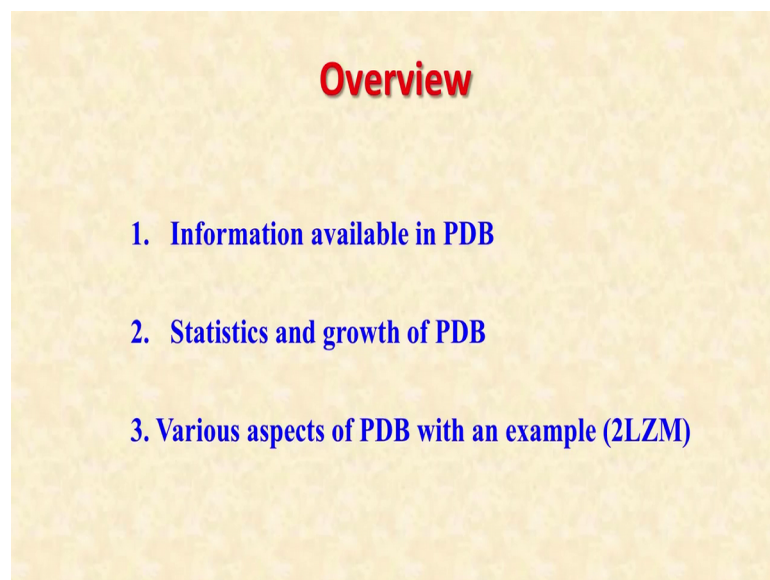


Bioinformatics
Prof. M. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 38
Demo of Protein Data Bank

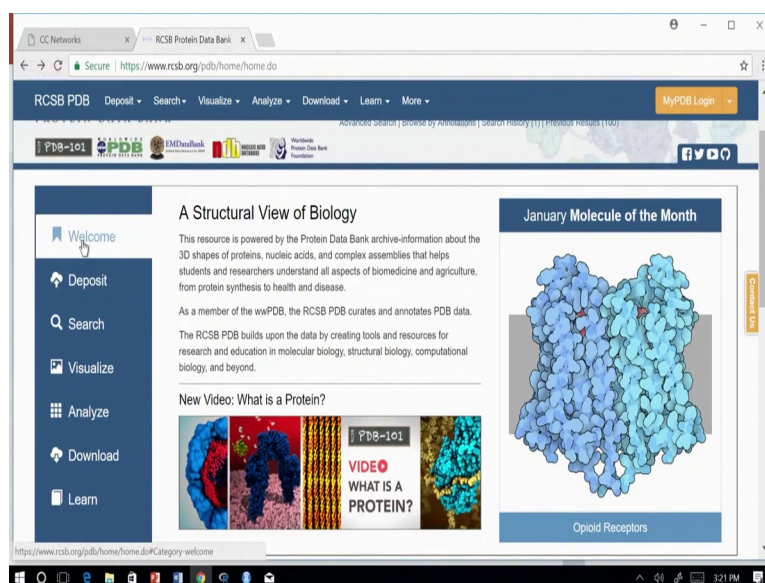
Demonstration on Protein Data Bank: in early lectures I discussed about various aspects of PDB that is Protein Data Bank and in this demonstration, we will mainly show about the information available in protein data bank.

(Refer Slide Time: 00:34)



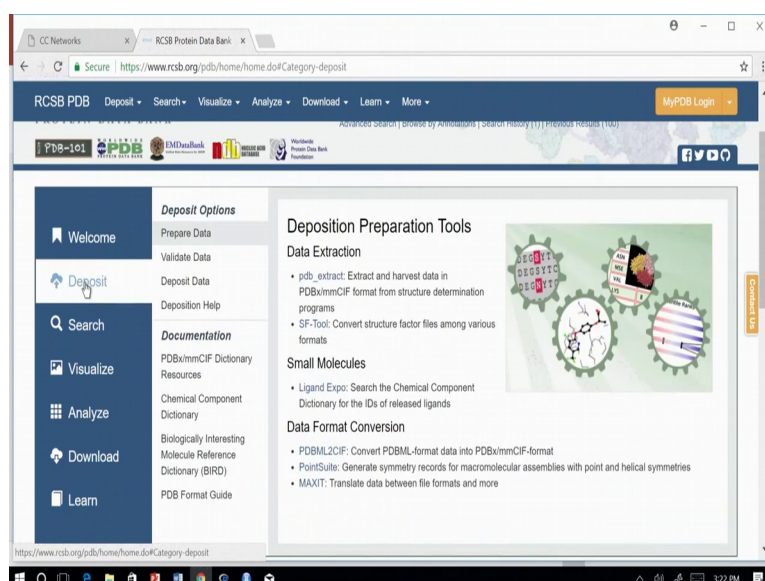
And the current statistics as well as the growth of PDB from is started in the national medical laboratory, and specifically various aspects and duplications of the protein data bank with a specific example say last time with the PDB code of 2 LZM. The PDB or protein data bank is maintained by the research collaboratory for structural bioinformatics or RCSB.

(Refer Slide Time: 01:01)



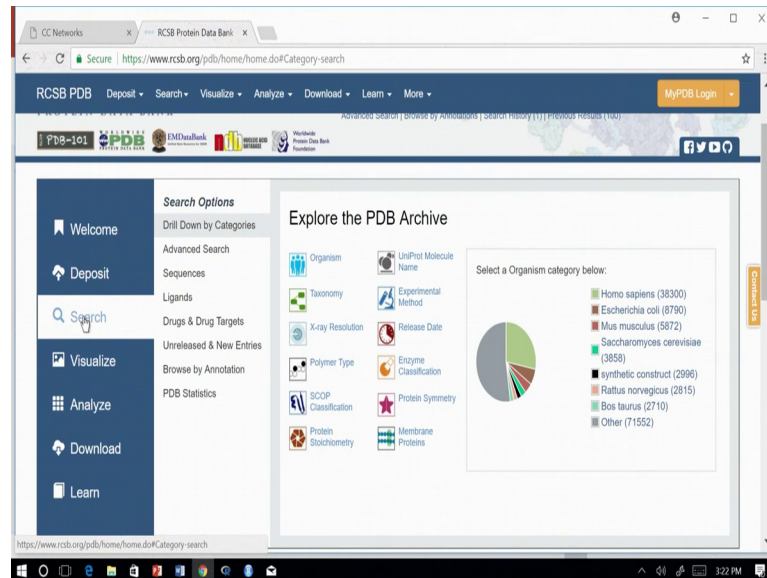
This structure are stored in terms of atom coordinates, each structures given a unique PDB id, consisting of a numeral and 3 alpha numeric characters such as 2 LZM or 1 a 4 y PDB is freely accessible to the public. We will learn about the PDB and how to navigate the website in the following demo. First go to the url rcsb dot org for slash PDB you can click on each menu item on the left side to learn about the resources available in PDB.

(Refer Slide Time: 01:41)



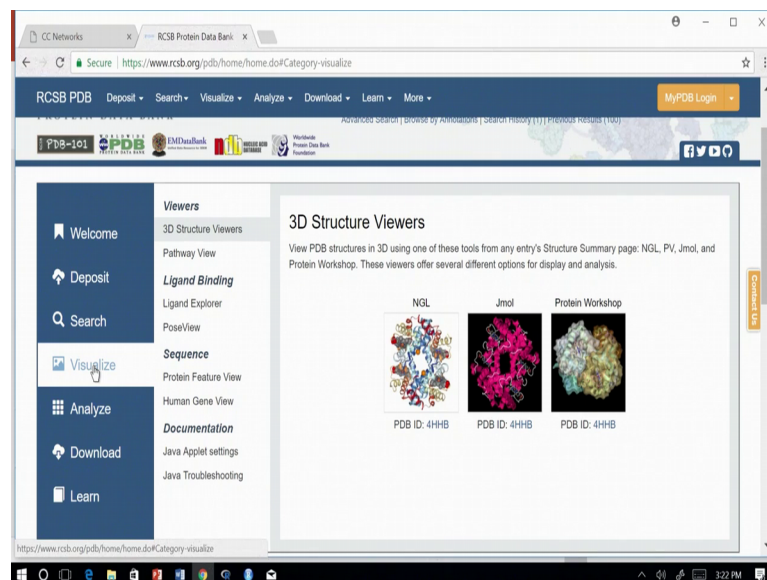
In the deposit menu item experimentalist can deposit their data, which is a biological structures illustrated by X-Ray crystallography or NMR etc.

(Refer Slide Time: 01:54)



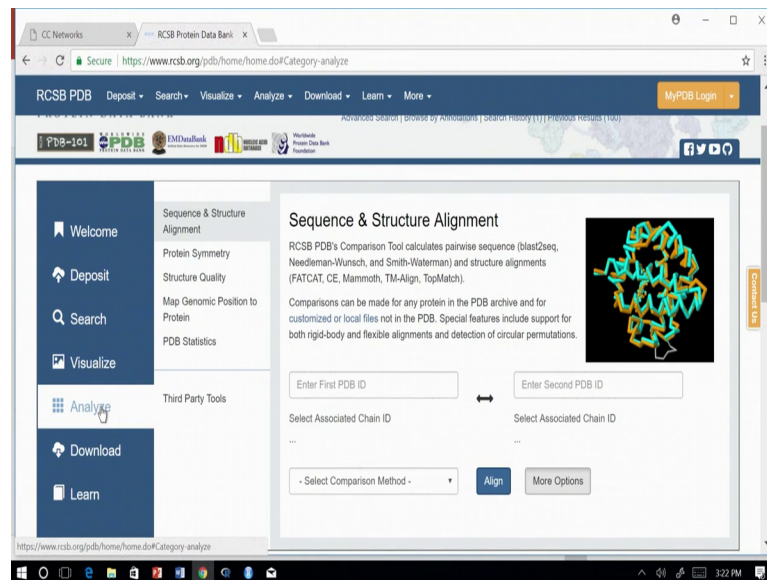
The search tab is for uses to search for a given crystal structure.

(Refer Slide Time: 01:58)



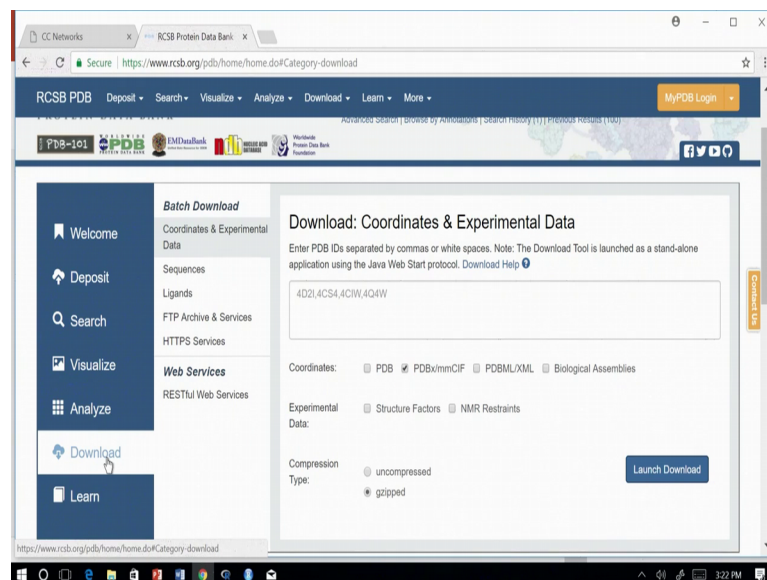
This visualized tab gives software to view the 3 D structures, sequence features or interactions with ligands.

(Refer Slide Time: 02:07)



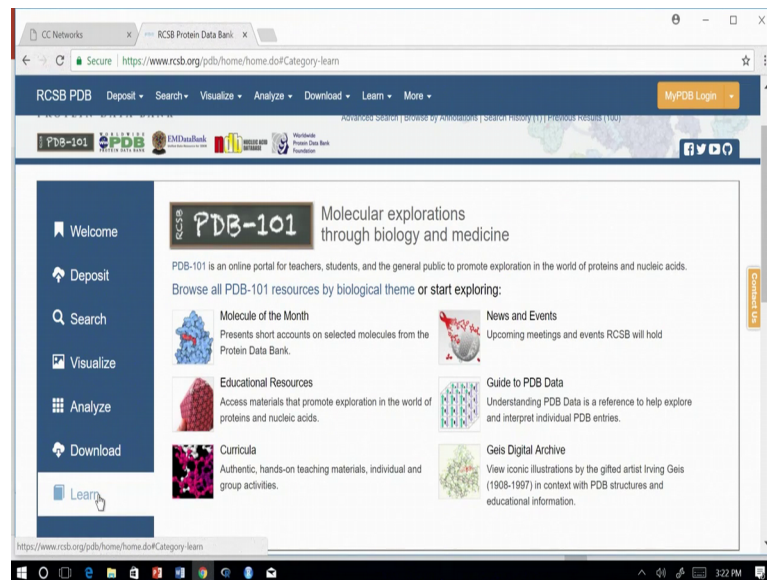
The analyze tab gives in house sequence and structure alignment tools links to the PDB statistics page as well as external tools.

(Refer Slide Time: 02:18)



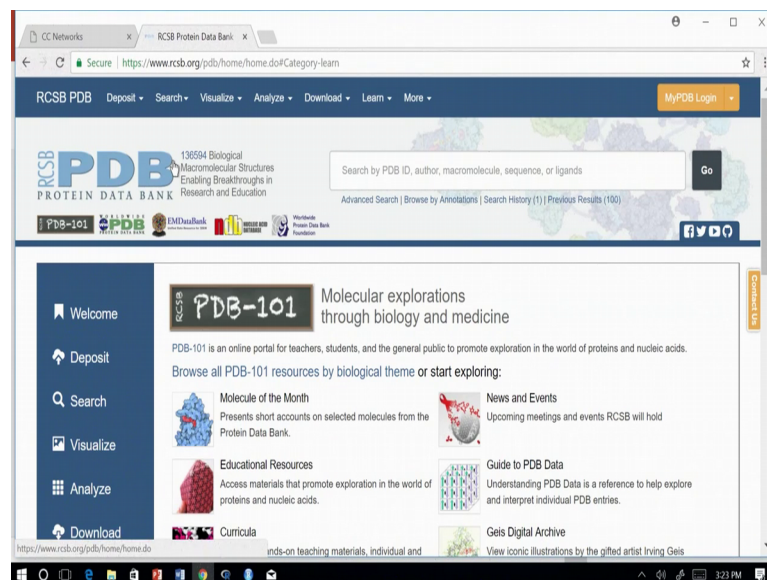
The download tab allow users to have bulk or customize data download and API services.

(Refer Slide Time: 02:27)



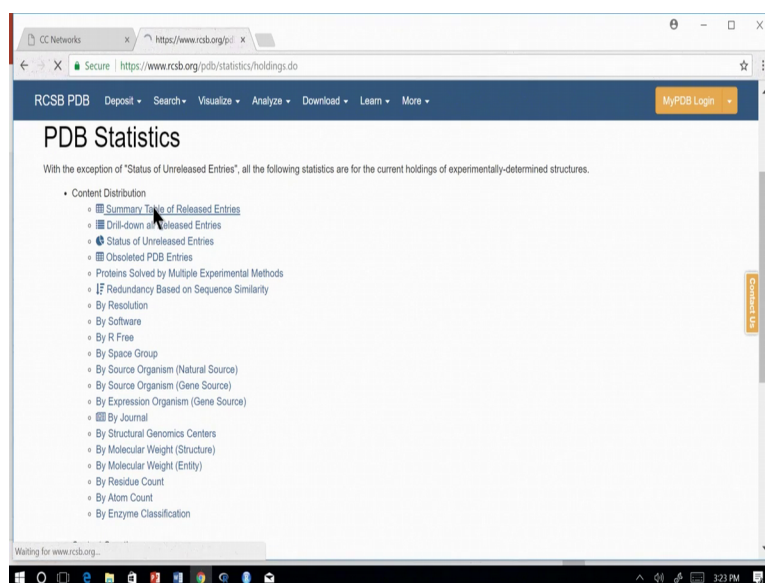
In the learn tab, users can access educational resources about PDB and molecular biology.

(Refer Slide Time: 02:35)



In the header tab here we can see the number of data available in PDB, the number of crystal structures which have been deposited. Now we will see the statistics of PDB data base this can be found here.

(Refer Slide Time: 02:56)



The statistics are grouped under two sections: content distribution and content growth content distribution. Content statistics on the data currently deposited in PDB; the summary table which is displayed here.

(Refer Slide Time: 03:15)

The screenshot shows the 'PDB Current Holdings Breakdown' table on the RCSB PDB website. The table has columns for 'Exp. Method', 'Proteins', 'Nucleic Acids', 'Protein/NA Complexes', 'Other', and 'Total'. The 'X-RAY' row is highlighted with a mouse cursor. Below the table, there is a note: '(Click on any number to retrieve the results from that category.)' followed by four lines of text: '112086 structures in the PDB have a structure factor file.', '9425 structures in the PDB have an NMR restraint file.', '3177 structures in the PDB have a chemical shifts file.', and '1887 structures in the PDB have a 3DEM map file.'

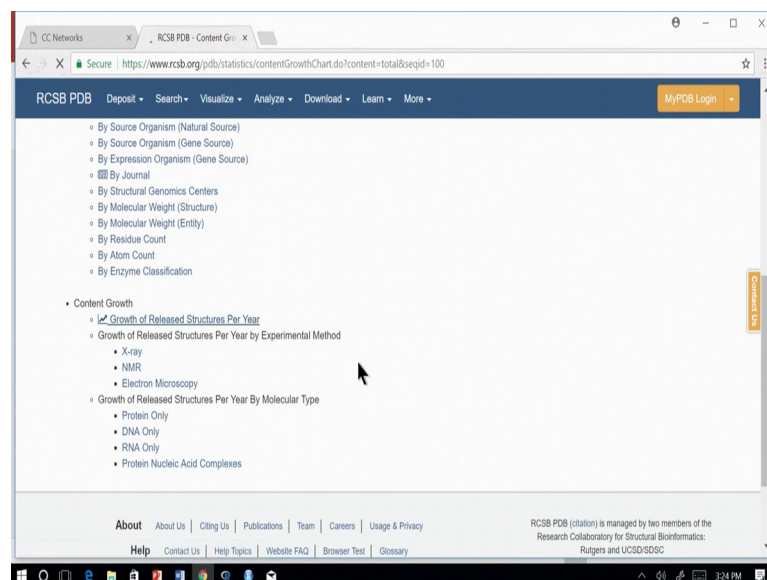
Exp. Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	114527	1907	5852	4	122290
NMR	10599	1231	247	8	12085
ELECTRON MICROSCOPY	1368	30	484	0	1882
HYBRID	106	3	2	1	112
other	202	4	6	13	225
Total	126802	3175	6591	26	136594

(Click on any number to retrieve the results from that category.)
112086 structures in the PDB have a structure factor file.
9425 structures in the PDB have an NMR restraint file.
3177 structures in the PDB have a chemical shifts file.
1887 structures in the PDB have a 3DEM map file.

Gives a number of structures under proteins, nucleic acids, protein nucleic acids complexes as well as other biological structures and also the breakdown by each experimental technique such as X-Ray NMR or electron microscopy and others the totals are given here. In total there are currently around 136000 structures available in PDB. To

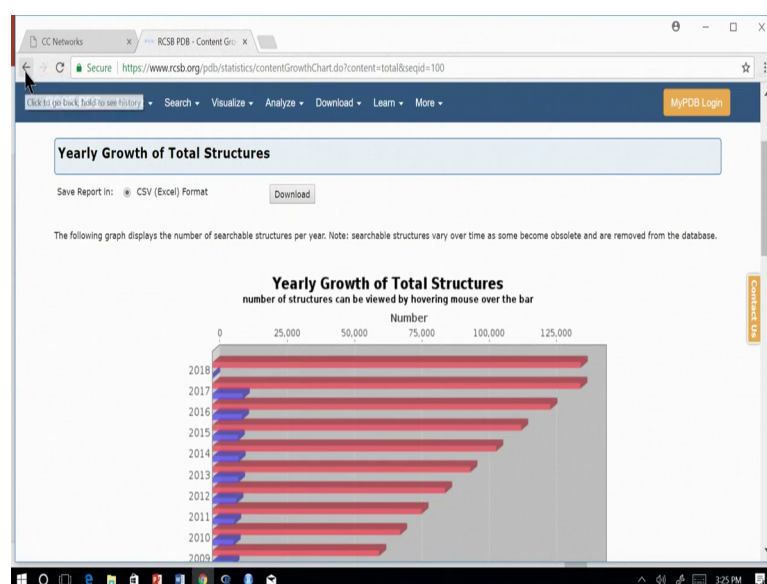
explore further you can see the data by molecular weight by residue count or by the resolution all these data is available for exploration and further study.

(Refer Slide Time: 03:59)



PDB also gives statistics by content growth. This shows how the data base has grown over time statistics have been kept every year since 1976. Let us see the statistics here. The records start from 1976. As you can see the amount of data being deposited in PDB has grown over the years and is currently around here.

(Refer Slide Time: 04:28)



Now, let us let us go for with an example. Now we type 2 LZM in the search bar and press enter.

(Refer Slide Time: 04:51)

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

Structure summary | 3D View | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment | Literature

2LZM

STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION

DOI: 10.2210/pdb2lzm/pdb Entry 2LZM supersedes 1LZM

Classification: [HYDROLASE \(O-GLYCOSYL\)](#)

Deposited: 1986-08-18 Released: 1986-10-24

Deposition author(s): [Weaver, L.H.](#), [Matthews, B.W.](#)

Organism: [Escherichia virus T4](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.7 Å

R-Value Observed: 0.193

wwPDB Validation

Metric	Percentile Ranks	Value
Clashscore	11	11
Ramachandran outliers	6.6%	6.6%
Sidechain outliers	13.8%	13.8%

Literature

Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution.

When we type any PDB id, the website gives a summary page view of the structure. So, we have a static view of the structure here the title as well as basic information about the experimental technique as well as protein structure which is given here. So, example this structure is classified as a hydrolase.

(Refer Slide Time: 05:14)

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

Structure summary | 3D View | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment | Literature

2LZM

STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION

DOI: 10.2210/pdb2lzm/pdb Entry 2LZM supersedes 1LZM

Classification: [HYDROLASE \(O-GLYCOSYL\)](#)

Deposited: 1986-08-18 Released: 1986-10-24

Deposition author(s): [Weaver, L.H.](#), [Matthews, B.W.](#)

Organism: [Escherichia virus T4](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.7 Å

R-Value Observed: 0.193

wwPDB Validation

Metric	Percentile Ranks	Value
Clashscore	11	11
Ramachandran outliers	6.6%	6.6%
Sidechain outliers	13.8%	13.8%

Literature

Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution.

[Weaver, L.H.](#), [Matthews, B.W.](#)

(1987) J Mol Biol. 193: 189-199

PubMed: 3586019 [Search on PubMed](#)

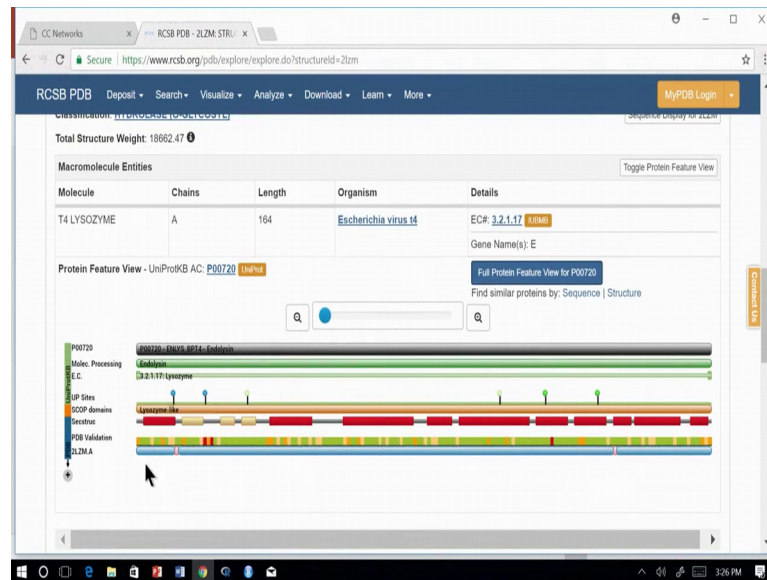
Also Cited By: 1 UQU 1 LLH

PubMed Abstract

The structure of the lysozyme from bacteriophage T4 has been refined at 1.7 Å resolution to a crystallographic residual of 19.3%. The final model has bond lengths and bond angles that differ from "ideal" values by 0.019 Å and 2.7°.

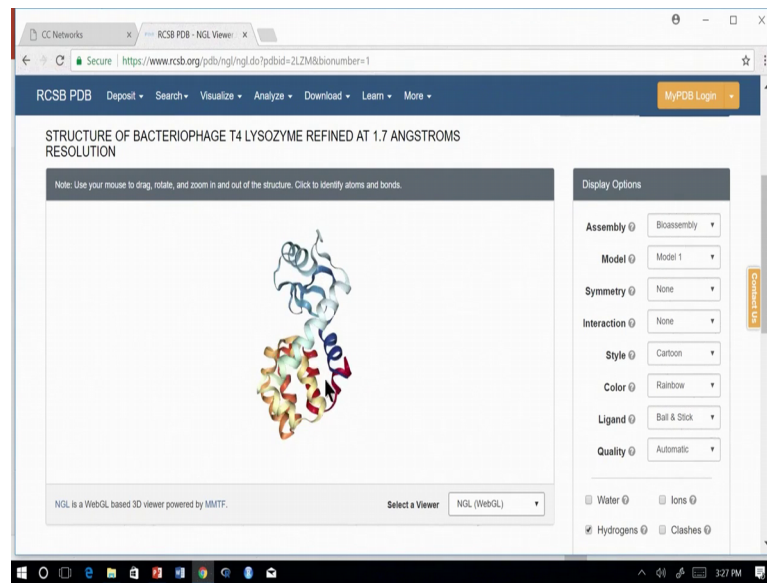
In this experimental method which was used to get the structure is X-Ray diffraction and the resolution is 1.7 angstroms. PDB also provides a literature reference from which the structure was taken.

(Refer Slide Time: 05:29)



Further you can find the macromolecules in this structure. So, in this structure we have only 1 entity that is the T4 LYSOZYME this is identified as chain A and it has 164 residues. You can also see the feature view here, in this page you can also link to other data bases such as pubmed and uniprot and you can also view the Ramachandran plot here. You will move on to see the 3D structure, we can also have 3D structure here or you can go and download the structure and view a 10 pi mole.

(Refer Slide Time: 06:17)



So, in the PDB website itself you can see the structure in the NGL viewer. So, you can drag rotate and see multiple views, we will move on to annotations.

(Refer Slide Time: 06:35)

STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION

Macromolecule Annotations for the Entities in PDB 2LZM

Domain Annotation: SCOP Classification

Chains	Domain Info	Class	Fold	Superfamily	Family	Domain	Species
A	d2lzmA_	Alpha and beta proteins (α+β)	Lysozyme-like	Lysozyme-like	Phage lysozyme	Phage T4 lysozyme	Bacteriophage T4 [Taxid: 10665]

Domain Annotation: CATH

Chains	Domain	Class	Architecture	Topology	Homology
A	2lzmA00	Mainly Alpha	Orthogonal Bundle	Lysozyme	

Protein Family Annotation

Chains	Pfam Accession	Pfam Identifier	Pfam Description	Type	Source
--------	----------------	-----------------	------------------	------	--------

Gene Product Annotation

Chains	Polymer	Molecular Function	Biological Process	Cellular Component
--------	---------	--------------------	--------------------	--------------------

For this entry we have annotations from SCOP CATH P FAM as well as gene product annotations.

(Refer Slide Time: 06:39)

The screenshot shows the RCSB PDB website for the entry 2LZM. The top navigation bar includes links for Deposit, Search, Visualize, Analyze, Download, Learn, and More. The main content area is divided into three sections: Domain Annotation, Protein Family Annotation, and Gene Product Annotation.

Domain Annotation: CATH (CATH Database (version 4.0.0) Homepage)

Chains	Domain	Class	Architecture	Topology	Homology
A	2lzmA00	Mainly Alpha	Orthogonal Bundle	Lysozyme	

Protein Family Annotation (Pfam Database Homepage)

Chains	Pfam Accession	Pfam Identifier	Pfam Description	Type	Source
A	T4 LYSOZYME (2LZM:A)				

Gene Product Annotation (Gene Ontology Consortium Homepage)

Chains	Protein	Molecular Function	Biological Process	Cellular Component
A	T4 LYSOZYME (2LZM:A)	<ul style="list-style-type: none"> Lysozyme Activity Catalytic Activity Hydrolase Activity Hydrolase Activity Acting On Glycosyl Bonds 	<ul style="list-style-type: none"> Metabolic Process Peptidolysan Catabolic Process Cell Wall Macromolecule Catabolic Process Viral Release From Host Cell Cytolysis Defense Response to Bacterium 	<ul style="list-style-type: none"> Host Cell Cytoplasm

So, SCOP and CATH gave an idea about the protein class, which 2 LZM belongs to and P FAM as well.

(Refer Slide Time: 06:59)

The screenshot shows the RCSB PDB website for the entry 2LZM, specifically the 'Sequence Similarity' tab. The title is '2LZM STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION'. Below the title, there are buttons for 'Display Files' and 'Download Files'.

Sequence Similarity Clusters for the Entities in PDB 2LZM

Entity #1 | Chains: A
T4 LYSOZYME protein, length: 164 (BLAST)

Sequence Similarity Cutoff	Rank	Chains in Cluster	Cluster ID / Name	Structural variation in cluster
100 %	7	11	5334	
95 %	232	51	9	Flexibility: Low Max RMSD: 6.3, Avg RMSD: 0.8 PDBlink
90 %	233	574	10	
70 %	233	574	16	
50 %	233	574	22	
40 %	233	574	35	
30 %	233	574	55	

Instructions

In the table for each entity, view a list of similar sequences by selecting the link associated with the percentage cutoff.

[View Table Legend](#)

View more detailed documentation on the redundancy reduction and sequence clustering procedure used by RCSB PDB.

You can also use the structure comparison tool to compare any 2

The sequence tab to know the structures or the sequences in that cluster structural similarity.

(Refer Slide Time: 07:13)

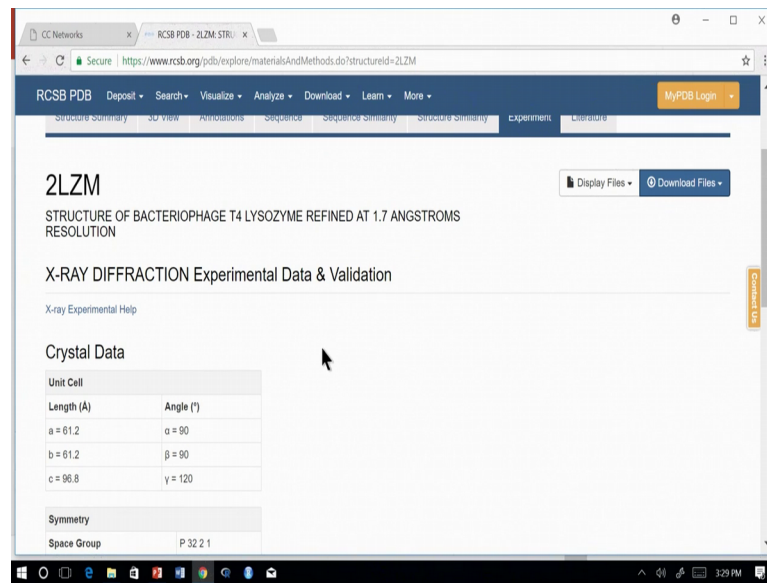
The screenshot shows the RCSB PDB Structure Cluster tool interface. The main heading is "STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION". Below this, the section "Structural Similarities for the Entities in PDB 2LZM" explains that similarities are found using the JFATCAT-rigid algorithm. It notes that a 40% sequence identity clustering has been applied to reduce hits. A button "Show structure comparison results for representative 5JDTA" is visible. On the right, an "Info & Help Documentation" sidebar includes a "View Table Legend" button and text about alignment procedures. The bottom of the page shows a Windows taskbar with the time 3:28 PM.

So, here we can see all the structurally similar entities in PDB which match 2 LZM, structure is done you are the structural, similarity is done using the J FAT CAT rigid algorithm, you can also choose other algorithms here.

(Refer Slide Time: 07:29)

This screenshot shows the same RCSB PDB Structure Cluster tool interface, but with the "Show structure comparison results for representative 5JDTA" button clicked. A dropdown menu labeled "Select Comparison Method" is now visible. Below this, a "References" section lists several scientific papers related to the tool and protein structure analysis. The bottom of the page features a footer with links for "About", "Help", "Citing Us", "Publications", "Team", "Careers", "Usage & Privacy", "Contact Us", "Help Topics", "Website FAQ", "Browser Test", and "Glossary". The Windows taskbar at the bottom shows the time 3:28 PM.

(Refer Slide Time: 07:38)



2LZM
STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION

X-RAY DIFFRACTION Experimental Data & Validation

X-ray Experimental Help

Crystal Data

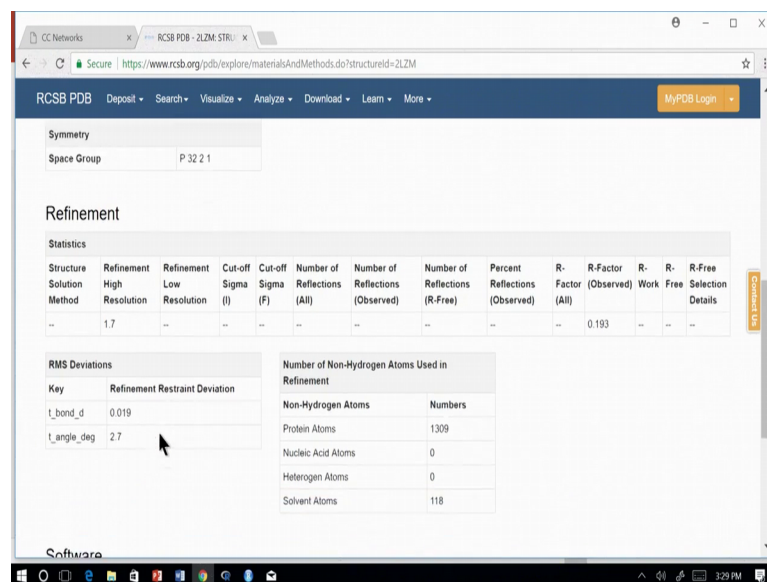
Unit Cell	
Length (Å)	Angle (°)
a = 61.2	α = 90
b = 61.2	β = 90
c = 96.8	γ = 120

Symmetry

Space Group
P 3 ₂ 2 1

Under the experiment tab you can see more metadata about the experiment, which was used to get the structure. Since, this is an X-Ray diffraction experiment you will find crystal data the space group.

(Refer Slide Time: 07:50)



Refinement

Statistics

Structure Solution Method	Refinement High Resolution	Refinement Low Resolution	Cut-off Sigma (I)	Cut-off Sigma (F)	Number of Reflections (All)	Number of Reflections (Observed)	Number of Reflections (R-Free)	Percent Reflections (Observed)	R-Factor (All)	R-Factor (Observed)	R-Work	R-Free Selection Details
--	1.7	--	--	--	--	--	--	--	--	0.193	--	--

RMS Deviations

Key	Refinement Restraint Deviation
t_bond_d	0.019
t_angle_deg	2.7

Number of Non-Hydrogen Atoms Used in Refinement

Non-Hydrogen Atoms	Numbers
Protein Atoms	1309
Nucleic Acid Atoms	0
Heterogen Atoms	0
Solvent Atoms	118

And, as well as others other RMS deviations and software.

(Refer Slide Time: 07:56)

The screenshot shows the RCSB PDB website interface for structure 2LZM. The 'Materials and Methods' tab is selected, displaying a table of refinement statistics. The table includes columns for Method, Resolution, Resolution (F), (F), (All), (Observed), (R-Free), (Observed), (All), and Details. The resolution is 1.7 Å. Below the table, there are sections for RMS Deviations, Number of Non-Hydrogen Atoms Used in Refinement, and Software.

Method	Resolution	Resolution (F)	(F)	(All)	(Observed)	(R-Free)	(Observed)	(All)	Details
--	1.7	--	--	--	--	--	--	0.193	--

RMS Deviations

Key	Refinement Restraint Deviation
t_bond_d	0.019
t_angle_deg	2.7

Number of Non-Hydrogen Atoms Used in Refinement

Non-Hydrogen Atoms	Numbers
Protein Atoms	1309
Nucleic Acid Atoms	0
Heterogen Atoms	0
Solvent Atoms	118

Software

Software	Purpose
TNT	refinement

(Refer Slide Time: 08:07)

The screenshot shows the RCSB PDB website interface for structure 2LZM, specifically the 'Literature' tab. The title is 'STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION'. It displays 'Citations in PubMed' with a primary citation of 3586019. Below this is a bar chart titled 'PDB ID Mentions in PubMed Central' showing the article count per year from 2005 to 2015. The chart shows a general upward trend, peaking in 2014.

STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION

Citations in PubMed

Primary Citation PubMed: 3586019

PDB ID Mentions in PubMed Central (Article count: 20)

Article count per year

Year	Article count
2005	1
2006	0
2007	2
2008	1
2009	0
2010	2
2011	2
2012	2
2013	4
2014	5
2015	1

The literature tab lists the citations from pubmed to the primary citation for this PDB id; this graph gives the mentions of this PDB id in pubmed central. Now how to view the PDB file, if you click here this display file you can see the PDB file.

(Refer Slide Time: 08:24)

The screenshot shows the RCSB PDB website interface. The main header includes the RCSB PDB logo and navigation links. The search bar is at the top right. The entry 2LZM is displayed, with the title "STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION". Below the title, there are links for "Citations in PubMed" and "PDB ID Mentions in PubMed Central". A sidebar on the right offers options to "Display Files" or "Download Files", with a list of available file formats including FASTA Sequence, PDB File, mmCIF File, and PDBML/XML File.

(Refer Slide Time: 08:31)

The screenshot shows the RCSB PDB website interface, specifically the "Header" section for entry 2LZM. The header information is displayed in a table format, including fields such as "TITLE", "AUTHOR", "REMARK", "PDB ID", "DOI", and "PDBML/XML File". The title is "STRUCTURE OF BACTERIOPHAGE T4 LYSOZYME REFINED AT 1.7 ANGSTROMS RESOLUTION". The author is "L.H. HEAVER, B.H. MATTHEWS". The remark section contains detailed information about the structure, including the resolution (1.7 A) and the PDB ID (2LZM).

Here we have the header information, so this gives the title the date it was deposited and it also gives the compound details, as we scroll down we will see a lot of information. So, here we can see the atom coordinates.

(Refer Slide Time: 08:53)

CC Networks

https://files.rcsb.org/view/2LZM.pdb

Secure | https://files.rcsb.org/view/2LZM.pdb

HELIX	1	H1	ILE	A	3	GLU	A	11	1	9	
HELIX	2	H2	LEU	A	39	ILE	A	50	1	11	
HELIX	3	H3	LYS	A	60	ARG	A	80	1	21	
HELIX	4	H4	ALA	A	82	SER	A	90	1	9	
HELIX	5	H5	ALA	A	93	MET	A	106	1	14	
HELIX	6	H6	GLU	A	108	GLY	A	115	5	6	
HELIX	7	H7	THR	A	115	GLN	A	123	1	9	
HELIX	8	H8	TRP	A	126	ALA	A	134	1	5	
HELIX	9	H9	ARG	A	137	GLN	A	141	1	5	
HELIX	10	H10	PRO	A	143	THR	A	155	1	13	
SHEET	1	S1	4	GLY	A	56	ILE	A	58	0	
SHEET	2	S1	4	ARG	A	14	ASP	A	20	-1	
SHEET	3	S1	4	TYR	A	24	ILE	A	27	-1	
SHEET	4	S1	4	MES	A	31	THR	A	34	-1	
CRYST1	61.200	61.200	96.800	96.800	96.800	120.00	P	32	2	1	
ORIGX1	1.154701	0.000000	0.000000	0.000000	0.000000						
ORIGX2	0.577350	1.000000	0.000000	0.000000	0.000000						
ORIGX3	0.000000	0.000000	1.000000	0.000000	0.000000						
SCALE1	0.018868	0.000000	0.000000	0.000000	0.000000						
SCALE2	0.009434	0.016340	0.000000	0.000000	0.000000						
SCALE3	0.000000	0.000000	0.010331	0.000000	0.000000						
ATOM	1	N	MET	A	1	36.644	-24.949	8.853	1.00	29.12	N
ATOM	2	CA	MET	A	1	36.942	-23.582	8.894	1.00	19.55	C
ATOM	3	C	MET	A	1	35.712	-22.887	9.526	1.00	22.27	C
ATOM	4	O	MET	A	1	34.626	-23.375	9.258	1.00	18.31	O
ATOM	5	C8	MET	A	1	37.385	-23.890	7.599	1.00	8.40	C
ATOM	6	CG	MET	A	1	37.639	-21.683	7.544	1.00	30.36	C
ATOM	7	SD	MET	A	1	39.389	-21.186	7.226	1.00	39.80	S
ATOM	8	CE	MET	A	1	40.241	-22.126	8.356	1.00	44.83	C
ATOM	9	N	ASN	A	2	35.890	-21.796	10.310	1.00	17.74	N
ATOM	10	CA	ASN	A	2	34.889	-21.015	10.910	1.00	5.91	C
ATOM	11	C	ASN	A	2	35.236	-19.557	10.931	1.00	11.34	C
ATOM	12	O	ASN	A	2	36.390	-19.244	10.620	1.00	9.58	O
ATOM	13	CB	ASN	A	2	34.487	-21.602	12.355	1.00	9.68	C
ATOM	14	CG	ASN	A	2	35.645	-21.566	13.309	1.00	11.82	C
ATOM	15	OD1	ASN	A	2	36.176	-20.496	13.515	1.00	18.42	O
ATOM	16	ND2	ASN	A	2	36.813	-22.685	13.919	1.00	9.43	N
ATOM	17	N	ILE	A	3	34.315	-18.689	11.287	1.00	6.65	N
ATOM	18	CA	ILE	A	3	34.543	-17.262	11.353	1.00	10.61	C

These are the this is the most important part of the PDB file these are the atom coordinates these are the x y and z coordinates, this is the occupancy and this is a B factor column. Here we can see the annotation for number of this is the atom number, this is the atom name, this is a residue the chain and this is the residue number. If you wish to download the PDB file, you can click here and download the file as a text or dot GZ file.