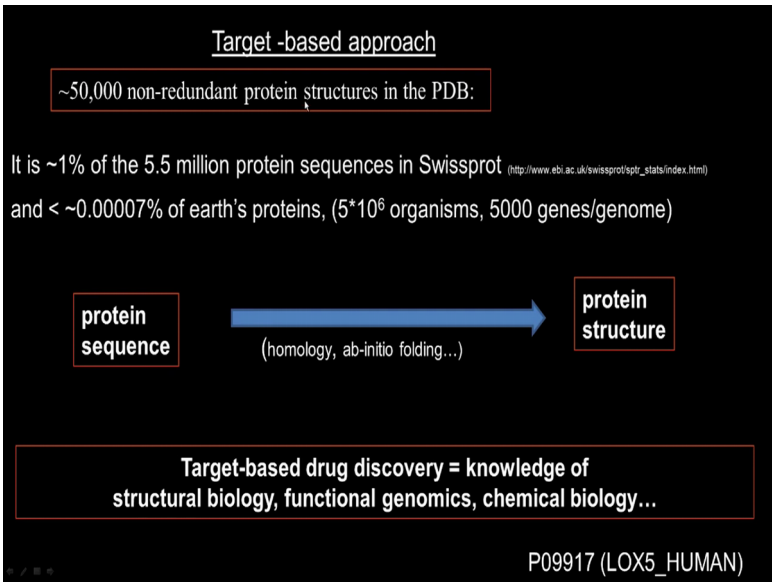
Computer Aided Drug Design
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology - Madras

Lecture - 35
Target Based Drug Design

Hello everyone, welcome to the course on Computer aided drug design, we will continue on the topic of target based drug design, that means the target could be a protein or enzyme. And you want to look at how the ligand goes and binds to that okay.

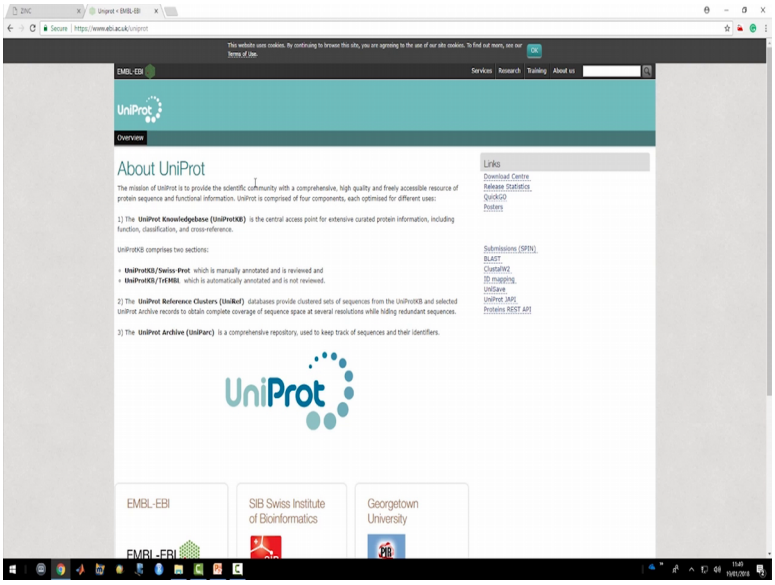
(Refer Slide Time: 00:34)



So in the target based approach, we need to understand this particular point okay, if you look at the protein databank it has 3D structure of 50000 protein structures, which is 1% of 5.5 million protein sequences that means 2 dimensional sequences available in this swiss-prot, and which is 0.00007% of earth proteins okay, because earth has lot of organism gene. So we have very little knowledge about the sequence of proteins found in the earth.

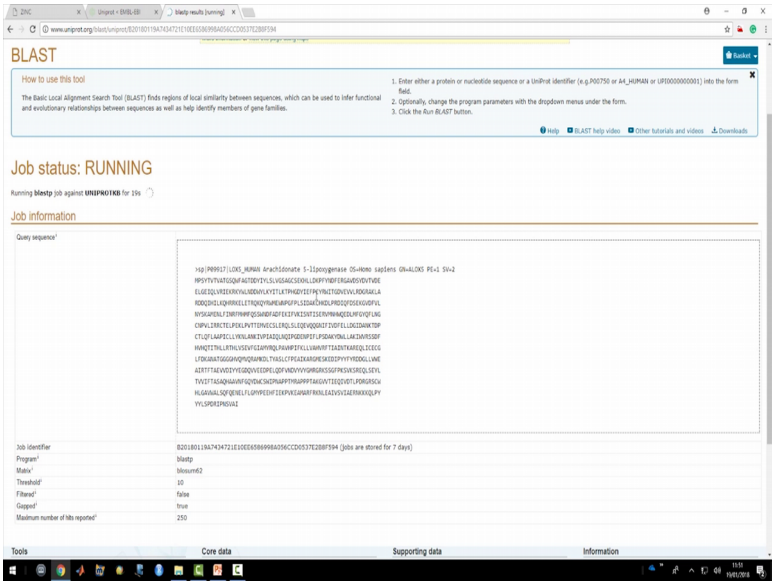
And we have still very, very little details about the 3 dimensional structure of the protein, so there is lot of scope, so for the next several 100 years we may be looking at 2D sequences and 3D crystal structures and so on. If you want to look at proteins involved in the disease okay, so that is still a long way off okay.

