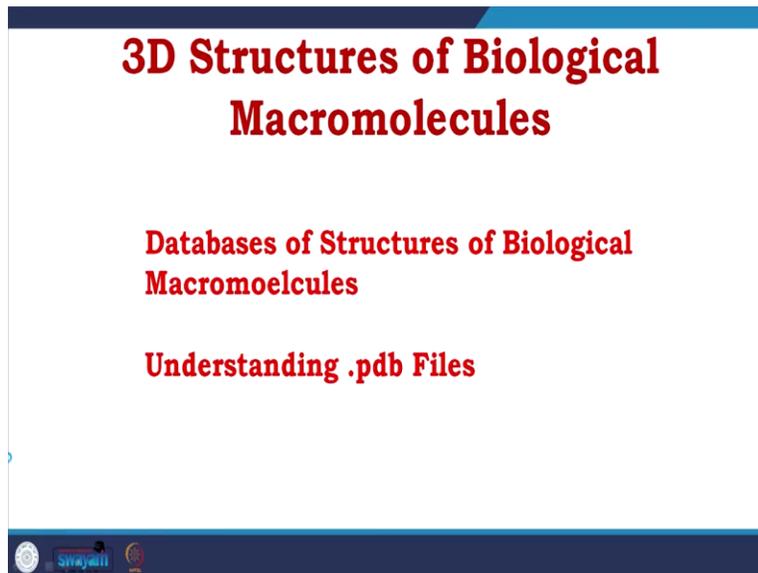


Structural Biology
Prof. Saugata Hazara
Department of Biotechnology
Indian Institute of Technology - Roorkee

Lecture – 40
Protein Data Bank

Hi everyone, welcome again to the course on structural biology. We are continuing with structural biology techniques. I have discussed high-resolution techniques like x-ray and NMR. In this module, we are talking about cryo-electron microscopy. This is the last class of this module. Here I will talk about protein data bank. I would also discuss some other relevant databases where structural data and structural information used to be deposited.

(Refer Slide Time: 01:10)



So, today's topic is the 3D structure of biological macromolecules, their databases of biological macromolecule structures, and understanding .pdb files.

(Refer Slide Time: 01:24)

Introduction:

The **Protein Data Bank (PDB)** is a database for the three-dimensional structural data of large biological molecules

The data, typically obtained by X-ray crystallography, NMR spectroscopy, or, increasingly, cryo-electron microscopy, and submitted by structural biologists from around the world, are freely accessible on the Internet

Any experimental information regarding biological macromolecules have to be submitted and validated first before scientific publication

- Currently this is considered as the only repository for 3-D biological macromolecular structure solved by using experimental techniques



The protein data bank is a database for the 3-dimensional structural data of large biological molecules or biological macromolecules. The data typically obtained by x-ray crystallography, NMR spectroscopy, or as now it is coming cryo-electron microscopy, and submitted by structural biologists from around the world are freely accessible on the internet.

Any experimental information regarding biological macromolecules' structural data must be submitted and validated before scientific publication. So, you cannot go for a scientific publication unless your data is deposited but deposition is not enough. After depositing people from PDB, will give you a validation certificate. So, when you show this validation certificate, only the journals will start reviewing your paper.

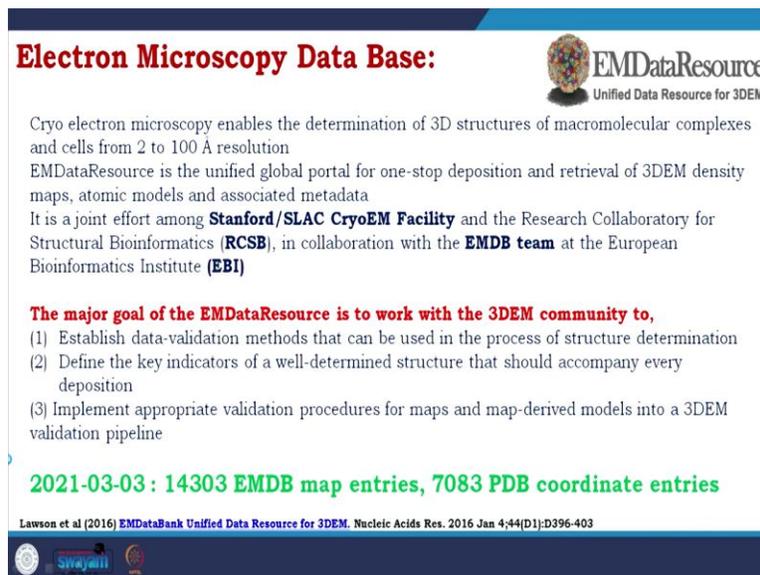
Currently, this is considered the only repository for 3D biological macromolecular structure solved by using an experimental technique. Anywhere, any structural information obtained through experiment has to be deposited in protein data banks that is why it is a very important database.

