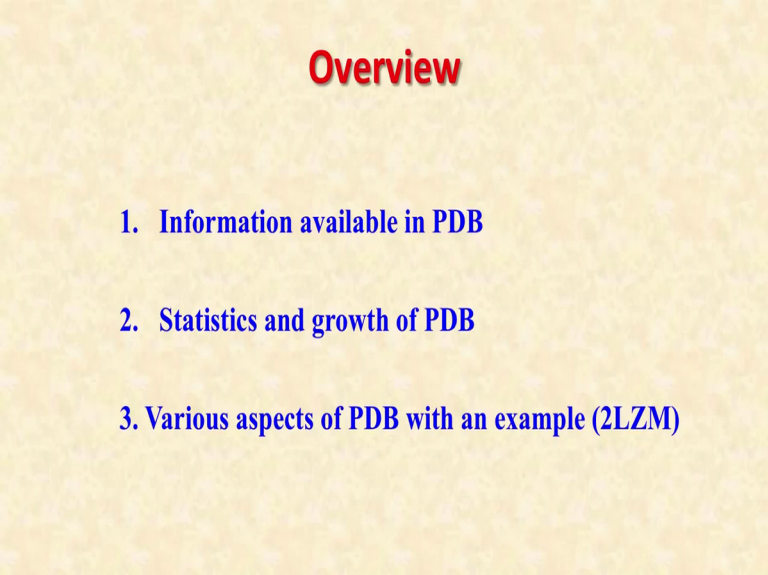
Bioinformatics
Prof. M. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 38
Demo of Protein Data Bank

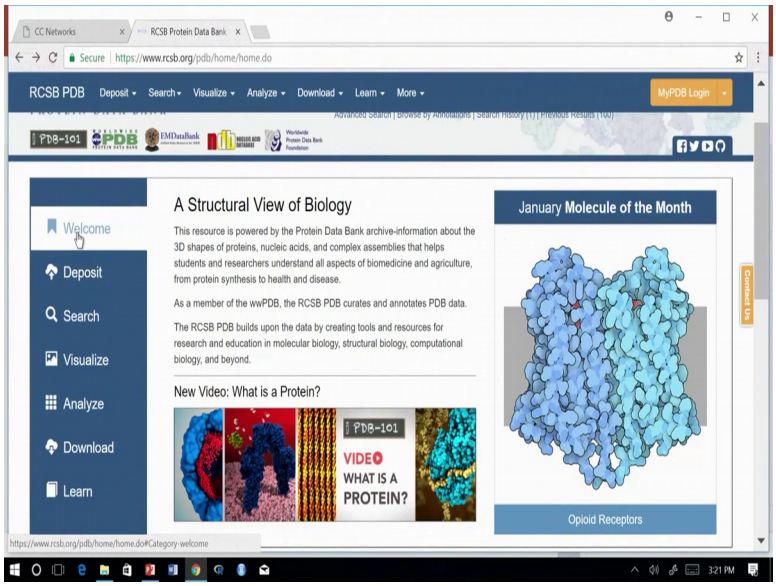
Demonstration on Protein Data Bank: in early lectures I discussed about various aspects of PDB that is Protein Data Bank and in this demonstration, we will mainly show about the information available in protein data bank.

(Refer Slide Time: 00:34)



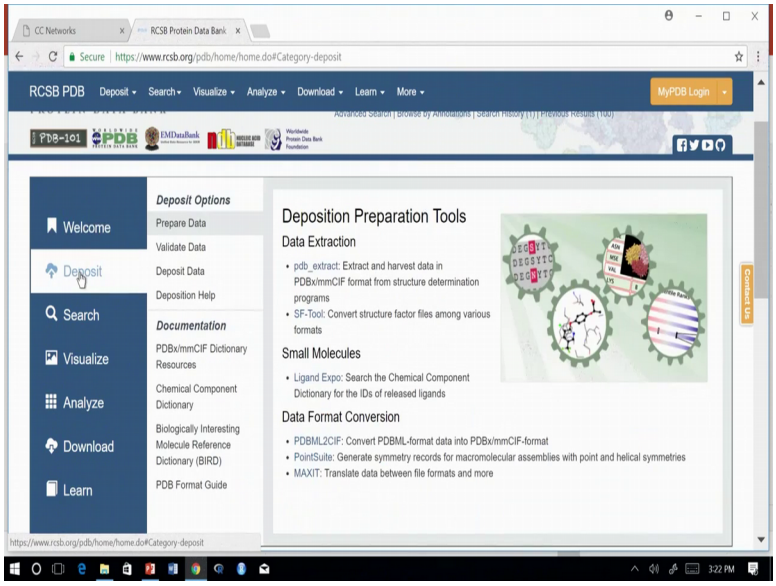
And the current statistics as well as the growth of PDB from is started in the national medical laboratory, and specifically various aspects and duplications of the protein data bank with a specific example say last time with the PDB code of 2 LZM. The PDB or protein data bank is maintained by the research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

(Refer Slide Time: 01:01)



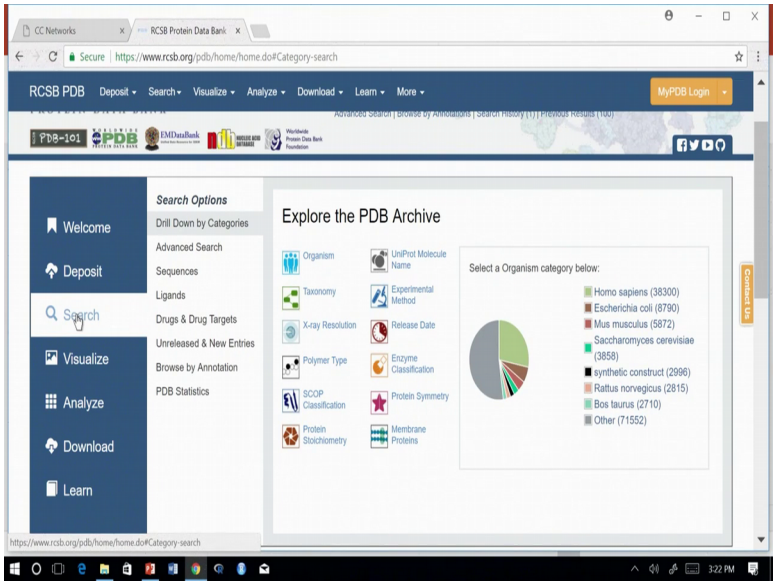
This structure are stored in terms of atom coordinates, each structures given a unique PDB id, consisting of a numeral and 3 alpha numeric characters such as 2 LZM or 1 a 4 y PDB is freely accessible to the public. We will learn about the PDB and how to navigate the website in the following demo. First 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)



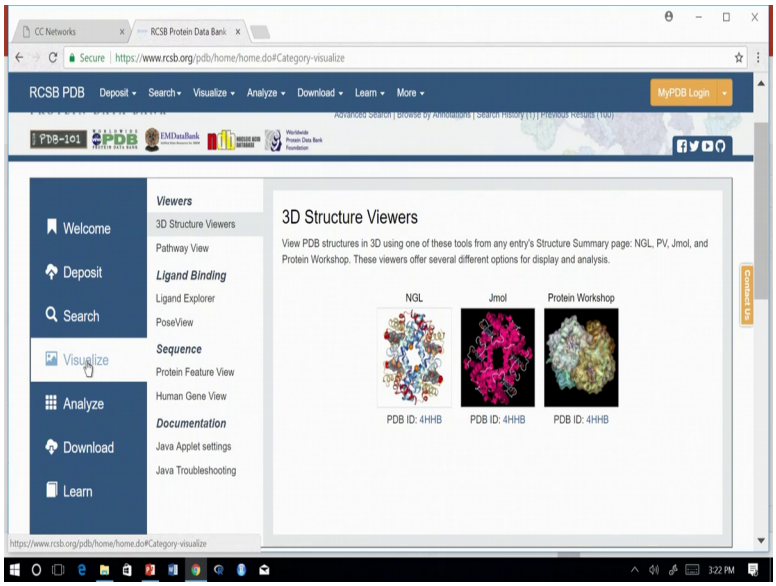
In the deposit menu item experimentalist can deposit their data, which is a biological structures illustrated by X-Ray crystallography or NMR etc.

(Refer Slide Time: 01:54)



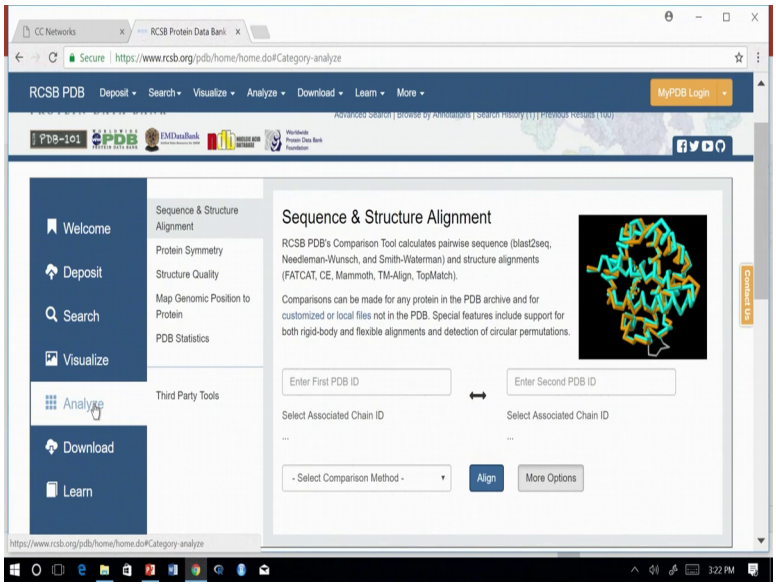
The search tab is for uses to search for a given crystal structure.

(Refer Slide Time: 01:58)



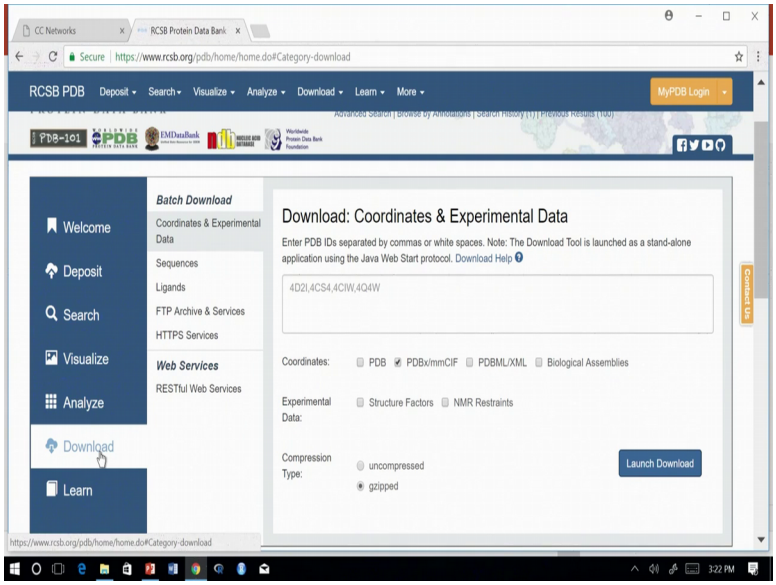
This visualized tab gives software to view the 3 D structures, sequence features or interactions with ligands.

(Refer Slide Time: 02:07)



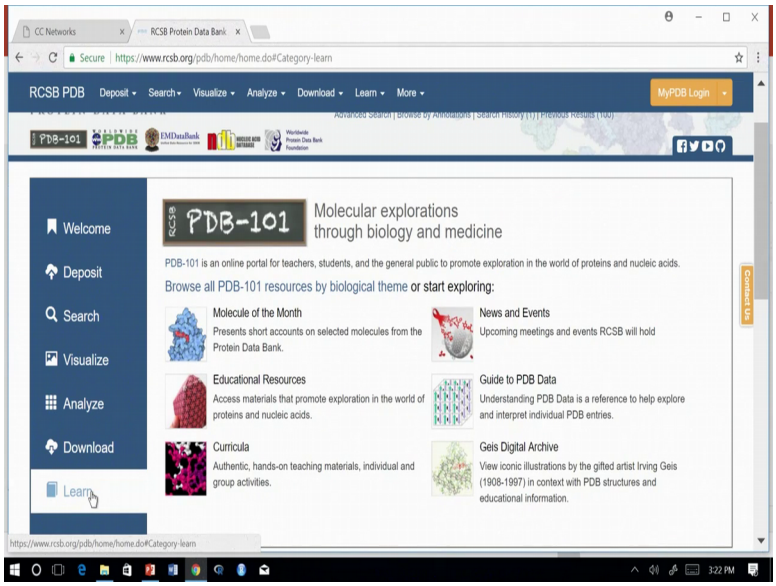
The analyze tab gives in house sequence and structure alignment tools links to the PDB statistics page as well as external tools.

(Refer Slide Time: 02:18)



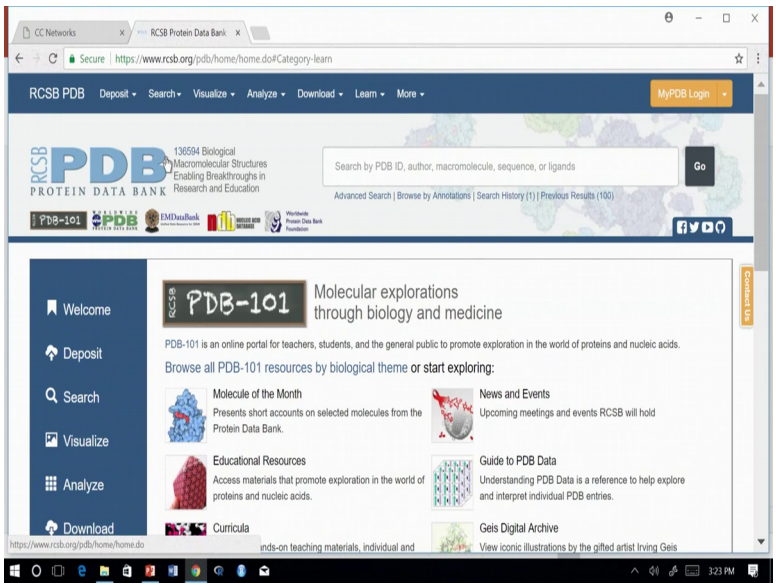
The download tab allow users to have bulk or customize data download and API services.

(Refer Slide Time: 02:27)



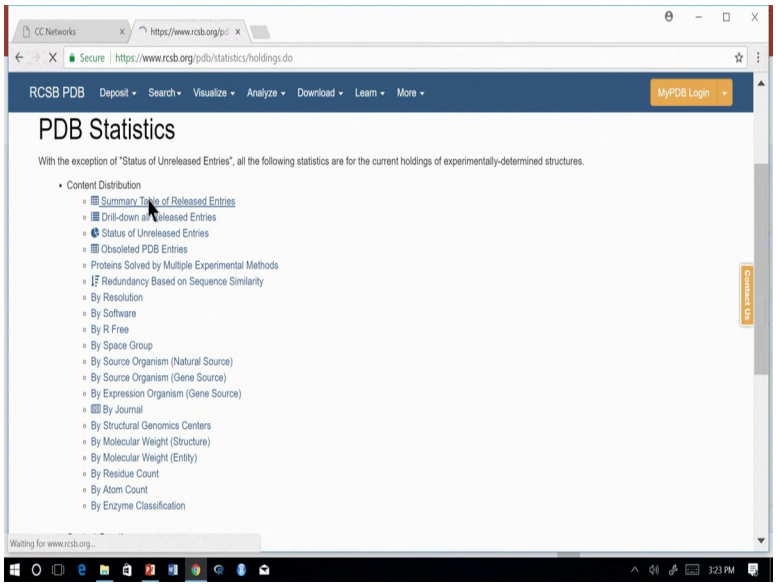
In the learn tab, users can access educational resources about PDB and molecular biology.

(Refer Slide Time: 02:35)



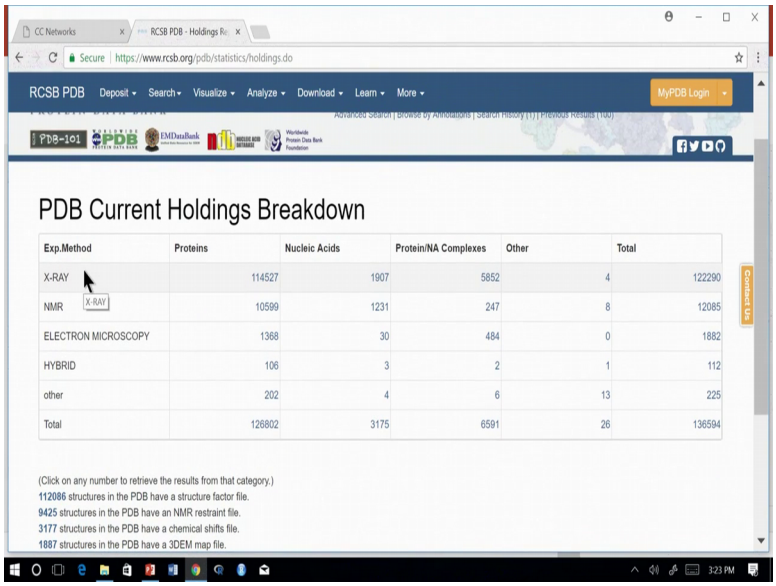
In the header tab here we can see the number of data available in PDB, the number of crystal structures which have been deposited. Now we will see the statistics of PDB data base this can be found here.

(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.

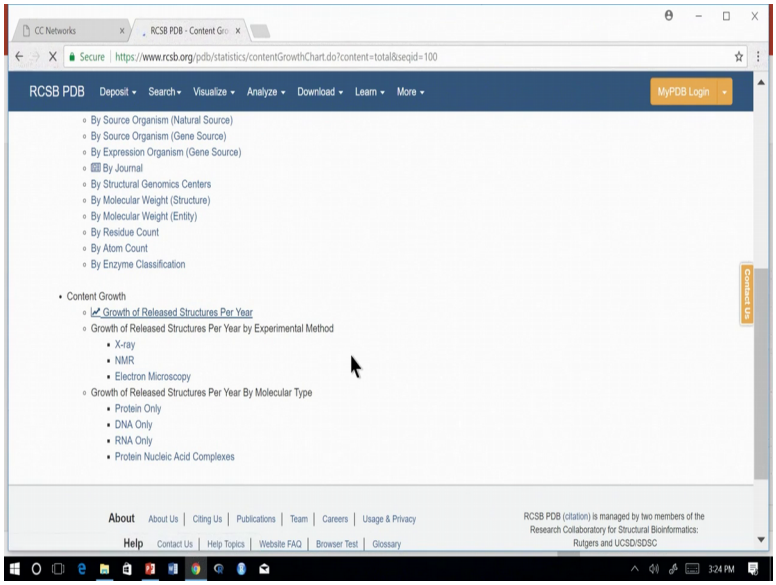
(Refer Slide Time: 03:15)



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

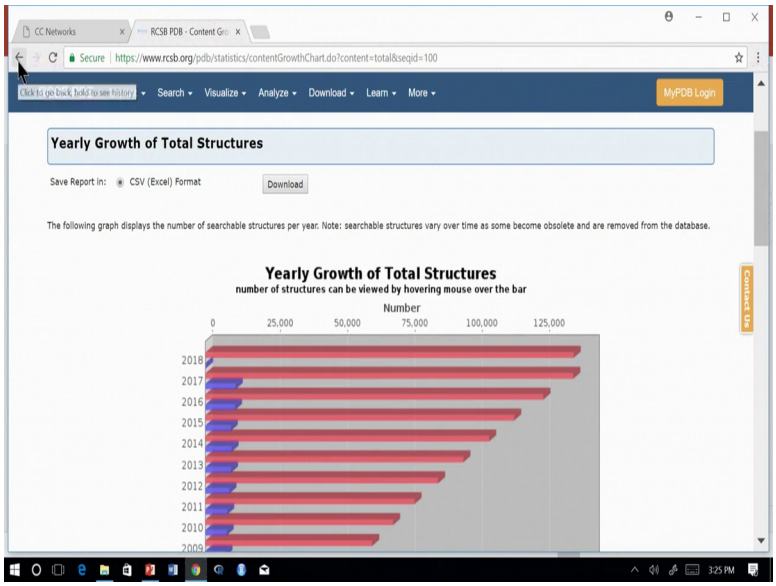
explore further you can see the data by molecular weight by residue count or by the resolution all these data is available for exploration and further study.

(Refer Slide Time: 03:59)



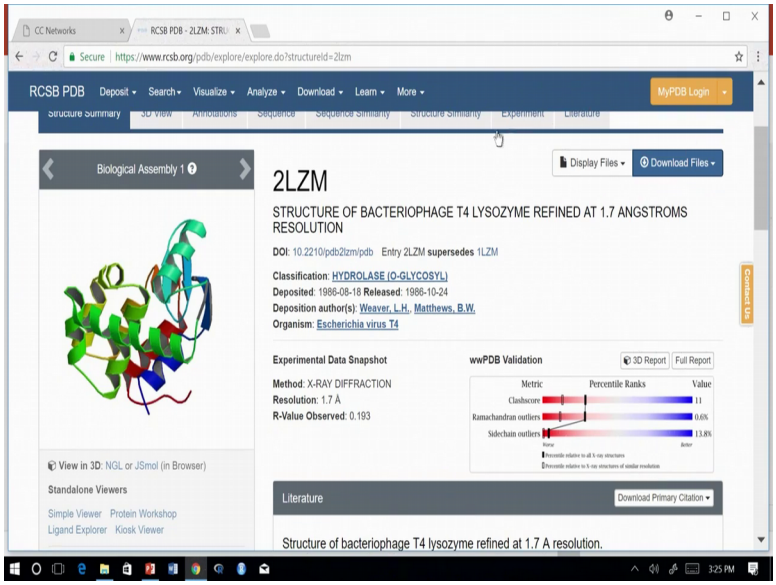
PDB also gives statistics by content growth. This shows how the data base has grown over time statistics have been kept every year since 1976. Let us see the statistics here. The records start from 1976. As you can see the amount of data being deposited in PDB has grown over the years and is currently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

(Refer Slide Time: 04:28)



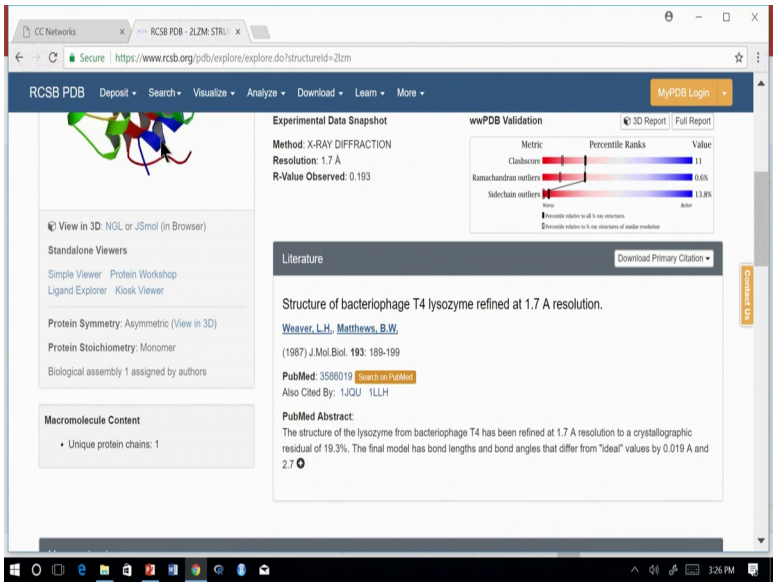
Now, let us let us go for with an example. Now we type 2 LZM in the search bar and press enter.

(Refer Slide Time: 04:51)



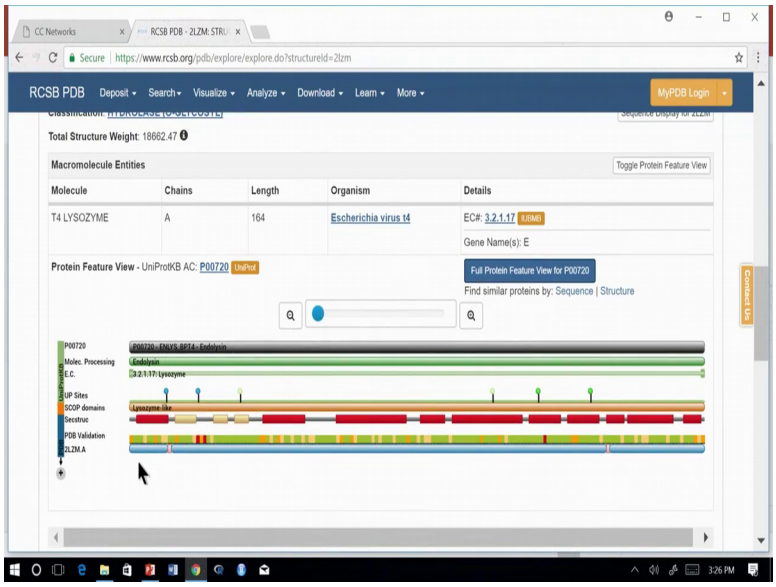
When we type any PDB id, the website gives a summary page view of the structure. So, we have a static view of the structure here the title as well as basic information about the experimental technique as well as protein structure which is given here. So, example this structure is classified as a hydrolase.

(Refer Slide Time: 05:14)



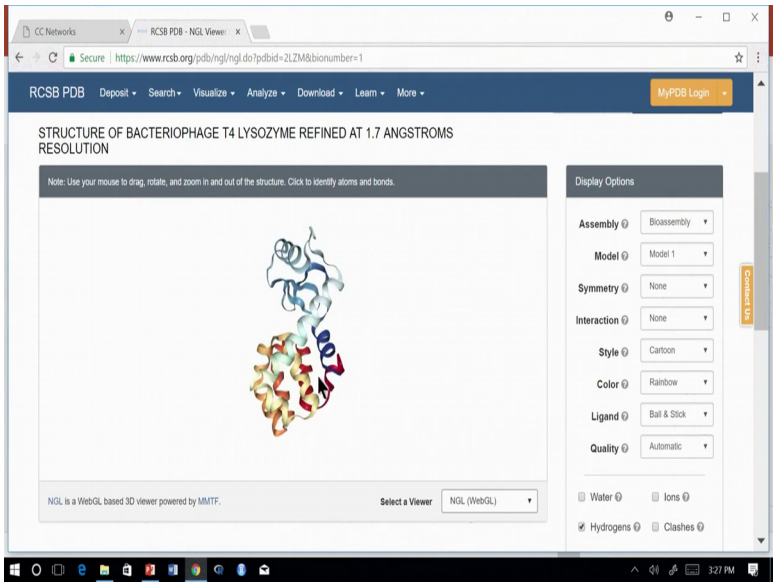
In this experimental method which was used to get the structure is X-Ray diffraction and the resolution is 1.7 angstroms. PDB also provides a literature reference from which the structure was taken.

(Refer Slide Time: 05:29)



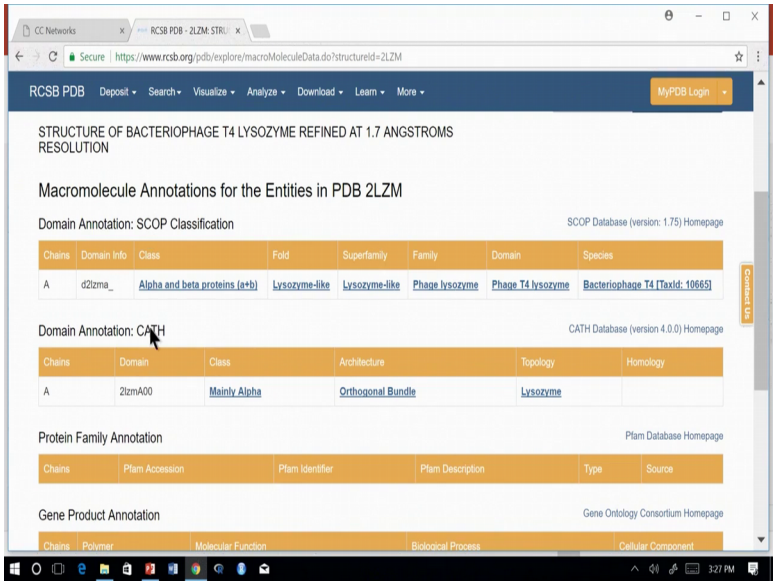
Further you can find the macromolecules in this structure. So, in this structure we have only 1 entity that is the T4 LYSOZYME this is identified as chain A and it has 164 residues. You can also see the feature view here, in this page you can also link to other data bases such as pubmed and uniprot and 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 a10 pi mole.

(Refer Slide Time: 06:17)



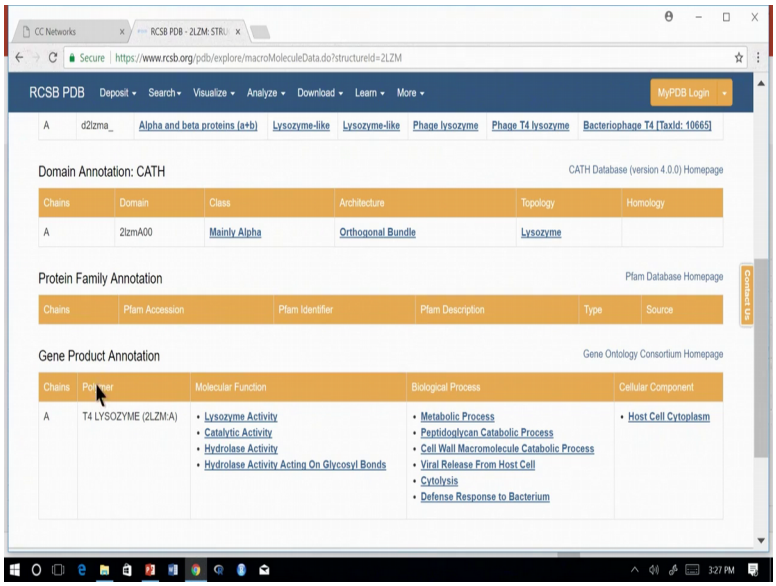
So, in the PDB website itself you can see the structure in the NGL viewer. So, you can drag rotate and see multiple views, we will move on to annotations.

(Refer Slide Time: 06:35)



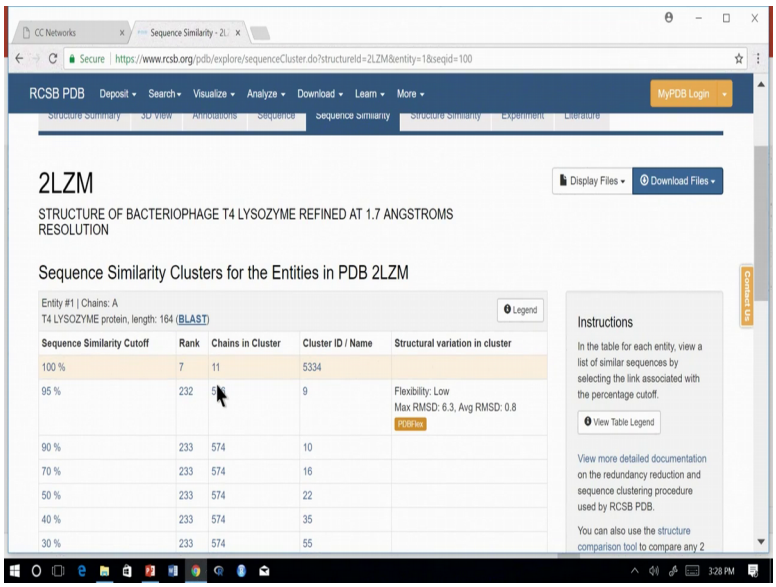
For this entry we have annotations from SCOP CATH P FAM as well as gene product annotations.

(Refer Slide Time: 06:39)



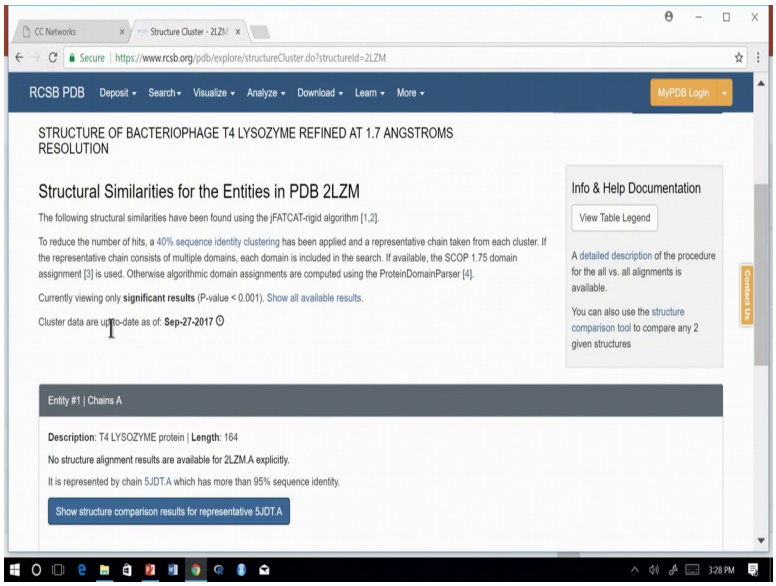
So, SCOP and CATH gave an idea about the protein class, which 2 LZM belongs to and P FAM as well.

(Refer Slide Time: 06:59)



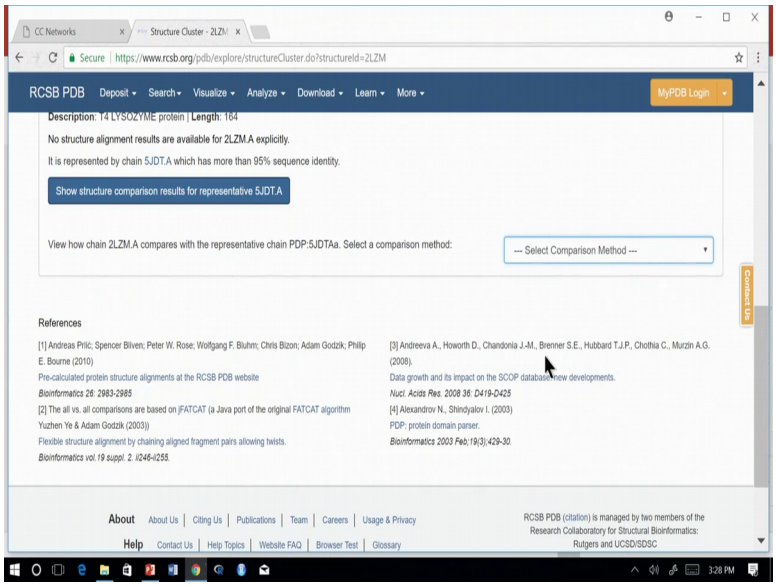
The sequence tab to know the structures or the sequences in that cluster structural similarity.

(Refer Slide Time: 07:13)

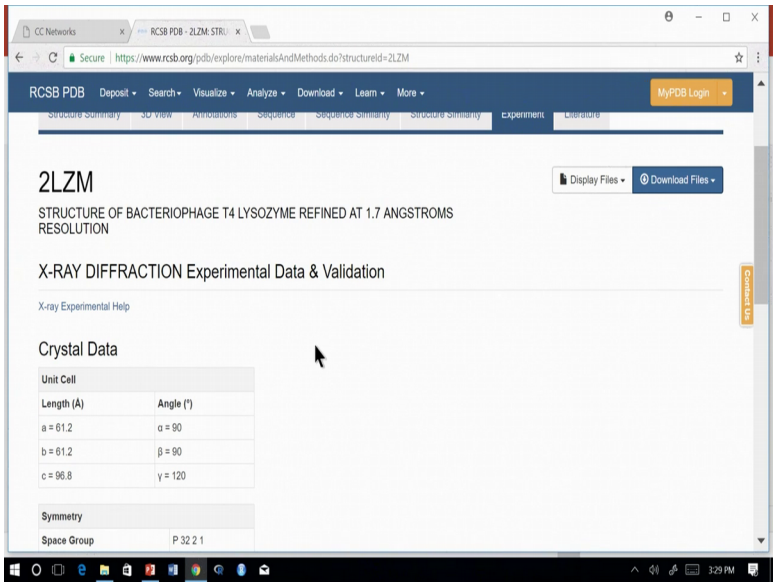


So, here we can see all the structurally similar entities in PDB which match 2 LZM, structure is done you are the structural, similarity is done using the J FAT CAT rigid algorithm, you can also choose other algorithms here.

(Refer Slide Time: 07:29)

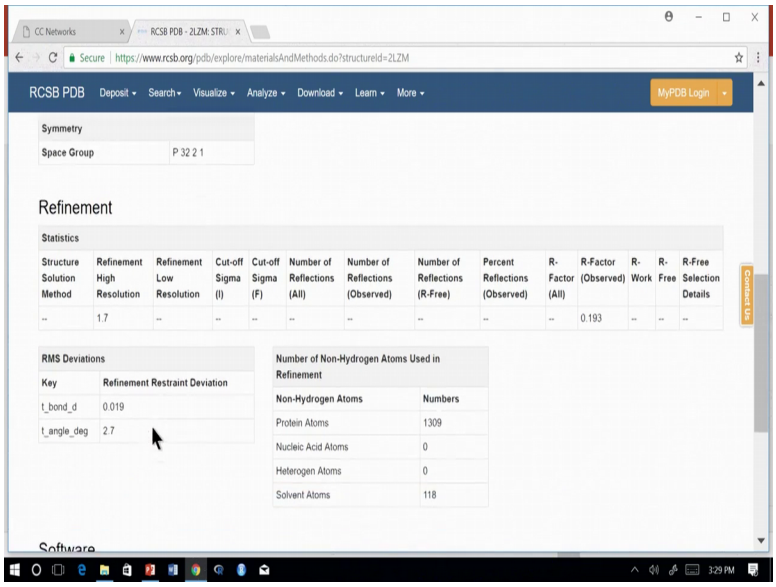


(Refer Slide Time: 07:38)



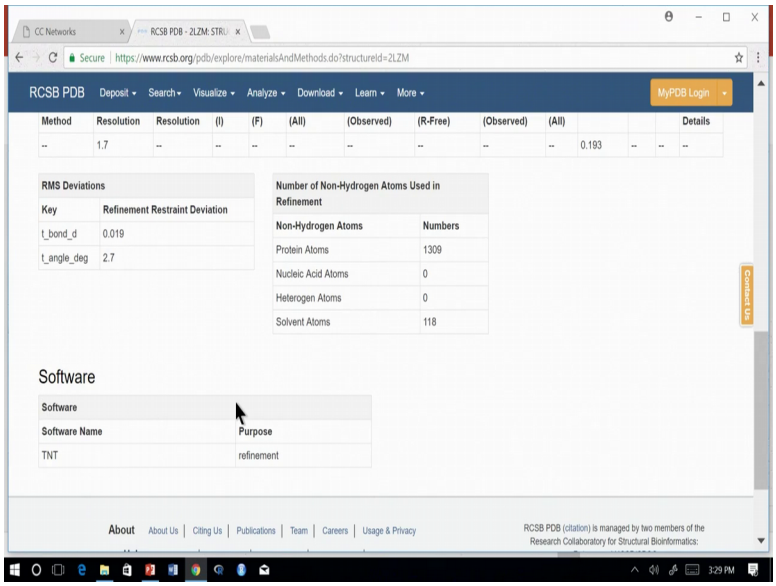
Under the experiment tab you can see more metadata about the experiment, which was used to get the structure. Since, this is an X-Ray diffraction experiment you will find crystal data the space group.

(Refer Slide Time: 07:50)

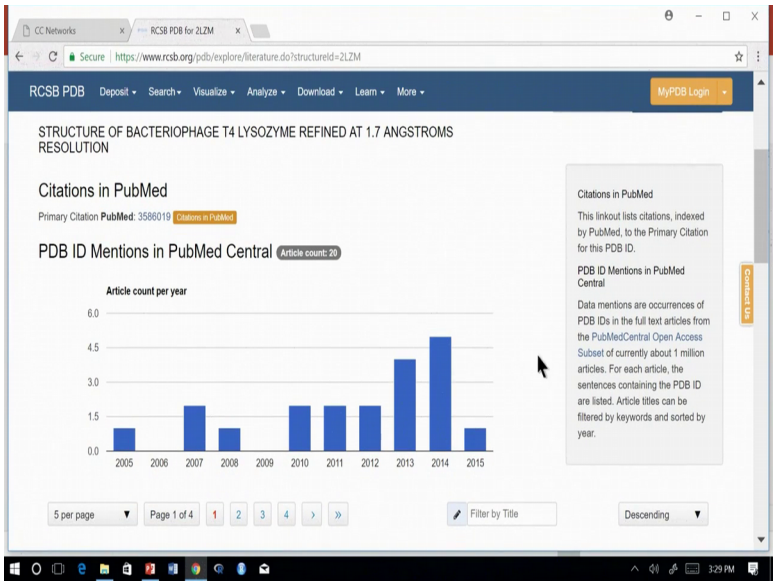


And, as well as others other RMS deviations and software.

(Refer Slide Time: 07:56)

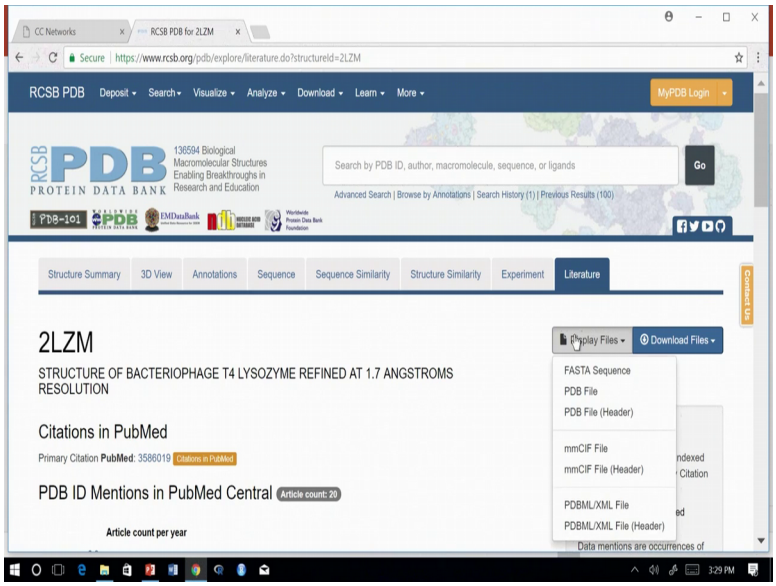


(Refer Slide Time: 08:07)

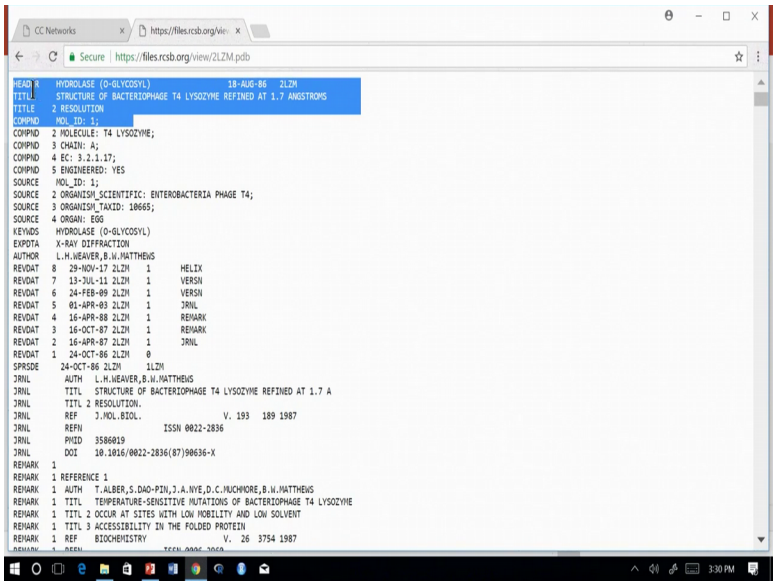


The literature tab lists the citations from pubmed to the primary citation for this PDB id; this graph gives the mentions of this PDB id in pubmed central. Now how to view the PDB file, if you click here this display file you can see the PDB file.

(Refer Slide Time: 08:24)

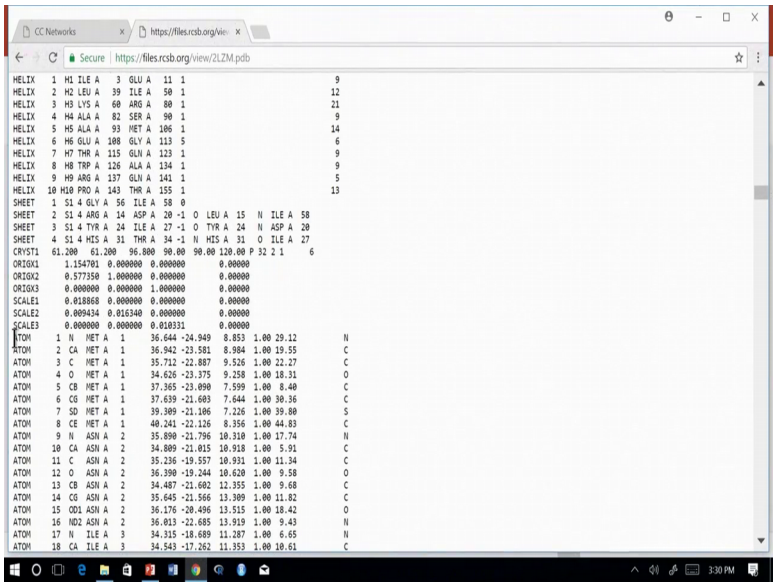


(Refer Slide Time: 08:31)



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)



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.