(Refer Slide Time: 01:43)



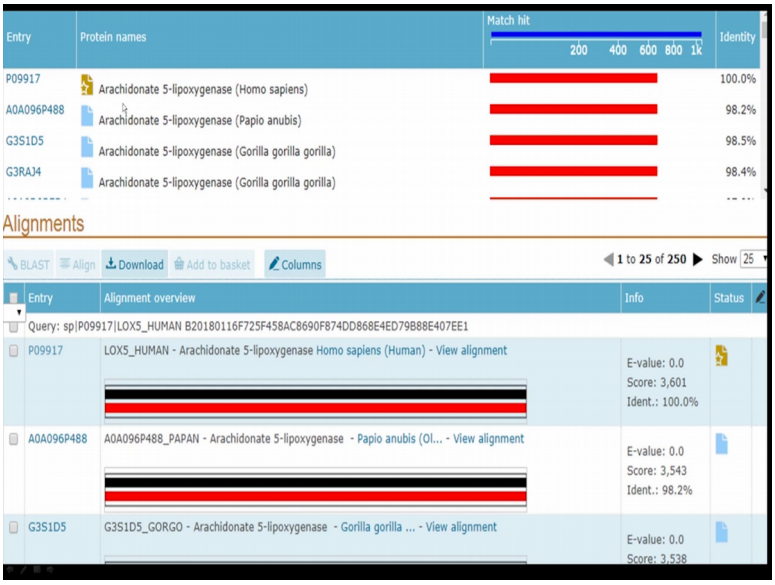
Let us look at this particular database called a swiss-prot okay, if this type is called uniprot okay uniprot, there are a specific ID for each one of the protein we can get the 2-dimensional sequence of all these proteins. It is for example look at this you know this is a lipoxygenase human lipoxygenase okay, as we have been telling this 5 human lipoxygenase is involved in the arachidonic pathway of the inflammation, and it leads to bronchitis and so on actually.

(Refer Slide Time: 02:27)



So let us look at the sequence of this, and we will make a comparison of other sequences of other protein that are found okay in the uniprot okay, so you give the uniprot number and then say BLAST okay, BLAST is a basic local algorithm search tool, and it looks at other proteins found okay job is running here. So it will look at 2 dimensional sequence of other proteins which are similar to the sequence of this particular 5LOX human which I had given.

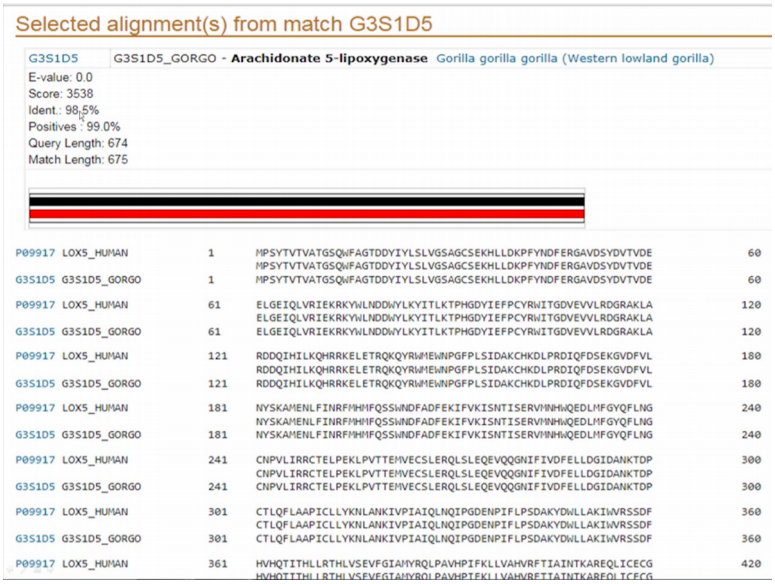
(Refer Slide Time: 03:08)



So it will take a long time, I have captured the screens before for quick okay, so this is how you will get okay, this is the protein this is the 5 lipoxygenase it is called arachidonic 5 lipoxygenase, because it arachidonic acid is substrate of this, and this particular protein is from homosapiens. So it compares other proteins in the uniprot 2-dimensional sequences and says this particular protein okay from papio anubis, has a sequence similarity of 98.2 percent identity.