(Refer Slide Time: 03:05)



This is the home page of PDB RCSB (research collaborative structural bioinformatics). I could provide you with the most recent structural information data by looking at one 175282 biological macromolecular structure entry currently deposited in the protein data bank.

(Refer Slide Time: 03:52)



So, as I told besides PDB there are other databases. One of them is the electron microscopy database. The electron microscopic data is deposited in an electron microscopy database or E M data resource. Cryo-electron microscopy enables the determination of the 3D structure of macromolecular complexes and cell's biological organelles from 2 to 100-angstrom resolution.

EM data resource is the unified global portal for one-step deposition and retrieving 3D electron microscopy density maps, atomic models, and associated metadata. It is a joint effort among Stanford slack cryo-EM facility. The research collaboratory for structural bioinformatics is RCSB, collaborating with the EMDB the Electron Microscopy Database team at the European Bioinformatics Institute or EBI.

The major goal of EM data resource is to work with the 3D electron microscopy community to establish a data validation method that can be used in structure determination, as we have discussed before. So, once the data is prepared, the data must be deposited, and then the data must be validated. Define the key indicator of a well-determined structure that should accompany every deposition.

So, comparing and getting the structure's quality to implement appropriate validation procedures for maps and map derived models into a 3D electron microscopy validation pipeline. Currently, this year, on 3rd March, there are 14303 EMDB map entries 7083 PDB coordinate entries, and if you want to know more, the link is given. You go to the link, and you will get the details about the electron microscopy database.

(Refer Slide Time: 06:22)

Nucleic Acid Database (NDB):

A Portal for Three-dimensional 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 contains information about experimentally-determined nucleic acids and complex assemblies

People could use the NDB to perform searches based on annotations relating to sequence, structure and function, and to download, analyze, and learn about nucleic acids

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

3-Mar-2021 number of released structures: 11234

The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Res.* (2014) 42, D114-22

7

Another related database is the nucleic acid database or NDB. So, a portal for 3-dimensional structural information about nucleic acids the NDB or nucleic acid database founded in 1992 by Helen M Barman from Rutgers University. Wilma K Olson again from Rutgers University, and

David Beveridge from West Lane University. The NDB contains information about experimentally determining nucleic acid and complex assemblies.

So, nucleic acid, nucleic acid small molecule, nucleic acid-protein, people could use the NDB to perform searches based on annotation relating to sequence, structure, and function, and download analyze and learn about nucleic acid acids. The NDB Project is funded by the National Institute of Health, NIH, and has been funded by the National Science Foundation, NSF, and the Department of Energy, DOE in the past. Again, 3rd match 2021 number of release structure the nucleic acid, nucleic acid complex with a small molecule with protein is 11234. Again about further details, the link is provided here. So, interested people could go through, and know more about it.

(Refer Slide Time: 07:58)

Biological Magnetic Resonance Data Bank (BMRB):

A Repository for Data from NMR Spectroscopy on Proteins, Peptides, Nucleic Acids, and other Biomolecules

The concept of a biomolecular NMR data bank was developed under a five-year research grant awarded to the University of Wisconsin-Madison from the National Library of Medicine, National Institutes of Health

BMRB at the University of Wisconsin-Madison has been supported since 1 September 1996 by the National Library of Medicine

BMRB collects, annotates, archives, and disseminates (worldwide in the public domain) the important spectral and quantitative data derived from NMR spectroscopic investigations of biologically relevant molecules

The goal is to empower scientists in their analysis of the structure, dynamics, and chemistry of biological systems and to support further development in the field of biomolecular NMR spectroscopy

<http://bmrbl.io>

Biological Magnetic Resonance Data Bank or BMRB: A repository for data from NMR spectroscopy on proteins, peptides, nucleic acids, and other biomolecules. The concept of a biomolecular NMR data bank was developed under a 5-year research grant awarded to the University Of Wisconsin Medicine from the National Library of Medicine, National Institute of Health, NIH. BMRB or Biological Magnetic Resonance Data Bank at the University Of Wisconsin Medicine has been supported since 1st September 1996 by the National Library of Medicine. BMRB collect annotated archive and disseminates them worldwide in the public domain.

The important spectral and quantitative data derived from NMR spectroscopic investigations of biologically relevant molecules. This data bank aims to empower scientists in their analysis of the structural dynamics and chemistry of biological systems and to support further development in the field of biomolecular NMR spectroscopy. So, this data bank is mostly specific to NMR spectroscopy. Again the link is provided, so, that you could go to the further details.

