PROTEIN DATABASE

PROTEIN DATABASE

Today, almost all amino acid sequence information arises from translation of gene sequences.

However the amino acid sequences is not in general inferrable with confidence from the gene sequence because of:

ambiguity in splicing in eukariotes

post-translational modifications

Protein sequence database collects these additional information from the literature and provide suitable annotation.

The first amino acid sequence database was developed by Margaret O. Dayhoff.

From this archive grew the

Protein Information Resource (PIR)

at the National Biomedical Research Foundation of the Georgetown University Medical Center in Washington DC, USA

In 2002 PIR, along with its international partners, EBI (European Bioinformatics Institute) and SIB (Swiss Institute of Bioinformatics), were awarded a grant from NIH to create UniProt.

UniProt is a single worldwide database of protein sequence and function, by unifying the PIR-PSD, Swiss-Prot, and TrEMBL databases.

EMBL-EBI	SIB	PIR
EMBL Outstation European Bioinformatics Institute (EBI)	SIB Swiss Institute of Bioinformatics Centre Medical Universitaire	Protein Information Resource (PIR) Georgetown University Medical Center

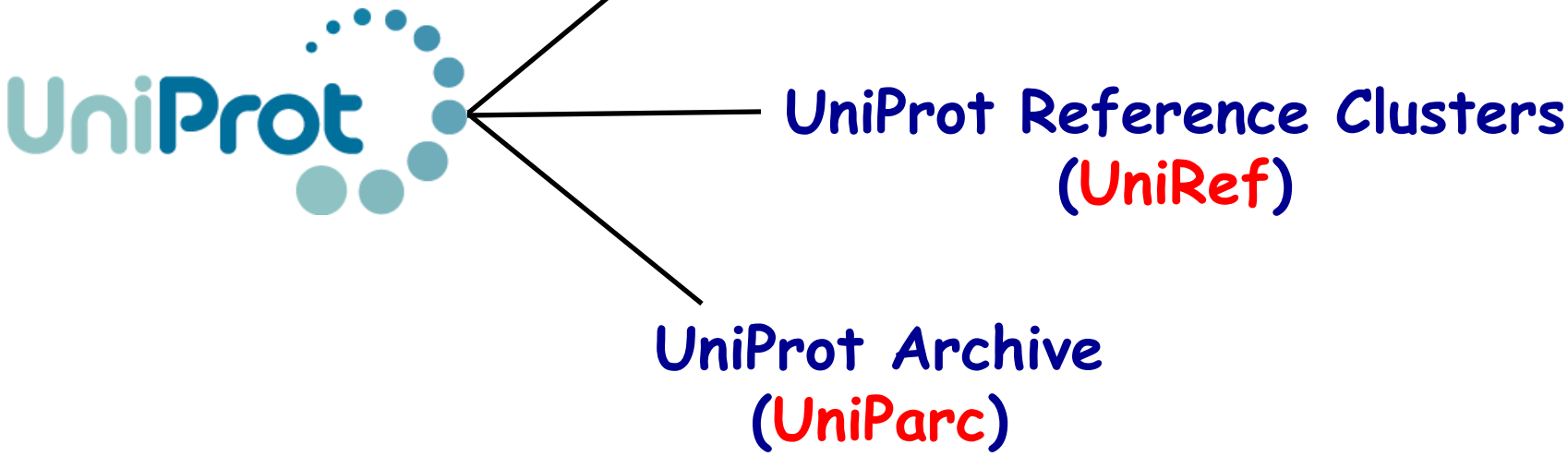
share the database but offer separate information and retrieval tools

The Universal Protein Resource (UniProt)

A comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

http://www.uniprot.org/

The UniProt databases



The UniProt Metagenomic and Environmental Sequences (UniMES) database is a repository specifically developed for metagenomic and environmental data.

UniProt Knowledgebase (UniProtKB) is the central access point for the collection of functional information on proteins, with accurate, consistent and rich annotation.

Each record contains:

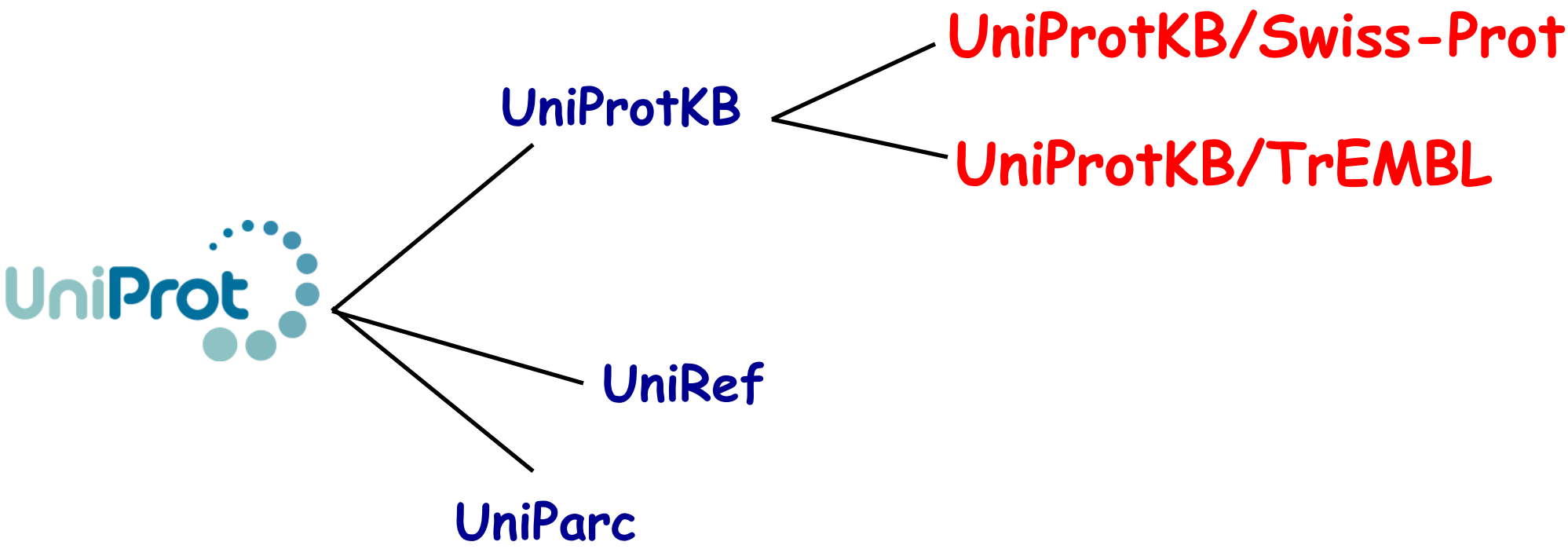
- **core data** (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information);
- **as much annotation information as possible is added.**

