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

Lecture – 17a
Protein Structure Analysis I

In this lecture, we will discuss about the properties, which can be derived from the protein 3D structure. In the last class we discussed about protein 3D structures right what did we discussed the last class, how to get the 3D structures?

Student: X-Ray Crystallography.

Experimental X-Ray Crystallography.

Student: NMR.

NMR spectroscopy and electron microscopy, and there are several Nobel laureates right for understanding the structural function of globular proteins using X-Ray Crystallography or NMR spectroscopy right.

So, all the structures which is solved by X-Ray or NMR or a electron microscopy are deposited in the database right what is the name of the database?

Student: The Protein Data Bank.

Is the Protein Data Bank right Protein Data Bank is maintained right by the Research Collaboratory in Structural Bioinformatics right. So, there are several sites in Japan, in US and in Europe right so that we can get the access from different parts of the world. So, what are the major contents of Protein Data Bank?

Student: There will be header information.

Right they have the header with the name and the source right and the resolution and the secondary structures, publications and we can have the data and the coordinates right. So, what are the information we can obtain from the coordinates.

Student: Atom name.

Atom name.

Student: Atomic number.

Atom number, residue name.

Student: Residue

Residue number.

Student: Chain

Chain information.

Student: And XYZ coordinates.

And the coordinates xyz coordinates.

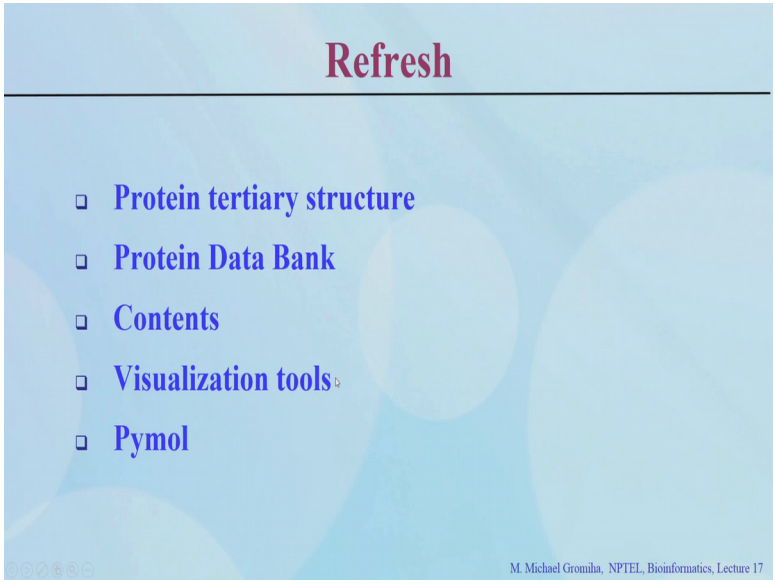
Student: And then B factor.

B factor.

Student: Occupancy

Occupancy right we will give all the information for all the atoms right in a protein.

(Refer Slide Time: 01:52)



Then we discussed about the visualization tools, there are several tools right which are commonly available to view the structures and you can also manipulate the structures and different directions, and you can calculate the bond length, bond angles, torsion angles mutations and so on. So, then we discussed in detail about Pymol. So, Pymol has several applications what are the potential applications of Pymol?

Student: you can alculate the distance.

We can get the different measurements and you can give the structures and you can make high quality figures right and we can do the mutation analysis, we can do the interactions, where hydrophobic or electrostatic or different interactions right. So, it has wide applications right. So, if you go through the Pymol right you can check the tutorials and we will get several options you can utilize Pymol effectively.

So, we take the protein structures, mainly they are classified into 4 different classes depending upon the secondary structures present in protein structures right. So, if you what is the secondary structures?

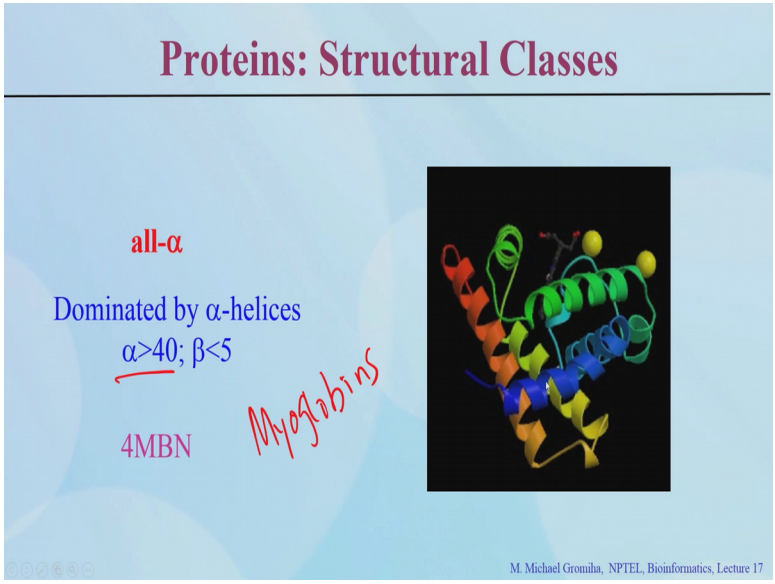
Student: Alpha helix.

And the secondary structures are alpha helices.

Student: Beta strand.

And beta strand right how these secondary structures are distributed in 3D structures, based on that we classify the proteins into 4 different groups. The major the major one that is all alpha proteins right all alpha protein contains mainly alpha helices.

(Refer Slide Time: 03:13)



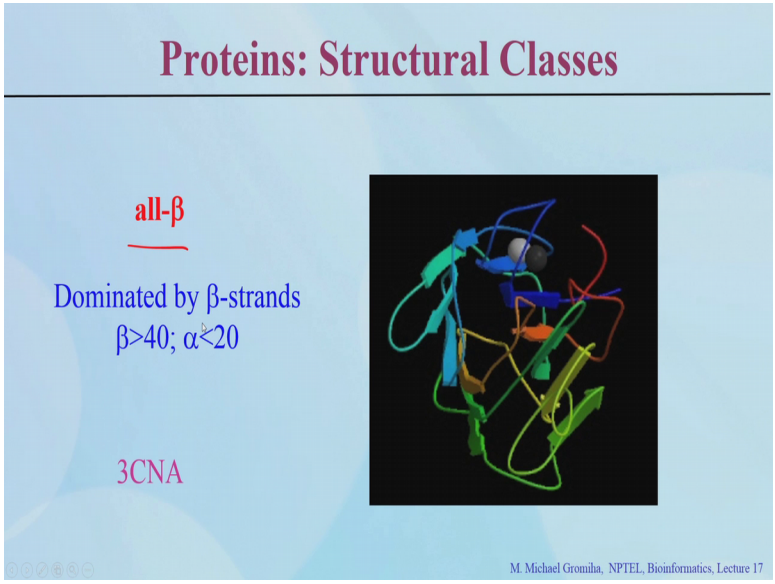
You can see dominated by alpha helices right. So, more than 40 percent alpha helices and less beta strands right we can see an example right here you can see several alpha alpha helices.

This is the structure for myoglobin right, this is a structure for the myoglobin right it contains alpha helices, but that no beta strand here. So, we can see the structures and all alpha proteins are getting popular because in the first solved structure of globular proteins right which contains.

Student: Alpha helices.

Alpha helices right. So, in this case this is popular and if you see the membrane proteins, here also we can see that the first solved membrane protein structures right and this photosynthetic reactions and right which also contains several alpha helices right. So, alpha helices we can see the all alpha proteins are predominant right. If you look at the all protein structures in the several structures they belong to all alpha class. Then the second class likewise all alpha, some proteins we have more number of beta strands predominantly you can see the occurrence of beta strands and these type of proteins are called all beta proteins.