(Refer Slide Time: 10:04)

History of Protein Data Bank:

Two major forces converged to initiate the PDB:

A small but growing collection of sets of protein structure data determined by X-ray diffraction;

The newly available (1968) molecular graphics display, the Brookhaven Raster Display (BRAD), to visualize these protein structures in 3-D

In 1969, with the sponsorship of Walter Hamilton at the Brookhaven National Laboratory (BNL), Edgar Meyer (Texas A&M University) began to write software to store atomic coordinate files in a common format to make them available for geometric and graphical evaluation

By 1971, one of Meyer's programs, SEARCH, enabled researchers to remotely access information from the database to study protein structures offline

SEARCH was instrumental in enabling networking, thus marking the functional beginning of the PDB

9

Two major forces converge to initiate the protein data bank. A small but growing collection of sets of protein structure data was determined by x-ray diffraction. Those data need to be kept somewhere, and that was one force to bring the idea about protein data bank. But this was also supported by newly available molecular graphics display the Brookhaven Raster Display or BRAD to visualize these protein structures in 3D. As new structures came, more and more scientists wanted to research them and compare them. So, visualization was a must.

In 1969 with the sponsorship of Walter Hamilton at the Brookhaven National Laboratory BNL, Edgar Mayor from Texas A M University began to write software to store atomic coordinate files in a common format to make them available for geometric and graphical evaluation.

By 1971 one of Meyer's programs called search enabled researchers to remotely access information from the database to offline study of protein structure. This is the first time people started getting all the data and the related software to analyze them making protein

crystallography and structural biology popular. SEARCH, the software, was instrumental in enabling networking, thus marking the functional beginning of the PDB.

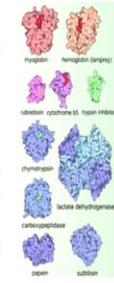
(Refer Slide Time: 12:38)

History of Protein Data Bank:

Founded in 1971 by Brookhaven National Laboratory, New York
PDB established at Brookhaven National Laboratory (USA) and the Cambridge Crystallographic Data Center (UK)

1973: Early structures
Deposited structures include carboxypeptidase, chymotrypsin, cytochrome b5, hemoglobin, lactate dehydrogenase, myoglobin, rubredoxin, subtilisin, trypsin inhibitor
Upon Hamilton's death in 1973, Tom Koeztle took over direction of the PDB for the subsequent 20 years

1974: First tRNA structures
1974 Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature*, 250, 546-55
1973 Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science*, 179, 285-288



It was founded officially in 1971 by Brookhaven National Laboratory, New York. It was established at Brookhaven National Laboratory with the Cambridge Crystallographic Data Center from the United Kingdom. In 1973 the early structures were myoglobin, hemoglobin as we know they are the first two structures. Then rubredoxin, cytochrome b5, trypsin inhibitor, chymotrypsin, carboxyl peptides, papain, subtilisin, lactate dehydrogenase are all early structures that came.

Tom Cosley took over the direction of the PDB for the subsequent 20 long years. In 1974 the first RNA structure appeared. The 1974 structure of yeast phenylalanine tRNA at 3-angstrom resolution was published in nature in 1973—the folding of the polynucleotide chain published in science.

(Refer Slide Time: 13:57)

History of Protein Data Bank:

1979: First Z-DNA published

1979 Molecular structure of a left-handed double helical DNA fragment at atomic resolution *Nature*, 282, 680-686

1981: First B-DNA published

1981 Structure of a B-DNA dodecamer: conformation and dynamics *Proceedings of the National Academy of Sciences of the United States of America*, 78, 2179-2183

1982: Nobel Prize in Chemistry awarded to Aaron Klug

Prize was awarded for his development of crystallographic electron microscopy and his structural elucidation of biologically important nucleic acid-protein complexes



11

In 1979 first Z-DNA published the molecular structure of a left-h, anded double helical DNA fragment at atomic resolution again published in nature. In 1981 first B-DNA dodecamer conformation and dynamics. 1982 Nobel Prize in chemistry awarded to Aaron Klug. The prize was awarded for developing crystallographic electron microscopy and structural elucidation of biologically important nucleic acid-protein complexes.

