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

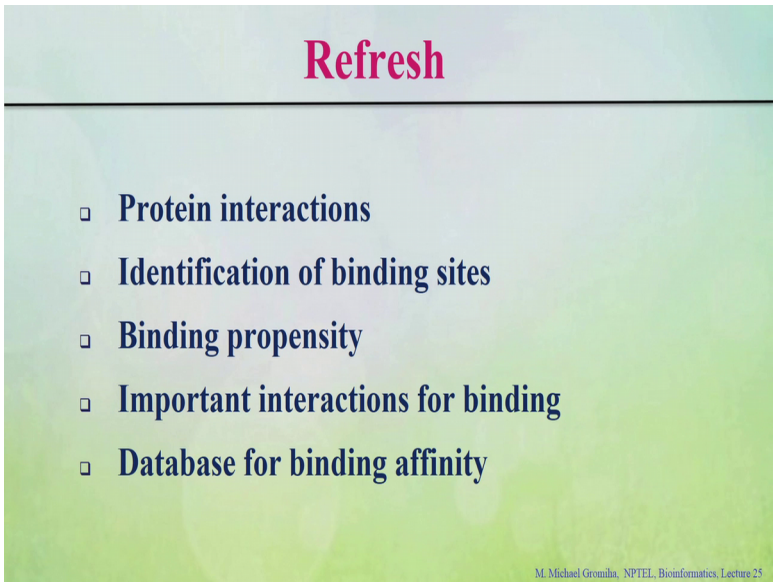
Lecture – 25a
Computer Aided Drug Design I

In this lecture, I will discuss about the applications of bioinformatics right in computer aided drug design. We have studied about various techniques in bioinformatics using sequences and the structures, and also we discussed about various types of interactions and how this information can be used to identify the lead molecules or the probable inhibitors right for any specific targets.

Because the current scenario it is very important for the development of drugs, because it takes about 10 years from the beginning to come to the market and many lead compounds they fail in various different clinical trials, and in this case it is very important to have a set of molecules right which could be a lead compounds right for initial start.

So, I will discussed different techniques for example, the docking and virtual screening and quantity structure activity relationship right to identify the probable inhibitors right generally used in computer aided drug design.

(Refer Slide Time: 01:20)



In last few classes we discussed about the proteins, protein folding rates, protein stability and protein interactions, and specific in the last class discussed about the protein interactions right what a different types of complexes?

Student: Protein-protein complex.

Protein-protein complexes.

Student: Protein.

Protein nucleic acid complexes protein cabohydrate complexes and protein-ligand complexes and all these complexes they are important for several functions right. Then we can identify the binding sites or the interface residues of any complex. So, we discussed about three different criteria to identify the binding sites, what are three different criteria we discussed?

Student: Distance based criteria.

Distance based criteria.

Student: Solvent accessibility.

Based solvent accessibility.

Student: Energy.

As well as interaction energy right.

And using these information we identify some residues which are preferred to be at the binding site and the binding partners right between two proteins there are some residue pairs are preferred, and they are involved in different types of interactions right. We discussed about three different types of interactions right what are different types of interactions we discussed.

Student: Aromatic.

Aromatic-aromatic interactions right and cation pi interaction.

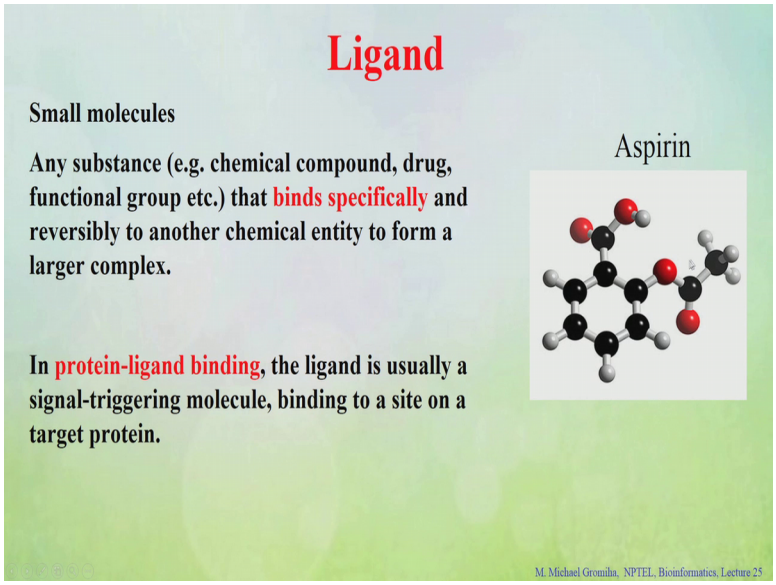
Student: Cation pi.

And electrostatic interaction, then to understand the importance of these interactions we need to see whether the mutation of these specific residues at the binding interface, change or alter the binding affinity right with the respect to other residues.

So, we discussed about a database called the proximate right, which contains information regarding the binding affinity upon mutations right. This could be useful resource right for researchers to understand the binding affinity right of any Protein-protein complexes upon mutations. So, when you go to the protein-ligand interactions right. So, what do you mean by ligand what is the meaning of ligand?

Generally we can say any small molecule, so ligands are small molecules or you can see any substance or any chemical compound right.

