Biological databases

Protein sequence databases

Swiss-Prot

https://www.uniprot.org/help/uniprotkb_sections

Biological databases

Protein sequence databases

Swiss-Prot

Swiss-Prot is an annotated protein sequence database, which was created at the Department of Medical Biochemistry of the University of Geneva and has been a collaborative effort of the Department and the EMBL, since 1987.

- Swiss-Prot (created in 1986) is a high quality manually annotated and non-redundant protein sequence database, which brings together experimental results, computed features and scientific conclusions.
- UniProtKB/Swiss-Prot is now the reviewed section of the UniProt Knowledgebase.
- The TrEMBL section of UniProtKB was introduced in 1996 in response to the increased dataflow resulting from genome projects.
- It was already recognized at that time that the traditional time- and labour-intensive manual curation process which is the hallmark of Swiss-Prot could not be broadened to encompass all available protein sequences.
- UniProtKB/TrEMBL contains high quality computationally analyzed records that are enriched with automatic annotation and classification.
- These UniProtKB/TrEMBL unreviewed entries are kept separated from the UniProtKB/Swiss-Prot manually reviewed entries so that the high quality data of the latter is not diluted in any way.
- Automatic processing of the data enables the records to be made available to the public quickly.

https://www.uniprot.org/help/uniprotkb_sections

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102476/#:~:text=SWISS%2DPROT%20(1)%20is,(EMBL)%2C%20since%201987.

Lecture – Structure databases

Biological databases

Types of data

Structure databases

- PDB
- NDB
- MMDB (from NCBI)
- SCOP
- CATH

Biological databases

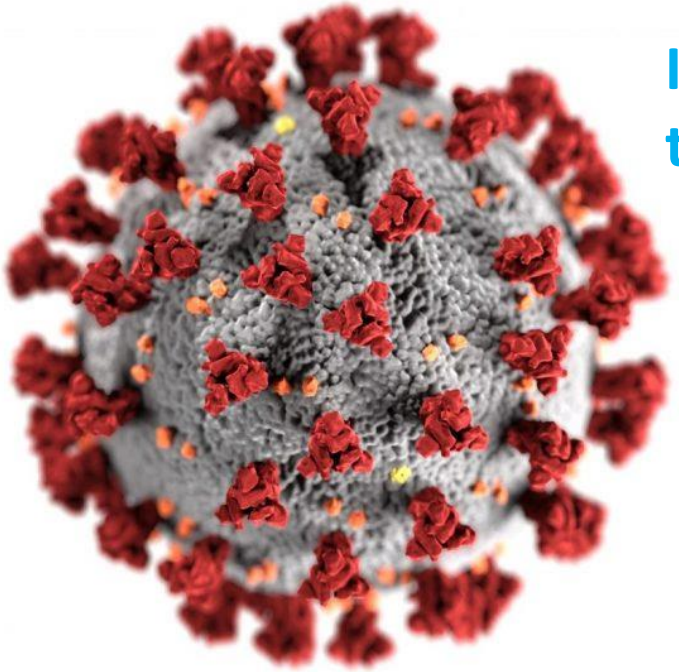
Structure databases

Protein Data Bank (PDB)

Biological databases

Structure databases

Protein Data Bank (PDB)



Identify the image?

Source:
https://www.clinicalomics.com/topics/patient-care/coronavirus/new-antibody-test-for-covid-19-targets-unique-region-of-spike-protein/

Spike protein

PDB code: 6XR8

https://www.rcsb.org/

Biological databases

Structure databases

Protein Data Bank (PDB)

What is it?

PDB is the single worldwide repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids.

These are the molecules of life that are found in all organisms including bacteria, yeast, plants, flies, other animals, and humans.

Understanding the shape of a molecule deduce a structure's role in human health and disease, and in drug development.

The structures in the archive range from tiny proteins and bits of DNA to complex molecular machines like the ribosome.

The PDB archive is available at no cost to users.

The PDB archive is updated weekly.

https://www.rcsb.org/

Biological databases

Structure databases

Protein Data Bank (PDB)

PDB – History

The PDB was established in 1971 at Brookhaven National Laboratory under the leadership of Walter Hamilton and originally contained 7 structures.

