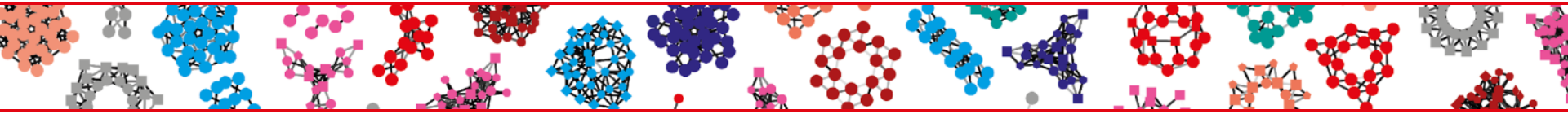
Swiss Institute of Bioinformatics

UniProt

Jerven Bolleman
With the extra help of: Marie Claude-Blatter & Anne Morgat

elixir SWITZERLAND swissprot

Overview



01 Biology

02 Data model

03 SPARQL

04 To Rhea

Protein sequence data: where does it come from ?

- > 190 billion ‘different’ proteins on earth (∑ N species x M genes)
- ~ 182 million ‘known and public’ protein sequences in now
  - 50 % more by next year !
- About 98% of the protein sequences are derived from the **translation of nucleotide sequences** (mRNA or DNA/genome)
- About 1 % come from direct protein sequencing (Edman, MS/MS…)

UniProt consortium : 

EBI : European Bioinformatics Institute (UK)
SIB : Swiss Institute of Bioinformatics (CH)
PIR : Protein Information Resource (USA)

www.uniprot.org

~180 millions of proteins/entries
derived from ~550’000 different species

Millions unique visitors/year

Very regular releases

sparql.uniprot.org
~60 billion triples

Thousands unique visitors/year

Very regular releases

UniProtKB: 2 sections

Major differences in the protein sequence and annotation accuracy !

UniProtKB

UniProt Knowledgebase

Swiss-Prot (561,568)

Manually annotated and reviewed.

Records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL (179,250,561)

Automatically annotated and not reviewed.

Records that await full manual annotation.

UniProtKB record

```
<Q5NUF3> rdf:type up:Protein ;
  up:reviewed true ;
  up:created "2013-10-16"^^xsd:date ;
  up:modified "2017-07-05"^^xsd:date ;
  up:version 69 ;
  up:mnemonic "HIDH_SOYBN" ;
  up:citation citation:15734910 ,
    citation:SIP6664389DA81EBC0D ,
    citation:20075913 ;
  rdfs:seeAlso <http://purl.uniprot.org/embl-cds/BAD80840.1> ,
    <http://purl.uniprot.org/embl-cds/ACU22699.1> ,
    <http://purl.uniprot.org/embl/CM000834#not-annotated-cds-Q5NUF3> ,
    <http://purl.uniprot.org/refseq/NP_001237228.1> ,
    <http://purl.uniprot.org/unigene/Gma.19376> ,
    <http://purl.uniprot.org/proteinmodelportal/Q5NUF3> ,
  ,
    <http://purl.uniprot.org/brenda/4.2.1.105> ,
    <http://purl.uniprot.org/sabio-rk/Q5NUF3> ,
    <http://purl.uniprot.org/genevisible/Q5NUF3> ,
    <http://purl.uniprot.org/gene3d/3.40.50.1820> ,
    <http://purl.uniprot.org/interpro/IPR029058> ,
    <http://purl.uniprot.org/interpro/IPR013094> ,
    <http://purl.uniprot.org/interpro/IPR002168> ,
    <http://purl.uniprot.org/pfam/PF07859> ,
    <http://purl.uniprot.org/supfam/SSF53474> ,
    <http://purl.uniprot.org/prosite/PS01173> ;
  up:recommendedName <Q5NUF3#SIP68A912E5EF4CD7F0> ;
  up:alternativeName <Q5NUF3#SIP07C683DCAE5FEC97> ;
  up:enzyme enzyme:3.1.1.1 ,
    enzyme:4.2.1.105 ;
  up:organism taxon:3847 ;
  up:isolatedFrom tissue:911 ;
  up:encodedBy <Q5NUF3#gene-MD50B25F40E3C6B5FA4BB91CCD8FFB7A35A> ;
  up:annotation <Q5NUF3#SIP4250CFDC6ABD38A1> ,
    <Q5NUF3#SIPEF90B73F14CBF384> ,
    <Q5NUF3#SIP7987EECCBE1E4CBC> ,
    <Q5NUF3#SIP9BA351214B8D4F3C> ,
    <Q5NUF3#SIP18111176CC8EA9445> ,
    <Q5NUF3#SIP6ACCCB66A18AC6BC> ,
    <Q5NUF3#SIP07A4E9E1D6504360> ,
    annotation:PRO_0000424101 ,
    <Q5NUF3#SIPEB602809AF7C0880> ,
    <Q5NUF3#SIPFC72514092BC05CE> ,
    <Q5NUF3#SIP7061CCEC43C5E467> ,
    <Q5NUF3#SIPB0F2DD2C1290461F> ;
  up:existence up:Evidence_at_Protein_Level_Existence ;
  up:classifiedWith keyword:284 ,
    keyword:378 ,
    keyword:456 ,
    keyword:1185 ,
    go:0033987 ,
    go:0052689 ,
    go:0009056 ,
    go:0009813 ,
    go:0009717 ,
    go:0046287 ;
  up:sequence isoform:Q5NUF3-1 ;
  up:attribution <Q5NUF3#attribution-FBA1515ECDA7D437CE9020783EBAA82C> ,
    <Q5NUF3#attribution-B1173712290FD2B95AA2CFBCD6874F84> ,
    <Q5NUF3#attribution-07EA7F9C35067AFDE011C65A7D7C4347> ,
    <Q5NUF3#attribution-4E2B9E0D3E37BD315B8F2E9BF674303D> ,
    <Q5NUF3#attribution-1F1D235CEE6E9F517881B603655DF439> ,
    <Q5NUF3#attribution-0CFF399BA45A6BFAB52E9F7FDC58CBD9> ;
  up:proteome <http://purl.uniprot.org/proteomes/UP000008827#Chromosome%201> .
citation:15734910 rdf:type up:Journal_Citation ;
  up:title "Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement
  up:submittedTo "EMBL/GenBank/DDBJ" .
```