(Refer Slide Time: 03:10)



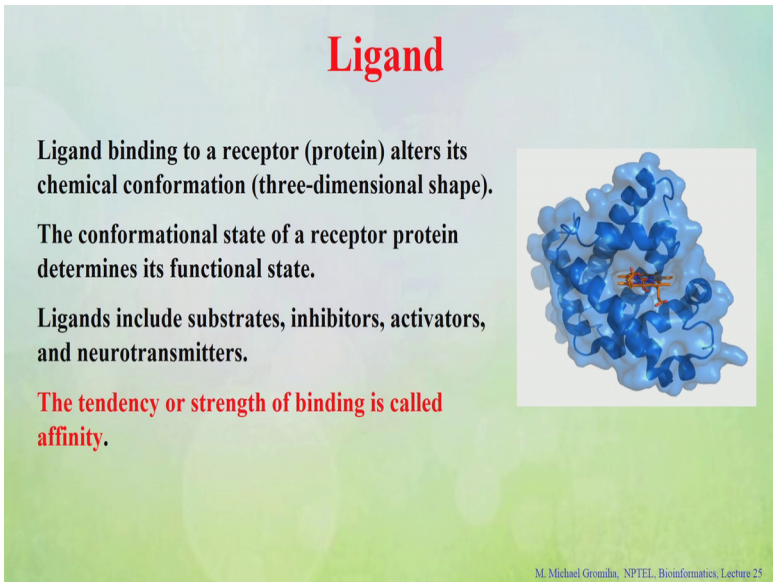
It can be a drug or any function group, that find specifically to any other entity to make a complex. For example, a small molecule that may interact with a protein like protein-ligand complex or a drug molecule can interact with a DNA like the DNA drug interactions that is so on.

Here is one example this is aspirin. what is the aspirin right?

Student: Pain killer.

It is a pain killer right, in that case if you have a headache, then you can use and have this aspirin right this can see it is a drug. So, if you look at the protein-ligand binding. So, you have protein and you have ligand, the if they binds the ligand is usually a signal triggering molecule, because this will go and attach with a pocket in a protein right and this will trigger right at the binding site. So, where are this will target right at the protein that site is called the binding site right in a target protein.

(Refer Slide Time: 04:08)



So, I show this ligand binding to any receptor any protein right. It changes the conformation of the protein, a protein has any specific conformation as it is, then when the ligand small molecule they interact with any pocket right this will change the conformation right and this will also make to have the interaction between the protein as well as the ligand. This ligand can be a substrates or this can be a inhibitors or activators as well as neurotransmitters, then how strong are also tendency of this specific ligand to interact with the protein right that tendency or the strength is called the affinity.

So, if you see affinity of a ligand. So, what is the meaning of the affinity of a ligand?

Student: The strength of binding.

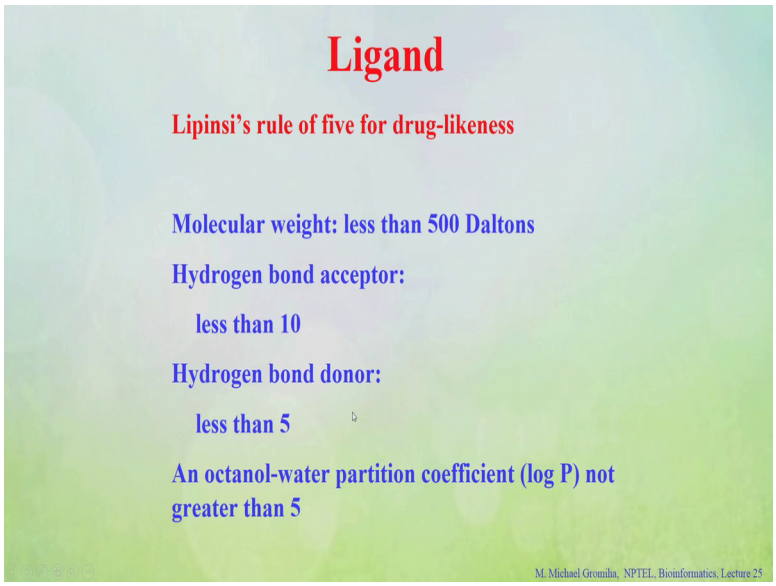
The strength of the binding right how far this ligand can interact with the Protein-protein molecule right that will give you the affinity. So, if it is high affinity what is meaning of that?

Student: Tightly bound.

It is tightly bound right like compare to low affinity. So, if you have several compounds then we can identify these different affinities of each compound then you can find the specificity right among these different small molecules, which one has the best preference right to interact with a particular protein.

So, if you see higher ligand generally it can be described with Lipinsi's rule of 5 right because if you see all the rules and the all the rules are the multiples of 5.

(Refer Slide Time: 05:35)



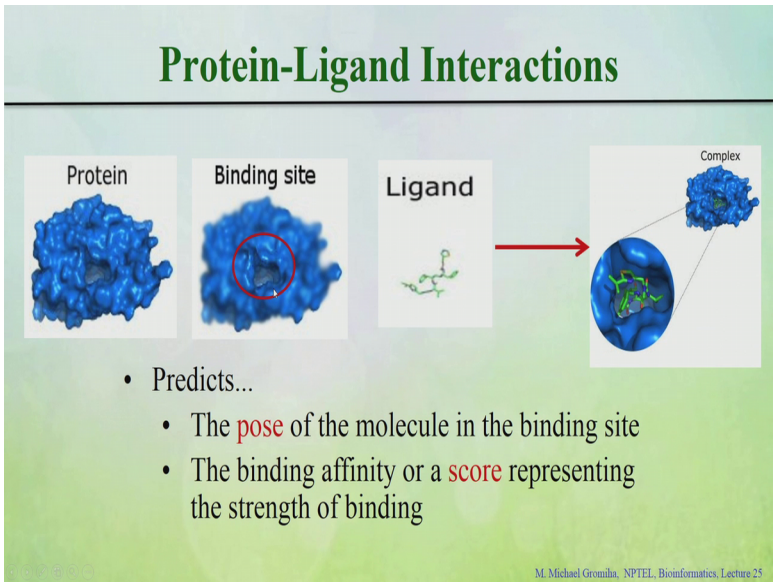
So, there are four rules all are multiples of 5. So, here Lipinsi stated these rules for a drug for a ligand to be kind of character of drug likeness right not necessarily all the drugs should have should follow this criteria, but most of the drugs right or any chemical compound which may act as a drug likeliness, they may have this type of conditions.

