Structural Bioinformatics Databases

Structural bioinformatics cannot be thought without the experimental studies. These experimental studies such as X-ray crystallogrphy, Cryo-EM, and NMR bring valuable information like 3D structure, sequence, functional indentifications, flexibility of molecules, ligand binding cavity and interactions, membrane localization, and so on. Among these information, it can be implied that 3D structure information is backbone of the structural bioinformatics. 3D structures are mostly provided as atomic coordinates which are stored in many databases. Most common databases are listed in Table 1 for general uses. Besides this, these databases provide not only 3D structure but also comprahansive informations for proteins, small-molecules and experimental protocols. Macromolecule 3D sturucture can be downloaded from these databases with different datafiles such as PDB, PDBx/mmCIF, SDF to visualize.

- Read more about datafiles link

Visualization Programs

Instalations

PyMol

https://pymol.org/dokuwiki/?id=installation
Education Licence Registration: https://pymol.org/edu/

VMD

https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD

Tutorial PDF

https://pymol.org/dokuwiki/doku.php?id=tutorials
https://www.ks.uiuc.edu/Training/Tutorials/vmd/vmd-tutorial.pdf

Table 1. Useful structural bioinformatics databases. The table was copied from "Structural Bioinformatics Databases of General Use". For detailed descriptions and explanations please read the full paper.

Database	Description	Web address	Ref.
Worldwide Protein Data Bank (wwPDB)		http://wwpdb.org/	[1]

Database	Description	Web address	Ref.
BMRB	Biological Magnetic Resonance Data Bank (NMR)	http://www.bmrb.wisc.edu/	[2]
PDBe	Protein Data Bank in Europe	http://www.ebi.ac.uk/pdbe/	[3]
PDBj	Protein Data Bank Japan	http://pdbj.org/	[4]
RCSB PDB	Research Collaboratory for Structural Bioinformatics Protein Data Bank	http://www.rcsb.org/	[5]
Other views on PDB data			
PDBsum	Pictorial analysis of macromolecular structures	http://www.ebi.ac.uk/pdbsum/	[6]
PDB_ REDO	Re-refined PDB files	https://xtal.nki.nl/PDB_REDO/	[7]
CCD	Chemical Component Dictionary	http://www.wwpdb.org/data/ccd/	[8]
Classification			

Database	Description	Web address	Ref.
CATH	Domain classification of structures	http://www.cathdb.info/	[9]
Pfam	Classification of sequence families	http://pfam.xfam.org/	[10]
Flexibility and disorder			
PDB Flex	Intrinsic flexibility in proteins	http://pdbflex.org/	[11]
PED3	Protein Ensemble Database	http://pedb.vib.be/	[12]
Pocketome	Encyclopedia of ensembles of druggable binding sites	http://www.pocketome.org/	[13]
DisProt	Database of Protein Disorder	http://www.disprot.org/	[14]
Membrane proteins			
OPM	Orientations of proteins in membranes	http://opm.phar.umich.edu/	[15]
MemProtMD	Membrane proteins models	http://sbcb.bioch.ox.ac.uk/memprotmd/	[16]

Database	Description	Web address	Ref.
Other biomacromolecules			
NDB	Nucleic Acids Database	http://ndbserver.rutgers.edu/	[17]
GFDB	Glycan Fragment Database	http://www.glycanstructure.org/	[18]
Other databases			
UniProt	All about Protein Sequences	http://www.uniprot.org/	[19]
ChEMBL	Small drug-like molecules and targets	https://www.ebi.ac.uk/chembl/	[20]
ChEBI	Chemical Entities of Biological Interest	https://www.ebi.ac.uk/chebi/	[21]
EMDataBank	Global resource for 3-Dimensional Electron Microscopy	https://www.emdataresource.org	[22]
EMPIAR	the Electron Microscopy Public Image Archive	https://www.ebi.ac.uk/empiar/	[23]