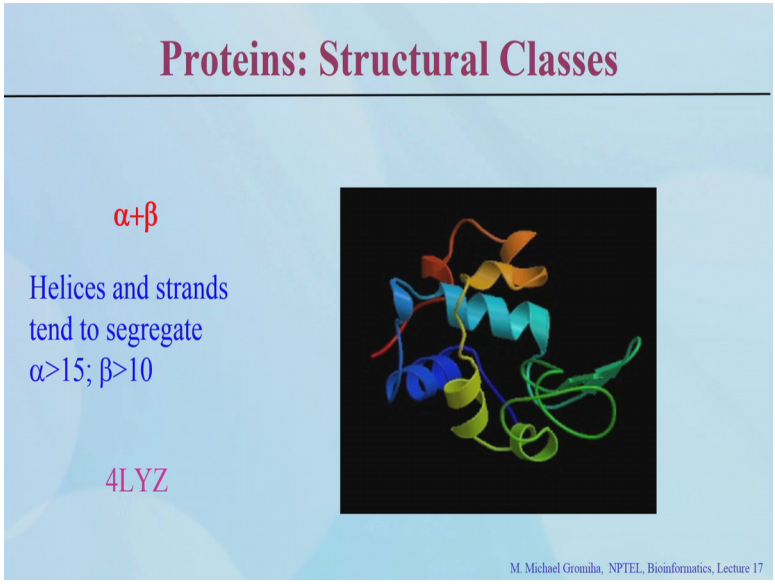
(Refer Slide Time: 04:18)



You can see the dominance of beta strands right we can see in the structures right.

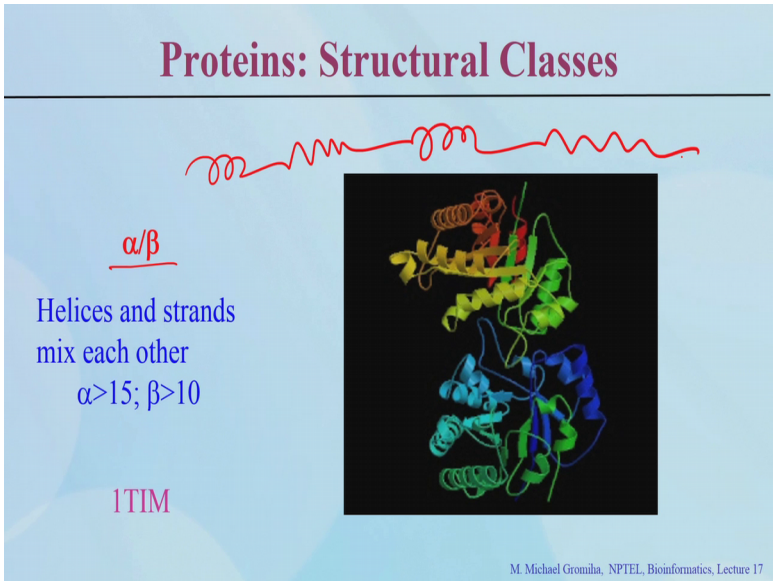
So, we can see lot of beta strands for example, more than 40 percent beta strands right and less alpha helices right this is an example this is concanavalin a. So, it contains several beta strands in this protein. So, one contains mainly alpha helices and the second class contains only beta strands. Now the next possibility is the proteins contain both both helices and strands right. Based on the location of these helices and strands they are classified in 2 groups one is alpha plus beta and the another is alpha by beta, depending upon the location of this beta strand alpha helices and beta strands.

(Refer Slide Time: 05:06)



Whether they are separate segregated separately or they are mixing with each other. So, one class is alpha plus beta proteins right in this case helices and strands tend to segregate you can see the presence of more than 15 percent alpha helices right and more than 10 percent beta strands, and the helices and strands oh here you can see the helices and you can see the strands right. So, they are segregate from each other and this is one example for lysozyme example for the alpha plus beta proteins.

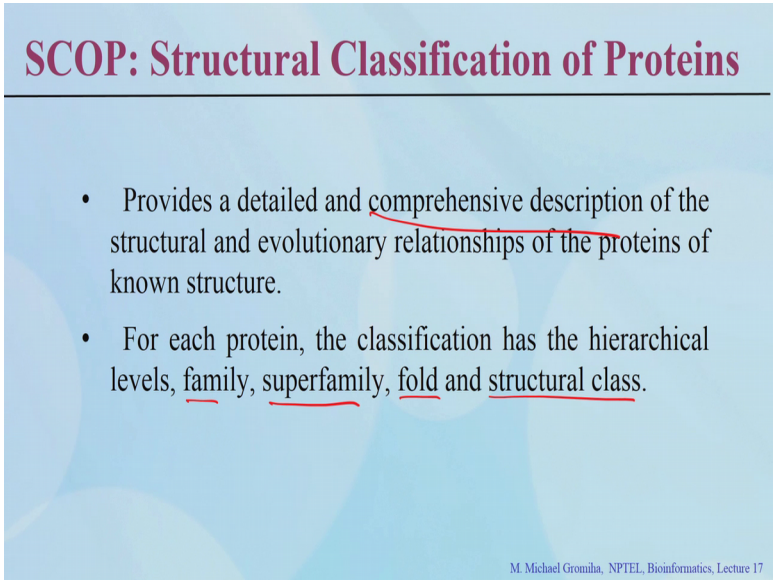
(Refer Slide Time: 05:34)



So, the forth type of proteins right their alpha by beta proteins right here also you can see the helices and strands, but they mix each other. We can see the more than 15 percent helices and more than 10 percent beta strands right. So, this is the numbers for understanding, this is not a hard and fast rules to have this specific numbers right. So, this helices and strands right which are mix with each other right for example, this is one team right it contains 8 alpha helices and 8 beta strands. So, we can see alpha helices and beta strands alpha helices beta strands and so on right you can see the structures in the TIM barrel proteins.

So, how can we get the information, whether a protein belongs to all alpha proteins or all beta or alpha plus beta or alpha by beta? So, to understand these characteristic features right there are several databases the major one is the S.

(Refer Slide Time: 06:20)

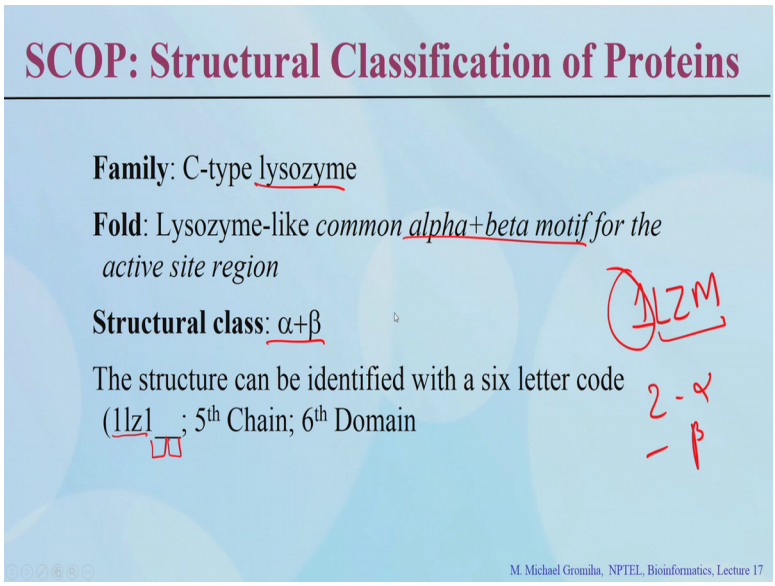


So, to classify the proteins based on different structure classes; not only the just structure classes then I also have several sub classifications, based on the how they fold which family they belong to, what are super families right all the information they give.

