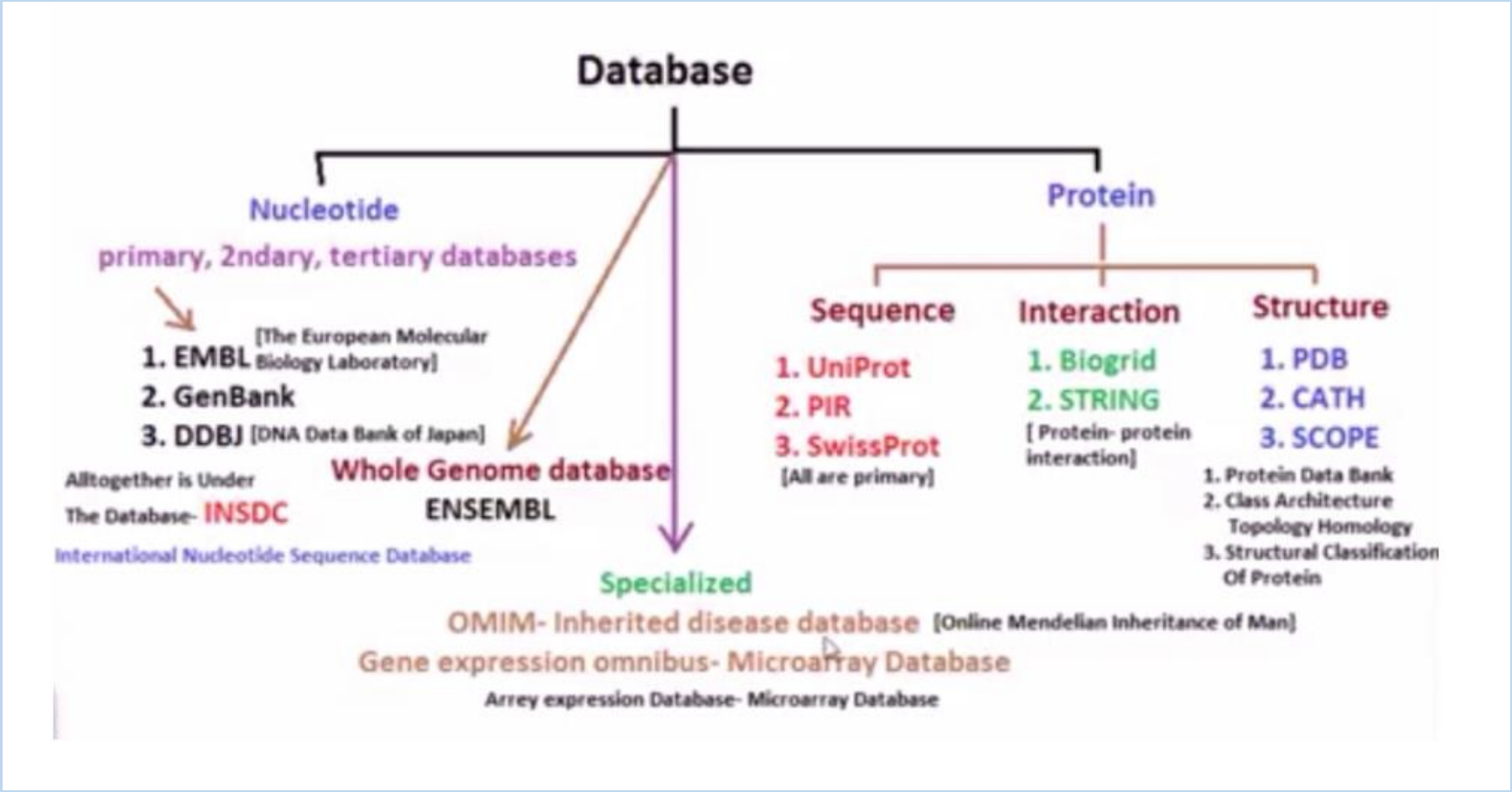
Structural Databases

MIC 405c| Microbial Genomics & Proteomics

Dr Shilpa Kaistha

Department of Microbiology

IBSBT, CSJMU, Kanpur

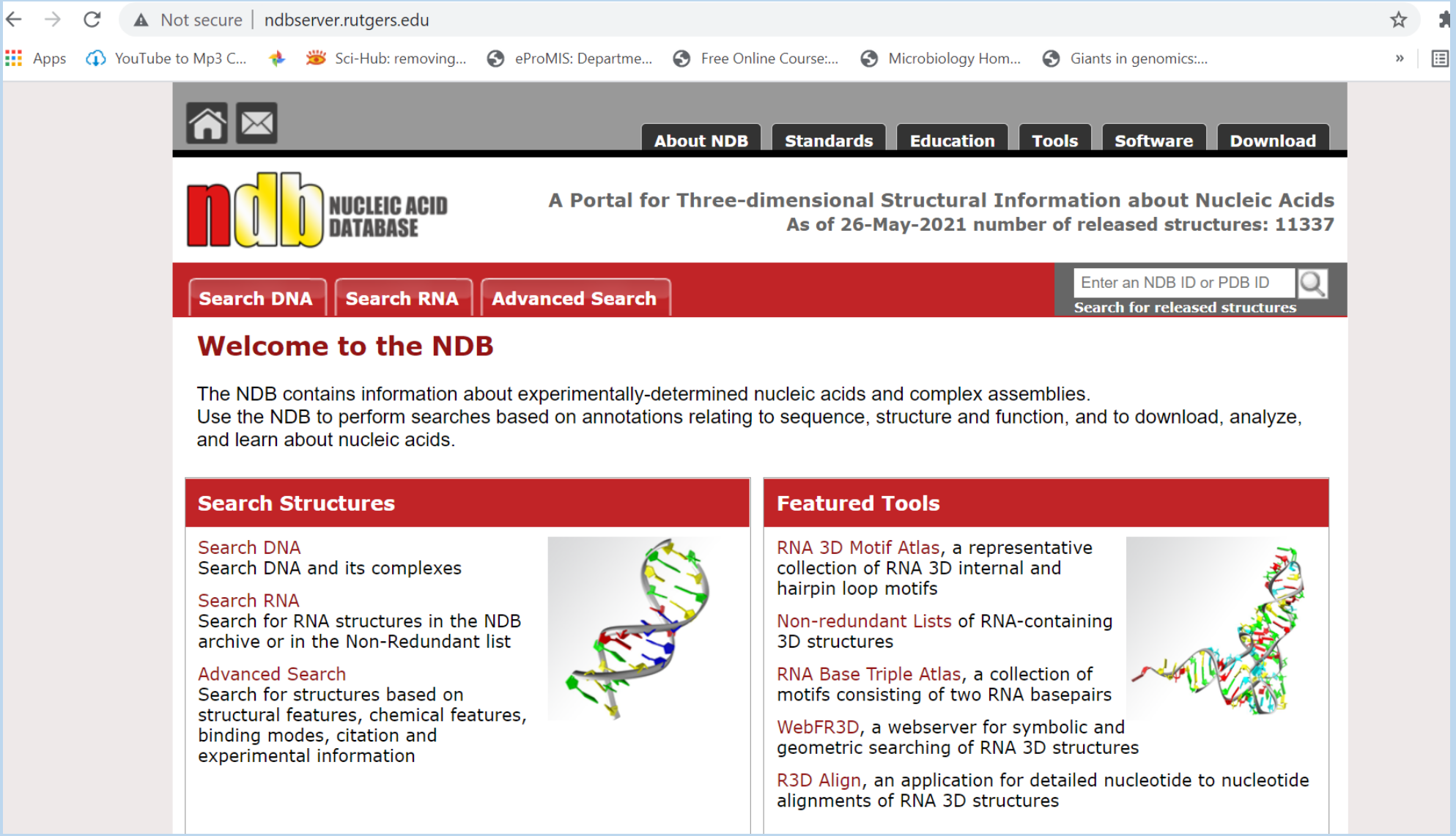


https://regiolearn.wordpress.com/2017/10/03/bioinformatic-databases/

Structure Databases

- NDB: Nucleic Acid Structure Database
- MSD:The Macromolecular Structure Database – A relational database representation of clean Protein Data Bank (PDB)
- 3DSeq: 3D sequence alignment server- Annotation of the alignments between sequence database and the PDB
- FSSP: Based on exhaustive all-against-all 3D structure comparison of protein structures currently in the Protein Data Bank (PDB)
- DALI: Fold Classification based on Structure-Structure Assignments
- 3Dee: Database of protein domain definitions wherein the domains have been clustered on sequence and structural similarity

Nucleic Acid Database



NUCLEIC ACID DATABASE

As of 26-May-2021 number of released structures: 11337

Search DNA | Search RNA | Advanced Search

Enter an NDB ID or PDB ID

Search for released structures

DNA Search Options:

Polymer

- (●) All
- ( ) DNA Only
- ( ) Protein DNA Complexes
- ( ) Drug DNA Complexes
- ( ) Hybrids and Chimera
- ( ) Peptide Nucleic Acid / Mimetics

Protein Function

- (●) All
- ( ) Enzymes
- ( ) Structural
- ( ) Regulatory

Structural Features

- (●) All
- ( ) Single Stranded
- ( ) A DNA
- ( ) B DNA
- ( ) Z DNA
- ( ) Other Double Helical Structures
- ( ) Triple helices
- ( ) Quadruple helices

Experimental Method

- (●) All

Polymer Type: All + **Protein Function:** All + **Structural Features:** All + **Experimental Method:** All

