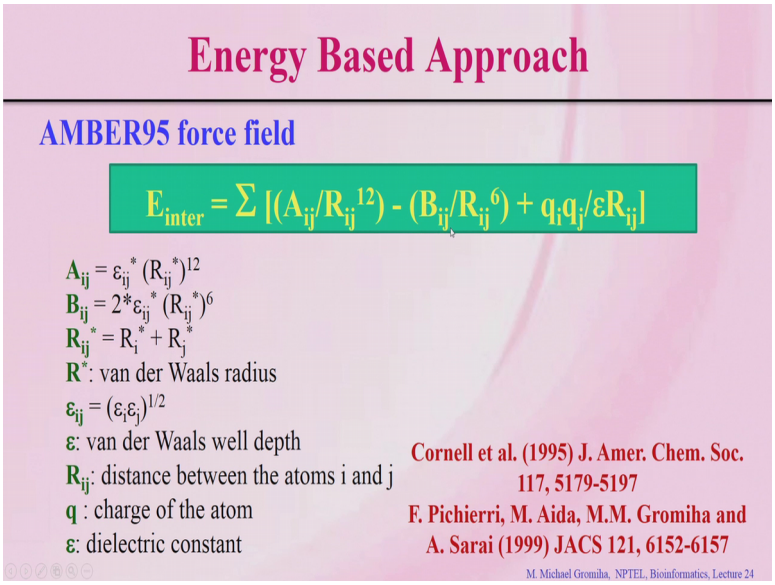
Bionformatics
Prof. M. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology, Madras

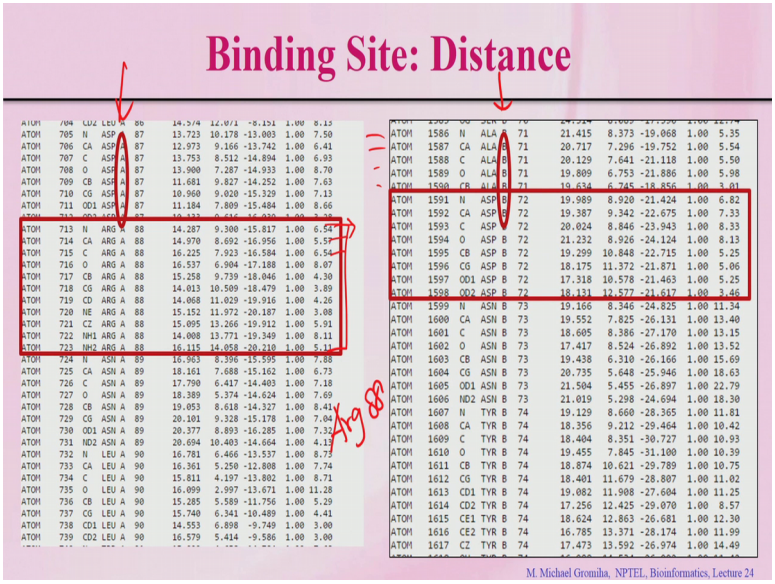
Lecture - 24b
Protein Interactions II

(Refer Slide Time: 00:17)



So this is another method. So, here they use the energy based criteria. So, mainly they calculate the van der waals energy and electrostatic energy, then using these interactions you can identify the binding sites how do you identify the binding sites.

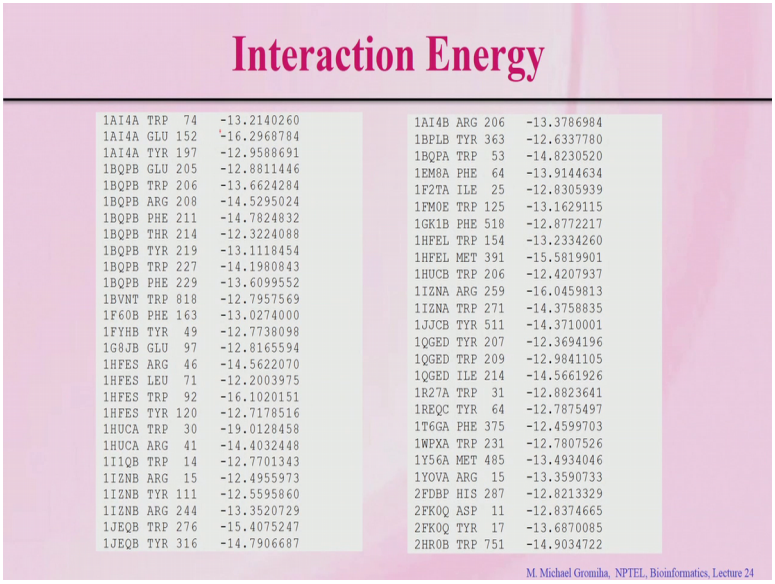
(Refer Slide Time: 00:32)



For example, if you have these PDB coordinates in all the cases we use the structures. In this case you can take these binding sites as a kind of experimental data because we are not doing experiments directly we have to identify these binding sites or binding affinity, but we use the known 3D structures and use this criteria to identify the binding sites. fine.

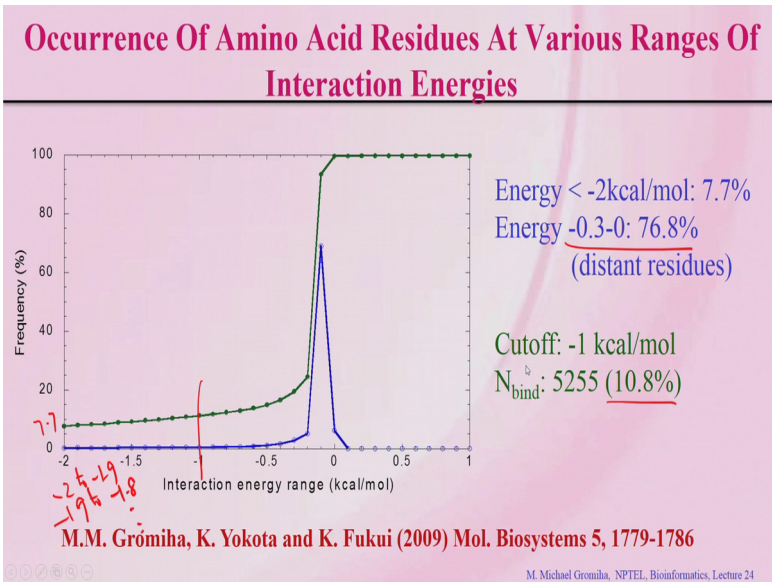
So, how to identify the binding sites using energy based criteria. So, here this is a chain A and chain B for example, if you take the arginine 88 right. So, there are different atoms in arginine. So, for each atom you calculate the energy with all atoms in the partner chain for all atoms in the partner chain. So, then you do with the second atom and the third atom. So, you calculate the energy then sum up all the energies in all the atoms in the arginine then you sum up the values that will give you the energy of these arginine 88. If you want to take the arginine 88, take all the atoms with each atom in chain A we calculate the energy with all atoms in chain B do it for all the atoms for example, N CA C O and so on. Then we sum up the values that will give you the energy of arginine 88 in chain A. Likewise you can do for all the residues for example, you have 100 residues. So, you get 100 energy values. From these energy values we can identify the binding sites based on the energy right.