After Hamilton's untimely death, Tom Koetzle began to lead the PDB in 1973, and then Joel Sussman in 1994.

Led by Helen M. Berman, the Research Collaboratory for Structural Bioinformatics (RCSB) became responsible for the management of PDB in 1998.

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.

wwPDB consists of organizations that act as deposition, data processing and distribution centers for PDB data.

Stephen K. Burley became Director in 2014.

https://www.rcsb.org/

Biological databases

Structure databases

Protein Data Bank (PDB)

What is it?

The RCSB PDB supports a website where visitors can perform simple and complex queries on the data, analyze, and visualize the results.

The RCSB PDB has an international community of users, including biologists (in fields such as structural biology, biochemistry, genetics, pharmacology); other scientists (in fields such as bioinformatics, software developers for data analysis and visualization); students and educators (all levels); media writers, illustrators, textbook authors; and the general public.

The website (rcsb.org) is accessed by >1 million unique visitors per year.

RCSB PDB services have broad impact across research and education.

https://www.rcsb.org/

Biological databases

Structure databases

Protein Data Bank (PDB)

What is it?

wwPDB

The RCSB PDB is a member of the wwPDB, a collaborative effort with PDBe (UK), PDBj (Japan), and Biological Magnetic Resonance Data Bank (BMRB, USA) to ensure the PDB archive is global and uniform.

As the wwPDB archive keeper, the RCSB PDB updates the PDB archive at ftp://ftp.wwpdb.org weekly.

The structures included in each release are highlighted on the RCSB PDB home page and clearly defined on the FTP site.

These sites are maintained 24 hours a day, seven days a week.

A failover system automatically redirects internet traffic to a mirror site, if needed.