The first one is molecular weight. So, what is the molecular weight? we can see the any compound you can see the weight, the molecule weight should be less than 500 Daltons, then we have the hydrogen bond acceptor the less than 10 and hydrogen bond donor less than 5. What difference between hydrogen bond acceptor and hydrogen bond donor, how to discriminate hydrogen bond acceptor and hydrogen bond donor?

Student: So, hydrogen (Refer Time: 00:27).

Yeah hydrogen bond attached hydrogen attached with the electronegative atom that is called hydrogen bond donor right which one receives this one is called acceptor then the hydrophobicity that is mentioned as octanol water partition coefficient right that is not greater than 5 we have the hydrophobic. So, this not be a not greater than 5.

(Refer Slide Time: 06:48)

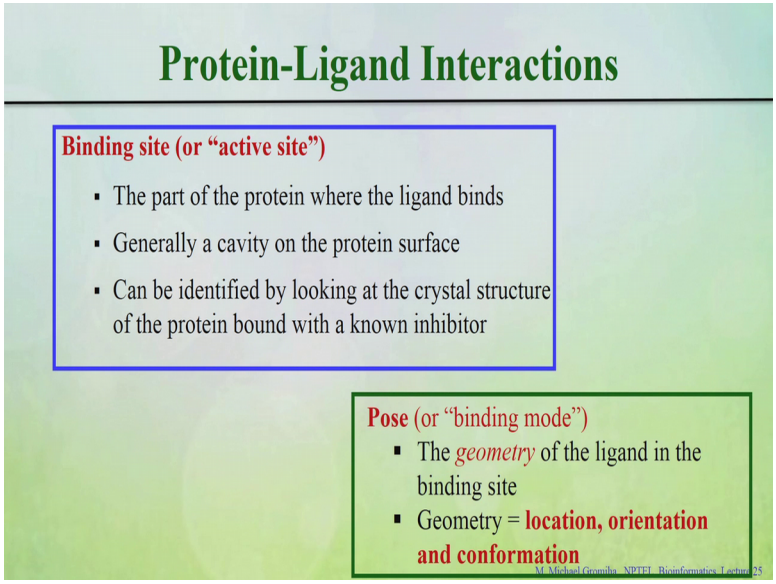


So, now if you are this is the characteristic features of ligands. Now when they try to interact right you can see a protein here we look at the protein. So, you can see a pocket here right. So, this is a pocket, this you can say kind of cavity, probably a binding site. So, if a ligand tries to goes near to this protein right may for example, here this is the ligand here this makes the complex. So, you have protein, protein has a binding site right and the ligand can go to the binding site, and it interact and finally make the complex.

Here we have to think about two different accepts one is the pose right. So, you have the cavity right acting as a binding site, what is the pose of this particular cavity, what is the conformation of that particular binding site. Depending on the conformation then the ligand can have different efficiencies right.

If it is fit well then it can tightly interact right this case the affinity is very high right or if does not fit well then this case the affinity is less. So, one is the pose the conformation of this particular binding site, and the second one you can see the binding affinity, we can relate in terms of a score which will give you the strength of the binding.

(Refer Slide Time: 08:02)



So, two different aspects one is the binding site or the pose, and the second thing is the scoring. So, what is the meaning of binding site or a pose right? Here you can see this is the part of the protein where the ligand binds, in this figure you can see here this is the place where the ligand can probably bind right. So, this you can say as the binding site.

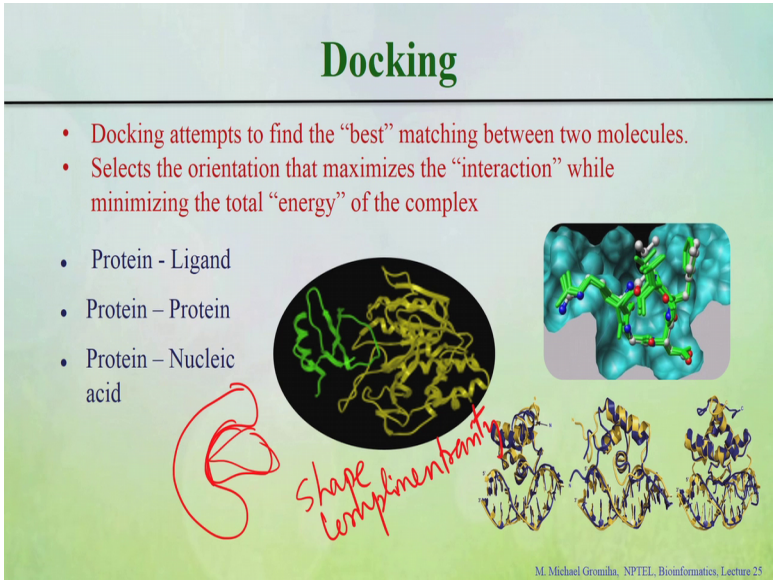
And generally this kind of a cavity on a protein surface right because in the interior part is highly covered, and they are mainly imparting the stability of the protein in this case you can see generally type of cavity right at the protein surface.

