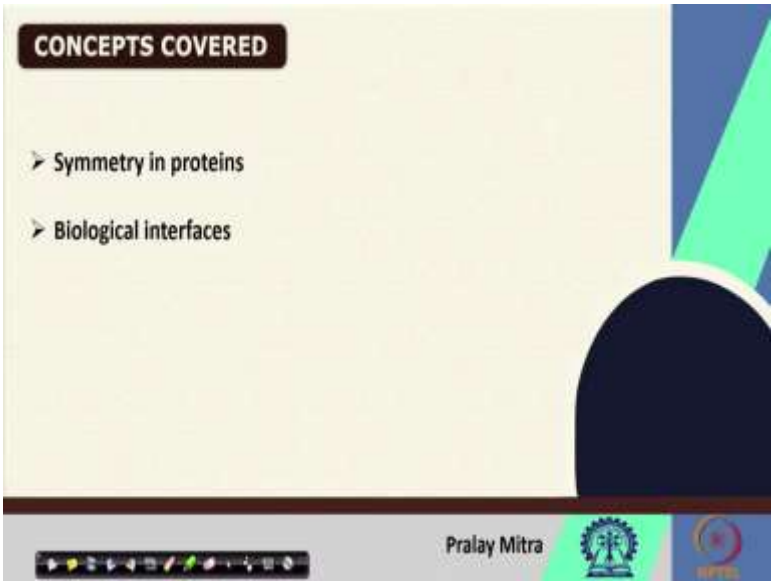
Algorithms for Protein Modelling and Engineering
Professor Doctor Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture 33
Symmetry in Proteins

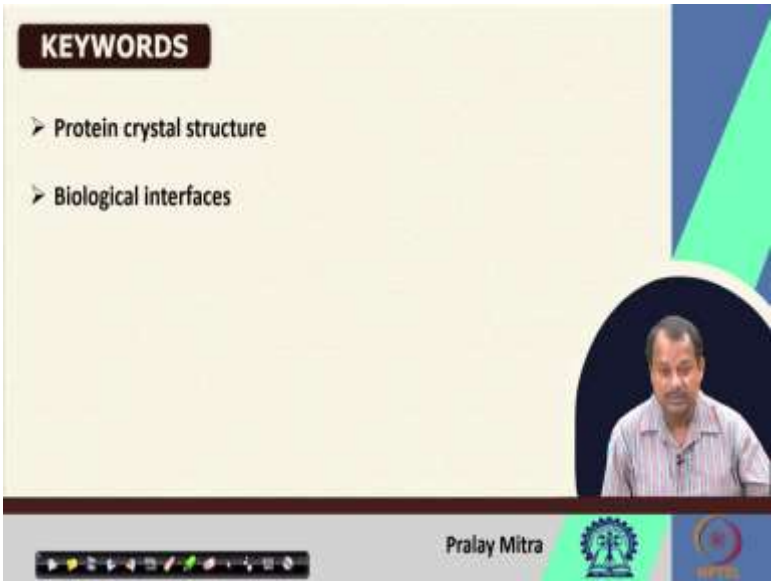
Welcome back. So, we are going to cover one interesting topic in this particular lecture that is symmetry in protein. That is interesting in the sense that when it is not one single protein but protein is interacting with, one chain is basically interacting with another chain that is called as a dimer, when three chains are in interaction then that is called trimer, then only it has some function. Otherwise it will not have some function.

Now, at the initial classes we have discussed the protein docking problem where given two protein structure I wish to determine their association. So, when only two such protein molecules are given and I wish to identify their association then, so there are several possibilities, and that is true. Now, in among those two possibilities, among those possibilities, so when it is two then there are several possibilities. So, when it is three then sometimes not all possibilities exist.

So, it has been observed long back that protein enjoys some sort of symmetry during their function. So, it is not enough only to look at the association or when, say we are designing some protein docking program in order to identify the association of more than one protein molecules then symmetry may play a crucial role. Why? That we will discuss and what kind of symmetry, what kind of symmetry exists in the protein structure? That we will discuss.

(Refer Slide Time: 01:55)





So, the concept we are going to cover today is the symmetry in proteins and the biological interfaces because there may be a number of interfaces which are not biological in nature but they exist as an artifact in the crystal structure; because you know that when it is the crystal structure then what happens?

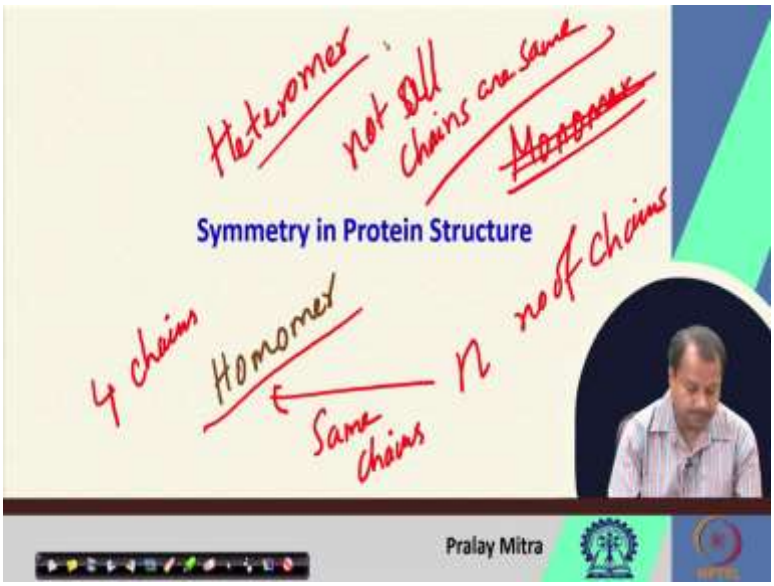
That the protein takes a crystal shape, so if it is in crystal then only the interface during the crystallization will be identified. Now, if it is a crystal then you also understand that, say if it is dimer, the simplest one then it will have in a lattice form. So, this is one occurrence, after that

this is another, this is another, this is another, this is another. That way it will be a long one, the lattice one.

Now, if it is, then this is one interface between this protein, this protein in a dimer. This is another interface between this protein and this protein in order to make it lattice. So, which one is correct, this one or this one? Or which one, by correct I wish to mean which one is functional, this one or this one? That we need to understand.

Also when it is in the crystal form then I may not guarantee that the functional form which will be there in the, say solution like water, inside the water, so it will be the same interface which will be retained. So, considering all those things so lot of variation will exist, which is the correct one from symmetry point of view? Let us analyze that one. So, before I proceed and, so keywords are also selected like this, protein crystal structure, biological symmetry.

(Refer Slide Time: 03:32)



So, before I proceed let me define you two things, homomer, another one is heteromer. So, first of all when there is only one chain or one connected component of a protein then that sometimes is called as the monomer. And monomer is not the functional form. So, in the context of the monomer, homomer and heteromer does not exist.

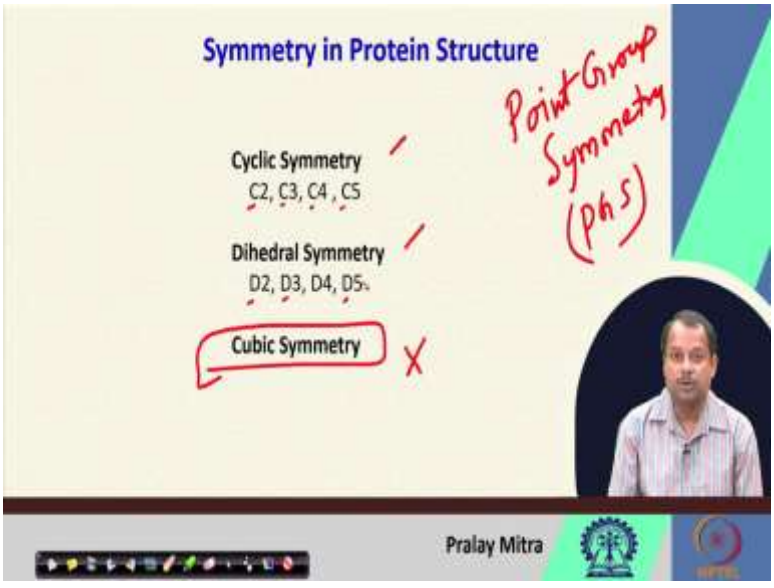
But when I am going more, means more than one chains are there. So, when there is a protein structure which consists of, say n number of chains, n number of chains then there are two

possibilities. All the n chains are same or they are not same. If they are same then that is called as the homomer, same chains. And heteromer integrates not all chains are same.

So, from this point of view we are defining two terms, one is homomer which consist of similar protein sequences or same chains or it is homomer where not all chains are same. But in this concept, in this context please note it down that it may possible, say there are, say there are 4 chains. Now, if you go by sequence alignment or, say if you try to do a string matching then they will not match, because it may possible that in one sequence few residues are missing. So, we discussed about this missing co-ordinate information etc at the initial lectures. So, you remember that one.

So, that way it may not be the same but that does not mean that it is a heteromer. That is a homomer. So, you have to check or you have to understand whether it is the missing co-ordinate or missing residues for which I am getting, say mismatch or it is indeed two different sequences. So, based upon that you have to understand. But whether it is a homomer or heteromer it is important, specifically when we are discussing the symmetry, because symmetry will only, we can discuss in the context of the homomer. For the heteromer it is not possible. That is trivial, I believe to you.

(Refer Slide Time: 06:18)

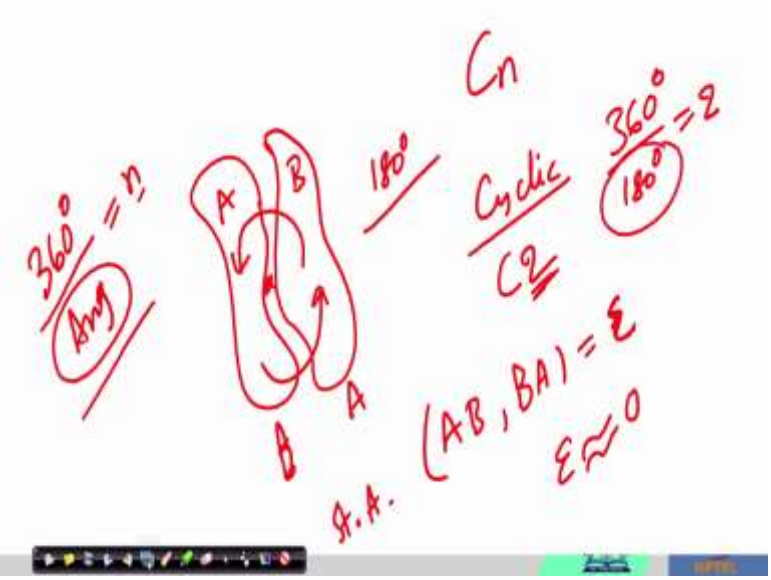


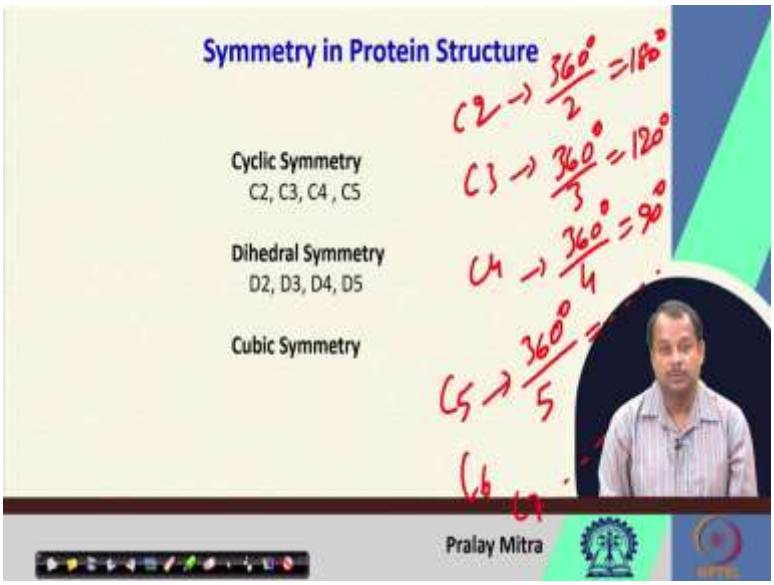
Usually in protein the symmetry we will be discussing that is called as the point group symmetry. There are three different kind of point group symmetries that we will be considering for protein.

So, one is cyclic symmetry. Another is dihedral symmetry and third one is the cubic symmetry. But cubic symmetry usually occurs for proteins whose dimension, whose number of chains are more than 10, so 12 or onwards. So, that is why we will not consider this cubic symmetry for our consideration. We will focus mostly on these two, dihedral symmetry and cyclic symmetry.

Now, in case of cyclic symmetry, so the notation it uses is C and in case of dihedral symmetry the notation it uses is D. So, C and D, now after C or D there are numbers. So, what is the meaning of that number? I will come to that. So, cyclic symmetry and dihedral symmetry is of our interest.

(Refer Slide Time: 07:42)





Now, let us start with one dimeric structure. So, if the dimeric structure looks like this, again symmetry means it will be homomer. Although my drawing is not very much identical, for that I will show you the protein structure but from here you can see that if, say kind of there is one axis somewhere here which is passing a perpendicular to this plane and about that one if I rotate, how much, say 180 degrees then this will go here and this will go here. So, if I say A and this B then after the rotation it will be say, B and it will be A.

Now, before the rotation A B and after the rotation B A if I perform some sort of structural alignment, structural alignment and then compute what is the deviation? Say if that is epsilon and if epsilon is very close to 0 then it is not possible for anybody after 180 degree rotation to identify whether they are the new transformed structure or the previous one.

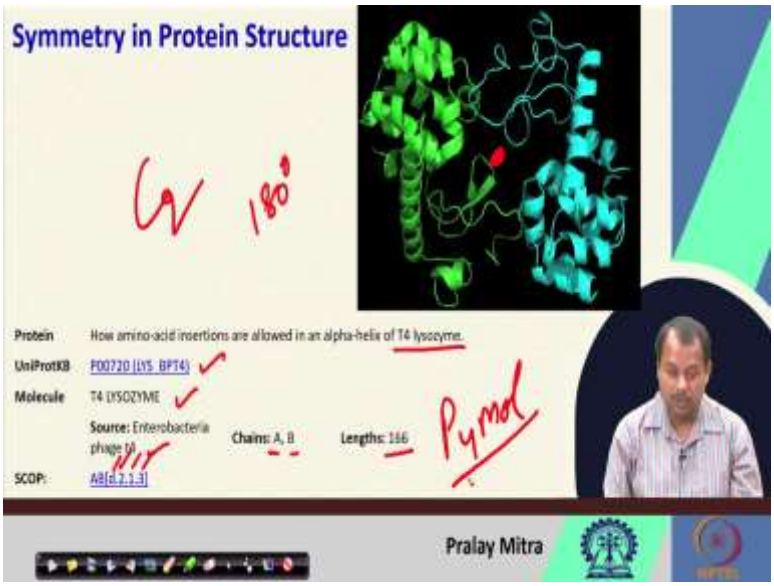
So, then we will call that there exists one cyclic symmetry with respect to angle 180 degree. Now, if I divide this 360 by this 180 we will get 2. So, the symmetry which exist here is called as the C2. So, this n, so in general for cyclic symmetry if I write Cn so n indicates that if I divide 360 degree by some angle then I will n. And what this Ang angle indicates? If there exists one symmetric axis about which if I rotate my total molecule by this Ang angle then after the rotation and before the rotation, the structural deviation epsilon is close to 0 or there is such no structural deviation.

There might be some minor structural deviation as you understand because, say in the side chain or, say in some co-ordinate information there may be some little mismatch because of that one

you may not get 0. But that is why I am telling that close to 0. If you get that one then I can say there exists one cyclic symmetry, and if there exist one symmetric axis about which if I give say Ang amount of angular rotation then I will not able to differentiate before the transformation, after the transformation.

Now, this C2, which means the amount of rotation will be 180 degree, if I write C3, so C2 so 360 by 2, 180 degree is my angle. If I write C3, 360 degree by 3, 120 degree is my angle of rotation. C4, 360 degree by 4, 90 degree; C5, 360 degree, 5 this way it will go. C6, C7 this way it will go.

(Refer Slide Time: 11:45)



Let me show you one example from the protein. So, this is the C2 symmetry. So, the protein is, the name of the protein as I noted in the Protein Data Bank is how amino-acid insertions are allowed in an alpha-helix of T4 lysozyme, this is the T4 lysozyme, and this is the UniProt ID and the molecule is T4 lysozyme. Its source is enterobacteria phage t4. There are two chains, A and B. Those are colored as green and cyan. Lengths are same, 166 because it is a homomer. And SCOP class now AB d 2 1 3.

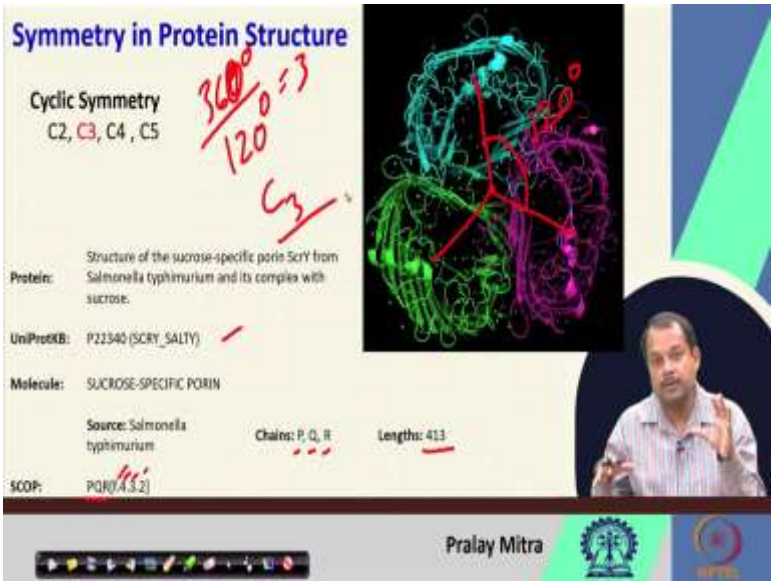
So, regarding this SCOP class we discussed extensively on the last two lectures. So, this is part of the class then, class of fold, then family, superfamily, domain. So, they belong to there. Now, regarding this structure, if you look at this structure then you will see that you can able to identify one symmetric axis which are passing perpendicular to this plane through some point something like this. Now, if you rotate this about that axis passing perpendicular to this plane by

180 degree then the cyan will go to green's color and green will go to cyan's color, and that way you cannot able to distinguish these two. So, there exist C2 symmetry.

Now, I have taken this image using the PyMOL software and after loading this molecule into the PyMOL I adjusted it so that I get a proper orientation and I can figure out, that so if I identify the axis which is perpendicular to this plane then after the rotation I cannot differentiate. So, that is why I fix it. But if you, in the PyMOL, since it is a three-dimensional structure if you rotate, then you may go for different axes. It may possible. So, based upon the molecule there may be one axis, symmetry; axis may be more than one. So, you have to figure it out.

Now, when it is basically on the, on the visual, based upon visual inspection then perhaps it is easy specifically when it is C2 C3 etc. But when we will go for, say higher dimension then what we have to do? So, you can think of some algorithm, where given two protein molecule you can say that what will be the axis and what sort of, say whether, since only two molecule is given to you, whether C2 exist or not.

(Refer Slide Time: 14:33)



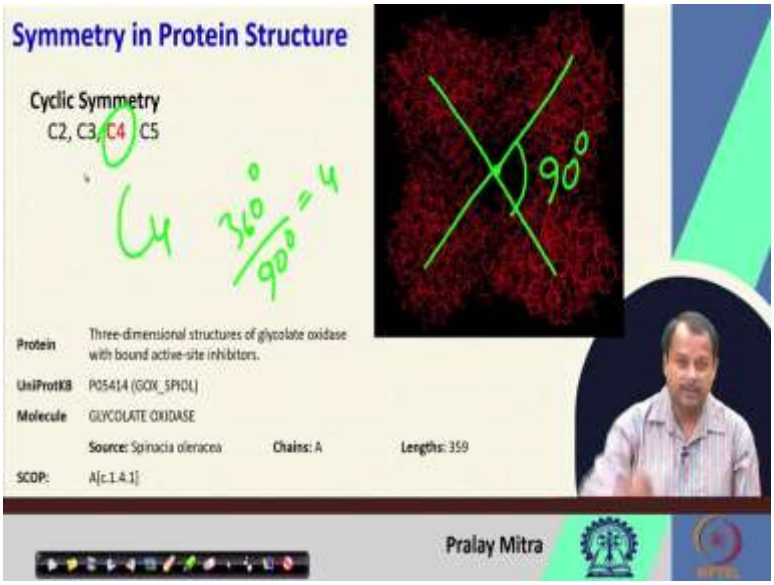
Now, if I go to the example of C3, the cyclic symmetry. So, this is C2. You see that how beautiful is the protein looks like. So, this is a very beautiful structure of a protein that I have taken. The molecule name is sucrose-specific porin, and its source is Salmonella typhimurium. There are three chains P, Q, R. All are with the same length, 413.

And SCOP class f 4 3 2. What is this P, Q, R? So, for P, Q and R, for all of them it is the same SCOP class. That is why it is written like this, PQR within square bracket f dot 4 dot 3 dot 2. Actually SCOP class is f dot 4 dot 3 dot 2, and for all of them it is same. That is why PQR is written like that. If it is different, separately it was written.

Now, its UniProt ID is given to you. It is trivial. And after looking this proteins structure also you can see that I can able to figure out one point here through which if I pass one perpendicular line, one axis perpendicular to this plane then I will get the symmetric axis. About that one if I rotate 120 degree, so this is 120? So, if I rotate about 120 degree then I will not able to understand that whether they are same or similar or not. So, that is why 360 degree by 120 degree that is 3. So, it belongs to C3. And here it is called as the trimer since three, three chains are there.

Again what I suggested that if you open this in some visualizing software like RasMol or PyMOL or say UCSF Chimera, so if you open that one it may possible that you have different orientation. So, you have to visualize that and then you have to identify what is the axis. Again how to define that axis or how to find that axis is up to you.

(Refer Slide Time: 16:56)



Next if I go for C4, so in this case I did not differentiate that four chains but I am telling you there are four chains. And also the axis is here. That you can figure out. It will be somewhere here through which you have to pass the axis. So, it is a glycolate oxidase. There is only one

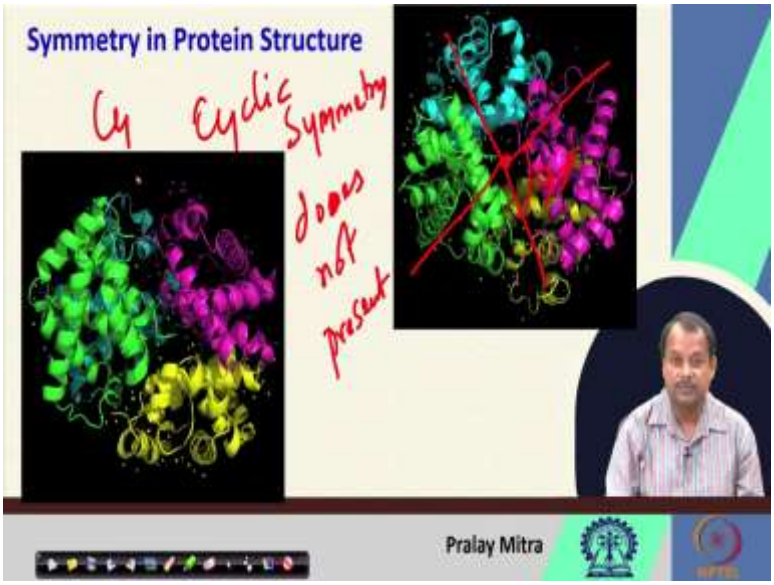
chain which means that in the PDB so there are actually four chains, sorry there is only one chain in the PDB and its length is 359. And this is the SCOP class. This is the UniProtKB.

Now, here comes one situation. In the Protein Data Bank only one chain is deposited whose length is 359. Since one chain is deposited how do you know that what is the functional form? So, it is not possible. So, that is one of the challenging problems that we will discuss right after this lecture that from a crystal structure or, say from the PDB the structure we are getting, if it is a crystal structure. Say if it is a NMR structure which means that is a solution structure, in solution, the experimental method NMR actually takes place in solution. So, I can with surety say that it is the functional form of the protein.

But if it is a crystal structure how do I know it is, whether it is a functional form or not? Specifically in this case, only one protein molecule is present, then the question if I ask to you that what is the functional form of this protein. From the PDB, so its PDB ID, sorry I missed to give you the PDB ID, so from the PDB ID, never-the-less with this information if you search then you will get the PDB ID. So, if you, if you basically look at that PDB then you will find there is only one chain. But actually there, it functions as a tetramer. So, there are four chains.

So, since there are four chains, sorry, since there are four chains, so you can see that if you have one point here, so if you give 90 degree rotation about the axis which is perpendicular to this plane then it is not possible for you to distinguish whether it is the same one or the transformed one. So, 90 degree, 360 degree, you are getting 4, which means it is C4. That is your symmetry. Like that way so C5, C6, C7; so those may come. I mentioned may come but sometimes may not come also. Why? That I will discuss later.

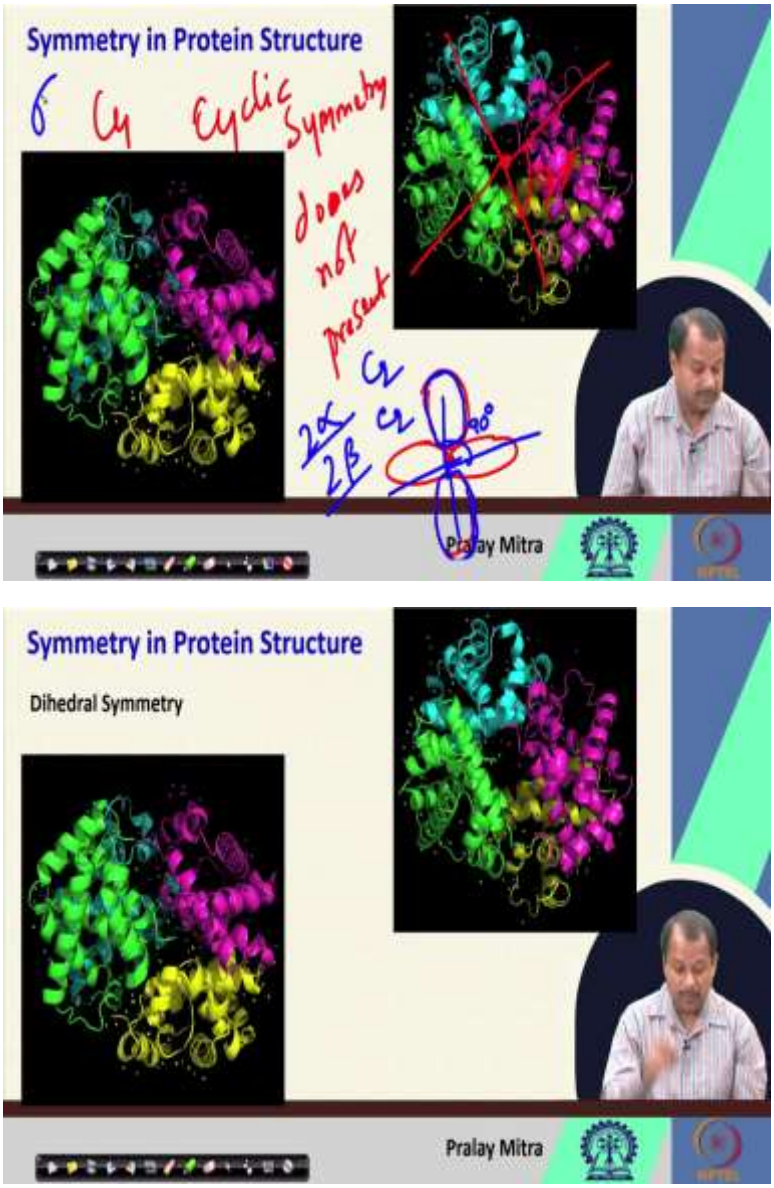
(Refer Slide Time: 19:50)



Now, this is the symmetry of the tetramer that you can see. This is another structure. Now, you see that as I mentioned that axis may be different, and in this case it is one situation which says that cyclic symmetry perhaps is not possible. So, if I look at this right top structure then if I draw one, say if I draw one point through which perpendicular axis is passing, then also if you rotate, then from the color you can see that there is a green, yellow, magenta and cyan.

So, there is no guarantee that after, so four module, so you are expecting that rotation by 90 degrees. But after rotating the 90 degree, so you can able to see that, so it is not the same structure. So, symmetry is not present. No symmetry present but cyclic symmetry do not present. Regarding the cyclic symmetry, so in the 4, so 90 degree I am considering which means C4 cyclic symmetry does not exist in this case. But there may be other symmetry.

(Refer Slide Time: 21:18)



Now, the other symmetry, regarding that one I am talking about; so if it is the case you consider, and here I am considering homomer only. So, four chains, four chains are there, and I am drawing like this. In this drawing, so one situation is all the four ellipse-like, say structures are staying on the same plane. If it is on the same plane then I can identify one axis which is passing perpendicular to this plane and then what I can do that, I can divide it here, and this is 90 degree. If I rotate then I will see that mostly it is identical.

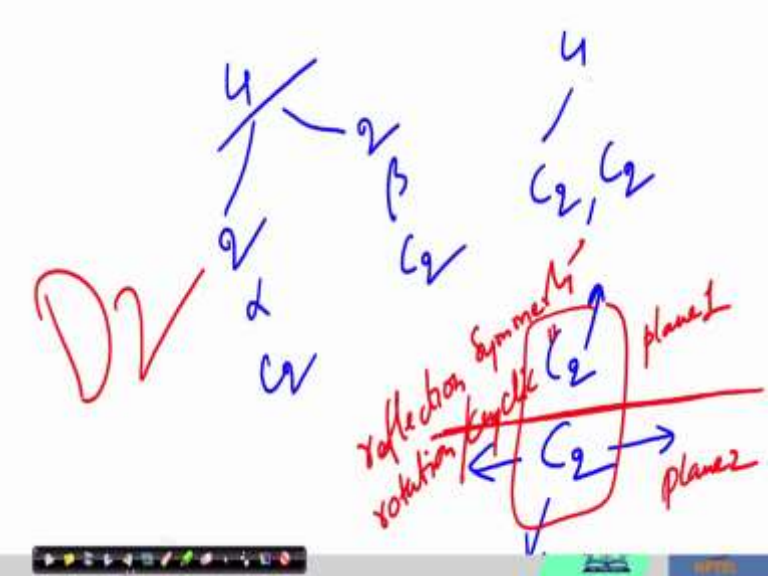
Now, if it is the case that, so there are four chains. If it is the case that there are, say two chains which are similar, say two alpha two beta. Then between these two alpha there can be C2, very

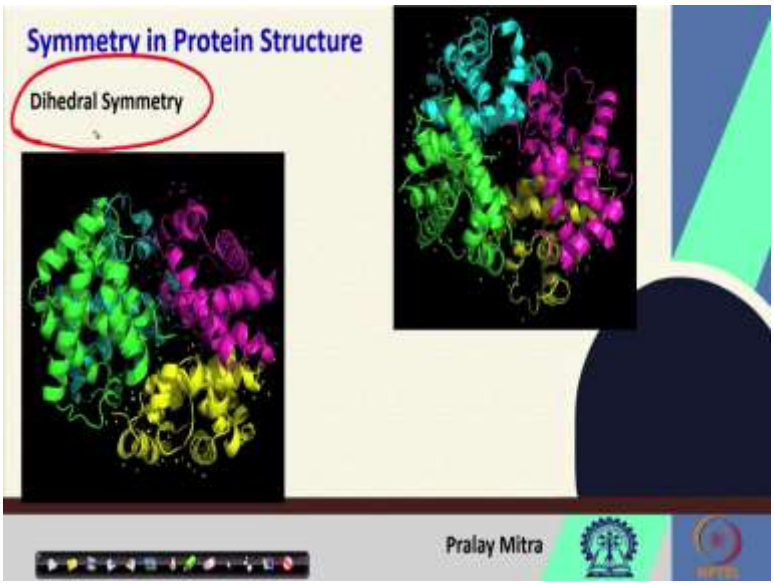
much. Between these two beta there can be C2 very much. What I am trying to say that if each two are same, this blue are same and these reds are same then basically there exist two C2 but not one C4. That is possible. But again C4 is not here.

Another possibility is that, in one plane, say these two are present. In another plane, on top of this, these two are present. So, here it is C2. Here it is C2. But C4 is not present. In this scenario, so these, all the four can be of same protein molecule, I mean the same chain, it may be different chain. If it is same chain, this C2 and this C2, then the situation is, so I can able to identify one axis about which, so two chains in one plane have C2.

Another two chain in another plane have another C2. And between these two planes, so this is in one plane, this is in another plane; between these two plane I can have another symmetric axis about which they are C2 or there is some reflection, so sigma. So, that kind of situation is being called as the dihedral symmetry. You got my point.

(Refer Slide Time: 24:04)

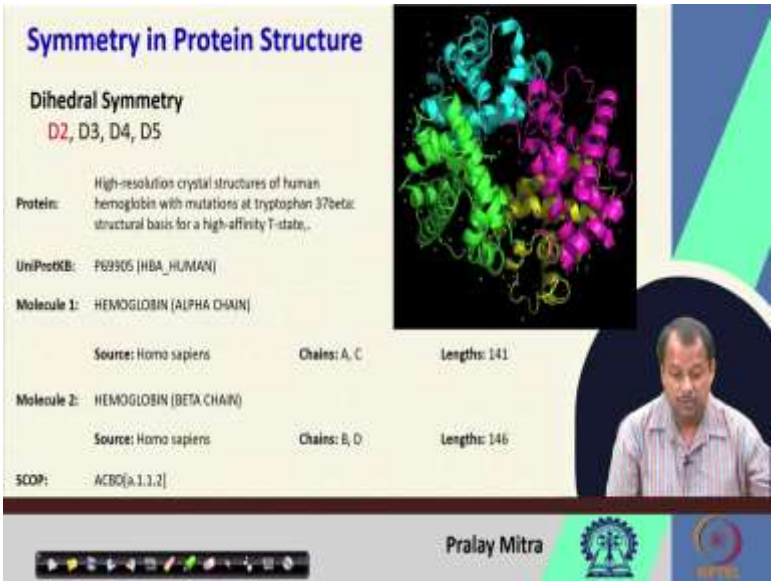




So, when I am considering, say 4, so 4 can be, again 4 homologous. So, it can be 4 consists of 2 and 2, say 2 alpha 2 beta then there can be C2, C2. That is one possibility. 4 can be, say consists of, say all 4, all 4 same, all 4 same but there can be only C2 to C2, or there can be C2 in one plane, C2 in another plane. And this is one plane, this is another plane. And there is one line, red line between these two planes.

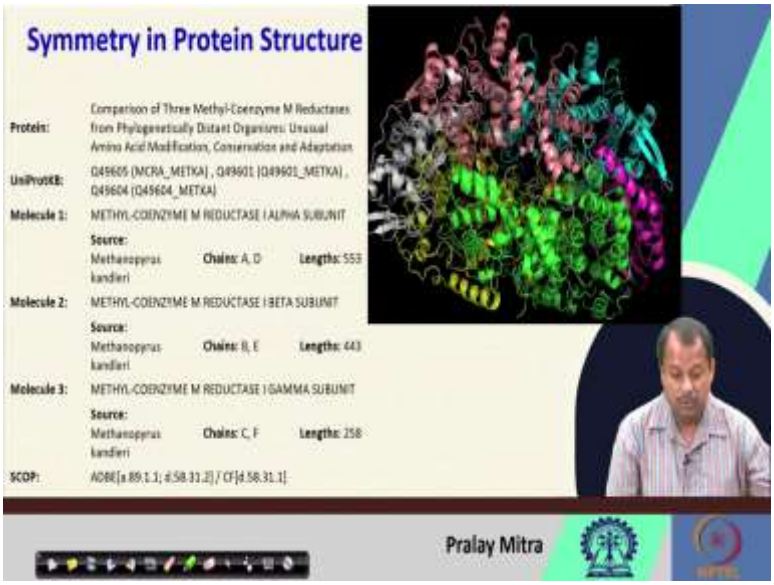
So, plane 1 and plane 2. About this, so either there can be reflection or rotation or which I am calling as a cyclic, it can be cyclic symmetry. So, about this red. Then I am getting what is called as the D2. That is the dihedral symmetry. That is the dihedral symmetry. So, dihedral starts with D2. Then very much there can be D6.

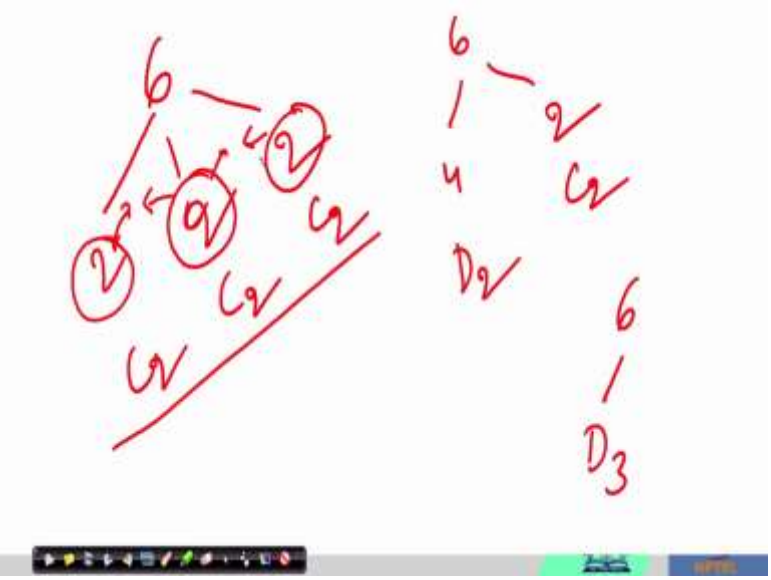
(Refer Slide Time: 25:40)



So, here there is an example of the dihedral, the same protein it is basically hemoglobin alpha chain, so there are two chains, two different chains, four chains actually. Two are alpha, two are beta. So, AC alpha and, say BD beta. So, that you can note here. Hemoglobin alpha chain AC, hemoglobin beta chain BD, so two different C2 exists. Next this is dihedral symmetry of class D2.

(Refer Slide Time: 26:17)

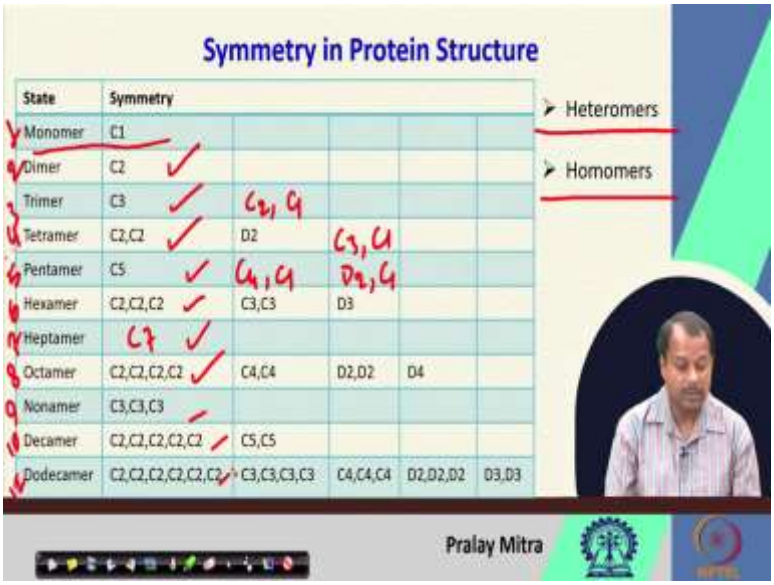




There is a possibility of D6 also. But when there are 6, again the possibilities considering homomer, so 2, 2, 2 which means C2, C2, C2. That is one possibility. Another 4, 2 so here I can have D2 C2, that is one possibility, or from 6 I can have D3 that is one possibility. So, these many number of possibilities exist. But symmetry is there if it is homomer. If it is heteromer then definitely how can I do that one? So, for heteromer it is not possible.

But once it is hexomer, so this, this is homomer, this is homomer, this is homomer. But this indicates if I take one chain from here, another chain from here then they are heteromer; one from here, another from here that is heteromer. That way it differs.

(Refer Slide Time: 27:28)



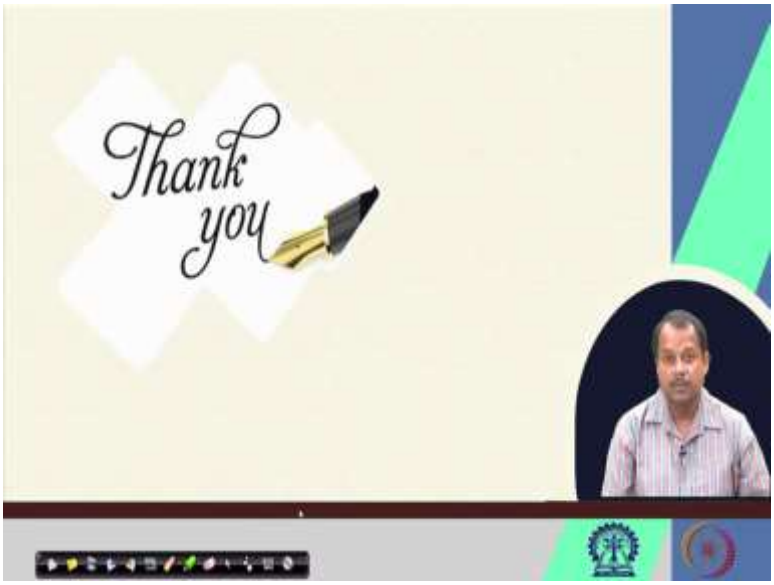
Now, this heteromer and homomer we have discussed. So, we discussed this heteromer and homomer. In case of symmetry if I look at the state, so monomer, dimer, trimer, tetramer, pentamer, hexamer, heptamer, octamer, nonamer, decamer, dodecamer. So, which means, usually, so this is, this is 1, 2, 3, 4, sorry, 4, 5, 6, 7, 8, 9, 10, 12. 11 usually does not exist. Now, what are the possibilities? That I have listed here.

So, for monomer it is C1 that is 360 by 360, 1. For dimer C2; for trimer C3 or it can be C2 and C1, so when there are two same chains and one different chain, homomer partly and heteromer then C2 and C1 can. For tetramer C2, C2, D2 or it can be, say C3, C1, that may possible, so if it is heteromer. But if not then it is not. So, if it is pentamer C5 is one possibility, then C4, C1 then D2, C1 so these number of possibilities are there.

So, ideally for heptamer C7 should be there but it has been noted long back that generally after trimer, after trimer odd number of mers does not occur frequently. It is the even number of mers which occurs. And that way if you look at the statistics, so mostly it is dimer, after the dimer it will be trimer, then tetramer, then it will be hexamer, then it will be octamer, then it will be pentamer, then it will be heptamer, of course I am not considering this monomer.

And then nonamer occurs very rarely, and decamer occurs, and sometimes mainly for the virus architecture the dodecamer occurs. So, that is mostly about the symmetry. We use this symmetry for two specific applications. That we will discuss in the next classes.

(Refer Slide Time: 30:08)



So, in summary what we have discussed here in this class is, in this lecture is basically the symmetry in proteins. We mention that there can be three different point group symmetry in protein apart from, say reflexive one, so cyclic, dihedral and cubic. Cubic usually does not occur for proteins with the size say up to 10 mer and since we are not currently dealing with the large size protein so we will exclude that one for the time being.

Now, after that one, what remains is a monomer, dimer, trimer, tetramer, heptamer hexamer, octamer, heptamer, nonamer, and decamer. So, all will be covered by cyclic and dihedral. If it is heteromer there is no question of symmetry. But if it is a homomer then symmetry must occur. Then the question is, whether the symmetry can be used to identify the biological form or the function of a protein. We will see two applications in our next lecture. Thank you very much.