Biological information / annotation

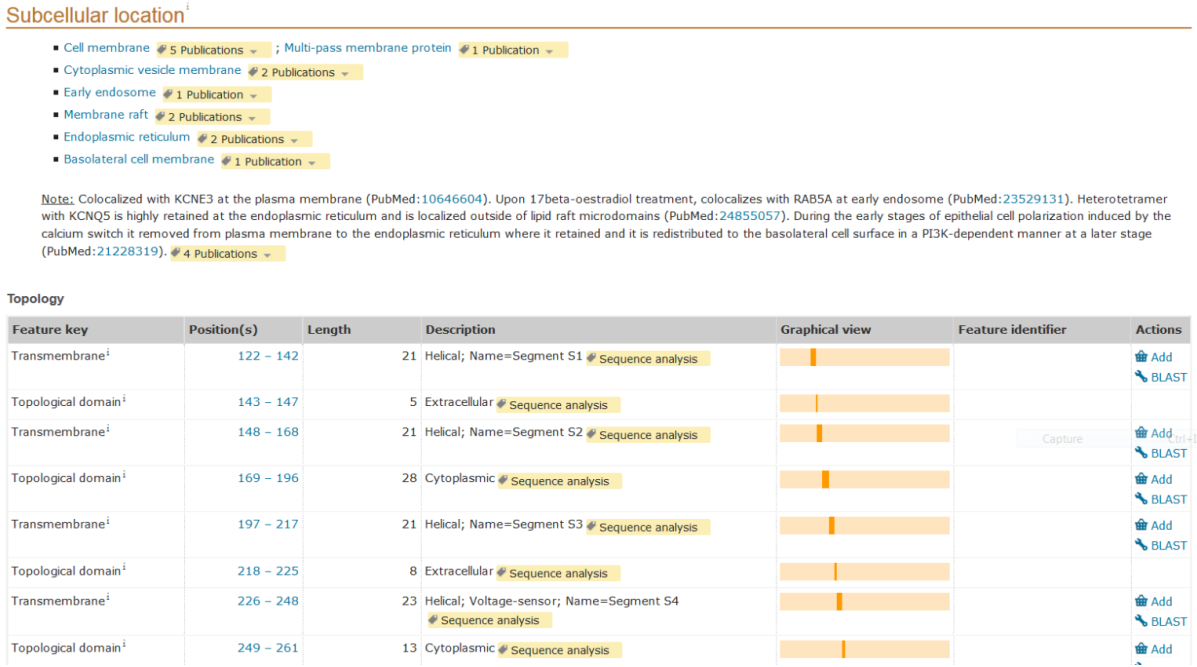
```
SQ   SEQUENCE   319 AA;  35138 MW;  E8333CF425FBA4A3 CRC64;
     MAKEIVKELL PLIRVYKDGS VERLLSSENV AASPEDPQTG VSSKDIVIAD NPYVSARIFL
     PKSHHTNNKL PIFLYFHGGA FCVESAFSFF VHRYLNILAS EANIIAISVD FRLLPHHPIP
     AAYEDGWTTL KWIASHANNT NTTNPEPWLL NHADFTKVYV GGETSGANIA HNLLLRAGNE
     SLPGDLKILG GLLCCPFFWG SKPIGSEAVE GHEQSLAMKV WNFACPDAPG GIDNPWINPC
     VPGAPSLATL ACSKLLVTIT GKDEFRDRDI LYHHTVEQSG WQGELQLFDA GDEEHAFQLF
     KPETHLAKAM IKRLASFLV
```

protein sequence

- Text search
- Training Dataset
- Statistics
- Genome annotation (Features)
- System biology
- …

- Proteomics
- BLAST
- Phylogeny
- Training datasets
- Domains
- …

Source of biological knowledge / annotation



Subcellular location

- Cell membrane — 5 Publications ; Multi-pass membrane protein — 1 Publication
- Cytoplasmic vesicle membrane — 2 Publications
- Early endosome — 1 Publication
- Membrane raft — 2 Publications
- Endoplasmic reticulum — 2 Publications
- Basolateral cell membrane — 1 Publication

Note: Colocalized with KCNE3 at the plasma membrane (PubMed:10646604). Upon 17beta-oestradiol treatment, colocalizes with RAB5A at early endosome (PubMed:23529131). Heterotetramer with KCNQ5 is highly retained at the endoplasmic reticulum and is localized outside of lipid raft microdomains (PubMed:24855057). During the early stages of epithelial cell polarization induced by the calcium switch it removed from plasma membrane to the endoplasmic reticulum where it retained and it is redistributed to the basolateral cell surface in a PI3K-dependent manner at a later stage (PubMed:21228319). 4 Publications

Topology

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Transmembrane	122 – 142	21	Helical; Name=Segment S1 Sequence analysis			Add BLAST
Topological domain	143 – 147	5	Extracellular Sequence analysis			
Transmembrane	148 – 168	21	Helical; Name=Segment S2 Sequence analysis			Add BLAST
Topological domain	169 – 196	28	Cytoplasmic Sequence analysis			Add BLAST
Transmembrane	197 – 217	21	Helical; Name=Segment S3 Sequence analysis			Add BLAST
Topological domain	218 – 225	8	Extracellular Sequence analysis			
Transmembrane	226 – 248	23	Helical; Voltage-sensor; Name=Segment S4 Sequence analysis			Add BLAST
Topological domain	249 – 261	13	Cytoplasmic Sequence analysis			Add

Experimental data

(publication)

Computational analysis

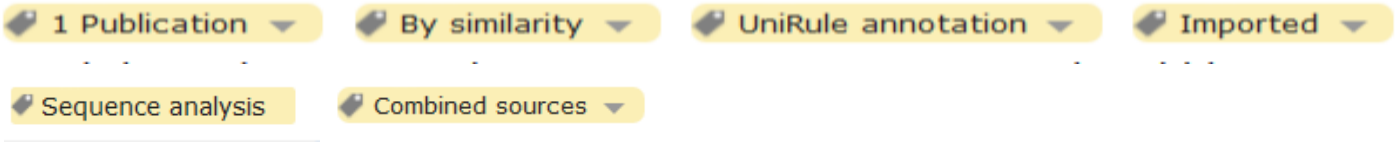
(curator-evaluated or not)

UniProtKB - P51787 (KCNQ1_HUMAN)

- a summary of knowledge in free text
- structured and machine-readable information

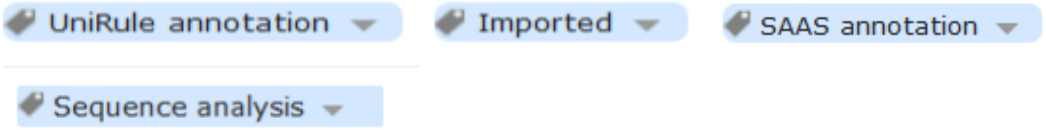
Source of annotation / Evidence statements

UniProtKB/Swiss-Prot: Manual insertion, color in yellow



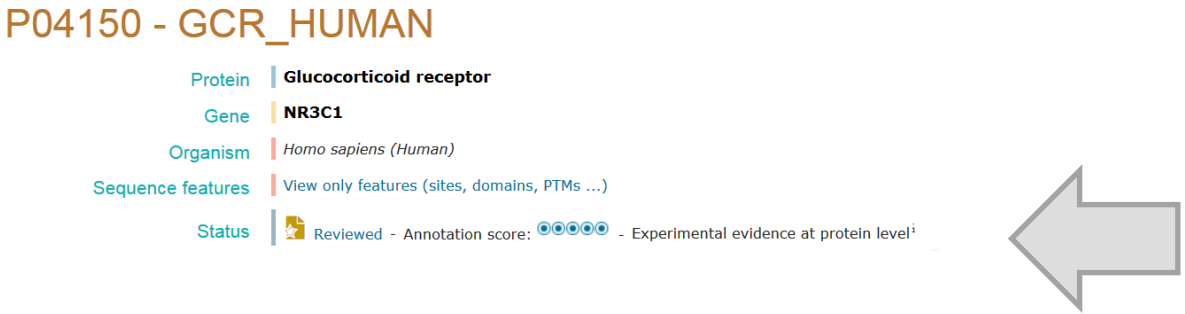
Computational analysis (curator-evaluated)