Then how to identify this active site, right. So, generally if you look at the crystal structure of a protein right bind with the any known inhibitor, and if we take the inhibitor you can see there is a cavity right where the ligand bound with this protein. So, you can look at structures you can see this binding site, and also a lot of several computer programs available. In this case you can give the PDB structure right you can identify the binding sites of any particular protein right.

The second thing is the binding mode like what is different binding mode? This will give you the scoring right. So, we need the binding site then the second one we need the binding mode or the pose and based on this the ligand can bind and we can score whether the ligand can properly fit or not.

So, what is the binding mode or the pose? This will give you the geometry of the ligand right in the binding site. So, geometry is terms as location, orientation as well as the conformation right you can see both in the case of the protein site as well as the ligand site. Ligand site also you can change the conformation and make various conformations right for example, if I see you show this ligand this is one conformation. During these translation rotations, we can change several conformation of this particular ligand then you can see whether this ligand can fit with this binding site right fine.

(Refer Slide Time: 10:01)



So, now, we have the protein, the protein we identified the binding site and we have a ligand, ligand we can also generate different conformations. Now the next aspect is whether these two will interact or how far these pose of these different ligands they can fit with the binding site right. For this we use set of docking right what it is a meaning of generally docking?

Student: (Refer Time: 10:29).

Is to get the interaction between the two molecules, the dock between two molecules. So, for example, the protein as well as the ligand. So, what the docking does? So, docking generally tries to find the best match if you have a protein as well as your ligand, it try to interact with the particular site with the best match.

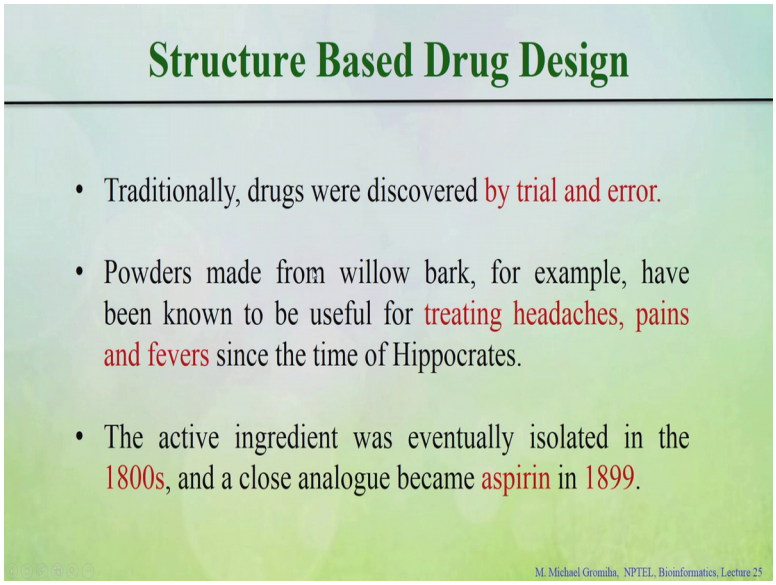
For example you can see any Protein-protein complexes or the protein ligand complexes or the protein-DNA complexes. It is a kind of lock and key mechanism, you can see a shape complementarity right for example, if you have a binding site like this, and if here you can see the binding sites right then you can find the ligand which right we can see to accept right to interact with that particular place, like the shape complementarity right to find the best match right.

So, then if you do like this what will happen? The orientation changes which maximizes the interaction between the protein and the ligand and minimize the total energy of the complex, this case it is it gives a strength of binding right. You can see various examples say protein-nucleic acid, protein-ligand the Protein-protein and so on.

So, these binding is very important and based on the strength of these interactions, you can relate with the activity of the compounds and this plays an important role in the structure based drug design; to identify a lead compounds right, to identify the inhibitors right for any specific targets right which are a involved in this different pathways or in different diseases fine.

So, how the terms structure based drug design emerge, how they started to identify some compounds and how they related these small molecules or ligands right to interact with the proteins and the discovery of structure based drug design.

(Refer Slide Time: 12:22)



Earlier days they tried to check the drugs by trial and error methods, in ancient days if they are sick they take some leaves right and they eat some of the leaves even nowadays right many medicinal compounds right. So, mainly the compounds are important, but if you eat this leaves right and several other parts of the of these plants right they are used as a potential drugs.

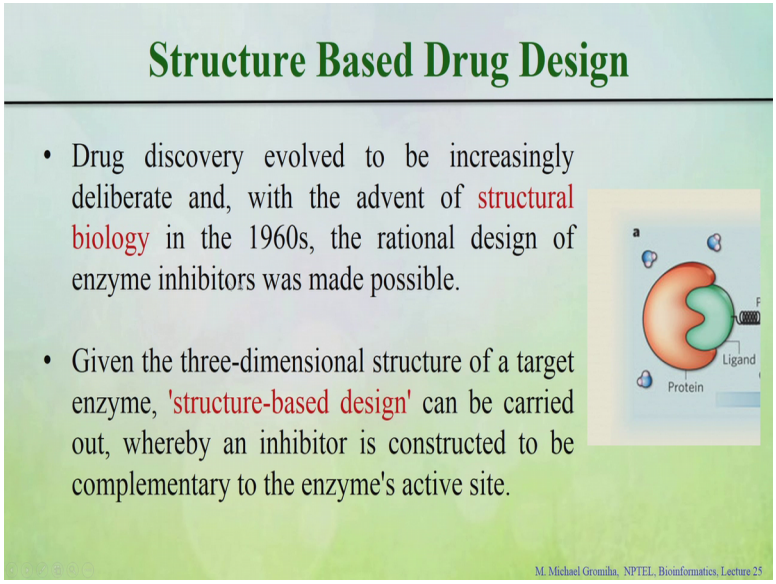
So, they get the powders made from willow bark for example, right useful for the treating headaches, pains, fevers and so on in the earlier days. Then they try to get what is the active ingredient, if you eat some of the leaves it can be medicinal, but what are the active compounds which one interact with the system right. Here they found in 1980s they found that this a close analog right then that became aspirin 1989 right.