The annotation information includes:

- widely accepted biological ontologies
- classifications and cross-references
- clear indications of the quality of annotation in the form of evidence attribution of experimental and computational data.

About 85% of the protein sequences provided by UniProtKB are derived from the translation of the coding sequences (CDS) which have been submitted to the public nucleic acid databases, as well as the related data submitted by the authors.

UniProt Knowledgebase (UniProtKB) consists of two sections:



UniProtKB/Swiss-Prot

This is a high quality manually annotated (reviewed) and non redundant protein sequence database, which brings together experimental results and computed features.

UniProtKB/TrEMBL

This is a computer-annotated (unreviewed) supplement to Swiss-Prot, which strives to gather all protein sequences that are not yet represented in Swiss-Prot.

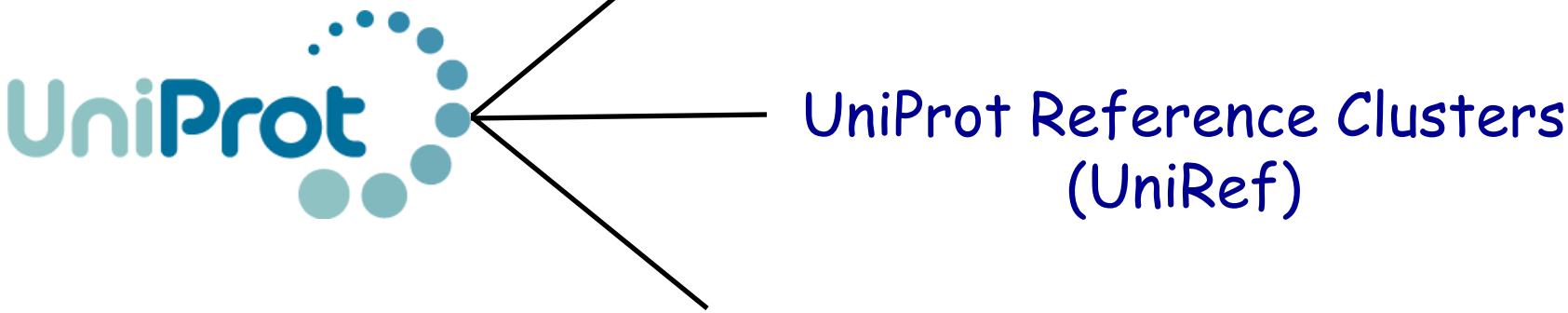
UniProtKB

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.

The UniProt databases



UniProt Archive
(UniParc)

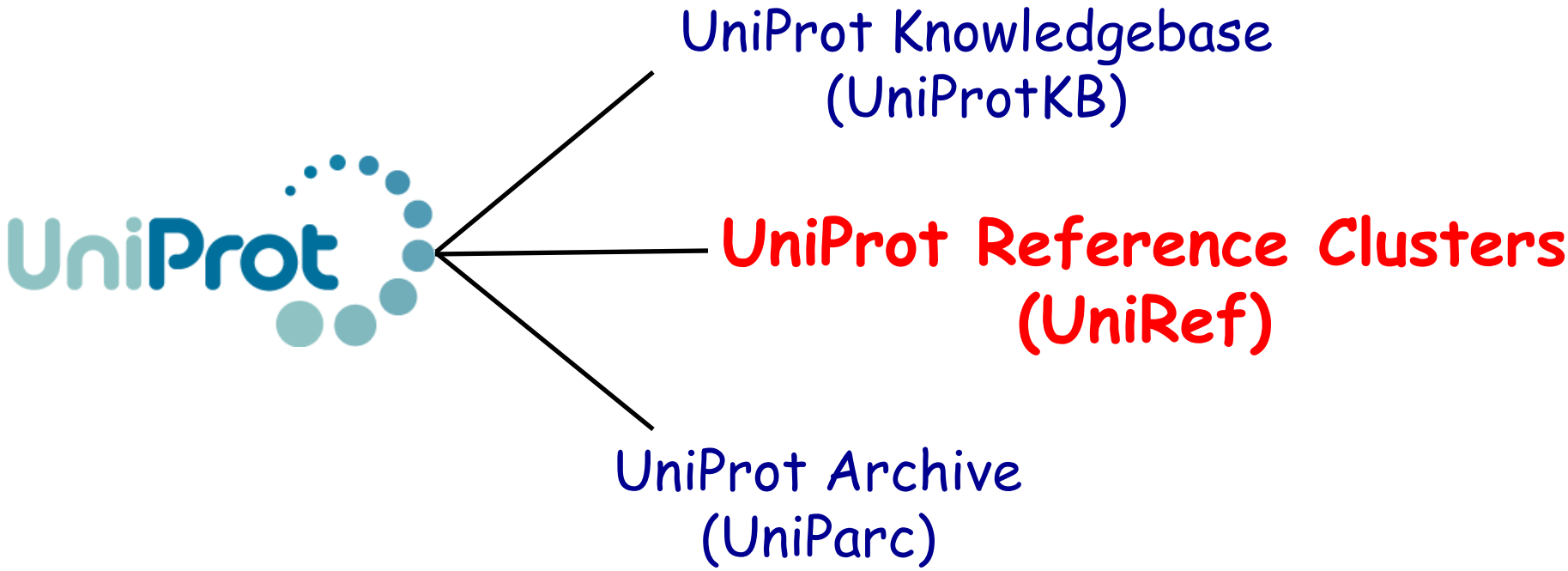
UniProt Archive (UniParc)

It is a comprehensive repository, used to keep track of sequences and their identifiers.