UniProtKB/TrEMBL: Automated insertion, color in blue



Computational analysis (NOT curator-evaluated)

Annotation score & evidence for protein existence

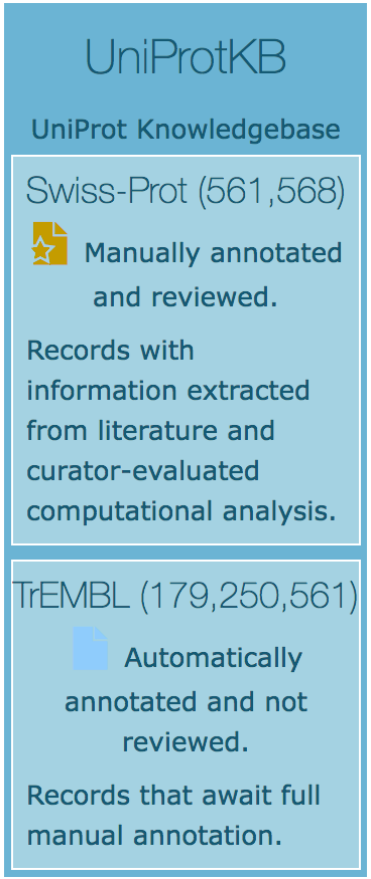


Status:

- Reviewed / Unreviewed
- Annotation score: http://insideuniprot.blogspot.ch/2014/10/introducing-annotation-scores-in-uniprot.html
- Evidence for protein existence

1. Experimental evidence at protein level
2. Experimental evidence at transcript level
3. Protein inferred from homology
4. Protein predicted
5. Protein uncertain

UniProtKB/TrEMBL

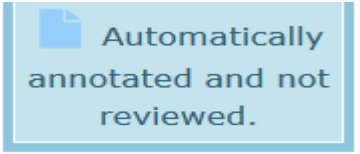


One protein sequence per entry

gene-centric **/protein-centric**

99 % of UniProtKB protein sequences

UniProtKB/TrEMBL
Automatic annotation



Protein sequence

- The quality of the protein sequences is dependent on the information provided by the submitter of the original nucleotide entry (CDS) or of the gene prediction pipeline (i.e. Ensembl).
- 100% identical sequences (same length, same organism are merged automatically).

Biological information

Sources of annotation

- Provided by the submitter (EMBL, PDB, TAIR…)
- Automated annotation

P73722 - P73722_SYNY3

Automatically annotated and not reviewed.

Protein	Submitted name: **SOS function regulatory protein**
Gene	**lexA**
Organism	*Synechocystis sp. (st… 803 / Kazusa)*
Status	Unreviewed - ●●○○○ - Protein inferred from homology[i]

Function[i]

Represses a number of genes involved in the response to DNA damage (SOS response), including recA and lexA. In the presence of single-stranded DNA, RecA interacts with LexA causing an autocatalytic cleavage which disrupts the DNA-binding part of LexA, leading to derepression of the SOS regulon and eventually DNA repair By similarity .

SAAS annotations

Catalytic activity[i]

Hydrolysis of Ala-|-Gly bond in repressor LexA. SAAS annotations

Keywords - Molecular function[i]

Hydrolase SAAS annotations , Repressor SAAS annotations

Keywords - Biological process[i]

DNA damage, DNA repair, DNA replication SAAS annotations , SOS response SAAS annotations , Transcription, Transcription regulation SAAS annotations

Keywords - Ligand[i]

DNA-binding SAAS annotations

Capture Ctrl+Ins

UniProtKB/Swiss-Prot

UniProtKB

UniProt Knowledgebase

Swiss-Prot (561,568)

Manually annotated and reviewed.

Records with information extracted from literature and curator-evaluated computational analysis.

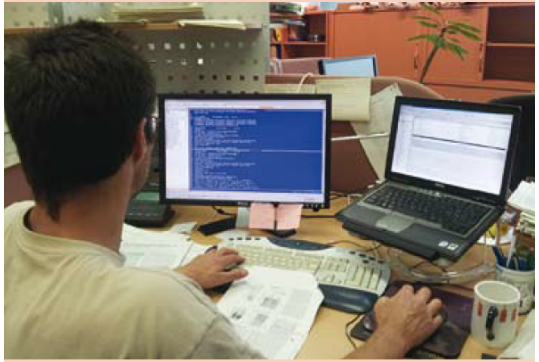
TrEMBL (179,250,561)

Automatically annotated and not reviewed.

Records that await full manual annotation.

1 % of UniProtKB protein sequences

What is a biocurator?



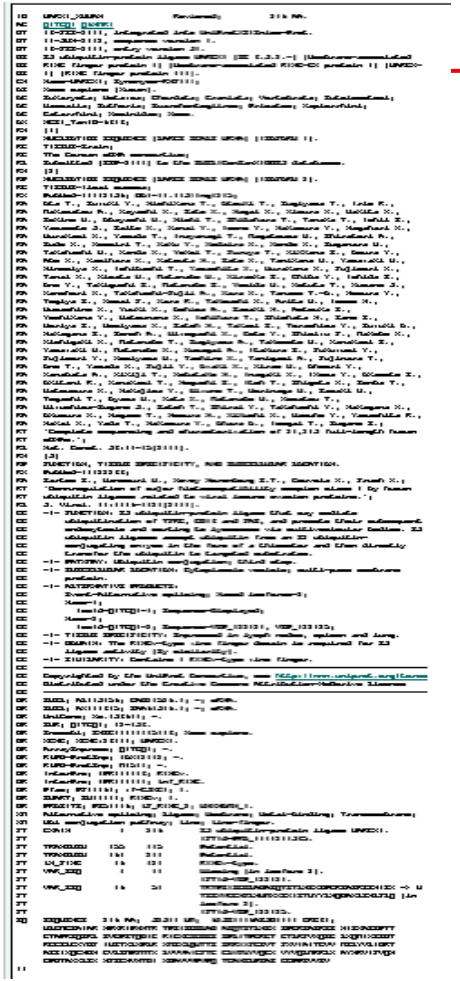
They spend their time behind the scenes, yet biocurators are essential to knowledge maintenance. The Swiss-Prot biocurators