So, they found that that there is a chemical compound that is present right in these plants, right this is close to the aspirin in 1989. Because of the width of the structure biology, then they you get the protein structures, which structure first they determine the structure which protein?

Student: Insulin.

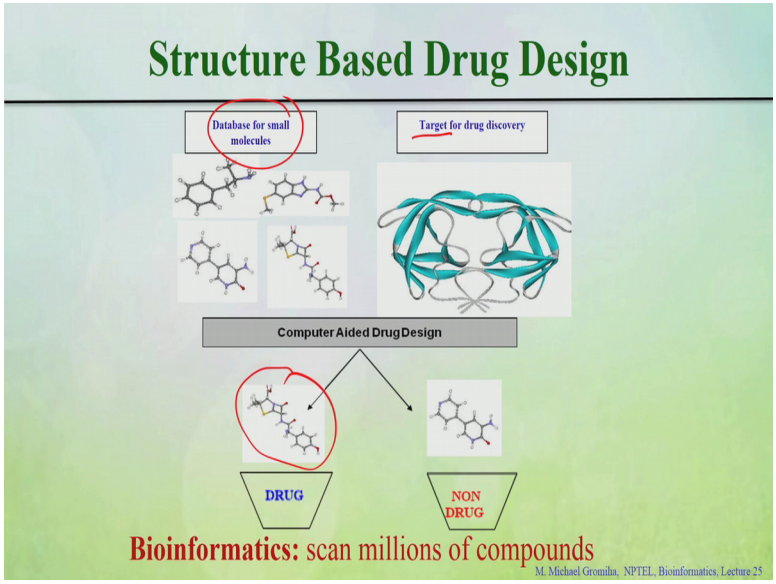
That is for a sequence. for the structures? that the hemoglobin myoglobin they got the structure, then they got the nobel prize. So, when we got the structures then we know the ligands which are working well as a drugs, then they use the structural biology right and then they see which ligands and how they fit with a structures. This gave this structure based drug design right.

(Refer Slide Time: 14:06)



So, we see given the 3D structure of any enzyme right we can use a structure based drug design right where the inhibitor can target right in the particular protein. For example, in this figure here if you see this is the protein and this the ligand right how this ligand can fit with the protein? Mainly the shape complementatirity and kind of the lock and key mechanism, how they will interact right and see how the activity changes fine.

(Refer Slide Time: 14:37)



So, they show an example. So, here left hand side, this is the database of a small molecules right there are many databases available for small molecules for example, a

pubchem, you have zinc database, the Chinese medicine, Enamin database right there are many databases likewise for different diseases for example, Chikungunya or dengue right recently this is going on and many other viruses right.

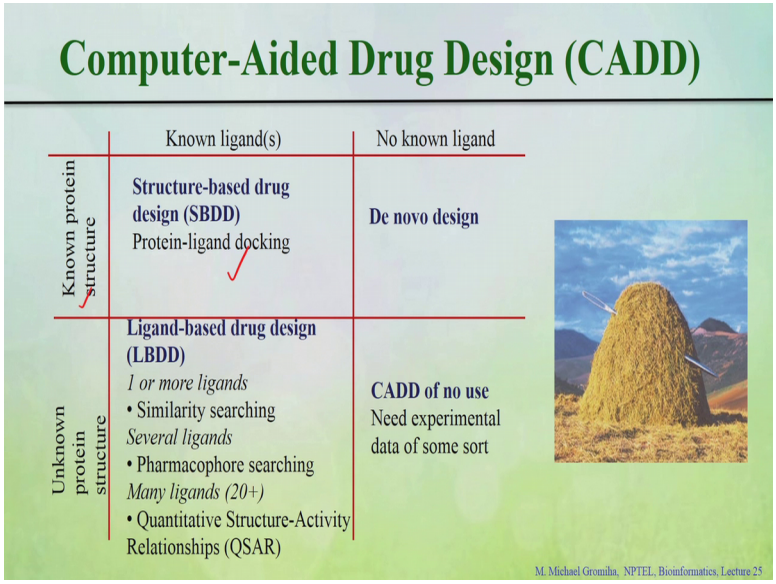
So, like there are several proteins which are involved right in these diseases right for example, RNA depend RNA polymerize right for the case of the dengue. So, in this case you can have the targets. So, now, the question is if you have the database of small molecules and we know the target right then how to identify the lead compounds which can fit with this target, and it could be useful as to design or get a drug molecule right. So, we try to use different compounds and this case we use computer drug design. So, there are various aspects like docking, screening, QSAR, I will explain the details.

So, using this information they will see some molecules. So, this can be a drug and other case this is not drug right. Also the drug likeness depends upon the molecular weight hydrogen donor acceptor as well as it should pass other trials right toxicity as well as the side effects right it has to pass all the clinical trial. So, using bioinformatics right we can scan millions of compounds, quickly you can scan millions of compounds and then can you can provide the information that these are the potential compounds right which could be used for the screening right.