(Refer Slide Time: 14:42)

History of Protein Data Bank:

1988: Nobel Prize in Chemistry for 3D structure of photosynthetic center

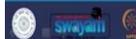
Noble prize was awarded to Johann Deisenhofer, Robert Huber and Hartmut Michel for the determination of the three-dimensional structure of a photosynthetic reaction center

1989: First NMR structure released in the PDB

First NMR structure was release in PDB with PDB ID 1BDS which belongs to the ANTIHYPERTENSIVE AND ANTIVIRAL PROTEIN BDS-I FROM THE SEA ANEMONE ANEMONIA SULCATA
Also IUCr policy on data deposition was published

1991: First 3D Electron Microscopy Data

First 3DElectron Mrcoscopy structure bacterio rhodopsin was released in the PDB



12

In 1988 Nobel Prize in chemistry again for the 3D structure of the photosynthetic center. A lot of structural biologists got Nobel Prize we have discussed. So, these are marked for PDB because each of these awards helps PDB to grow more. The Nobel Prize was awarded to Johan Dyson

Hoffer, Robert Hoover, and Hot Mood Missile to determine the 3-dimensional structure of a photosynthetic reaction center.

In 1989 the first NMR structure was released in PDB. So, the first NMR structure was released in PDB with pdb id1BDS which belongs to the anti-hypertensive and antiviral protein bds I or one from the sea anemone anemonia sulcata.

Also, in 1989, the IUCr International Union of Crystallographic research made fixed policies on data deposition and published the data. In 1991 the first 3D electron microscopy data, the first 3 electron microscopy structure of bacterial rhodopsin, is one of the innovations that made the mark of electron microscopy towards solving protein structure by Henderson, who is also part of the Nobel Prize in 2017.

(Refer Slide Time: 17:17)

History of Protein Data Bank:

- 1997: Nobel Prize for mechanism of adenosine synthesis**
Noble prize was awarded in Chemistry to Paul D. Boyer and John E. Walker for their elucidation of the enzymatic mechanism underlying the synthesis of adenosine
- 1998: Inclusion of other databases**
Database at European Bioinformatics Institute (EBI) become PDBe 
Transferred to the Research Collaboratory for Structural Bioinformatics (RCSB) 
- 1999: 10,000th structure is released**
- 2000: First ribosome structures in the PDB**
Osaka University opens a PDB data deposition center in Japan

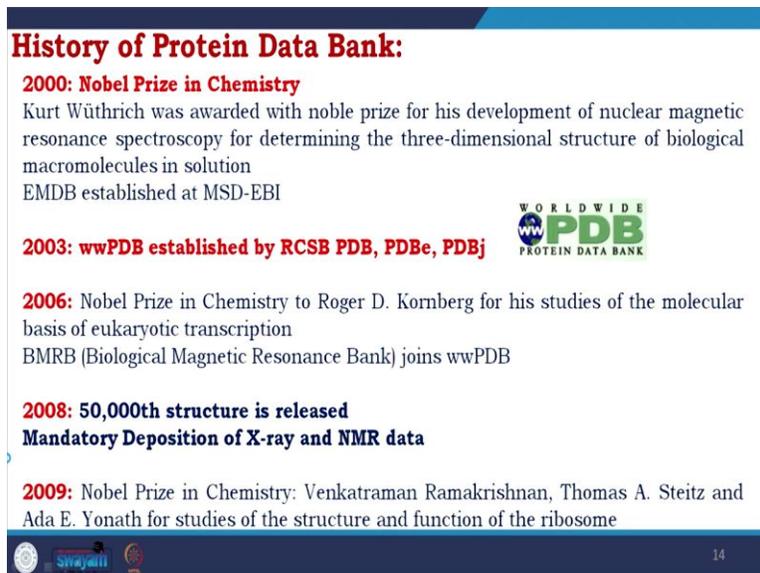
  13

But now, it is also connected in terms of PDB in the 1997 Nobel Prize for the mechanism of adenosine synthesis. So, the Nobel Prize was awarded in chemistry to Paul D Boyer, and John E Walker for their elucidation of the enzymatic mechanism underlying the synthesis of adenosine. Again, it was critical because this is the first elucidated mechanism about the nucleic acids.

1998: inclusion of other databases. In 1998 that was started the inclusion of other databases. The EBI database at European Bioinformatics Institute becomes PDBe or PDB Europe. It is transferred to the Research Collaboratory for Structural Bioinformatics or RCSB.

In 1999 10000 structure was released from PDB. In 2000 the first ribosome structure in the PDB and Osaka University opened a PDB Data Deposition Centered in Japan called PDBj.

(Refer Slide Time: 19:11)



History of Protein Data Bank:

2000: Nobel Prize in Chemistry
Kurt Wüthrich was awarded with noble prize for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution
EMDB established at MSD-EBI

2003: wwPDB established by RCSB PDB, PDBe, PDBj

2006: Nobel Prize in Chemistry to Roger D. Kornberg for his studies of the molecular basis of eukaryotic transcription
BMRB (Biological Magnetic Resonance Bank) joins wwPDB

2008: 50,000th structure is released
Mandatory Deposition of X-ray and NMR data

2009: Nobel Prize in Chemistry: Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath for studies of the structure and function of the ribosome

14

2000 a Nobel Prize again in chemistry. Kurt Wuthrich was awarded the Nobel Prize for developing nuclear magnetic resonance spectroscopy to determine the 3-dimensional structure of biological macromolecules in solution. We have talked about that when we are going through NMR spectroscopy, it is a critical event in structural biology. EMDb established the electron microscopy database we talked about at MSD-EBI, The European Bioinformatics Institute.