UniProtKB/Swiss-Prot

UniProtKB/TrEMBL

```
ID   Q8TCG1_HUMAN            Unreviewed;       289 AA.
AC   Q8TCG1
DT   01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT   01-JUN-2002, sequence version 1.
DT   08-JAN-2007, entry version 23.
DE   Hypothetical protein DKFZp564M1682.
GN   Name=MARCH1; Synonyms=DKFZp564M1682;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
OC   Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RG   The German cDNA Consortium;
RA   Fosatka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Cytoplasm (Probable).
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AL713759; CAD28529.1; -; mRNA.
DR   EMBL; Q8TCG1; 72-126.
DR   Ensembl; ENSG00000145416; Homo sapiens.
DR   HGNC; HGNC:26077; MARCH1.
DR   ArrayExpress; Q8TCG1; -.
DR   RZPD-ProtExp; IOH42112; -.
DR   RZPD-ProtExp; W1571; -.
DR   GO; GO:0046872; F:metal ion binding; IEA:UniProtKB-KW.
DR   GO; GO:0005515; F:protein binding; IEA:InterPro.
DR   GO; GO:0008270; F:zinc ion binding; IEA:InterPro.
DR   GO; GO:0006512; P:ubiquitin cycle; IEA:UniProtKB-KW.
DR   InterPro; IPR011016; RING-v.
DR   InterPro; IPR001841; Znf_RING.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00744; RINGv; 1.
DR   PROSITE; PS00097; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Ubl conjugation pathway; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   289 AA;  32308 MW;  923E1809AB3D7087 CRC64;
     MLGWCEAIAR NDMRIPNNTR TPEISGDLAD ASQTSTLNEK SPGRSASRSS NISKASSPTT
     GTAPRSQRRL SVCPSPTQDI C KICHCEGDES SPLITPCRCT GTLNFVHQSC LNQWIKSSDT
     RCCELCKYDF IMETKLKPLR KWEKLQMTTS ERRKIFCSVT FHVIAITCVV WSLVVLIDRT
     AEEIKQGNDN GVLEWEPWTK LVVVAIGFTG GLVFMVVQCR VVVQLWRRLK AYNRVIFVQN
     CPDTAKLEK NFSCNVNTDI KDAVVVPVPQ TGANSLPEAE GSPEVVSSV
//
```

Manual annotation of the sequence and associated biological information



- **At least 20% of UniProtKB/Swiss-Prot entries required curation effort to “correct” the sequences.**

- **Typical problems**
  - unsolved conflicts;
  - uncorrected initiation sites;
  - frameshifts;
  - other ‘problems’

UniProtKB/Swiss-Prot

Biological knowledge / annotation / GO annotation

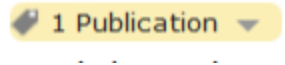
Knowledge:

- comprehensive summary (free text) that provides a complete overview of the information available
- standardized vocabularies to facilitate subsequent retrieval whenever possible

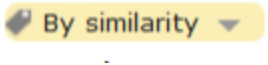
Source of annotation / Evidence statements

Every piece of knowledge is associated with the source of the information and the type of supporting evidence, using the evidence ontology (ECO)

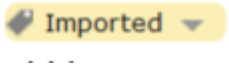
- Selected Publication (experimental)



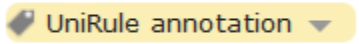
- Another UniProtKB entry (orthologs): *by similarity*



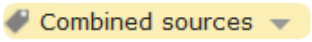
- An entry from another database: *imported*



- Curator-evaluated computational analysis



- Combined sources



Selected Publication

PubMed=16595657

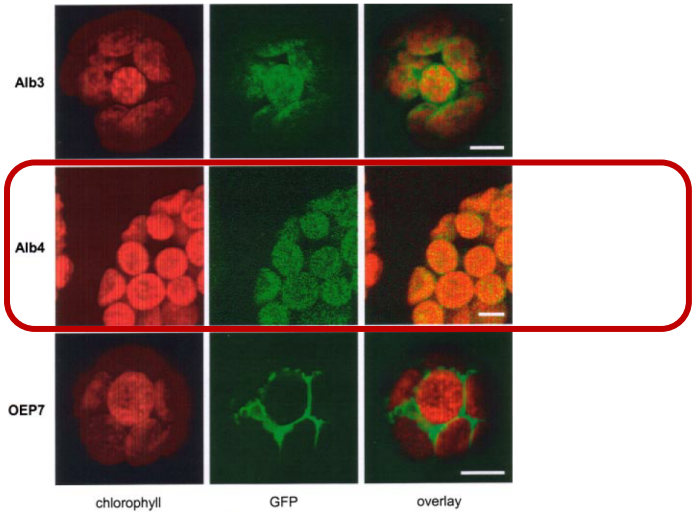


FIGURE 3. **Subcellular localization of Alb3-GFP and Alb4-GFP fusion proteins.** *Arabidopsis* mesophyll protoplasts were transiently transformed with constructs for Alb3- or Alb4-GFP. Maximum intensity signals from confocal images are shown for chlorophyll autofluorescence, GFP fluorescence, and an overlay of both. OEP7-GFP is included as a marker for the chloroplast envelope. *Bar*, 5 μm.

comprehensive and computer friendly representation of biological knowledge

corresponds to the predicted length of the Alb4 mRNA. Even after prolonged exposure of the blots treated with the Alb4 probe, no signal could be found at ~3.5 kb, the predicted size of the Artemis transcript.

Alb4 Is a Thylakoid Membrane Protein—Alb4 is predicted to have a chloroplast transit peptide with a processing site after amino acid residue 45 based on the ChloroP prediction program (33). To test this prediction, *Arabidopsis* mesophyll protoplasts were transiently transformed with constructs for Alb3-GFP or Alb4-GFP. Merging of the GFP and autofluorescence images indicated a thylakoid localization of Alb4. The GFP distribution for Alb4 is similar to that of Alb3 and not to that of outer envelope protein AtOEP7 (Fig. 3). To test this assumption, *in vitro* translated radiolabeled Alb4 was imported into isolated pea

UniProtKB Q9FYL3

PubMed=16595657

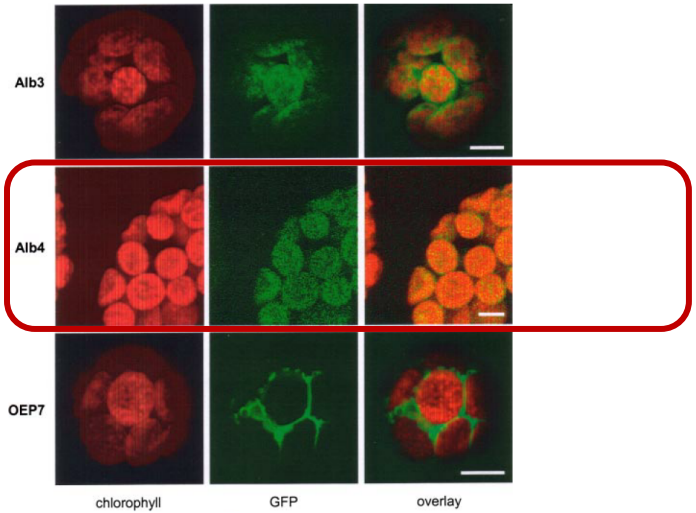


FIGURE 3. **Subcellular localization of Alb3-GFP and Alb4-GFP fusion proteins.** *Arabidopsis* mesophyll protoplasts were transiently transformed with constructs for Alb3- or Alb4-GFP. Maximum intensity signals from confocal images are shown for chlorophyll autofluorescence, GFP fluorescence, and an overlay of both. OEP7-GFP is included as a marker for the chloroplast envelope. *Bar*, 5 μm.