And then from that screening you can identify the compounds, you can reduce enormously and see this can be used right for the further experiments right to see the activity as well as for the other different animal studies.

So, if you want to carry out computer aided drug design, are structure based drug design. So, what are the features what are the input information we require right. If you there are two different aspects one we need structure one is a ligand if the structure is known and the ligand is also known, then can we do this you can do that right if it is here the known structure right.

(Refer Slide Time: 16:55)



And here the known ligand right then you can see these ligands can interact with this protein right you can see the docking, and see whether this ligand will fit right with the particular protein right and this could be a probable lead compound or not.

Now, the second question is if we know the protein structure, and we do not know the ligand can we use structural based drug design? We can do right because we have a set of libraries for the ligands, protein structure is known. So, we can screen all these small molecules with respect to a protein, and then we can see these are the probable compounds right which could be a lead compounds right.

It is a kind of the needle in haystack to finding this one structure from a set of structures right. If you have pool of a small molecules and from that the set we have to identify the probable compounds and the next option is we know the ligand, but we do not know protein structure.

In this case can we you use the computer aided drag design right. If protein structure is not known, but ligands are known. In this case if you unknown the protein structure can we predict the protein structure? Yes right we discussed about various types of a modeling what are different types of a methods to get the protein modeling.

Student: (Refer Time: 18:23).

Homology modeling.

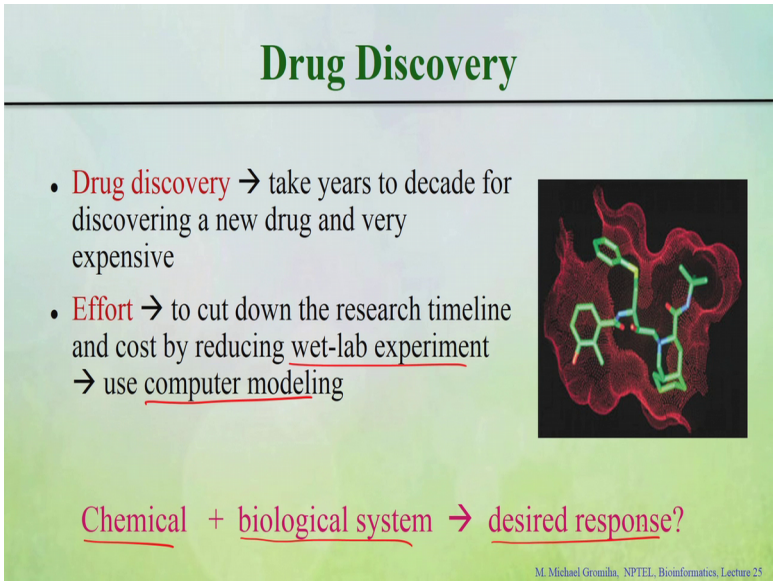
Student: (Refer Time: 18:24).

Ab initio fold recognition, and the hybrid methods right. So, if we do not know the structure we can get the structure right and the structures are known then we can try to interact to the ligand and we can a do that. But if structures are not known, ligand sets are not known then can we a do right.

But we can do that, but that is not accurate, but will give mainly wrong results in this case that cannot be used right because we need at least some sort of experimental data right without any experimental information right. We can get some models, but the models are not reliable, in this case we cannot trust this computer aided drug design to get the probable lead compounds fine.

So, we look into the drug discovery right if you want to have a drug, normally how many years you will take to get a new drug? It is almost about 10 years or 10 to 15 years it will take right from the lead compound and then do all these experiments, get the activities then the animal studies then you have to do with these clinical trials right phase one, phase two, phase three then finally, if you go with the several million compounds ended up one or two compounds right as a potential drug.

(Refer Slide Time: 19:38)



But it involves lot of time right and also lot of efforts we need and it also involves lot of money right because it is very expensive. So, to do all these trials. So, if we start

everything from this experiment they do the wet lab experiment from the beginning then it involves lot of time right it is a big process. So, to reduce these samples, to reduce the probable potential compounds right we can use the computer modeling right.

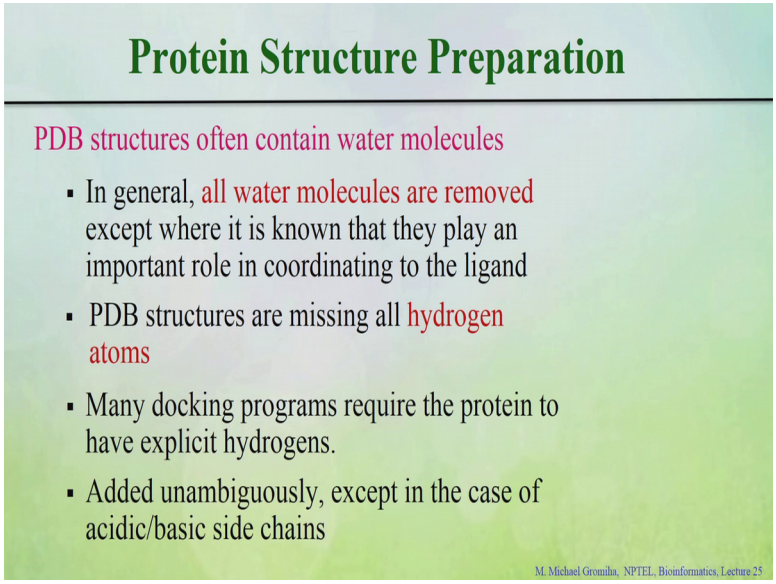
So, to see or screen the probable compounds which could be act has lead compounds. For example, you have the chemicals say ligands with the biological system right it may interact with the different types of the proteins right and whether we give the desired response or not right. You should work right, but it should not give any side effects right it should not give any toxic it should not be toxic. So, whether you can give. So, bioinformatics can do right to identify the initial set of compounds at least right we can identify these lead compounds.

So, now we have the protein and we have ligands right how to do the docking. Before doing docking. So, you need to prepare the protein also we need to prepare the ligands right because it earlier lot of other issues in the case of proteins as well as the ligands right. So, you have to work on that. So, let me explain what are the various options. So, how to get the PDB structures from x ray crystallography? So, you will get the direct structure from x ray crystallography? No we do not get right we do not get the structures what will we get?

Student: Map.

Map you will get the electron density map. From the electron density map we derive the structure, but how accurate the structures from the electron density map. If the map is not clear then the resolution is very low, in this case we are not sure about the atomic positions of these all the atoms right. So, then second case the water molecules right sometimes they put lot of water molecules, to crystalize and to solve the structure right.

(Refer Slide Time: 21:43)



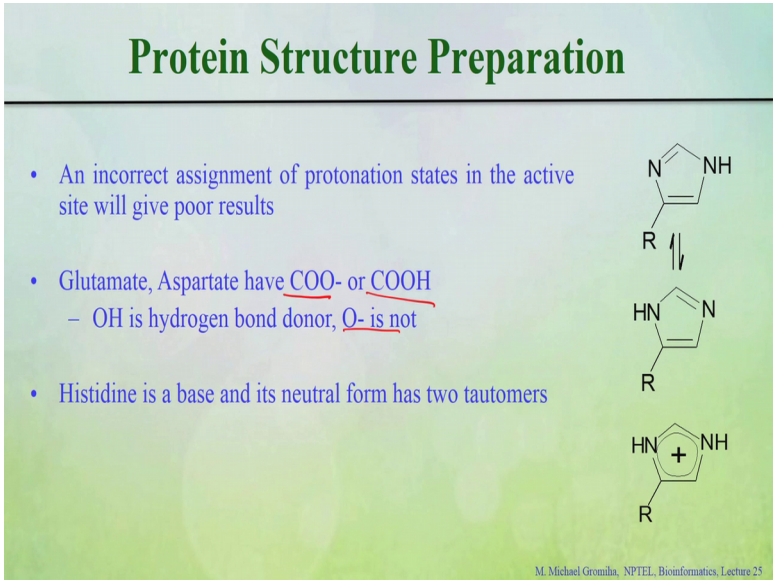
So, in the case of the protein preparation right. So, we if the water molecules are important; we keep the water molecule if the water molecules are not important for the ligand to interact, then this case we can remove the water molecules. Second thing is most of the PDB structures if you see, the hydrogen atoms are missing mainly in the case of x ray structures.

If you see the NMR structures you can see the hydrogen atoms, but if we look into the x ray structures, most of the structures the hydrogen atoms are missing. So, in this case you have to add the hydrogen atoms right all the missing atoms because most are docking programs require the explicit hydrogen why do you need explicit hydrogen? At least for getting these hydrogen bonds; hydrogen bonds are very important. So, in this case we need these hydrogen to see the explicit where we can a have these hydrogen bond.

So, we require the hydrogen atoms we can add hydrogen bonds, but some many cases we can add for example, the CH 2 group. So, three hydrogen 2 or 3 hydrogens you can see the positions, you can add easily. But if you say the case of the acidic and basic side chains, the question is whether this is charged or this neutral. So, that is a confusion this an ambiguity so adding all these things.

So, then the main aspect is if you incorrectly assign the protonation states, then it will give the poor results mainly the active sites when the ligand interacts with the protein.

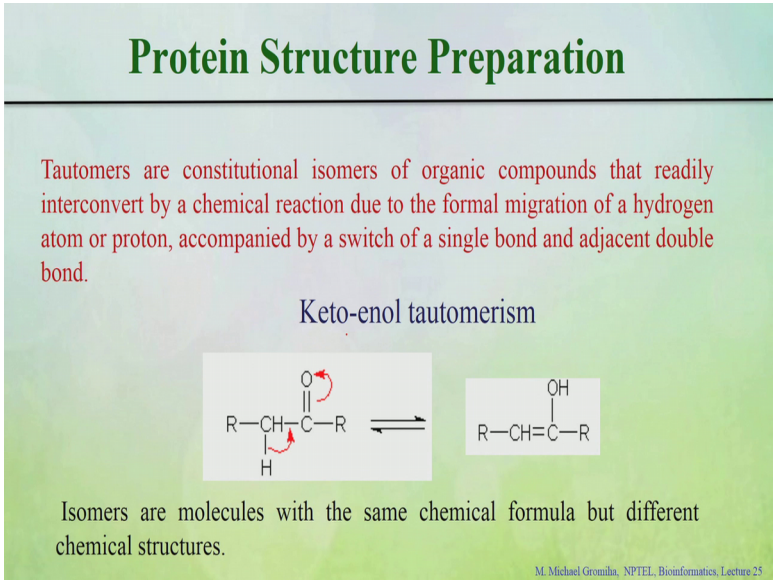
(Refer Slide Time: 23:16)



So, you have to assign the exactly protonated state for example, if you take the glutamic acid or aspartic acid, either we can give coo minus or COOH, but either same or different they are different right. Here if you see the COOH it is OH is hydrogen bond donor, but this is not right.

So, we need to be careful to assign all these protonation state at the active site right that is the place where the ligand can interact. Then if you take the histidine right it is a base right and is a neutral form it has the either the base right a positive charge and neutral form there is two tautomers right. In this case we have to check which one is the preferred ones for the ligand to interact. So, what is tautomers?

(Refer Slide Time: 23:56)



These are isomers are organic compounds right they inter convert the chemical reaction right for example, from the Keto to enol; due to the migration of hydrogen atom or proton right from the single bond and adjacent double bond. See here this H this is CO, so it is Keto group.

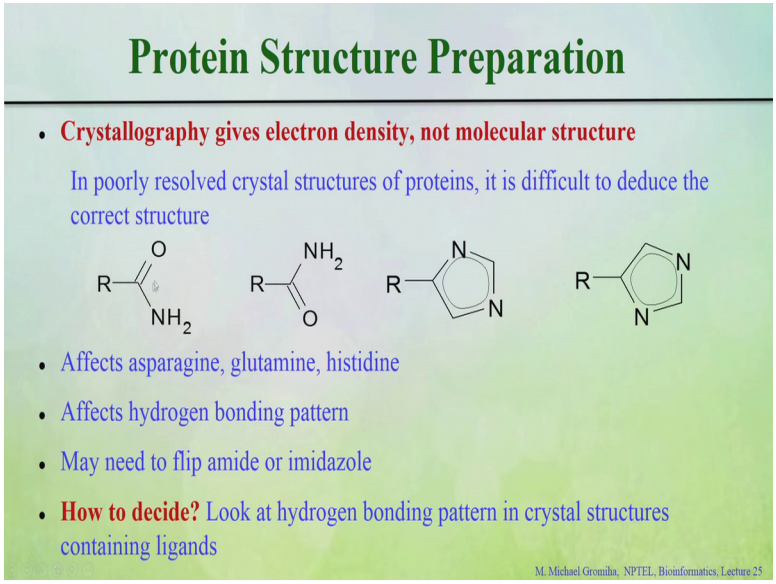
So, this protonation happen here then finally, they see the single bond is OH, this is called the keto-enol tautomerism right. So, what is the isomers?

Student: (Refer Time: 24:31).

Right they have the same chemical formula, but different chemical structure right. So, now, we need to consider mainly about these tautomers specifically in the case of these histidines and the conversion of 1 chemical group to the another chemical group.

So, now the next aspect is as I discussed earlier, the crystallography gives electron density, but not the structure right. In fully resolved crystal structures, it is very difficult for example, see the asparagine are the glutamine here in case is down, but the OE is up.

(Refer Slide Time: 24:57)



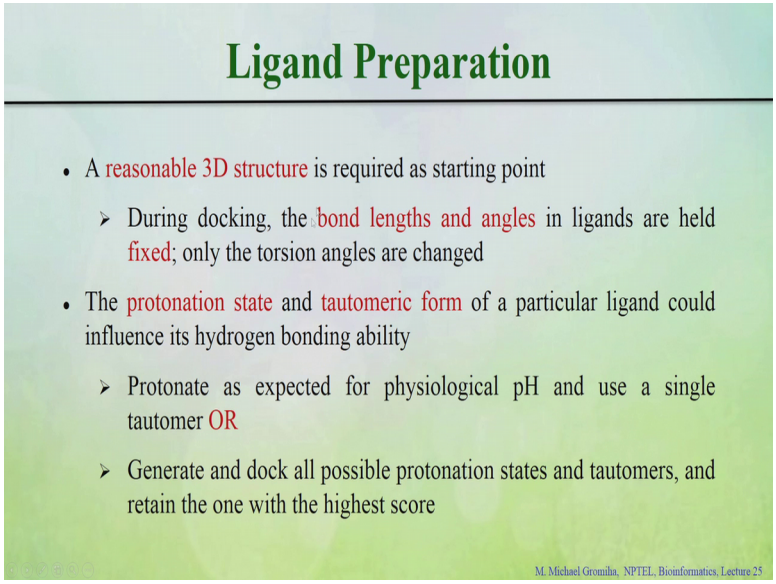
It is also possible in other way around right we change the conformation then the ligand also the interaction will be different right. Because if the ligand we gave one pose with oxygen at this side we have different interactions, if the other side you have different interactions right.

So, this orientation that is important then we have the we need to flip the amide or imidazole. So, mainly in the case of the histidines also we need to see where are these nitrogens, then how to decide which one is a best right. So, we can decide based on the structures with ligands. If you see the PDB several structures we can give free structures also many cases you have the structures with ligands.

With the ligands you can look at the structure and see what are the probable orientations probable conformation and what are the probability of having this histidines right in which state. Then look into this information and then you can decide for example, your case this could be the most optimal one fine.

So, if you go the proteins side is fine, then go to ligand side, ligand side also we need to have a probable 3D structure right.

(Refer Slide Time: 26:14)



Because the docking mainly bond lengths bond angles would not change much. So, they do not consider these, but torsion angles are important right. The torsion angles are important that we give the rotations around the different atoms.

See here also we have to see the protonation state and tautomeric form right. This will change or influence the hydrogen bonding ability, whether they share the strong bonds are the tight bonds or it is loose. So, here we see the protonated state right either is a physiological pH and use single tautomer, or use all possible protonation states and then look at the lowest one at the highest score, then take that for the further processing so that you can do right.

So, now we have the protein, we analyze a protein and prepaid the protein, then we have the ligand, ligand also we have choose the confirmation ligand right then what we can do then you start docking.