The basic information stored within each UniParc entry is:

- the identifier
- the sequence
- cyclic redundancy check number
- source database(s) with accession and version numbers
- time stamp

The UniProt databases



UniProt Reference Clusters (UniRef)

This databases provide clustered sets of sequences from the UniProtKB and selected UniProt Archive records to obtain complete coverage of sequence space at several resolutions while removing sequence redundancy and reducing the number of sequences.

UniRef database consists of three sub-databases:

UniRef100 database combines identical sequences and sub-fragments with 11 or more residues (from any organism) into a single UniRef entry. The sequences are derived from UniProtKB and UniParc databases.

UniRef90 is build by clustering UniRef100 sequences such that each cluster is composed of sequences that have at least 90% identity.

UniRef50 is build by clustering UniRef100 sequences such that each cluster is composed of sequences that have at least 50% identity.

Such clustering allowed to reduce a size of UniRef100 database of approximately

40% (UniRef90)
65% (UniRef50)

Thus the time needed for similarity searches is significantly reduced.

WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: ★ Swiss-Prot, which is manually annotated and reviewed. ★ TrEMBL, which is automatically annotated and is **not** reviewed. Includes complete and reference proteome sets.
UniRef	Sequence clusters, used to speed up sequence similarity searches.

NEWS

UniProt release 2011_09 - Sep

Reference proteomes in UniProt

› Statistics for UniProtKB: Swiss-Prot · TrEMBL
› Forthcoming changes
› News archives

Follow @uniprot · 71 followers

SITE TOUR

Search in Protein Knowledgebase (UniProtKB)

Query beta-carotene hydroxylase

Search | Advanced | Clear

1 - 25 of 370 results for **beta-carotene** ☒ **AND hydroxylase** ☒ in **UniProtKB** sorted by **score** descending ☒

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway |

Reduce sequence redundancy to 100%, 90% or 50% |

Page 1

Results Customize

› Show only reviewed (12) ★(UniProtKB/Swiss-Prot) or unreviewed (358) ★(UniProtKB/TrEMBL) entries

› Restrict term "beta carotene" to protein name (361)

› Restrict term "hydroxylase" to gene ontology (3), protein name (360)

› Restrict term ""beta carotene"" to pathway

	Accession	Entry name	Status	Protein names	Gene names	Organism
☐	Q9SZZ8	BCH1_ARATH	★	**Beta-carotene 3-hydroxylase 1, chloroplastic**	**BETA-OHASE 1** B1 CHY1 At4g25700 L73G19.80	Arabidopsis thaliana (Mouse-ear cress)
☐	Q9LTG0	BCH2_ARATH	★	**Beta-carotene 3-hydroxylase 2,**	**BETA-OHASE 2** B2 CHY2	Arabidopsis thaliana (Mouse-ear cress)

Q9SZZ8 (BCH1_ARATH) Reviewed, UniProtKB/Swiss-Prot

Last modified September 21, 2011. Version 58. This entry in the past...

Contribute
Send feedback
Read comments (0) or add you

Clusters with 100%, 90%, 50% identity | Third-party data

text xml rdf/xml gff

Names and origin

Protein names	*Recommended name:* **Beta-carotene 3-hydroxylase 1, chloroplastic** Short name=AtB1 EC=1.14.13.129
Gene names	Name: **BETA-OHASE 1** Synonyms: B1, CHY1 Ordered Locus Names: At4g25700 ORF Names: L73G19.80
Organism	**Arabidopsis thaliana (Mouse-ear cress)**
Taxonomic identifier	3702 [NCBI]
Taxonomic lineage	Eukaryota › Viridiplantae › Streptophyta › Embryophyta › Tracheophyta › Spermatophyta › Magnoliophyta › eudicotyledons › eudicotyledons › rosids › malvids › Brassicales › Brassicaceae › Camelineae › Arabidopsis

Organism	**Arabidopsis thaliana (Mouse-ear cress)**
Taxonomic identifier	3702 [NCBI]
Taxonomic lineage	Eukaryota › Viridiplantae › Streptophyta › Embryophyta › Tracheophyta › Spermatophyta › Magnoliophyta › eudicotyledons › eudicotyledons › rosids › malvids › Brassicales › Brassicaceae › Camelineae › Arabidopsis

Protein attributes

Sequence length	310 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.
Protein existence	Evidence at protein level

General annotation (Comments)

Function	Nonheme diiron monooxygenase involved in the biosynthesis of xanthophylls. Specific for beta-ring hydroxylations of beta-carotene. Has also a low activity toward the beta- and epsilon-rings of alpha-carotene. No activity with acyclic carotenoids as lycopene and neurosporene. Uses ferredoxin as an electron donor Probable.
Catalytic activity	Beta-carotene + 2 NADH + 2 O_2 = zeaxanthin + 2 NAD^+ + 2 H_2O.
Subunit structure	Homodimer Probable.
Subcellular location	Plastid › chloroplast membrane; Multi-pass membrane protein Potential.
Tissue specificity	Expressed in leaves, flowers, stems, roots and siliques.

Protein existence

Last modified February 17, 2011

This subsection of the 'Protein attributes' section indicates the type of evidence that supports the existence of the protein. Note that this subsection does not give information on the accuracy or correctness of the sequence(s) displayed. While it gives information on the existence of a protein, it may happen that the sequence slightly differ, especially for sequences derived from gene model predictions from genomic sequences.