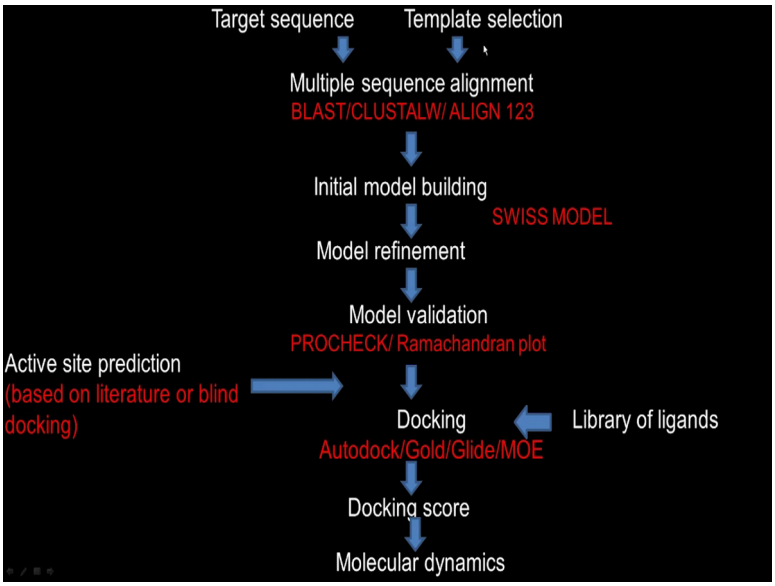
And then another one gorilla 98.5%, and then another one 98.5 and so on. So as you can see it makes a comparison it gives something called E-value, score, identity okay and so on actually. So we can look at large number of proteins and we can select the one which has a very high percentage identity or a very high score okay.

(Refer Slide Time: 04:09)



So we can also look for example look at this particular one okay 98.2, and then make a comparison of the 2D sequence alignment okay. So you can look at here okay it is got an identity of 98.5 and so on actually. So let me again check whether a software is done its job okay it is still running, so it takes a little bit long time, and you will get the results here with a comparison of many things okay, we will do that later because it is still running here okay.

(Refer Slide Time: 04:50)



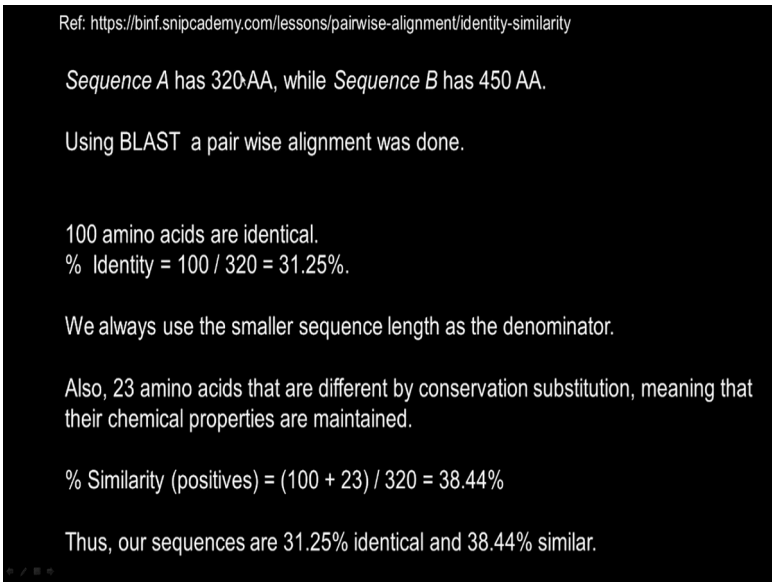
So that is very, very important for you to identify is there any protein or proteins which have similar sequence identity as the protein of your selection okay. So we have the target sequence, template sequence, then we do multiple sequence alignment, and then if the 3 dimensional structure of that protein say like a gorilla or some other animal is available okay, then we prepare the 3 dimensional structure of our protein of interest using that particular template okay.

And then we see whether the protein which you have created is okay using a various bioinformatics checking okay, this is called the homology modeling okay once we have that okay, so we use the software there is a software called Swiss model we look at it, I just showed you the BLAST also. So we can do a homology modeling using the Swiss model okay then we can predict the active site of the protein.

And then if you have a library of ligands, you start docking all the ligands into that it is called Autodock, Gold, Glide, so many different software's commercial, non-commercial software are there, so we can make use of them, and then we get something called docking score

which tells you how good the interaction is, we can also do molecular dynamics so that is how the target based drug design happens actually okay.