Subcellular location

- Plastid › chloroplast thylakoid membrane — 1 Publication ; Multi-pass membrane protein — 1 Publication

Topology

Feature key	Position(s)	Length	Description	Graphical view
Transmembrane	115 – 135	21	Helical — Sequence analysis	
Transmembrane	184 – 204	21	Helical — Sequence analysis	
Transmembrane	263 – 283	21	Helical — Sequence analysis	
Transmembrane	302 – 322	21	Helical — Sequence analysis	

GO - Cellular component

- chloroplast — Source: TAIR
- chloroplast thylakoid membrane — Source: TAIR
- integral component of membrane — Source: UniProtKB-KW
- thylakoid — Source: TAIR

Controlled vocabulary
GO annotation

2. **"A second thylakoid membrane-localized Alb3/OxaI/YidC homologue is involved in proper chloroplast biogenesis in Arabidopsis thaliana."**
Gerdes L., Bals T., Klostermann E., Karl M., Philippar K., Huenken M., Soll J., Schuenemann D.
J. Biol. Chem. 281:16632-16642(2006) [PubMed] [Europe PMC] [Abstract]
Cited for: SEQUENCE REVISION, TISSUE SPECIFICITY, SUBCELLULAR LOCATION.

corresponds to the predicted length of the Alb4 mRNA. Even after prolonged exposure of the blots treated with the Alb4 probe, no signal could be found at ~3.5 kb, the predicted size of the Artemis transcript.

Alb4 Is a Thylakoid Membrane Protein—Alb4 is predicted to have a chloroplast transit peptide with a processing site after amino acid residue 45 based on the ChloroP prediction program (33). To test this prediction, *Arabidopsis* mesophyll protoplasts were transiently transformed with constructs for Alb3-GFP or Alb4-GFP. Merging of the GFP and autofluorescence images indicated a thylakoid localization of Alb4. The GFP distribution for Alb4 is similar to that of Alb3 and not to that of outer envelope protein AtOEP7 (Fig. 3). To test this assumption, *in vitro* translated radiolabeled Alb4 was imported into isolated pea

Curator-evaluated computational analysis

PubMed=16595657

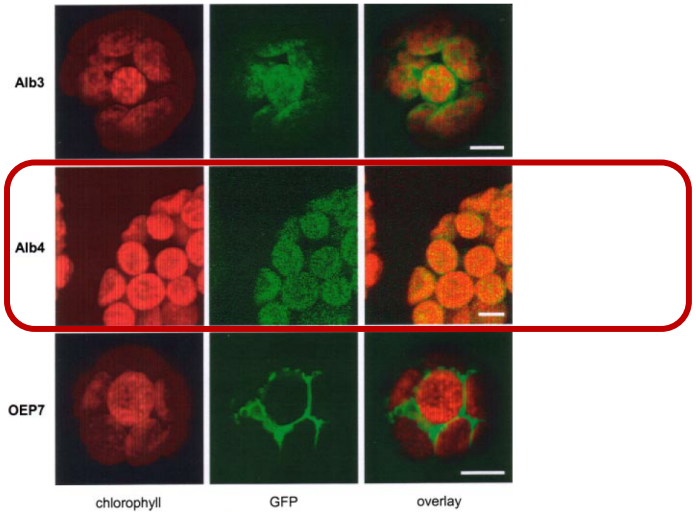


FIGURE 3. **Subcellular localization of Alb3-GFP and Alb4-GFP fusion proteins.** *Arabidopsis* mesophyll protoplasts were transiently transformed with constructs for Alb3- or Alb4-GFP. Maximum intensity signals from confocal images are shown for chlorophyll autofluorescence, GFP fluorescence, and an overlay of both. OEP7-GFP is included as a

corresponds to the predicted length of the Alb4 mRNA. Even after prolonged exposure of the blots treated with the Alb4 probe, no signal could be found at ~3.5 kb, the predicted size of the Artemis transcript.

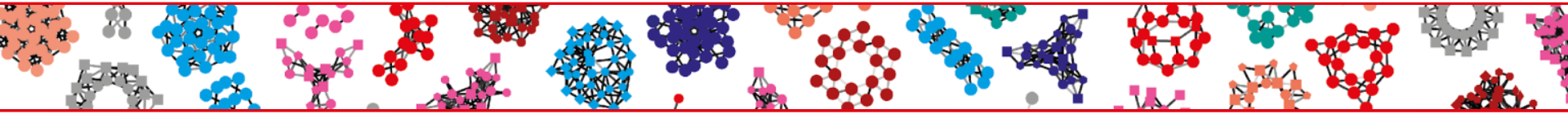
Alb4 Is a Thylakoid Membrane Protein—Alb4 is predicted to have a chloroplast transit peptide with a processing site after amino acid residue 45 based on the ChloroP prediction program (33). To test this prediction, *Arabidopsis* mesophyll protoplasts were transiently transformed with constructs for Alb3-GFP or Alb4-GFP. Merging of the GFP and autofluorescence images indicated a thylakoid localization of Alb4. The GFP distribution for Alb4 is similar to that of Alb3 and not to that of outer envelope protein AtOEP7 (Fig. 3). To test this assumption, *in vitro* translated radiolabeled Alb4 was imported into isolated pea

UniProtKB
Q9FYL3

Molecule processing

Feature key	Position(s)	Length	Description
Transit peptide [i]	1 – 45	45	Chloroplast Sequence analysis
Chain [i]	46 – 499	454	ALBINO3-like protein 1, chloroplastic

Overview



01 Biology

02 Data model

03 SPARQL

04 To Rhea

The FAIRest format of them all

UniProt Code

```
<P05067> rdf:type up:Protein ;
 up:reviewed              true ; #This is a Swiss-Prot entry
 up:organism       taxon:9606 ; #Talking about a human entry
 up:classifiedWith
         [ go:GO_0043198 rdf:type        owl:Class ;
                         rdfs:label      "dendritic shaft" ;
                         rdfs:subClassOf go:GO_0044463 ,
                                         go:GO_0097458 ,
                                         go:GO_0030425 ,
                                         [ owl:restriction
[ owl:onProperty obo:BFO_0000050 ;
owl:someValuesFrom go:GO_0030425 .

]]
                  ]
```