In UniProtKB there are 5 types of evidence for the existence of a protein:

- 1. Evidence at protein level
- 2. Evidence at transcript level
- 3. Inferred from homology
- 4. Predicted
- 5. Uncertain

The value **'Evidence at protein level'** indicates that there is clear experimental evidence for the existence of the protein. The criteria include partial or complete Edman sequencing, clear identification by mass spectrometry, X-ray or NMR structure, good quality protein-protein interaction or detection of the protein by antibodies.

The value **'Evidence at transcript level'** indicates that the existence of a protein has not been strictly proven but that expression data (such as existence of cDNA(s), RT-PCR or Northern blots) indicate the existence of a transcript.

The value **'Inferred by homology'** indicates that the existence of a protein is probable because clear orthologs exist in closely related species.

- 2. Evidence at transcript level
- 3. Inferred from homology
- 4. Predicted
- 5. Uncertain

The value **'Evidence at protein level'** indicates that there is clear experimental evidence for the existence of the protein. The criteria include partial or complete Edman sequencing, clear identification by mass spectrometry, X-ray or NMR structure, good quality protein-protein interaction or detection of the protein by antibodies.

The value **'Evidence at transcript level'** indicates that the existence of a protein has not been strictly proven but that expression data (such as existence of cDNA(s), RT-PCR or Northern blots) indicate the existence of a transcript.

The value **'Inferred by homology'** indicates that the existence of a protein is probable because clear orthologs exist in closely related species.

The value **'Predicted'** is used for entries without evidence at protein, transcript, or homology levels.

The value **'Uncertain'** indicates that the existence of the protein is unsure.

Only the highest or most reliable level of supporting evidence for the existence of a protein is displayed for each entry. For example, if the existence of a protein is supported by both the presence of ESTs and direct protein sequencing, the protein is assigned the value 'Evidence at protein level'.

Link to relevant document

Criteria description for protein existence.

Go to:

```
LOCUS       BAI47579                309 aa            linear   PLN 30-MAR-2010
DEFINITION  beta-carotene hydroxylase [Ipomoea obscura].
ACCESSION   BAI47579
VERSION     BAI47579.1  GI:262036876
DBSOURCE    accession AB499057.1
KEYWORDS    .
SOURCE      Ipomoea obscura
  ORGANISM  Ipomoea obscura
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Convolvulaceae; Ipomoeeae; Ipomoea.
REFERENCE   1
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Carotenoid composition and carotenogenic gene expression during
            Ipomoea petal development
  JOURNAL   J. Exp. Bot. 61 (3), 709-719 (2010)
   PUBMED   19933319
REFERENCE   2  (residues 1 to 309)
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2009) Contact:Chihiro Yamamizo National Institute
            of Floricultural Science; Fujimoto 2-1, tsukuba, Ibaraki 305-8519,
            Japan
FEATURES             Location/Qualifiers
     source          1..309
                     /organism="Ipomoea obscura"
                     /db_xref="taxon:89652"
     Protein         1..309
                     /product="beta-carotene hydroxylase"
     Region          56..308
                     /region_name="FA_hydroxylase"
                     /note="Fatty acid hydroxylase superfamily; cl01132"
                     /db_xref="CDD:194046"
     CDS             1..309
                     /gene="CHYB"
                     /coded_by="AB499057.1:1..930"
ORIGIN
        1 mavgisiaas sgnvyncqfs lvrpathsas ppslfspls rrfrssvlss rrkprltvcf
       61 vledeklesg vqiraeeiek aiekqisasr laeklarkrs erstylvaav msslgitsma
      121 vlavyyrfaw qmeggavpyt emfgtfalsv gaavgmefwa rwahralwha slwhmheshh
      181 kpregpfeln dvfaiinavp aiallsygff hkglvpglcf gaglgitvfg maymfvhdgl
      241 vhkrfpvgpi advpyfrrva aahqlhhtdk fngvpyglfl gpkeleevgg lndlevevsr
      301 rikmsstgr
```

Run BLAST

Identify Conserved Domains

Find in this Sequence

Related information

BLink

Related Sequences

CDD Search Results

Conserved Domains (Concise)

Conserved Domains (Full)

Domain Relatives

Encoding mRNA

Full text in PMC

Nucleotide

PubMed

Taxonomy

Recent activity

Turn Off Clear

beta-carotene hydroxylase [Ipomoea obscura]
Protein

beta carotene hydroxylase (996)
Protein

Homo sapiens dystrophin (DMD), transcript variant Dp140c, mRNA
Nucleotide

(human Duchenne muscular dystrophy) AND "Homo sapiens"[porgn] (431)
Nucleotide

human Duchenne muscular dystrophy (989)
Nucleotide

See more...

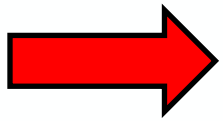
Search for gene and associated protein defective in the Marfan Syndrome

*134797

FIBRILLIN 1; FBN1

Alternative titles; symbols

FIBRILLIN; FBN



UniProt

HGNC Approved Gene Symbol: *FBN1*

Cytogenetic location: *15q21.1* ***Genomic coordinates (GRCh37):*** *15:48,700,502 - 48,937,984* (from NCBI)

Table of Contents – *1347
External Links:
Genome
DNA
Protein
UniProt
HPRD
Gene Info
Clinical Resources
Variation
Animal Models
Cellular Pathways

Gene Phenotype Relationships

Location	Phenotype	Phenotype MIM number
15q21.1	Acromicric dysplasia	102370
	Aortic aneurysm, ascending, and dissection	
	Ectopia lentis, familial	129600
	Geleophysic dysplasia 2	614185
	Marfan syndrome	154700
	MASS syndrome	604308
	Shprintzen-Goldberg syndrome	182212
	Stiff skin syndrome	184900
	Weill-Marchesani syndrome 2, dominant	608328

TEXT

Description

Fibrillin is the major constitutive element of extracellular microfibrils and has widespread distribution in both elastic and nonelastic connective tissue throughout the body. The cDNA was identified in 1991 and was mapped coincident with the locus for Marfan syndrome. Subsequent studies confirmed that mutations in the FBN1 gene are the major cause of Marfan syndrome (MFS; 154700).