2003 is a very important year for PDB wwPDB established by RCSB PDB, PDBe, and PDBj. In 2006 Nobel Prize in chemistry to Roger D Kornberg for his stories of the molecular basis of eukaryotic transcription, another landmark discovery in structural biology.

BMRB, which we talked about initially, the biological magnetic resonance bank joins the wwPDB. In 2008, a 50000 structure was released, and if another very, very significant thing happened, it became mandatory to deposit the x-ray and NMR data if you want to publish. In 2009 Nobel Prize in chemistry Venkatraman Ramakrishnan, Thomas A Steitz, and Ada E Yonath for studies of the structure and function of the ribosome.

(Refer Slide Time: 21:33)

History of Protein Data Bank:

2010: Worldwide Protein Data Bank Foundation established
Deposition of chemical shift data mandatory for NMR data

2011: First structure from an **X-ray Free Electron Laser** (Chapman *et al.*, 2011)

2012: Nobel Prize in Chemistry: Robert J. Lefkowitz and Brian K. Kobilka *for studies of G-protein-coupled receptors*

2013: Mandatory validation report

Validation Reports required for manuscript submission to Journal of Biological Chemistry

Nobel prize in Chemistry was awarded to Martin Karplus, Michael Levitt and Arieh Warshel for the development of multiscale models for complex chemical systems



In 2010 worldwide protein data bank foundation was established. So, now they make a worldwide protein data bank foundation wwPDB was there it is Worldwide Protein Data Bank Foundation, Deposition of Chemical Shift Data mandatory for NMR data.

In 2011 first structure from an x-ray free-electron lesser. In 2012, again, Nobel Prize in chemistry to Robert Lepkowitz , and Brian K Kobielka for G protein coupled receptors studies. Again, a very, very landmark discovery because G protein coupled receptors are already known to be associated with a lot of drug transport and a lot of disease-related phenomenas.

In 2013, structural biology is definitely still in the leading role, but many computational initiatives are grown. In 2013, with a thorough meeting between the structural biologists and computational biologists worldwide, they made the validation report mandatory.

The same year, the Nobel Prize in chemistry was awarded to Martin Kerplus, Michael Levitt , and Arieh Warshel to develop multiscale models for complex chemical systems. It is given to dynamic study, and biology could never be understood perfectly at the static level.

(Refer Slide Time: 27:37)

History of Protein Data Bank:

2014: Integration of large structures with the main PDB archive using PDBx/mmCIF (previously split across multiple entries in traditional PDB file format)

>100,000 structures in the archive

OneDep: New system for Deposition-Validation-Biocuration
UNESCO celebrates the International Year of Crystallography

2016: Deposition of EM maps mandatory
New Validation Reports provided for NMR, 3DEM

2017: Nobel Prize in Chemistry Jacques Dubochet, Joachim Frank, Richard Henderson *for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution*

PDB-Dev: A Prototype System for Integrative/Hybrid Structural Models announced in *Structure*



In 2014 integration of large structures with the main PDB archive using PDBx/mm CIF (previously split across multiple entries in traditional PDB file format). So, now you get them together with the introduction of mm CIF. Now greater than one lakh structures in the archive, another significant step introduction of OneDep, a global deposition system called a new system for deposition, validation, and biocuration. It becomes user-friendly. It becomes automatically validated. So, it improves the quality of the structure. UNESCO celebrates the year 2014, the international year of crystallography. In 2016 the deposition of electron microscopy maps mandatory. The year 2016 was considered cryo-electron microscopy as a technique of the year, followed by the Nobel Prize to Dubochet, Henderson and Joachim Frank.

In 2016 new validation report was provided for NMR, and 3-dimensional electron microscopy. In 2017, I talked about the Nobel Prize in chemistry to Jack Dubochet, Joachim Frank, Richard Henderson for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution.