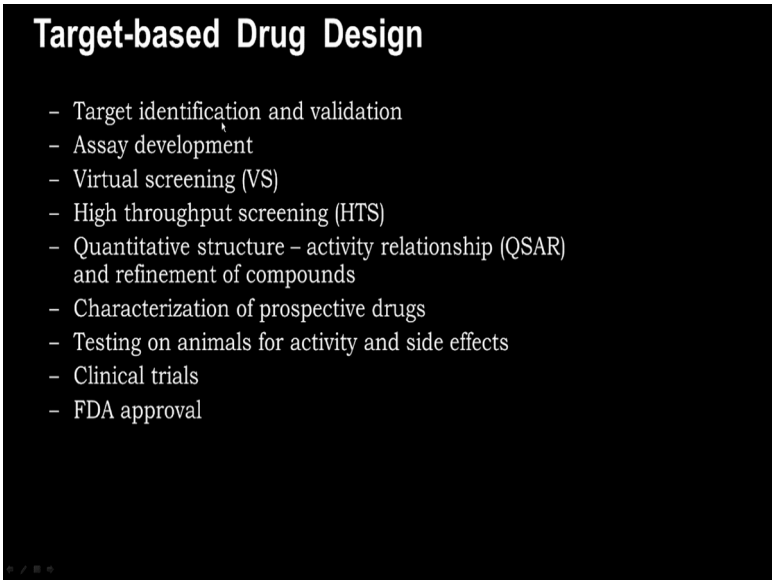
(Refer Slide Time: 06:38)



Okay, so as I showed you here, what you do? So we have a sequence A say imagine 322 amino acids, sequence B 450 amino acids. So I just showed you BLAST, so it makes a pairwise alignment of both A and B, 100 amino acids are identical, so it is 100/320 because 320, generally we always use the smallest sequence as in the denominator, so identity will be 31%. Also 23 amino acids are different by conservation substitution meaning that the chemical properties are maintained okay.

So we have 100+23/320 that comes 38.4, so that is called similarity, identity is exactly matching, similarities we also add other amino acids which may have similar chemical properties. So sequence identity is 31%, sequence similarity is 38.4%, that is how the software's like BLAST calculated, actually I took this calculation from this particular reference okay.

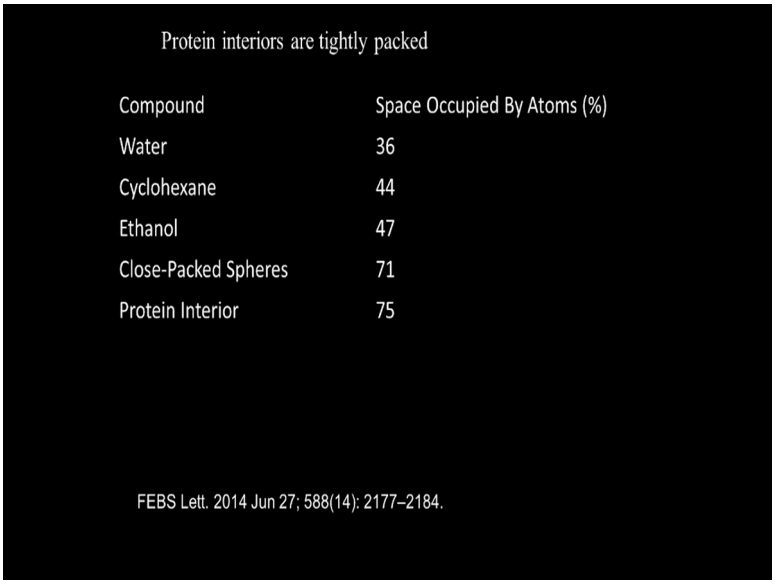
(Refer Slide Time: 07:41)



So in target based drug design, we identify target identification validation that means if I know the disease mechanism I decide on which target I focus on. We are talking about inflammation, arachidonic acid pathway, so your target could be a lipoxygenase which produces leukotrienes or it could be prostaglandins like cox-2 prostaglandin or E-synthase towards prostaglandin. And then we should also of course have experimental approach assays for measuring the interactions between both.

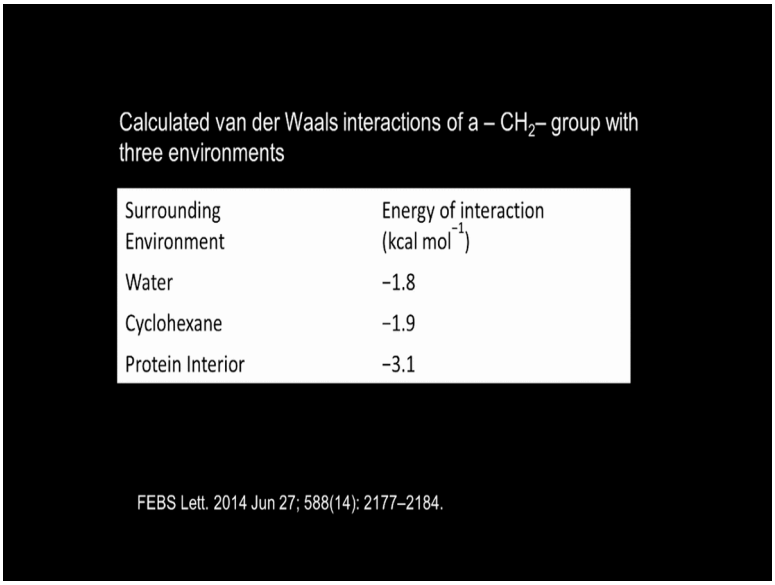
And then we need to do the virtual screening, take database of ligands like all zinc or drugs databank or so on, and then keep docking them. And then find out how docking happens, you can also do QSAR, refined the compounds, and then we can use the Lipinski's rule and other rules, shortlist compound. Then go to your animal studies, in vitro studies, clinical trials, FDA approval and so on. So this is how things happen.