Cloning

Search | Blast | Align | Retrieve | ID Mapping

Search in: Protein Knowledgebase (UniProtKB)

Query: FBN1 AND organism:"Homo sapiens [9606]"

Search | Advanced Search » | Clear

Search through gene and organism

WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: Swiss-Prot, which is manually annotated and reviewed. TrEMBL, which is automatically annotated and is **not** reviewed. Includes complete and reference proteome sets.
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations, taxonomy, keywords, subcellular locations and more.

Getting started

- Text search
- Sequence similarity searches (BLAST)
- Sequence alignments
- Batch retrieval
- Database identifier mapping (ID Mapping)

UniProt

NEWS

› Statistics for UniProtKB: Swiss-Prot · TrEMBL

› Forthcoming changes

› News archives

Follow @uniprot · 71 followers

SITE TOUR

Learn how to make best use of the tools and data on this site.

PROTEIN SPOTLIGHT

life's tremors
September 2011

Destruction is sometimes necessary for life to continue. It may sound paradoxical but examples are many. Our body shreds the food we eat to use the parts to feed itself. Certain cells commit suicide when they are of no use anymore…

Search in: Protein Knowledgebase (UniProtKB)

Query: FBN1 AND organism:"Homo sapiens [9606]"

Search | Advanced Search » | Clear

12 results for **FBN1** ⊠ **AND organism:"Homo sapiens (Human) [9606]"** ⊠ in **UniProtKB** sorted by **score** descending ⊠

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% |

Results Customize

› Show only reviewed (5) (UniProtKB/Swiss-Prot) or unreviewed (7) (UniProtKB/TrEMBL) entries

› Restrict term "fbn1" to gene name (6)

› Show only entries from a complete proteome set (8)

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
P35555	FBN1_HUMAN	★	**Fibrillin-1**	**FBN1** FBN	Homo sapiens (Human)	
Q75N89	Q75N89_HUMAN	☆	**Fibrillin 1**	**FBN1**	Homo sapiens (Human)	
Q07092	COGA1_HUMAN	★	**Collagen alpha-1(XVI) chain**	**COL16A1** FP1572	Homo sapiens (Human)	
Q14766	LTBP1_HUMAN	★	**Latent-transforming growth factor beta-bindin...**	**LTBP1**	Homo sapiens (Human)	
F5H2N7	F5H2N7_HUMAN	☆	**Uncharacterized protein**	**FBN1**	Homo sapiens (Human)	
D2JYH6	D2JYH6_HUMAN	☆	**Fibrillin 1**	**FBN1**	Homo sapiens (Human)	
Q75N88	Q75N88_HUMAN	☆	**Fibrillin 1**	**FBN1**	Homo sapiens (Human)	
P36897	TGFR1_HUMAN	★	**TGF-beta receptor type-1**	**TGFBR1** ALK5 SKR4	Homo sapiens (Human)	
Q8N2S1	LTBP4_HUMAN	★	**Latent-transforming growth factor beta-bindin...**	**LTBP4**	Homo sapiens (Human)	
F8W7L2	F8W7L2_HUMAN	☆	**Uncharacterized protein**	**FBN1**	Homo sapiens (Human)	
Q9NP01	Q9NP01_HUMAN	☆	**Fibrillin 15**		Homo sapiens (Human)	
Q59HB9	Q59HB9_HUMAN	☆	**Fibrillin 1 variant**		Homo sapiens (Human)	

P35555 (FBN1_HUMAN) Reviewed, UniProtKB/Swiss-Prot

Last modified September 21, 2011. Version 145. This entry in the past...

Names and origin

Protein names	*Recommended name:* **Fibrillin-1**
Gene names	Name: **FBN1** Synonyms:FBN
Organism	Homo sapiens (Human)
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo

Protein attributes

Sequence length	2871 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.
Protein existence	Evidence at protein level

General annotation (Comments)

Function	Fibrillins are structural components of 10-12 nm extracellular calcium-binding microfibrils, which occur either in association with elastin or in elastin-free bundles. Fibrillin-1-containing microfibrils provide long-term force b support. Regulates osteoblast maturation by controlling TGF-beta bioavailability and calibrating TGF-beta and BMP levels, respectively By similarity. Ref.22
Subunit structure	Interacts with COL16A1. Interacts with integrin alpha-V/beta-3. Ref.11 Ref.22
Subcellular location	Secreted › extracellular space › extracellular matrix.
Post-translational modification	Forms intermolecular disulfide bonds either with other fibrillin-1 molecules or with other components of the microfibrils.
Involvement in disease	Defects in FBN1 are a cause of Marfan syndrome (MFS) [MIM:154700]. MFS is an autosomal dominant disorder that affects the skeletal, ocular, and cardiovascular systems. A wide variety of skeletal abnormalities occu including scoliosis, chest wall deformity, tall stature, abnormal joint mobility. Ectopia lentis occurs in up to about 80% of MFS patients and is almost always bilateral. The leading cause of premature death in MFS patients dilation of the aortic root and ascending aorta, causing aortic incompetence and dissection. Note=The majority of the more than 600 mutations in FBN1 currently known are point mutations, the rest are frameshifts and s mutations. Marfan syndrome has been suggested in at least 2 historical figures, Abraham Lincoln and Paganini. Ref.2 Ref.24 Ref.25 Ref.26 Ref.27 Ref.28 Ref.29 Ref.30 Ref.32 Ref.33 Ref.34 Ref.35 Ref.36 Ref.38 Ref.42 Ref.43 Ref.45 Ref.48 Ref.49 Ref.50 Ref.51 Ref.52 Ref.53 Ref.54 Ref.55 Ref.57 Ref.58 Ref.59 Ref.60 Ref.61 Defects in FBN1 are a cause of isolated ectopia lentis (EL) [MIM:129600]. The symptoms of this autosomal dominant fibrillinopathy overlap with those of Marfan syndrome, with the exclusion of the skeletal and cardiova manifestations. Ref.31 Ref.53 Ref.54 Ref.55 Ref.57