Results: 7678 Download results as an excel file Gallery view

NDB ID: 6X1J
PDB ID: 6X1J
Release: 2021-05-26

Title: The homing endonuclease I-WcaI bound to its DNA recognition sequence
Classification: HYDROLASE
Authors: Nawimanage, R., Yuan, Z., Schultz, M., Joshi, R., Lohman, J.R., Gimble, F.S.
Citation: The homing endonuclease I-WcaI bound to its DNA recognition sequence
To Be Published pp. - 0
Experiment: X-RAY DIFFRACTION **Resolution:**1.945Å **R work:**0.227 **R free:**0.261

NDB ID: 6ZPF
PDB ID: 6ZPF
Release: 2021-05-26

Title: Racemic compound of RNA duplexes.
Classification: RNA
Authors: Rypniewski, W., Kiliszek, A., Blaszczyk, L., Bejger, M.
Citation: Racemic compound of RNA duplexes.
To Be Published pp. - 0
Experiment: X-RAY DIFFRACTION **Resolution:**1.442Å **R work:**0.201 **R free:**0.249

NDB ID: 6ZQ9
PDB ID: 6ZQ9
Release: 2021-05-26

Title: Racemic compound of RNA duplexes.
Classification: RNA
Authors: Rypniewski, W., Kiliszek, A., Blaszczyk, L., Bejger, M.

NDB ID: 6X1J **PDB ID: 6X1J**

Title:

THE HOMING ENDONUCLEASE I-WCAI BOUND TO ITS DNA RECOGNITION SEQUENCE

Molecular Description:

Probable intron-encoded endonuclease 1/DNA Complex

Deposited:

2020-05-19

Released:

2021-05-26

Structural Keywords:

B DOUBLE HELIX

Nucleic Acid Sequence:

Click to show/hide 2 nucleic acid sequences

Protein Sequence:

Click to show/hide 1 protein sequences

Primary Citation:

Nawimanage, R., Yuan, Z., Schultz, M., Joshi, R., Lohman, J.R., Gimble, F.S.
The homing endonuclease I-WcaI bound to its DNA recognition sequence
To Be Published, , pp. - , 0.

