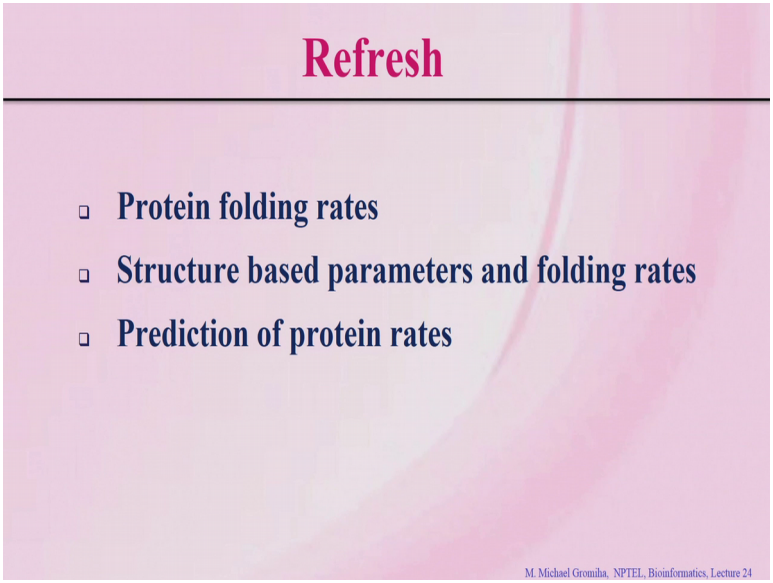
BioInformatics: Algorithms and Applications
Prof. M. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 24a
Protein Interactions I

In this lecture we will discuss about the interaction of proteins with other molecules, what are the functions of the proteins interacting with other biological molecules or small molecules? And how to identify the binding sites and are there any specific preference of interactions at the binding sites. In the last class, what did we discussed in the last class?

(Refer Slide Time: 00:41)



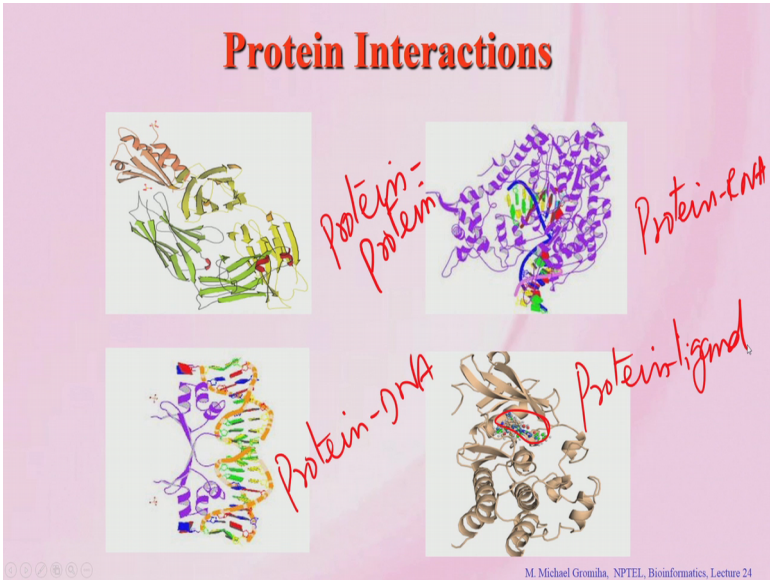
So, we mainly focus on protein folding rates right. What is protein folding rates; it is a measure right you see whether a protein can take less time or more time to fold from this amino acid sequence to 3 dimensional structure, its order of 8 order of magnitude right from milliseconds to an hour right. So, in there are various parameters right which can relates the protein folding rates.

Specifically the interaction between amino acid residues in protein structures for example medium and long range long range interactions right. So, medium range interactions influence right mainly in the case of all alpha proteins, and the long range interactions in case of all beta proteins, and we see that the all alpha proteins fold faster than all beta proteins right.

Likewise we related several structure based parameters try to understand the protein folding rates right. What are various structure based parameters we discussed in the last class; like the contact order long range order like the multiple contact index right. So, we derived various parameters and related the parameters with folding rates. If we showed that these parameters right have inverse relationship with the protein folding rates.

Then we try to utilize these parameters to predict the holding rates as well as predicting the folding rates from amino acid sequence in 2 different ways, one is directly you can use the properties or you can predict the contacts and from the contacts you can calculate the long range order and you can get these folding rates fine .

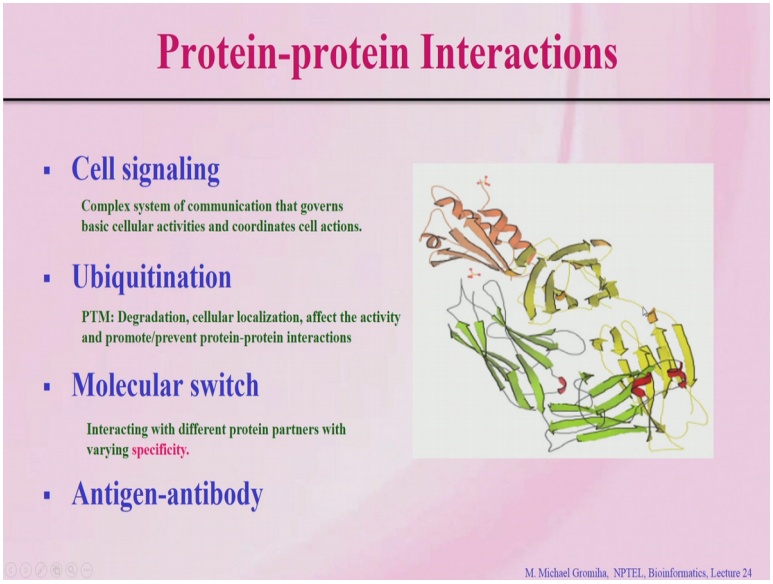
(Refer Slide Time: 02:36)



So, today we will discuss about the protein interactions right proteins interact with several other molecules right biological molecules, chemical compounds right for various functions, like protein interactions, protein RNA interactions right. So, protein DNA interactions, here you can see the protein as well as see this protein here and this is RNA here. So, this is protein RNA interactions right. You can see the protein here right and you can see the DNA. So, this protein DNA interactions right. Here you can see the proteins right and see the ligands here. So, this is protein ligand interactions right.

So, these interactions are important right for various cellular process right.