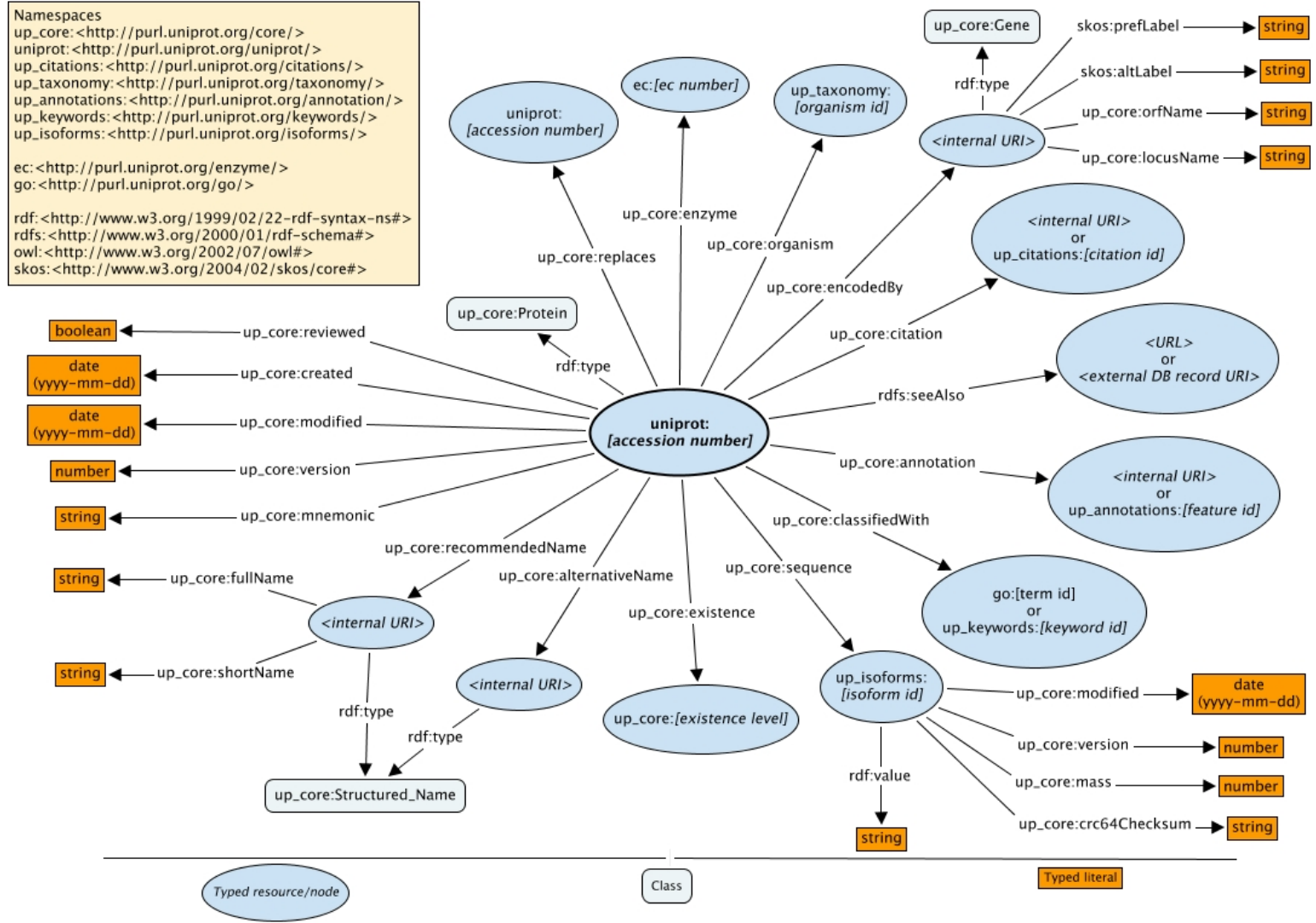
GO Code

UniProt RDF

- **https://sparql.uniprot.org**
- **400+ files on FTP**
  - **Fast loading into your own DB**
  - **Categorized by taxonomy**
  - **Or dataset**
- **Downloadable subsets on our website**
  - **Slice and dice as you wish**

UniProt RDF

- **It’s huge !**
  - **Terabytes !**
- **Query for something**
  - **Get back 10 billion rows**
  - **Protection for this in the HTML view**


Namespaces
up_core:<http://purl.uniprot.org/core/>
uniprot:<http://purl.uniprot.org/uniprot/>
up_citations:<http://purl.uniprot.org/citations/>
up_taxonomy:<http://purl.uniprot.org/taxonomy/>
up_annotations:<http://purl.uniprot.org/annotation/>
up_keywords:<http://purl.uniprot.org/keywords/>
up_isoforms:<http://purl.uniprot.org/isoforms/>
ec:<http://purl.uniprot.org/enzyme/>
go:<http://purl.uniprot.org/go/>
rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
rdfs:<http://www.w3.org/2000/01/rdf-schema#>
owl:<http://www.w3.org/2002/07/owl#>
skos:<http://www.w3.org/2004/02/skos/core#>
uniprot: [accession number]
ec:[ec number]
up_taxonomy: [organism id]
up_core:Gene
skos:prefLabel
string
skos:altLabel
string
up_core:orfName
string
up_core:locusName
string
rdf:type
<internal URI>
up_core:enzyme
up_core:organism
up_core:replaces
up_core:encodedBy
<internal URI> or up_citations:[citation id]
up_core:citation
up_core:Protein
rdf:type
boolean
up_core:reviewed
date (yyyy-mm-dd)
up_core:created
date (yyyy-mm-dd)
up_core:modified
number
up_core:version
string
up_core:mnemonic
uniprot: [accession number]
rdfs:seeAlso
<URL> or <external DB record URI>
up_core:annotation
<internal URI> or up_annotations:[feature id]
up_core:classifiedWith
go:[term id] or up_keywords:[keyword id]
up_core:recommendedName
up_core:alternativeName
up_core:sequence
up_core:existence
string
up_core:fullName
<internal URI>
string
up_core:shortName
rdf:type
rdf:type
<internal URI>
up_core:existence
up_core:[existence level]
up_isoforms: [isoform id]
up_core:modified
date (yyyy-mm-dd)
up_core:version
number
up_core:mass
number
up_core:crc64Checksum
string
rdf:value
string
up_core:Structured_Name
Typed resource/node
Class
Typed literal