So, SCOP gives the comprehensive description of the structural and evolutionary relationship of the proteins of a known structures. They started from the structures available in Protein Data Bank, and they classify the proteins right based on the information available in the structural data. So, they have the hierarchy levels. So, go with the family and then super family for example, take a protein right first the small one

they belong to some family, then different families they come close together and they have a super family, that different super families they belong to same fold. So, go with the fold and the different folds right they have the similar class right for example, the structural class all alpha right we can see a 4 helical bundles fold and different types of helical folds right, they are mainly a fold these all the helical proteins having different folds right they belong together and under the class all alpha.

(Refer Slide Time: 07:25)



So, for example, he give one example say here lysozyme right, the families lysozyme then you can see the fold right you can see the lysozyme like common alpha plus beta motif. And the structural class alpha plus beta right and you can identify these protein in a 6 letter code, first 4 are the PDB code right how to represent the PDB ID?

Student: First one.

First one is the.

Student: Numeric.

Numeric way, first numeric and then 3 letters.

Student: Can be

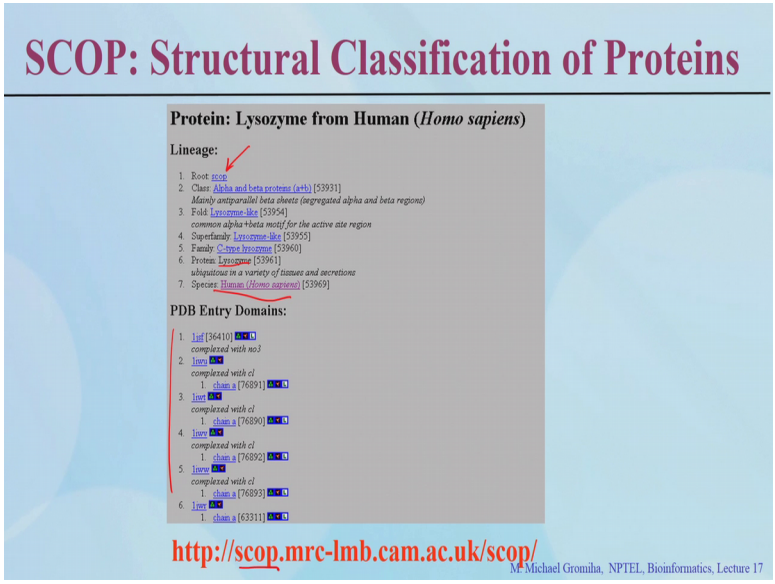
That can be numeric or any alphabets right the for example, 1LZM. So, you can this is always a numeric and look at this 3 either this is alphabets or numeric right. So, you have

the first 4 letters, and if you have any chain information for example, a protein contains different chains right for example, hemoglobin how many chains in a hemoglobin?

Student: 4.

Four chains right 2 alpha and 2 beta. In this case 4 chains for example, if is a ABCD, then you can see the chain information here. If it is a chain it is A B chain it is B and so on. And the sixth one you can see there are several domains for example, here a large protein, it contains several domains right which are stable right. If you know the domain information then we can put the number 1 2 3 right they have domain information. This is how they represent the structure of any protein.

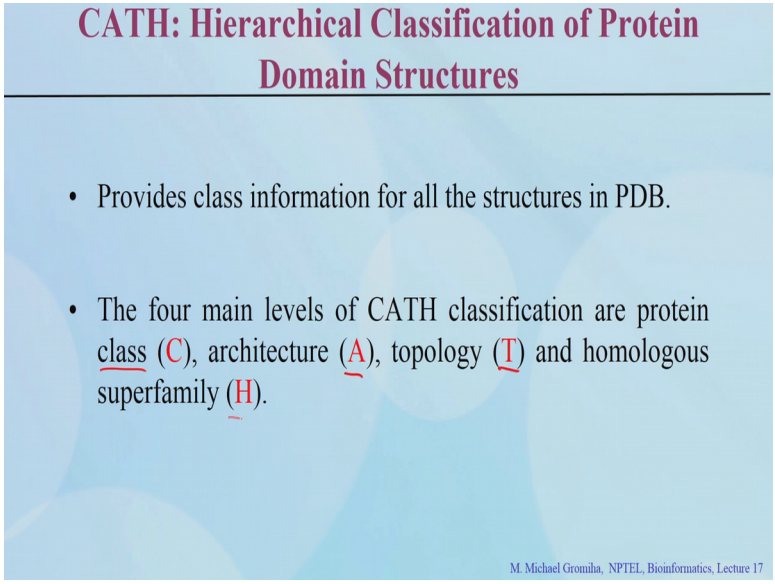
(Refer Slide Time: 08:47)



So, this is the web server right the database of SCOP, you can search the database right with any PDB id right or with the names. So, here you can the root is the SCOP is the protein belongs to alpha plus beta, and the fold is lysozyme like right and you have the superfamily and the and the protein name is lysozyme, and the species *Homo sapiens* this belongs to the human. Then they give the other entries which are relevant to put the particular protein for example, in this case it is lysozyme. So, what the other proteins right which is similar to lysozyme, you can give the details about the similar entries.

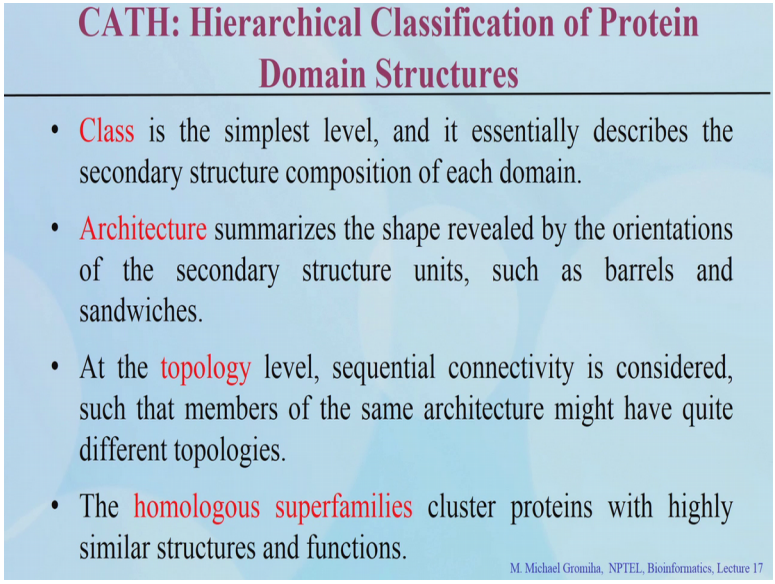
So, here is a web server the SCOP server. So, access this site, you can access this database, and you can see the structure class information for several proteins.

(Refer Slide Time: 09:33)



Similarly, there is another one server, right this is also called kind of the database right they called a CATH. CATH here also they tried to classify the structures based on the class similar to SCOP, and the architecture and the topology and the superfamily. They use this word CATH which represent class A for architecture T for topology and H for homologous superfamily.

(Refer Slide Time: 10:07)



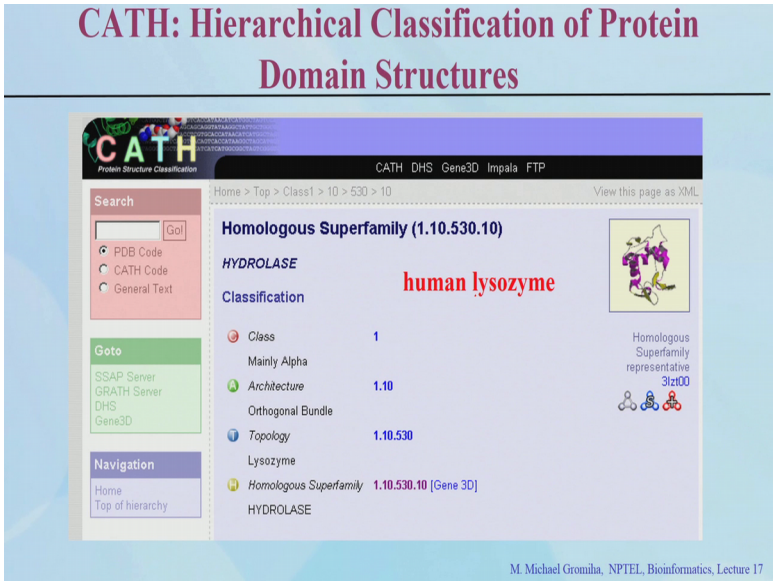
So, what is C, what is A, what is T and what is H. The C is their class is a simplest level for different classes just we discussed what are the 4 different classes?