Database	Description	Web address	Ref.
AlphaFold Protein Structure Database	DeepMind and EMBL's European Bioinformatics Institute EMBL-EBI partnared AF2 database	https://alphafold.ebi.ac.uk/	[24]

References

1. Berman, H., Henrick, K., Nakamura, H.: Announcing the worldwide Protein Data Bank. Nat. Struct. Biol. **10**(12), 980–980 (2003). doi: 10.1038/nsb1203-980 CrossRefPubMedGoogle Scholar

2. Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C.F., Tolmie, D.E., Kent Wenger, R., Yao, H., Markley, J.L.: BioMagResBank. Nucl. Acids Res. **36**(Database), D402–D408 (2007). doi: 10.1093/nar/gkm957 Google Scholar

3. Velankar, S., van Ginkel, G., Alhroub, Y., Battle, G.M., Berrisford, J.M., Conroy, M.J., Dana, J.M., Gore, S.P., Gutmanas, A., Haslam, P., Hendrickx, P.M.S., Lagerstedt, I., Mir, S.: Fernandez Montecelo, M.A., Mukhopadhyay, A., Oldfield, T.J., Patwardhan, A., Sanz-García, E., Sen, S., Slowley, R.A., Wainwright, M.E., Deshpande, M.S., Iudin, A., Sahni, G., Salavert Torres, J., Hirshberg, M., Mak, L., Nadzirin, N., Armstrong, D.R., Clark, A.R., Smart, O.S., Korir, P.K., Kleywegt, G.J.: PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. Nucl. Acids Res. **44**(D1), D385–D395 (2016). doi: 10.1093/nar/gkv1047 Google Scholar

4. Kinjo, A.R., Suzuki, H., Yamashita, R., Ikegawa, Y., Kudou, T., Igarashi, R., Kengaku, Y., Cho, H., Standley, D.M., Nakagawa, A., Nakamura, H.: Protein data bank Japan (PDBj): maintaining a structural data archive and resource description framework format. Nucl. Acids Res. **40**(D1), D453–D460 (2012). doi: 10.1093/nar/gkr811 CrossRefPubMedGoogle Scholar

5. Berman, H.M.: The protein data bank. Nucl. Acids Res. **28**(1), 235–242 (2000). doi: 10.1093/nar/28.1.235 CrossRefPubMedPubMedCentralGoogle Scholar

6. de Beer, T.A.P., Berka, K., Thornton, J.M., Laskowski, R.A.: PDBsum additions. Nucl. Acids Res. **42**(D1), D292–D296 (2014). doi: 10.1093/nar/gkt940 CrossRefPubMedGoogle Scholar

7. Joosten, R.P., Joosten, K., Murshudov, G.N., Perrakis, A.: PDB_REDO: constructive validation, more than just looking for errors. Acta Crystallogr. Sect. D Biol. Crystallogr. **68**(4), 484–496 (2012). doi: 10.1107/S0907444911054515 CrossRefGoogle Scholar

8. Westbrook, J.D., Shao, C., Feng, Z., Zhuravleva, M., Velankar, S., Young, J.: The chemical component dictionary: complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the protein data bank. Bioinformatics **31**(8), 1274–1278 (2015). doi: 10.1093/bioinformatics/btu789 CrossRefPubMedGoogle Scholar

9. Sillitoe, I., Lewis, T.E., Cuff, A., Das, S., Ashford, P., Dawson, N.L., Furnham, N., Laskowski, R.A., Lee, D., Lees, J.G., Lehtinen, S., Studer, R.A., Thornton, J., Orengo, C.A.: CATH: comprehensive structural and functional annotations for genome sequences. Nucl. Acids Res. **43**(D1), D376–D381 (2015). doi: 10.1093/nar/gku947 CrossRefPubMedGoogle Scholar

10. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M.: Pfam: the protein families database. Nucl. Acids Res. **42**(D1), D222–D230 (2014). doi: 10.1093/nar/gkt1223 CrossRefPubMedGoogle Scholar