Annotation

- 27 subtypes

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX up:<http://purl.uniprot.org/core/>
SELECT (COUNT(?class) AS ?classes)
FROM <http://purl.uniprot.org/core/>
WHERE
{
    ?class a owl:Class ;
             rdfs:subClassOf up:Annotation .
}
```

Annotation

http://biohackathon.org/resource/faldo#location	faldo:location
http://purl.uniprot.org/core/catalyticActivity	up:catalyticActivity
http://purl.uniprot.org/core/catalyzedPhysiologicalReaction	up:catalyzedPhysiologicalReaction
http://purl.uniprot.org/core/cofactor	up:cofactor
http://purl.uniprot.org/core/conflictingSequence	up:conflictingSequence
http://purl.uniprot.org/core/disease	up:disease
http://purl.uniprot.org/core/frameshift	up:frameshift
http://purl.uniprot.org/core/locatedIn	up:locatedIn
http://purl.uniprot.org/core/maximum	up:maximum
http://purl.uniprot.org/core/measuredActivity	up:measuredActivity
http://purl.uniprot.org/core/measuredAffinity	up:measuredAffinity
http://purl.uniprot.org/core/measuredError	up:measuredError
http://purl.uniprot.org/core/measuredValue	up:measuredValue
http://purl.uniprot.org/core/method	up:method
http://purl.uniprot.org/core/range	up:range
http://purl.uniprot.org/core/sequence	up:sequence
http://purl.uniprot.org/core/substitution	up:substitution
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	a
http://www.w3.org/2000/01/rdf-schema#comment	rdfs:comment
http://www.w3.org/2000/01/rdf-schema#seeAlso	rdfs:seeAlso
http://www.w3.org/2004/02/skos/core#related	skos:related

Function annotation (web view)

Functions as a cell surface receptor and performs physiological functions on the surface of neurons relevant to neurite growth, neuronal adhesion and axonogenesis. Interaction between APP molecules on neighboring cells promotes synaptogenesis (PubMed:25122912). Involved in cell mobility and transcription regulation through protein-protein interactions. Can promote transcription activation through binding to APBB1-KAT5 and inhibits Notch signaling through interaction with Numb. Couples to apoptosis-inducing pathways such as those mediated by G(O) and JIP. Inhibits G(o) alpha ATPase activity (By similarity). Acts as a kinesin I membrane receptor, mediating the axonal transport of beta-secretase and presenilin 1 (By similarity). By acting as a kinesin I membrane receptor, plays a role in axonal anterograde transport of cargo towards synapes in axons (PubMed:17062754, PubMed:23011729). Involved in copper homeostasis/oxidative stress through copper ion reduction. In vitro, copper-metallated APP induces neuronal death directly or is potentiated through Cu^{2+}-mediated low-density lipoprotein oxidation. Can regulate neurite outgrowth through binding to components of the extracellular matrix such as heparin and collagen I and IV. The splice isoforms that contain the BPTI domain possess protease inhibitor activity. Induces a AGER-dependent pathway that involves activation of p38 MAPK, resulting in internalization of amyloid-beta peptide and leading to mitochondrial dysfunction in cultured cortical neurons. Provides Cu^{2+} ions for GPC1 which are required for release of nitric oxide (NO) and subsequent degradation of the heparan sulfate chains on GPC1. By similarity 3 Publications

Function annotation (RDF)

<P05067#SIP8BC30A3A03BD108B> rdf:type up:Function_Annotation ;

rdfs:comment "Functions as a cell surface receptor and performs physiological functions on the surface of neurons relevant to neurite growth, neuronal adhesion and axonogenesis. Interaction between APP molecules on neighboring cells promotes synaptogenesis (PubMed:25122912). Involved in cell mobility and transcription regulation through protein-protein interactions. Can promote transcription activation through binding to APBB1-KAT5 and inhibits Notch signaling through interaction with Numb. Couples to apoptosis-inducing pathways such as those mediated by G(O) and JIP. Inhibits G(o) alpha ATPase activity (By similarity). Acts as a kinesin I membrane receptor, mediating the axonal transport of beta-secretase and presenilin 1 (By similarity). By acting as a kinesin I membrane receptor, plays a role in axonal anterograde transport of cargo towards synapes in axons (PubMed:17062754, PubMed:23011729). Involved in copper homeostasis/oxidative stress through copper ion reduction. In vitro, copper-metallated APP induces neuronal death

Sequence

- Isoforms and Canonical materialized
- IUPAC code
  - No spaces

Cross-reference

- 173 databases `rdfs:seeAlso` → `up:database db:{DB}`
- Ensembl (like)
  - `up:translatedTo`
  - `up:transcribedFrom`
- PDB
  - `http://rdf.wwpdb.org/pdb/1AAP`
- All others
  - `rdfs:comment`
- Try to use their (your? IRI) else purl.uniprot.org/{DB}/{ID}
- HAMAP, InterPro, PFam etc.
  - `up:signatureSequenceMatch`

Cross-reference: Ensembl (and ensembl like)

```
<http://rdf.ebi.ac.uk/resource/ensembl.transcript/ENST00000233072> ↷
  rdf:type ↷
    up:Transcript_Resource ;
  up:database <http://purl.uniprot.org/database/Ensembl> ;
  up:translatedTo <http://rdf.ebi.ac.uk/resource/ensembl.protein/ENSP00000233072> ;
  up:transcribedFrom <http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000021826> ;
  rdfs:seeAlso isoform:P31327-1 .
```

Cross-reference: PDB

```
<http://rdf.wwpdb.org/pdb/2YVQ> rdf:type up:Structure_Resource ;
  up:database <http://purl.uniprot.org/database/PDB> ;
  up:method up:X-Ray_Crystallography ;
  up:resolution "1.98"^^xsd:float ;
  up:chainSequenceMapping ↴
       <http://purl.uniprot.org/isoforms/P31327-1#PDB_2YVQ_tt1343tt1478> .
<http://purl.uniprot.org/isoforms/P31327-1#PDB_2YVQ_tt1343tt1478> up:chain ↴
       "A=1343-1478" .
```

Evidence

Reifies the statement ↻

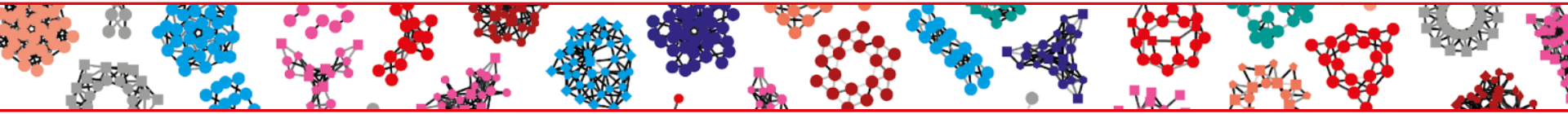
Triple ?s ?p ?o ↷

```
[] a rdf:Statement  ;
   rdf:subject ?s ;
   rdf:predicate ?p ;
   rdf:object ?o ;
   up:attribution ?whoWhatWhy .
```

Evidence (example)

```
<P05067#attribution-7B6976B7E1FAA17744B06EA4C9F47A94>↴
    dcterms:creator↴
        <http://purl.uniprot.org/goa-projects/ARUK-UCL> ;
  up:manual true ;
  up:evidence ECO:0000314 ;
  up:source citation:18723082 .
```

Overview



01 Biology

02 Data model

03 SPARQL

04 To Rhea

Annotation

- 27 subtypes

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX up:<http://purl.uniprot.org/core/>
SELECT (COUNT(?class) AS ?classes)
FROM <http://purl.uniprot.org/core/>
WHERE
{
    ?class a owl:Class ;
             rdfs:subClassOf up:Annotation .
}
```