(Refer Slide Time: 30:21)

History of Protein Data Bank:

2018: Nobel Prize awarded to Frances H. Arnold for the Directed Evolution of Enzymes

2019: Improved ligand validation
>150,000 structures in the PDB archive

2020: Nobel Prize in Chemistry: Emmanuelle Charpentier and Jennifer A. Doudna *for the development of a method for genome editing*
Nobel in Physiology or Medicine: Harvey J. Alter, Michael Houghton and Charles M. Rice *for the discovery of hepatitis C virus*
First **SARS-CoV-2 structure** in the PDB (February 5); COVID-19 main protease
Validation Reports for 3DEM model coordinates and maps
Improved carbohydrate data and representation released

2021: PDB50: Golden Anniversary of the PDB



In 2018 Nobel Prize was awarded to Francis H Arnold for the directed evolution of enzyme, or enzyme engineering. So, enzyme engineering is a very modern technology with modern demands, and this is only possible by utilizing high-resolution structures. In 2019 there is improved ligand validation, greater than one lakh fifty thousand structure entries in the PDB archive.

In the 2020 Nobel Prize in chemistry, Emmanuel Charpentier and Jennifer A Doudna for developing a method of genome editing the CRISPR technology revolutionized the world of molecular biology the target based modifications. In the same year Nobel Prize in physiology medicine to Harvey Alter, Michael Houghton, and Charles M Rice for the discovery of the hepatitis C virus.

Validation reports for the 3D electron microscopy model coordinate and maps. Improve carbohydrate data and representation released, and 2021 this year where you are standing is PDB 50 the golden anniversary of the protein data bank starting in 1971. It touched the 50 years of its journey.

(Refer Slide Time: 32:47)

Missions of PDB:

Manage the wwPDB Core Archives as a public good according to the FAIR Principles

Provide expert deposition, validation, biocuration, and remediation services at no charge to Data Depositors worldwide

Ensure universal open access to public domain structural biology data with no limitations on usage

Develop and promote community-endorsed data standards for archiving and exchange of global structural biology data

What is the vision of PDB? Sustained, freely accessible, interoperating core archives of structured data and metadata for biological macromolecules as an enduring public good to promote basic and applied research and education across the science. What are the missions of PDB? According to the fair principles, managing the wwPDB core archive as a public good provides expert deposition, validation, biocuration, and remediation services at no charge to data depositors worldwide.

Ensure universal open access to public domain structural biology data with no limitation on usage. For this reason we could access any structure at any moment at free of cost.

Develop, and promote community, endorse data standards for archiving an exchange of global structural biology data.

(Refer Slide Time: 34:13)

PDB Functional Components:

- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

Deposit: Prepare, Validate and Deposit

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

20

So, if you look at the functional component of how PDB works, they could divide into six parts deposit, search, visualize, analyze, download, and learn. The deposit has three stages prepare, validate, and deposit for data extraction `pdb_extract` (extract and harvest data in PDBx / mmCIF format from structure determination programs) `SF-tool` (converts the structure factor files among various formats). So, these are for large biological macromolecules or biological molecules. For small molecules, you have `Ligand expo`(it 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, `MAZIT` translate data between file formats and more.

(Refer Slide Time: 35:55)

Validation reports: Deposit: Prepare, Validate and Deposit

 Deposit	<p>Validation reports contain an assessment of the quality of a structure and highlight specific concerns by considering the coordinates of the model, the experimental data and the fit between the two</p>
 Search	<p>Easily interpretable summary information that compares the quality of a model with that of other models in the archive will help users of PDB data to critically assess archived entries and to select the most appropriate structural models for their needs</p>
 Visualize	<p>These reports are developed using the recommendations of the wwPDB Validation Task Forces</p>
 Analyze	<p>Reports for released entries are available from Structure Summary pages</p>
 Download	<p>Validation reports for manuscript reviewers are created during annotation of deposited structures</p>
 Learn	


21

The validation report assesses the quality of a structure and highlights specific concerns by considering the coordinates of the model, the experimental data, and the fit between the two. So, it needed to be fit between your experimental data, and the model you have developed. Easily interpretable summary information that compares the quality of a model with that of other models in the archive will help users of PDB data assess archive entries critically and select the most appropriate structural models for their needs.

These reports are developed using the recommendation of the wwPDB validation task forces. Reports for released entries are available from structure summary pages. Validation reports for manuscript reviewers are created during the annotation of deposited structures.

(Refer Slide Time: 37:50)

wwPDB OneDep System 

This is the video tutorial of OneDep
Deposition system

Deposition ID

Password

Step-by-Step Tutorials:
[Coordinate Deposition](#)
[NMR-specific Depositions](#)
[Electron Microscopy Volume Map Depositions](#)

[Link:https://www.youtube.com/watch?v=yFMG_cFebTA&t=3s](https://www.youtube.com/watch?v=yFMG_cFebTA&t=3s)

22

Then as I told the deposition is in one dep system. You have to log in, and you have to deposit.
(Refer Slide Time: 38:49)

Search:

Search Options

- Basic Search
- Advanced Search
- Sequence Search
- Unreleased & New Entries
- Browse by Annotation
- Explore the Archive
- PDB Statistics
- Ligands
- Drugs & Drug Targets

23

PDB has made a lot of different options for searching basic search, advanced search, unreleased, and new entries, browse by annotation, explore the archive, PDB statistics, ligand, drug and drug targets.