(Refer Slide Time: 02:15)



So, how to do that? This is one example now for example, the 1AI4A tryptophan 74 has minus 13.2 kilo cal per mole right. Likewise phenylalanine 211 has 14.78 kilo cal per mole fine. So, now, from these energy values you can see which residues are high preference to interact with the binding with a partner.

(Refer Slide Time: 02:38)



So, we have the energy values we divide the different bins for example, minus 2 to minus 1.9, minus 1.9 1.8, see each bins you can see the frequency of occurrence of the residues with our particular energy. For example, if you see minus 2 to minus 1.9, minus 1.9 to minus 1.8. So,

you can go with up to +1 right.

So, here you have two different colors, one is in blue and one is in green. The blue one used the data for each bin and the green one with the cumulative values because we can go even less than minus 2 here I show the value many values less than minus 2 right. So, these value if it is less than minus 2 you can see that is around 7.7 percent because the bins are small as this where did not a broad that.

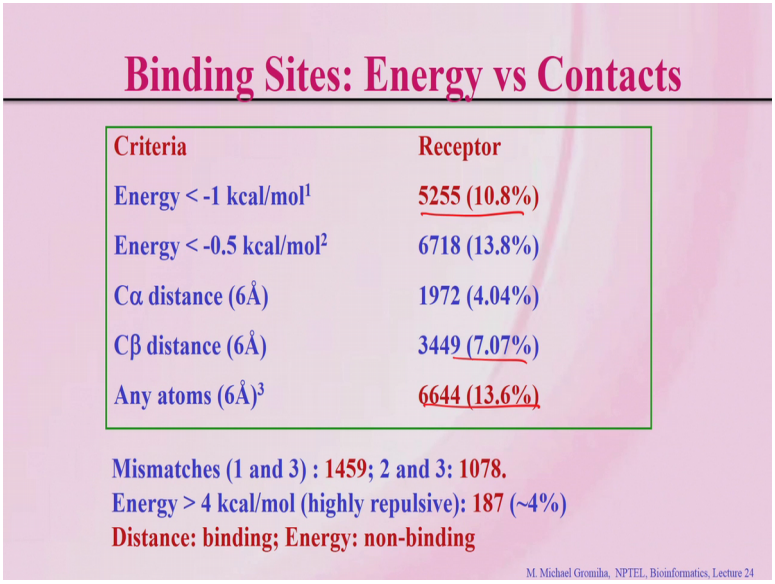
So, finally, if you see minus 2 there very less number then up to minus 0.5 you can see the rise and there is a peak and then it is going down. So, if you take a cutoff of 1 kilocal per mole for example, take this cutoff then how many residue that is what is the procedure of residues is identify the binding sites. If you do this there is about 10.8 percent of residues which identified as binding site residues if this is reasonable there is relevant to the report and literature and another case if you see there is a peak. For example, take the energy of minus 0.320, it is 77 percent. Why this is a peak here? Because a energy values is very less it is not significant, so in this case if the two atoms or the two residues which are far away in the structures they are not making any contacts. They are not making any interactions that is the reason why the values are around 0, that is about 77 percent of the residues they are not in contact with each other fine.

So, here we get the data based on the energy; when you compare the data with the distance because energy and distance are almost similar to each other because energy values are computed with two types of interactions. What are the interactions we use?

Student: (Refer Time: 04:57).

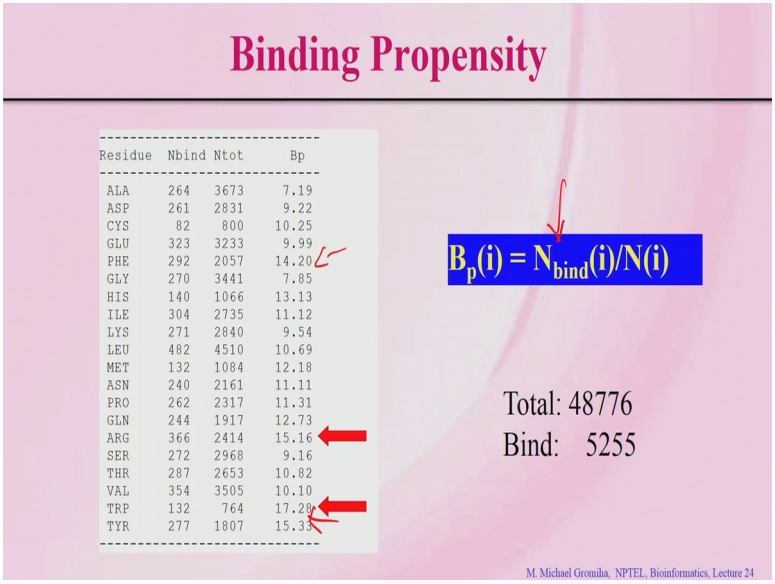
Coulombs law as well as this van der waals both are based on distance right. So, somewhat advantages some disadvantages right.

(Refer Slide Time: 05:09)



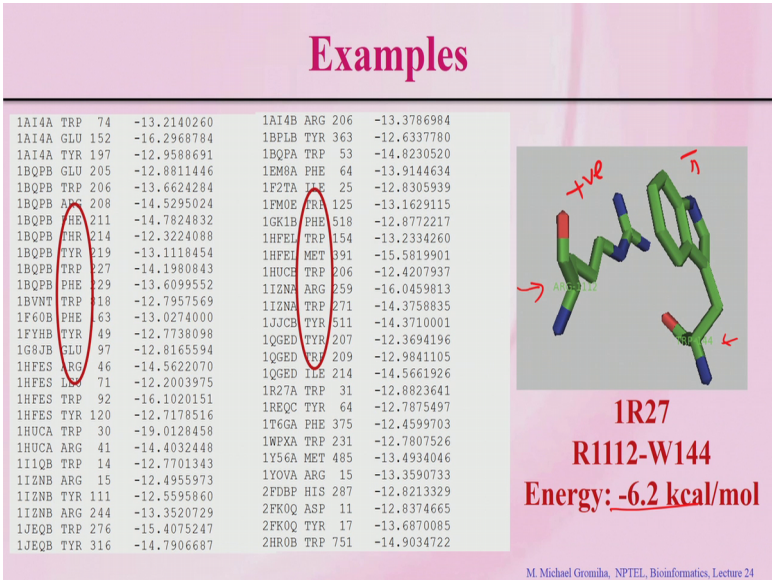
Now, if you see the energy you can say about 10.8 percent of residues which are identify binding sites and the distance we use different types of distances. We can say, if you use the 6 angstrom with the any atoms we can say with 13.6, if it is C alpha it is 4 and the C beta it is 7 right. So, (Refer Time: 05:24) 1 kilocal per mole it is in between this C beta distance as well as with any atoms. And if you see there will many with mismatches about more than 80 percent, 90 percent you can see the matches and some which mismatches mainly because of the some of the repulsion interactions are also taken into account in this type of energy based criteria, that is the reason right.