(Refer Slide Time: 08:50)



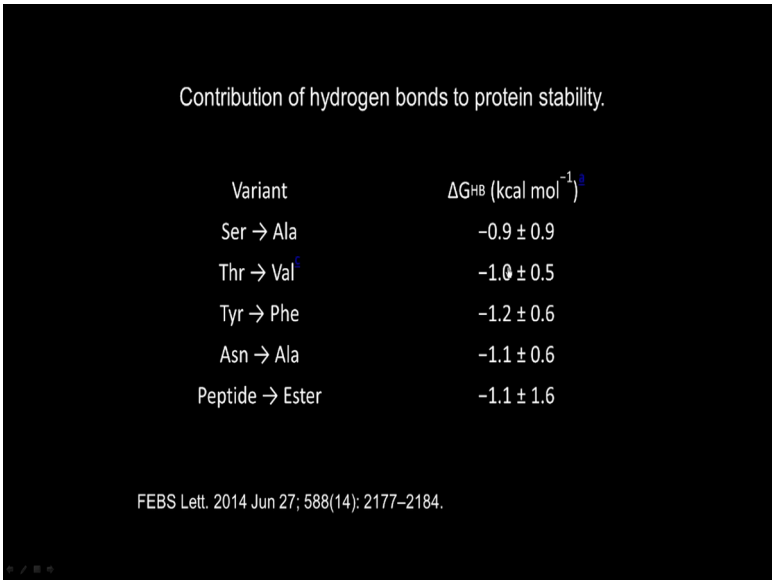
Now if you look at proteins their interiors are tightly packed okay, if you take water 36% cyclohexane 44%, ethanol 47, close-packed spheres 71%. Protein interior 75%, that means proteins are very tightly packed interiorly, so the ligands do not actually penetrate into and bind generally they go up to some active sites mostly on the surface okay or on the leaf or gap between 2 proteins and so on actually.

(Refer Slide Time: 09:25)



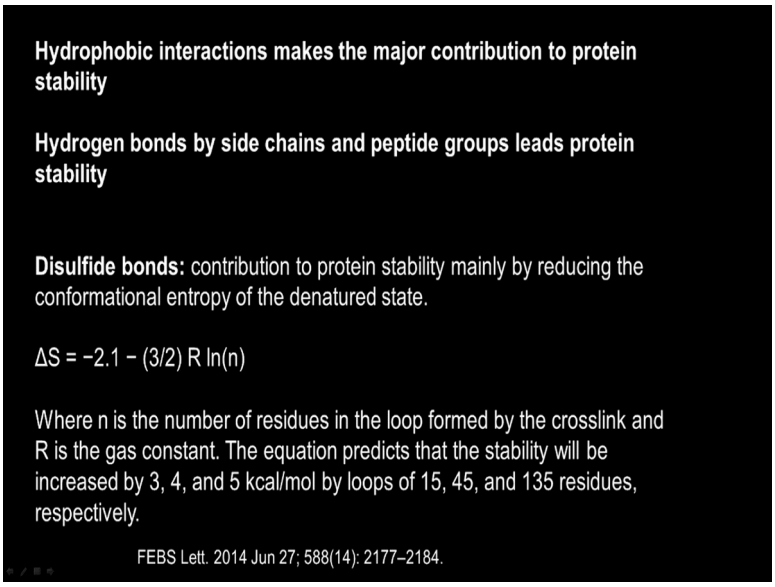
So look at Van der waal interaction of CH2 group with 3 environments okay, so CH2 interacting with water you get -1.8, CH2 interacting with cyclohexane -1.9, CH2 interacting with protein interior -3.1. So interior is generally very hydrophobic, because protein is generally taken in aqueous medium, so hydrophobic regions are always buried inside the protein okay, so that is another point you need to keep in mind.

(Refer Slide Time: 09:58)



So how does hydrogen bonds improve the stability of protein? If you look at Serine, alanine, this is about -0.9 kilocalories okay, so valine and proline -1, tyrosine, phenylalanine -1.2 and so on. So this is how the interactions is being the hydrogen bonds in various amino acids contribute to the energy okay.