Experimental Information:

X-RAY DIFFRACTION

Space Group:

P 1 2 1

Cell Constants:

a = 67.601 b = 37.281 c = 94.064 (Ångstroms)

α = 90.0 β = 100.44 γ = 90.0 (degrees)

Refinement:

he REFMAC program. The R value is 0.229 for 31986 reflections

as/xray/structures/6/6x1j/6X1J-biol1.jpg

Structural Features

RNAML

Base Pair Hydrogen Bonding Classification

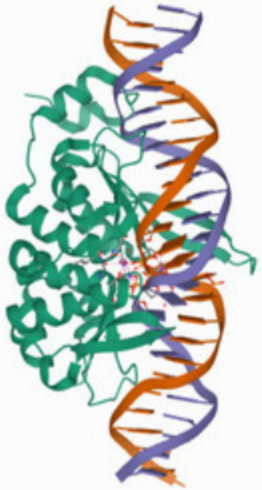
Nucleic Acid Backbone Torsions

Base Pair Morphology Parameters

Base Pair Morphology Step Parameters

Conformer Analysis (DNATCO)

Biological Assembly 1



RNA View

Protein DataBank (PDB)

- Important in solving real problems in molecular biology
- Protein Databank
  - PDB Established in 1972 at Brookhaven National Laboratory (BNL)
  - Sole international repository of macromolecular structure data
  - Moved to Research Collaboratory for Structural Bioinformatics

http://www.rcsb.org/

PDB

Repository of

- 3D structural data of protein/protein-NA complexes (atomic coordinated, chemical data
- Data from X ray crystallography, XRD, NMR, eM etc
- Submit Data: PDB AutoDep Input Tool
- Visualise
- Analyse

Welcome
Deposit
Search
Visualize
Analyze
Download
Learn

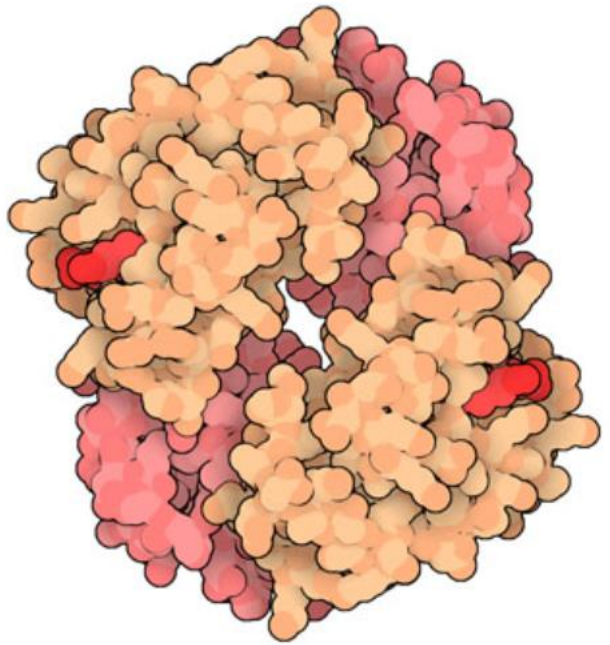
A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

May **Molecule of the Month**



Fetal Hemoglobin

178229
PDB entries from
2021-05-26

PDBj
Protein Data Bank Japan

Worldwide Protein Data Bank Foundation

English 日本語 简体中文 繁體中文 한국어

Search pdbj.org

wwPDB RCSB PDB PDBe BMRB Adv. Search Search help

7MSW

Latest new entries

- Home
  - Top Page
  - Statistics
  - Help
  - FAQ
  - Contact Us
  - Cite Us / Terms and Conditions
  - Links
  - Settings
- Data deposition (OneDep)
  - Help
  - Deposition to PDB, EMDB or BMRB
- Download

PDBj (Protein Data Bank Japan) is a project team operating under the Joint Usage and Research activities of the Institute for Protein Research, Osaka University. We maintain the single global PDB/BMRB/EMDB archives of macromolecular structures and provide integrated tools, in collaboration with RCSB PDB and BMRB in the USA, and PDBe in the EU. PDBj is supported by JST-NBDC and AMED-BINDS. PDBj's new logo design incorporates cytochrome c (PDBID: 1cyc), which was the first structure determined in Japan.

Guide for first time visitors

The material of this site is aimed mainly for researchers. For general people, please refer PDBj numon ("numon" means introduction or beginner) site.
When accessed with old web browser, the web pages may be displayed under limited function mode. In order to use the full function mode, refer prerequisite page.
For details about this site, see interactive tutorials.

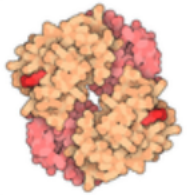
Find the service you need

Choose a keyword listed below or input keywords into the textbox at the right of the keyword list. The brief explanation of the matched services will be displayed.

- Click the 'Show all services' button to display the explanation for all services.
- Input some keywords into the 'Word Search Box' to narrow down the search results.

Molecule of the Month

Fetal Hemoglobin



Article List

PROTEIN DATA BANK 50 Years

Hot Structural News on COVID-19

New PDBe-KB COVID-19 Data Portal

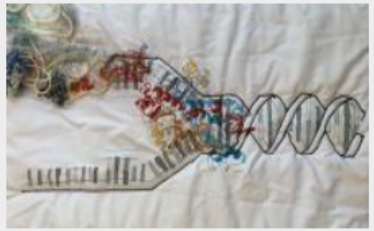
Our new PDBe-KB COVID-19 data portal brings together all available PDB data from SARS-CoV-2 structures, to help researchers easily identify important structural features to support the development of treatments and vaccines.