(Refer Slide Time: 03:32)

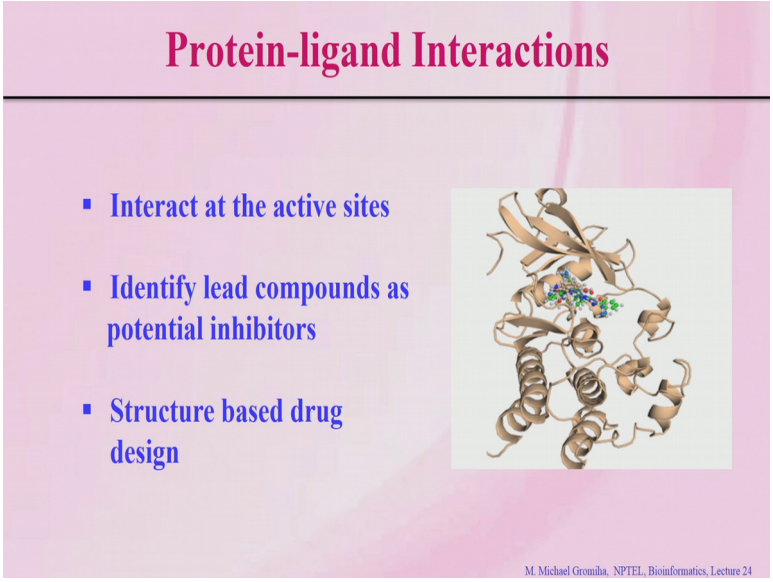


For example if we take the protein-protein interactions. So, they involved in various process for example, cell signaling right this is a complex system of communication right which governs the basic cellular activities as well as the coordinating cellular actions see ubiquitination. So, this a post translational modification right, this cells affects the activity and promote or prevent the formation of this protein-protein interactions. So, also the several protein-protein complexes, they are acting as molecular switches right this interact with the different proteins with the varying specificity right.

So, first for example, say protein "a" interact with "b" with specific specificity as well as with C with different specificity. In this case they have the bias right to each molecule it can interact with high specificity. Likewise antigen antibody interactions right there is most of this immunological process. So, we can see the interaction between the antigen as well as this antibody.

So, for understand the inviting the bacteria and viruses fine, likewise in protein nucleic acid interactions right this protein nucleic acid interactions are important for the regulation of gene expression, a DNA replication and repair also transcription packaging so on right.

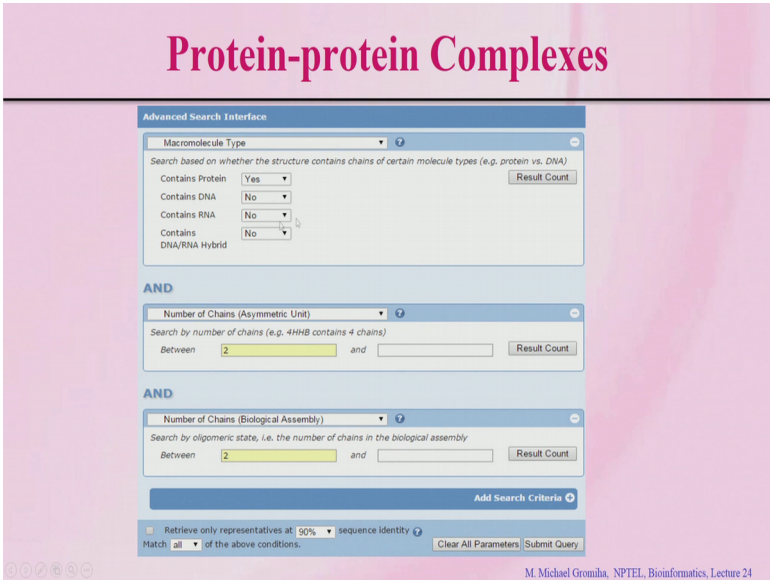
(Refer Slide Time: 04:46)



So, we look into these protein ligand interactions right this ligand interactions are also important, because these ligands trigger the affinity of this compounds, this can interact the active sites right and change the conformation the protein right and then inhibit several activities. So, the protein ligand interactions are important in structure based drug design right for identifying these lead compounds right are the proteins inhibitors for various targets which involved in different biological pathways as well as in different diseases.

So, now we have these different complexes you have the protein-protein interactions, protein DNA interactions, protein RNA interactions and protein ligand interactions. How to extract the information right for known structures right. So, these structures are deposited in protein data bank, right and how to obtain the structures right.

(Refer Slide Time: 05:41)



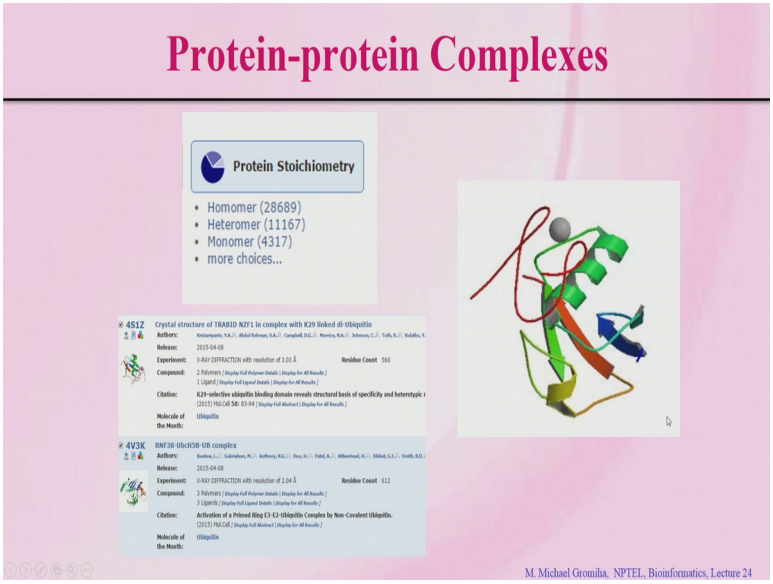
For example if you are interested in protein-protein complexes, which specifically on the heterodimers right; what is heterodimers?

Student: Different molecules.

Yeah there 2 chains which are different, the molecules are different rights either can the dimer or trimer and so on. So, in this case you go to these protein data bank and the advanced search options, and see this is contains protein right yes or no its yes because we are looking for the protein-protein complexes. Then contains DNA? no DNA right. So, RNA no RNA then the hybrid that is we do not need then it contains protein means you can get all these structures with proteins right. Sometimes right you can get the monomers and the dimers, and same proteins right dimers and trimers and so on.

So, now we to define how many chains do you need right some asymmetric units right, at least we need 2 right. Then you get see the number of chains based on biological assembly right. Here also we take 2 and then if you submit the query right then you will get the data.