(Refer Slide Time: 10:26)

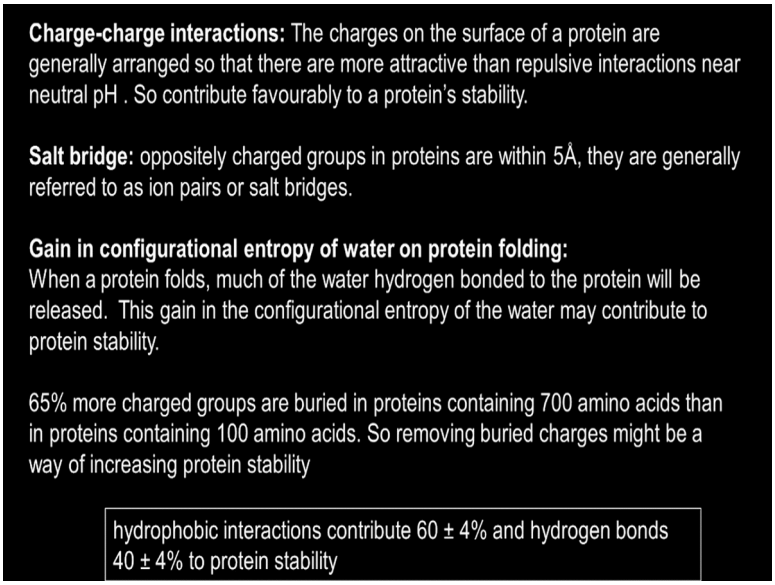


So hydrophobic interactions make the major contributions to protein stability, like I said the interiors are tightly packed and all the hydrophobic regions are inside, the hydrophilic that is the regions which interact with water are outside okay, so hydrophobic regions. Hydrogen bonds by side chains and peptide groups leads to protein stability of course okay. Disulfide bonds also are there, contribute to protein stability mainly by reducing the conformational entropy of the denatured state okay.

So it prevents the changes in the conformation of the protein okay, so that is disulphide bond okay. So if you look at the delta S=-2.1-3/2 R ln n okay. So this is number of conformation that are possible, n is number of residues in the loop formed by the cross-linked, and n is the number of residues in the loop formed with the cross-linked R is is the gas constant. So the equation predicts the stability will be increased by 3, 4 and 5 kilocalories by loops 15, 45 and 135 residues respectively okay.

So if you have a lot of residue in the loops the stability of the protein gets increased based on this particular equation, here n is the number of residues in the loop okay.

(Refer Slide Time: 11:52)

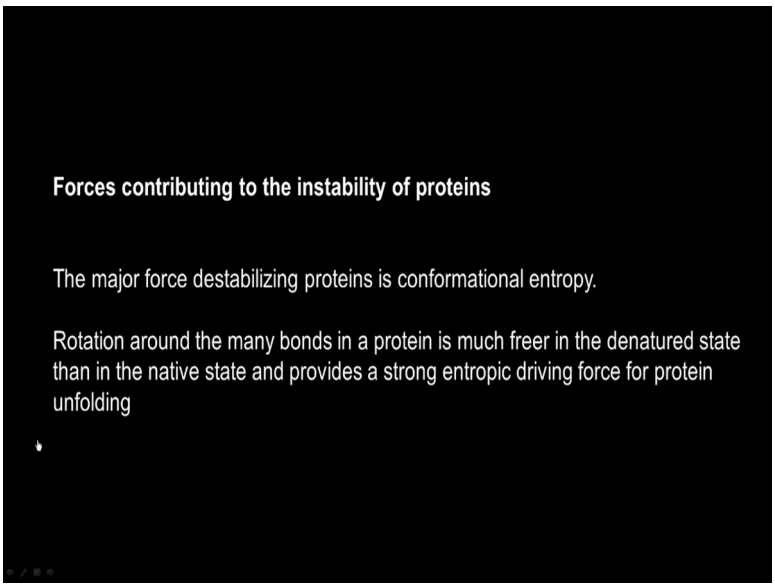


So charge-charge interactions, the charges in the surface of the protein are generally arranged, so that there are more attractive than repulsive interactions near neutral pH okay. So instead of repelling each other they will be attracting each other okay neutral pH, so they contribute favourably to protein stability. Salt bridges okay, so these are oppositely charged groups in proteins within 5 angstroms, they are also called ion pairs or salt bridges.

So salt bridges also contribute to the stability of protein okay. Gain in configurational entropy of water on protein folding, that is another point when a protein folds much of the water hydrogen bonded to the protein will be released, so this gain in the configurational entropy of the water may contribute to protein stability. Otherwise too much water is there protein can take too much of conformation okay.