(Refer Slide Time: 39:08)

Viewers: 3D Structure Viewers Provides a graphical summary of biological and structural protein features of PDB entities and how they correspond to UniProtKB sequences. It loads features from RCSB and third party resources such as UniProtKB, CATH or SCOPe

Sequence
Protein Feature View
Genome View

Illustrates the correspondences between PDB Entity sequences and genomes. When the relationships between PDB Entities and genes are available, PDB Entity sequences are mapped to their genome positions to show which regions of a gene are available in the Entity coordinates.



PDB ID: 4HHB **PDB ID: 3J3Q** **PDB ID: 1BTN**

24

Viewers mean the platform to make the view. So, one is a 3D structure viewer, and there are sequence viewers also. One is the protein feature view. It provides a graphical summary of PDB entities' biological and structural protein features and how they correspond to the UniProtKB sequences.

(Refer Slide Time: 41:02)

Download: **Download Coordinate & Experimental Data Files** ^{ST/2/GDN}

By entering PDB IDs, multiple files can be downloaded in batches containing one or more file formats

Coordinate Data Files can be downloaded in the following formats:

- PDB
- PDBx/mmCIF
- PDBML/XML
- PDBML/XML (Header only)
- Biological Assemblies in PDB
- Biological Assemblies in PDBx/mmCIF

Experimental Data Files can be downloaded in the following formats:

- Structure Factors
- NMR Restraints
- Chemical Shifts
- NMR Restraints v2

25

You could download coordinate and experimental data files. What you have to do is you have to enter the PDB id. Coordinate data files can be downloaded in the following formats PDB, PDBx, mmCIF, PDBML XML with the header, biological assemblies in PDB, biological assemblies in PDBx mmCIF. Experimental data files can be downloaded in the following formats—structure factors NMR restraints, chemical shifts, and NMR restraint in a new version.

(Refer Slide Time: 42:33)

Learn:

- [Deposit](#)
- [Search](#)
- [Visualize](#)
- [Analyze](#)
- [Download](#)
- [Learn](#)

[Molecule of the Month](#)
Presents short accounts on selected molecules from the Protein Data Bank

[Educational Resources](#)
Access materials that promote exploration in the world of proteins and nucleic acids

[Curricula](#)
Authentic, hands-on teaching materials, individual and group activities

[News and Events](#)
Upcoming meetings and events RCSB will hold

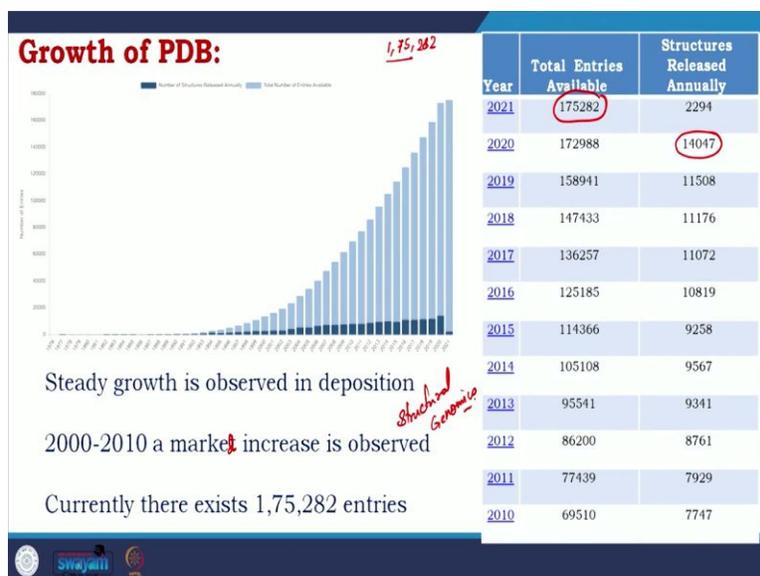
[Guide to PDB Data](#)
Understanding PDB Data is a reference to help explore and interpret individual PDB entries

[Geis Digital Archive](#)
View iconic illustrations by the gifted artist Irving Geis (1908-1997) in context with PDB structures and educational information

