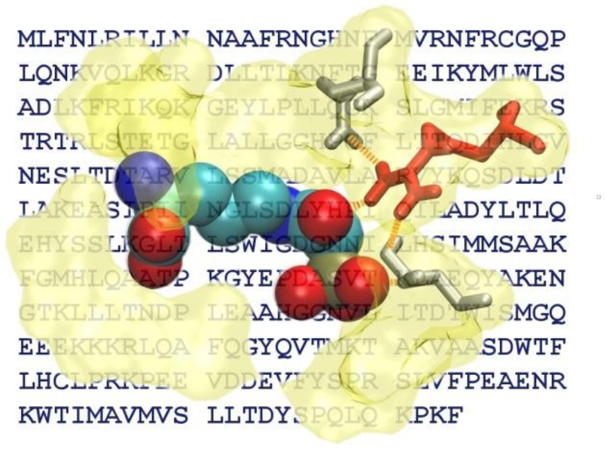
UGC SPONSORED NATIONAL WORKSHOP ON

BIOINFORMATICS AND GENOME ANALYSIS

FOR COLLEGE TEACHERS



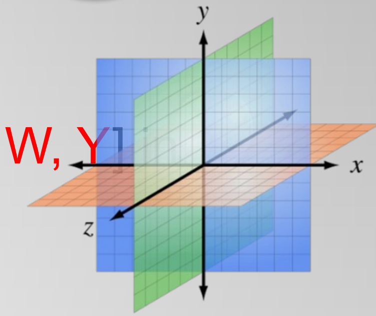
LECTURE TOPIC: **PROTEIN DATABASE**

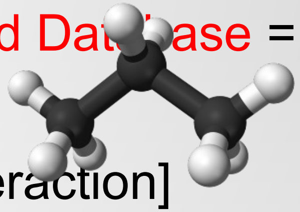
T. ASHOK KUMAR
HEAD, DEPARTMENT OF BIOINFORMATICS
NOORUL ISLAM COLLEGE OF ARTS AND SCIENCE
KUMARACOIL, THUCKALAY - 629180

TOPICS COVERED

- **Protein Terms & Definitions** – Computational biology aspect of protein
- **ExPASy** – SIB Bioinformatics Resource Portal (http://www.expasy.org)
- **UniProt/Swiss-Prot** – A comprehensive, non-redundant, expert manually annotated protein sequence database (http://www.uniprot.org/)
- **NBRF/PIR** – A comprehensive, non-redundant, expertly manually annotated, fully classified and extensively cross-referenced protein sequence database (http://pir.georgetown.edu/)
- **PDB** – A single worldwide repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids (http://rcsb.org/pdb)
- **SCOP** – Knowledge-based expert analysis and classification of proteins that are structurally characterized and deposited in the Protein Data Bank (http://scop2.mrc-lmb.cam.ac.uk/)
- **CATH** – A hierarchical domain classification of protein structures in the Protein Data Bank (http://www.cathdb.info/)
- **MOTIF** – Finds sequence motifs in a query sequence, also provides functional and genomic information of the found motifs using DBGET and LinkDB as the hyperlinked annotations (http://www.genome.jp/tools/motif/)
- **Pfam** – Database of protein HMM profiles that define domain families (http://pfam.xfam.org/)
- **PROSITE** – Database of protein motifs expressed as patterns or profiles (http://prosite.expasy.org/)

PROTEIN TERMS & DEFINITIONS





- Protein Sequence – 20 a.a. characters [A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y] sequence
- Protein Structure – 3D of atomic co-ordinates [x-axis, y-axis, z-axis]
- Types of Biological Databases – [Raw Database = Plain text, Object-oriented Database = Table (Records), Relational Database = Table of tables]
- 3D Atom Model – [Sphere = Atom, Cylinder = Bond, Dotted Line = Bond Interaction]
- Sequence Alignment – [Match = Similar Character, Mismatch = Dissimilar Character, Gap = No Substitute Character, Word = Sub-string, Sequence = Super-string, Score = Rating, Identity = Similar in function]
- Motif – Short, conserved sequence associated with a distinct function.
- Domain – Evolutionarily conserved sequence region that corresponds to a structurally independent 3D unit associated with a particular functional role. It is usually much larger than a motif.
- Pattern – Sequence with symbol representation for a expression. Example: N{P}[ST]{P}
- Regular Expression – Representation format for a sequence motif, which includes positional information for conserved and partly conserved residues. Similar to Pattern, but applies to MSA.
- Profile – Scoring matrix that represents a multiple sequence alignment. It contains probability or

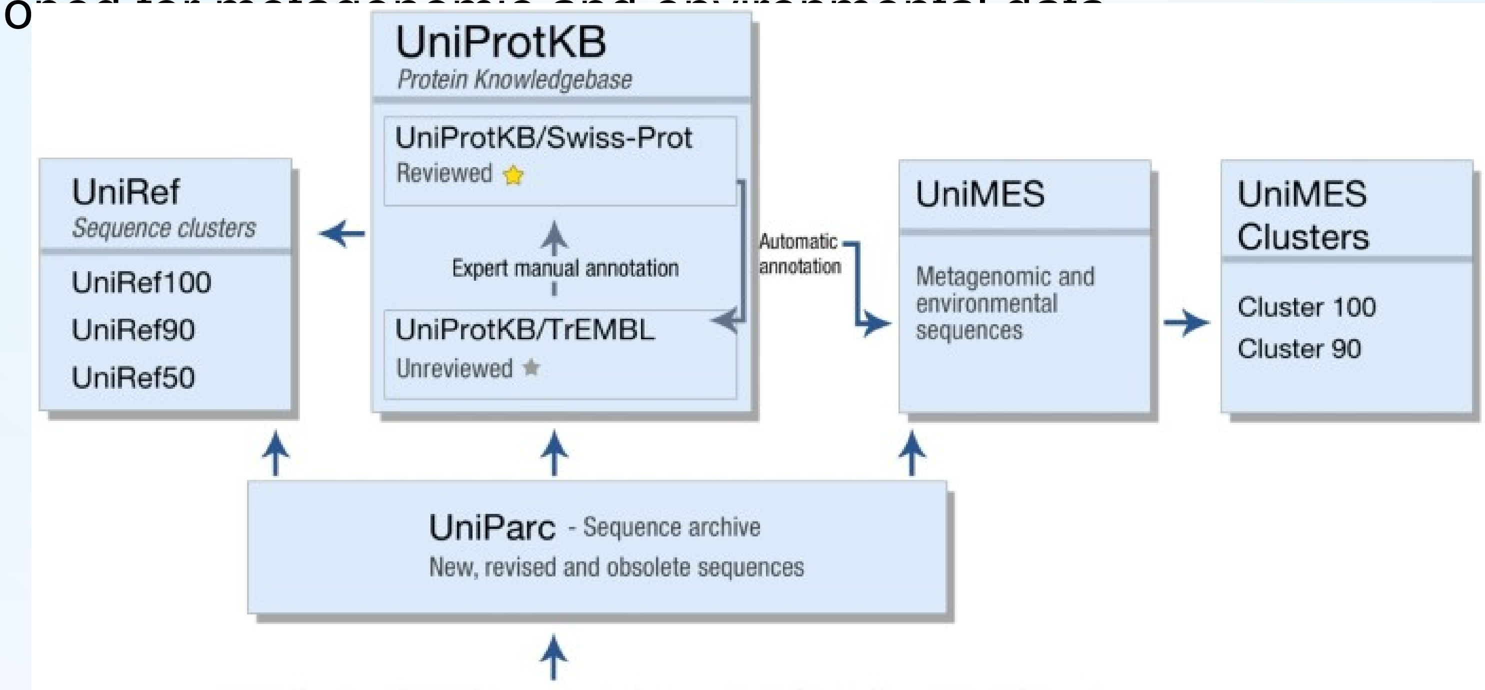
EXPASY

- ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) is a bioinformatics resource portal operated by the Swiss Institute of Bioinformatics (SIB).
- ExPASy was the first website of the life sciences.
- Extensible and integrative portal for accessing many scientific resources, databases and software tools.
- Wide range of resources in many different domains, such as proteomics, genomics, phylogeny/evolution, systems biology, population genetics, transcriptomics, etc.
- Proteomics server to analyze protein sequences and structures and 2D Page gel electrophoresis.
- Databases, online and offline software tools are hosted by different groups of the SIB and partner institutions.   --- CFSSP
- ExPASy references the pr e knowledgebase, UniProtKB/Swiss-Prot, and its computer annotated suppl tKB/Trembl.

SIB ExPASy Bioinformatics Resource Portal

ARCHITECTURE OF UNIPROT/SWISS-PROT

- Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data
- The UniProt databases are the UniProt Knowledgebase (**UniProtKB**), the UniProt Reference Clusters (**UniRef**), and the UniProt Archive (**UniParc**)
- UniProt Metagenomic and Environmental Sequences (UniMES) database is a repository specifically developed for metagenomic and environmental data



BACKGROUND OF UNIPROT/SWISS-PROT

- UniProt is a collaboration between the European Bioinformatics Institute (**EMBL-EBI**), the Swiss Institute of Bioinformatics (**SIB**) and the Protein Information Resource (**PIR**)
- EMBL-EBI and SIB together used to produce **Swiss-Prot** and **TrEMBL**, while PIR produced the Protein Sequence Database (**PIR-PSD**)
- Translated EMBL Nucleotide Sequence Data Library (**TrEMBL**) was originally created because sequence data was being generated at a pace that exceeded Swiss-Prot's ability to keep up
- PIR maintained the PIR-PSD and related databases, including iProClass, a database of protein sequences and curated families

http://www.uniprot.org/

Search in: Protein Knowledgebase (UniProtKB)

Query: serpin

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, cross-referenced databases and more.

Getting started

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

TRY THE NEW UNIPROT WEBSITE

NEWS

UniProt release 2014_07 - Jul 9, 2014

Lark or owl? PER3 is the answer | Cross-references to CCDS and GeneReviews | UniParc cross-references with protein and gene names

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

Follow @uniprot 1,018 followers

SITE TOUR

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

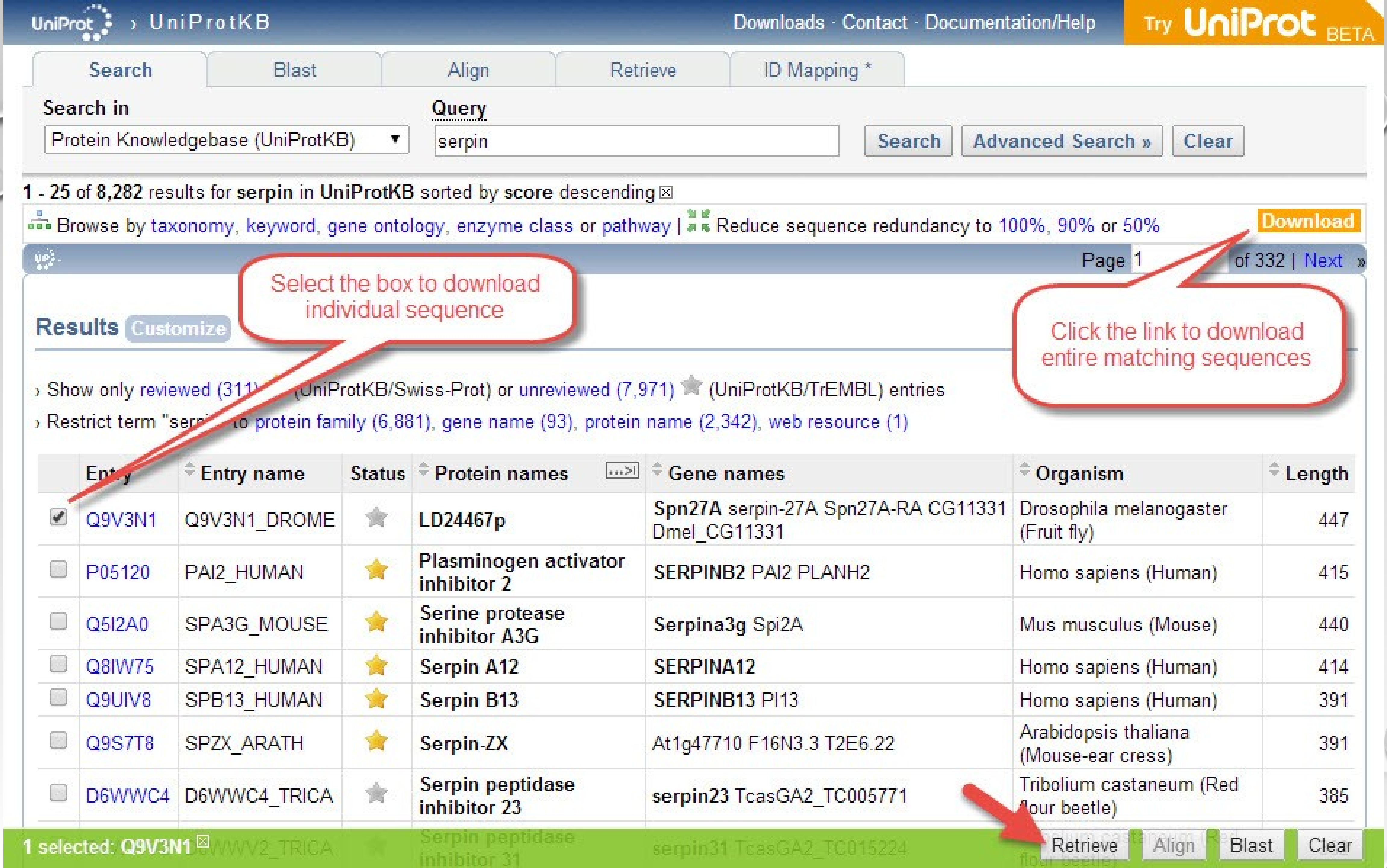
PROTEIN SPOTLIGHT

it did it its way

July 2014

Temperatures can get cold. And living organisms have to find ways of keeping themselves warm. Humans use clothes. Polar bears grow fur. Whales are lined with blubber. And many animals avoid the cold by migrating to warmer parts of the planet…

EMBL-EBI PIR SIB



Query: serpin

1 - 25 of **8,282** results for **serpin** 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 of 332 | Next »