11. Hrabe, T., Li, Z., Sedova, M., Rotkiewicz, P., Jaroszewski, L., Godzik, A.: PDBFlex: exploring flexibility in protein structures. Nucl. Acids Res. **44**(D1), D423–D428 (2016). doi: 10.1093/nar/gkv1316 CrossRefPubMedGoogle Scholar

12. Varadi, M., Kosol, S., Lebrun, P., Valentini, E., Blackledge, M., Dunker, A.K., Felli, I.C., Forman-Kay, J.D., Kriwacki, R.W., Pierattelli, R., Sussman, J., Svergun, D.I., Uversky, V.N., Vendruscolo, M., Wishart, D., Wright, P.E., Tompa, P.: pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. Nucl. Acids Res. **42**(D1), D326–D335 (2014). doi: 10.1093/nar/gkt960 CrossRefPubMedGoogle Scholar

13. Kufareva, I., Ilatovskiy, A.V., Abagyan, R.: Pocketome: an encyclopedia of small-molecule binding sites in 4D. Nucl. Acids Res. **40**(D1), D535–D540 (2012). doi: 10.1093/nar/gkr825 CrossRefPubMedGoogle Scholar

14. Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V.N., Obradovic, Z., Dunker, A.K.: DisProt: the database of disordered proteins. Nucl. Acids Res. **35**(Database), D786–D793 (2007). doi: 10.1093/nar/gkl893 Google Scholar

15. Lomize, M.A., Lomize, A.L., Pogozheva, I.D., Mosberg, H.I.: OPM: Orientations of proteins in membranes database. Bioinformatics **22**(5), 623–625 (2006). doi: 10.1093/bioinformatics/btk023 CrossRefPubMedGoogle Scholar

16. Stansfeld, P.J., Goose, J.E., Caffrey, M., Carpenter, E.P., Parker, J.L., Newstead, S., Sansom, M.S.: MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. Structure **23**(7), 1350–1361 (2015). doi: 10.1016/j.str.2015.05.006 CrossRefPubMedPubMedCentralGoogle Scholar

17. Narayanan, B., Westbrook, J., Ghosh, S., Petrov, A.I., Sweeney, B., Zirbel, C.L., Leontis, N.B., Berman, H.M.: The nucleic acid database: new features and capabilities. Nucl. Acids Res. **42**(D1), D114–D122 (2014). doi: 10.1093/nar/gkt980 CrossRefGoogle Scholar

18. Jo, S., Im, W.: Glycan fragment database: a database of PDB-based glycan 3D structures. Nucl. Acids Res. **41**(D1), D470–D474 (2013). doi: 10.1093/nar/gks987 CrossRefPubMedGoogle Scholar

19. UniProt: a hub for protein information. Nucl. Acids Res. **43**(D1), D204–D212 (2015). doi: 10.1093/nar/gku989

20. Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., Overington, J.P.: ChEMBL: a large-scale bioactivity database for drug discovery. Nucl. Acids Res. **40**(D1), D1100–D1107 (2012). doi: 10.1093/nar/gkr777 CrossRefPubMedGoogle Scholar

21. Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., Steinbeck, C.: The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucl. Acids Res. **41**(D1), D456–D463 (2013). doi: 10.1093/nar/gks1146 CrossRefPubMedGoogle Scholar

22. Lawson, C. L., Patwardhan, A., Baker, M. L., Hryc, C., Garcia, E. S., Hudson, B. P., ... & Chiu, W. (2016). EMDataBank unified data resource for 3DEM. *Nucleic acids research*, *44*(D1), D396-D403. doi:https://doi.org/10.1093/nar/gkv1126 CrossRef.Google Scholar

23. Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J., & Patwardhan, A. (2016). EMPIAR: a public archive for raw electron microscopy image data. *Nature methods*, *13*(5), 387-388. doi:[https://doi.org/10.1038/nmeth.3806]. CrossRefGoogle Scholar

24. Varadi, M *et al*. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* (2021). doi:[https://doi.org/10.1093/nar/gkab1061] CrossRef