Swayamii 26

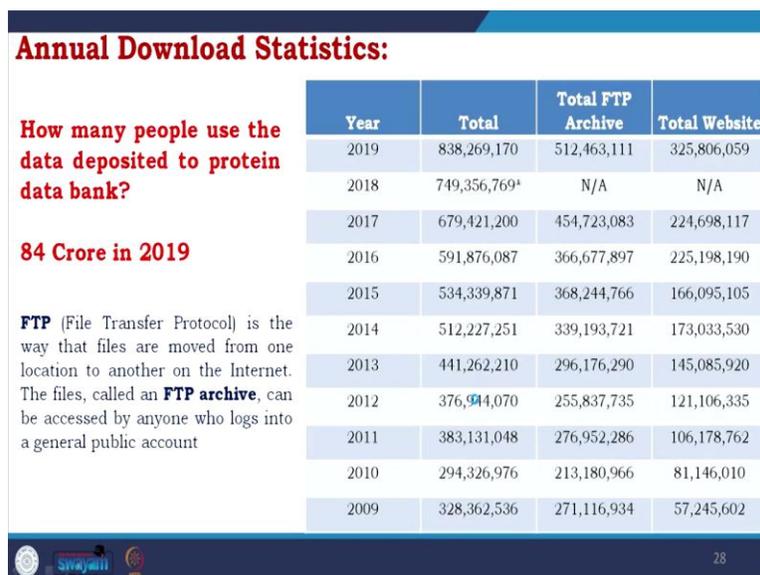
Learning protein data bank has contributed significantly towards the learning of common students regarding structural development and how those biological macromolecular structures have contributed to biology. David Goodsell is a name who is continuously writing about the molecule of the month. Molecule of the Month is a learning session of PDB where the best or potentially most important structures are solved in a current month have immense biological importance. David Goodsell writes about that.

(Refer Slide Time: 44:37)



Now I will talk about the growth of PDB as you see starting in 1970 now it is growing, and if you look at the number as I have already told the number is 175282 total entry. And each year, the entry is increasing currently. In the last year 2020 we have 14000 new entries. So, steady growth is observed in the deposition of different structural entries from 2000 to 2010. A marked increase is observed because of the introduction of structural genomics, which I talked about in the x-ray crystallography part.

(Refer Slide Time: 46:16)



So, when they have downloaded structure, all this comes here this is 51 crore, and through website around 32.6 crores.

(Refer Slide Time: 47:12)

PDB Entries Available (Year Wise):

Year	Total Entries Available	Structures Released Annually
2010	69510	7747
2009	61763	7291
2008	54472	6904
2007	47568	7134
2006	40434	6407
2005	34027	5334
2004	28693	5149
2003	23544	4147
2002	19397	2994
2001	16403	2814
2000	13589	2627
1999	10962	2356
1998	8606	2057
1997	6549	1565
1996	4984	1172
1995	3812	941
1994	2871	1289
1993	1582	696
1992	886	192

↑
Shuc!

It is a huge number as I said, there is a marked increase you see in 1992 192 696 1289 it reduced a little bit but then a steady increase. So, this is the major time of structural genomics.

(Refer Slide Time: 47:38)

PDB Current Holdings Breakdown:

Molecular Type	X-ray	NMR	EM	Multiple methods	Neutron	Other	Total
Protein (only)	136983	11613	4725	166	67	32	153586
Protein/Oligosaccharide	8093	31	655	5	0	0	8784
Protein/NA	7226	270	1632	3	0	0	9131
Nucleic acid (only)	2166	1347	53	7	2	1	3576
Other	149	31	3	0	0	0	183
Oligosaccharide (only)	11	6	0	1	0	4	22
Total	154628	13298	7068	182	69	37	175282

This is very important this is the current holding breakdown at the protein data bank. If you look at that x-ray 136983 protein only NMR 11613 only protein, electron microscopy 4725 there are other methods, neutron diffraction combined methods, and all but if you combine protein, oligosaccharide for x-ray 8000, electron microscopy 655 versus NMR 31. Protein nucleic acid again 7226 for x-ray 270 for NMR, and 1632 for electron microscopy showing the steady increase of a potential technique discussed in our previous classes.

So, total through x-ray 154628 structures through NMR 13298 electron microscopy 7000, but if you remember, we have shown in 2017 to 18 EM have already started superseding NMR.

(Refer Slide Time: 49:36)

Experimental Methods:	Experimental Method	Entries
<p>There are number of methods which contributes to PDB</p> <p>X-ray clearly dominates, followed by Solution state NMR</p> <p>Cryo EM is upcoming with amazing potential</p>	X-ray Crystallography	154693
	Solution NMR	13223
	Cryo EM	6911
	Neutron Diffraction	177
	Electron Crystallography	172
	Solid State NMR	140
	Solution Scattering	71
	Fiber Diffraction	39
	Powder Diffraction	20
	EPR	08
	Infrared Spectroscopy	04
	Fluorescence Transfer	01