Select the box to download individual sequence

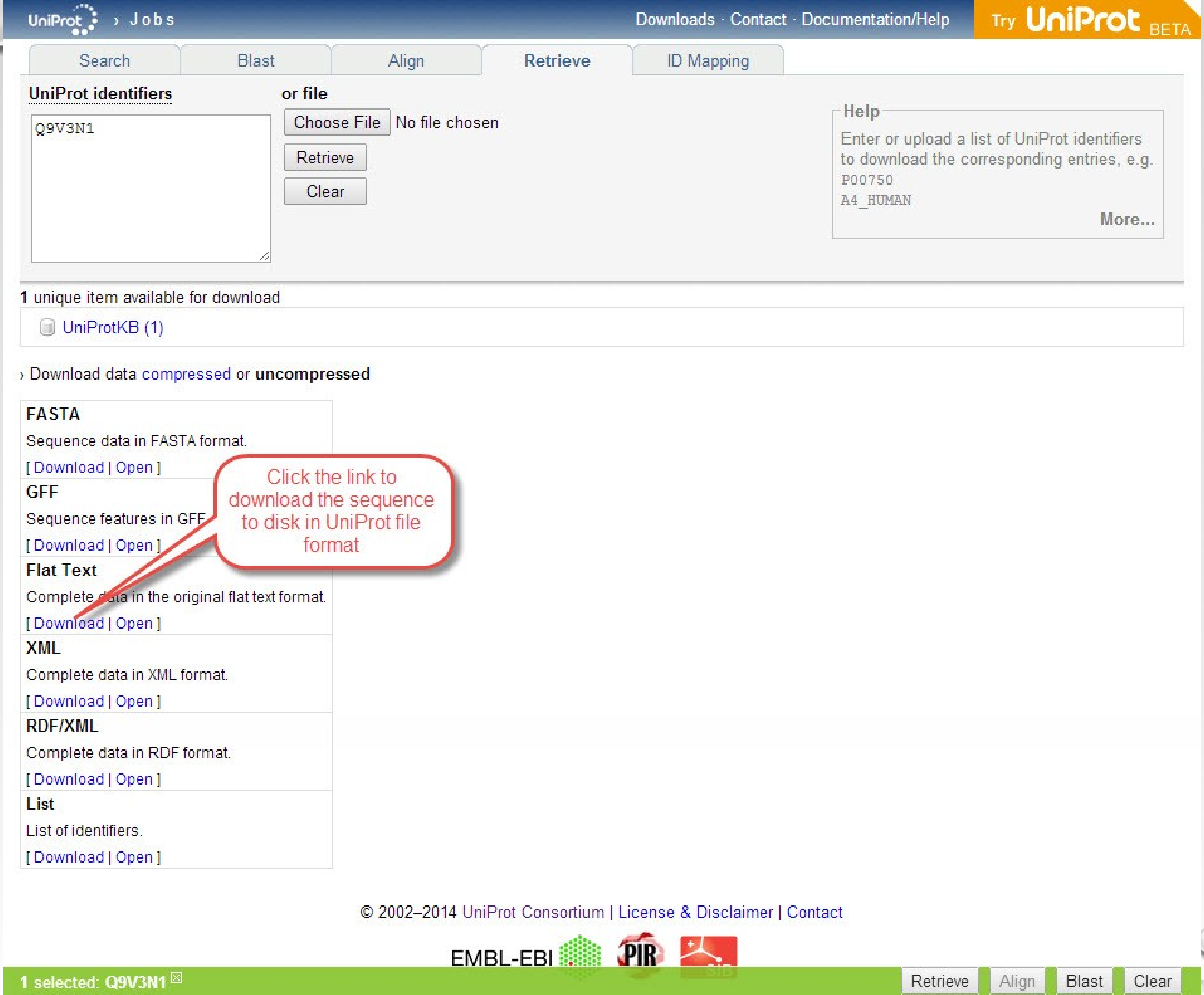
Click the link to download entire matching sequences

› Show only reviewed (311) (UniProtKB/Swiss-Prot) or unreviewed (7,971) (UniProtKB/TrEMBL) entries

› Restrict term "serpin" to protein family (6,881), gene name (93), protein name (2,342), web resource (1)

	Entry	Entry name	Status	Protein names	Gene names	Organism	Length
☑	Q9V3N1	Q9V3N1_DROME	☆	**LD24467p**	**Spn27A** serpin-27A Spn27A-RA CG11331 Dmel_CG11331	Drosophila melanogaster (Fruit fly)	447
☐	P05120	PAI2_HUMAN	★	**Plasminogen activator inhibitor 2**	**SERPINB2** PAI2 PLANH2	Homo sapiens (Human)	415
☐	Q5I2A0	SPA3G_MOUSE	★	**Serine protease inhibitor A3G**	**Serpina3g** Spi2A	Mus musculus (Mouse)	440
☐	Q8IW75	SPA12_HUMAN	★	**Serpin A12**	**SERPINA12**	Homo sapiens (Human)	414
☐	Q9UIV8	SPB13_HUMAN	★	**Serpin B13**	**SERPINB13** PI13	Homo sapiens (Human)	391
☐	Q9S7T8	SPZX_ARATH	★	**Serpin-ZX**	At1g47710 F16N3.3 T2E6.22	Arabidopsis thaliana (Mouse-ear cress)	391
☐	D6WWC4	D6WWC4_TRICA	☆	**Serpin peptidase inhibitor 23**	**serpin23** TcasGA2_TC005771	Tribolium castaneum (Red flour beetle)	385
		WWV2_TRICA		**Serpin peptidase inhibitor 31**	**serpin31** TcasGA2_TC015224		

1 selected: Q9V3N1 ☒



UniProt identifiers

Q9V3N1

or file

No file chosen

Help

Enter or upload a list of UniProt identifiers to download the corresponding entries, e.g.

`P00750`
`A4_HUMAN`

1 unique item available for download

UniProtKB (1)

› Download data compressed or **uncompressed**

FASTA Sequence data in FASTA format. [Download | Open]
GFF Sequence features in GFF [Download | Open]
Flat Text Complete data in the original flat text format. [Download | Open]
XML Complete data in XML format. [Download | Open]
RDF/XML Complete data in RDF format. [Download | Open]
List List of identifiers. [Download | Open]

Click the link to download the sequence to disk in UniProt file format

© 2002–2014 UniProt Consortium | License & Disclaimer | Contact

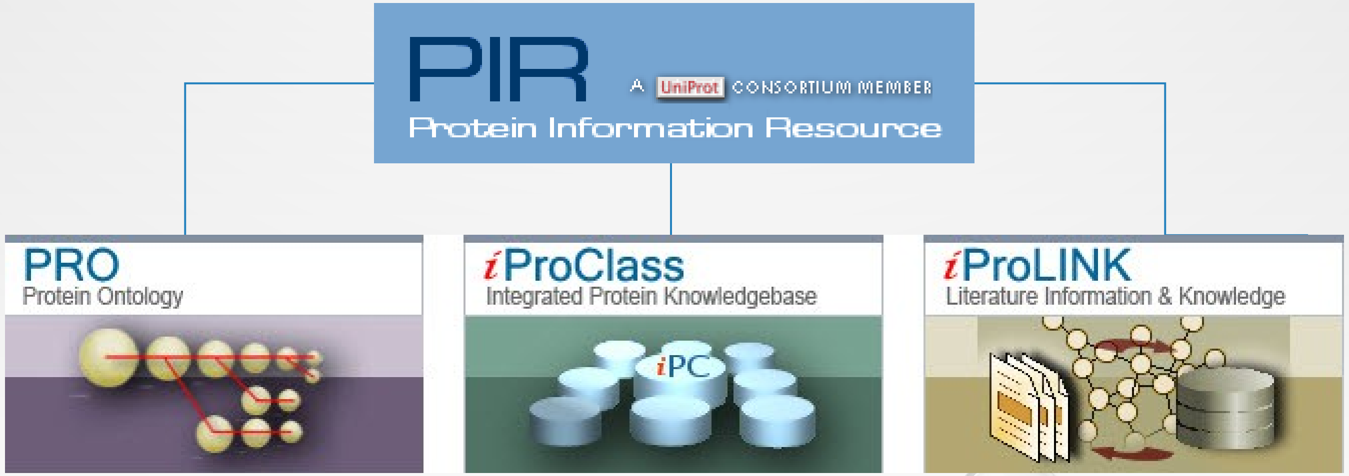
1 selected: Q9V3N1

UNIPROT/SWISS-PROT FILE FORMAT

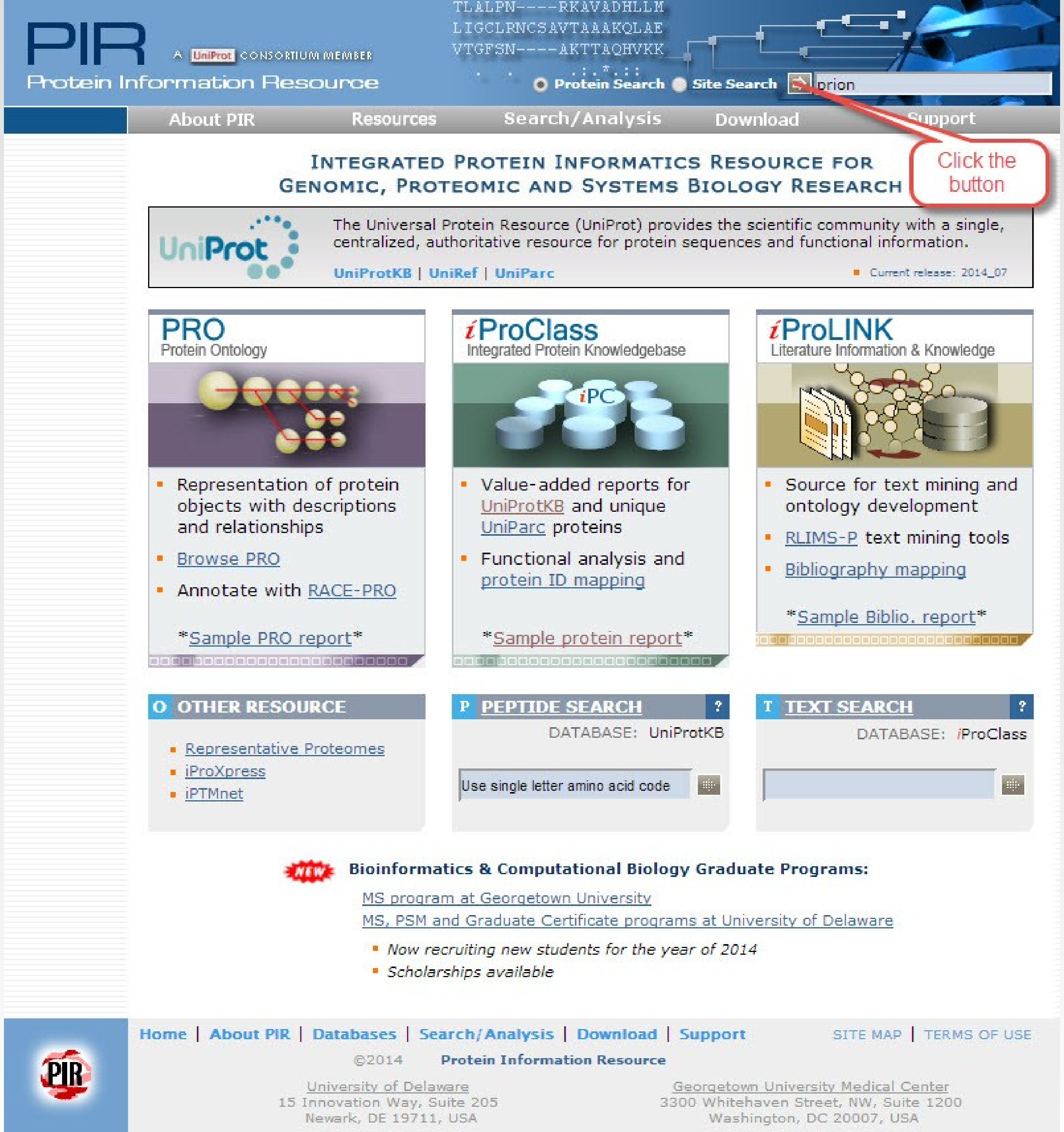
Line code	**Content**	**Occurrence in an entry**
ID	Identification	Once; starts the entry
AC	Accession number(s)	Once or more
DT	Date	Three times
DE	Description	Once or more
GN	Gene name(s)	Optional
OS	Organism species	Once or more
OG	Organelle	Optional
OC	Organism classification	Once or more
OX	Taxonomy cross-reference	Once
OH	Organism host	Optional
RN	Reference number	Once or more
RP	Reference position	Once or more
RC	Reference comment(s)	Optional
RX	Reference cross-reference(s)	Optional
RG	Reference group	Once or more (Optional if RA line)
RA	Reference authors	Once or more (Optional if RG line)
RT	Reference title	Optional
RL	Reference location	Once or more
CC	Comments or notes	Optional
DR	Database cross-references	Optional
PE	Protein existence	Once
KW	Keywords	Optional
FT	Feature table data	Once or more in Swiss-Prot, optional in TrEMBL
SQ	Sequence header	Once
(blanks)	Sequence data	Once or more
//	Termination line	Once; ends the entry

NBRF/PIR

- The Protein Information Resource (**PIR**) was established in 1984 by the National Biomedical Research Foundation (**NBRF**) as a resource to assist researchers in the identification and interpretation of protein sequence information.
- In 2002 PIR, along with its international partners, EBI and SIB, were awarded a grant from NIH to create UniProt, by unifying the PIR-PSD, Swiss-Prot, and TrEMBL databases.
- As of 2010, PIR offers a wide variety of resources mainly oriented to assist the propagation and standardization of protein annotation: PRO, iProClass, iProLINK.



http://pir.georgetown.edu/



Click the button

INTEGRATED PROTEIN INFORMATICS RESOURCE FOR GENOMIC, PROTEOMIC AND SYSTEMS BIOLOGY RESEARCH

The Universal Protein Resource (UniProt) provides the scientific community with a single, centralized, authoritative resource for protein sequences and functional information.

UniProtKB | UniRef | UniParc

Current release: 2014_07

PRO
Protein Ontology

- Representation of protein objects with descriptions and relationships
- Browse PRO
- Annotate with RACE-PRO

Sample PRO report

iProClass
Integrated Protein Knowledgebase

- Value-added reports for UniProtKB and unique UniParc proteins
- Functional analysis and protein ID mapping

Sample protein report

iProLINK
Literature Information & Knowledge

- Source for text mining and ontology development
- RLIMS-P text mining tools
- Bibliography mapping

Sample Biblio. report

O OTHER RESOURCE

- Representative Proteomes
- iProXpress
- iPTMnet

P PEPTIDE SEARCH ?

DATABASE: UniProtKB

Use single letter amino acid code

T TEXT SEARCH ?

DATABASE: iProClass

Bioinformatics & Computational Biology Graduate Programs:

MS program at Georgetown University
MS, PSM and Graduate Certificate programs at University of Delaware

- *Now recruiting new students for the year of 2014*
- *Scholarships available*

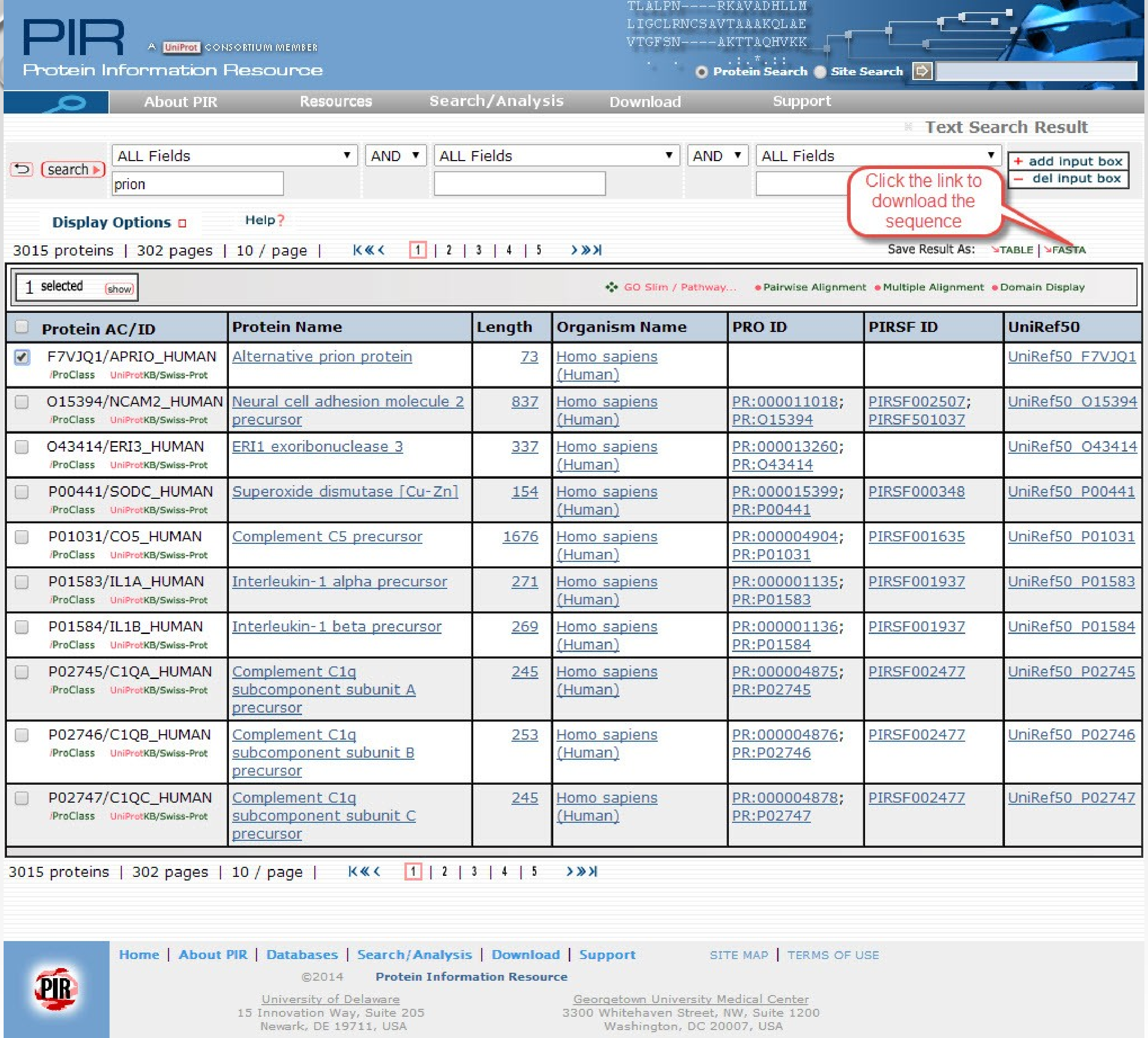
Home | About PIR | Databases | Search/Analysis | Download | Support

SITE MAP | TERMS OF USE

©2014 Protein Information Resource

University of Delaware
15 Innovation Way, Suite 205
Newark, DE 19711, USA

Georgetown University Medical Center
3300 Whitehaven Street, NW, Suite 1200
Washington, DC 20007, USA



Text Search Result

ALL Fields: prion AND ALL Fields AND ALL Fields

Click the link to download the sequence

3015 proteins | 302 pages | 10 / page | 1 | 2 | 3 | 4 | 5

Save Result As: TABLE | FASTA

1 selected (show)

☐	Protein AC/ID	Protein Name	Length	Organism Name	PRO ID	PIRSF ID	UniRef50
☑	F7VJQ1/APRIO_HUMAN iProClass UniProtKB/Swiss-Prot	Alternative prion protein	73	Homo sapiens (Human)			UniRef50_F7VJQ1
☐	O15394/NCAM2_HUMAN iProClass UniProtKB/Swiss-Prot	Neural cell adhesion molecule 2 precursor	837	Homo sapiens (Human)	PR:000011018; PR:O15394	PIRSF002507; PIRSF501037	UniRef50_O15394
☐	O43414/ERI3_HUMAN iProClass UniProtKB/Swiss-Prot	ERI1 exoribonuclease 3	337	Homo sapiens (Human)	PR:000013260; PR:O43414		UniRef50_O43414
☐	P00441/SODC_HUMAN iProClass UniProtKB/Swiss-Prot	Superoxide dismutase [Cu-Zn]	154	Homo sapiens (Human)	PR:000015399; PR:P00441	PIRSF000348	UniRef50_P00441
☐	P01031/CO5_HUMAN iProClass UniProtKB/Swiss-Prot	Complement C5 precursor	1676	Homo sapiens (Human)	PR:000004904; PR:P01031	PIRSF001635	UniRef50_P01031
☐	P01583/IL1A_HUMAN iProClass UniProtKB/Swiss-Prot	Interleukin-1 alpha precursor	271	Homo sapiens (Human)	PR:000001135; PR:P01583	PIRSF001937	UniRef50_P01583
☐	P01584/IL1B_HUMAN iProClass UniProtKB/Swiss-Prot	Interleukin-1 beta precursor	269	Homo sapiens (Human)	PR:000001136; PR:P01584	PIRSF001937	UniRef50_P01584
☐	P02745/C1QA_HUMAN iProClass UniProtKB/Swiss-Prot	Complement C1q subcomponent subunit A precursor	245	Homo sapiens (Human)	PR:000004875; PR:P02745	PIRSF002477	UniRef50_P02745
☐	P02746/C1QB_HUMAN iProClass UniProtKB/Swiss-Prot	Complement C1q subcomponent subunit B precursor	253	Homo sapiens (Human)	PR:000004876; PR:P02746	PIRSF002477	UniRef50_P02746
☐	P02747/C1QC_HUMAN iProClass UniProtKB/Swiss-Prot	Complement C1q subcomponent subunit C precursor	245	Homo sapiens (Human)	PR:000004878; PR:P02747	PIRSF002477	UniRef50_P02747

3015 proteins | 302 pages | 10 / page | 1 | 2 | 3 | 4 | 5

©2014 Protein Information Resource

University of Delaware
15 Innovation Way, Suite 205
Newark, DE 19711, USA

Georgetown University Medical Center
3300 Whitehaven Street, NW, Suite 1200
Washington, DC 20007, USA

iProClass Summary Report for UniProtKB Entry: F7VJQ1

BioThesaurus *ID Mapping*

GENERAL INFORMATION	
Protein Name and ID	UniProtKB ID: APRIO_HUMAN; UniProtKB Accession: F7VJQ1; Protein Name: Alternative prion protein RefSeq: NP_001258490.1 GenPept: DAA34790.1 IPI: IPI01026282
Taxonomy	*Source Organism:* Homo sapiens (Human) *Taxon Group:* Euk/mammal *NCBI Taxon:* 9606 *Lineage:* Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
Keywords	complete proteome; membrane; mitochondrion; mitochondrion outer membrane; reference proteome; transmembrane; transmembrane helix
Function	Involved in the secretion of alkaline protease.
Tissue Specificity	Detected in brain homogenate, primary neurons, and peripheral blood mononuclear cells (at protein level).

CROSS-REFERENCES	
Bibliography	▸ View Bibliography Information *Annotated references:* PMID: 11120925; 11220690; 11244488; 11559357; 11584448; 11593450; 11684342; 11704923; 11756421; 11775001; 11787070; 11833672; 11840201; 11961239; 11986958; 12070046; 12084159; 12161431; 12186633; 12205650 [GeneRIF] More *Other references:* PMID: 16004966; 2180366; 16294306; 7913755; 10970892; 1683708; 8105771; 1736177; 1357663; 12679034; 2564168; 12514748; 12356908; 17500595; 11438139; 18051367; 1351748; 9786248; 10988071; 1363802; More
DNA Sequence	EMBL: AL133396; BK007887 DDBJ: AL133396; BK007887 GenBank: AL133396; BK007887
Genome/Gene	*Gene Name:* PRNP; prion protein *Locus Tag*: RP5-1068H6.2; *Synonyms*: ASCR; AltPrP; CD230; CJD; GSS; KURU; PRIP; PrP; PrP27-30; PrP33-35C; PrPc; p27-30 ; *Map Location*: 20p13 Entrez Gene: 5621 UniGene: Hs.472010; Hs.610285; Hs.721670 RefSeq: NM_000311.3 NP_000302.1 [Map Viewer]; NM_001080121.1 NP_001073590.1 [Map Viewer]; NM_001080122.1 NP_001073591.1 [Map Viewer]; NM_001080123.1 NP_001073592.1 [Map Viewer]; NM_001271561.1 NP_001258490.1 [Map Viewer]; NM_183079.2 NP_898902.1 [Map Viewer] UCSC: uc002wkw.4 human KEGG: hsa:5621
Gene Expression	GEO
Ontology	Gene Ontology: *Cellular Component* GO:0005739:mitochondrion [GO_REF:0000037; evidence:IEA] GO:0005741:mitochondrial outer membrane [GO_REF:0000037; evidence:IEA][GO_REF:0000039; evidence:IEA][PMID:21478263; evidence:IDA] GO:0016021:integral component of membrane [GO_REF:0000037; evidence:IEA] GO:0016020:membrane [GO_REF:0000037; evidence:IEA]
Structure	ProteinModelPortal: F7VJQ1
Other Databases	neXtProt: NX_F7VJQ1 NextBio: 35510517 CTD: 5621 GenomeRNAi: 5621 GeneWiki: PRNP

FAMILY CLASSIFICATION	
UniRef	UniRef100_F7VJQ1; UniRef90_F7VJQ1; UniRef50_F7VJQ1
Other Classification	KO: K05634

FEATURE & SEQUENCE DISPLAY

Length = 73 Click on a bar to show its sequence; to copy and paste it, press ctl-c then ctl-v.

```
 1  MEHWGQPIPGAGQPWRQPLPTSGRWWLGAASWWWLGAASWWWLGAAPWWWLGTASWWWLG
61  SRRWHPQSVEQAE
```

University of Delaware
15 Innovation Way, Suite 205
Newark, DE 19711, USA

Georgetown University Medical Center
3300 Whitehaven Street, NW, Suite 1200
Washington, DC 20007, USA

SEQUENCE RETRIEVED FROM NBRF/PIR IN FASTA FILE FORMAT

```
>F7VJQ1 APRIO_HUMAN Alternative prion protein [Homo
sapiens]
MEHWGQPIPGAGQPWRQPLPTSGRWWLGAASWWWLGAASWWWLGAAPWWWLGTASWWWL
G
SRRWHPQSVEQAE
```

PDB

- The Protein Data Bank (**PDB**) archive is the single worldwide repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids.
- The PDB was established in 1971 at Brookhaven National Laboratory (**BNL**) under the leadership of Walter Hamilton and originally contained 7 structures.
- In 1998, the Research Collaboratory for Structural Bioinformatics (**RCSB**) became responsible for the management of the PDB.
- In 2003, the **wwPDB** was formed to maintain a single PDB archive of macromolecular structural data that is freely and publicly available to the global community.
- The RCSB PDB supports a website where visitors can perform simple and complex queries on the data, analyze, and visualize the results.
- Members of wwPDB are: **RCSB PDB** (USA), **PDBe** (Europe) and **PDBj** (Japan), and Biological Magnetic Resonance Data Bank **BMRB** (USA).

http://rcsb.org/pdb/

An Information Portal to Biological Macromolecular Structures

As of **Tuesday Jul 29, 2014 at 5 PM PDT** there are **102158** Structures | **PDB Statistics**

Biological Macromolecular Resource

Full Description

- Homo sapiens (26245)
- Escherichia coli (4855)
- Mus musculus (4419)
- Saccharomyces cerevis
- Bos taurus (2321)
- Rattus norvegicus (215
- Escherichia coli K-12 (2
- Other (56239)

4TP8 : Crystal structure of the E. coli ribosome bound to dalfopristin and quinupristin. This file contains the 30S subuni...

Noeske, J., Huang, J., Olivier, N.B., Giacobbe, R.A., Zambrowski, M., Cate, J.H.

Synergy of streptogramin antibiotics occurs independently of their effects on translation.

check out all of the **latest** structures released.

New user? Try the browser **Compatibility Check** and view information on **Getting Started**.

Latest release: April 2014

2014-08-01

wwPDB Events at IUCr (August 5-12)

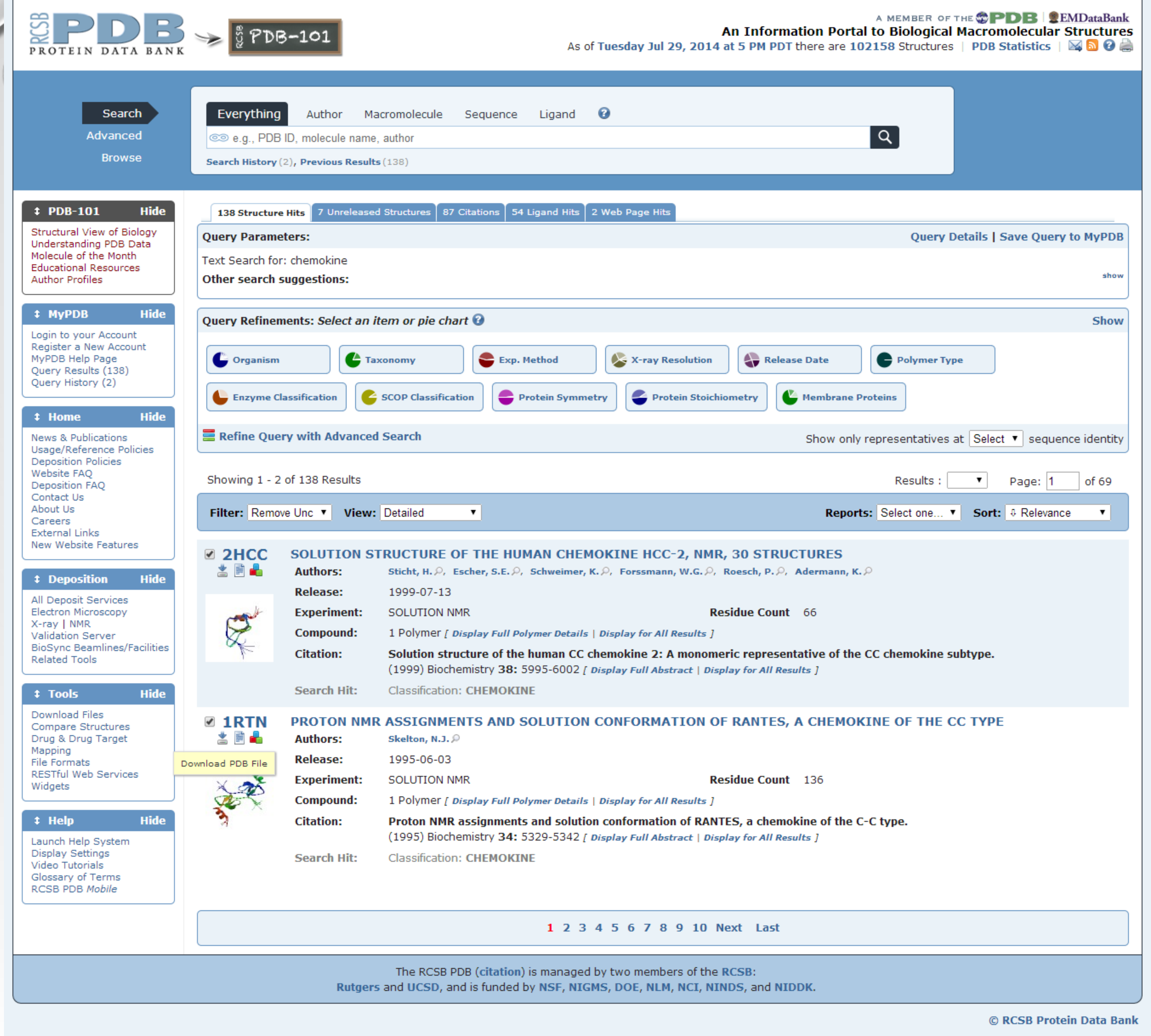
- **Improved Representation of Large Structures in the PDB Archive**
- **Full wwPDB News**
- **Statement on Retraction of PDB Entries**

Current PDB FTP Archive:
ftp.wwpdb.org

Yearly FTP Snapshots:
ftp://snapshots.wwpdb.org

The RCSB PDB (**citation**) is managed by two members of the **RCSB**:
Rutgers and **UCSD**, and is funded by **NSF**, **NIGMS**, **DOE**, **NLM**, **NCI**, **NINDS**, and **NIDDK**.

© RCSB Protein Data Bank



An Information Portal to Biological Macromolecular Structures

As of Tuesday Jul 29, 2014 at 5 PM PDT there are 102158 Structures | PDB Statistics

Search History (2), Previous Results (138)

Query Parameters:

Text Search for: chemokine

Other search suggestions:

Query Refinements: *Select an item or pie chart*

Show only representatives at Select sequence identity

Showing 1 - 2 of 138 Results

Results : Page: 1 of 69

Filter: Remove Unc **View:** Detailed **Reports:** Select one... **Sort:** Relevance

☑ **2HCC** SOLUTION STRUCTURE OF THE HUMAN CHEMOKINE HCC-2, NMR, 30 STRUCTURES

Authors: Sticht, H., Escher, S.E., Schweimer, K., Forssmann, W.G., Roesch, P., Adermann, K.

Release: 1999-07-13

Experiment: SOLUTION NMR **Residue Count** 66

Compound: 1 Polymer [*Display Full Polymer Details* | *Display for All Results*]

Citation: **Solution structure of the human CC chemokine 2: A monomeric representative of the CC chemokine subtype.** (1999) Biochemistry **38:** 5995-6002 [*Display Full Abstract* | *Display for All Results*]

Search Hit: Classification: **CHEMOKINE**

☑ **1RTN** PROTON NMR ASSIGNMENTS AND SOLUTION CONFORMATION OF RANTES, A CHEMOKINE OF THE CC TYPE

Authors: Skelton, N.J.

Release: 1995-06-03

Experiment: SOLUTION NMR **Residue Count** 136

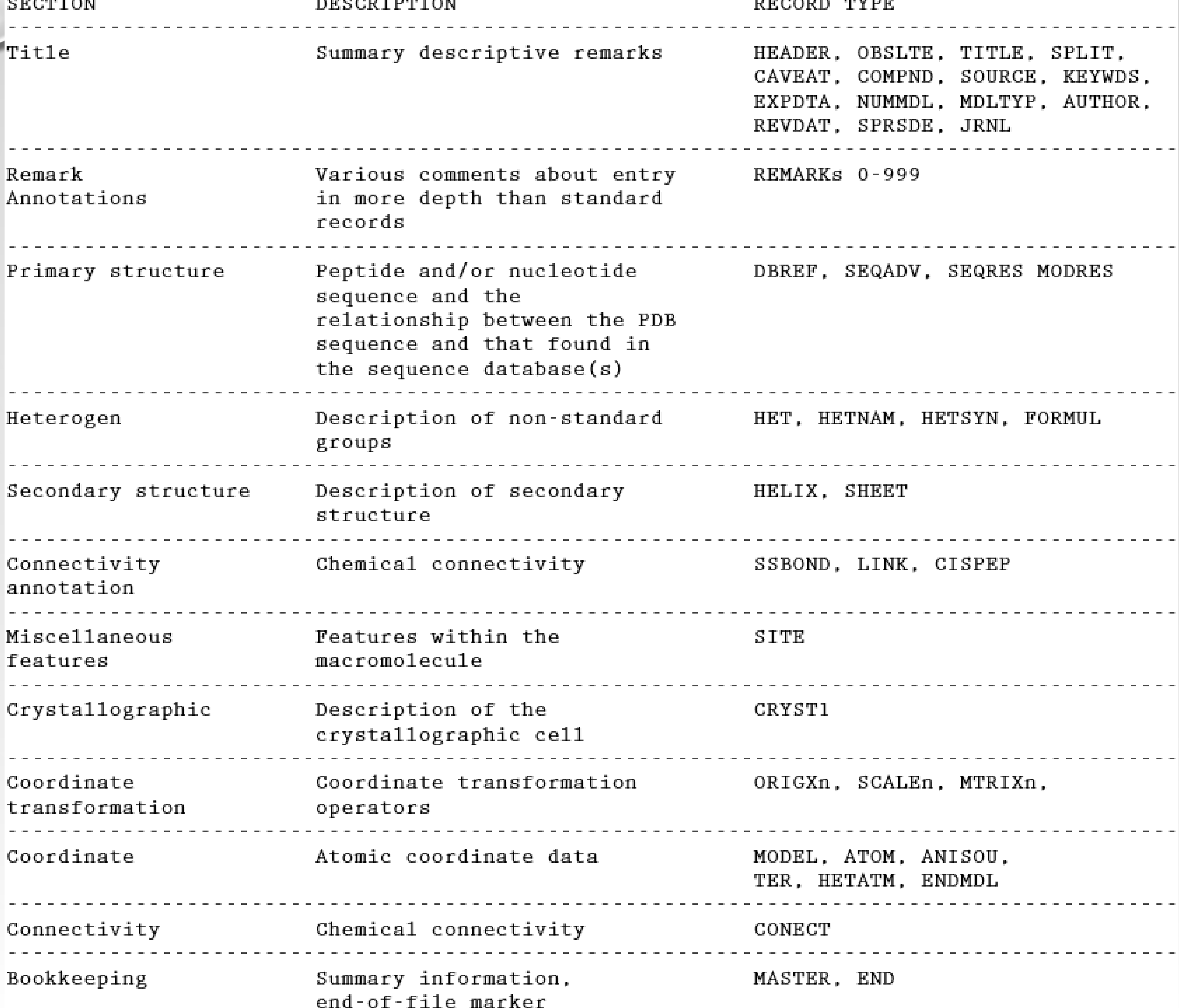
Compound: 1 Polymer [*Display Full Polymer Details* | *Display for All Results*]

Citation: **Proton NMR assignments and solution conformation of RANTES, a chemokine of the C-C type.** (1995) Biochemistry **34:** 5329-5342 [*Display Full Abstract* | *Display for All Results*]

Search Hit: Classification: **CHEMOKINE**

The RCSB PDB (citation) is managed by two members of the RCSB: Rutgers and UCSD, and is funded by NSF, NIGMS, DOE, NLM, NCI, NINDS, and NIDDK.

© RCSB Protein Data Bank



SECTION	DESCRIPTION	RECORD TYPE
Title	Summary descriptive remarks	HEADER, OBSLTE, TITLE, SPLIT, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, NUMMDL, MDLTYP, AUTHOR, REVDAT, SPRSDE, JRNL
Remark Annotations	Various comments about entry in more depth than standard records	REMARKs 0-999
Primary structure	Peptide and/or nucleotide sequence and the relationship between the PDB sequence and that found in the sequence database(s)	DBREF, SEQADV, SEQRES MODRES
Heterogen	Description of non-standard groups	HET, HETNAM, HETSYN, FORMUL
Secondary structure	Description of secondary structure	HELIX, SHEET
Connectivity annotation	Chemical connectivity	SSBOND, LINK, CISPEP
Miscellaneous features	Features within the macromolecule	SITE
Crystallographic	Description of the crystallographic cell	CRYST1
Coordinate transformation	Coordinate transformation operators	ORIGXn, SCALEn, MTRIXn,
Coordinate	Atomic coordinate data	MODEL, ATOM, ANISOU, TER, HETATM, ENDMDL
Connectivity	Chemical connectivity	CONECT
Bookkeeping	Summary information, end-of-file marker	MASTER, END

SCOP

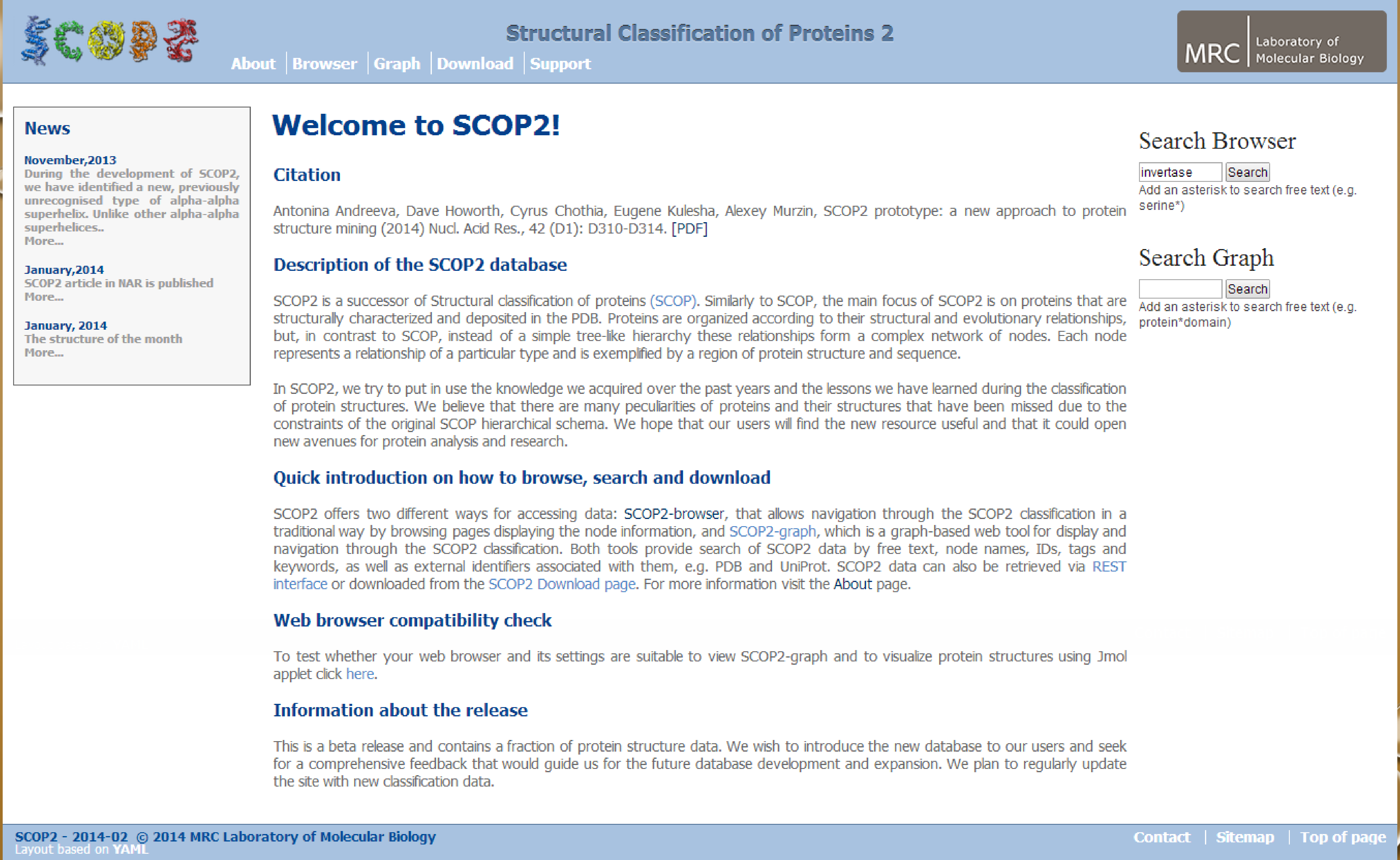
- The Structural Classification of Proteins (**SCOP**) database is a largely manual classification of protein structural domains based on similarities of their structures and amino acid sequences.
- A motivation for this classification is to determine the evolutionary relationship between proteins.
- Proteins with the same shapes but having little sequence or functional similarity are placed in different "superfamilies", and are assumed to have only a very distant common ancestor.
- Proteins having the same shape and some similarity of sequence and/or function are placed in "families", and are assumed to have a closer common ancestor.
- SCOP has been discontinued and the last official version of SCOP is 1.75. SCOP1.75 is also known as SCOP2.
- SCOP2 offers two different ways for accessing data: SCOP2-browser, and SCOP2-graph.
- SCOP2-browser allows navigation in a traditional way by browsing pages displaying the node information.
- SCOP2-graph is a graph-based web tool for display and navigation.
- The source of protein structures is the Protein Data Bank.

HIERARCHICAL STRUCTURE OF SCOP

- The unit of classification of structure in SCOP is the protein domain.
- The levels of SCOP are as follows.

1. **Class**: Types of folds, e.g., all α, all β, α/β, α+β, α&β, etc.
2. **Fold**: The different shapes of domains within a class, e.g., 2 helices; antiparallel hairpin, left-handed twist, etc.
3. **Superfamily**: The domains in a fold are grouped into superfamilies, which have at least **a** distant common ancestor.
4. **Family**: The domains in a superfamily are grouped into families, which have **more** recent common ancestor.
5. **Protein domain**: The domains in families are grouped into protein domains, which are essentially the same protein.
6. **Species**: The domains in "protein domains" are grouped according to species.
7. **Domain**: It is part of a protein. For simple proteins, it can be the entire protein.

http://scop2.mrc-lmb.cam.ac.uk/



Structural Classification of Proteins 2

News

November,2013
During the development of SCOP2, we have identified a new, previously unrecognised type of alpha-alpha superhelix. Unlike other alpha-alpha superhelices..
More...

January,2014
SCOP2 article in NAR is published
More...

January, 2014
The structure of the month
More...

Welcome to SCOP2!

Citation

Antonina Andreeva, Dave Howorth, Cyrus Chothia, Eugene Kulesha, Alexey Murzin, SCOP2 prototype: a new approach to protein structure mining (2014) Nucl. Acid Res., 42 (D1): D310-D314. [PDF]

Description of the SCOP2 database

SCOP2 is a successor of Structural classification of proteins (SCOP). Similarly to SCOP, the main focus of SCOP2 is on proteins that are structurally characterized and deposited in the PDB. Proteins are organized according to their structural and evolutionary relationships, but, in contrast to SCOP, instead of a simple tree-like hierarchy these relationships form a complex network of nodes. Each node represents a relationship of a particular type and is exemplified by a region of protein structure and sequence.

In SCOP2, we try to put in use the knowledge we acquired over the past years and the lessons we have learned during the classification of protein structures. We believe that there are many peculiarities of proteins and their structures that have been missed due to the constraints of the original SCOP hierarchical schema. We hope that our users will find the new resource useful and that it could open new avenues for protein analysis and research.

Quick introduction on how to browse, search and download

SCOP2 offers two different ways for accessing data: SCOP2-browser, that allows navigation through the SCOP2 classification in a traditional way by browsing pages displaying the node information, and SCOP2-graph, which is a graph-based web tool for display and navigation through the SCOP2 classification. Both tools provide search of SCOP2 data by free text, node names, IDs, tags and keywords, as well as external identifiers associated with them, e.g. PDB and UniProt. SCOP2 data can also be retrieved via REST interface or downloaded from the SCOP2 Download page. For more information visit the About page.

Web browser compatibility check

To test whether your web browser and its settings are suitable to view SCOP2-graph and to visualize protein structures using Jmol applet click here.

Information about the release

This is a beta release and contains a fraction of protein structure data. We wish to introduce the new database to our users and seek for a comprehensive feedback that would guide us for the future database development and expansion. We plan to regularly update the site with new classification data.

Search Browser

invertase

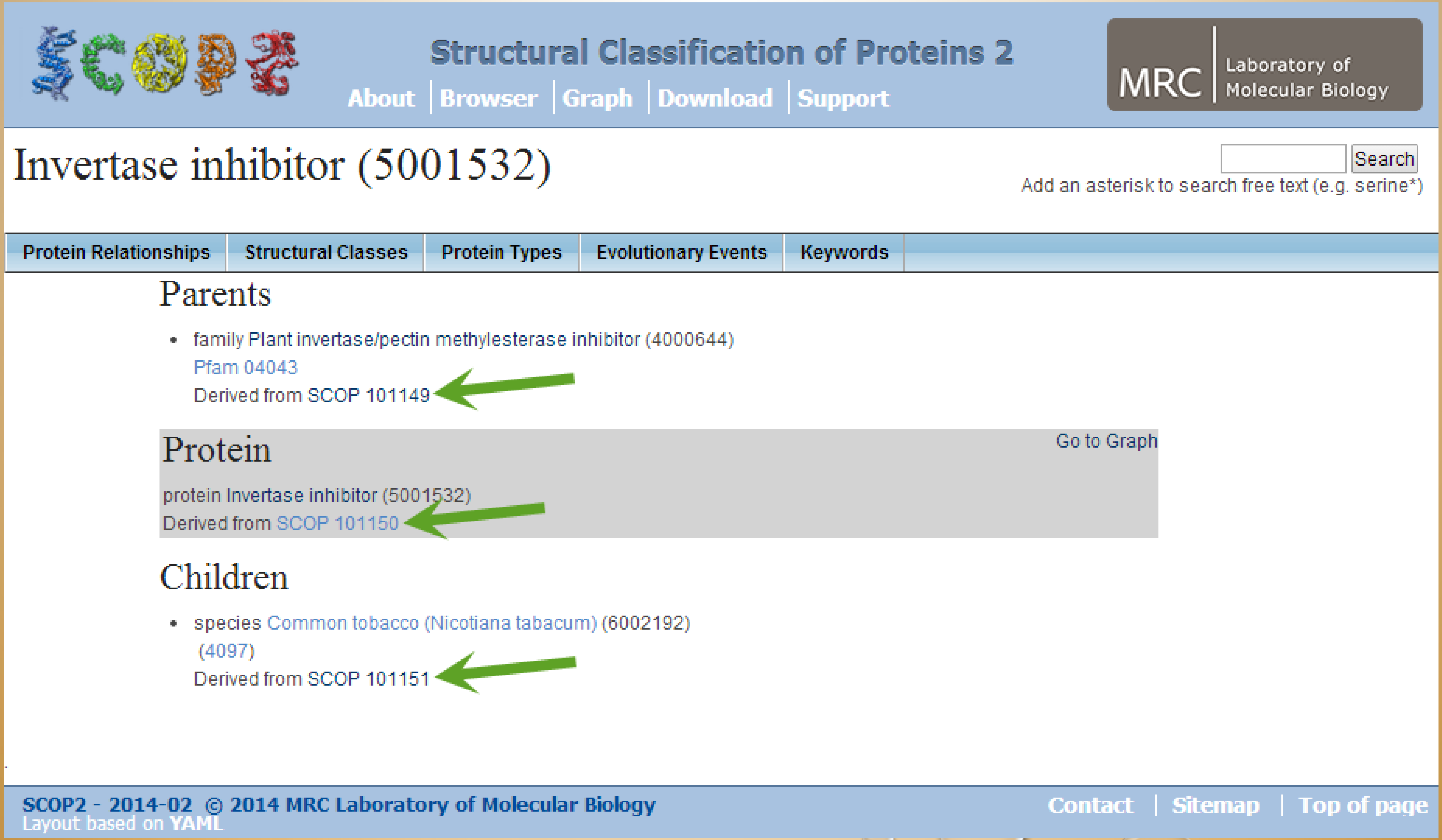
Add an asterisk to search free text (e.g. serine*)

Search Graph

Add an asterisk to search free text (e.g. protein*domain)

SCOP2 - 2014-02 © 2014 MRC Laboratory of Molecular Biology
Layout based on YAML

OUTPUT OF SCOP



Invertase inhibitor (5001532)

Add an asterisk to search free text (e.g. serine*)

Parents

- family Plant invertase/pectin methylesterase inhibitor (4000644)
  Pfam 04043
  Derived from SCOP 101149

Protein

protein Invertase inhibitor (5001532)
Derived from SCOP 101150

Children

- species Common tobacco (Nicotiana tabacum) (6002192)
  (4097)
  Derived from SCOP 101151

SCOP2 - 2014-02 © 2014 MRC Laboratory of Molecular Biology
Layout based on YAML

Structural Classification of Proteins

Family: Plant invertase/pectin methylesterase inhibitor

Pfam 04043

Lineage:

1. Root: scop
2. Class: All alpha proteins [46456]
3. Fold: Bromodomain-like [47363]
   4 helices; bundle; minor mirror variant of up-and-down topology
4. Superfamily: Plant invertase/pectin methylesterase inhibitor [101148]
   contains a short alpha-hairpin at the N-terminal extension
   Superfamily
5. Family: Plant invertase/pectin methylesterase inhibitor [101149]
   Pfam 04043

Protein Domains:

1. Invertase inhibitor [101150]
   1. Common tobacco (Nicotiana tabacum) [TaxId: 4097] [101151] (9)
2. Pectin methylesterase inhibitor 1, PMEI1 [116887]
   forms dimers and trimers by swapping the extra N-terminal hairpins
   1. Thale cress (Arabidopsis thaliana) [TaxId: 3702] [116888] (3)
      SQ Q9LNF2 28-176

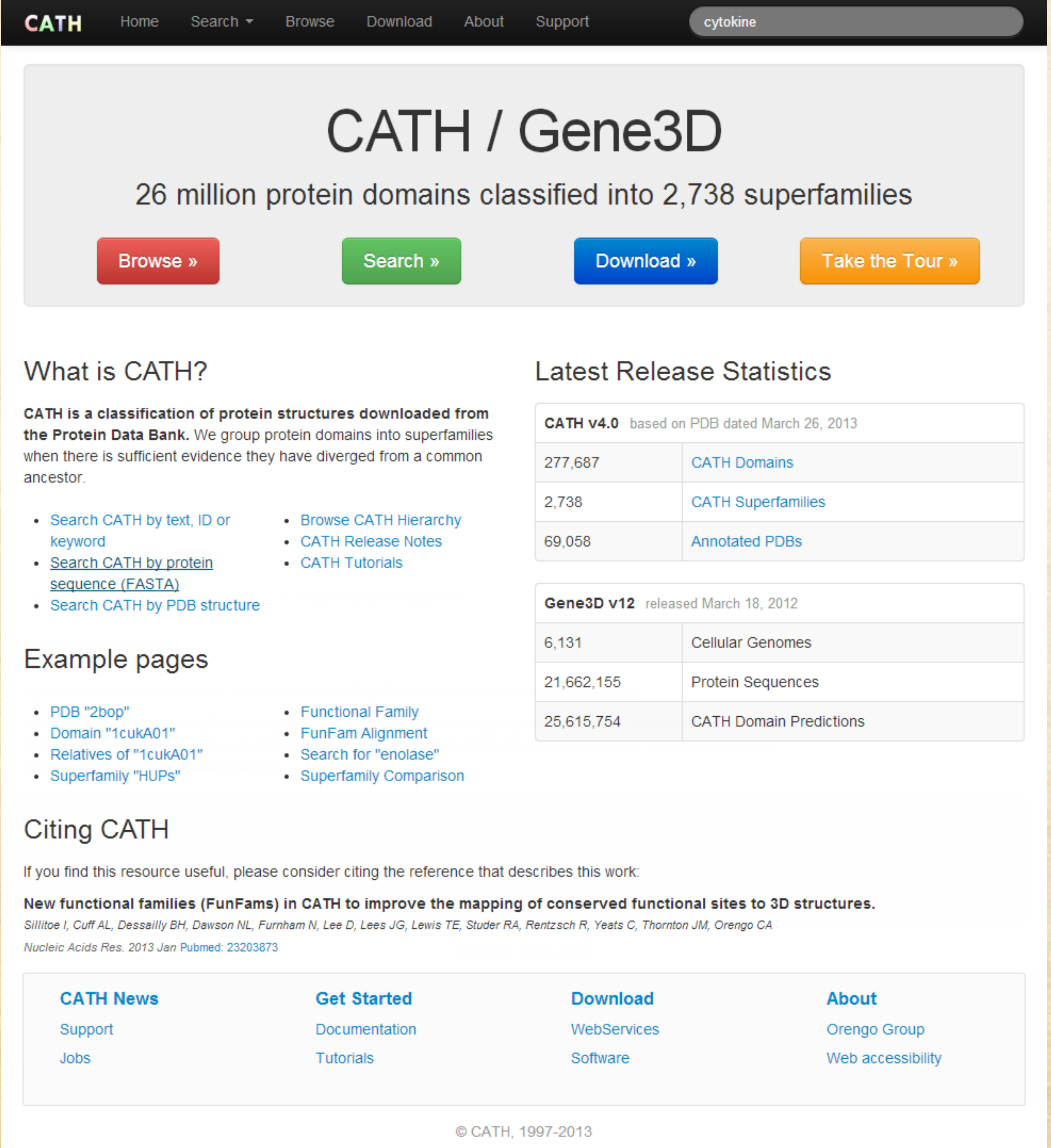
Enter search key: Search

Generated from scop database 1.75 with scopm 1.101 on Wed Jun 3 10:42:06 2009

CATH

- The CATH (**C**lass, **A**rchitecture, **T**opology, and **H**omologous superfamily) is a semi-automatic, hierarchical classification of protein domains.
- CATH shares many broad features with its principal rival, SCOP.
- The four main levels of the CATH hierarchy are as follows:
  - **Class:** the overall secondary-structure content of the domain. e.g., all α, all β, α/β, α+β, α&β, etc.
  - **Architecture:** high structural similarity but **no evidence of homology**. Equivalent to a fold in SCOP.
  - **Topology:** a large-scale grouping of topologies which share particular structural features
  - **Homologous superfamily:** indicative of a demonstrable evolutionary relationship. Equivalent to the superfamily level of SCOP.

http://www.cathdb.info/



CATH / Gene3D

26 million protein domains classified into 2,738 superfamilies

What is CATH?

CATH is a classification of protein structures downloaded from the Protein Data Bank. We group protein domains into superfamilies when there is sufficient evidence they have diverged from a common ancestor.

- Search CATH by text, ID or keyword
- Search CATH by protein sequence (FASTA)
- Search CATH by PDB structure
- Browse CATH Hierarchy
- CATH Release Notes
- CATH Tutorials

Example pages

- PDB "2bop"
- Domain "1cukA01"
- Relatives of "1cukA01"
- Superfamily "HUPs"
- Functional Family
- FunFam Alignment
- Search for "enolase"
- Superfamily Comparison

Latest Release Statistics

CATH v4.0 based on PDB dated March 26, 2013	
277,687	CATH Domains
2,738	CATH Superfamilies
69,058	Annotated PDBs

Gene3D v12 released March 18, 2012	
6,131	Cellular Genomes
21,662,155	Protein Sequences
25,615,754	CATH Domain Predictions

Citing CATH

If you find this resource useful, please consider citing the reference that describes this work:

New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures.

Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, Lee D, Lees JG, Lewis TE, Studer RA, Rentzsch R, Yeats C, Thornton JM, Orengo CA

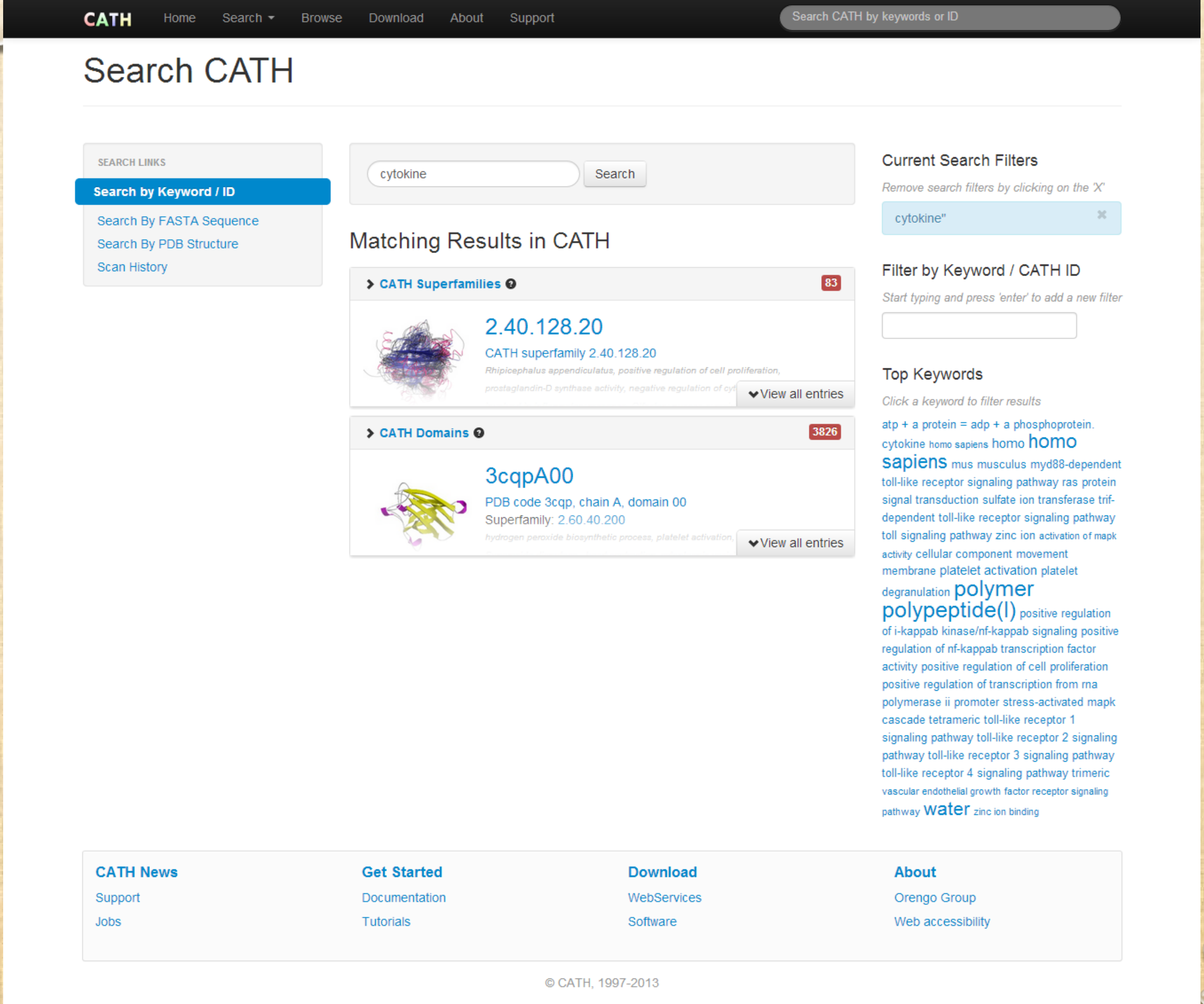
Nucleic Acids Res. 2013 Jan Pubmed: 23203873

CATH News
Support
Jobs

Get Started
Documentation
Tutorials

Download
WebServices
Software

About
Orengo Group
Web accessibility



Search CATH

SEARCH LINKS

- **Search by Keyword / ID**
- Search By FASTA Sequence
- Search By PDB Structure
- Scan History

cytokine

Matching Results in CATH

CATH Superfamilies 83

2.40.128.20

CATH superfamily 2.40.128.20

Rhipicephalus appendiculatus, positive regulation of cell proliferation, prostaglandin-D synthase activity, negative regulation of cyt

CATH Domains 3826

3cqpA00

PDB code 3cqp, chain A, domain 00

Superfamily: 2.60.40.200

hydrogen peroxide biosynthetic process, platelet activation,

Current Search Filters

Remove search filters by clicking on the 'X'

cytokine" ✖

Filter by Keyword / CATH ID

Start typing and press 'enter' to add a new filter

Top Keywords

Click a keyword to filter results

atp + a protein = adp + a phosphoprotein. cytokine homo sapiens homo homo sapiens mus musculus myd88-dependent toll-like receptor signaling pathway ras protein signal transduction sulfate ion transferase trif-dependent toll-like receptor signaling pathway toll signaling pathway zinc ion activation of mapk activity cellular component movement membrane platelet activation platelet degranulation polymer polypeptide(l) positive regulation of i-kappab kinase/nf-kappab signaling positive regulation of nf-kappab transcription factor activity positive regulation of cell proliferation positive regulation of transcription from rna polymerase ii promoter stress-activated mapk cascade tetrameric toll-like receptor 1 signaling pathway toll-like receptor 2 signaling pathway toll-like receptor 3 signaling pathway toll-like receptor 4 signaling pathway trimeric vascular endothelial growth factor receptor signaling pathway water zinc ion binding

CATH News
Support
Jobs

Get Started
Documentation
Tutorials

Download
WebServices
Software

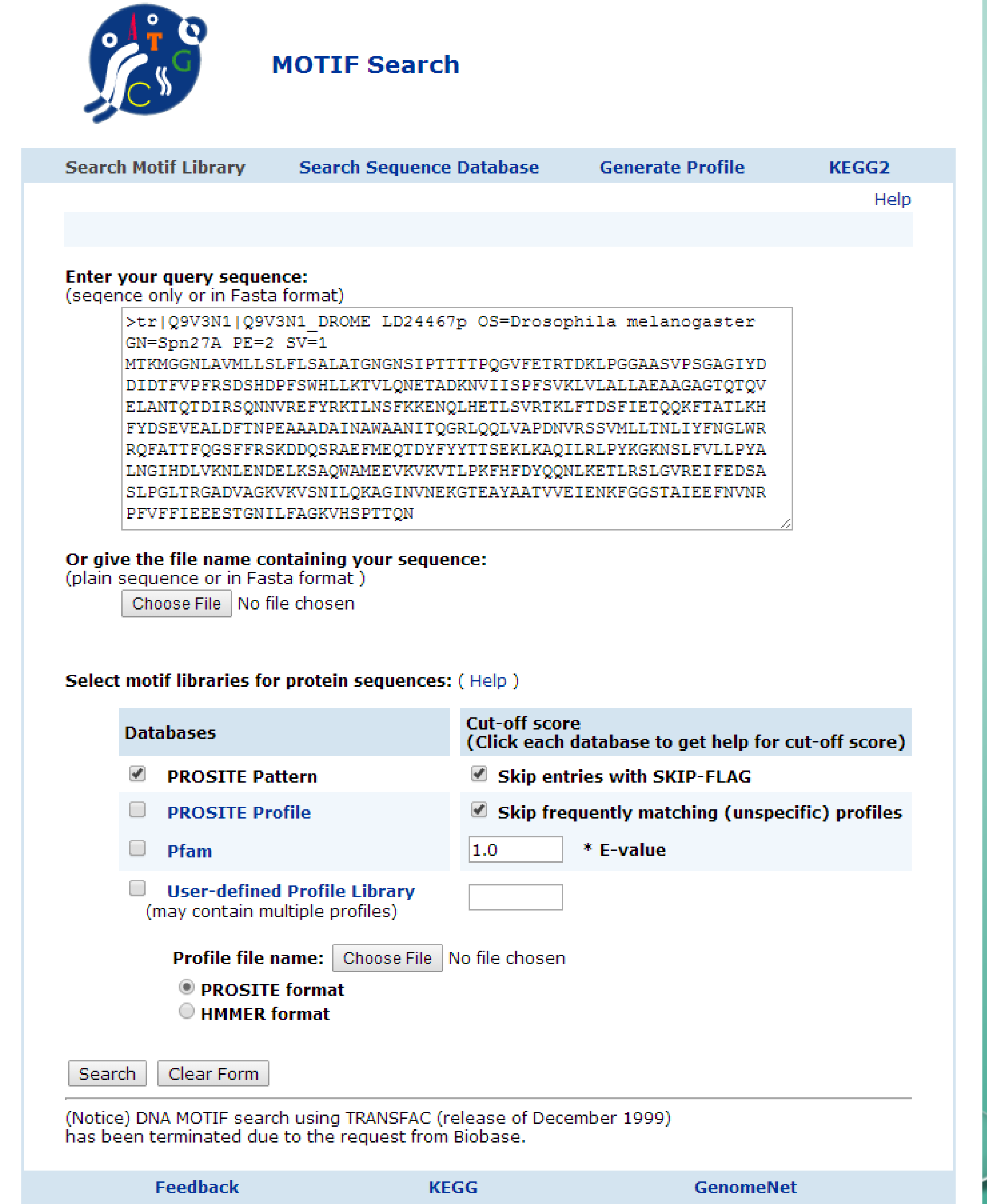
About
Orengo Group
Web accessibility

MOTIF

- Motif is a search service provided by GenomeNet to search with a protein query sequence against Motif Libraries.
- Supports several motif databases such as Prosite, BLOCKS, ProDom, Pfam, and PRINTS.
- Allows you to search protein sequence libraries with your patterns.
  - Each residue must be separated with - (minus sign).
  - x represents any amino acids.
  - [DE] means either D or E.
  - {FWY} means any amino acids except for F, W and Y
  - A(2,3) means that A appears 2 to 3 times consecutively.
  - The pattern string must be terminated with . (period).

    For example, C-x-{C}-[DN]-x(2)-C-x(5)-C-C.
- Generates a profile from a set of multiple aligned sequences using PFMake or HMMBuild, respectively.

http://www.genome.jp/tools/motif/



Enter your query sequence:
(seqence only or in Fasta format)

```
>tr|Q9V3N1|Q9V3N1_DROME LD24467p OS=Drosophila melanogaster
GN=Spn27A PE=2 SV=1
MTKMGGNLAVMLLSLFLSALATGNGNSIPTTTTPQGVFETRTDKLPGGAASVPSGAGIYD
DIDTFVPFRSDSHDPFSWHLLKTVLQNETADKNVIISPFSVKLVLALLAEAAGAGTQTQV
ELANTQTDIRSQNNVREFYRKTLNSFKKENQLHETLSVRTKLFTDSFIETQQKFTATLKH
FYDSEVEALDFTNPEAAADAINAWAANITQGRLQQLVAPDNVRSSVMLLTNLIYFNGLWR
RQFATTFQGSFFRSKDDQSRAEFMEQTDYFYYTTSEKLKAQILRLPYKGKNSLFVLLPYA
LNGIHDLVKNLENDELKSAQWAMEEVKVKVTLPKFHFDYQQNLKETLRSLGVREIFEDSA
SLPGLTRGADVAGKVKVSNILQKAGINVNEKGTEAYAATVVEIENKFGGSTAIEEFNVNR
PFVFFIEEESTGNILFAGKVHSPTTQN
```

Or give the file name containing your sequence:
(plain sequence or in Fasta format)

Choose File No file chosen

Select motif libraries for protein sequences: (Help)

Databases	**Cut-off score** **(Click each database to get help for cut-off score)**
☑ **PROSITE Pattern**	☑ **Skip entries with SKIP-FLAG**
☐ **PROSITE Profile**	☑ **Skip frequently matching (unspecific) profiles**
☐ **Pfam**	1.0 * **E-value**
☐ **User-defined Profile Library** (may contain multiple profiles)	

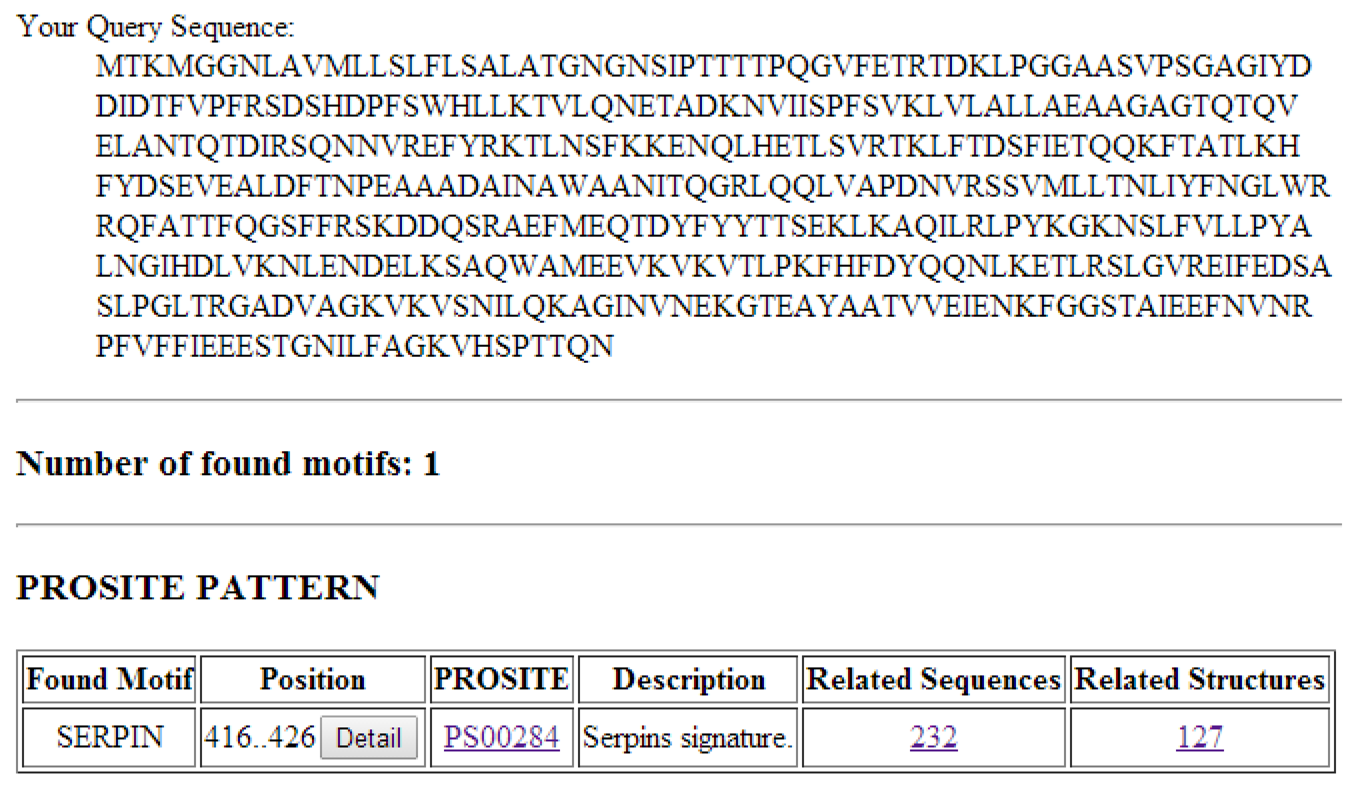
Profile file name: Choose File No file chosen

● **PROSITE format**
○ **HMMER format**

Search Clear Form

(Notice) DNA MOTIF search using TRANSFAC (release of December 1999)
has been terminated due to the request from Biobase.

MATCHING MOTIF HITS



Your Query Sequence:

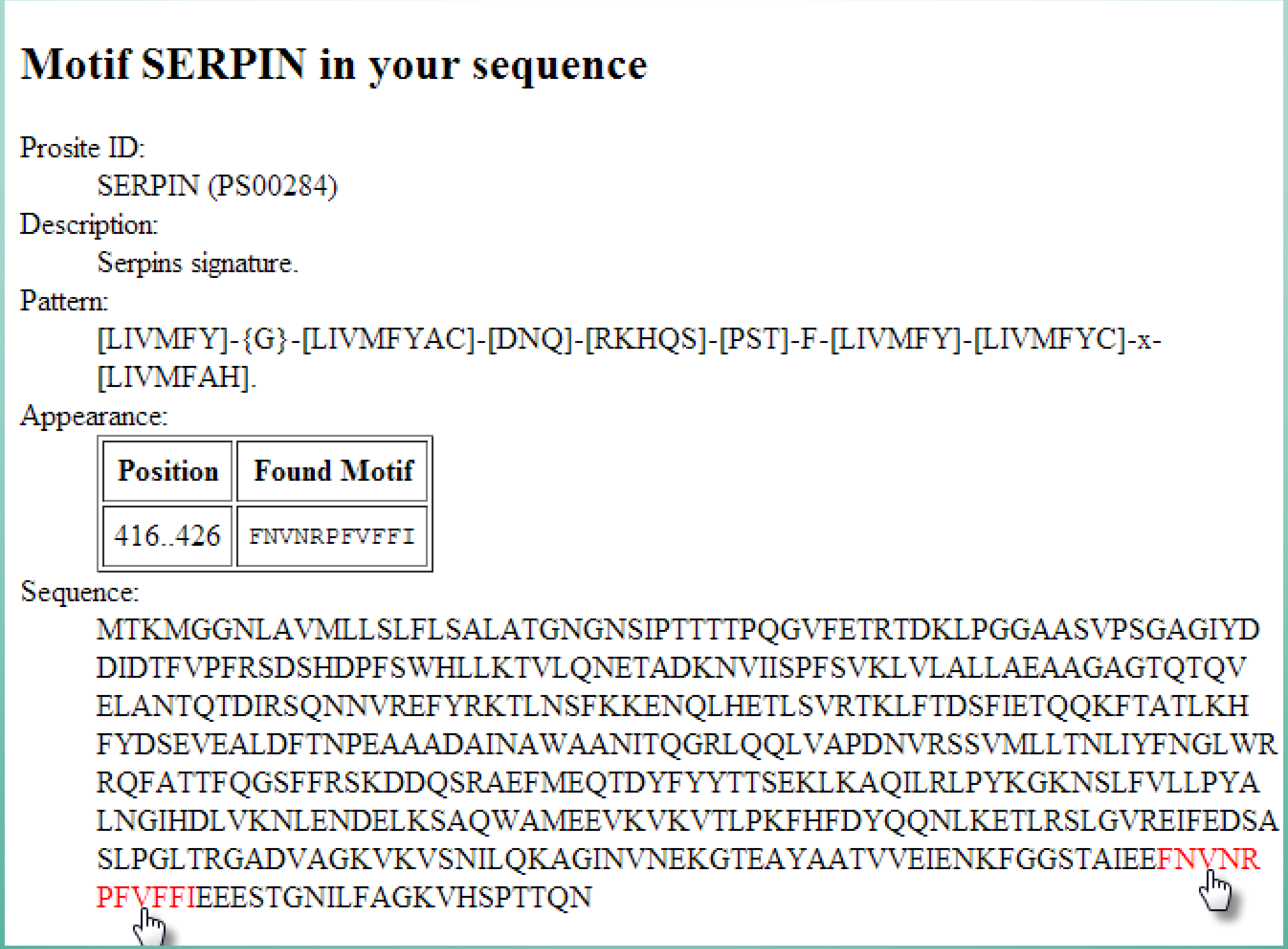
```
        MTKMGGNLAVMLLSLFLSALATGNGNSIPTTTTPQGVFETRTDKLPGGAASVPSGAGIYD
        DIDTFVPFRSDSHDPFSWHLLKTVLQNETADKNVIISPFSVKLVLALLAEAAGAGTQTQV
        ELANTQTDIRSQNNVREFYRKTLNSFKKENQLHETLSVRTKLFTDSFIETQQKFTATLKH
        FYDSEVEALDFTNPEAAADAINAWAANITQGRLQQLVAPDNVRSSVMLLTNLIYFNGLWR
        RQFATTFQGSFFRSKDDQSRAEFMEQTDYFYYTTSEKLKAQILRLPYKGKNSLFVLLPYA
        LNGIHDLVKNLENDELKSAQWAMEEVKVKVTLPKFHFDYQQNLKETLRSLGVREIFEDSA
        SLPGLTRGADVAGKVKVSNILQKAGINVNEKGTEAYAATVVEIENKFGGSTAIEEFNVNR
        PFVFFIEEESTGNILFAGKVHSPTTQN
```

Number of found motifs: 1

PROSITE PATTERN

Found Motif	Position	PROSITE	Description	Related Sequences	Related Structures
SERPIN	416..426 Detail	PS00284	Serpins signature.	232	127

PATTERN OF MATCHING MOTIF HITS



Motif SERPIN in your sequence

Prosite ID:

SERPIN (PS00284)

Description:

Serpins signature.

Pattern:

[LIVMFY]-{G}-[LIVMFYAC]-[DNQ]-[RKHQS]-[PST]-F-[LIVMFY]-[LIVMFYC]-x-[LIVMFAH].

Appearance:

Position	Found Motif
416..426	FNVNRPFVFFI

Sequence:

```
MTKMGGNLAVMLLSLFLSALATGNGNSIPTTTTPQGVFETRTDKLPGGAASVPSGAGIYD
DIDTFVPFRSDSHDPFSWHLLKTVLQNETADKNVIISPFSVKLVLALLAEAAGAGTQTQV
ELANTQTDIRSQNNVREFYRKTLNSFKKENQLHETLSVRTKLFTDSFIETQQKFTATLKH
FYDSEVEALDFTNPEAAADAINAWAANITQGRLQQLVAPDNVRSSVMLLTNLIYFNGLWR
RQFATTFQGSFFRSKDDQSRAEFMEQTDYFYYTTSEKLKAQILRLPYKGKNSLFVLLPYA
LNGIHDLVKNLENDELKSAQWAMEEVKVKVTLPKFHFDYQQNLKETLRSLGVREIFEDSA
SLPGLTRGADVAGKVKVSNILQKAGINVNEKGTEAYAATVVEIENKFGGSTAIEEFNVNR
PFVFFIEEESTGNILFAGKVHSPTTQN
```

PFAM

- The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs).
- Pfam version 27.0 was produced at the European Bioinformatics Institute using a sequence database called **Pfamseq**, which is based on UniProt.
- The descriptions of Pfam families are managed by the general public using Wikipedia.
- The Pfam database contains information about protein domains and families.
- Pfam-A is the manually curated portion of the database that contains over 10,000 entries.
- Pfam-B contains a large number of small families derived from clusters produced by an algorithm called ADDA (<u>for automatic generation</u>).
- Pfam-B families can be useful when no Pfam-A families are found (**but low quality**).

http://pfam.xfam.org/

Pfam 27.0 (March 2013, 14831 families)

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. More...

QUICK LINKS	YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...
SEQUENCE SEARCH	Analyze your protein sequence for Pfam matches
VIEW A PFAM FAMILY	View Pfam family annotation and alignments
VIEW A CLAN	See groups of related families
VIEW A SEQUENCE	Look at the domain organisation of a protein sequence
VIEW A STRUCTURE	Find the domains on a PDB structure
KEYWORD SEARCH	Query Pfam by keywords
JUMP TO	insulin Go Example Enter any type of accession or ID to jump to the page for a Pfam family or clan, UniProt sequence, PDB structure, etc. Or view the help pages for more information

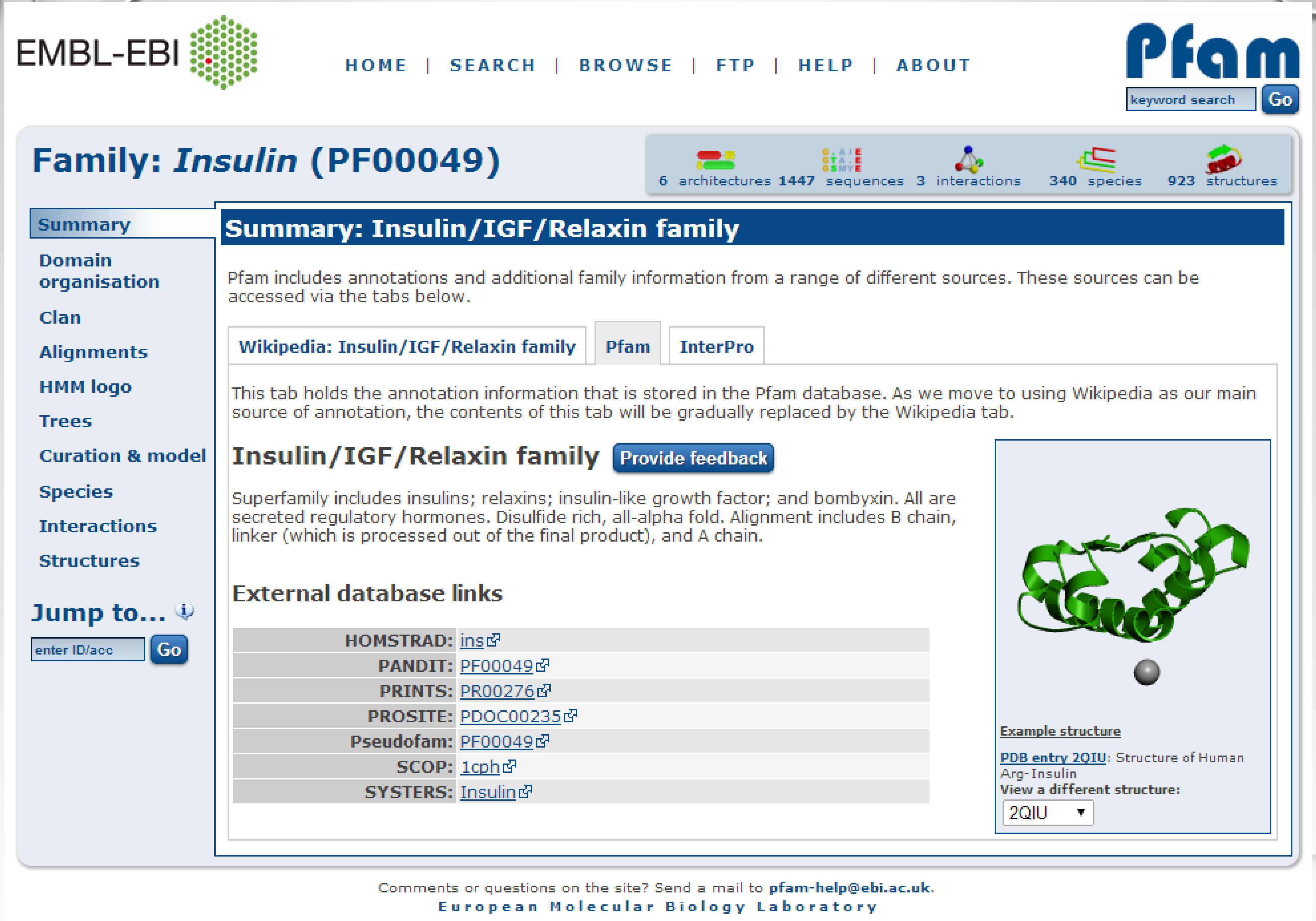
Citing Pfam

If you find Pfam useful, please consider citing the reference that describes this work:

The Pfam protein families database: R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, M. Punta **Nucleic Acids Research** (2014) Database Issue 42:D222-D230

You have hidden the blog posts section. You can restore it **here**.

Comments or questions on the site? Send a mail to **pfam-help@ebi.ac.uk**.
European Molecular Biology Laboratory



Family: *Insulin* (PF00049)

Summary: Insulin/IGF/Relaxin family

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

This tab holds the annotation information that is stored in the Pfam database. As we move to using Wikipedia as our main source of annotation, the contents of this tab will be gradually replaced by the Wikipedia tab.

Insulin/IGF/Relaxin family

Superfamily includes insulins; relaxins; insulin-like growth factor; and bombyxin. All are secreted regulatory hormones. Disulfide rich, all-alpha fold. Alignment includes B chain, linker (which is processed out of the final product), and A chain.

External database links

HOMSTRAD:	ins
PANDIT:	PF00049
PRINTS:	PR00276
PROSITE:	PDOC00235
Pseudofam:	PF00049
SCOP:	1cph
SYSTERS:	Insulin

Example structure

PDB entry 2QIU: Structure of Human Arg-Insulin

View a different structure: 2QIU

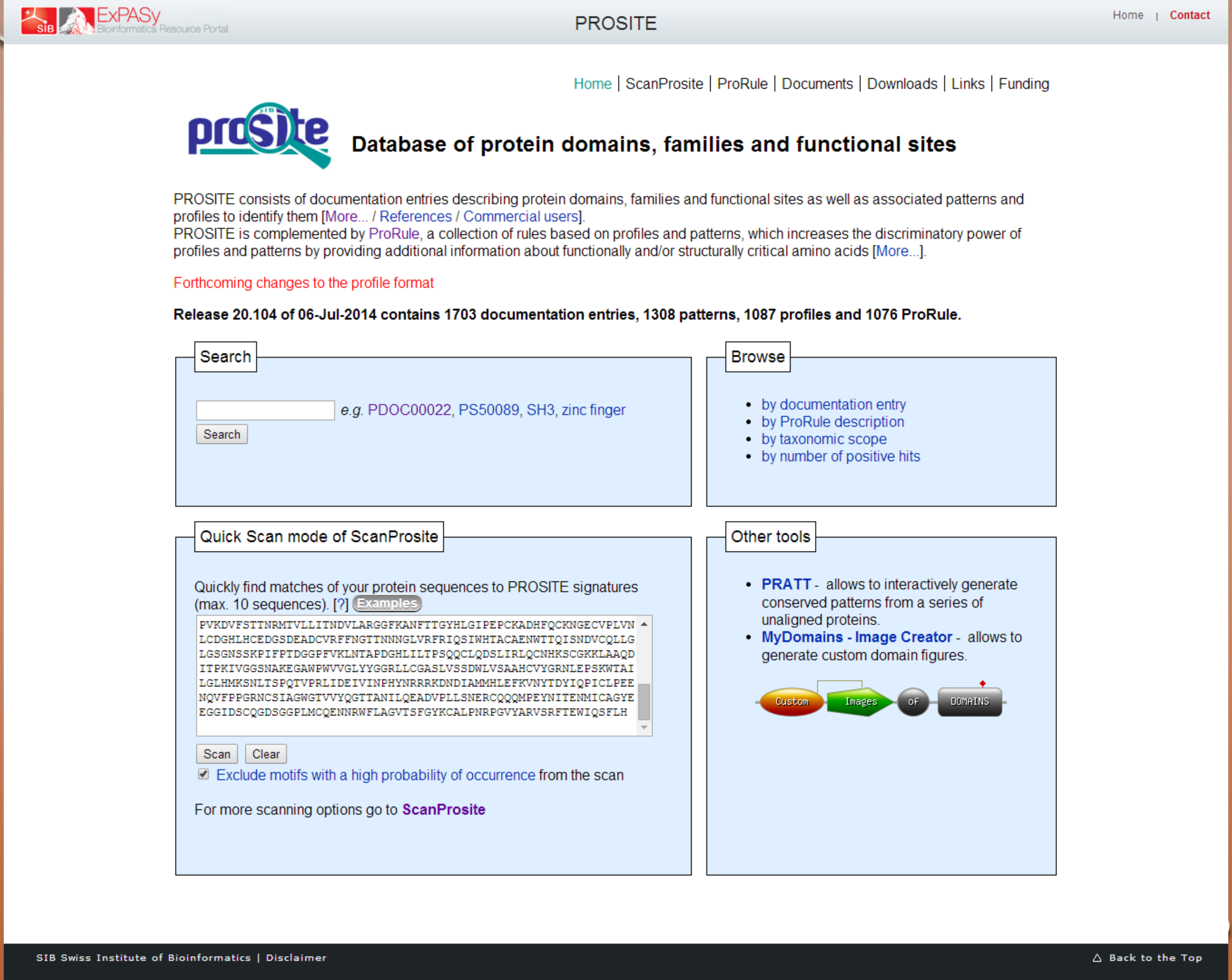
Comments or questions on the site? Send a mail to **pfam-help@ebi.ac.uk**.

European Molecular Biology Laboratory

PROSITE

- PROSITE, a protein domain database for functional characterization and annotation.
- PROSITE consists of entries describing the protein families, domains and functional sites as well as amino acid patterns and profiles in them.
- PROSITE is manually curated by a team of the Swiss Institute of Bioinformatics and tightly integrated into Swiss-Prot protein annotation.
- PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns.
- The rules contain information about biologically meaningful residues, like active sites, substrate- or co-factor-binding sites, posttranslational modification sites or disulfide bonds, to help function determination.

http://prosite.expasy.org/



Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users].
PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...].

Forthcoming changes to the profile format

Release 20.104 of 06-Jul-2014 contains 1703 documentation entries, 1308 patterns, 1087 profiles and 1076 ProRule.

Search

e.g. PDOC00022, PS50089, SH3, zinc finger

Browse

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

Quick Scan mode of ScanProsite

Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?]

```
PVKDVFSTTNRMTVLLITNDVLARGGFKANFTTGYHLGIPEPCKADHFQCKNGECVPLVN
LCDGHLHCEDGSDEADCVRFFNGTTNNNGLVRFRIQSIWHTACAENWTTQISNDVCQLLG
LGSGNSSKPIFPTDGGPFVKLNTAPDGHLILTPSQQCLQDSLIRLQCNHKSCGKKLAAQD
ITPKIVGGSNAKEGAWPWVVGLYYGGRLLCGASLVSSDWLVSAAHCVYGRNLEPSKWTAI
LGLHMKSNLTSPQTVPRLIDEIVINPHYNRRRKDNDIAMMHLEFKVNYTDYIQPICLPEE
NQVFPPGRNCSIAGWGTVVYQGTTANILQEADVPLLSNERCQQQMPEYNITENMICAGYE
EGGIDSCQGDSGGPLMCQENNRWFLAGVTSFGYKCALPNRPGVYARVSRFTEWIQSFLH
```

☑ Exclude motifs with a high probability of occurrence from the scan

For more scanning options go to **ScanProsite**

Other tools

- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** - allows to generate custom domain figures.

found: 13 hits in 1 sequence

sp-P98073-ENTK_HUMAN (1019 aa)

```
MGSKRGISSRHHSLSSYEIMFAALFAILVVLCAGLIAVSCLTIKESQRGAALGQSHEARATFKITS
GVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNSRVLQFENGSIIVVFDLFFAQWVS
DENVKEELIQGLEANKSSQLVTFHIDLNSVDILDKLTTTSHLATPGNVSIECLPGSSPCTDALTCI
KADLFCDGEVNCPDGSDEDNKMCATVCDGRFLLTGSSGSFQATHYPKPSETSVVCQWIIRVNQGLS
IKLSFDDFNTYYTDILDIYEGVGSSKILRASIWETNPGTIRIFSNQVTATFLIESDESDYVGFNAT
YTAFNSSELNNYEKINCNFEDGFCFWVQDLNDDNEWERIQGSTFSPFTGPNFDHTFGNASGFYIST
PTGPGGRQERVGLLSLPLDPTLEPACLSFWYHMYGENVHKLSINISNDQNMEKTVFQKEGNYGDNW
NYGQVTLNETVKFKVAFNAFKNKILSDIALDDISLTYGICNGSLYPEPTLVPTPPPELPTDCGGPF
ELWEPNTTFSSTNFPNSYPNLAFCVWILNAQKGKNIQLHFQEFDLENINDVVEIRDGEEADSLLLA
VYTGPGPVKDVFSTTNRMTVLLITNDVLARGGFKANFTTGYHLGIPEPCKADHFQCKNGECVPLVN
LCDGHLHCEDGSDEADCVRFFNGTTNNNGLVRFRIQSIWHTACAENWTTQISNDVCQLLGLGSGNS
SKPIFPTDGGPFVKLNTAPDGHLILTPSQQCLQDSLIRLQCNHKSCGKKLAAQDITPKIVGGSNAK
EGAWPWVVGLYYGGRLLCGASLVSSDWLVSAAHCVYGRNLEPSKWTAILGLHMKSNLTSPQTVPRL
IDEIVINPHYNRRRKDNDIAMMHLEFKVNYTDYIQPICLPEENQVFPPGRNCSIAGWGTVVYQGTT
ANILQEADVPLLSNERCQQQMPEYNITENMICAGYEEGGIDSCQGDSGGPLMCQENNRWFLAGVTS
FGYKCALPNRPGVYARVSRFTEWIQSFLH
```

Legend:

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function.
For more information about how these graphical representations are constructed, go to http://prosite.expasy.org/mydomains/.

hits by profiles: [8 hits (by 6 distinct profiles) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.

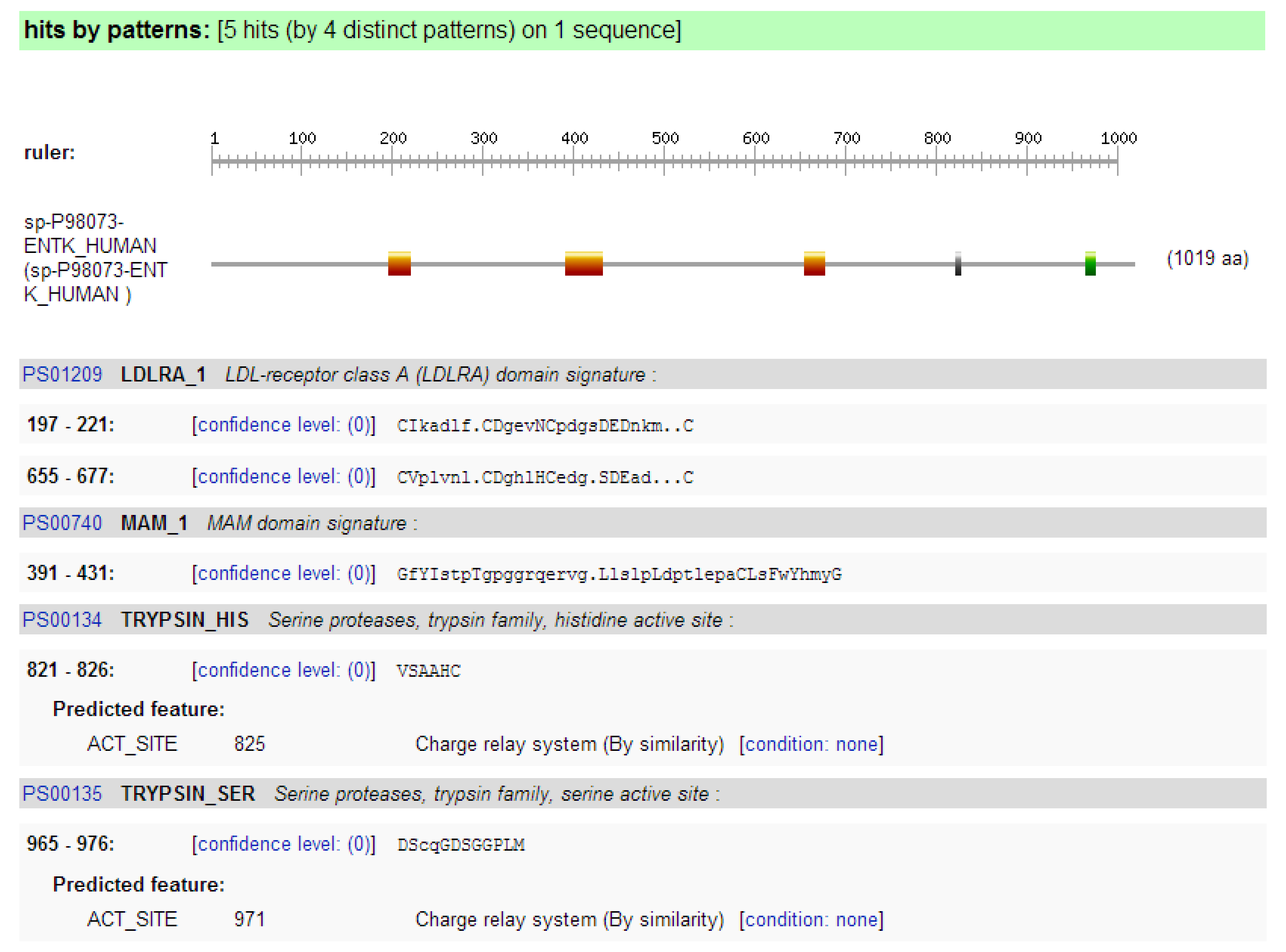
PS50024 **SEA** *SEA domain profile :*

54 - 169: **score** = 32.979

```
QSHEARATFKITsGVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNSRVLQ
FEN--GSIIVVFDLFFAQWVSD--ENVKEELIQGLEANKssQLVTFHIDLNSVDILDKLT
```

Predicted feature:

DOMAIN 54 169 SEA [condition: none]



hits by patterns: [5 hits (by 4 distinct patterns) on 1 sequence]

sp-P98073-ENTK_HUMAN (sp-P98073-ENTK_HUMAN) (1019 aa)

PS01209 **LDLRA_1** *LDL-receptor class A (LDLRA) domain signature :*

197 - 221: [confidence level: (0)] `CIkadlf.CDgevNCpdgsDEDnkm..C`

655 - 677: [confidence level: (0)] `CVplvnl.CDghlHCedg.SDEad...C`

PS00740 **MAM_1** *MAM domain signature :*

391 - 431: [confidence level: (0)] `GfYIstpTgpggrqervg.LlslpLdptlepaCLsFwYhmyG`

PS00134 **TRYPSIN_HIS** *Serine proteases, trypsin family, histidine active site :*

821 - 826: [confidence level: (0)] `VSAHC`

Predicted feature:

ACT_SITE 825 Charge relay system (By similarity) [condition: none]

PS00135 **TRYPSIN_SER** *Serine proteases, trypsin family, serine active site :*

965 - 976: [confidence level: (0)] `DScqGDSGGPLM`

Predicted feature:

ACT_SITE 971 Charge relay system (By similarity) [condition: none]

THANK YOU