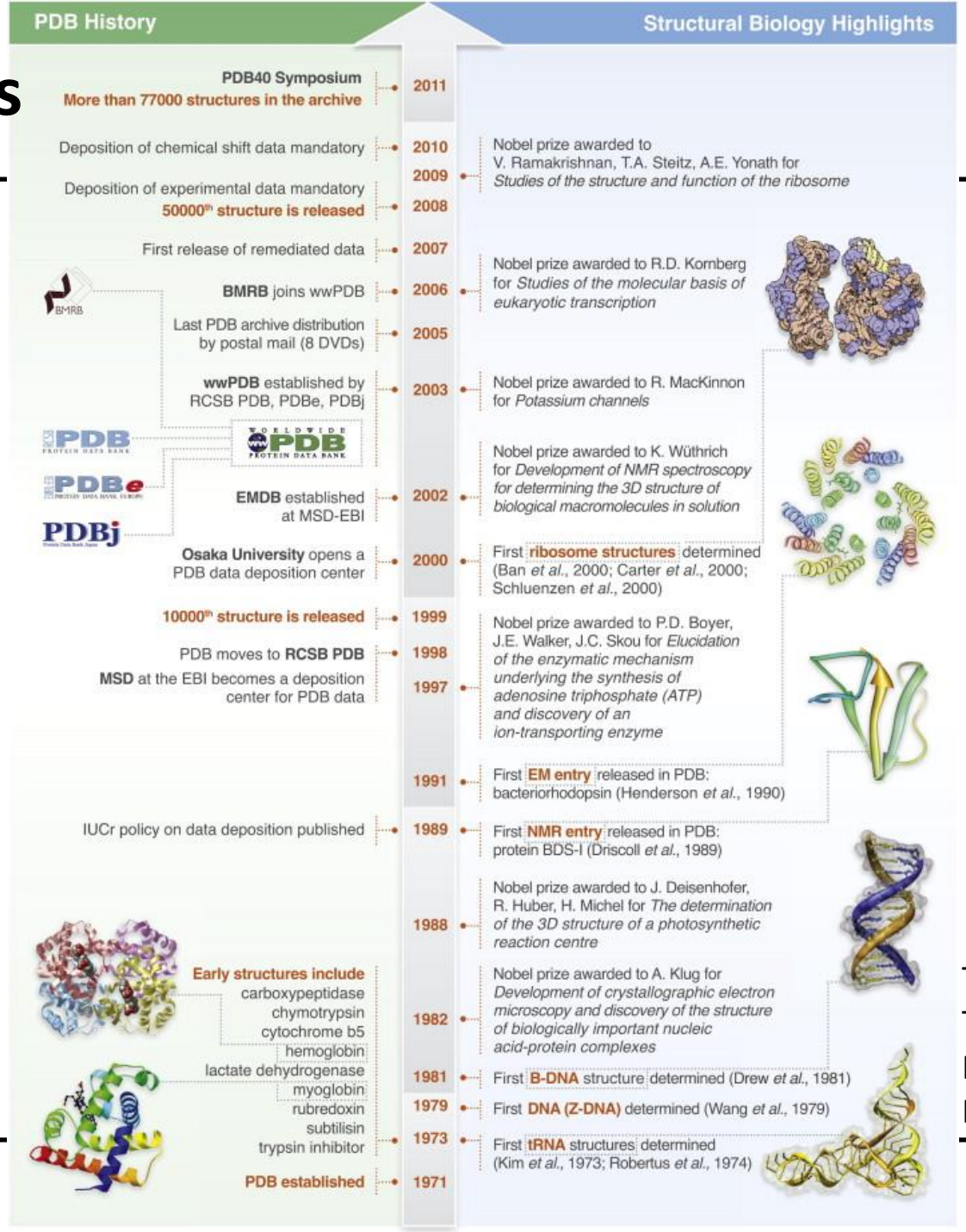
https://www.rcsb.org/

Biological databases

Structure databases

Protein Data Bank (PDB)

RCSB PDB History overview



Taken from Berman, 2012, The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future.

https://www.rcsb.org/

Biological databases

Structure databases

Protein Data Bank (PDB)

Experience PDB

COVID-19/SARS-CoV-2 Resources

PDB ID: 4 character alphanumeric code

https://www.rcsb.org/news?year=2020&article=5e74d55d2d410731e9944f52&feature=true