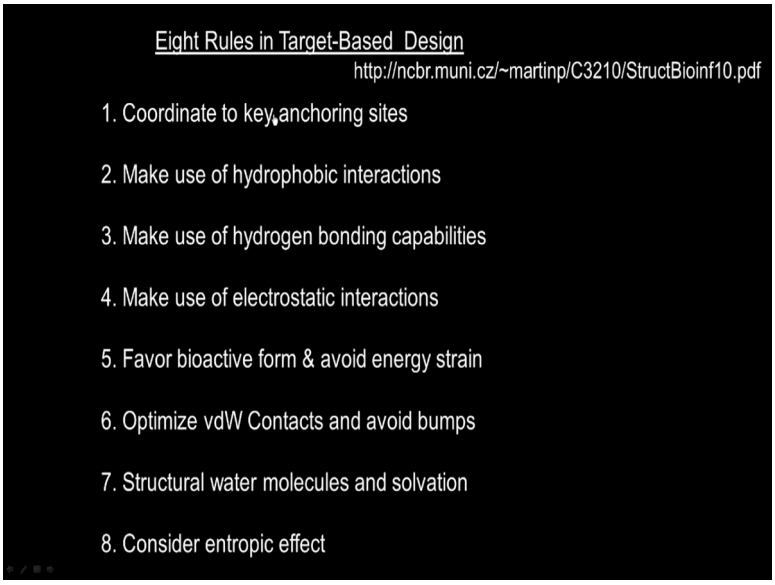
Then 65% of more charge groups are buried in protein containing 700 amino acids than in proteins containing 100 amino acids okay, so removing buried charges might be a way of increasing protein stability. So hydrophobic interactions contribute to 60% and hydrogen bonds contribute to 40% to protein stability, so these points are very important because when you are designing drugs we need to understand that hydrophobicity and hydrogen bond contributes quite a lot than many other types of forces.

(Refer Slide Time: 13:25)



The forces contributing to the instability of proteins, the major one is the conformational entropy that means rotation around many bonds in the protein is much freer in the denatured state than in the native state, and provides a strong entropic driving force for protein unfolding okay. So if it has got too many rotatable bonds, so once it gets denatured the rotatable bonds leads to a lot of conformational changes, so protein becomes more unstable.

(Refer Slide Time: 13:55)

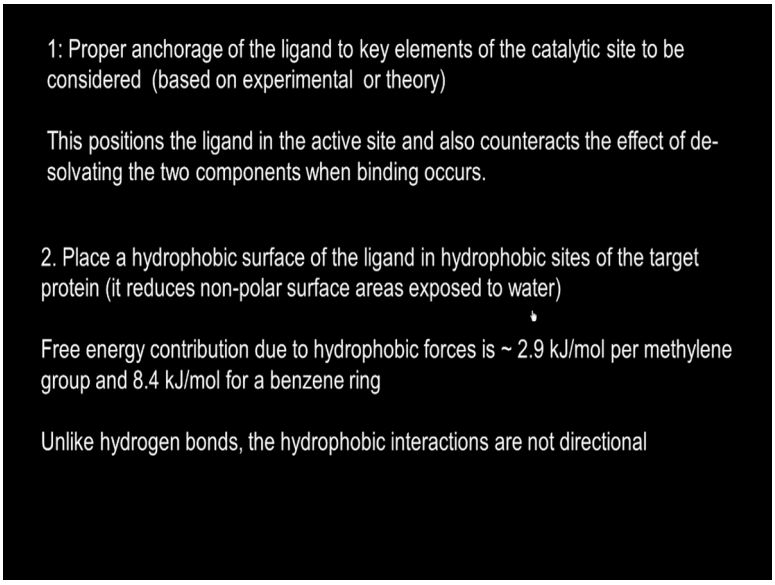


So there are several rules when we talk about target based design okay, so this is taken from this particular reference. Coordinates to key anchoring sites, so you should identify what are the key anchoring sites like sometimes they have catalytic triad for example in biotransformation, so those are the key. Make use of hydrophobic interactions, because as I said there are a lot of hydrophobic interactions and protein inside so we need to.

Make use of hydrogen bonding capabilities okay, this is the second, first is the hydrophobic hydrogen bonding, then make use of electrostatic interaction. Favor bioactive form and avoid energy strain, you should try to avoid conformation of the ligand which will lead to higher energies which is energy strain, so you should look at that drug or the ligand which has the lowest energy strain or in the form of best bioactive form.

Optimize Van der waal contacts and avoid bumps okay. Structural water molecules and solvation, you should look at structure water molecules and solvation, we will talk a little bit more. And then finally consider entropic effect we will talk a little bit more on each one of these headings okay.

