



IIT Guwahati

Lecture 35

Course BT 631

Protein Structure, Function and Crystallography

Prof. Arun Goyal

Dept. of Biosciences and Bioengineering



Protein Data Bank (PDB)

The Protein Data Bank (PDB) at Brookhaven National Laboratory (BNL) is a database containing experimentally determined 3-dimensional structures of proteins, nucleic acids and other biological macromolecules, over 8000 entries.

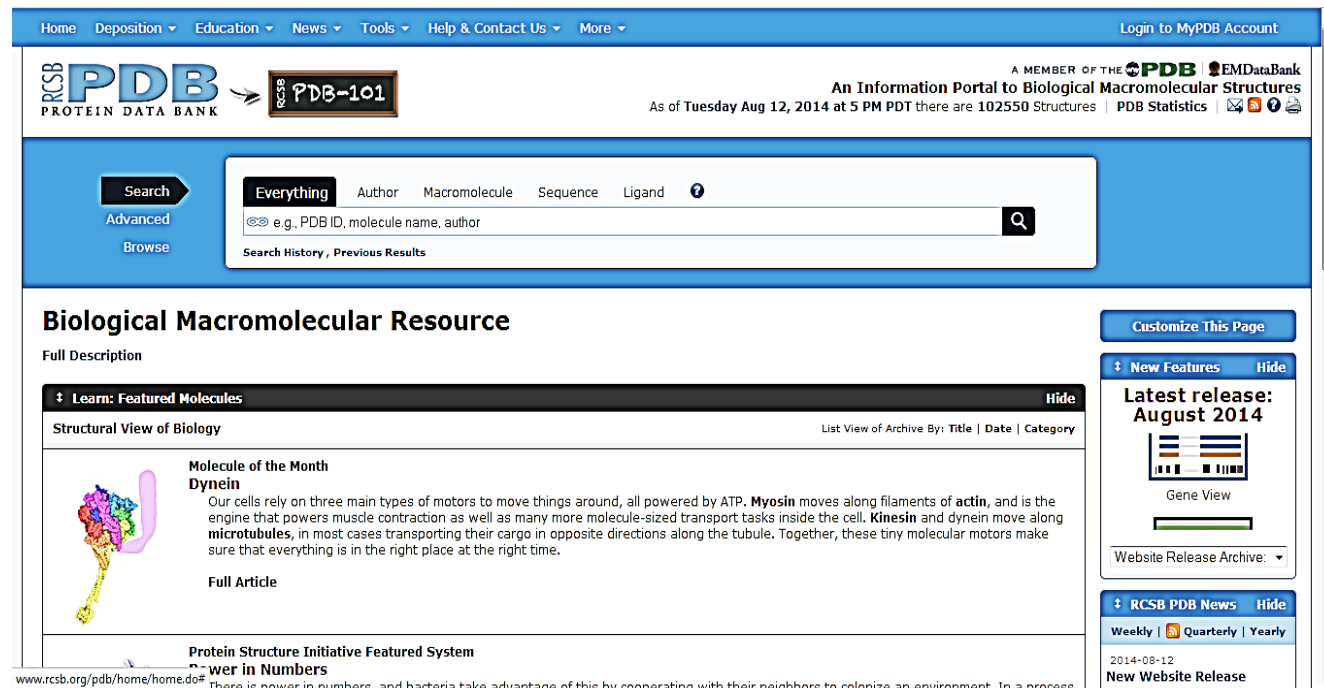
Data are easily submitted via PDB's WWW-based tool AutoDep, in either mmCIF or PDB format and are most conveniently examined via PDB's WWW-based tool 3DB Browser.

In parallel, PDB's new AutoDep facility allows researchers to deposit their data quickly and accurately using World Wide Web directly into the PDB either in the European Bioinformatics Institute (EBI) or BNL. Data are then processed by the PDB staff at Brookhaven.

Protein Data Bank (PDB)

The PDB was established in 1971 by Dr Walter Hamilton, at the suggestion of members of the American Crystallographic Association (ACA) and participants at the 1971 Cold Spring Harbor Symposium.

Introduction of the current PDB format in 1972 ensured that these data were readily accessible in a convenient and standard form, not only to crystallographers but also to biologists and chemists.