Visit the PDBe-KB COVID-19 data portal

PDBe is the European resource for the collection, organisation and dissemination of data on biological macromolecular structures. Read more about PDBe.

Featured structure

The unzipping enzyme



01 May 2021

There is an enzyme that plays a role in all cellular DNA transactions, including DNA replication and repair, transcription, translation, ribosome synthesis, RNA maturation and splicing, and nuclear export processes.

Unwind and read more...

Popular

- PDBe-KB
- EMsearch
- PDBeFold
- PDBePISA
- PDBeChem
- PDBe REST API
- EM resources
- NMR resources
- EMPIAR
- Coordinate Server
- PDB Component Library
- News
- Events
- Training
- Contact us

Latest archive statistics

As of 26 May 2021 the PDB contains 178229 entries (latest PDB entries, chemistry, biology) and EMDB contains 15243 entries (latest map releases, latest updates).

Welcome
Deposit
Search
Visualize
Analyze
Download
Learn

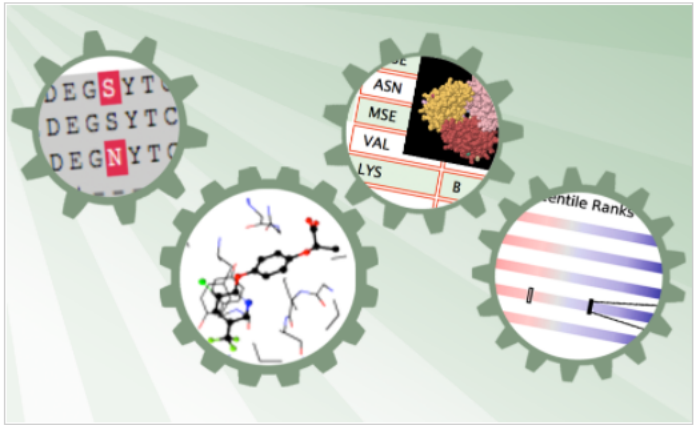
Deposit Options
Prepare Data
Validate Data
Deposit Data
Deposition Help

Documentation
PDBx/mmCIF Dictionary Resources
Chemical Component Dictionary
Biologically Interesting Molecule Reference Dictionary (BIRD)
PDB Format Guide

Deposition Preparation Tools

Data Extraction

- pdb_extract: Extract and harvest data in PDBx/mmCIF format from structure determination programs
- SF-Tool: Convert structure factor files among various formats



Small Molecules

- Ligand Expo: Search the Chemical Component Dictionary for the IDs of released ligands

Data Format Conversion

- PDBML2CIF: Convert PDBML-format data into PDBx/mmCIF-format
- PointSuite: Generate symmetry records for macromolecular assemblies with point and helical symmetries
- MAXIT: Translate data between file formats and more

Effective use of PDB

- Queries are of three types
  - PDBid - As quoted in paper
  - Search Lite - one or more keywords
  - Search Fields - A detailed query form
- Query results
  - Structure Explorer - details of the structure
  - Query Result Browser - for multiple structures
- PDB Viewer

Welcome
Deposit
Search
Visualize
Analyze
Download
Learn

Viewers
3D Structure Viewers
Sequence
Protein Feature View
Genome View

Mol* 3D Viewer

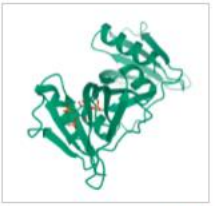
Launch Mol* from the Structure Summary page for any entry.

A standalone version of Mol* is available. Users can upload their custom files into this tool.

A detailed Mol* User Guide is available.

PDB ID: 4HHB

PDB ID: 3J3Q



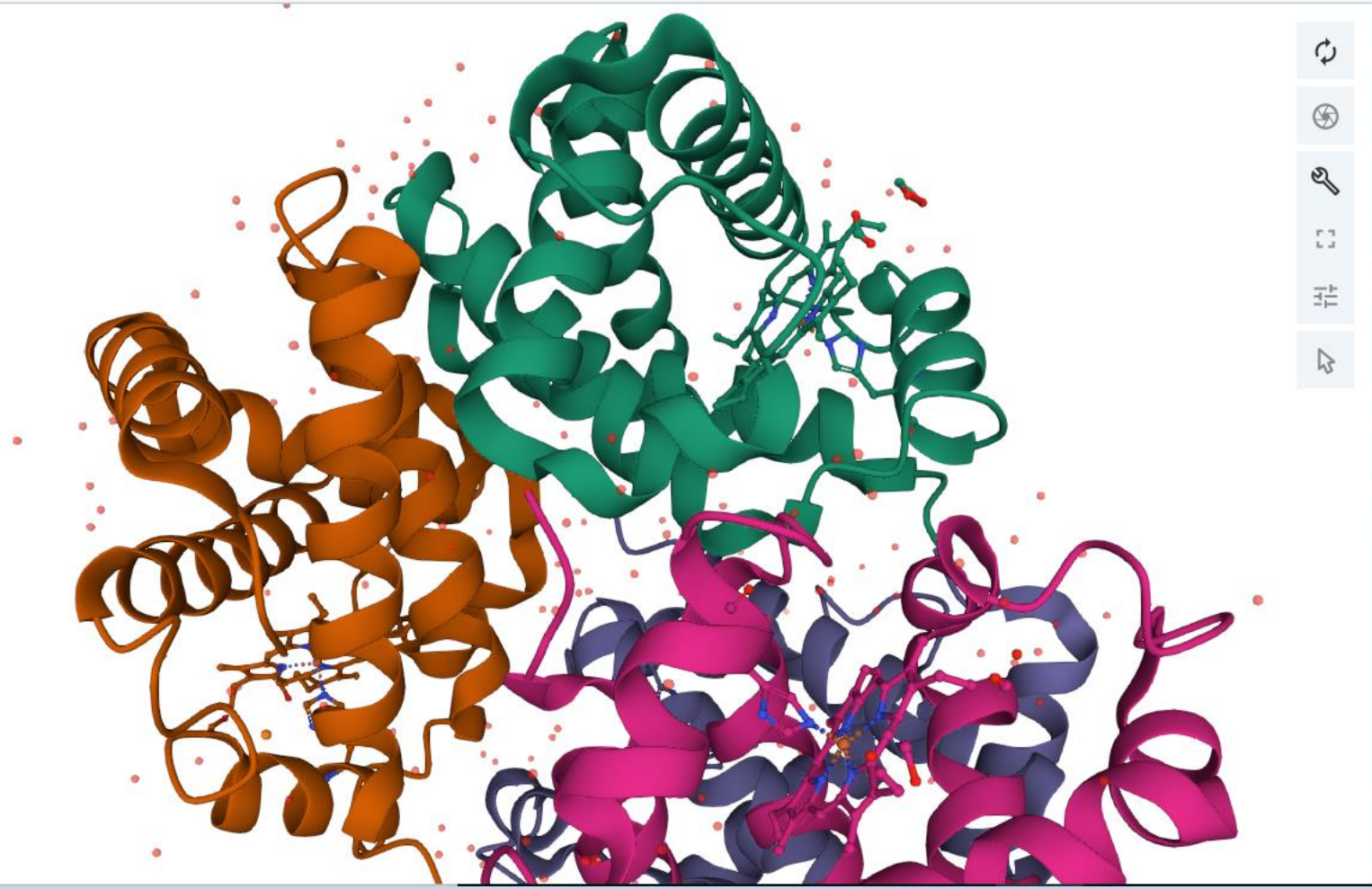
PDB ID: 1BTN

rcsb.org/3d-view/4hhb/1

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

RCSB PDB | Deposit | Search | Visualize | Analyze | Download | Learn | More | Documentation | MyPDB

```
1           11          21          31          41          51          61          71          81          91          101         111
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP
        121         131         141
AEFTPAVHASLDKFLASVSTVLTSKYR
```



4HHB | THE CRYSTAL STRUCTURE OF...

Type	Assembly
Asm Id	1: Author And Software ...

Nothing Focused

Measurements

Structural Motif Search

Submit Search (3 selections required)

Add one or more selections (toggle mode)

Components 4HHB

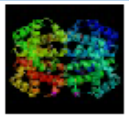
Preset | + Add

Polymer	Cartoon	
Ligand	Ball & Stick	
Water	Ball & Stick	
Ion	Ball & Stick	

Unit Cell P 1 21 1

PDB
PROTEIN DATA BANK

Structure Explorer - 4HHB



Summary Information

Summary Information

View Structure

Download/Display File

Structural Neighbors

Geometry

Other Sources

Sequence Details

Crystallization Info

Previous version(s): 1HHB

Explore SearchLite SearchFields

Compound: **Hemoglobin (Deoxy)**

Authors: **G. Fermi, M. F. Perutz**

Exp. Method: **X-ray Diffraction**

Classification: **Oxygen Transport**

Source: **Homo Sapiens**

Primary Citation: **Fermi, G., Perutz, M. F., Shaanan, B., Fourme, R.: The crystal structure of human deoxyhaemoglobin at 1.74 A resolution. *J Mol Biol* 175 *pp.* 159 (1984)**
[Medline]

Deposition Date: **07-Mar-1984** *Release Date:* **17-Jul-1984**

Resolution [Å]: **1.74** *R-Value:* **0.135**

Space Group: **P 21**

Unit Cell: *dim [Å]:* *a* **63.15** *b* **83.59** *c* **53.80**
angles [°]: *alpha* **90.00** *beta* **99.34** *gamma* **90.00**

Polymer Chains: **A, B, C, D** *Residues:* **574**

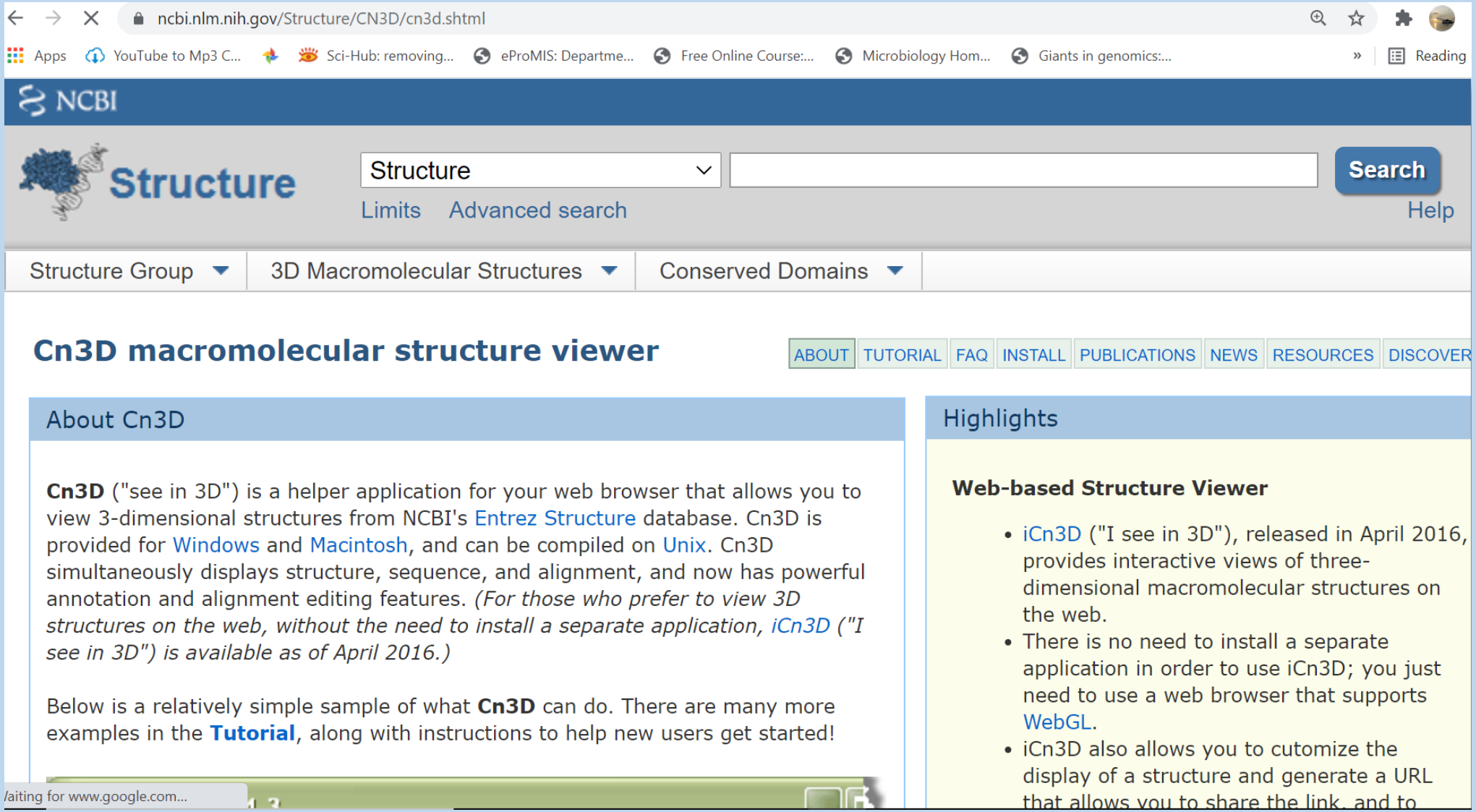
Atoms: **4779**

HET groups:

ID	*Name*	*Formula*
HEM	PROTOPORPHYRIN IX CONTAINING FE	$C_{34}H_{32}N_4O_4FE_1$
PO4	PHOSPHATE ION	O_4P_1

Other Versions: **2HHB, 3HHB**

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

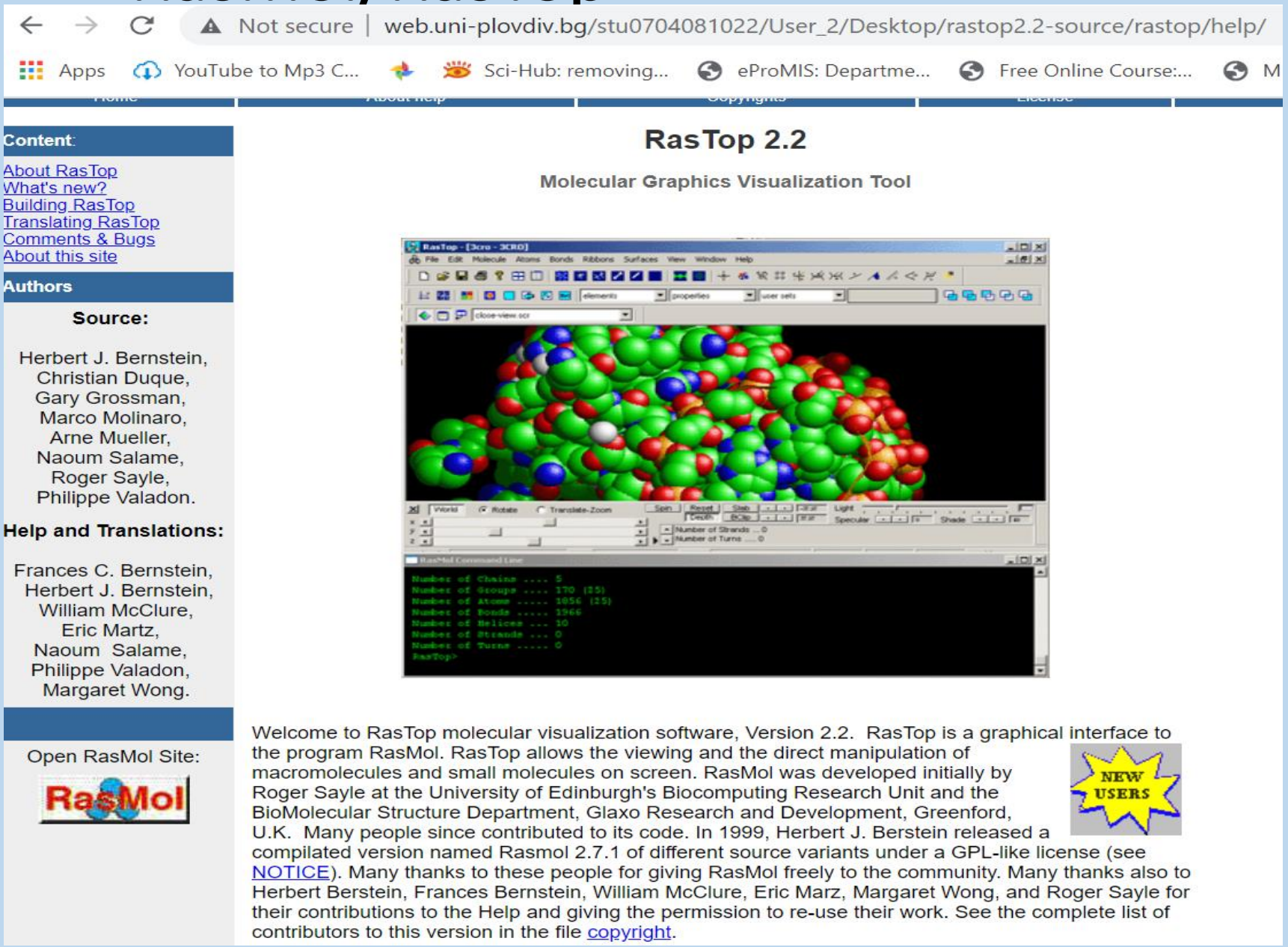
Structure comparisons

VAST Search

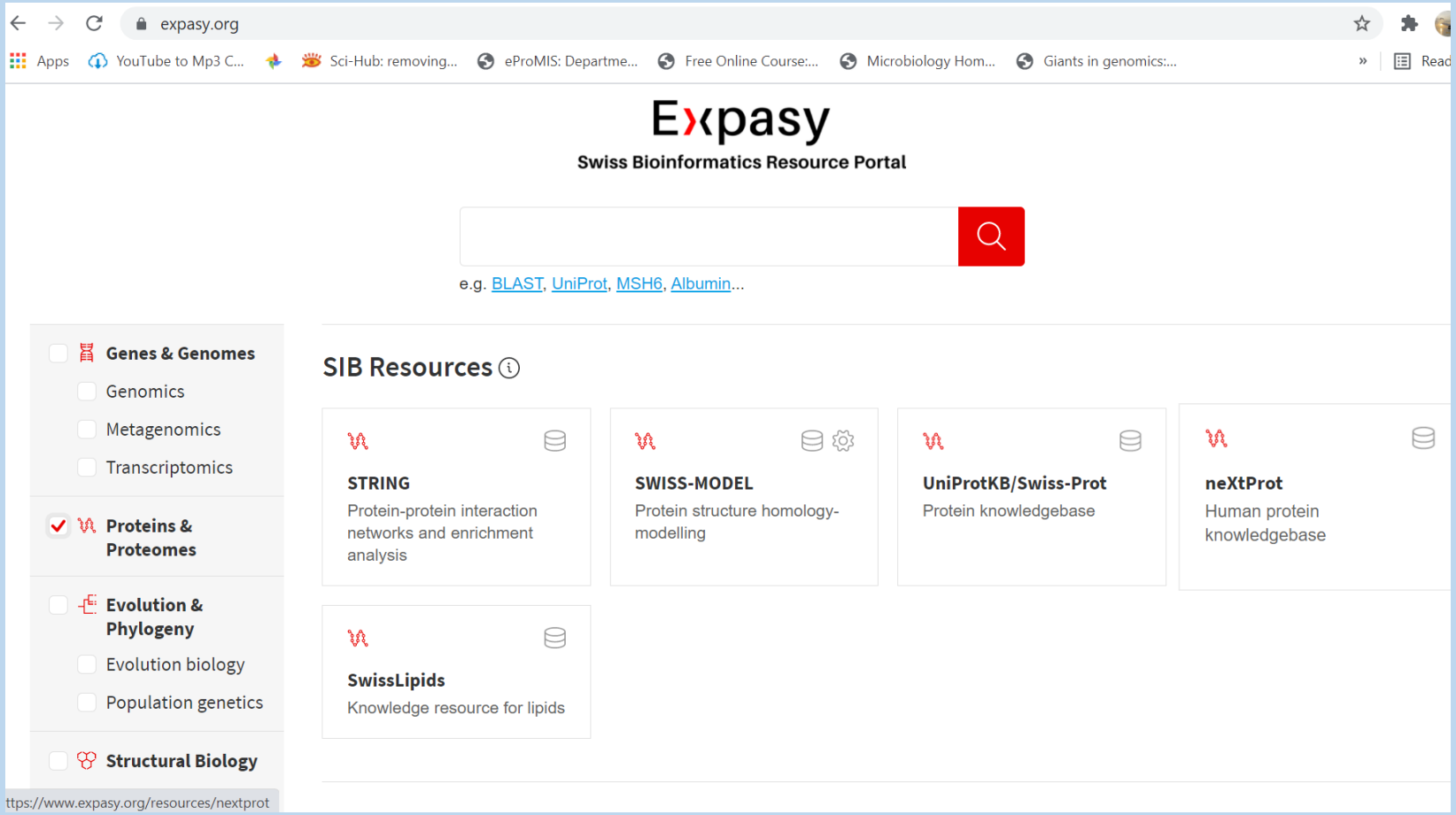
Rasmol/RasTop

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

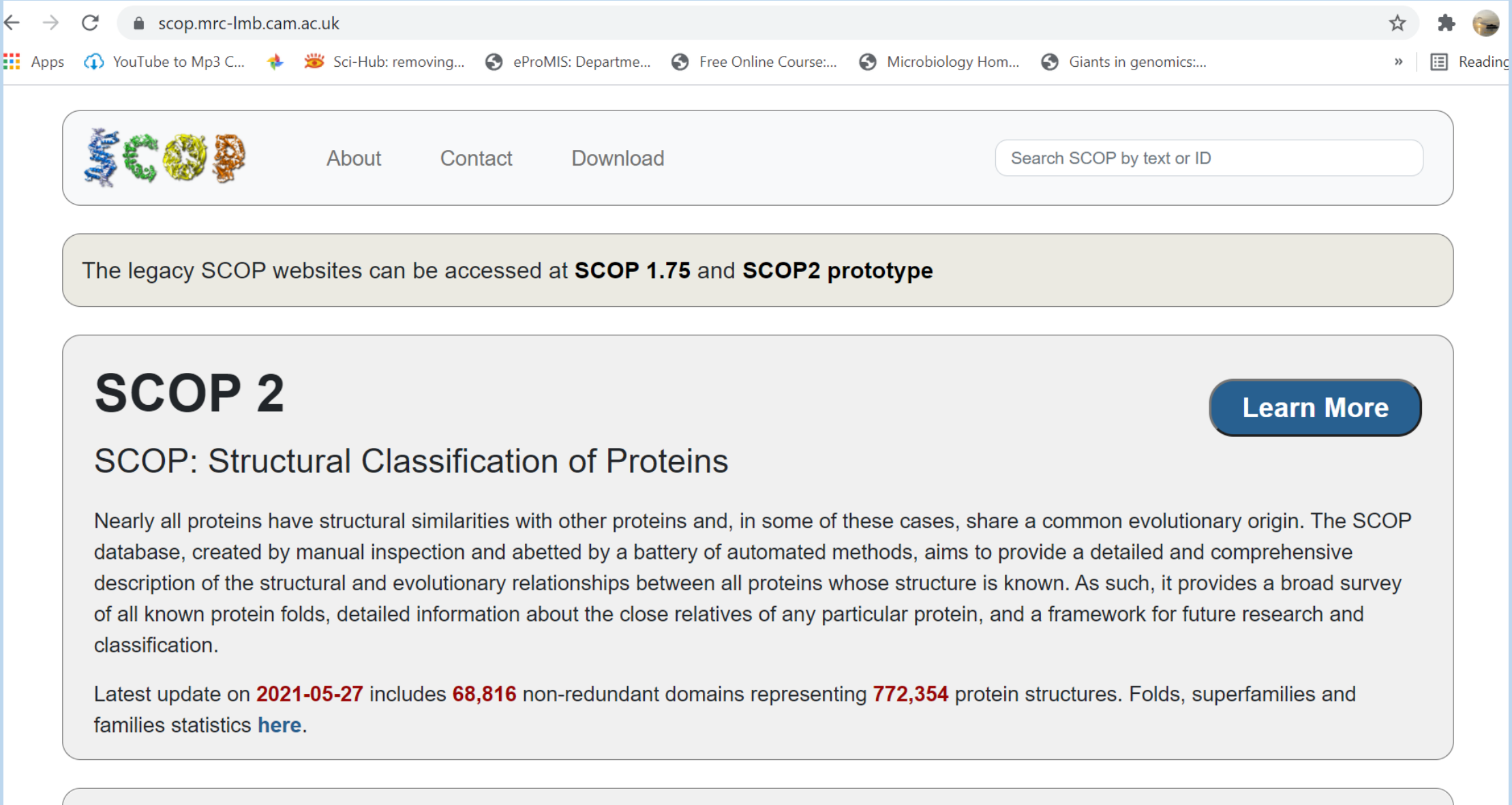
Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in **SWISS-MODEL Repository**?

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://h178133.nist.gov:4400/bmcd/bmcd.html