	Disulfide bond	2674 ↔ 2686		By similarity		

Natural variations

	Natural variant	20	1	Y → C in MFS. Ref.59		VAR_023859
	Natural variant	27	1	A → T. [dbSNP:rs25397]		VAR_014663
	Natural variant	62	1	R → C in MFS; also in a patient with ectopia lentis and retinal detachment. [dbSNP:rs25403] Ref.54		VAR_017967
	Natural variant	89	1	C → F in MFS. Ref.53		VAR_017968
	Natural variant	111	1	C → R in MFS. Ref.45		VAR_002276
	Natural variant	114	1	R → C in MFS. Ref.55		VAR_017969
	Natural variant	115	1	S → C in EL. Ref.54		VAR_017970
	Natural variant	122	1	R → C in MFS. Ref.33 Ref.49 Ref.53		VAR_002277
	Natural variant	123	1	C → Y in MFS. Ref.59		VAR_023860
	Natural variant	129	1	C → Y in MFS; severe neonatal. Ref.38		VAR_002278
	Natural variant	133	1	H → Q. [dbSNP:rs363850]		VAR_055723
	Natural variant	154	1	C → S in MFS. Ref.57		VAR_017971
	Natural variant	166	1	C → F in MFS. Ref.38		VAR_002279
	Natural variant	166	1	C → S in MFS. Ref.57		VAR_002280
	Natural variant	177	1	C → R in MFS. [dbSNP:rs363853] Ref.59		VAR_023861
	Natural variant	217	1	W → G in MFS. Ref.30 Ref.36		VAR_002281
	Natural variant	224	1	C → R in MFS. Ref.59		VAR_023862
	Natural variant	240	1	R → C in MFS and EL. Ref.53 Ref.55 Ref.57		VAR_017972
	Natural variant	329	1	I → T. [dbSNP:rs12324002]		VAR_055724
	Natural variant	363	1	G → S. [dbSNP:rs363855]		VAR_055725
	Natural variant	366	1	W → C in MFS. Ref.53		VAR_017973
	Natural variant	439	1	R → G in MFS. Ref.59		VAR_023863
	Natural variant	472	1	C → Y. [dbSNP:rs4775765] Ref.1 Ref.4 Ref.5 Ref.6		VAR_058090
	Natural variant	476	1	C → G in MFS.		VAR_002282
	Natural variant	490	1	D → Y in MFS.		VAR_002283
	Natural variant	504	1	C → F in MFS. Ref.52		VAR_010776
	Natural variant	507	1	Missing in MFS.		VAR_023864
	Natural variant	541	1	C → Y in MFS. Ref.60		VAR_023865
	Natural variant	545	1	R → C in MFS. Ref.45 Ref.53		VAR_002284
	Natural variant	548	1	N → I in MFS. Ref.27		VAR_002285
	Natural variant	560	1	G → S in MFS. Ref.53		VAR_017974
	Natural variant	570	1	C → Y in MFS. Ref.53		VAR_017975
	Natural variant	587	1	C → Y in MFS. Ref.43 Ref.54		VAR_002286

Database	Threshold	Matrix	Filtering	Gapped	Hits
UniProtKB	10	Auto	None	yes	250

P35555[129], Fibrillin-1, Homo sapiens

```
        10         20         30         40         50         60
MRRGRLLEIA LGFTVLLASY TSHGADANLE AGNVKETRAS RAKRRGGGGH DALKGPNVCG

        70         80         90        100        110        120
SRYNAYCCPG WKTLPGGNQC IVPICRHSCG DGFCSRPNMC TCPSGQIAPS CGSRSIQHCN

       130        140        150        160        170        180
IRCMNGGSCS DDHCLCQKGY IGTHCGQPVC ESGCLNGGRC VAPNRCACTY GFTGPQCERD

       190        200        210        220        230        240
YRTGPCFTVI SNQMCQGGLS GIVCTKTLCC ATVGRAWGHP CEMCPAQPHP CRRGFIPNIR

       250        260        270        280        290        300
TGACQDVDEC QAIPGLCQGG NCINTVGSFE CKCPAGHKLN EVSQKCEDID ECSTIPGICE

       310        320        330        340        350        360
GGECTNTVGS YFCKCPPGFY TSPDGTRCID VRPGYCYTAL TNGRCSNQLP QSITKMQCCC

       370        380        390        400        410        420
DAGRCWSPGV TVAPEMCPIR ATEDFNKLCS VPMVIPGRPE YPPPPLGPIP PVLPVPPGFP

       430        440        450        460        470        480
PGPQIPVPRP PVEYLYPSRE PPRVLPVNVT DYCQLVRYLC QNGRCIPTPG SCRCECNKGF

       490        500        510        520        530        540
QLDLRGECID VDECEKNPCA GGECINNQGS YTCQCRAGYQ STLTRTECRD IDECLQNGRI

       550        560        570        580        590        600
CNNGRCINTD GSFHCVCNAG FHVTRDGKNC EDMDECSIRN MCLNGMCINE DGSFKCICKP

       610        620        630        640        650        660
GFQLASDGRY CKDINECETP GICMNGRCVN TDGSYRCECF PGLAVGLDGR VCVDTHMRST

       670        680        690        700        710        720
CYGGYKRGQC IKPLFGAVTK SECCCASTEY AFGEPCQPCP AQNSAEYQAL CSSGPGMTSA

       730        740        750        760        770        780
GSDINECALD PDICPNGICE NLRGTYKCIC NSGYEVDSTG KNCVDINECV LNSLLCDNGQ

       790        800        810        820        830        840
CRNTPGSFVC TCPKGFIYKP DLKTCEDIDE CESSPCINGV CKNSPGSFIC ECSSESTLDP

       850        860        870        880        890        900
TKTICIETIK GTCWQTVIDG RCEININGAT LKSQCCSSLG AAWGSPCTLC QVDPICGKGY

       910        920        930        940        950        960
SRIKGTQCED IDECEVFPGV CKNGLCVNTR GSFKCQCPSG MTLDATGRIC LDIRLETCFL

       970        980        990       1000       1010       1020
RYEDEECTLP IAGRHRMDAC CCSVGAAWGT EECEECPMRN TPEYEELCPR GPGFATKEIT

      1030       1040       1050       1060       1070       1080
NGKPFFKDIN ECKMIPSLCT HGKCRNTIGS FKCRCDSGFA LDSEERNCTD IDECRISPDL

      1090       1100       1110       1120       1130       1140
CGRGQCVNTP GDFECKCDEG YESGFMMMKN CMDIDECQRD PLLCRGGVCH NTEGSYRCEC
```