Student: Alpha.

Alpha all alpha all beta alpha plus beta and alpha by beta right we can see the 4 different this say the simplest way right. And then go with this architecture this summarizes the orientation of the structure units like this is the barrel or sandwich and so on. Then the topology level here you can see the sequence of connectivity, how the members of this architecture right how they are connected sometimes you have the same architecture, but have the different topologies right depending upon the alpha helices beta strands how they are connected with each other. Then the last one is homologous superfamilies that is H a similar structure and function.

So, based on the connectivity and the proteins which have a similar structure function or the orientation of secondary structure right they classify a different groups right. So, this were they called as CATH C for class, A for architecture, T for topology, and H for homologous superfamilies ok.

(Refer Slide Time: 11:03)

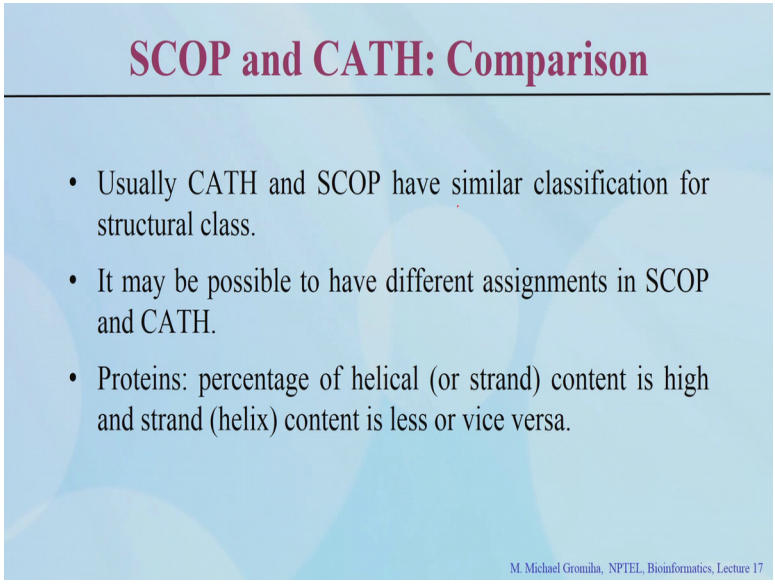


This is the database CATH database you can here for the same human lysozyme. So, it is the class is mainly alpha here, and the architecture they put the number 1.10 this orthogonal bundle and the lysozyme right and the super-families belongs to hydrolase.

In case of SCOP right, what is the classification for the class? Lysozyme they put alpha plus beta, but here they put all alpha right. Most of the cases we can see the similar

classifications and some cases we can see the differences right I will explain why there is a some difference right for the same protein in different databases. If you look into several structures right almost all this structure in the Protein Data Bank.

(Refer Slide Time: 11:49)



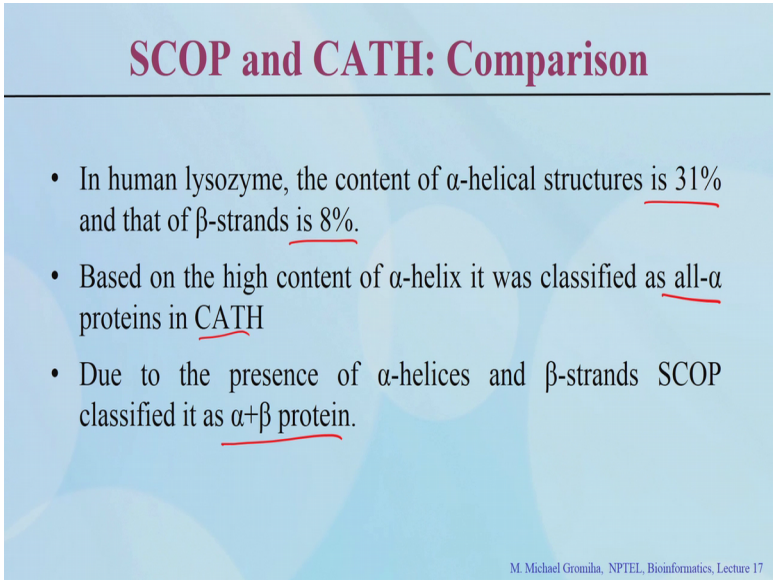
And if we classify based on SCOP or the CATH, most of the cases we can have the similar classification right that not much different.

In some cases it is possible to the different assignments in SCOP and CATH; especially for the case of this a alpha plus beta or alpha beta proteins right how this happens right how do you classify the all alpha?

Student: more than 40 percent.

Mostly alpha helices how you classify alpha plus beta? It has both alpha helices and beta strands right depending upon this cutoff values right. For example, if you take 10 percent or 15 percent right then you can say either they belong to all alpha or belong to another mixed class alpha plus beta or alpha beta right. If the helical content is high then we can see this can be all alpha proteins. Sometimes the helical content is high, but also it contains beta strands. In this case we have the conflict if we consider the high helical content right in this case you can see this can be classified as all alpha.

(Refer Slide Time: 12:48)



So, in this case human lysozyme, if you look at the look into the contents alpha helical structure is 31 percent and beta strand is the 8 percent. So, if high content of alpha that more than 30 percent. So, classified as a alpha all alpha protein right in CATH. Due to the presence of helices and strands more than 5 percent strand and more than 30 percent helix right it is classified as alpha plus beta in the case of SCOP. This happens only for few structures. So, in this case we need to check these databases see whether there is a consistency or not, then based on your requirements we can classify this as the all alpha or alpha plus beta. Then if you have the structures right we can get the structures from Protein Data Bank right then you can derive several parameters, like the from sequence we derived various factors or the various properties we derived from sequence?

Student: Amino acid

Occurrence composition.

Student: Molecular weight

Molecular weight.

Student: Average property

Average property values.

Student: Hydrophobicity.

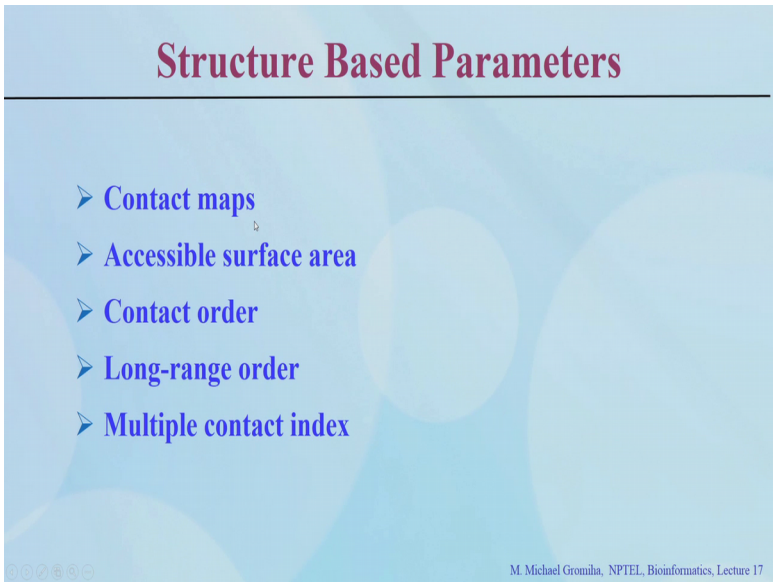
And then hydrophobicity profile.