The screenshot shows the PDB website interface. At the top, there is a navigation bar with links: Home, Deposition, Education, News, Tools, Help & Contact Us, and More. A 'Login to MyPDB Account' link is on the right. Below the navigation bar is the PDB logo and a banner for 'PDB-101'. A status bar indicates 'As of Tuesday Aug 12, 2014 at 5 PM PDT there are 102550 Structures' and provides links to 'PDB Statistics' and social media icons. The main content area features a search bar with a 'Search' button and a dropdown menu for search criteria (Everything, Author, Macromolecule, Sequence, Ligand). Below the search bar is a 'Biological Macromolecular Resource' section with a 'Full Description' link. This section includes a 'Learn: Featured Molecules' link, a 'Structural View of Biology' link, and a 'Molecule of the Month' section featuring 'Dynein'. The 'Dynein' section includes a 3D molecular model and a text description: 'Our cells rely on three main types of motors to move things around, all powered by ATP. Myosin moves along filaments of actin, and is the engine that powers muscle contraction as well as many more molecule-sized transport tasks inside the cell. Kinesin and dynein move along microtubules, in most cases transporting their cargo in opposite directions along the tubule. Together, these tiny molecular motors make sure that everything is in the right place at the right time.' Below this is a 'Full Article' link. At the bottom of the main content area is a 'Protein Structure Initiative Featured System' section with a link to 'Power in Numbers'. On the right side of the page, there is a 'Customize This Page' button and a 'New Features' section with a 'Hide' link. Below this is a 'Latest release: August 2014' section with a 'Gene View' link and a 'Website Release Archive' dropdown menu. At the bottom right is an 'RCSB PDB News' section with a 'Hide' link, a 'Weekly' link, and a 'Quarterly' link. The footer of the page contains the URL 'www.rcsb.org/pdb/home/home.do' and a copyright notice: 'There is power in numbers, and bacteria take advantage of this by cooperating with their neighbors to colonize an environment. In a process'.

Fig: PDB introduction page

Protein Data Bank (PDB)