Swiss-Prot variant: VAR_002278 in UniProtKB/Swiss-Prot **P35555**

General Information · Information on the variant · Sequence features · Structural features · References for the variant · Cross-references for the variant

Top

General information

UniProtKB/Swiss-Prot	**FBN1_HUMAN (P35555)**
Gene symbol(s)	Official: FBN1 Synonym(s): FBN
Chromosomal location	15q21.1
Protein name	Fibrillin-1
Length of the protein	2871

Top

Information on the variant

FTId	VAR_002278
Amino acid position of the variant	129
Residue change	From **Cysteine** (C) to **Tyrosine** (Y), C129Y, p.Cys129Tyr
Physico-chemical property	Change from medium size and polar (C) to large size and aromatic (Y)
BLOSUM score	-2
Status	Disease
Disease	**Marfan syndrome (MFS)** Defects in FBN1 are a cause of Marfan syndrome (MFS) [MIM:154700]. MFS is an autosomal dominant disorder that affects the skeletal, ocular, and cardiovascular systems. A wid abnormalities occurs with MFS, including scoliosis, chest wall deformity, tall stature, abnormal joint mobility. Ectopia lentis occurs in up to about 80% of MFS patients and is almost cause of premature death in MFS patients is progressive dilation of the aortic root and ascending aorta, causing aortic incompetence and dissection. Note=The majority of the more currently known are point mutations, the rest are frameshifts and splice site mutations. Marfan syndrome has been suggested in at least 2 historical figures, Abraham Lincoln and P
Comment	Severe neonatal

Disclaimer: Variants classification is intended for research purposes only, not for clinical and diagnostic use. The label disease variant is assigned according to literature reports on probable diseas on theoretical reasons. Therefore this label must not be considered as a definitive proof for the pathogenic role of a variant.

T

38]"Fifteen novel FBN1 mutations causing Marfan syndrome det
heteroduplex analysis of genomic amplicons."
Nijbroek G., Sood S., McIntosh I., Francomano C.A., Bull E., F
L., Ramirez F., Pyeritz R.E., Dietz H.C.
Am. J. Hum. Genet. 57:8-21(1995) [PubMed: 7611299] [Abstra
Cited for: VARIANTS MFS TYR-129; PHE-166; CYS-746; ARG
ARG-1013; LYS-1073; SER-1382 AND ARG-1928.

Cited for: VARIANTS MFS THR-705; TYR-711; GLY-1055 AND TYR-1153.

[43] **"A novel de novo mutation in exon 14 of the fibrillin-1 gene associated with delayed secretion of fibrillin in a patient with a mild Marfan phenotype."**
Booms P., Withers A.P., Boxer M., Kaufmann U.C., Hagemeier C., Vetter U., Robinson P.N.
Hum. Genet. 100:195-200(1997) [PubMed: 9254848] [Abstract]
Cited for: VARIANT MFS TYR-587.

[44] **"The pathogenicity of the Pro1148Ala substitution in the FBN1 gene: causing or predisposing to Marfan syndrome and aortic aneurysm, or clinically innocent?"**
Schrijver I., Liu W., Francke U.
Hum. Genet. 99:607-611(1997) [PubMed: 9150726] [Abstract]
Cited for: VARIANT ALA-1148.

[45] **"Mutation screening of all 65 exons of the fibrillin-1 gene in 60 patients with Marfan syndrome: report of 12 novel mutations."**
Hayward C., Porteous M.E.M., Brock D.J.H.
Hum. Mutat. 10:280-289(1997) [PubMed: 9338581] [Abstract]
Cited for: VARIANTS MFS ARG-111; CYS-545; CYS-627; GLY-750; ARG-1074; HIS-1170; TRP-1171; LYS-1173; TYR-1404; GLY-1610; LYS-1893; TRP-2099; TYR-2111; ARG-2258; TRP-2282 AND ARG-2489.

[46] **"P1148A in fibrillin-1 is not a mutation leading to Shprintzen-Goldberg syndrome."**
Watanabe Y., Yano S., Koga Y., Yukizane S., Nishiyori A., Yoshino M., Kato H.
Hum. Mutat. 10:326-327(1997) [PubMed: 9338588] [Abstract]
Cited for: VARIANT ALA-1148.

[47] **"P1148A in fibrillin-1 is not a mutation anymore."**
Wang M., Mathews K.R., Imaizumi K., Beiraghi S., Blumberg B., Scheuner M., Graham J.M. Jr., Godfrey M.
Nat. Genet. 15:12-12(1997) [PubMed: 8988160] [Abstract]
Cited for: VARIANT ALA-1148.

[48] **"Multiple molecular mechanisms underlying subdiagnostic variants of Marfan syndrome."**
Montgomery R.A., Geraghty M.T., Bull E., Gelb B.D., Johnson M., McIntosh I., Francomano C.A., Dietz H.C.
Am. J. Hum. Genet. 63:1703-1711(1998) [PubMed: 9837823] [Abstract]
Cited for: VARIANT MFS ARG-1265.