(Refer Slide Time: 05:46)



Now, I give you an example how to get these interactions right.

(Refer Slide Time: 05:48)

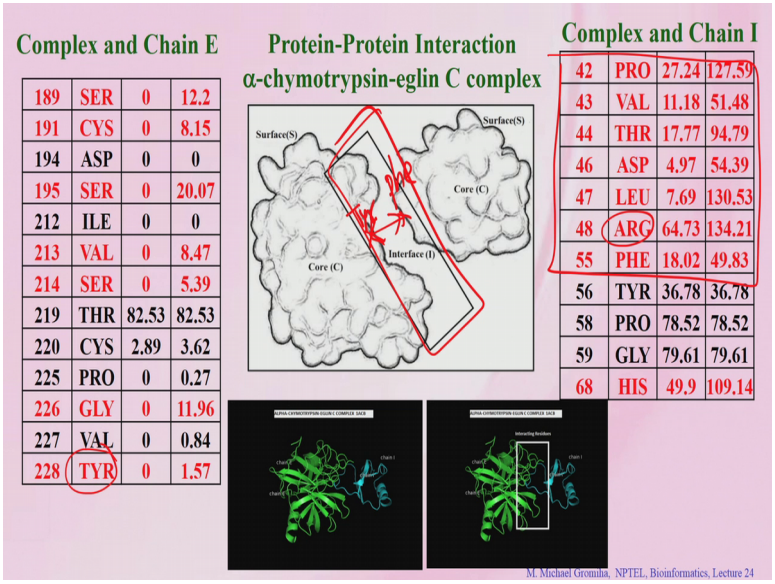


So, here these are the values the energy values some residues the energy values and if you look into these structures for example ratio one example in 1R27, this arginine 1112 and this is a tryptophan 144. These two residues they try to interact, they are interacting with an energy of minus 6.2 kilocal per mole this is the positive charge and this is the aromatic ring right. So, they are making a sort of interactions. Which interaction they make?

Student: Cation pi.

Cation-pi interaction, say this is the pi system this is a positive charge. So, they make a cutoff an interactions. Then the question is there are some specific interactions are there any preference of amino acid residues to form these type of interactions right. So, in this case I show the propensity of different residues to be involved at the binding defix. So, in this case we just calculated the binding propensity of this residue "i" using the number of residues of type i in the binding site with normalized with the number of residues in the whole protein. So, what is n bind of i. Right for example, if you see show this figure I show this figure right.

(Refer Slide Time: 07:14)



How many residues of type I are the interface these the interface residues for example, alanine, how many alanine in the interface compared with the all alanines the particular protein right.

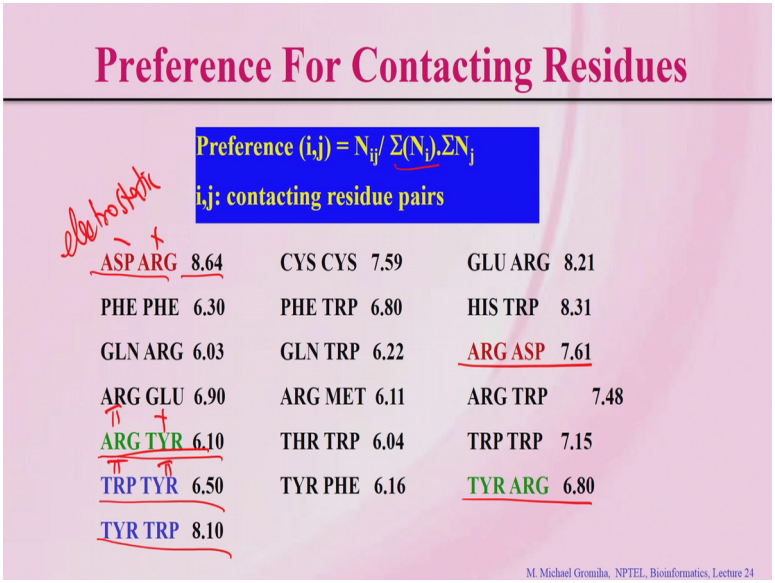
So, if we do this you can see some preferences for example, the arginine which has value of 15.1 and the tryptophan 17.28 and the tyrosine 15.33 you can see phenylalanine also very high right. So, you can see the preference of some specific residues. From these residues we can see that there are some sort of interactions they are preferred to be at interface due to type of this residues right.

Now, the question is here we have only one residue propensity how can we estimate how can we identify the interacting partner to identify the type of interactions. In this case what do you need we need the partner because the binding propensity will tell you which residues are preferred to be interface like arginine, tryptophan and tyrosine to understand the type of interactions we need the partner, for example, tyrosine or arginine trying to interact to with which type the residues. In this case you take these residues in these complex for example, this residue and this residue interact, this is which type of residue and this is which type of residue.

For example, this kept to be the phenylalanine, here you can see a tyrosine, then you can see both are in the ring system they can be pi interactions or here you can see the arginine here and here you can see the tyrosine here, then you can say which type of interactions they can

make.

(Refer Slide Time: 09:19)



So, in this case do we need the pairs how to identify the pairs right. So, you can see preference of i comma j using this equation nij this a number of types of i paired with j, i from the protein one and j from the protein two in the partner and N i is a number of types residue i and j is the number of residues of type j right.

When you look into the binding partners you can see some preferences for example, you see aspartic acid arginine you can the highest value of 8.64 then you can see the a arginine and tyrosine 6.1, likewise a tryptophan and tyrosine 6.5 tyrosine tryptophan 8.1. So, interestingly if you see these numbers you can find some sort of interactions are important. For example, if we take aspartic acid and arginine, which stable interactions are important? Electrostatic interaction this is minus, this is plus.

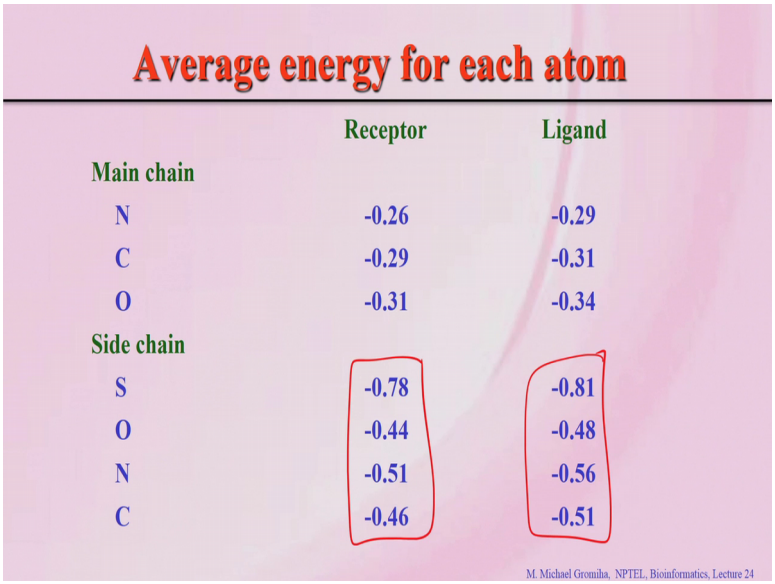
Also its does not depend on the type of the protein where take one or two even if you take the other way around that is also very high arginine aseptic acid. If you see arginine and tyrosine this is a ring and here this is the positive charge and if you see in other way around you can see tyrosine arginine also having high preference then tryptophan tyrosine this is also pi and this also pi we can see the pi pi interactions with a aromatic aromatic interactions

Now, the question is you identify the residues which are at the interface and you identify the pairs when the partner proteins protein one and protein two right. So, now, the question is

whether these interactions are specific are not specific, how can you estimate whether this residues from electrostatic interactions or not right.