Data deposition

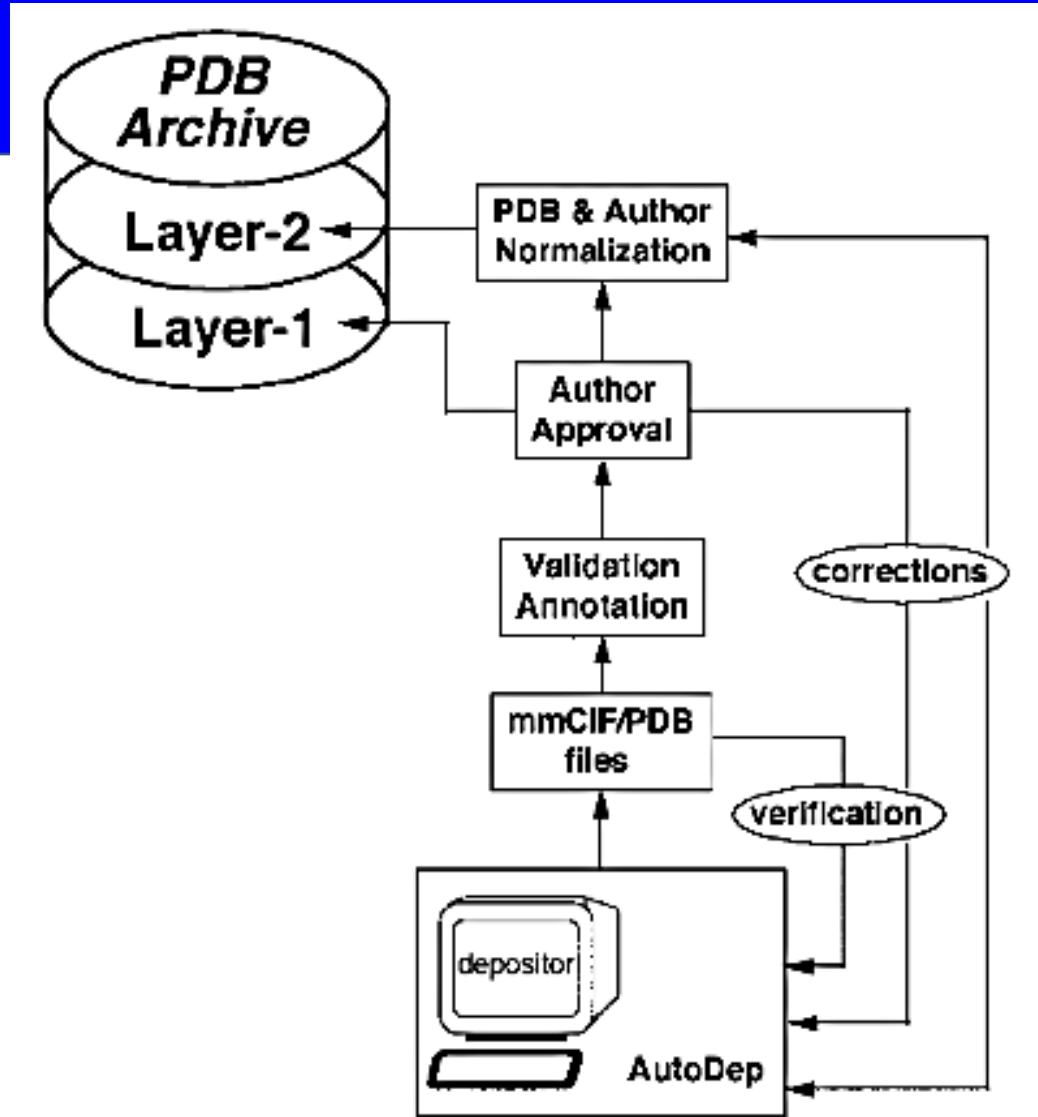


Fig. Schematic presentation of data deposition in PDB.

Protein Data Bank (PDB)

Rapid developments in the preparation of crystals of macromolecules and in experimental techniques for structure analysis and refinement have led to a revolution in structural biology. These factors have contributed significantly to an enormous increase in the number of structural studies of macromolecules to atomic resolution.

Protein Data Bank (PDB)

Advances in developments include:

- (1) recombinant DNA techniques that permit almost any protein or nucleic acid to be produced in large amounts;
- (2) rapid protein and DNA (gene) sequencing techniques that have made protein sequencing routine;
- (3) better X-ray detectors;
- (4) real-time interactive computer graphics systems, together with more automated methods for structure determination and refinement;
- (5) synchrotron radiation, allowing the use of extremely tiny crystals, multiple-wavelength anomalous dispersion (MAD) phasing, and time-resolved studies via Laue technique (*It is a method of X-ray diffraction for the analysis of crystal structure, using a single, fixed crystal irradiated by a beam of a continuous spectrum of X-ray, the spots pattern produce is called X-ray diffraction. It gives information about the crystallographic symmetry of the crystal structure. Also called Lauégrams.*);
- (6) NMR methods permitting structure determination of macromolecules in solution;
- (7) electron microscopy (EM) techniques, for obtaining high-resolution structures.

These dramatic advances produced an abrupt transition from the linear growth of 15-25 new structures deposited per year in the PDB before 1987 to a rapid exponential growth reaching the current rate of about 50 submissions per week.

Contents and access to the PDB Archives

The archives contain atomic coordinates, bibliographic citations, primary and secondary structure information as well as the crystallographic structure factors and NMR experimental data.

Annotations in the structure entries include amino-acid or nucleotide sequences (with notes of any conflicts between the structure in the PDB and sequence databases), source organism from which the biological material was derived, references to the papers, secondary structure, complexes with small molecules included within the structure.

Databases linked with PDB

BioMagResBank: Relational database for sequence-specific protein NMR Data

BLOCKS: Database of conserved regions in groups of proteins

CATH: Protein structure classification

Dali/FSSP: Families of structurally similar proteins

EMBL: European Molecular Biology Laboratory sequence database

Entrez: NCBI's documentation database

ENZYME: Enzyme nomenclature database

ESTHER: ESTerases and alpha/beta hydrolase enzymes database

GenBank: NIH genetic sequence database

GDB: Genome database

Kinase: Protein kinase database project

KineMage: Protein Sciences Kinemage server

Databases linked with PDB

LPFC: Library of protein family cores

MacroMolecule: [Crystal macromolecule files](#)

MMDB Molecular modelling database **NDB** Nucleic Acid Database

OLDERADO: Core, domain and representative structure database

PDBObs: Archive of Obsolete PDB Entries at SDSC

PDBREPORT: [Structure verification reports for X-ray structures](#)

PIR: Protein Information Resource

PROSITE: [Dictionary of protein sites and patterns](#)

ProtMotDB: Protein Motions Database

PubMed: Medline bibliographic database

SCOP: [Structural Classification of Proteins](#)

Swiss 3D-Image: [Three-dimensional images of proteins](#) and other biological macromolecules

SwissProt: [Annotated protein sequence database](#)

TREMBL: Translation from EMBL sequence database