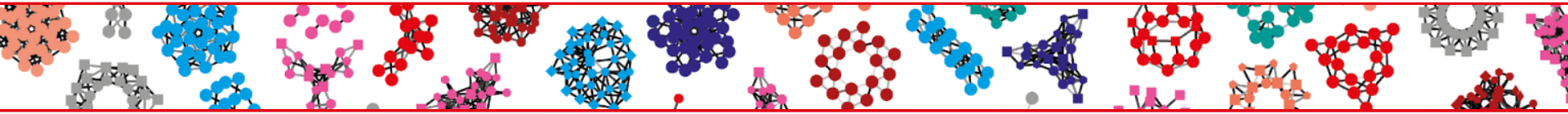
Sequence

```
BASE <http://purl.uniprot.org/uniprot/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX up:<http://purl.uniprot.org/core/>
SELECT ?protein ?aaSequence
FROM <http://sparql.uniprot.org/uniprot>
WHERE
{
  BIND (<P05067> AS ?protein)
  ?protein up:sequence ?sequence .
  ?sequence rdf:value ?aaSequence .
}
```

Evidence

- 4,258,699,129
  - Often needs at least 3 joins
  - Probably slowest

Overview



01 Biology

02 Data model

03 SPARQL

04 To Rhea

Query: Retrieve the evidences of Rhea reactions annotated in UniProtKB/Swiss-Prot

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>

SELECT DISTINCT
    ?upProtein
    ?rheaReaction
    ?rheaReaction_evt
    ?source
WHERE {
  ?upProtein up:reviewed true .
  #
  ?upProtein up:annotation ?annotation .
  ?annotation a up:Catalytic_Activity_Annotation ;
         up:catalyticActivity ?catalyticActivity .
  ?catalyticActivity up:catalyzedReaction ?rheaReaction .
  #
  ?upProtein up:attribution ?attribution .
  ?attribution up:evidence ?rheaReaction_evt .
  [] rdf:subject ?annotation ;
     rdf:predicate up:catalyticActivity ;
     rdf:object ?catalyticActivity ;
     up:attribution ?attribution .
  OPTIONAL {?attribution up:source ?source .}
}
```

Result: Retrieve the evidences of Rhea reactions annotated in UniProtKB/Swiss-Prot

Results

Sparql XML | Sparql JSON | CSV | Show query | Share

upProtein	rheaReaction	rheaReaction_evt	source
http://purl.uniprot.org/uniprot/Q9UXZ0	http://rdf.rhea-db.org/10164	http://purl.obolibrary.org/obo/ECO_0000255	http://purl.uniprot.org/hamap-rule/MF_00318
http://purl.uniprot.org/uniprot/A0A087WNH6	http://rdf.rhea-db.org/21708	http://purl.obolibrary.org/obo/ECO_0000255	http://purl.uniprot.org/hamap-rule/MF_00563
http://purl.uniprot.org/uniprot/A0A078BQP2	http://rdf.rhea-db.org/13665	http://purl.obolibrary.org/obo/ECO_0000250	http://purl.uniprot.org/uniprot/Q19187
http://purl.uniprot.org/uniprot/B0YJ81	http://rdf.rhea-db.org/45812	http://purl.obolibrary.org/obo/ECO_0000269	http://purl.uniprot.org/citations/18554506
http://purl.uniprot.org/uniprot/B0YJ81	http://rdf.rhea-db.org/45812	http://purl.obolibrary.org/obo/ECO_0000269	http://purl.uniprot.org/citations/23933735
http://purl.uniprot.org/uniprot/A3BF39	http://rdf.rhea-db.org/12132	http://purl.obolibrary.org/obo/ECO_0000269	http://purl.uniprot.org/citations/22123790
http://purl.uniprot.org/uniprot/A3BF39	http://rdf.rhea-db.org/20621	http://purl.obolibrary.org/obo/ECO_0000269	http://purl.uniprot.org/citations/22123790
http://purl.uniprot.org/uniprot/A0A024B7W1	http://rdf.rhea-db.org/60860	http://purl.obolibrary.org/obo/ECO_0000255	http://purl.uniprot.org/prosite-prorule/PRU00924
http://purl.uniprot.org/uniprot/A0A024B7W1	http://rdf.rhea-db.org/23680	http://purl.obolibrary.org/obo/ECO_0000250	http://purl.uniprot.org/uniprot/Q32ZE1
http://purl.uniprot.org/uniprot/A0A024B7W1	http://rdf.rhea-db.org/60856	http://purl.obolibrary.org/obo/ECO_0000255	http://purl.uniprot.org/prosite-prorule/PRU00924

[…]

UniProt Team

PIs: Alex Bateman, Alan Bridge, Cathy Wu

Key staff: Cecilia Arighi (Curation), Lionel Breuza (Curation), Elisabeth Coudert (Curation), Hongzhan Huang (Development), Damien Lieberherr (Curation), Michele Magrane (Curation), Maria Martin (Development), Peter McGarvey (Content), Darren Natale (Content), Sandra Orchard (Content), Ivo Pedruzzi (Curation), Sylvain Poux (Curation), Manuela Pruess (Coordination), Shriya Raj (Coordination), Nicole Redaschi (Development)

Content / Curation: Lucila Aimo, Ghislaine Argoud-Puy, Andrea Auchincloss, Kristian Axelsen, Emmanuel Boutet, Emily Bowler, Ramona Britto, Hema Bye-A-Jee, Cristina Casals-Casas, Anne Estreicher, Livia Famiglietti, Marc Feuermann, John S. Garavelli, Penelope Garmiri, George Georghiou, Arnaud Gos, Nadine Gruaz, Emma Hatton-Ellis, Chantal Hulo, Nevila Hyka-Nouspikel, Florence Jungo, Kati Laiho, Philippe Lemercier, Yvonne Lussi, Alistair MacDougall, Patrick Masson, Anne Morgat, Sandrine Pilbout, Lucille Pourcel, Catherine Rivoire, Karen Ross, Christian Sigrist, Elena Speretta, Shyamala Sundaram, Nidhi Tyagi, C. R. Vinayaka, Qinghua Wang, Kate Warner, Lai-Su Yeh, Rossana Zaru

Development: Shadab Ahmed, Emanuele Alpi, Leslie Arminski, Parit Bansal, Delphine Baratin, Teresa Batista Neto, Jerven Bolleman, Borisas Bursteinas, Chuming Chen, Yongxing Chen, Beatrice Cuche, Edouard De Castro, Tunca Dogan, Elisabeth Gasteiger, Sebastien Gehant, Leonardo Gonzales, Alexandr Ignatchenko, Giuseppe Insana, Rizwan Ishtiaq, Vishal Joshi, Dushyanth Jyothi, Arnaud Kerhornou, Thierry Lombardot, Jie Luo, Mahdi Mahmoudy, Andrew Nightingale, Joseph Onwubiko, Monica Pozzato, Sangya Pundir, Guoying Qi, Daniel Rice, Rabie Saidi, Edward Turner, Preethi Vasudev, Yuqi Wang, Xavier Watkins, Hermann Zellner, Jian Zhang

EMBL-EBI

European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK

PIR

Protein Information Resource (PIR), Washington DC and Delaware, USA

SIB

SIB Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland