Structural Databases II

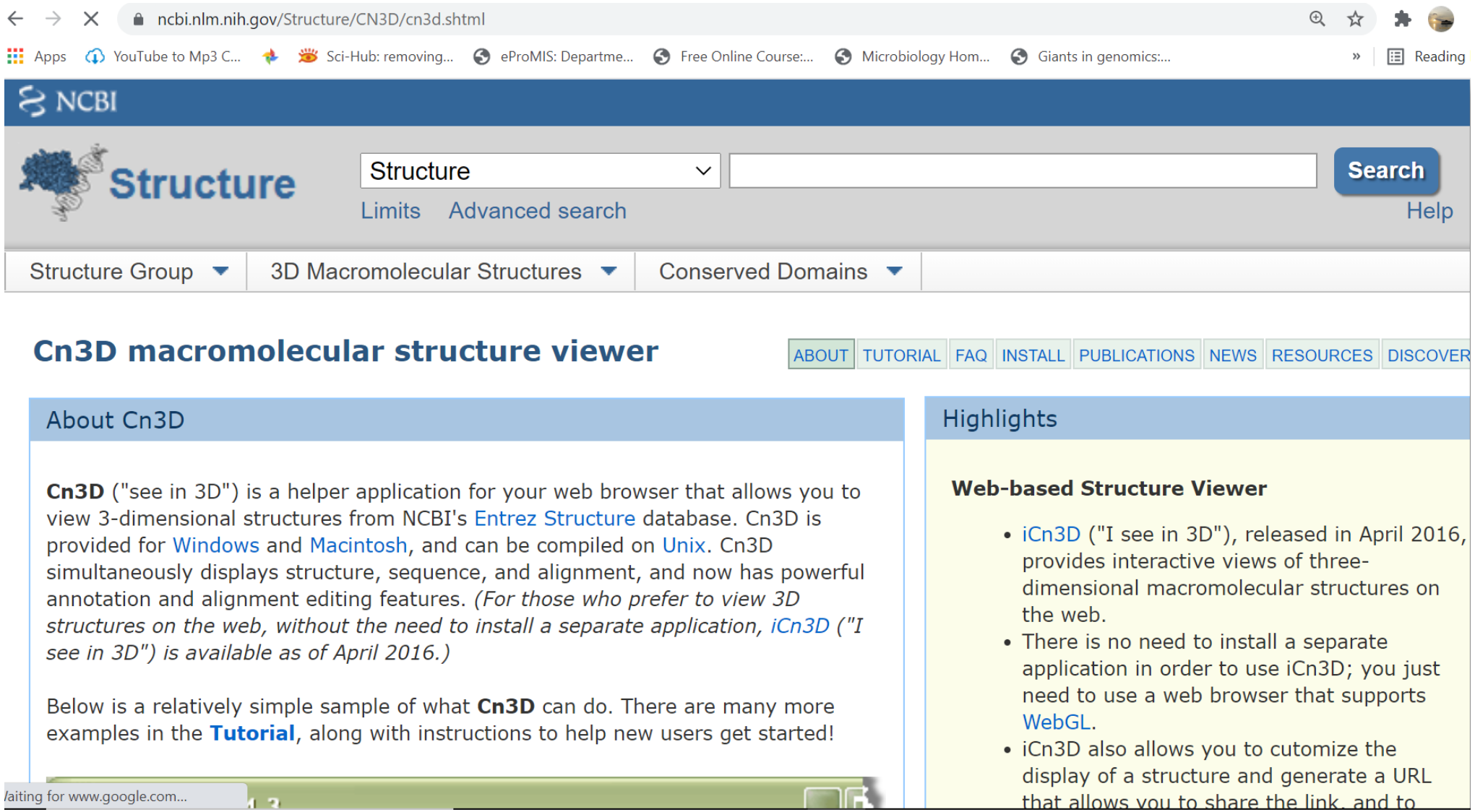
MIC 405c| Microbial Genomics & Proteomics

Dr Shilpa Kaistha

Department of Microbiology

IBSBT, CSJMU, Kanpur

Cn-3D- structure viewers



A comprehensive help document

Cn3D FAQ

Frequently Asked Questions

Cn3D Install

Installation and Configuration

MMDB

NCBI's structure database

PDBeast

Taxonomy in MMDB

VAST

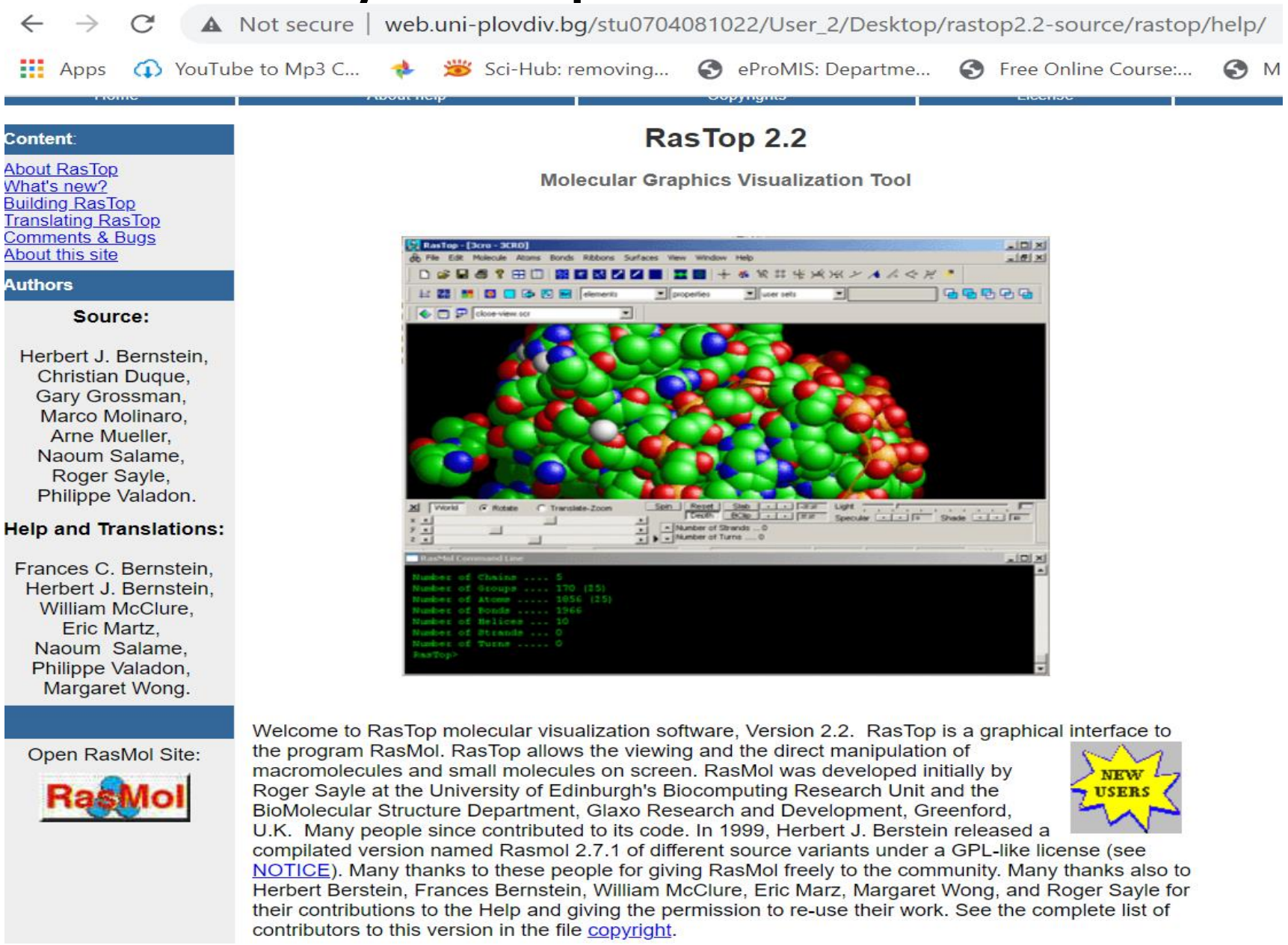
Structure comparisons

VAST Search

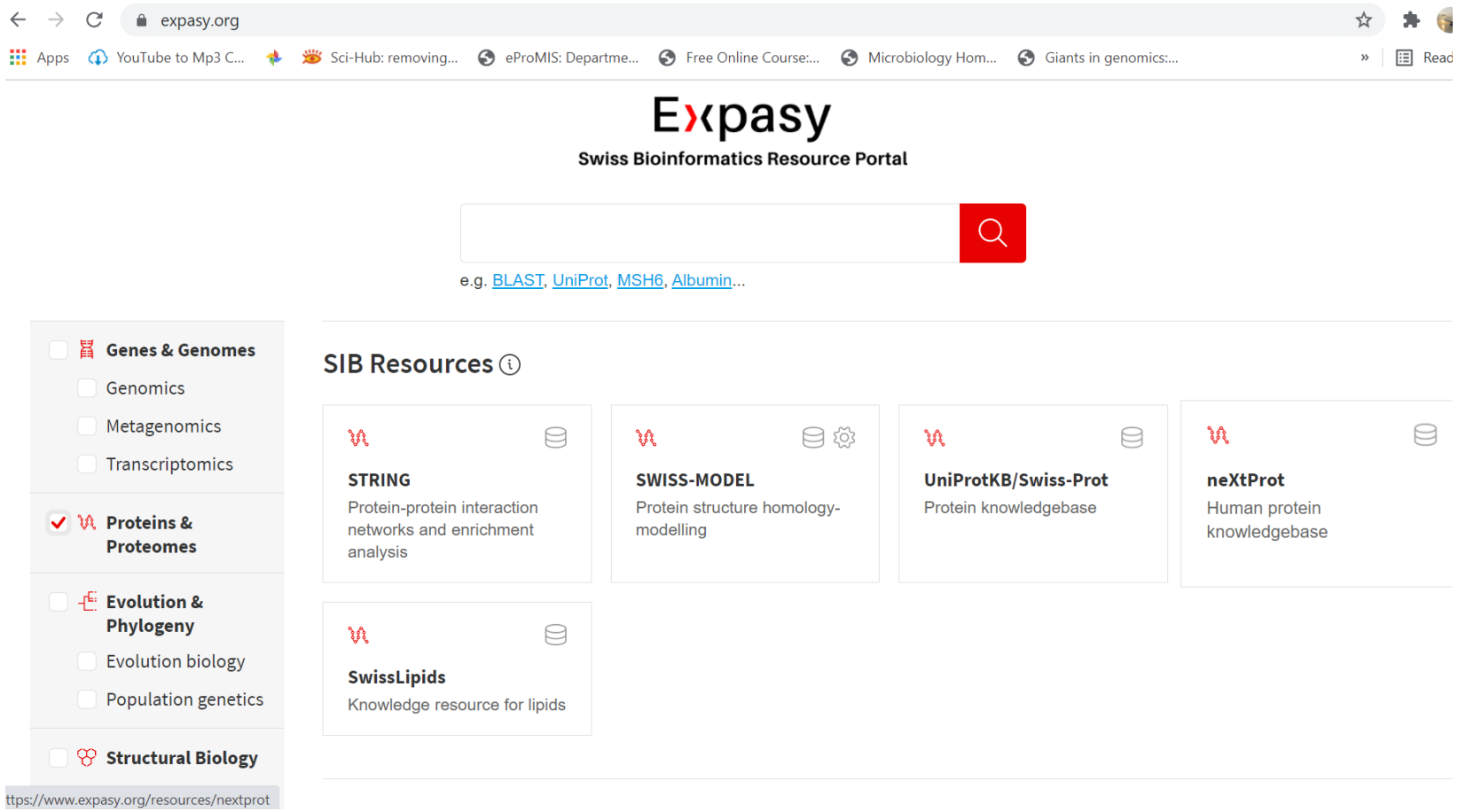
Rasmol/RasTop

RasMol is a molecular graphics program intended for the visualisation of proteins, nucleic acids and small molecules.

The program reads in a molecule coordinate file and interactively displays the molecule on the screen in a variety of colour schemes and molecule representations.
Currently available representations include depth-cued wireframes, 'Dreiding' sticks, spacefilling (CPK) spheres, ball and stick, solid and strand biomolecular ribbons, atom labels and dot surfaces.



Expert Protein Analysis System (Original): ExPASY-integrated genomics, proteomics resource of SIB



swissmodel.expasy.org

Apps | YouTube to Mp3 C... | Sci-Hub: removing... | eProMIS: Departme... | Free Online Course:... | Microbiology Hom... | Giants in genomics:... | Reading

BIOZENTRUM
University of Basel
The Center for Molecular Life Sciences

SWISS-MODEL

Modelling Repository Tools Documentation Log in Create Account

SWISS-MODEL

is a fully automated protein structure homology-modelling server, accessible via the **Expasy web server**, or from the program DeepView (Swiss Pdb-Viewer).

The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

Start Modelling

Repository

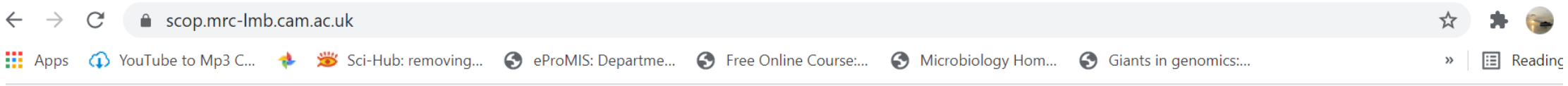
Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in **SWISS-MODEL Repository**?

Search SWISS-MODEL Repository

Selected WWW database resources for macromolecular structures.

Databases	URL
Structure and sequence/structure databases	
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/
CATH	http://www.biochem.ucl.ac.uk/bsm/cath/
FSSP	http://www2.ebi.ac.uk/dali/fssp/
Molecular Modeling Database	http://www.ncbi.nlm.nih.gov/Structure/
CAMPASS	http://www-cryst.bioc.cam.ac.uk/~campass/
ISSD	http://www.protein.bio.msu.su/issd/
Library of Protein Family Cores (LPFC)	http://WWW-SMI.Stanford.EDU/projects/helix/LPFC/
3D_ALI (a database of aligned protein structures and related sequences)	http://www.embl-heidelberg.de/argos/ali/ali_info.html
IDITIS (relational database and query tool for proteins)	http://www.oxmol.co.uk/prods/iditis/
HSSP	http://www.sander.embl-heidelberg.de/hssp/
Speciality databases	
HIV Protease Database	http://www-lbsc.ncifcrf.gov/HIVdb/
Nucleic Acid Database	http://ndbserver.rutgers.edu/
Prolysis (protease and protease inhibitor Web server)	http://delphi.phys.univ-tours.fr/Prolysis/
International Immunogenetics Database (IMGT)	http://imgt.cnusc.fr:8104/
Enzyme Structures Database	http://www.biochem.ucl.ac.uk/bsm/enzymes/
Features databases	
Molecular Movements Database	http://bioinfo.mbb.yale.edu/MolMovDB/
OLDERADO	http://neon.chem.le.ac.uk/olderado/
PROCAT	http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html
Protein Quaternary Structures (PQS)	http://pqs.ebi.ac.uk/
ReLIBase (receptor–ligand complexes database)	http://www2.ebi.ac.uk:8081/home.html
PROMISE	http://bioinf.leeds.ac.uk/promise/
PDBSum	http://www.biochem.ucl.ac.uk/bsm/pdbsum/
Biological Macromolecule Crystallization Database (BMCD)	http://[illegible]:4400/bmcd/bmcd.html
Resources	
Protein Data Bank	htttp://www.rcsb.org/pdb/

SCOP-Structural Classification of Proteins



About Contact Download

Search SCOP by text or ID

The legacy SCOP websites can be accessed at **SCOP 1.75** and **SCOP2 prototype**

SCOP 2

Learn More

SCOP: Structural Classification of Proteins

Nearly all proteins have structural similarities with other proteins and, in some of these cases, share a common evolutionary origin. The SCOP database, created by manual inspection and abetted by a battery of automated methods, aims to provide a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known. As such, it provides a broad survey of all known protein folds, detailed information about the close relatives of any particular protein, and a framework for future research and classification.

Latest update on **2021-05-27** includes **68,816** non-redundant domains representing **772,354** protein structures. Folds, superfamilies and families statistics **here**.

SCOP

- 1995 (yearly updated) Manual classification of protein structure domains
- Class- general structural architecture of protein (β sheet, α helix, membrane protein, coiled, peptide fragments, multidomain, non natural derived
- Fold- Similar arrangements of secondary structures
- Superfamily- structural and functional similarity
- Family- sequence similarity shared

classification.

Latest update on **2021-05-27** includes **68,816** non-redundant domains representing **772,354** protein structures. Folds, superfamilies and families statistics **here**.

Keyword and ID search Sequence search

Enter free text, SCOP ID, PDB ID or UniProt ID Go

Browse by structural class

- **All alpha proteins**
- **All beta proteins**
- **Alpha and beta proteins(a/b)**
- **Alpha and beta proteins(a+b)**
- **Small proteins**

Browse by protein type

- **Globular proteins**
- **Membrane proteins**
- **Fibrous proteins**
- **Non-globular/Intrinsically unstructured proteins**

Please cite: 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 **and** Antonina Andreeva, Eugene Kulesha, Julian Gough, Alexey Murzin, The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. (2020) Nucl. Acid Res., 48 (D1): D376-D382

Show ancestry ☐

Families [1 entry]

- **SARS coronavirus accessory protein X4 (ORF8, ORF7a)** SCOP ID 4007538

Domains [3 entries]

	ID	Region	Links
Protein **Protein 7a** Species ***Severe acute respiratory syndrome-related coronavirus*** **Representative domain 8055375** Represented structures [1]	**P59635** **1XAK**	15-82 A:-1-67	UniProt PDBe RCSB PDB
Protein **Protein 7a** Species ***Severe acute respiratory syndrome coronavirus 2*** **Representative domain 8092974**	**P0DTC7** **6W37**	16-81 A:1-66	UniProt PDBe RCSB PDB
Protein **Accessory protein 7a** Species ***Severe acute respiratory syndrome coronavirus 2*** **Representative domain 8102631**	**A0A6C0X2S1** **7CI3**	14-82 A:14-82	UniProt PDBe RCSB PDB

https://scop.mrc-lmb.cam.ac.uk/term/8055374

FOLD

TIM beta/alpha-barrel

SCOP ID: 2000031

(beta-alpha)8; parallel beta-sheet barrel, closed, n=8, S=8, strand order 12345678 (anticlockwise); the first seven superfamilies have similar phosphate-binding sites

Keywords **beta-barrel** **parallel beta-sheet** **anticlockwise**

Superfamilies [34 entries]

- **(Trans)glycosidases** SCOP ID 3000313
  Families: 25
- **Metallo-dependent hydrolases** SCOP ID 3000428
  the beta-sheet barrel is similarly distorted and capped by a C-terminal helix has transition metal ions bound inside the barrel
  Families: 18
- **Xylose isomerase-like** SCOP ID 3000560
  different families share similar but non-identical metal-binding sites
  Families: 9
- **Aldolase** SCOP ID 3000445
  Common fold covers whole protein structure
  Families: 8
- **Phosphoenolpyruvate/pyruvate domain** SCOP ID 3000510

CATH:

-University College of London - Evolutionary relationships of protein domains

-1990 established

-

The four main levels of the CATH hierarchy:

#	Level	Description
1	**C**lass	the overall secondary-structure content of the domain. (Equivalent to the SCOP Class)
2	**A**rchitecture	high structural similarity but no evidence of homology. (Equivalent to the 'fold' level in SCOP)
3	**T**opology/fold	a large-scale grouping of topologies which share particular structural features
4	**H**omologous superfamily	indicative of a demonstrable evolutionary relationship. (Equivalent to SCOP superfamily)

CATH

- Hierarchical semi-automatic
- Classes derived from secondary protein structure and packing (all alpha, all beta, a mixture of alpha and beta, or little secondary structure)
- Architecture: derived from secondary structure arrangement in three-dimensional space
- Topology: information on how the secondary structure elements are connected and arranged is used
- Homology: assignments are made to the Homologous superfamily (H) level if there is good evidence that the domains are related by evolution

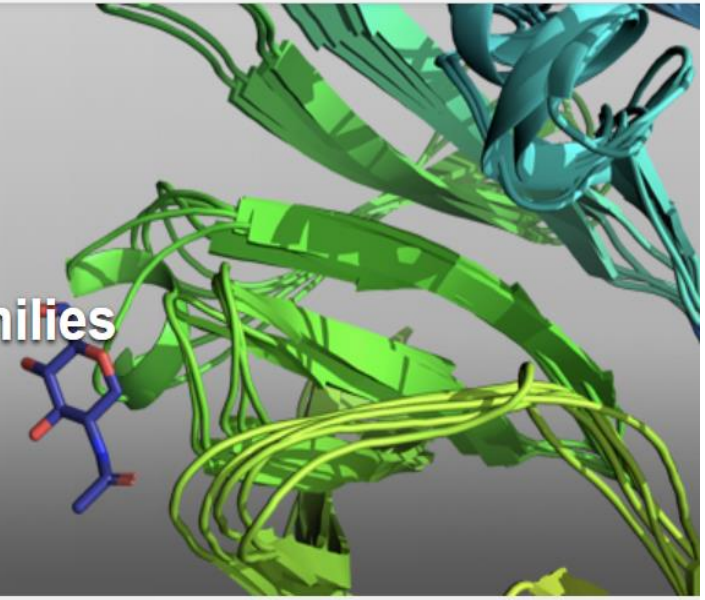
cathdb.info

Apps YouTube to Mp3 C... Sci-Hub: removing... eProMIS: Departme... Free Online Course:... Microbiology Hom... Giants in genomics:... Readin

C A T H Home Search Browse Download About Support Search CATH by keywords or ID

CATH / Gene3D v4.3

151 million protein domains classified into 5,481 superfamilies



Search by keywords, PDB code, GO term, etc Search

22-23 July 2020 The CATH website experienced some technical issues during this period as a result of a power outage. Everything should now be working as expected now - apologies for the inconvenience.

Core classification files for the latest version of CATH-Plus (v4.3) are now available to download. Daily updates of our very latest classifications are also available.

cathdb.info

Apps | YouTube to Mp3 C... | Sci-Hub: removing... | eProMIS: Departme... | Free Online Course:... | Microbiology Hom... | Giants in genomics:... | Reading list

CATH | Home | Search | Browse | Download | About | Support | Search CATH by keywords or ID

3D Structure

Find out what 3D structure your protein adopts

Find out more | Go

Protein Evolution

Learn about a particular protein family and how it evolved

Find out more

Protein Function

Investigate the function of your protein

Find out more | Go

Conserved Sites

Look at protein sites that are highly conserved and implicated in function

Find out more | Go

Download Data

Download data files and query CATH via webservices

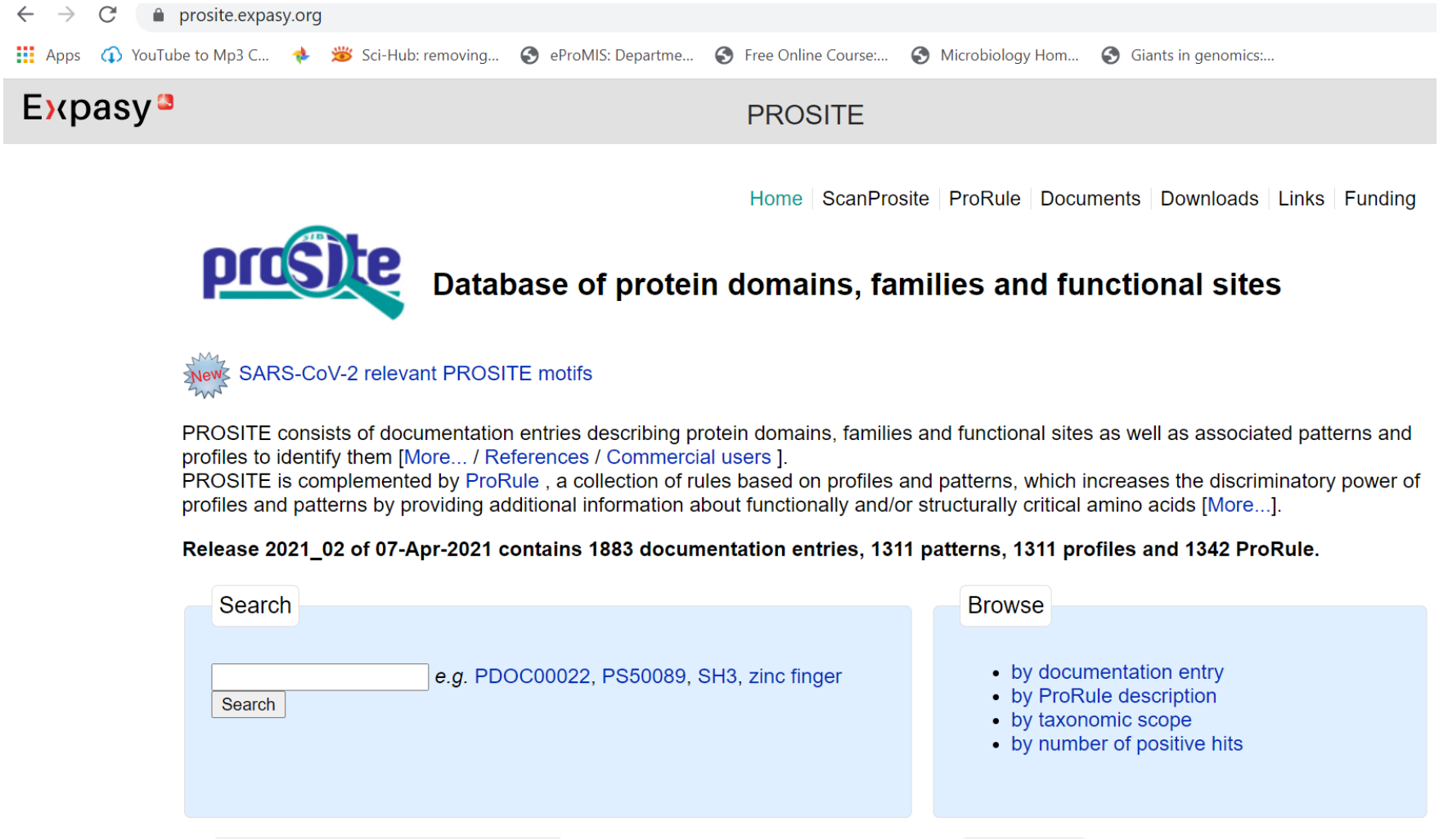
Go

Learn more

Find out how CATH is created and maintained, how to link to CATH and more

Go

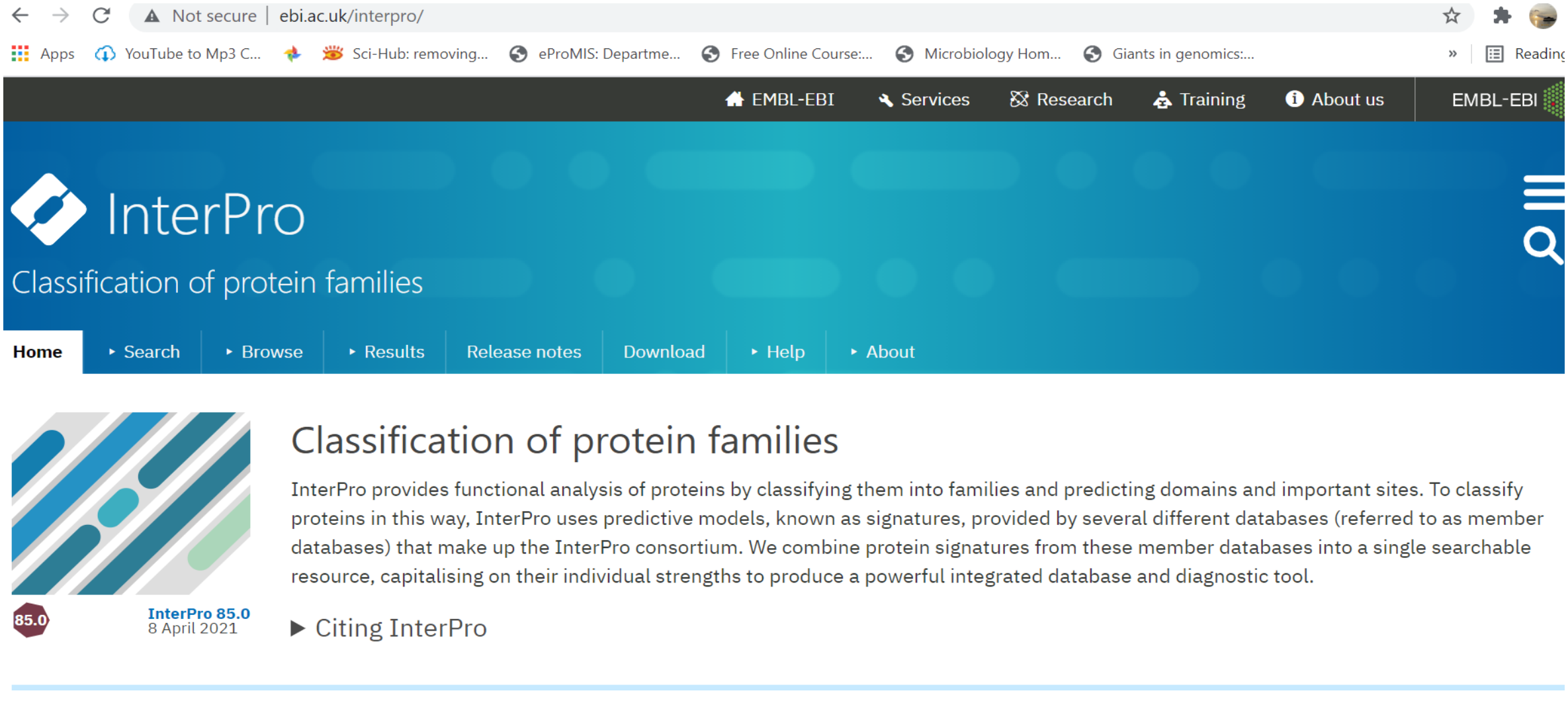
Specialized Protein Database



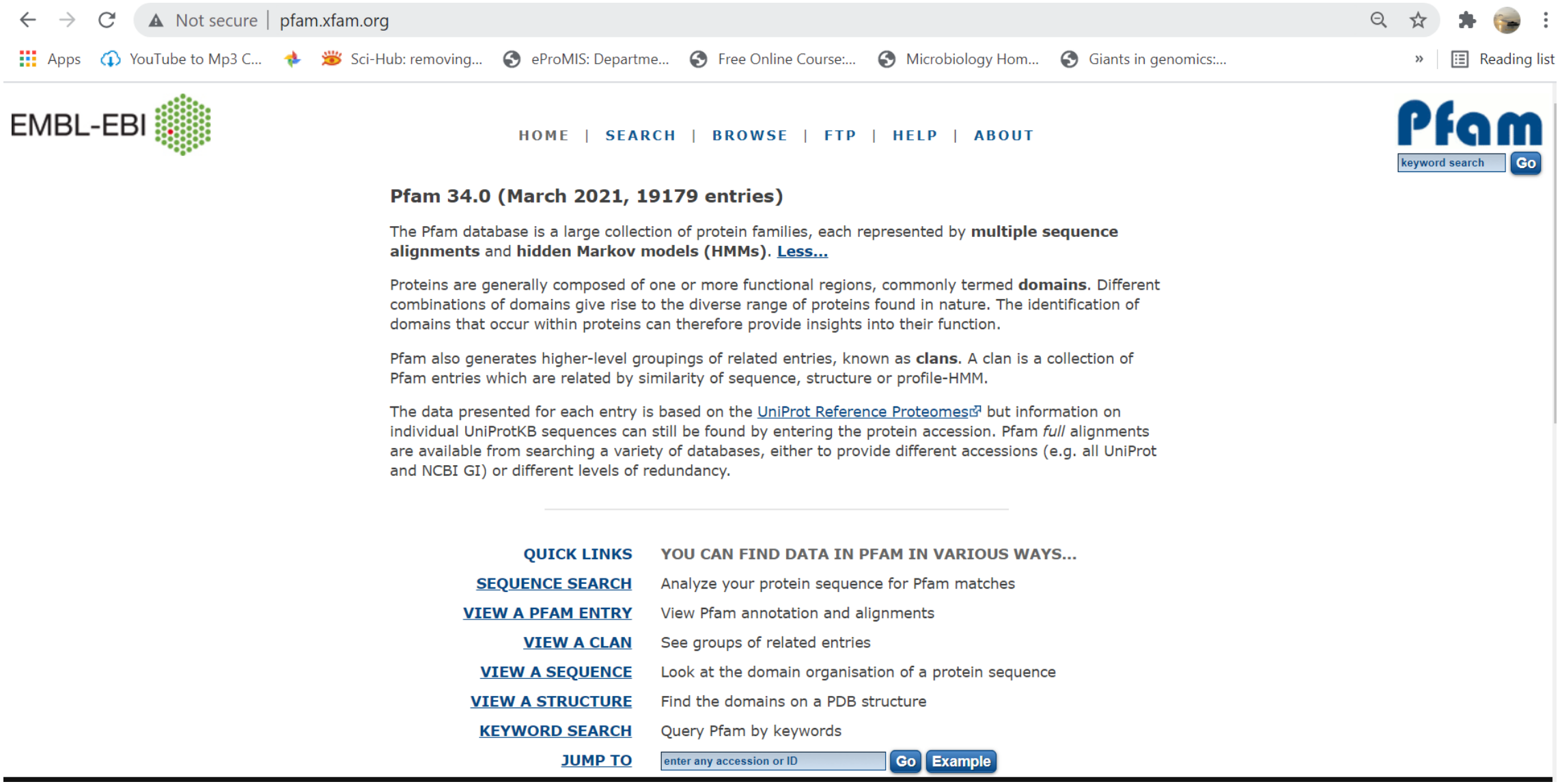
PROSITE

- 1988 by Amos Bairoch
- It is part of the ExPASy proteomics analysis servers Hosted by SIB
- PROSITE offers tools for protein sequence analysis and motif detection (see sequence motif, PROSITE patterns)

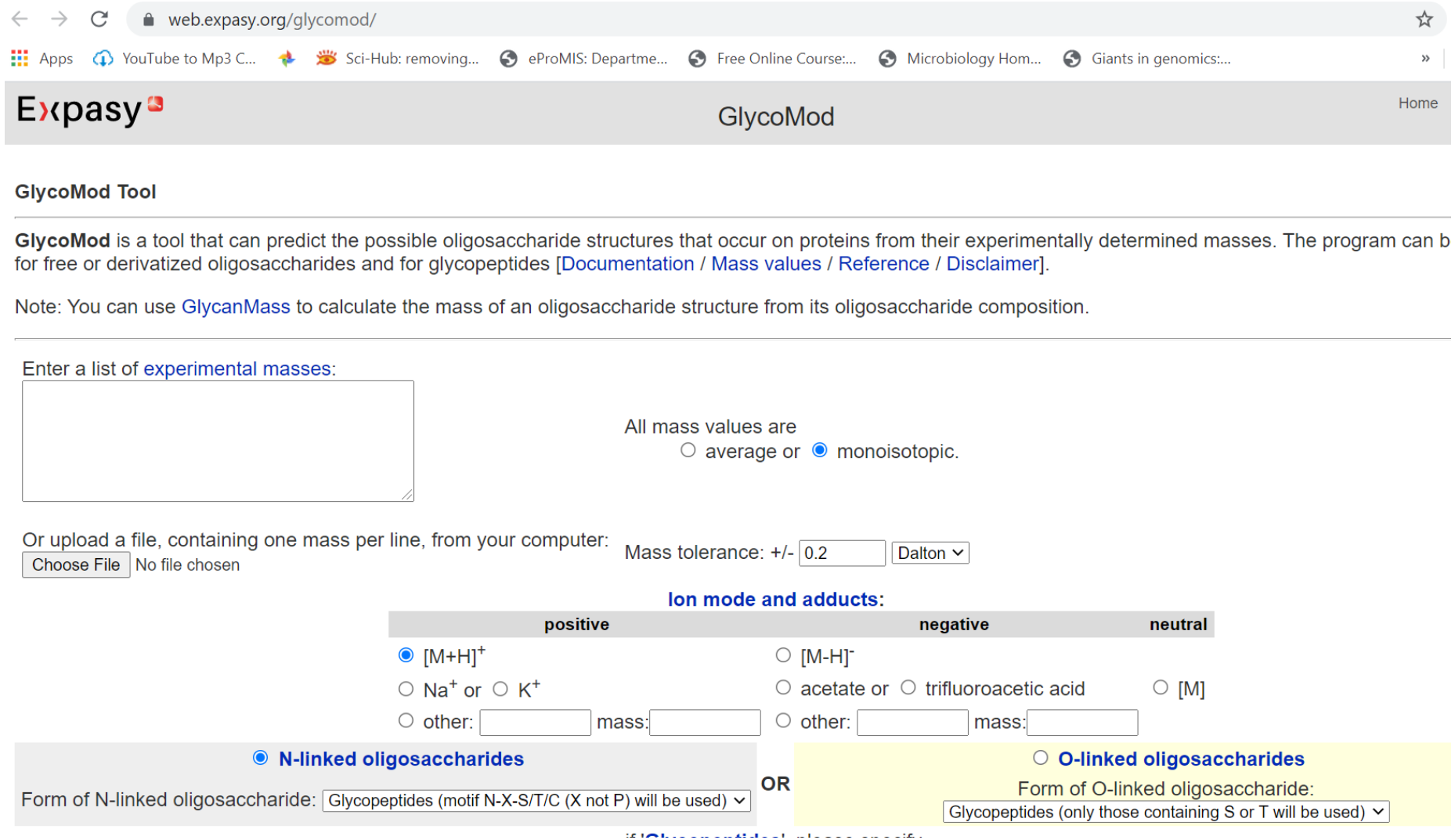
Interpro- functionally analyzes protein sequences and classifies them into protein families while predicting the presence of domains and functional sites.



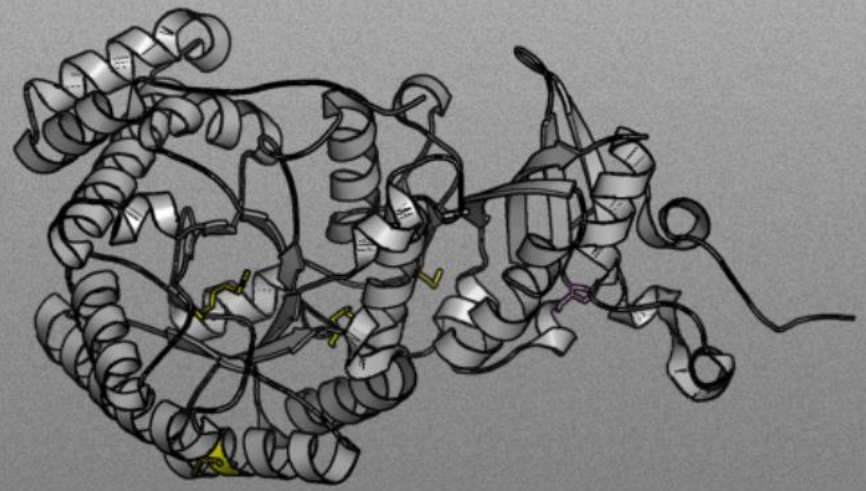
PFAM


Not secure | pfam.xfam.org
Apps
YouTube to Mp3 C...
Sci-Hub: removing...
eProMIS: Departme...
Free Online Course:...
Microbiology Hom...
Giants in genomics:...
Reading list
EMBL-EBI
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT
Pfam
keyword search
Go
Pfam 34.0 (March 2021, 19179 entries)
The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs). Less...
Proteins are generally composed of one or more functional regions, commonly termed domains. Different combinations of domains give rise to the diverse range of proteins found in nature. The identification of domains that occur within proteins can therefore provide insights into their function.
Pfam also generates higher-level groupings of related entries, known as clans. A clan is a collection of Pfam entries which are related by similarity of sequence, structure or profile-HMM.
The data presented for each entry is based on the UniProt Reference Proteomes but information on individual UniProtKB sequences can still be found by entering the protein accession. Pfam full alignments are available from searching a variety of databases, either to provide different accessions (e.g. all UniProt and NCBI GI) or different levels of redundancy.
QUICK LINKS
YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...
SEQUENCE SEARCH
Analyze your protein sequence for Pfam matches
VIEW A PFAM ENTRY
View Pfam annotation and alignments
VIEW A CLAN
See groups of related entries
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
enter any accession or ID
Go
Example

Post translational modifications



PTM STRUCTURAL DATABASE



Hydroxylation on P-104 in P00877

Welcome to the Post Translational Modification Structural Database (PTM-SD).

Proteins are composed of a succession of amino acid residues. Their 3D structures are the support of major crucial biological functions.

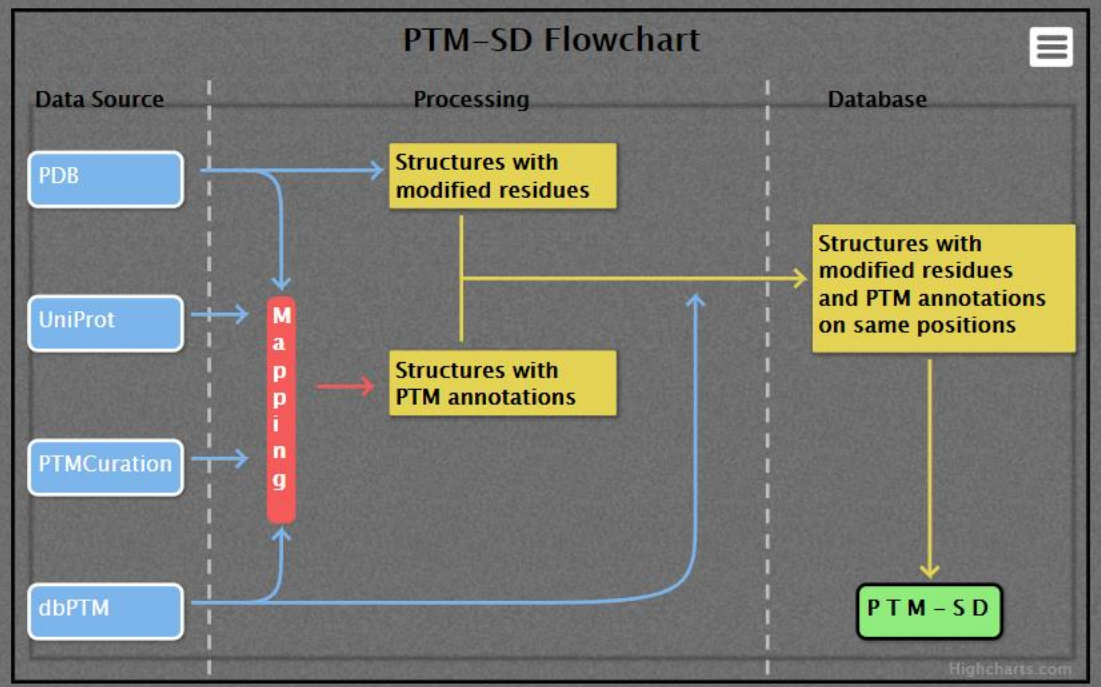
Post translational modifications (PTMs) are covalent chemical modifications of proteins that typically occur after the protein synthesis.

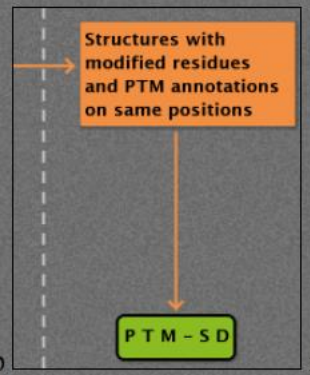
They allow a fine and sometimes obligatory modulation of biological functions; they can determine proteins activity state, localization, turnover, and interactions with other proteins.

PTM-SD provides an access to proteins for which PTMs are both experimentally annotated and structurally resolved.

Read more...

PTM STRUCTURAL DATABASE





For the last step it was necessary
to verify if the PTM annotations correspond to the chemical structure of the modified r

We use automatic and manual verification processes, which are based on the atom infor
and a correspondence annotation table.

Databases related to Proteomics

- Contain information obtained by 2D-PAGE: master images of the gels and description of identified proteins
- Examples: SWISS-2DPAGE, ECO2DBASE, Maize-2DPAGE, Sub2D, Cyano2DBase, etc.
- Format: composed of image and text files
- Mass Spectrometry (MS) database

SWISS-2DPAGE

Two-dimensional polyacrylamide gel electrophoresis database

SWISS-2DPAGE

Search by

[accession number]
[description, ID or gene]
[author names]
[spot ID / serial number]
[identification methods]
[pI / Mw range]
[combined fields]

Maps

[experimental info]
[protein list]
[graphical interface]

Select Remote Interfaces

[All Interfaces]	☐
World-2DPAGE Portal	☐
World-2DPAGE Repository	☐

☐ Exclude local DBs
has only effect if a remote interface is selected

SWISS-2DPAGE contains data on proteins identified on various 2-D PAGE and SDS-PAGE reference maps. You can locate these proteins on the 2-D PAGE maps or display the region of a 2-D PAGE map where one might expect to find a protein from UniProtKB/Swiss-Prot [More details / References / Linking to SWISS-2DPAGE / Commercial users].

Release 19.00, 23rd of May 2011, and updates up to the 9th of November 2011 (containing 1265 entries in 36 reference maps from human, mouse, *Arabidopsis thaliana*, *Dictyostelium discoideum*, *Escherichia coli*, *Saccharomyces cerevisiae*, and *Staphylococcus aureus (N315)*).

Access to SWISS-2DPAGE

- **[How to use this interface]**

- by description (any word in the ID, DE, GN and KW lines)
- by accession number (AC lines)
- by clicking on a spot: select one of our 2-D PAGE or SDS-PAGE reference maps, click on a spot and then get the corresponding information from the SWISS-2DPAGE database.
- by author (RA lines)
- by spot serial number (2D and 1D lines)
- by experimental pI/Mw range
- by experimental identification methods
- by full text search
- retrieve all the protein entries identified on a given reference map
- user defined / complex queries (SRS like)

SWISS-2DPAGE documents

- **[Facts and statistics]**
- User manual
- Release notes (September 26, 2006)
- FAQ (Frequently Asked Questions about SWISS-2DPAGE)

- **Protocols:**
  - Technical information about 2-D PAGE (IPG's, silver staining, protocols, etc)

- **Figure captions of SWISS-2DPAGE maps available from publications:**
  - Human CSF, ELC, HEPG2, HEPG2SP, LIVER, LYMPHOMA, PLASMA, PLATELET, RBC, U937, CEC, KIDNEY.
  - *Dictyostelium discoideum*, *Escherichia coli*, *Saccharomyces cerevisiae*.

Services

Software

world-2dpage.expasy.org/swiss-2dpage/viewer

Apps | YouTube to Mp3 C... | Sci-Hub: removing... | eProMIS: Departme... | Free Online Course:... | Microbiology Hom... | Giants in

ECOLI { Escherichia coli } (All identified proteins)

search engine

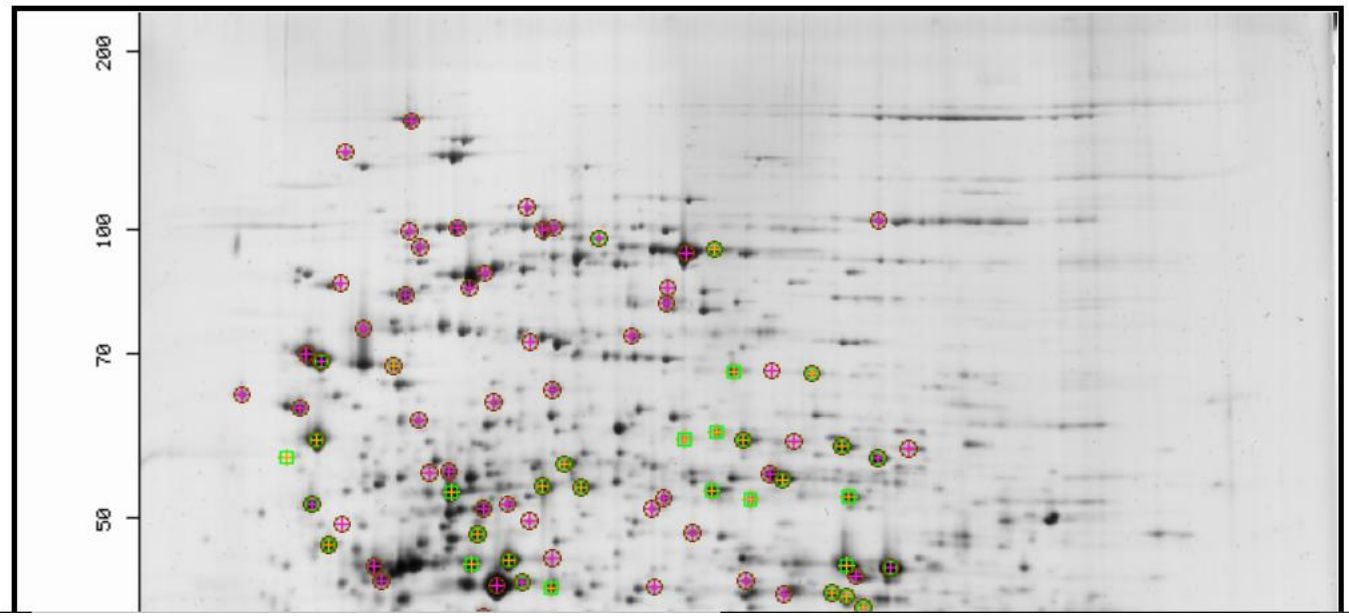
Switch to Gel: Escherichia coli -- ECOLI { Escherichia coli }

Re-scale Gel from 100% to: 100% **View:** all identified proteins

Refresh **Display:** + Identified spots **Identification details:** show hide

details: PMF Tandem MS AA Composition Micro-Sequencing / Tagging Gel Matching Comigration Immunobloting

-Get more information by dragging your mouse pointer over any highlighted spot -Click on a highlighted spot to access all its associated protein entries



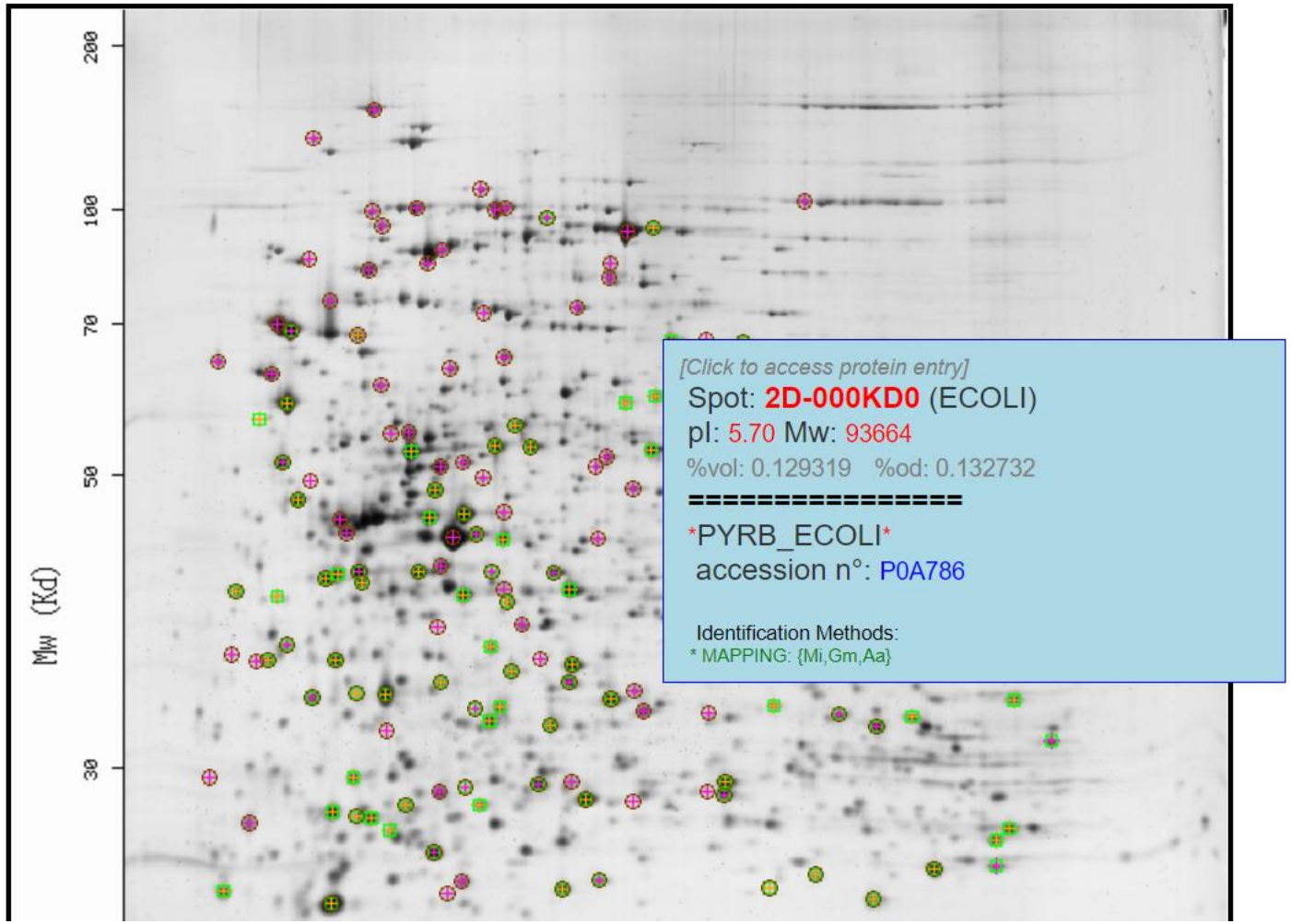
world-2dpage.expasy.org/swiss-2dpage/viewer

Apps | YouTube to Mp3 C... | Sci-Hub: removing... | eProMIS: Departme... | Free Online Course:... | Microbiology Hom... | Giants in genomics:...

Refresh **Display:** ☑ + Identified spots **Identification details:** ◉ show ○ hide

details: ☑ × PMF ☑ □ Tandem MS ☑ + AA Composition ☑ □ Micro-Sequencing / Tagging ☑ ○ Gel Matching ☑ ▫ Comigration ☑ ◦ Immunobloting

-Get more information by dragging your mouse pointer over any highlighted spot -Click on a highlighted spot to access all its associated protein entries



e.g. BLAST, UniProt, MSH6, Albumin...

MELANIE

Proteins & Proteomes Software tool

Melanie combines a unique and flexible interface for the visualization, exploration and analysis of 2D gel and blot images. The software allows differential protein expression analysis of conventional 2-DE and 2D-DIGE gels, to detect protein abundance changes for biomarker discovery. It also features dedicated functionality for Host Cell Protein (HCP) antibody coverage analysis to support development of immunoassays for HCP detection in biopharmaceuticals.

Browse the resource website

What you can do with this resource

Image analysis,
Protein quantification,
Protein property calculation,
Statistical calculation,
Differential protein expression analysis,
Expression analysis

Browse these keywords in Expasy

2D PAGE image,
Experimental design and

Melanie

The Proteome Imaging group, affiliated to the Swiss-Prot Group of SIB Swiss Institute of Bioinformatics has been developing **Melanie** since more than three decades rendering this two-dimensional electrophoresis (2-DE) gel analysis platform one of today's market leaders for analyzing, annotating and querying complex 2-D gel samples.

The latest versions of **Melanie** were developed to replace the **DeCyder™ 2D** and **ImageMaster™ 2D Platinum** applications, with a single goal in mind: help you draw more reliable conclusions from all your 2-D electrophoresis data. With its intuitive workflow and its 100% matching algorithm, analyzing your protein experiment has never been so easy.

The most recent product releases are Melanie 9 enabling our users to seamlessly perform both **differential protein expression analysis** (for single stain and DIGE gels) but also include a dedicated workflow for analysis of **Host Cell Protein (HCP) antibody Coverage analysis**. These versions provide powerful and innovative solutions to shorten the path from data acquisition to protein information. The superior spot detection and matching algorithms facilitate the extraction of statistically valid differences between groups of 2-D gels, while requiring minimal user intervention and therefore speeding up analysis time. The application integrates filtering, querying, reporting, statistical and graphing options so that you can easily view, compare, analyze and present your results.

Melanie is jointly distributed by SIB Swiss Institute of Bioinformatics and by Cytiva (formerly **GE Healthcare Life Sciences**), offering to scientists a worldwide network specialists helping them using our software wherever they are and give us a unique way to get user's day-to-day feedback to enhanced our solutions.

For more information, benefits, features, quotes, documentation and a free demo version, we invite you to visit www.2d-gel-analysis.com

Expression analysis

Host cell protein analysis