Right, then you need to see whether the contribute energetic contributions are from the main chain or from the side chain. If the contribution mainly from the main chain then you can this is not specific if that is from the side chain then they are specific fine.

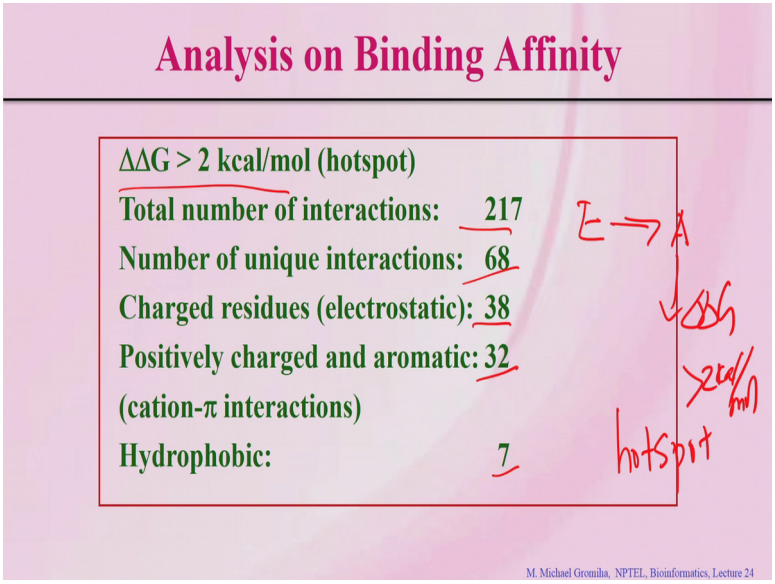
(Refer Slide Time: 11:41)



So, in this case we can see the contribution for the main chain atoms and the side chain atoms and we compare the values in the main chain and side chain you can see side chain values are at least twice then that of these main chain atoms both in the case of the protein 1 or the protein 2. Here I define ligand as the smaller portion of this complex and receptor as a bigger part of this complex right.

So, whatever this you consider you can see this side chain contribution is higher than that of these main chain atoms. This case you can see there is these interactions are very specific. Now, we need to experimentally verify whether these residues they form typically interactions or not fine. So, in this case we need to identify the residues which are serving as the hotspot residues right.

(Refer Slide Time: 12:37)



So, in this can you put a cut off of 2 kilocal per mole, we mutate any specific residue. So, what is mutation?

Student: Change of residues.

Change of residues for example, if you change any specific residues for example, you change a glutamic acid to alanine. If you change as the binding affinity of more than 2 kilocal per mole then we can say these residues as hot spot residues right. So, then we scan the literature many data available in the literature and from these data we can find about 217 types of interactions right; that means, they can reduce the binding affinity at least 2 kilocal per mole.

So, there are many mutants are common. So, we take the unique interactions you will get 68 interactions and we look into the residues which residues are mutated for this drastic change. Most of them are charged residues, about 38 residues are charged residues and if you take the positive charge and aromatic they are 32 mutations and the hydrophobics are only 7. This will again confirm that what are the interactions we observed from these pairs aspartic acid, arginine electrostatic cut and pi and the aromatic aromatic interactions these residues are important to form in typical type of interactions right.

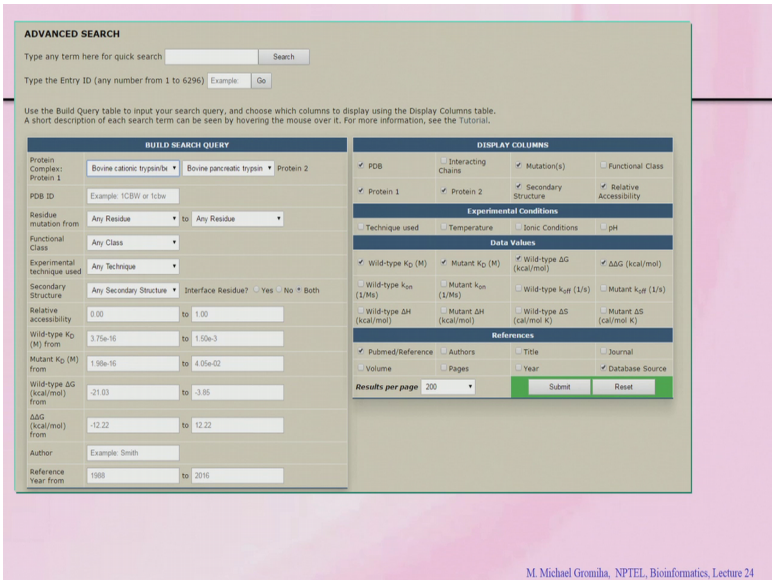
So, to get more information regarding these type of interactions recently we developed database this is called the Proximate database. This will give you the binding affinity of different mutants because currently it has more than 6000 mutants and we will use this

information available in the database to understand which type of interactions are important at different locations. Here I general generalize with a type of interaction and the specific pairs right.

(Refer Slide Time: 14:38)

I will explain little bit about the database and this contains the binding affinity of protein-protein complexes along with other different types of structural information as well as the literature information right.

(Refer Slide Time: 14:52)



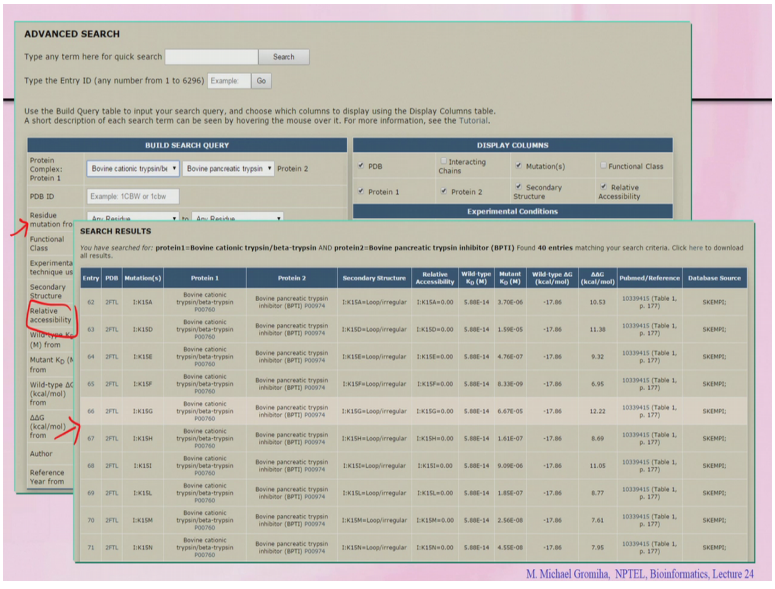
You have a search option here. So, we use the different information you can search based on

the proteins if you know the complexes protein 1, protein 2, you can get all the data for this particular complex or if you want to get the hot spot residues you can see the a mutant K D or mutant delta G, we have the delta G give the range and you can see mutants which are serving as hotspots.

So, here we can give the mutation type either you can use from any residues to any residues you can take the information. And you can search with experimental techniques and specific secondary structure helix or strand and also you can see whether you need only the interface residues. Mainly if you see the binding affinities affected by the interface residues and hence most of data you can see they replaced only at the interface region fine.

So, you can get the hotspot residues giving the value of 2 kilocal per mole delta G values and get the hot spot residues and you can also analyze whether this is secondary structure or how it effect in the different functional class or how it effect with the location of residues for example, you can see accessible surface area and so on right.

(Refer Slide Time: 16:15)

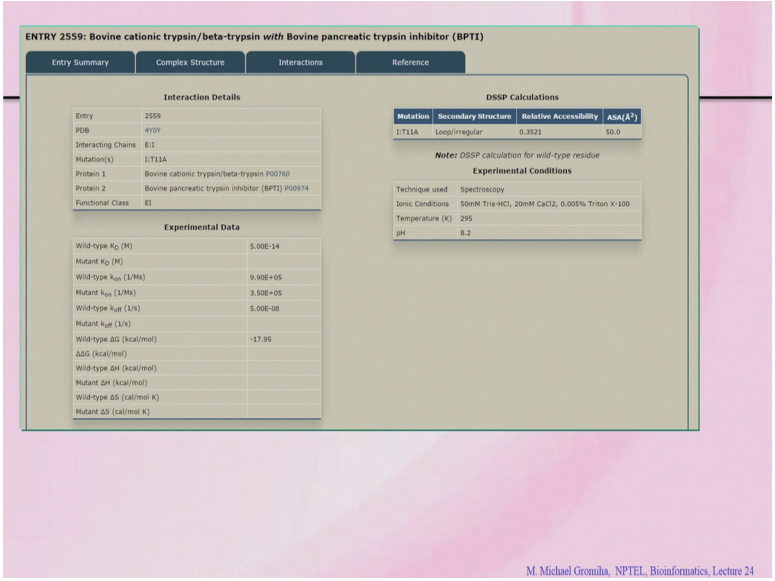


Now, this is the result for example, if you put the data bovine cationic trypsin and this protein to right. So, you can this see inhibitor. So, you can say this is a trypsin this in trypsin inhibitor then get all the data.

So, this is a secondary structure is looped, accessible surface area is completely buried and you can see the wild type K D and mutant K D and have the delta G values you can see all

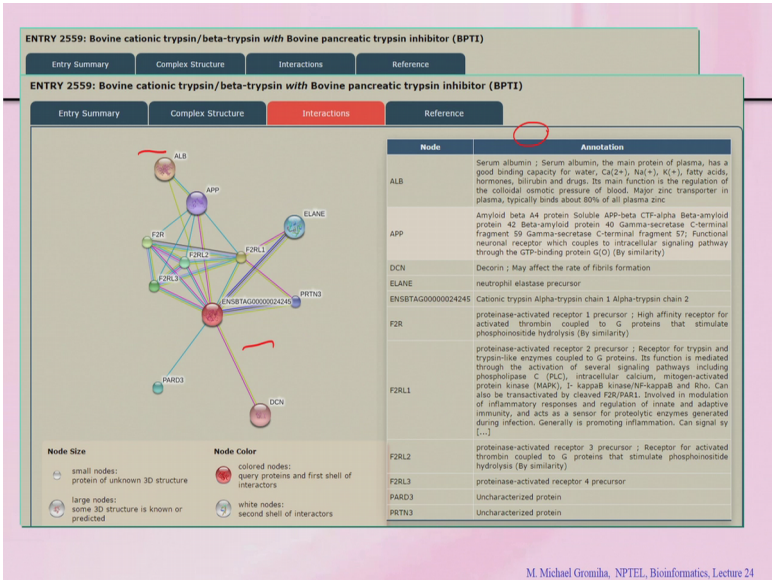
the values. Then we give the pubmed reference so that you can link to the pubmed what is the pubmed? literature database and here you also a given the exact wire you can get the data table one page 177 you will get that exactly the same data in literature. This will help the users to verify the data as well as to get the more information regarding any particular complex fine.

(Refer Slide Time: 17:06)



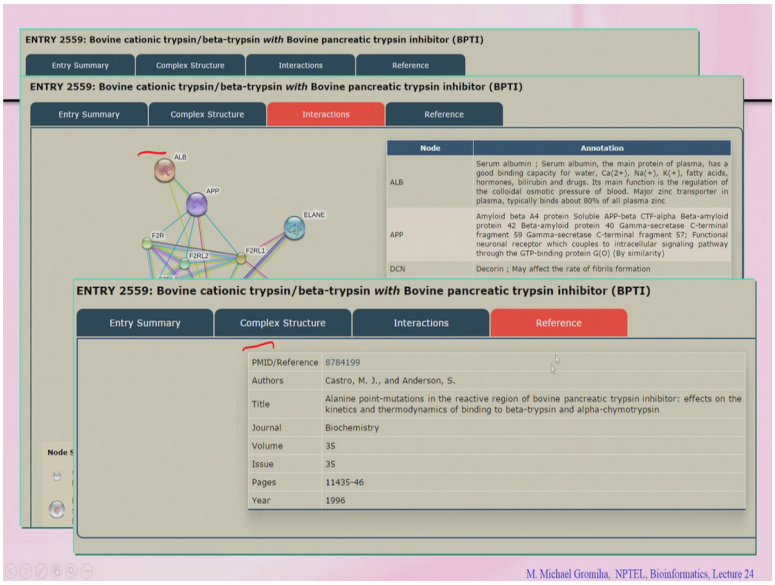
So, now we give the more details about the interactions here there are two chains chain E and chain I, these two residues are making the making the complex these equation in enzyme inhibitor complex and here is the data is a value of K D and you have the mutant values and delta G is minus 17.95 kilocal per mole. This is the loop region and the relative accessible surface area is 0.35 that is 35 percent the area is 50 angstrom square that is fine.

(Refer Slide Time: 17:42)



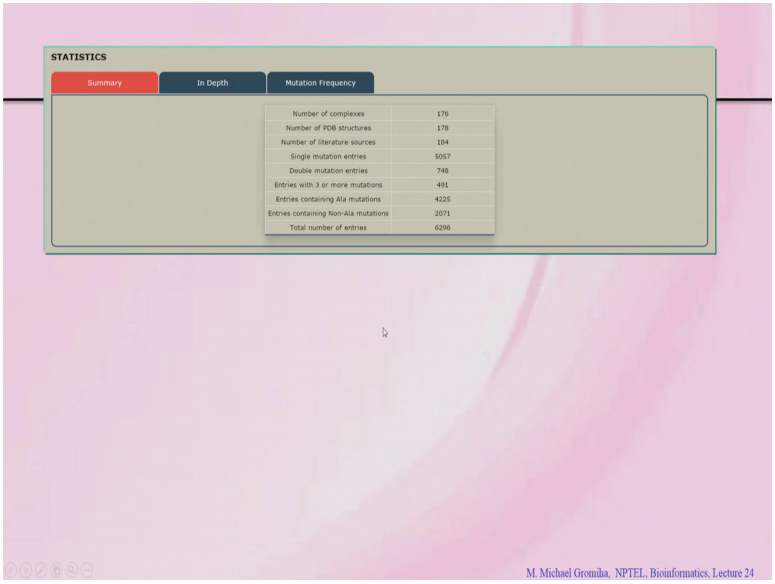
So, when you go to the details you can get the summary and the complex structure and the interactions as well as the reference. The interactions if this is your protein right. So, now, you can have what other proteins they are interacting with the particular protein. So, you can have the value from the string database right. So, and have the all the annotations and what the residues which are interacting with the particular proteins and so on.

(Refer Slide Time: 18:08)



So, you get is references these are the authors and the title as well as the details of the particular paper fine.

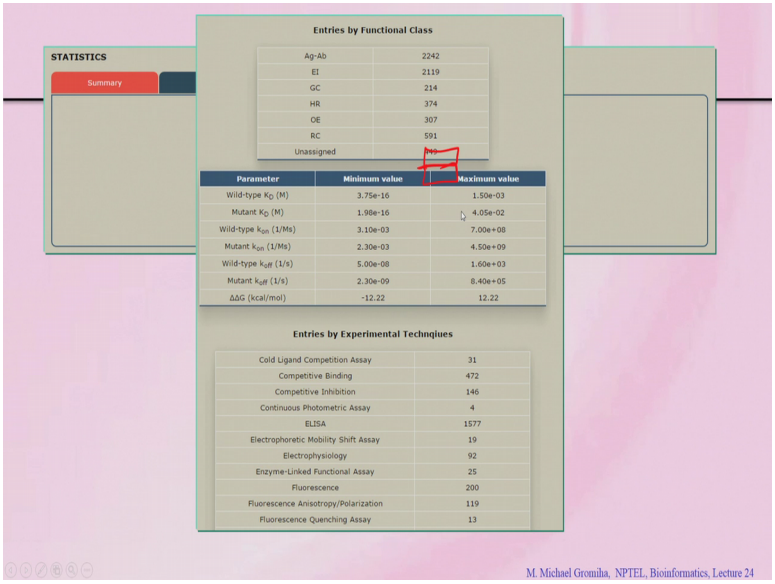
(Refer Slide Time: 18:17)



So, now I give the statistics currently we have 176 complexes and the literature resource is about 184 and there are more than 5000 single mutation entries. So, these entries are sufficient to do any analysis for example, if you want to analyze hot spot residues or you want the analyze the residues which will not make any changes in the binding affinity or you can also relate to the binding affinity with other features.

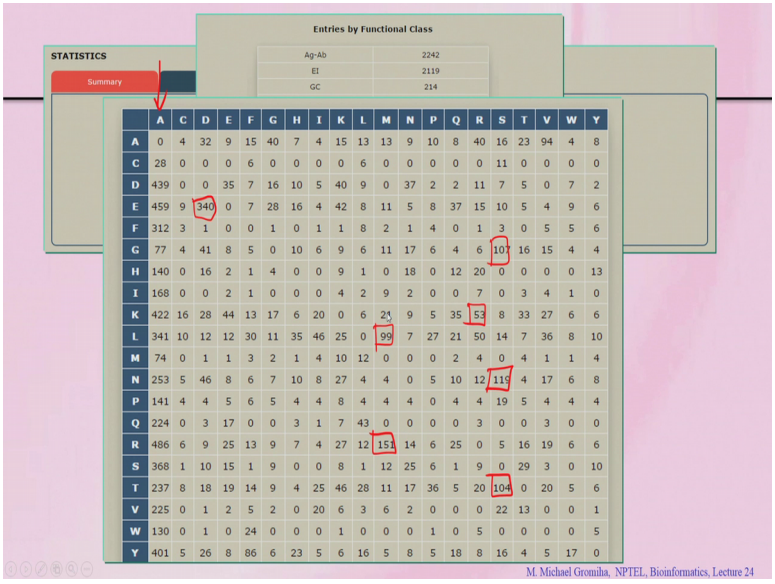
To understand the features which affect the binding affinity or you can derive models, different methods you can derive to predict the binding affinity of one mutation because sufficient number of data in this case it is possible. Then also you can see a double mutation 748 the double mutations you can use and see whether the binding affinity of a double mutation can be explained with the some of these single mutations or they are totally difference if it is totally different what factors influence this difference and if is additive or why they are additive. So, you can do the analysis using this data right.

(Refer Slide Time: 19:28)



So, with respect to the function class there is mainly the antigen antibody complexes more than 2000 entries and enzyme if it are also about 2000 followed by the other classes GC, coupling receptor, receptor complexes other enzymes and so on. So, then with respect to different parameters like delta G we can have the many value of minus 12.2 to the maximum value of plus 12.2 and also you can have the different data based on the experimental techniques.

(Refer Slide Time: 19:57)



	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0	4	32	9	15	40	7	4	15	13	13	9	10	8	40	16	23	94	4	8
C	28	0	0	0	6	0	0	0	0	6	0	0	0	0	0	11	0	0	0	0
D	439	0	0	35	7	16	10	5	40	9	0	37	2	2	11	7	5	0	7	2
E	459	9	340	0	7	28	16	4	42	8	11	5	8	37	15	10	5	4	9	6
F	312	3	1	0	0	1	0	1	1	8	2	1	4	0	1	3	0	5	5	6
G	77	4	41	8	5	0	10	6	9	6	11	17	6	4	6	107	16	15	4	4
H	140	0	16	2	1	4	0	0	9	1	0	18	0	12	20	0	0	0	0	13
I	168	0	0	2	1	0	0	0	4	2	9	2	0	0	7	0	3	4	1	0
K	422	16	28	44	13	17	6	20	0	6	24	9	5	35	53	8	33	27	6	6
L	341	10	12	12	30	11	35	46	25	0	99	7	27	21	50	14	7	36	8	10
M	74	0	1	1	3	2	1	4	10	12	0	0	0	2	4	0	4	1	1	4
N	253	5	46	8	6	7	10	8	27	4	4	0	5	10	12	115	4	17	6	8
P	141	4	4	5	6	5	4	4	8	4	4	4	0	4	4	19	5	4	4	4
Q	224	0	3	17	0	0	3	1	7	43	0	0	0	0	3	0	0	3	0	0
R	486	6	9	25	13	9	7	4	27	12	151	14	6	25	0	5	16	19	6	6
S	368	1	10	15	1	9	0	0	8	1	12	25	6	1	9	0	29	3	0	10
T	237	8	18	19	14	9	4	25	46	28	11	17	36	5	20	104	0	20	5	6
V	225	0	1	2	5	2	0	20	6	3	6	2	0	0	0	22	13	0	0	1
W	130	0	1	0	24	0	0	0	1	0	0	0	1	0	5	0	0	0	0	5
Y	401	5	26	8	86	6	23	5	6	16	5	8	5	18	8	16	4	5	17	0

So, this is a mutation matrix right. So, what can you infer from this mutation matrix like

protein stability. Protein stability mainly we obtain the data for alanine mutations here also you can see the alanine mutation because the alanine scanning mutation is, so you can see the plenty of data whether these residues are mutated to alanine which way there are many data for alanine.

Right, because we can see the difference because if you take 20 different amino acid residues glycine is the simplest one, but the problem with using Glycine is there is no side chain hydrogen is side chain. So, it is completely flexible in this case flexibility is also important plays a role for this binding affinity we take the alanine. So, all the 18 amino acid they contain the series two and the others are difference. So, alanine content CH 2 like all the other 18 amino acids right. So, if you mutate any residue with alanine you can see the influence or the contribution of different amino acid residues. That is the reason why they use generally its alanine to get the difference in binding affinity right.

So, what are the other mutations which are frequently done? So, they use glycine to serine, then asparagine into serine in it makes sense because asparagine is a polar residue we might asparagine to serine to see the binding affinity and the threonine in right. So, here arginine to methionine.

Student: E to.

E to.

E to D. So, if you do E to D what will happen?

Student: Length of the -CH2 group

Length of the CH 2 group right. So, I can say the how they length I want to say it is a group a reducing one from E to D, change the binding affinity. Likewise they did the lysine to arginine, not right, they increase lysine to arginine, arginine to lysine not right which one.

Student: Leucine.

Leucine yeah, or leucine to methionine right. So, there are many mutations. So, this will tell you how far the binding affinity changes with the respect to amino acid mutation fine.

So, you can do the average assignment method like here also, arginine the values and then see whether we can identify the mutants which increase the binding affinity decrease the binding

affinity or whether we can able to predict the binding affinity. And currently there are several methods are available in the literature right, but now the available amount of data we can make a good model for predicting the binding affinity of our mutation right.

(Refer Slide Time: 23:08)

So, we give the links. So, link to the different other databases for example, protein data bank this will for the protein structures. What is uniprot?

Student: Protein sequence database.

Protein sequence database, and the pubmed this literature database and string.

Protein interaction networks you can see if you have any particular protein is linked with R. This is interact with what are the other proteins in the network. For example, if we construct of a human protein-protein interaction network there are many proteins are making as a hubs interacting other proteins that information you can get from the sting database.

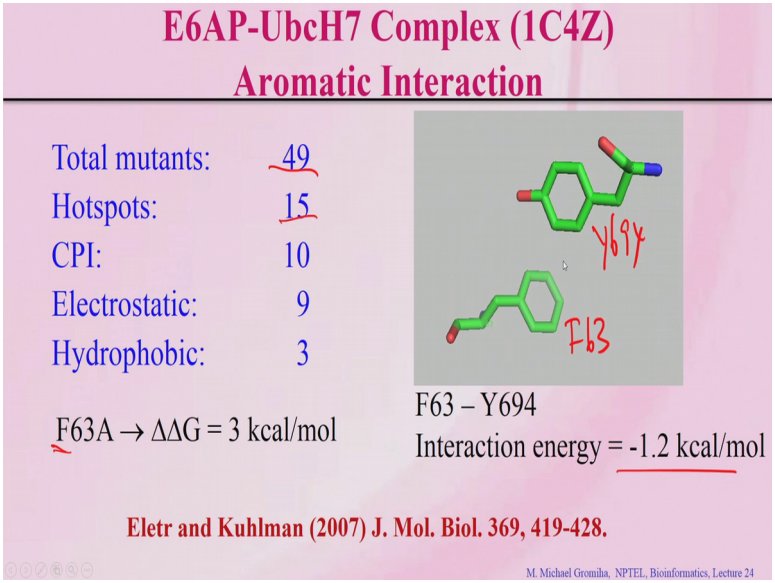
Also you can get the dictionary of protein secondary structure DSSP from this site and you can get this affinity data from different other resources, one is the ASEdb alanine scanning energetics database and the PINT, these are developed earlier and you can see this computer a ways is also contains the energetics of mutation protein interaction.

But compared with the other databases our database Proximate has several advantages where we have the interface and we can get different types of data with the functional classes and

more number of data. So, currently it is including the wild type as well as the mutant data fine right. So, from this database you can identify the mutants which are termed as hotspot residues. What are hotspot residues?

Yeah, at least the binding affinity change by 2 kilocal per mole. So, earlier we discussed and we identified some type of interactions then we will see whether these interactions are important or not, I show of you few examples.

(Refer Slide Time: 24:52)



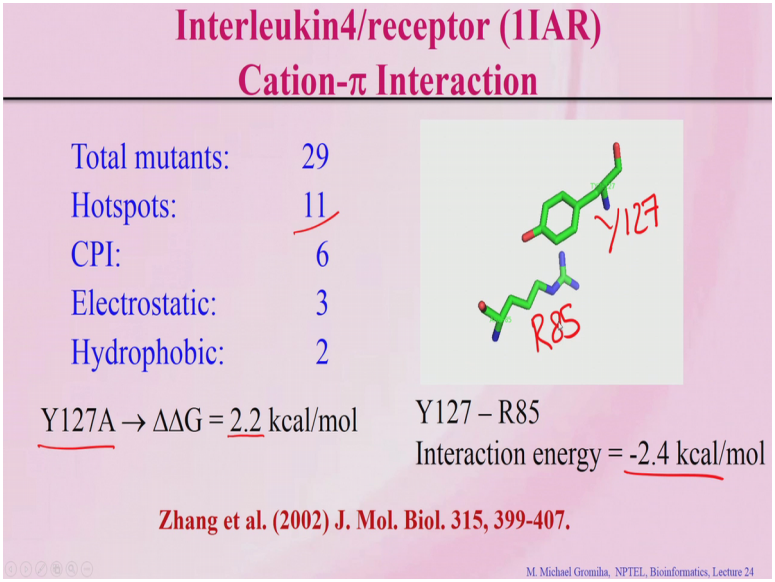
So, this is a complex E6AP-UbcH7 complex, this will tell you the importance of aromatic interactions right. So, here Eletr and Kuhlman, they experimentally measured the binding affinity of 49 mutants in this particular protein right. And they showed that 15 residues are hotspots. Among the 15 residues if you see 10 residues which are involved in cutoff an interactions and 5 are no electrostatic interactions, only 3 are hydrophobic residues.

There is specifically if you tell look at this replaced plain alanine at the position number 63 to alanine they change at the free energy change of 3 kilocal per mole. Now, the question is whether this F63 is important or not. You see the structural analyses, we also showed the phenylalanine, its important for the binding affinity.

So, you replaces this one that should reduce binding free energy. So, take these complex structure is known and get the F63 and how these F63 is interacting with other residues right. So, you can see these phenylalanine 63 is interacting with tyrosine 694. Now, if we see the

contributions all the main chain contribution and the side chain contributions and you see most of the contributions are due to the side chains and you can see the interaction energy of -1.2 kilocal per mole. In this case if you mutate this residue this will change the binding this free energy and this eventually reflected in the case of these solution studies like the free energy change of 3 kilocal per mole. That means, if we mutate the phenyl alanine change of free energy by 3 kilocal per mole because these free analyze very important even they structure they are making good interactions specifically aromatic aromatic interactions with tyrosine 694.

(Refer Slide Time: 26:54)



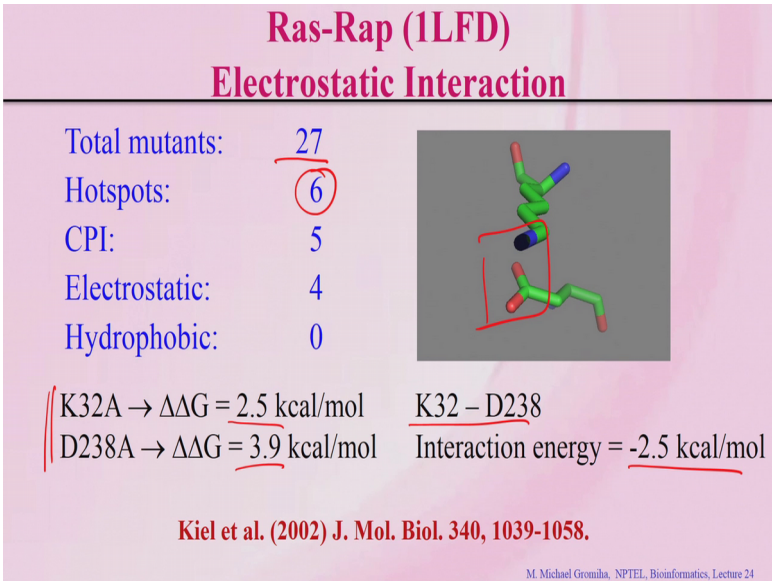
I show another example these were cation pi interactions here the Zhang et al they analyzed 29 different mutants. Among 29 mutants 11 mutants are identified in hotspots and 6 are cation pi interaction forming residues, 3 electrostatic interactions forming residues; that means, charged residues they mutailed Y127A. This case its change the free energy of 2.2 kilocal per mole; that means, there are also identified as a hotspot residue.

See if you see whether these tyrosine is really important for the binding affinity then they should make good type of interactions with the binding partner. See we see here this is the tyrosine 127, this is tyrosine 127, if you look into the binding partners for example, in the Interleukin 4 and receptor, so this arginine 85.

So, having the energy of 2.4 kilocal per mole, if we calculate the interaction energy right, this is 2.4 kilocal per mole these the van der waals plus the electrostatic interactions right. So, in

this case these two residues they are very important to form the interaction specifically a cation-pi interaction in this particular protein interleukin 4 receptor this complex fine.

(Refer Slide Time: 28:15)



So, in those, another example this is the Ras-Rap protein. So, here this is the importance of electrostatic interaction. So, here this group they mutated 27 residues, among the 27 mutants they I do not only 6 are hotspots and if you do the analysis on the hot spot residues 5 are involved in cation pi interactions and 4 are the electrostatic interaction forming residues there are charged residues.

So, here they mutated lysine 30 into alanine and they reduce the free energy of 2.5 kilocal per mole and D238A you can see here also they showed the free energy change of 3.9 kilocal per mole. If you look into the structure now this is the ITG studies and they go to the structure, so lysine a 32 and the aspartic acid 238. So, these two residues, these are two residues they are try to mutate and they interact with each other mainly because of this side chain with the free energy of -2.5 kilocal per mole they are making very strong interactions.

This shows that these residues K32 to D238 they are important to form the electrostatic interaction this is a reason why they are having the favourable interaction energy of 2.5 kilocal per mole and if we mutate either K or D they change the free energy of more than 2.5 kilocal per mole; that means, you can see there is an agreement between the computational approaches obtained from structures as well as the free energy change they measured from these experimental techniques. So, what are the various aspects we discussed today?

Or different types of complexes, what are different types of complexes?

Student: Protein.

Protein complexes.

Student: Protein.

Protein DNA complexes, protein RNA complexes and protein ligand complexes. What are the specific functions of protein complexes.

Student: cell signalling.

Student: Ubiquitination.

ubiquitination.

Student: Molecular switching.

Molecular switching and in antibody interactions and so on, immunological response and all. For the case of the protein - nucleic acid interactions.

Student: Transcription.

Transcription, DNS repair packaging and so on. Protein ligand interactions, mainly they are acting as an or triggering as an inhibitors for any activity and we can also use the protein ligand interactions to identify inhibitors set basically in the case of the structures based drug design or ligand based drug design right.

So, now if we have the complex it is very essential to identify the binding sites right. So, how to define the binding sites?

So, the two a proteins or the two binding partners whether they share a common interface that region you can see the binding sites because they are close to each other, they share the common interface interface area. How to identify the binding sites?

Student: Distance based criteria.

Distance based criteria.

Accessible surface based criteria.

Student: Energy based.

Energy based criteria. How to identify the binding site using distance based criteria?

You need the atom coordinates we need X, Y, Z coordinates, see their protein 1 and the partner protein 2 and get the distance rate of any atoms in protein 1 and protein 2 that is very important. If we set up any distance say 5 angstrom we need a heavy atoms with there within the distance you can identify these has binding site residues right. So, how to identify binding site residues using accessible surface area based criteria?

Student: Reduction accessibility.

Reduction accessibility, either they have the complex and if you take out the protein and see whether any change in accessible surface area upon complex formation. If there is a change then you can say that these residues are involved in binding. Then how to identify the binding site using energy based criteria.

Student: Calculate energy.

Calculate the energy like this sense you can calculate the energy for each atom and sum up the atoms for each residue and see which residues they have the free energy of minus 2 cal kilocal per mole then you identify the binding site residues right.

So, when you identify the binding site residues either you say distance based criteria or accessible surface area base criteria or the energy based criteria you can find the binding propensity right. So, if you do the binding propensity which residues are preferred at the interface? arginine and you can see the tyrosine, tryptophan and so on right. So, then we see whether these residues making any specific partners we identify some partners right. So, these partners they tend to form some specific types of interactions for example.

Student: Aromatic.

Aromatic, aromatic interactions, cation-pi interactions, electrostatic interactions, we discuss with the example. Experimentally they prove that this binding affinity drastically changes if the residues are involved in any specific types of interactions and we mutate this will change the binding affinity right. So, then we discussed about database which contains the binding

affinity of complexes and the mutants. Which database we discussed?

Student: Proximate.

Proximate, proximate contains which type of data?

Mainly binding affinity. So, you will get you will get the binding affinity of any mutations we can try to used has values for any understanding the binding affinity of any complexes or the affinity of on mutations right.

So, in the following classes we will then move on to the protein ligand-interaction and how to use our computing techniques for designing the inhibitors using structure based drug design or using the QSAR techniques and so on and after that we will discuss about some of the common problems and how to approach the problems using bioinformatics approaches.

Thanks for your kind attention.