Student: dipeptide.

And we can do the; a dipeptide composition and we can do the alignment multiple sequence alignment conservation right several features we can do it. Likewise if you have the 3D structures we can do better because 3D structures contain more information than amino acid sequences.

So, we can derive various parameters right and the important aspect is whatever the parameter we derive from this structures, they have some applications to understand the structure or function or anything related with diseases. I will explain some of the features for example, contact maps right.

(Refer Slide Time: 14:22)

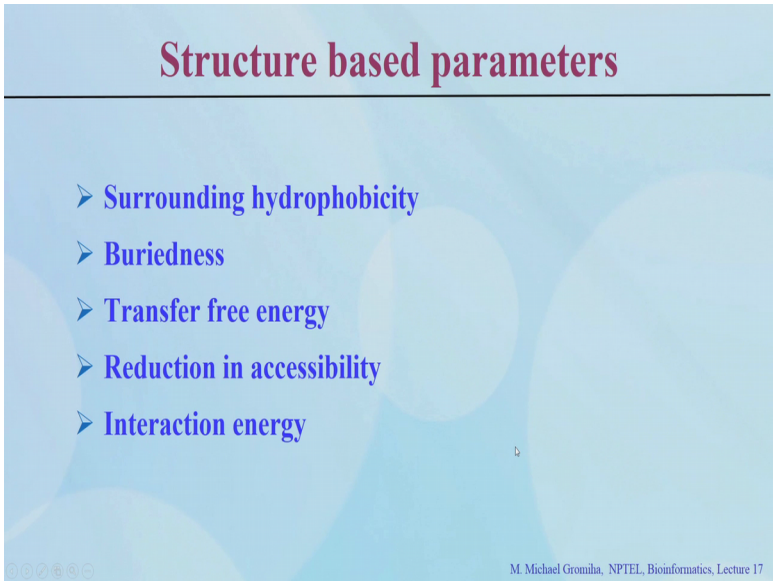


This is the simplest one we can construct from protein 3D structures, and accessible surface area, contact order right depending upon the how the 2 residues are in contact in protein structures, and long range order they depending upon the contacts which are closed in space, but they are far in the sequence level, and some cases some residues you have more number of contacts some case less number of contacts right some case more number of contacts.

For example in a class; for example, here class representative right keep contacts large more number of contacts right some students they have less number of contacts right. So,

then these contacts are important right. So, the residues which have more number of contacts they have a higher influence right this is the principle of multiple contact index.

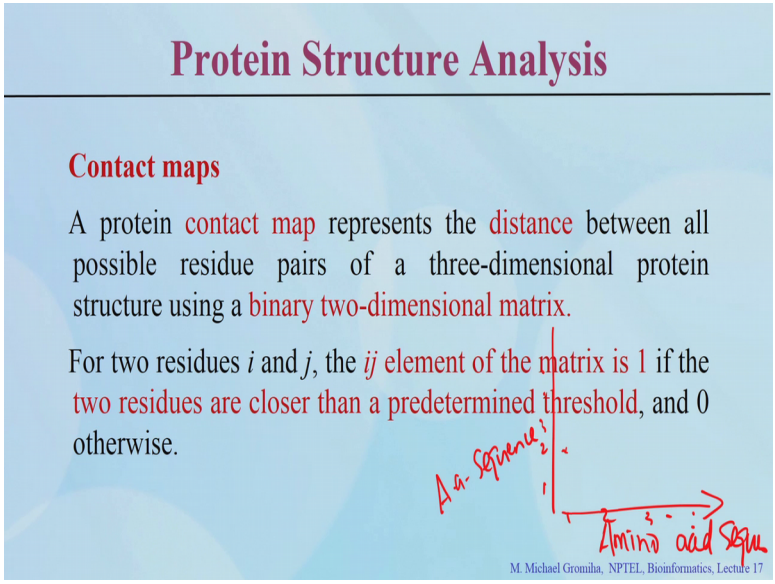
(Refer Slide Time: 15:06)



Then we can derive other factors like hydrophobicity, buriedness, transfer free energy how the accessibility is reduced from the unfolded state as it goes to the folded state, how they get the different interactions; cation-pi interactions, electrostatic interaction, hydrophobic interactions and so on.

So, we will explain some of the parameters right and derive how to do it. So, first one is this simplest one, we can obtain from 3D structures, it is the contact maps right. So, what is a contact map? It simply represents the distance between different residues in protein structures. So, we have a 3D structures the atoms and the residues are located with the xyz coordinates right and we in the contact map, we can see how residue 1 and 2 are in contact with each other 1 and 5 are in contact or not. So, the representation of 3D structures into 2D graph.

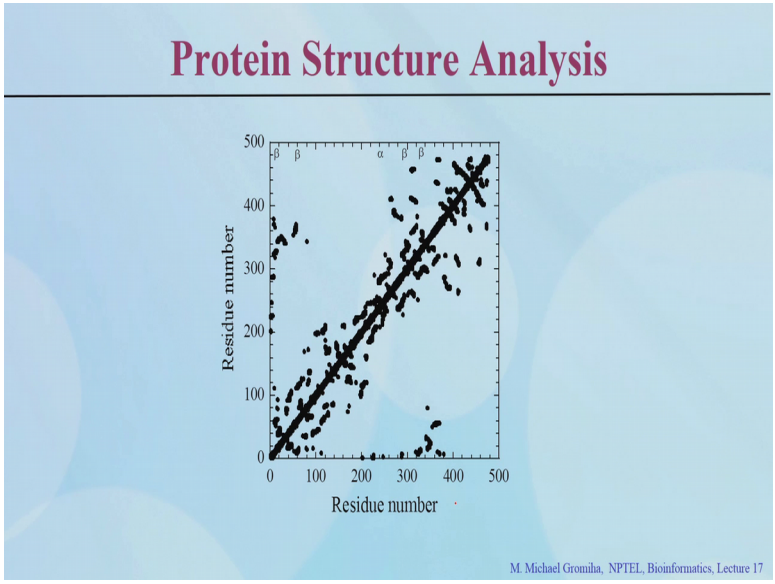
(Refer Slide Time: 16:03)



For example x axis if we have the amino acid sequence right. So, amino acid sequence here, and y axis here the amino acid sequence. So, 1 2 3 like that right.

Now, the question is whether this residues are in contact or not. If they are in contact you put a dot if no contact leave it then you will get a matrix right will show you whether which residues are in contact right I will show a some example. So, the 2 residues i and j one the residue i second another residue j right if it is equal to if this matrix is one if the residues are closer than any specific threshold like any distance. In this case we need the information regarding a distance right and then getting the distance which atoms we need to consider these are 2 different aspects we need to construct the contact map.

(Refer Slide Time: 17:09)



So, if you look this is the example for the contact map just I showed earlier. So, here you have the amino acid sequence x axis, y axis amino acid sequence right and you can put a dot if i and j are closed in the specific threshold right otherwise its a blank. If you see this graph what you can infer from this graph. So, diagonals you can see it is always present; what does it mean?

Student: Nearby residue.

Nearby residues right because they are near the van der Waals contacts. So, in this case 1 and 2 1 and 3 like 2 and 3, 2 and 1. So, they are always in contact right. So, this way it is in the diagonal you can see always represent. And if you close to the diagonal some cases we can see this some residues are present right some cases no. In this case there is no this case there is no right most of the case it is yes, but if you look into these specific cases for example, here it is the very far in this case it is will as a around 10 this is around 300.

These residues they are close in space, but they far away in the sequence right I will explain in the details now when we make the contact right as I discussed earlier the 2 different aspects, one is we need to fix the distance second one we need to fix the threshold. If we take the atoms there are various ways to define these either you can consider Cα atoms that is the simplest one take all the consider only Cα atoms and see the distance or we can see Cβ atoms right because Cβ atoms they can see the interior of the protein right this way a many many researchers they use Cβ atoms.

So, can we use Cβ atoms for all the residues?

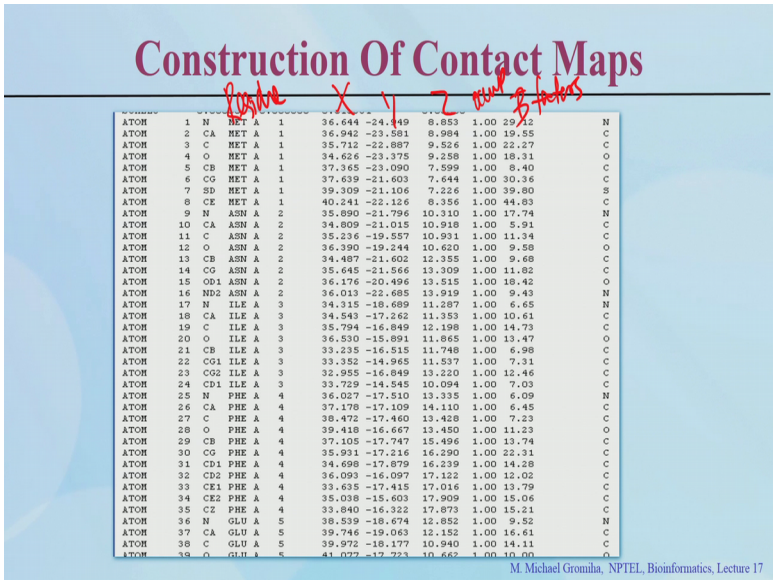
Student: No except glycine

No right because one residue right for glycine does not have the Cβ. So, that case they use all alpha and all other residues they beat beta right they represent better than the Cα. So, they use Cβ atoms or you can has any atoms for example, residue 5 and residue 10, they are in contact or not. You can see any of the heavy atoms which are within then specific distance or we can see centroid any atom any residue we can get a centroid XYZ coordinates right, now all the residues right are represented by centroids then we can calculate the distance and then you can see the cutoff right.

So, there various ways we can consider the atoms, either Cα or we can get Cβ or all heavy atoms or you can use centroid. Then the distance which distance we need to consider? 4 5 6 which distance you want to consider depending upon the atoms you consider for example, if we use a Cα or Cβ, there you can use the distance of 6 to 12 Å right because we consider only one atom. In the case if you have the all atoms then you can reduce right go for 4 Å or 5 Å otherwise you will get more number of contacts right.

So, depending upon the atoms you use either Cα or Cβ or all atoms you can define it rest of all.

(Refer Slide Time: 20:07)

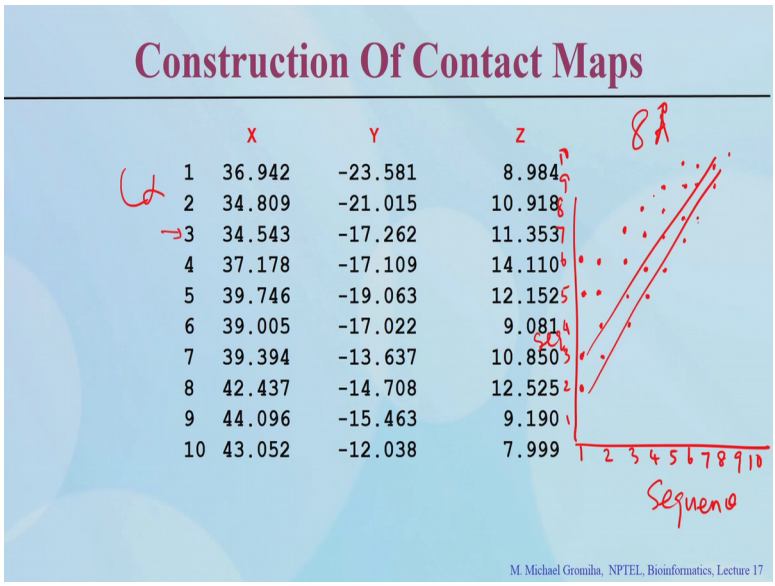


So, now here this is a coordinates. So, this is the residue name where are the coordinates? Here right XYZ coordinates, what is the next one? Occupancy this one.

Student: B factor.

B factor right. So, you have the XYZ coordinates right any protein structure if you go I discuss I showed earlier the both the Protein Data Bank right one example. So, we will gets the same the level of this representation, you can see XYZ coordinates you can use these coordinates you consider contact maps right.

(Refer Slide Time: 20:42)



For example I show the coordinates these XYZ coordinates. So, all this is a Cα atoms because I extract the Cα atoms from here right; this is 36.42 see the Cα atoms right. So, here this is the coordinates 36 minus 23 and minus 8 right I extract all the Cα atom Cα atoms.

In this case which distance it is better to define.

Student: 6 to 12.

6 to 12 or 12 Å till 6 to 14 Å. So, I use the C is the distance of 8 Å right we can these are the Cα atoms and you can construct maps how to construct a contact map? First this is a; what is a x axis, this is a sequence right this is a sequence here also you have the sequence. So, now, we have the a residues right you can see this is the one its

methionine, 2 is asparigine, 3 is isolysine right we have the residues or you can put the numbers 1 2 3 4 5 6 7 8 9 10.

So, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. So, take 1,1 and 2 are in contact yes right 1 and 3 yes, 1 and 4 about 1 and 5.

Student: 1and 4 not in contact

Not in contact 1 and 5.

Student: Not in contact.

1 and 5 also you know here 3 9 36 plus 9 equal to 45, 16 maybe in contact we will see about 1 and 6.

Student: 1 and 6 not in contact.

In contact 1 and 7 no.

Student: No

1 and 8 no.

Student: not.

1 and 9.

Student: no.

No 1 and 10, no because 12 and 20 is already right now then the 2, 2 and 3.

Student: Yes.

Yes 2 and 4.

Student: Yes.

2 and 5 yeah 2 and 5 yes.

Student: Yes.

2 and 6.

Student: Yes.

Yes; 2 and 7, 2 and 7 no right.

Student: No.

2 and 8.

Student: No.

2 and 9.

Student: No sir.

No 2 and 10 no 34, 43. So, this no then now goes to 3; 3 and 4 yes.

Student: Yes.

3 and 5 yes 3 and 6.

Student: Yes.

Yes 3 and 7 yes 3 and 8.

Student: No

No 3 and 9 no 3 and 10 no then 4 and 5.

Student: Yes.

Yes 4 and 6 yes 4 and 7.

Student: yes

4 and 7 yes, 4 and 8 probably yes, 4 and 9 no we know right yeah 4 and 10 no.

Student: No.

Right then 5; 5 and 6 yes, 5 and 7 yes, 5 and 8 yes, 5 and 9 yes 5 and 10.

Student: Yes

We know then 6 and 7 yes, 6 and 8, 6 and 9 no 36 39 43.

Student: Yes.

Yes.

Student: Yes.

So, the 6 and 10, 4 yes.

Student: Yes.

About this then 7; 7 and 8 yes, 7 and 9 yes, 7 and 10 yeah yes then 8; 8 and 9.

Student: Yes.

Eight and 10, 9, 9 and 10 right. So, you consider the matrix. So, we will get the symmetrical matrix or the non symmetrical matrix which one.

Student: symmetrical.

Yes symmetric right why it is symmetric.

Student: We did not draw the shape.