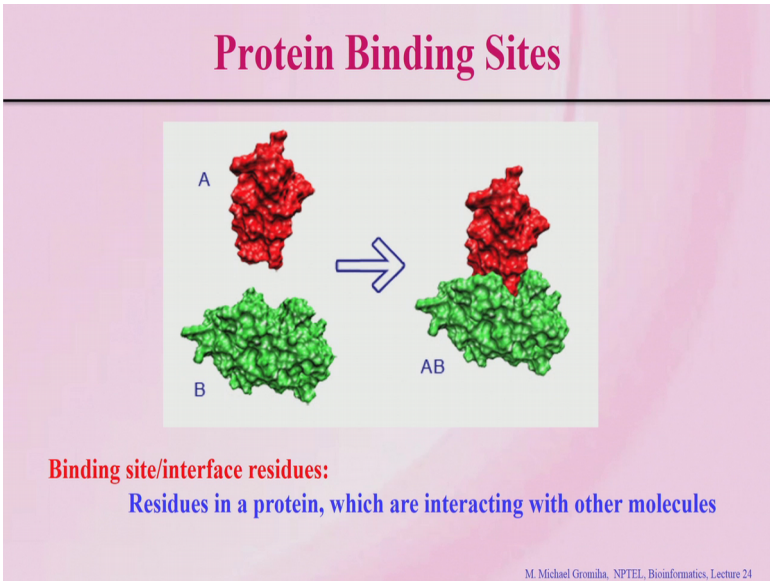
(Refer Slide Time: 06:48)



So, we can see this is a heteromer, this is a homomer right and you can see how many monomer and more choices.

So, then we will give the list right and then we can get the structures, that you can see the structures and you can carry out your analysis right; for example, if have protein-protein complexes right.

(Refer Slide Time: 07:03)



For example here it is A 7 protein and B is another protein. So, these proteins A and B, they form the complex rights AB complex and if you look at these structures you can see

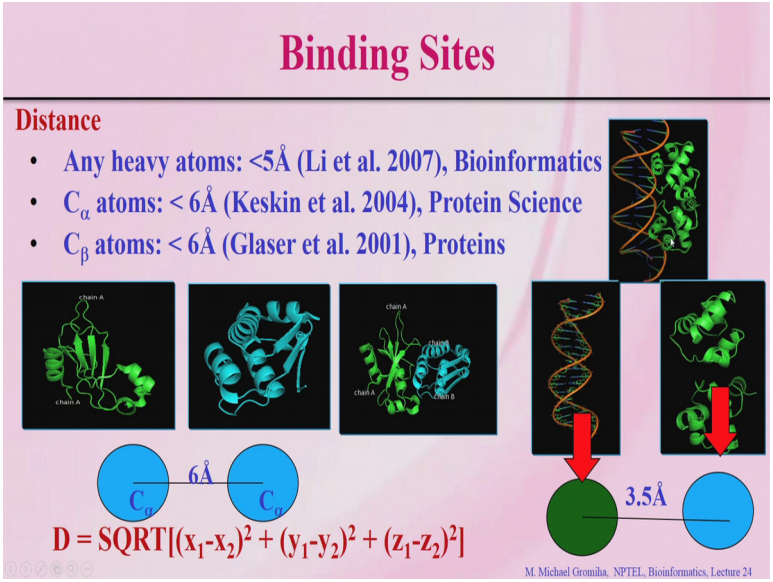
the some regions where these proteins can interact right for example this region right. So, you this can interact and this region right we called as binding sites right. So, then what is a binding site?

Student: Site where protein-protein.

Interact. There is residue in the protein where right these 2 proteins are interacting right, this is called the binding site or you can say interface residues right, because they are sharing the common interface between these 2 proteins.

So, now the question is if there are several complexes protein-protein complexes, protein DNA complexes, protein RNA complexes and so on. How to identify the binding sites, which residues are located at the interface area? So, there are various methods to identify the binding sites right. Major method is you can use the distance based criteria, and you can use the solvent accessibility based criteria, and also you can use energy based criteria right.

(Refer Slide Time: 08:19)



So, how to identify the binding sites based on distance right? For example, if you see this is the protein DNA complex right. Here I show one example is protein-protein complex right. How to identify the binding sites? If you look into these structures we can tell right because here you can see this is a region where the DNA can interact right. Here also you can see here this is the this area where the proteins can interact, but how we quantify

how to exact identify, which residues are interacting which residues not interacting right. While having the uniform representation of these binding sites residues right you can define the binding sites using distance based criteria.

For example either you can use the C alpha atoms or you can use C beta atoms or you can use any heavy atoms right, based on that we define the distance right between these 2 atoms. For example, if you take the C alpha atoms right now generally they use the distance of about 6 angstrom and the case of this C beta atoms here also 6 angstrom is used, and any heavy atoms that depends upon the type of the complex right you can change the distance right.

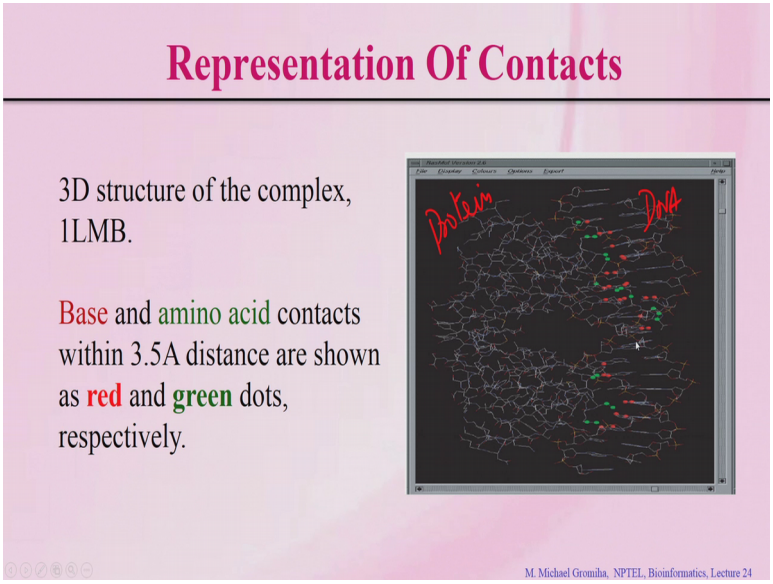
So, for example, if we say 5 angstrom right let us see any heavy atoms right which is a distance of 5 angstrom. You can use different criteria if you understand the binding site residues and the type of interactions, which one will be better which one gives the best data either to use C beta any heavy atoms now any heavy atoms. Because if you want to see the electrostatic interactions that is between the atoms right. So, the oxygen in the negative charged residues, and the NH2 group in the positive charged residues right.

If we take the C alpha – C alpha, right they may be far, but if we see the charged atoms these atoms are close to each other. So, in this case if uses the any heavy atoms right you can get the beta information, then using either C alpha atoms or C beta atoms right then how to define? So, for example, the any atom in a DNA and this protein, for mainly in this case or if in this case you can take any of these 2 proteins. So, first way is you can just have the complex and cut it into 2 pieces. Now you go 2 change either one is DNA is here and the proteins here, this case is a protein one and the protein 2 right.

Now you can calculate the distance between any at heavy atoms in the protein 1 as well as protein 2 and in this case DNA and the protein. So, we have the equation to calculate the distance right the square root of x1 minus x2 the whole square right plus y1 minus y2 the whole square plus z1 minus z2 the whole square right calculate distance.

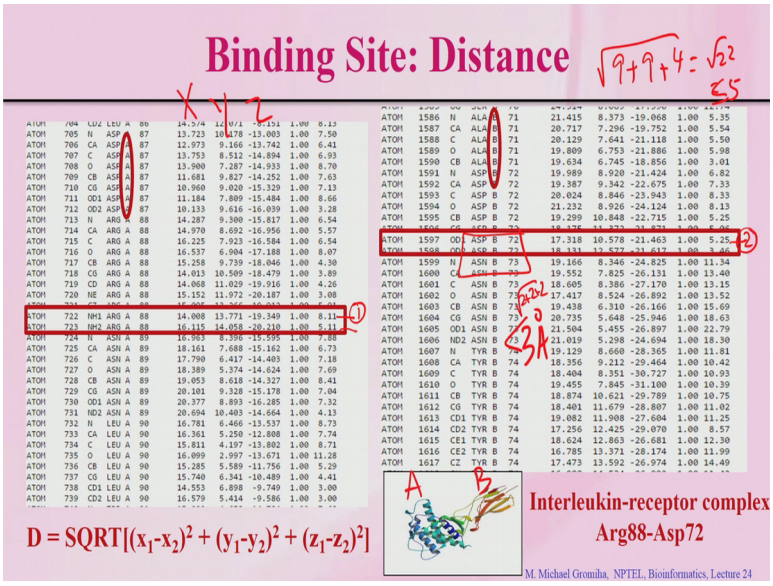
Then you keep the distance cutoff say 5 angstrom any heavy atoms or 6 angstrom any C alpha atoms right and if this is less than the distance, then we see that 2 residues are in contact and the 2 residues may be form the interface residues right. Because they share the common interface one residue from protein 1 another residue from protein 2, because they are very close right in space fine. How to do that?

(Refer Slide Time: 11:29)



This is one example for a protein DNA complex right. So, we see this is the DNA is here you can see the DNA here right and the protein is here right and you can see the interface residues right, you can see the residues marked in green, and you can see the contacts which I have marked in red right. So, the green one are the amino acid contacts from the protein side, and the red one are the base contacts from the DNA side right you can see the contacts fine.

(Refer Slide Time: 12:01)



So, now you show the PDB coordinates right. The PDB, which information get right we get the x, y, z coordinates, this is chain A and this is chain B right here if we see the complex this complex, interleukin receptor complex, we can say this is one is chain A this chain B for example, right you have the coordinates here for chain A this is the coordinates for the chain B right.

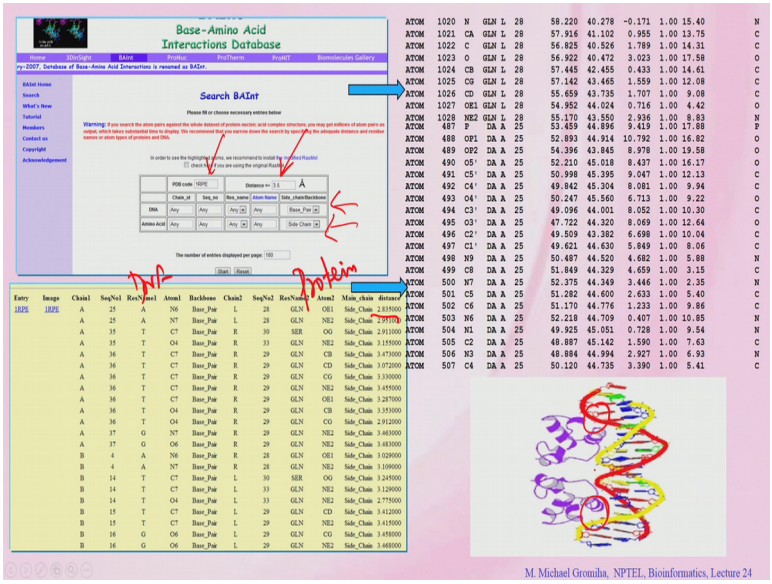
Now, if you see the atoms of any residue and calculate the distance with any of the heavy atoms in the second chain, and set the distance of for example, 5 angstrom and if this distance is less than 5 angstrom, then you can say the residues in a chain A and the residue in chain B are in contact and they are interacting with each other.

For example if you see this 88 arginine 88 here, this is the chain A right and here this is the aspartic acid 72 right in chain B we get the contacts right. you can see this should be within the specific cutoff distance for example, if we take this one NH1 of arginine and the OD1 of aspartic acid right. what is the distance? Now you take this and this. So, 14 and 17 this should be around 9 plus 19, 13 and 10 right this will be 3 this plus 9 plus 19 and 21 that is 2 that is around 4 right. So, this will be around 22. This is less than 5 right, fine.

So, you can say these residues are in contact with each other and specifically if you see this is the arginine. It's a positive charges residue with NH1 and NH2. It is a negative charge with OD1 or OD2 right. So, we are having this electrostatic interactions right especially if you see the second one the NH2 and the OD2, 16 and 18 is equal to 2 its very close 14 and 12.5, this is also less than 2 there is 4, 21 21 that is 1. So, there is very this is less than root 2 plus 2 plus 2 there is the root 6 right.

In this case it is less than or equal to 3 angstrom right it can form the strong end pairs right. So, if you see the distance between any 2 atoms, and if we less than any specific cutoff right. So, then we can see that this is less than, 3 angstrom. You see then this is when in contact right there also you can see whether they form any specific types of interactions right. You can look into this is an example fine.

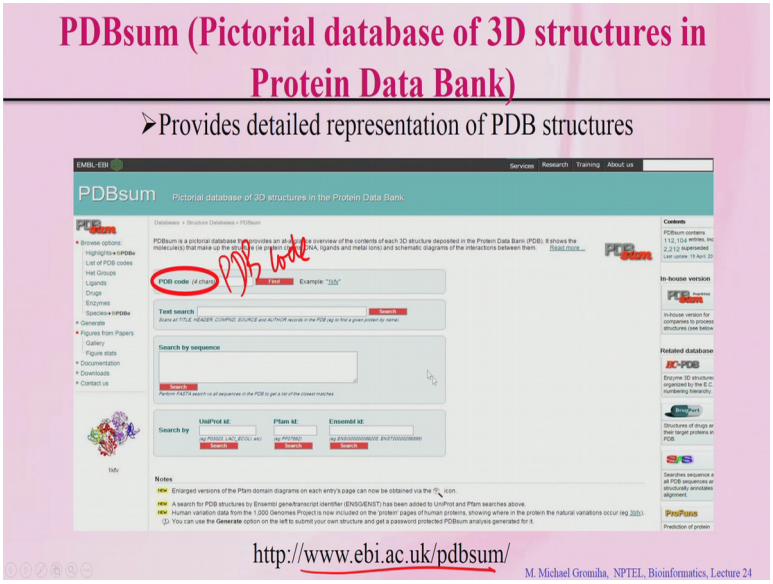
(Refer Slide Time: 15:00)



So, this is the server for the getting the base amino acid interaction database right. So, if you give any PDB id is you give the PDB code here, and we can the specify the distance, then it ask for the contacts whether we need the base contacts right or the backbone side chain contacts or any contacts right you can you can define from the DNA side which is the contacts you require right and the protein side which contacts like you like to see right then if you see anything then you can get the data right.

For example this is the PDB coordinates for this particular protein DNA complex right and here this is the sequence number 1, residue a, right this is the base pair this is again one this is the protein this is the DNA, you can see the distance right. So, in this case you can easily see these are the regions where this DNA can contact with the particular protein right this is identify the binding site fine.

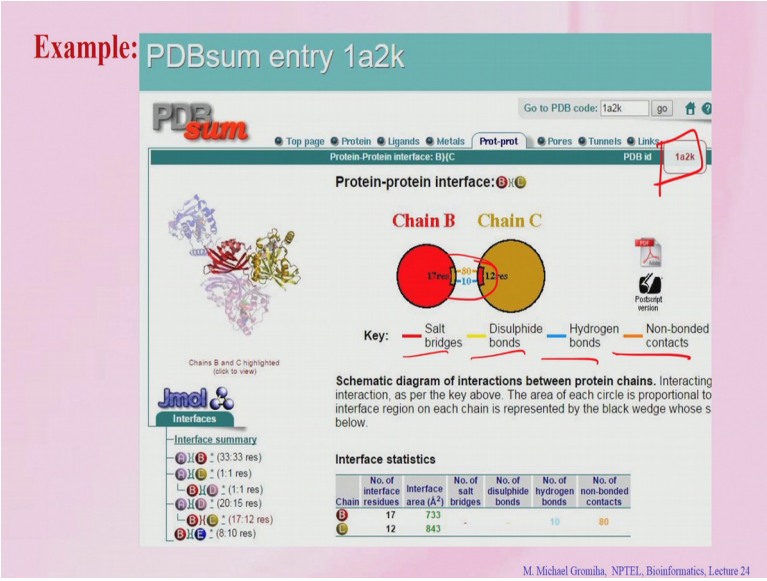
(Refer Slide Time: 16:04)



1. So, this is from the distance based contact right you can get from this a coordinates from the PDB, and you can get the distance based contacts. So, if you get the information on the PDB. So, right this is the additional information, which I include the PDB that is called PDBsum, this will give you the pictorial presentation of the 3D structures in protein data bank right. They give you several information regarding on secondary structure and the types of interactions right and so on.

So, if you go to the PDB sum you can see the contact between 2 residues, and which type of interactions these residues make where the electrostatic interaction or Van Der Waals interactions or hydrophobic interactions and so on right. For example, if you go to the PDB sum this is website for the PDBsum right go to the websites and then you can give the PDB id right if you can give any PDB id here right you can give the PDB code right and then if you click this submit button right then search then you will get this type of information.

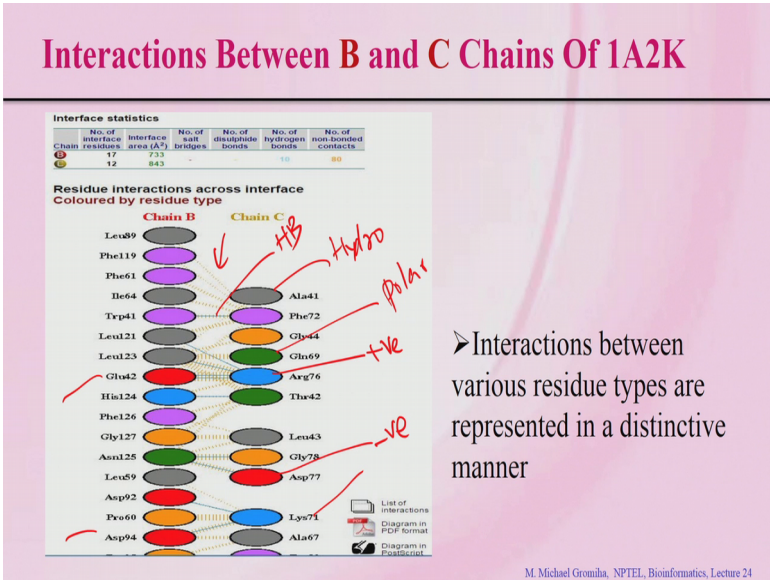
(Refer Slide Time: 17:03)



This is for 1A2K right and you can see this is how they repair the chains B and C and. So, the they these are the residues how many residues and how many contacts there is 17 residues in chain B, 12 residues in chain C right they are making about 80 plus 10 contacts right 80 non bonded contacts right. You can see non bonded contacts in yellow and the hydrogen bonds in blue and disulfide bonds this is also yellow right this is orange and the salt bridges in red right.

So, you can see here you have only the 80 non bonded contacts plus 10 hydrogen bonds, then if you go through the details we will get the more detailed information right.

(Refer Slide Time: 17:46)



You can see the here is a chain B of 1A2K this is chain C right. So, we are make a different types of contacts right they use different colors what is the meaning of different colors.

For example, this is a blue for the.

Student: Arginine.

This is arginine and lys this make the positive charge and the red for.

Student: Glutamic acid.

Glutamic acid aspartic acid this was the negative charge and the gray for.

Student: Hydrophobic.

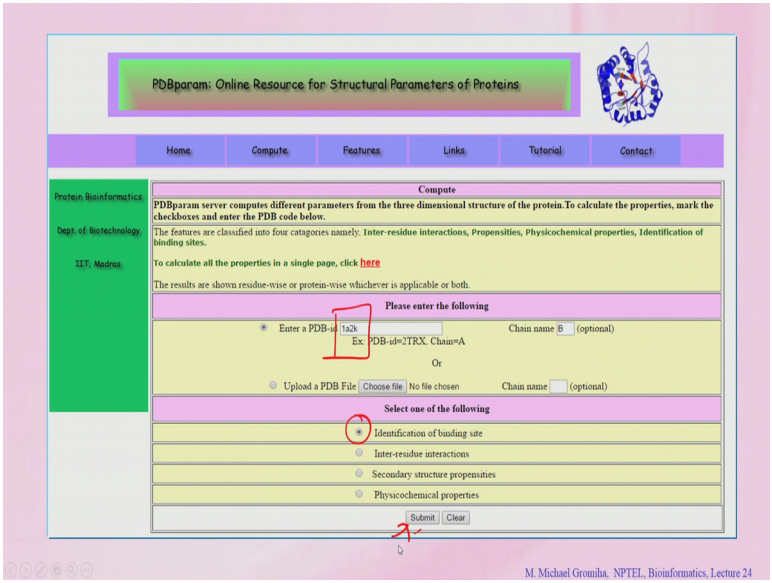
Hydrophobic right and the mainly that the green for the polar right now we can see the type of interactions here right. So, what is this green the blue one?

Student: Hydrogen bonds.

Hydrogen bond the blue one is for the hydrogen bonds and you can see the other contacts right this mainly the hydrophobic interactions right. In this case mainly the hydrophobic interactions as well as the hydrogen bonds. So, if you look at the PDBsum right. So, it

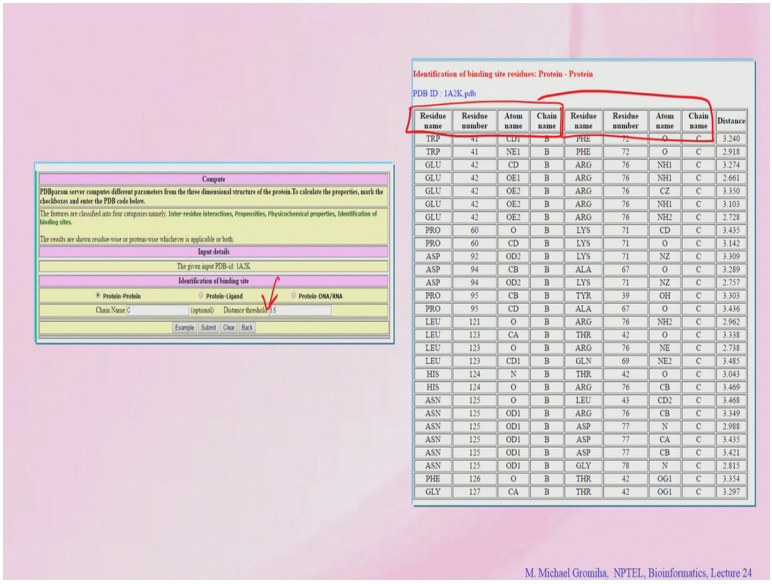
will identify the probable contacts and based on that classify the contacts in different types of interactions either electrostatic interactions or the hydrogen bonds right or non-bonded contacts or salt bridges right for any case you can see that fine.

(Refer Slide Time: 19:11)



So, now I will see earlier discussed about the PDBparam, it contains more than 50 structure, based parameters we utilize the PDBparam to get the surrounding hydrophobicity long range order right and different other structure based parameters. You can use the same for identifying the bindings sites also right for example, if you go to the PDBparam and you want to get the binding sites of a particular protein, you go to the website and then just click on identification of binding site right. So, and then submit right.

(Refer Slide Time: 19:47)

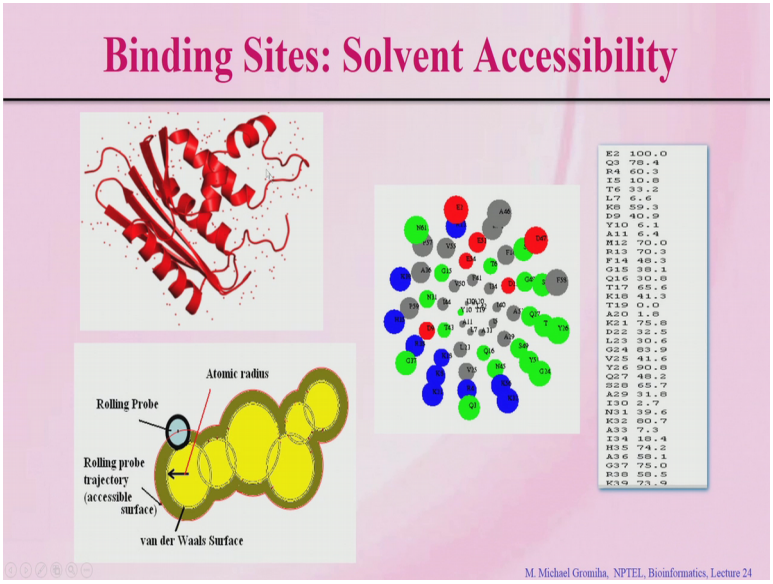


Then this will ask you for the distance threshold, which distance a threshold a you want to calculate a the identify the binding site.

So, if we see this is a 3.5 angstrom right. So, we will get the get the interactions like the residue name. This residue for number 41, this atom name CD 1 and chain B this is from the one chain right and this is from the another chain right. So, phenylalanine is 72 right the chain C if you have the distance of 3.2, 3 angstrom right the same 1A2K right the trictrophan and phenylalanine right they are within the distance of 3.2 angstrom right they are having the interaction.

The same information you can also get from this a PDBsum, they are making the hydrophobic contacts right you can get from this information right fine. So, in this case you can also get all the contacts, from this contacts you can also identify which type of interactions they are possibly to make right between these 2 chains for example, chain B as well as the chain C, this is one concept. So, we used the concept of distance right distance based criteria to identify the binding sites right you can use any distance criteria.

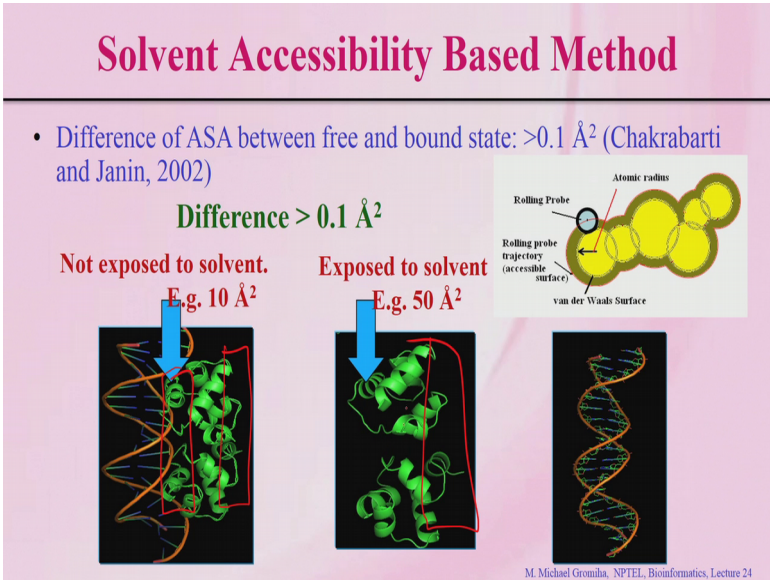
(Refer Slide Time: 21:15)



So, next is next concept its mainly based on solvent accessibility right what is solvent accessibility? They accessible to solvent for example, this is a protein right. So, you can make these residues and atoms in interlocking spheres, this is solvent molecule. So, you can role over the residue residues and the atoms right I can see how far this can penetrate in the Van der waals surface right. So, based on that we identify the residues which are the surface and some residues which are in the core right. These are the values right for each a residues fine already we discussed in detail in the earlier classes right.

So, how to use the concept of solvent accessibility to identify the binding sites right.

(Refer Slide Time: 21:51)



And for example, here I show the same figure again. So, this is the complex the which complex this one? Protein DNA complex right. So, we cut into 2 pieces right here we are not taking the free protein or the free DNA, take the complex right and we cut it into 2 pieces. This is the protein part the same right and then this is a DNA part.

Now, we can calculate the accessible surface area right in the complex states or in the unbound state right. You cannot see the free state we can say this it the bound this you can see the unbound, but same a structure right. So, if we look into these 2 proteins right if you take some region for example, this region right how about the accessible surface area will vary with now is you know varies right, because this is accessible to solvent here also accessible to solvent right in this case you do not see any change right.

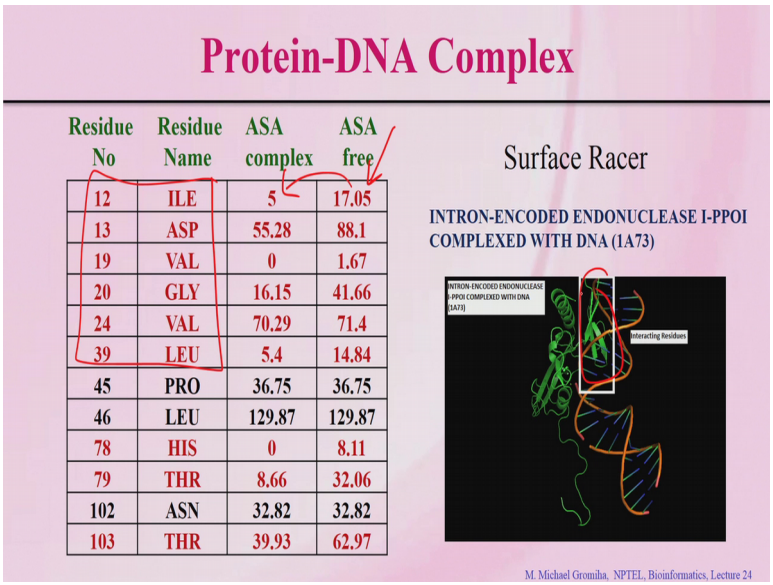
If we take this region right then you can see this is the complex state, this is the unbound state which one you have the higher value, unbound and in this case it is accessible to solvent. So, this is have more have accessible surface area than in the complex area because here it is hindered by this DNA, because this region right is held by the DNA this is in the hence the area will be less right.

So, if you compare the solvent accessibilities with the complex, and the unbound state then if there is a difference in accessibility then what can we say.

Student: (Refer Time: 23:26).

Yeah there is because of the reduce reduction in the accessibility, its mainly because of something else right. In this case DNA because of this process of DNA they reduce accessibility right from the unbound state to the unbound state this is the unbound state fine.

(Refer Slide Time: 23:44)

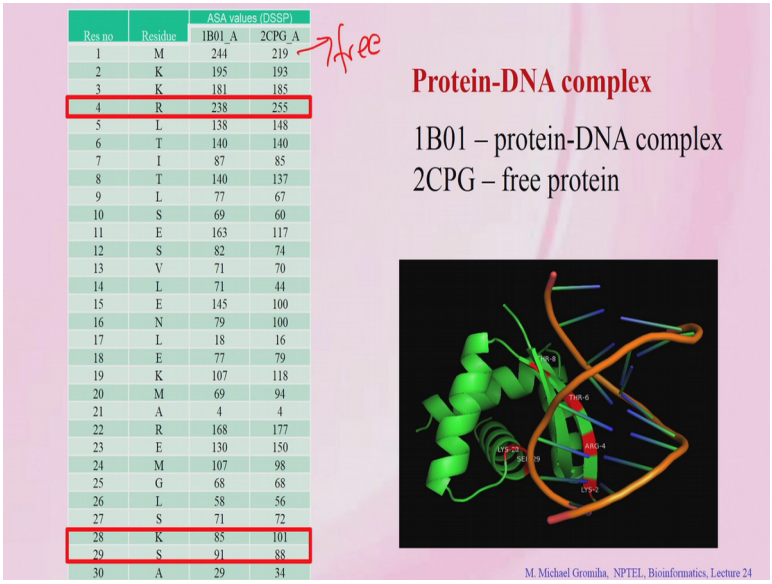


How to do this now I show one example, here is this is the protein DNA chain this is the place where we can see the interacting residues right.

For example is we see these protein this is the ASA of this complex right as well as in 12 5 angstroms square and ASA is 55 and so on. Now this is with the free form this unbound form right. So, in the case of protein 45 right any find any difference no difference Leucine 46 no difference, but the other hand if you see other residues like isoleucine 12, aspartic acid thirteen and 19 valine, 20 glycine.

So, this is the unbound form and we make the complex, you can see the difference right then if this is the difference right and then we can see that these residues are interacting right in the bound the complex. In this case what are the residues which are interactive with DNA isolation 12 these are the residues which are marked in red, these residues we can find in this region right they should write me the DNA fine.

(Refer Slide Time: 24:54)

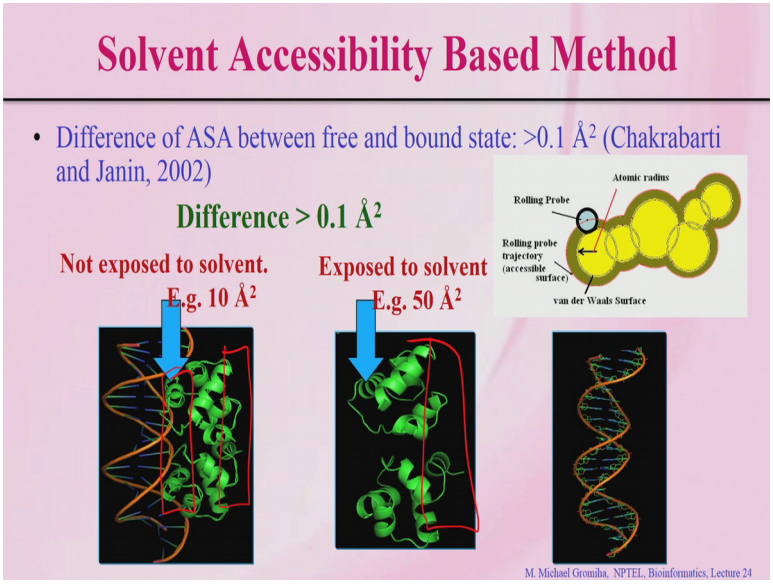


Res no	Residue	1B01_A	2CPG_A
1	M	244	219
2	K	195	193
3	K	181	185
4	R	238	255
5	L	138	148
6	T	140	140
7	I	87	85
8	T	140	137
9	L	77	67
10	S	69	60
11	E	163	117
12	S	82	74
13	V	71	70
14	L	71	44
15	E	145	100
16	N	79	100
17	L	18	16
18	E	77	79
19	K	107	118
20	M	69	94
21	A	4	4
22	R	168	177
23	E	130	150
24	M	107	98
25	G	68	68
26	L	58	56
27	S	71	72
28	K	85	101
29	S	91	88
30	A	29	34

I show one more example right in this case I put 1B01 this is the complex, and here this is the free protein this is the free protein right and we see there are several cases right for example, this is arginine they reduce to 258 to 238 right and some cases it is also possible that the complex case right that you have more values than the free one how this is possible. So, here I like to mention the difference between the unbound and the free one right. In the unbound case we tell you that is a complex, the complex we cut into pieces and take the protein.

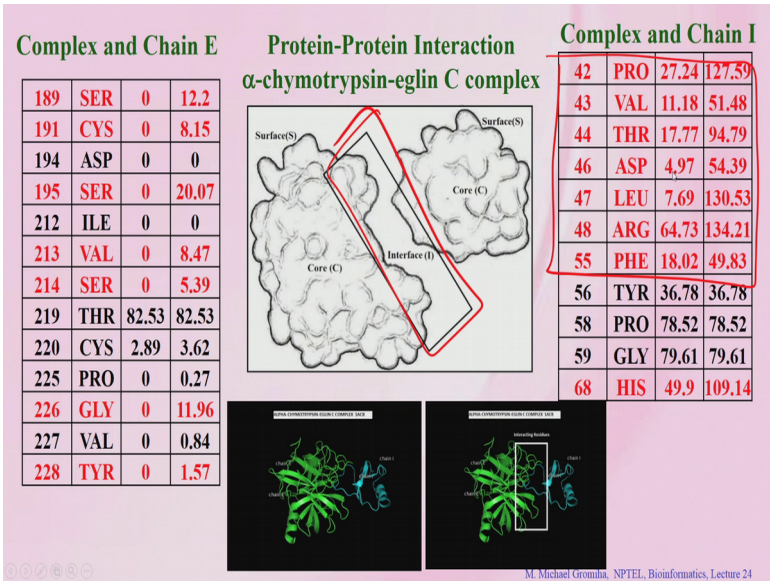
In this case if we superimpose this protein as well as to one in the complex their image exactly match right there is no difference in the conformational change. But in the free protein what we do is we take the amino acid sequence and check for the structures without any ligands without any DNA. In this case you can see the change in conformation because of change in conformation you can see the difference in accessible surface area right sometimes it is higher in the case of the complex than in the case of this free protein.

(Refer Slide Time: 26:11)



Then I will show one example to the protein-protein interaction complexes right this is the alpha chymotrypsin eglin C complex right.

(Refer Slide Time: 26:16)

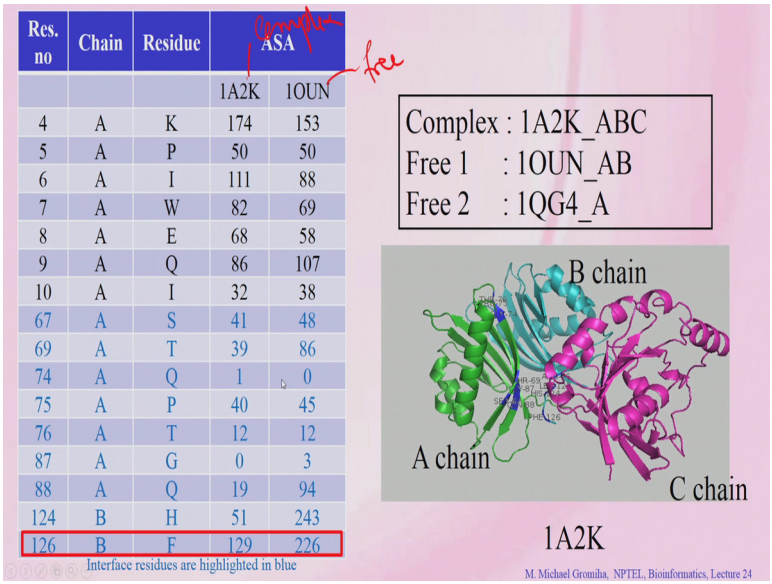


So, this is the protein 1 this is the protein 2 right with the complex and the chain E right where this is the ASP with a complex is a chain E right and if we see there are many cases only in this region interface region is easily you can see the lock and key mechanism right is you can see this is the convex and concave surface here right . So, 12.2 reduced to 0, because it is completely inaccessible right here the 8.15 is completely

0. So, in this case these residues are identified as binding site residues or the interface residues. Also if we in this chain I continuously a stretch 42 to 48.

All these residues right they change the accessible surface area very very drastic change where on127to 27, 54 to 4 and so on. So, these residues we identifying the binding sites right this is also you can do in the case of the protein a protein complexes.

(Refer Slide Time: 27:17)



This is another example the same I explained earlier. So, this is the complex 1A2K this is the complex, and this is the free protein right. Some case if you see the free protein higher than these complex and also vice versa. So, we can see in some other cases right. So, if it is reduce to the complex right then you can see this residues or identified as the binding site residues fine.

So, we discuss about 2 different approaches right what are 2 different approaches we discussed distance based criteria.

Student: ASA based criteria.

And ASA based criteria distance based criteria how to identify the binding sites.

Simply based by the cut off distance and can identify within that limit, the ASA how to identify the binding sites? Change in the accessible surface area right between the unbound form as well as the complex form.