[49] **"Correlation of a recurrent FBN1 mutation (R122C) with an atypical familial Marfan syndrome phenotype."**
Black C., Withers A.P., Gray J.R., Bridges A.B., Craig A., Baty D.U., Boxer M.
Hum. Mutat. Suppl. 1:S198-S200(1998) [PubMed: 9452085] [Abstract]

Search in: Protein Knowledgebase (UniProtKB)

Query: marfan AND syndrome

Marfan syndrome

10 results for **marfan** ☒ **AND syndrome** ☒ in **UniProtKB** sorted by **score** descending ☒

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% |

Results Customize

› Show only reviewed (7) ★ (UniProtKB/Swiss-Prot) or unreviewed (3) ★ (UniProtKB/TrEMBL) entries

› Quote terms: "marfan syndrome"

› Restrict term "syndrome" to keyword (2)

	Accession	Entry name	Status	Protein names	Gene names	Organism
☐	P35555	FBN1_HUMAN	★	**Fibrillin-1**	**FBN1** FBN	Homo sapiens (Human)
☐	P36897	TGFR1_HUMAN	★	**TGF-beta receptor type-1**	**TGFBR1** ALK5 SKR4	Homo sapiens (Human)
☐	P37173	TGFR2_HUMAN	★	**TGF-beta receptor type-2**	**TGFBR2**	Homo sapiens (Human)
☐	P35556	FBN2_HUMAN	★	**Fibrillin-2**	**FBN2**	Homo sapiens (Human)
☐	P08123	CO1A2_HUMAN	★	**Collagen alpha-2(I) chain**	**COL1A2**	Homo sapiens (Human)
☐	P35520	CBS_HUMAN	★	**Cystathionine beta-synthase**	**CBS**	Homo sapiens (Human)
☐	Q75N90	FBN3_HUMAN	★	**Fibrillin-3**	**FBN3** KIAA1776	Homo sapiens (Human)
☐	Q75N89	Q75N89_HUMAN	☆	**Fibrillin 1**	**FBN1**	Homo sapiens (Human)
☐	Q9NP01	Q9NP01_HUMAN	☆	**Fibrillin 15**		Homo sapiens (Human)
☐	Q75N88	Q75N88_HUMAN	☆	**Fibrillin 1**	**FBN1**	Homo sapiens (Human)

Cellular component	microfibril Traceable author statement. Source: BHF-UCL
Molecular function	calcium ion binding Inferred from electronic annotation. Source: InterPro extracellular matrix structural constituent Traceable author statement. Source: ProtInc

Complete GO annotation...

Alternative products

This entry describes 2 isoforms produced by **alternative splicing**. [Align] [Select]

Isoform 1 (identifier: **P35556-1**)

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.

Isoform 2 (identifier: **P35556-2**)

The sequence of this isoform differs from the canonical sequence as follows:
113-145: Missing.
1491-1506: DIDECSFQNICVFGTC → GGSPGFQLIFKLDQPQ
1507-2912: Missing.

Sequence annotation (Features)

Feature key	Position(s)	Length	Description	Graphical view
Molecule processing				
Signal peptide	1 – 28	28	Potential	
Chain	29 – 2912	2884	Fibrillin-2	
Regions				
Domain	111 – 142	32	EGF-like 1	
Domain	145 – 176	32	EGF-like 2	
Domain	176 – 208	33	EGF-like 3	
Domain	214 – 266	53	TB 1	
Domain	276 – 317	42	EGF-like 4; calcium-binding	

Sequence Clusters UniRef

Search | Blast | Align | Retrieve | ID

Search in: Sequence Clusters (UniRef)
Query: FBN1

14 results for **FBN1** in **UniRef** sorted by **score** descending

Browse by taxonomy | Map to UniProtKB or UniParc

Page 1

Results Customize

› Show only clusters with an identity of 50%, 90%, 100%
› Restrict term "fbn1" to cluster name (1)
› Repeat search in UniProtKB (60)

UniRef cluster name	Status	UniRef cluster name	Size	UniRef cluster member(s)	Organisms	Length	Identity
UniRef50_P35556	★	Cluster: Fibrillin-2	203	P35556 UPI00017C2F73 UPI0002264C52 UPI00005A2410 F7HHZ8 UPI0001D3647C F7BMQ7 G1PBK9 UPI0001D57536 +193	Homo sapiens (Human) ? Bos taurus (Bovine) Canis familiaris (Dog) (Canis lupus familiaris) Callithrix jacchus (White-tufted-ear marmoset) Equus caballus (Horse) Myotis lucifugus (Little brown bat) Oryctolagus cuniculus (Rabbit) Mus musculus (Mouse) +23	2,912	50
UniRef90_P35555	★	Cluster: Fibrillin-1	38	P35555 Q75N89 Q75N88 UPI00020AC550 F1MZC8 F7CKI7 A2AQ53 F7GUT7 UPI0001D37633 +28	Homo sapiens (Human) ? Bos taurus (Bovine) Equus caballus (Horse) Mus musculus (Mouse) Callithrix jacchus (White-tufted-ear marmoset) Rattus norvegicus (Rat) Macaca mulatta (Rhesus macaque) Sus scrofa (Pig) +4	2,871	90
UniRef100_Q99L19	☆	Cluster: Fbn1 protein (Fragment)	2	Q99L19 Q7TMG6	Mus musculus (Mouse)	272	100
UniRef90_Q59HB9	☆	Cluster: Fibrillin 1 variant (Fragment)	14	Q59HB9 O88840 Q91WZ4 Q63956 Q3TRZ5 Q3TNW1 Q60784 F1PIA3 F7DVX9 +4	Homo sapiens (Human) Mus musculus (Mouse) Mus sp. Canis familiaris (Dog) (Canis lupus familiaris) Callithrix jacchus (White-tufted-ear marmoset) Gallus gallus (Chicken)	830	90
UniRef50_A8WM46	☆	Cluster: CBR-FBN-1 protein	6	A8WM46 A2G8J0 G0PN10	Caenorhabditis briggsae Trichomonas vaginalis Caenorhabditis brenneri (Nematode worm)	3,928	50

The UniProt Consortium. 2010. The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Research.Vol. 38:D142–D148