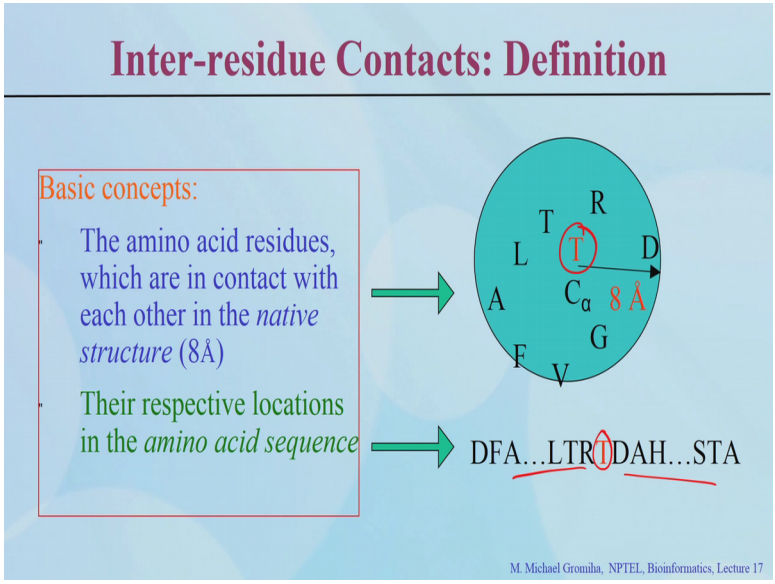
Yeah we did not draw another shape because if 2 and 3 are in contacts, then 3 and 2 are also in contact right not like the amino acid sequence that is we talk about the neighboring residues. So, in this case you do not get a symmetrical matrix right, but here if 2 residues are in contact right, 2 and 3 in contact 3 and 2 are also in contact. So, in this case you can get the symmetrical matrix. So, here you can see the diagonal, from this one you can easily say that the residues are influenced with the short range contacts because the residues are very close.

Then some residues right you can see even near the diagonal, 2 3 residues which are far away. They are also contributing right, they are also interacting in protein structures right, and here we consider only 10 residues, this is why we will shown we could see the long range contacts, but in this figure right, we can see the long range contacts for this. So, depending upon these contacts right based on the space, and how they are located in the sequence we can classify into different types of interactions right different types of

contacts. Whether these contacts are short range; that means, short in terms of sequence right because we fix the space, because when you when you construct a map the space is fixed any distance, 6 Å or 7 Å or 8 Å we fix.

Now, their difference is only the sequence level. So, the sequence level we see whether they are very close in sequence right and how far they are apart in the sequence for example, if we see this one right.

(Refer Slide Time: 26:28)



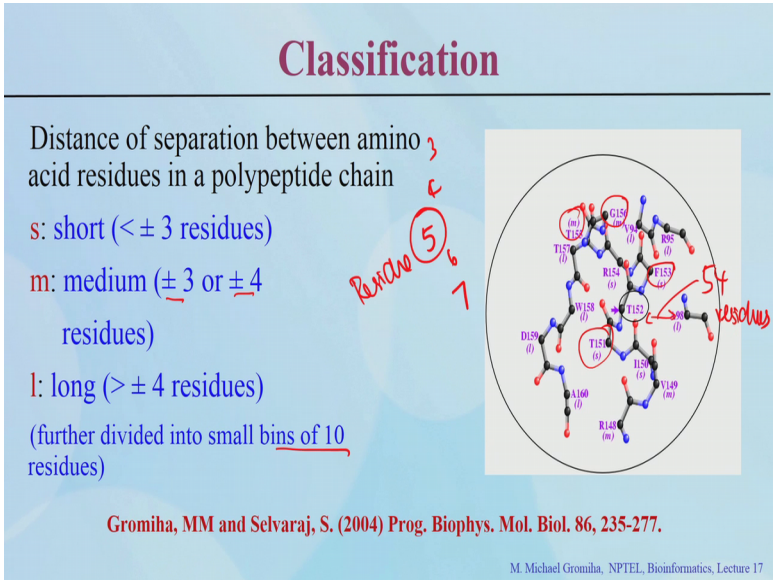
The T is the central one, and the radius is 8 Å we construct this sphere and there are several residues. Now these residues are located in the sequence which I give below right. We can see this the T is central, we can see the 3 here and how they are distributed along the sequence. Based on the distribution of the residues along the sequence we can classify into different types.

When we can see short range contacts, that we just we want to see the residues which are close to the central residue either +2 +1 or +2 almost all residues they have a short range contacts right. So, if we take any residue how many short range contacts each residue will have.

Student: 3, 4.

Four residues right for example, you take this 5.

(Refer Slide Time: 27:15)



For example residue number 5. So, residue number 5. then you can see 4 3 6 7. So, 4 contacts then you can see whether any residue the residues, they have contact with at least ±3 or ±4 residues.

In this case you can see medium range contacts, we use 3 or 4 because we need to represent the type of alpha helices right what how the alpha helices are found, what is the range? i and i plus?

Student: 4.

4 right in this case we can see it will come within this range, and you can see long range contacts if it is more than 4, but this limit is so far, the because we can see there are very large. So, we can see even 4 we can divide into small bins, 4 to 10 11 to 20 and so on and then see which range you can see the contacts in different types of a structural classes. So, here I show a figure this is the contact a T 152 of lysozyme, the real contacts if you look into the PDB structure, and take the Threonine 152 and if you get the residues which are within the limit of 8 Å you can get this figure.

You see this one you can define the short, medium and long range contacts. Some residues for example, F 153 and 151 they are just neighboring residues, these residues form short range contacts. And some cases for example, 156 and 155 they are 3 to 4

residues far apart. So, they make medium range contacts and some residues which are far away in the sequence, but they are close in space for example, T 152 and.

Student: A 98.

A 98 right what is the distance in the sequence level.

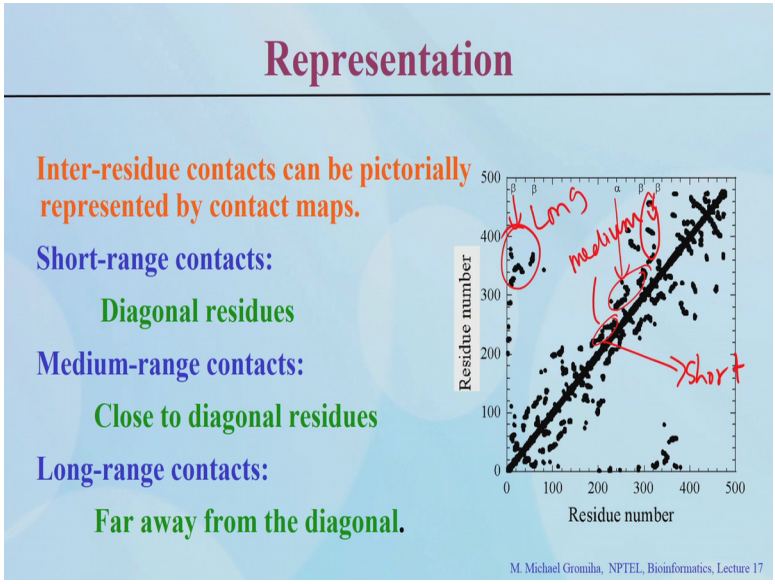
Student: 60, 54.

54 residues right. So, they are far away in the sequence. So, this residues contribute long range contacts right. So, if you take the structures then you can see the contacts based on the short range medium range and long range we can also interpret in terms of alpha helix and beta strands alpha helices we can see it is dominated mainly by the medium range interactions and the beta strands, which are dominated by long range interactions why.

Student: Because the.

Because of the hydrogen bonding patterns right in alpha helix and the beta, beta strands right.

(Refer Slide Time: 29:29)



So, now if you see this one, diagonal ones, they are mainly these are the short range these are the short range contacts and you can see the residues very close by the diagonal, these are medium range and here you can see they are long range and if you

see the secondary structures for example, we see these are the beta strands right we can see the beta strands here because there are many long range contacts and several because these are all mainly alpha helices.

Now, if you see the distribution of secondary structures and the contacts, easily you can say that, which region belongs to helices or which region belongs to beta strands based on this patterns, the long range contacts mainly they are far away from the diagonal. So, now, you show the data for the 4 structured classes, we discussed 4 structured classes right what the 4 structured classes we discussed?

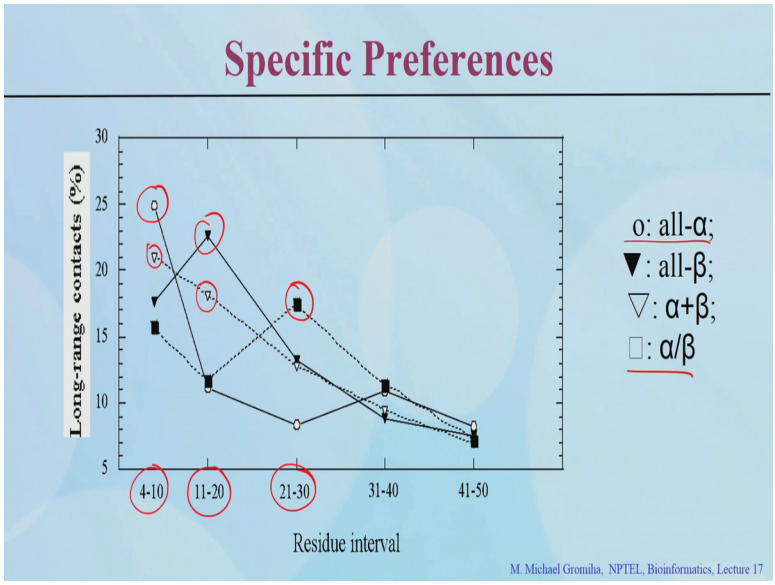
Student: All alpha all beta.

All alpha, all beta, alpha plus beta and alpha beta. So, if you see the all alpha class, all alpha class dominated with helices right. Helices are mainly dominated with?

Student: Medium range.

Medium range interactions right.

(Refer Slide Time: 30:37)



So, in this case if you see you have a different residue intervals namely 3 to 4, and then you can see one or 2 residues apart 5 or 6. So, we can see more than 25 percent of these a helices we can see mainly at this level 4 to 10 interval, and we can see it is very very less for the different guesses it is going down right. There are the contacts from the long

range it is very less in the in case of all alpha proteins, but if we look into the all beta proteins right this is the all beta proteins.

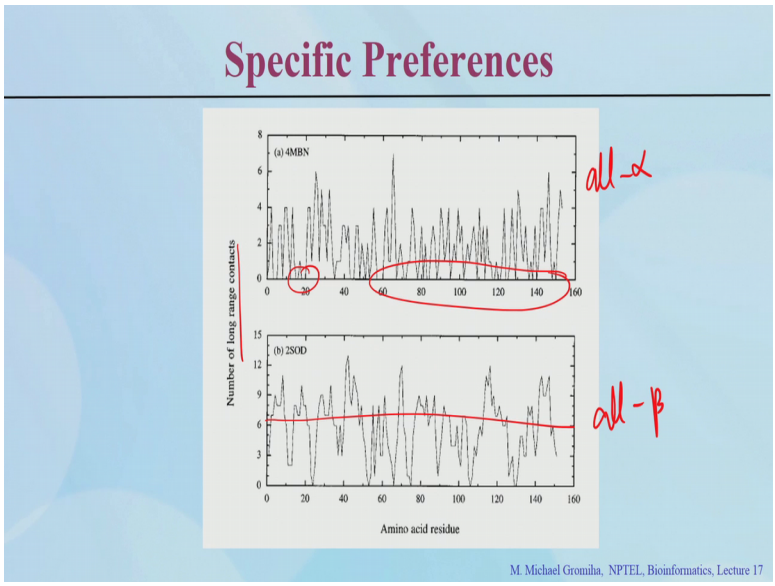
So, here all beta proteins you can see the range is 11 to 20, because mainly these anti parallel beta strands, they have the hydrogen bonding pattern with respect to more than 10 residues. So, you can see 11 to 20 residues that is the dominance in case of all beta proteins. You look in to alpha plus beta proteins right there is in between right you can see both the cases this is the all alpha is all beta we can see it is in between that. Here also all beta, here it is all alpha in between all alpha and all beta, but look at the alpha by beta proteins. So, what is the specialty of the all beta alpha by beta proteins?

Student: Alternate

Alternate one alpha one beta in this case if you take 2 beta or if you take 2 alpha, the distance will be more right if we take the beta strands right. So, beta alpha and beta. So, in between one alpha it crosses more than 10 15 residues, this is a reason if you take this type of proteins right we can see this is dominant in the 21 to 30 range. Several TIM barrel proteins right in this you if you see the hydrogen bonding pattern, they are between 21 to 30; this way it is dominant the in 21 to 30 range.

So, it makes sense and if you see the structures, you can relate the secondary structures the structural class as well as the number of contacts in protein structures.

(Refer Slide Time: 32:30)

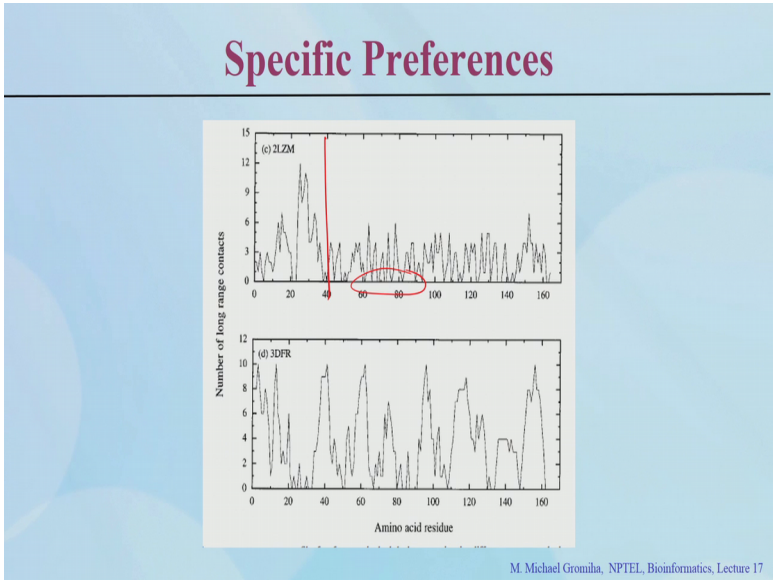


I show an example this is all alpha this is all beta right. If you look at these figures and the based on the number of contacts mainly the long range contacts right. So, which one has more number of long range contacts?

Student: All beta.

All beta its expected right. So, you can see the more number there up to 13 contacts, but here it is very less and most of the case it is 0. In any case we can see 0, there is no long range contact in the case of all alpha proteins, but here there is very less most of them are having more than 6 long range contacts what do you expect for the alpha plus beta and alpha by beta? Alpha plus beta if you take you can see the mixed combination of these two; one part you should be less number of long range contacts and another part you can see the more number of long range contacts.

(Refer Slide Time: 33:12)



You can see these right if you see here, here you have more number of long range contacts here we have less number of long range contacts.

Here most of the case they have zero long range contacts, that is fine right you can understand. So, then the alpha by beta you can see it is exact right you have some high and low high low that is you can see the patterns. So, in terms of these contacts, you can easily understand different structure classes and how they interact and how they make the patterns, then easily you can see the different functional aspects right. This is the

contact they will make and how we can superimpose the different structures and to understand the different functions.

So, now we have the contact maps to construct the contact map what are the different information do you need?

Student: Coordinates.

You need the coordinates and we need to consider the distance plus the atoms right depending upon the atom type right we can fix a distance, and we can see if this is in contact we can put a dot, and you can see, get the construct the contact maps that is simply the depreciation right of the 3D structures in 2D space. Then we can understand how they residues the contact with each other at different structural classes.