Resources	
Protein Data Bank	htttp://www.rcsb.org/pdb/

SCOP-Structural Classification of Proteins



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

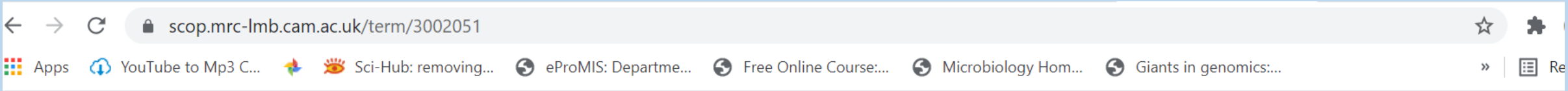
Browse by structural class

- **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

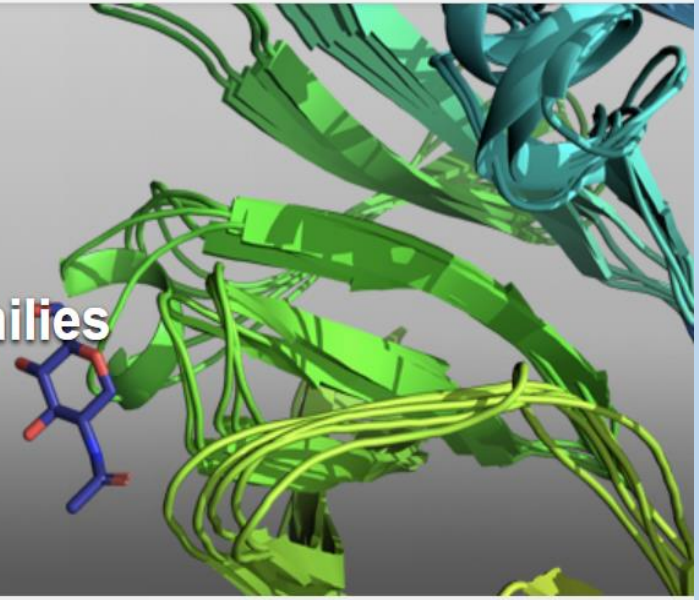
Apps | YouTube to Mp3 C... | Sci-Hub: removing... | eProMIS: Departme... | Free Online Course:... | Microbiology Hom... | Giants in genomics:... | Readin

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