QUERY: PDB ID(s) **IN** (6XLU, 6XM0, 6XM3, 6XM4, 7CAH, 7JN2, 6ZME, 6ZRT, 6ZRU, 6WC1, 7JIR, 7JIT, 7JIV, 7JIW, 6ZWV, 6XEZ, 6XM5, 6ZDG, 6ZOW, 6ZP5, 6ZP7, 6XQB, 6ZSL, 7C7P, 7JFQ, 6ZBP, 6ZHD, 6ZOK, 6ZLW, 6ZM7, 6ZN5, 6ZON, 6ZP4, 6XEY, 6XR8, 6XRA, 6XS6, 6ZP0, 6ZP1, 6ZP2, 6ZOX, 6ZOY, 6ZOZ, 6ZOJ, 6XQS, 6XQT, 6XQU, 6XKL, 6XOA, 6XC2, 6XC3, 6XC4, 6XC7, 6XHM, 6XKM, 6XKF, 6XKH, 6XMK, 6ZFO, 6XCM, 6XCN, 6XE1, 6Z97, 6ZDH, 6ZGE, 6ZGG, 6ZGH, 6ZGI, 7C2L, 6XHU, 6XIP, 6ZCO, 6XFN, 6XG2, 7C8U, 6XG3, 6ZCT, 6ZCZ, 6ZER, 7C8W, 7C8V, 7CAN, 6XDG, 6X2G, 6XB0, 6XB1, 6XB2, 6XA4, 6XAA, 6XA9, 6XCH, 6XBG, 6XBH, 6XBI, 6XDC, 6XDH, 6Z2E, 6Z4U, 6M5I, 7BQ7, 7C8R, 7C8T, 6X6P, 7BYR, 5RHB, 5RHC, 5RHD, 5RHE, 5RHF, 6X4I, 6YZ5, 6YZ7, 6Z2M, 6Z43, 6M1V, 7BWJ, 7BZF, 7C2K, 6WPS, 6WPT, 6X29, 6X2A, 6X2B, 6X2C, 6WZO, 6WZQ, 6X1B, 7BW4, 7C2I, 7C2J, 7C01, 6WZU, 5RGT, 5RGU, 5RGV, 5RGW, 5RGX, 5RGY, 5RGZ, 5RH0, 5RH1, 5RH2, 5RH3, 5RH4, 5RH5, 5RH6, 5RH7, 5RH8, 5RH9, 5RHA, 6Y2G, 6Y2F, 6Y2E, 6W02, 6W01, 6Y84, 6W41, 6W4H, 6VSB, 6W4B, 6W61, 6W63, 6W75, 6VW1, 6W6Y, 6VXS, 6VWW, 6VYO, 6VYB, 6VXX, 6YB7, 5R84, 5R83, 5R7Y, 5R80, 5R82, 5R81, 5R7Z, 5REA, 5REC, 5REB, 5REE, 5RED, 5REG, 5REF, 5RE9, 5RE8, 5RE5, 5RE4, 5RE7, 5RE6, 5RFB, 5RFA, 5RFD, 5RFC, 5RFF, 5RFE, 5RFH, 5RFG, 5REY, 5REX, 5RF9, 5REZ, 5RF2, 5REP, 5RF1, 5RES, 5RF4, 5RER, 5RF3, 5REU, 5RF6, 5RET, 5RF5, 5REW, 5RF8, 5REV, 5RF7, 5REI, 5REH, 5REK, 5REJ, 5REM, 5REL, 5REO, 5RF0, 5REN, 5RFZ, 5RFY, 5RFR, 5RFQ, 5RFT, 5RFS, 5RFV, 5RFU, 5RFX, 5RFW, 5RFJ, 5RFI, 5RFL, 5RFK, 5RFN, 5RFM, 5RFP, 5RFO, 5RG0, 6M03, 6M17, 6M0J, 6M3M, 6LU7, 6LVN, 6LXT, 6LZG, 6W9C, 5R8T, 6M71, 6W9Q, 6YI3, 7BTF, 6WEN, 6WCF, 5RG1, 5RG2, 5RG3, 5RGG, 5RGH, 5RGI, 5RGJ, 5RGK, 5RGL, 5RGM, 5RGN, 5RGO, 5RGP, 5RGQ, 5RGR, 5RGS, 6M2N, 6M2Q, 6YLA, 6WIQ, 6WJI, 6WJT, 7BQY, 7BV2, 7BV1, 6LZE, 6M0K, 7BUY, 6W37, 6WEY, 6WKP, 6WKQ, 6WLC, 6YHU, 6YM0, 6YNQ, 6YOR, 6WKS, 6WNP, 6WOJ, 6WQF, 6WQ3, 6WQD, 6WRH, 6YT8, 6YWK, 6YWL, 6YWM, 6WRZ, 6WTC, 6WVN, 6YVA, 6YYT, 6YZ1, 7BRO, 7BRP, 7BRR, 7BZ5, 6WTJ, 6WTK, 6WTM, 6WTT, 6YVF, 6YZ6, 6WUU, 6WX4, 6WXC, 6WXD, 6YUN, 7C22)

Distinct conformational states of SARS-CoV-2 spike protein

https://www.rcsb.org/structure/6XRA

Biological databases

Structure databases

Nucleic Acids Database (NDB)

Biological databases

Structure databases

Nucleic Acids Database (NDB)

- A Portal for Three-dimensional Structural Information about Nucleic Acids.
- As of 9-Dec-2020 number of released structures: 11094

- The NDB contains information about experimentally-determined nucleic acids and complex assemblies.
- The goal of the NDB is to archive and distribute structural information about nucleic acids.
- The NDB was founded in 1992 by Helen M. Berman, Rutgers University, Wilma K. Olson, Rutgers University, and David Beveridge, Wesleyan University.
- The NDB Project is funded by the National Institutes of Health and has been funded by National Science Foundation and the Department of Energy in the past.

http://ndbserver.rutgers.edu/

Biological databases

Structure databases

Molecular Modeling Database (MMDB)

Biological databases

Structure databases

Molecular Modeling Database (MMDB)

Maintained by NCBI

- **Contains experimentally resolved structures of proteins, RNA, and DNA, derived from the Protein Data Bank (PDB),**
- **with value-added features such as explicit chemical graphs, computationally identified 3D domains (compact substructures) that are used to identify similar 3D structures, as well as links to literature, similar sequences, information about chemicals bound to the structures, and more.**
- **These connections make it possible, for example, to find 3D structures for homologs of a protein sequence of interest, then interactively view the sequence-structure relationships, active sites, bound chemicals, journal articles, and more.**

More at: https://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml