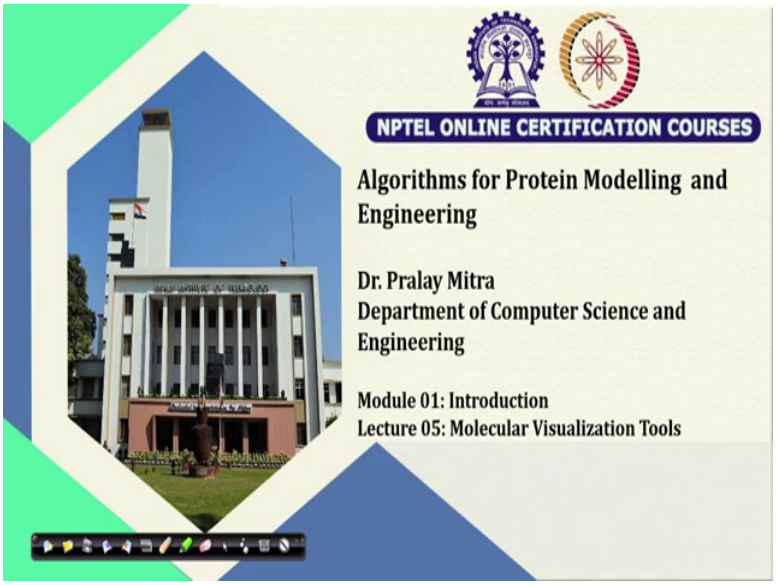
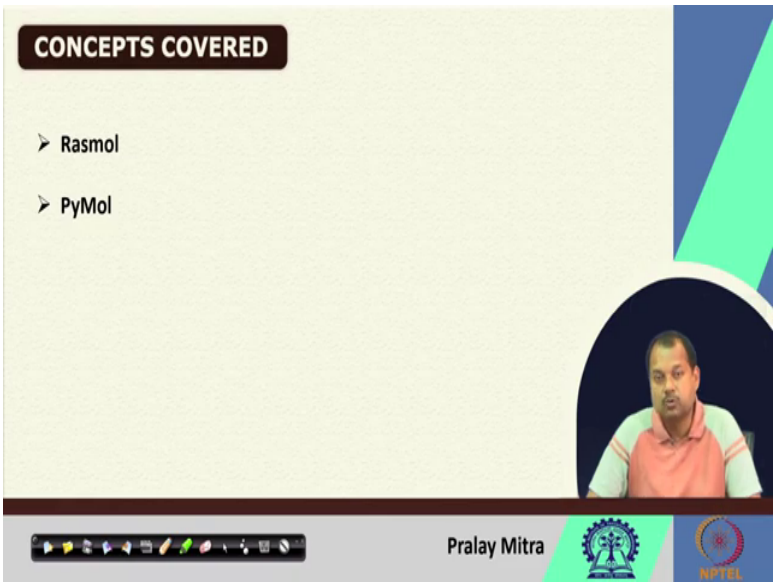
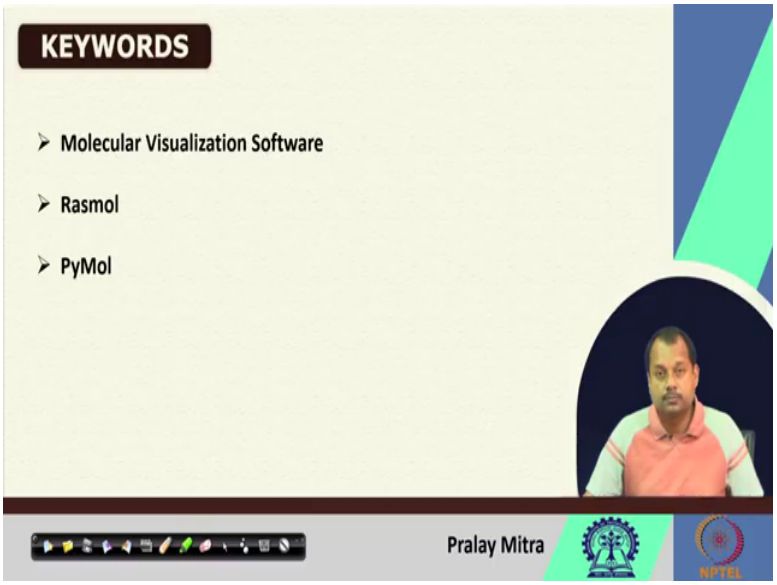
Algorithms for Protein Modelling and Engineering
Professor Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture: 05
Molecular Visualization Tools

(Refer Slide Time: 00:21)



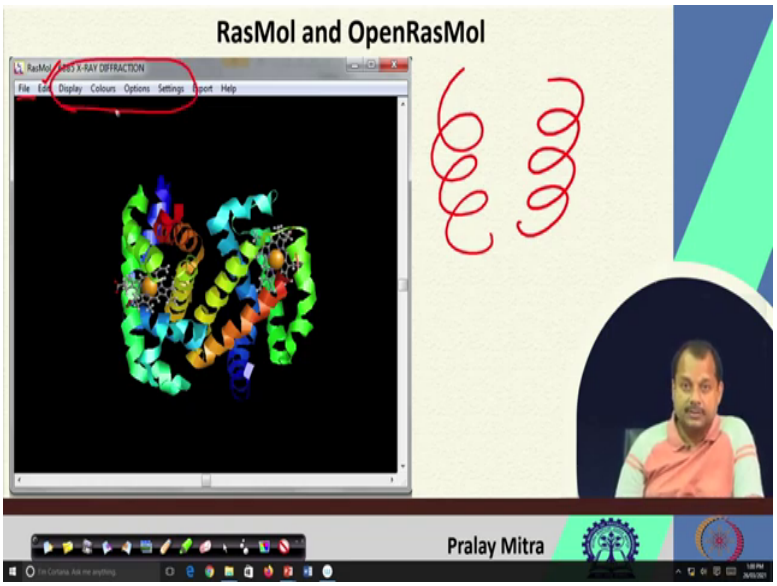




Welcome back. Today in this lecture, we are planning to discuss molecular visualization tools. We shall be focusing on two tools RasMol and PyMol that are most widely used. Specifically, the advantage of RasMol is that it is very lightweight and mostly if your system is not that advanced one, I mean that your system is slow or RAM is not much then also you can run the RasMol.

But PyMol is a very advanced one specifically when you are generating in the molecular image or you wish to do a lot of plots using the molecules, etcetera. And mostly whatever the figures you see in the literature are taken from the PyMol. We shall discuss both, one after another.

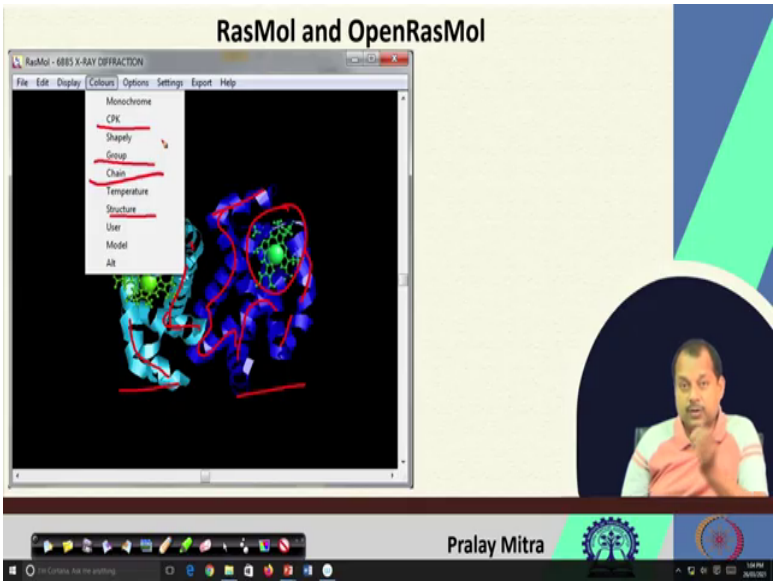
(Refer Slide Time: 01:26)

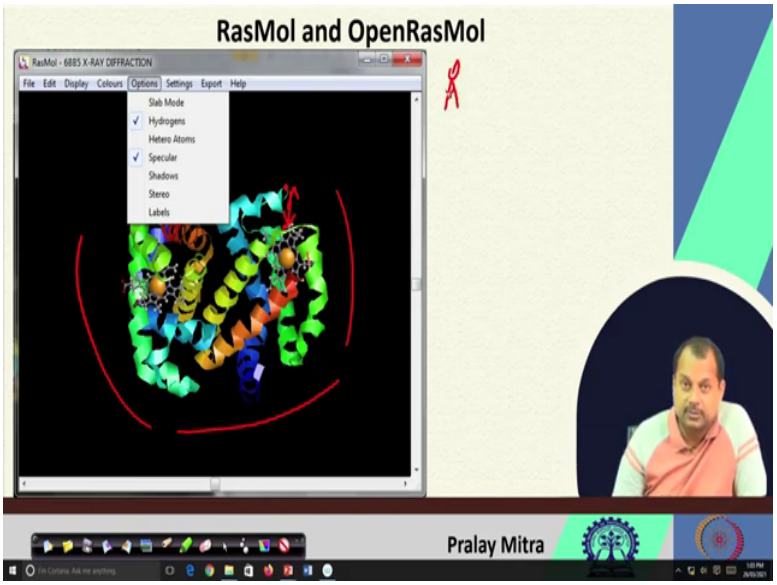


Let us start with RasMol and PyMol both software are free for academic purposes. Thus, you can download that one and install it. The advantage will be that if you download and install it in your system locally then you can look at the protein structure and when we design some algorithm, we model some protein or we engineer some protein then what is the difference between the previous protein. I mean before and after the modeling how it will look like or after the designing, how the structure changes, etcetera. So that it will be easy for you to visualize. Nevertheless, if you are looking at the protein data bank website then using that online software which is available as a plugin in the protein data bank you can visualize the protein structures.

RasMol or PyMol or say OpenRasMol will be useful for you when you download protein structure in your system locally and wish to visualize that one or say you would tweak the structure, you model the structure, you engineer the structure or design the structure that is why we are discussing this one. When we load the same molecule 6BB5, the human Oxy-Hemoglobin, then the initial view looks like this. Here, the coloring is done by the secondary structure. As I mentioned the helix format or say spiral, which is one secondary structure that we call an H or helix. Now based upon that coloring is done. On top, you will see options for opening, closing, and saving files. Some edit options like copy, cut, etcetera is there. Mostly we shall be using display, color option, and some settings.

(Refer Slide Time: 03:45)





When you click on the options, then you will see the slab mode, hydrogens, heteroatoms, specular, shadow, stereo, labels, etcetera. Here, the structure which you opened or uploaded in this RasMol or PyMol for the visualization will be PDB format. There is another support also, but we are discussing them in the context of the PDB format.

When you load the PDB file, then you remember that I mentioned that hydrogen atoms may or may not be present. Now, if it is present and you check it, then all the hydrogen bonds will also be shown to you. However, after checking this hydrogen bond, I cannot see it here. The reason is that at the secondary structure level hydrogen bonds are not visible. When I shall go for all-atom representation, then only I can able to see the hydrogen atoms, then the specular indicates.

If you look at the structure then you will see that there is some reflection that is coming out. That reflection is because I selected it as specular. If you select the shadow then some shadow will appear, the stereo is interesting. Corresponding to one molecule another copy will be created and if you look at the left side copy using your right eye and right side copy using your left eye, then at the crossing point you can able to see their 3-dimensional structure. It is a very tricky thing for which you need to practice or else you can use some spec. Also, there is a specific camera for visualizing this stereo mode.

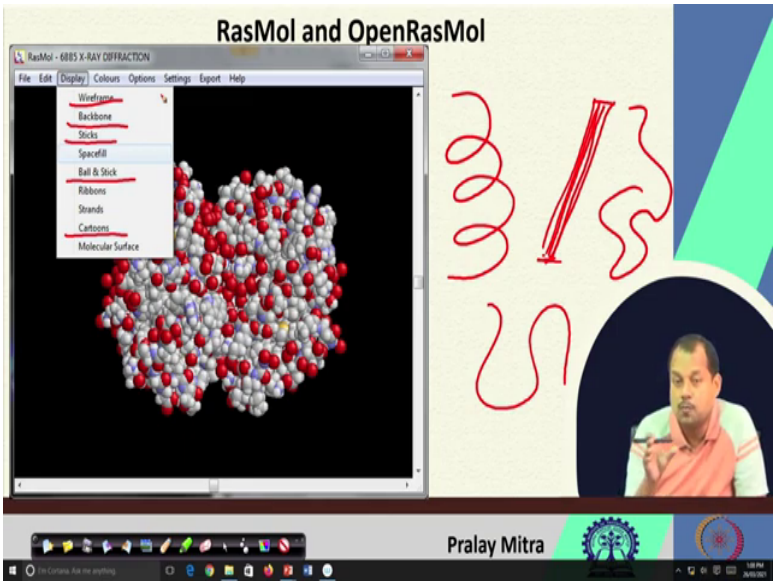
Regarding the settings, I do not have much here, but you can pick the distance. If you select the distance and then click on one atom, and then on another atom, the distance between these

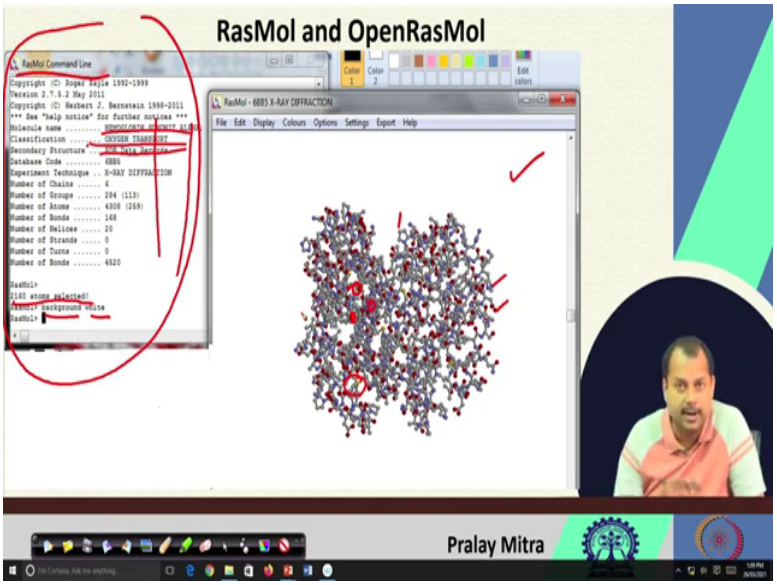
two atoms (in angstrom; 10^{-10} m=1 Å) will be displayed here. Similar to that other settings are also there. I do not have more settings for you, but I have the color for you.

Mostly we go for this chain coloring. You can see that if I select the color by the chain, and as you know by this time that this particular molecule has only two chains (alpha and beta, chain A and chain B), that's why there are two colors. Here is one color, and here is another. Inside this one, it is heteroatom that is not part of the 20 amino acids, but they are present. Thus, they are colored separately. Now, this color, if you follow although it is not very easy, but neither it is difficult that this is one connected component, this blue color, and this light blue is another connected component. Similarly, you have two chains, chain A and chain B.

Two connected components are there. I selected by chain coloring. If I select by structure, then it is a previous one. Helix will be colored. I can select by group then based upon the group it will be colored, I can select my CPK then atom-wise it will be colored. Carbon will get one color, nitrogen another, oxygen another that way it will be colored and if I go for monochrome then only one color will be present.

(Refer Slide Time: 08:01)





Now, if I go to display a lot of options will appear in front of me in RasMol. Among the options, which are there, you see the spacefill (that is also selected here) which displays all atoms as a hard-sphere following van der Waal radii.

The spacefill you see here and on top of this red color is the hydrogen atoms that are present. If you go for ball and stick, (I do not have the visualization) then all the covalent bonds will be represented by one stick along with atoms as balls.

If I use the cartoon (my first picture was the cartoon representation) the helix will be represented by this one and then the sheet will be represented by something like this. And for the coil, there will be no such representation it will be something like this.

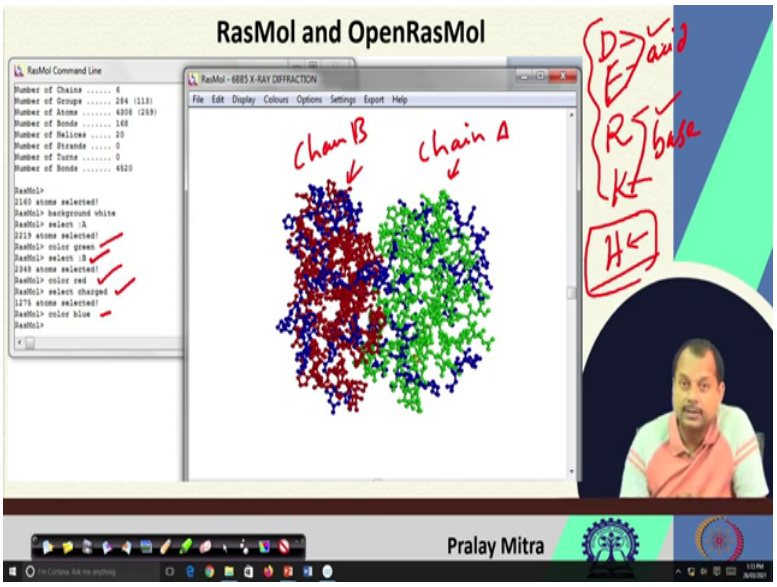
The difference between the coil and the sheet is that in a sheet some thickness like a tip will be there but for the coil, there will be no thickness. If I go for the backbone, then the only connected component will be shown, and that way you will get the fold information. If this is my fold then I will see that one. So, N C-alpha C N C-alpha C and along with that carboxyl C that CO will be attached. But the side chain information, atomic information like after C alpha, that CB, C-gamma, C-delta that information will be excluded from the backbone information. And wireframe is more or less similar to the stick. But for the stick, it has some thickness. You can consider that as a kind of a rod, but when it is a wireframe as the name suggests it is like a wire which is a different representation, and whenever it is required based upon that one you can go for that representation.

This is the menu from where you can operate. The same operations can also be done from the command line. Where will the command line appear? The moment you will open a molecule using the RasMol, two pop-ups will appear - one is this one and another will be this one. Here, you can write RasMol command line and here you will see that after the loading.

This molecule name and classification information all are taken from PDB file if it is present. Else it will not be displayed. That does not make much difference, since it is there, so it is displaying that one.

It will also mention how many atoms are there. What do I do from the command prompt? I write the background white and that is why the black background is now changed to white. This is what I mentioned from the display that you can go for the ball and stick color by the CPK and color by the CPK means this red, blue, and then white. So, red will, and this yellow. Yellow is sulfur and this is carbon, this is nitrogen, and this is oxygen that way it will be presented. I think in the last time when I show you the red color that is the oxygen. Because hydrogen I believe will be represented by white only. You can also say control that using the command prompt.

(Refer Slide Time: 13:03)



Here, what have I done? I select one chain - *select :A*. Then all the atoms for which column number 22 (you remember that is chain ID column) will be selected. After selecting that one I did color green, then those atoms are colored as green, it is easy and simple.

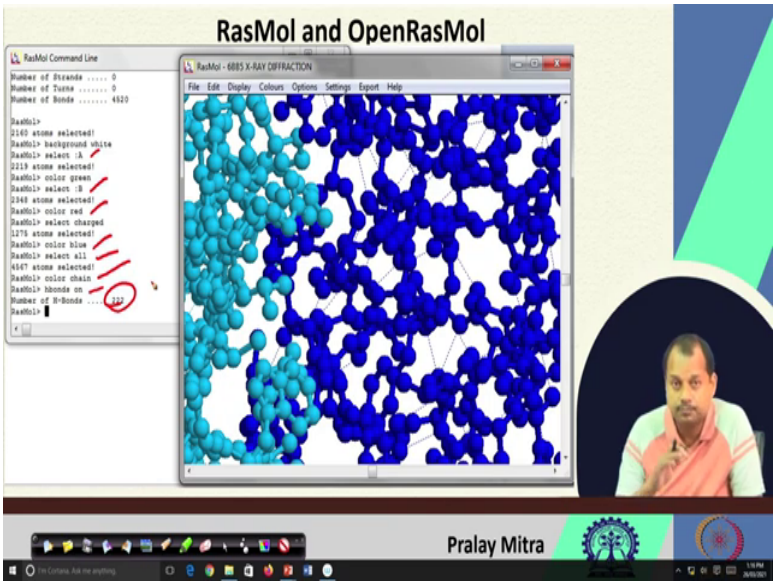
Up to this, I demonstrated select B. After the color green, I mentioned *select :B*. You remember 6BB5 contains two chains, chain ID A, and chain ID B. When I select A, 2219 atoms were selected. There are 2219 rows, which start with ATOM and whose twenty-second column contains chain ID as *A*.

When I *select :B* then 2348 atoms are selected which means 2348 lines are there which starts with ATOM and whose twenty-second column is B or whose twenty-second column contains the character B. After that, I said the *color red.* So, you can see this red. Do not look at the blue one for that another command is there. I shall show you that later. This red is chain B, and this green is chain A. After that, I use *select charge*. Charge means among the 20 amino acids, acid, and basic amino acids like aspartic acid, glutamic acid, arginine, and lysine (sometimes histidine is also considered as charge residues).

This as the name also suggests is acid, this is base. This is more or less neutral sometimes it shows little charge nature based upon the presence/absence of one oxygen atom that is not relevant right now. Still, you remember this. These are the charged residues. If I go for *select charge*, then all those residues which are either D or E or R or K will be selected and then when I say *color blue*, the charged residues will be colored as blue.

While doing this *select charge*, I did not mention the select charge of chain A. That is why all charge residues inside this molecule which are either part of A chain or part of B chain have been selected. When I say *color blue* they are colored here and that is the justification for why partly it is green, partly red, and partly blue. First I select chain A, color green, select chain B, the color red, select charge for chain A and B, and color blue that is why is that coloring scheme.

(Refer Slide Time: 17:00)

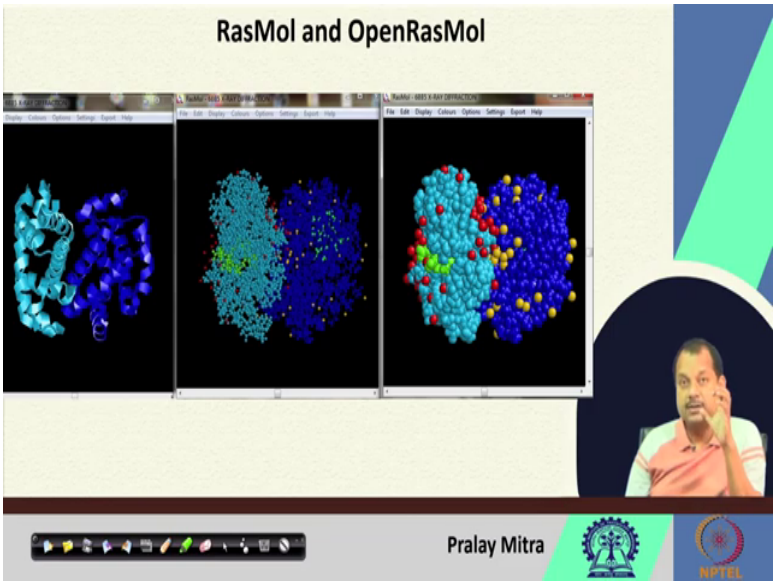


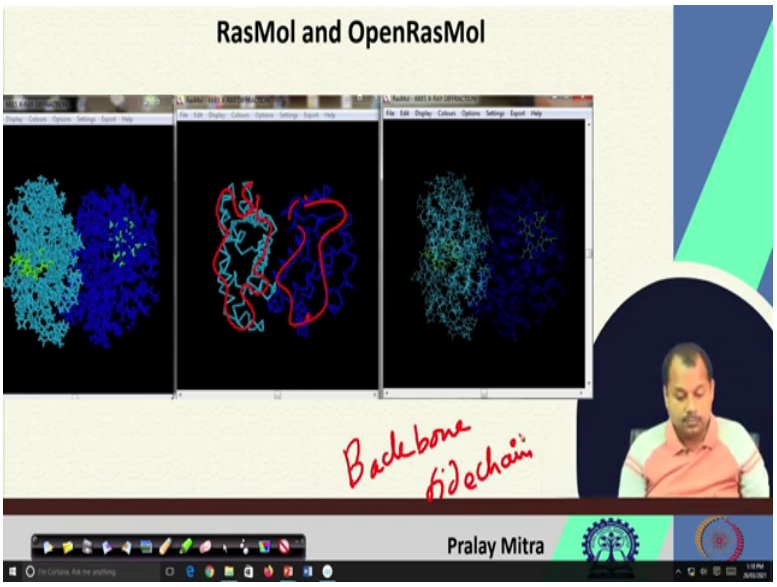
Similary, there are a lot of command-line options. One such option is *hbonds on*, which means if there is any hydrogen bond(s) then show those. I can select A, select B then color red, the color blue, and then I go for *select all* to select everything. Then I said color chain-wise so that one chain gets blue color, another gets light blue color, and finally, *hbonds on*.

If there is a provision for forming the H-bond following the specific criteria - there should be a donor, there should be an acceptor, and there should be one hydrogen. In case of the absence of hydrogen atom in the data, but should be there as per molecular structure, then you can assume the existence of hydrogen bond but the donor and the acceptor must be there. Also, there should be a specific geometry between the donor and acceptor, and between the acceptor and the atom with which the acceptor is forming the covalent bond. If all of the above-mentioned information and geometry of the atoms hold then it will form one hydrogen bond (dotted line) will be given. It will also tell you how many hydrogen bonds are formed.

Now, these number of hydrogen bonds are from the complete molecule which means, intra-chain as well as inter-chain. For any reason, if you want an inter-chain then it is simple. You select only one change switch on hydrogen bond, another chain, switch on hydrogen bond, count their number, select all the atoms, switch on hydrogen bond, take the subtraction to know the inter-chain hydrogen bonds. Or you can write a separate program when you know what is the geometric property of the hydrogen bond.

(Refer Slide Time: 19:22)





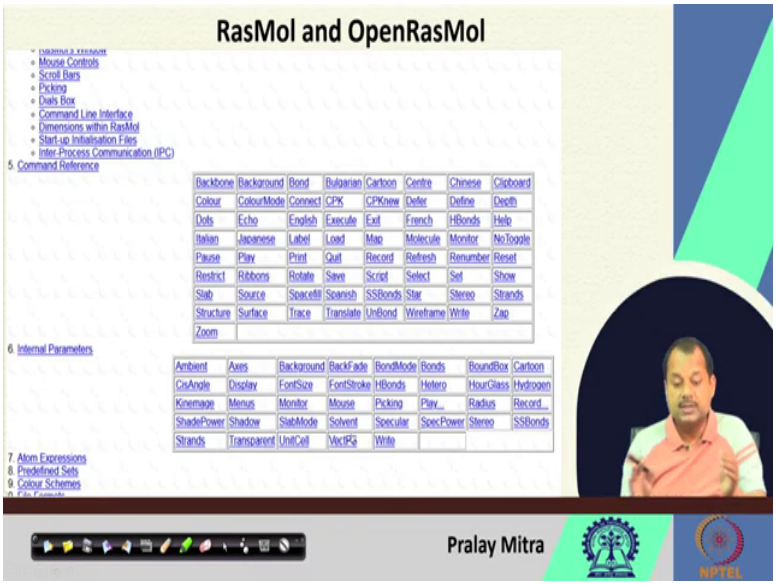
These are some of the visualizations which are available using the RasMol. So, on the left-hand side, you are getting the secondary structure representation color by the chain. The second middle one is the ball-and-stick model where the atoms are represented as a ball and covalent bonds are represented as sticks. Again, they are colored by the chain. Then some heteroatoms are there they are colored separately.

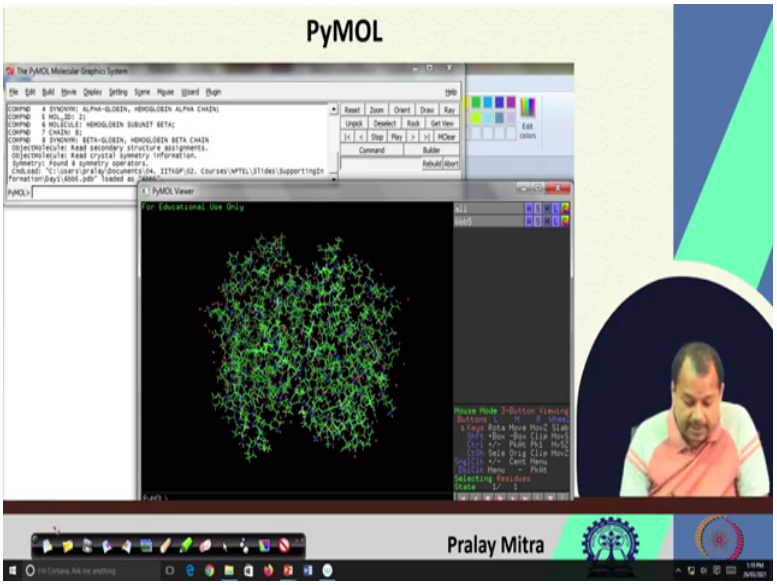
Then the third one is the spacefill, where you can see only the atoms as spheres. On the left-hand side is the stick model, where the atoms and the covalent bonds are connected by a stick. On the rightmost side is the wireframe. Between the leftmost and rightmost, you will see the difference is in thickness, and one is of rod shape and another one is wire shape.

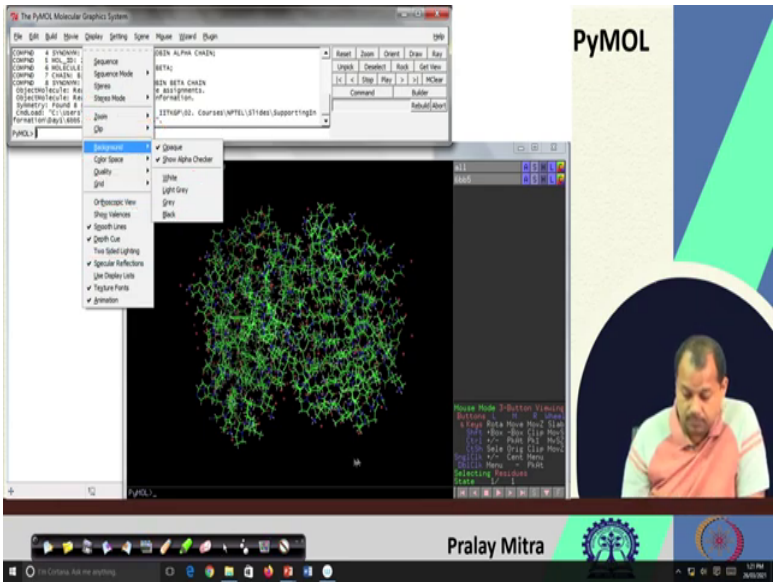
In the middle is the backbone representation. As I mentioned that in backbone representation side chain information (R1, R2 part) or in the atomic coordinate level C-beta onwards will be chopped off. The rest of the part is there demonstrating only one connected part - the sky-blue if you track. Starting from here it is going, going, going, going, going, going, going, going, going and then coming here, and if you track it, I can see one loose end here, another loose end. I cannot able to see it - as it is occluded now, but it will be somewhere.

Likewise, this is one connected component and this is another connected component, two connected components, and only the backbone information. This means N, C-alpha, C that way the backbone information is present. This backbone and side-chain are very useful for testing the goodness of our model after the design or after the engineering or after the modeling, then we shall compare the backbone-wise or all atom-wise before the modeling, or before the change and after the change. Hence, these representations are very useful. I got this list of commands from the help page of the RasMol/OpenRasMol that you can make use of.

(Refer Slide Time: 21:54)







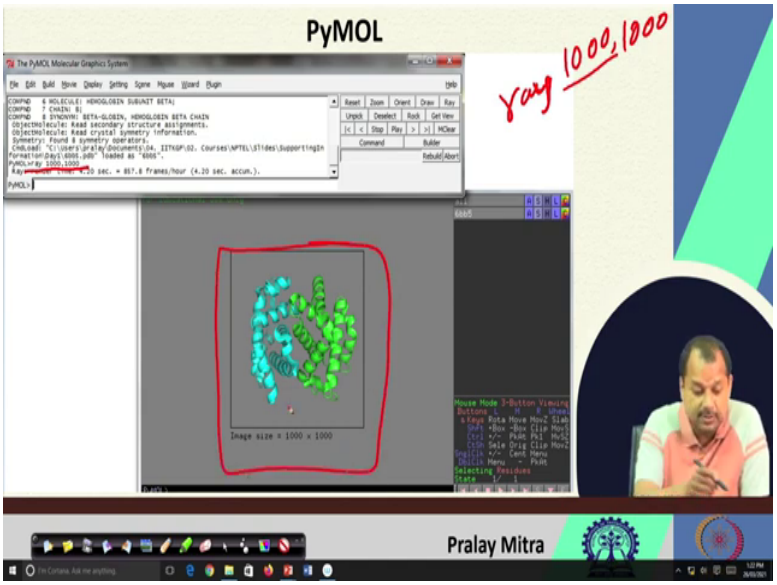
In PyMol the features are enhanced. You can consider PyMol as a feature-heavy one compared to RasMol. Once you load the protein structure, you will see something like this. Here there is an option under File for saving an image, saving a session, or saving one particular molecule after editing the molecule. Next, if you go to display, you will see the background for selecting your choice of background color - white, light gray, gray, or black.

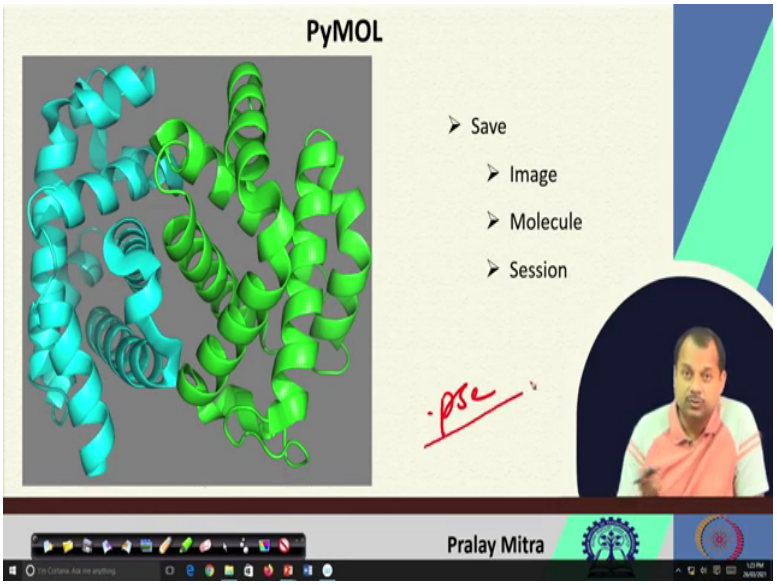
There is a provision for two-sided lighting that will make your figure spectacular. Then definitely you go for this depth cue instead of orthoscopic view unless otherwise required. At color space, you have the option for CMYK for publication purposes, RGB for display purposes, or PyMol has its own color space, you can use that one. Quality, you can go for say maximum performance or maximum quality.

Similar to the RasMol you will also get stereo mode. At the zoom option, you can have the complete molecule, or 8Å zoom, or 10Å zoom. Those are part of the display.

After the display and setting two-sided lighting, shadow, etcetera, if you wish to take the photograph then in PyMol you can go for one command that is called as the ray, so r a y, followed by you mention the resolution. I have used 1000 comma 1000 but you can have 2000 comma 2000. Be careful since it is very computing-intensive. If you go higher then your system will slow down. Thus, use some reasonable value like from 1000 to 2000. After the rendering, your image will look like this. If I stretch it then also the quality of the image will not degrade. Now, you save and compare it with the RasMol image.

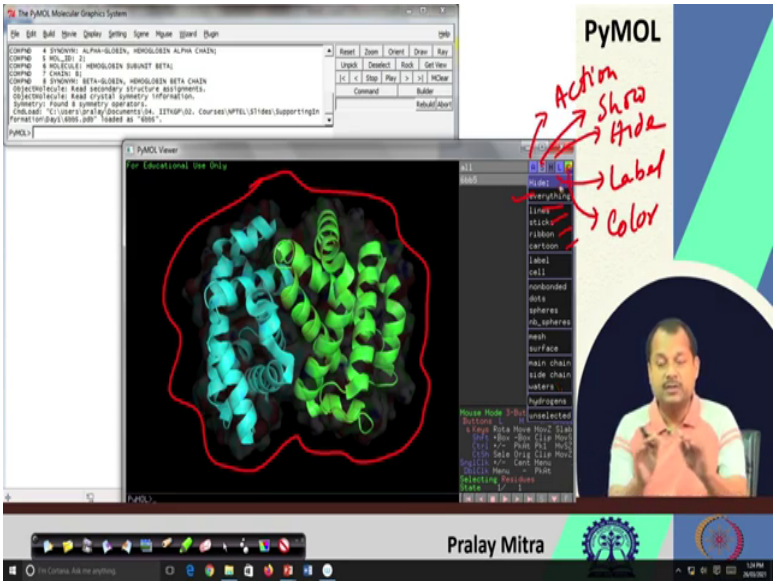
(Refer Slide Time: 24:14)

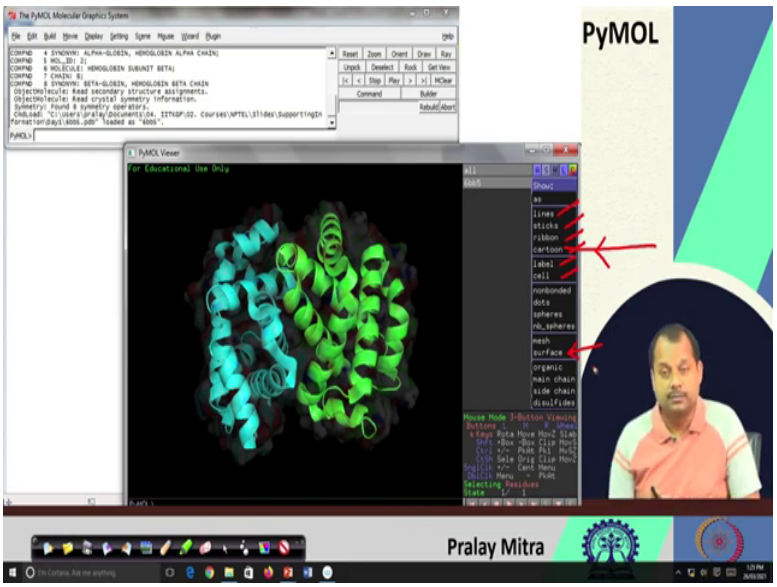


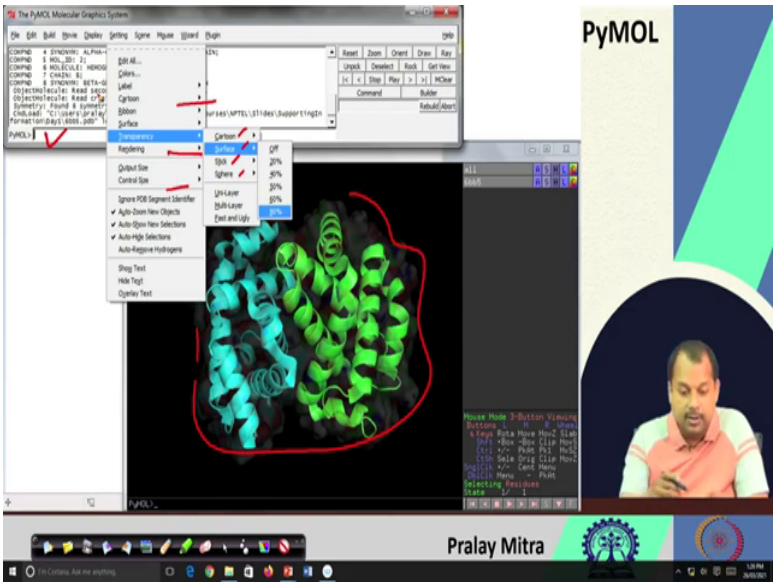


So, it is always advisable if you wish to put your figure for publication, it is better to go by PyMol. RasMol is also good. PyMol has a lot of features, you can save the image, save the molecule, or save the session. After your work, you can save the session before exiting PyMol. The session will be saved with the extension dot PSE. Next time you open dot PSE then the previous save state will be retained. Undoubtedly, this is advantageous.

(Refer Slide Time: 26:24)







In this figure, I can easily see one contour. That contour is for this protein molecule. But, this representation is the cartoon representation that I mentioned to you previously. Now, these options are also available on the right-hand side here. A stands for Action, S stands for Show, H stands for Hide, L stands for Label, and C for Color. Under each several options are there.

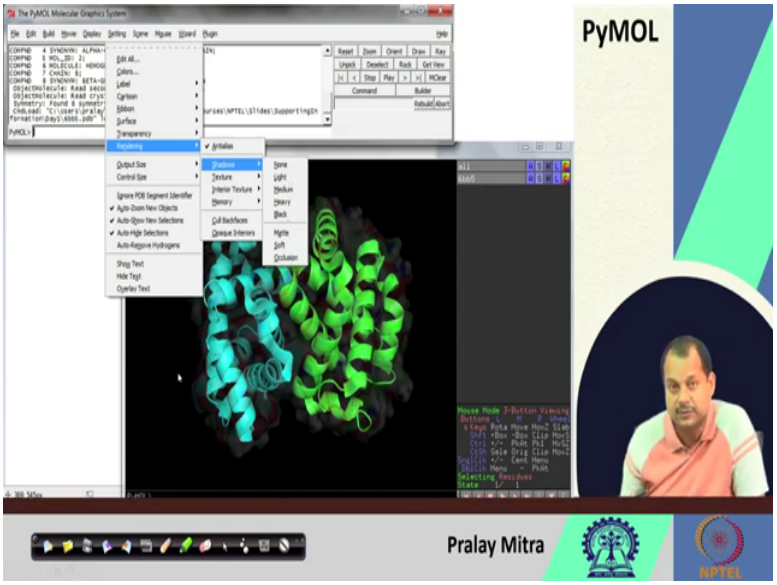
The best is at the beginning you hide everything. Selectively if you wish to hide the line, stick, ribbon, cartoon then you can go by that one. But if you hide everything, everything will be blank. Next, you can start your representation. If you go to the show, then you will also get the option of Line, Sticks, Ribbon, Cartoon, Label, Cell, etcetera.

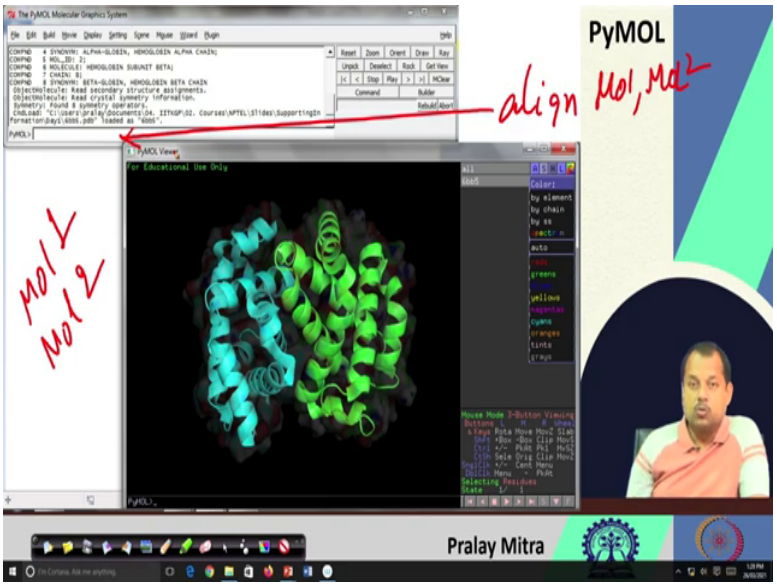
Here, along with a cartoon I also selected the surface that is why I am getting two different representations in a combined way. This kind of combination you can do here and after that,

if you wish to save this information or save this image then from setting you to choose transparency. I set transparency at 80%. Else, the surface image will occlude cartoon image. Transparency is important when you are having multiple representations like having cartoon and surface or stick and cartoon. Rendering you can do, you can control the output size, etcetera.

Here, you can control the size of the cartoon, its looks. If you wish to give some bond between two atoms you can do that from here and from the command line also by a pro or advanced user. In this user interface, a lot of facilities are provided you can make use of.

(Refer Slide Time: 29:39)





In rendering, you can see that shadows are there. You can give light, medium, heavy shadows. In the black background, the shadow will not be a very good option. But if you have a light background you will see that the quality of the image will also improve with shadow. Also, I wish to mention that the label option is here so that you can see it. You can also mention color by the element, by a chain, by secondary structure.

There is an option in PyMol to align two structures using command *align*. If you have two molecules viz., mol1 and mol2, then I can write, *align mol1, mol2* in this command prompt for aligning them. Also, their alignment score will be displaced in angstrom. This feature is specifically important before and after modeling, or after designing or engineering molecule to check what is the deviation or how similar it is structure-wise or fold-wise before the modeling and after the modeling. In summary, this is about an introduction of the protein molecule, the protein databases, and the viualization software - RasMol, and PyMol.

From the next week, we shall start developing the algorithm for modeling the protein molecules. And in the context whenever it will be required, we shall elaborate more on the use of the PyMol or RasMol or provide you more detail about the PDB structure parsing or defining the data structure for storing the sequence and the structure, etcetera. That is it.

(Refer Slide Time: 31:58)

These are the two web-links from where you can download RasMol and the PyMol. And as I mentioned, they are free for academic purposes. Thank you very much.