(Refer Slide Time: 15:17)

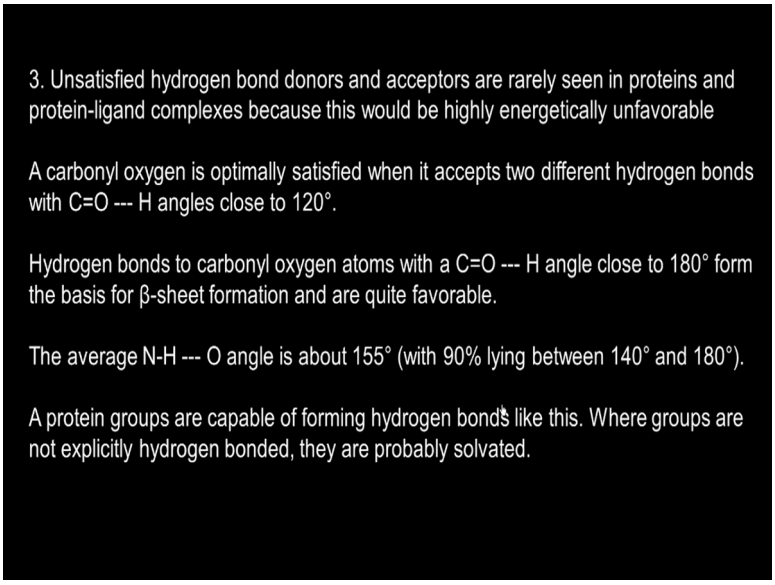


So what is this proper anchorage of the ligand okay, so we should know the catalytic site based on experimental or theory, so we should adjust our ligand so that they make use of those catalytic site. This positions the ligand in the active site and also counteracts the effect of de-solvating the 2 components when binding occurs. Place a hydrophobic surface of the ligand in hydrophobic sites of the target okay, that is very important.

Otherwise, there could be a non-polar surface exposed to water, you want to avoid that. Free energy contribution due to hydrophobic forces is 2.9 kilo joule per mole per methylene group, and 8.4 kilo joule per mole for a benzene ring okay. Unlike hydrogen bonds, the hydrophobic interactions are not a directional okay, like I talked about the pi-pi interactions and so on, hydrogen bonds are little bit more directional.

Generally, non-bonded interactions are not so directional, but hydrogen bonds are little bit directional and hydrophobic interactions are not directional okay.

(Refer Slide Time: 16:25)



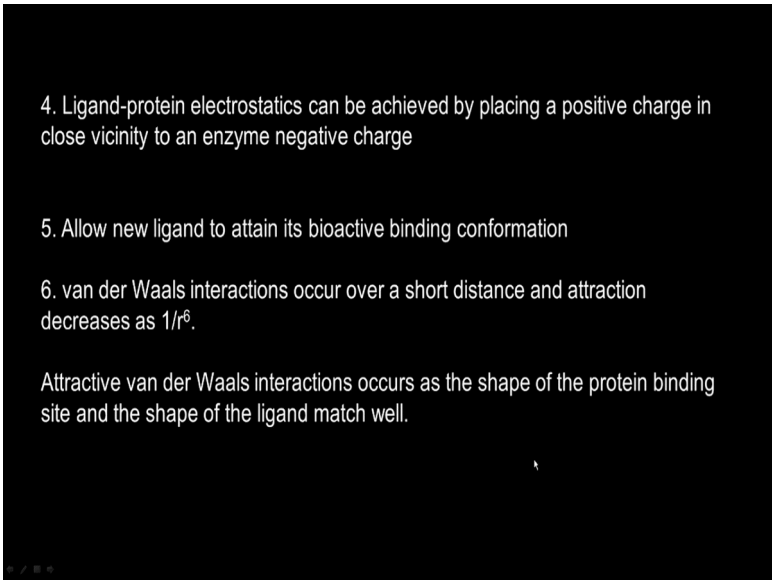
Unsatisfied hydrogen bond donors and acceptor are rarely seen in proteins and protein ligand complexes okay, because this would be highly energetically unfavourable. So if there is a ligand which has hydrogen bond donor or hydrogen bond acceptor group, you need to position that so that it has an interaction with the active site corresponding acceptor and donor and the amino acids in the active site okay.

So you will rarely have unsatisfied condition. A carbonyl oxygen is optimally satisfied when it accepts 2 different hydrogen bonds okay close to 120 degrees, this point you need to keep in mind. So if you have a carbonyl in your ligand, and if there are 2 hydrogen bond donors, so ideally if it goes and fits there, then that conformation is very suitable. Hydrogen bonds to carbonyl oxygen atoms with a C double bond O H and close it to 180 form the basis for beta sheet formation, and that is also quite favorable okay.

The average N H O angle is about 155, 90% lying between 140 and 180, so ideally the N H O angle should be 155 okay. So if you are designing ligand and if there are hydrogen bond acceptors in the amino acids in the active site, then look for that sort of angle. The protein groups are capable of forming hydrogen bonds like this, where groups are not explicitly hydrogen bonded, they are probably solvated.

So you need to remember that if there are some groups which are not hydrogen bond forming, then there could be water coming in, and they maybe interacting with the water.

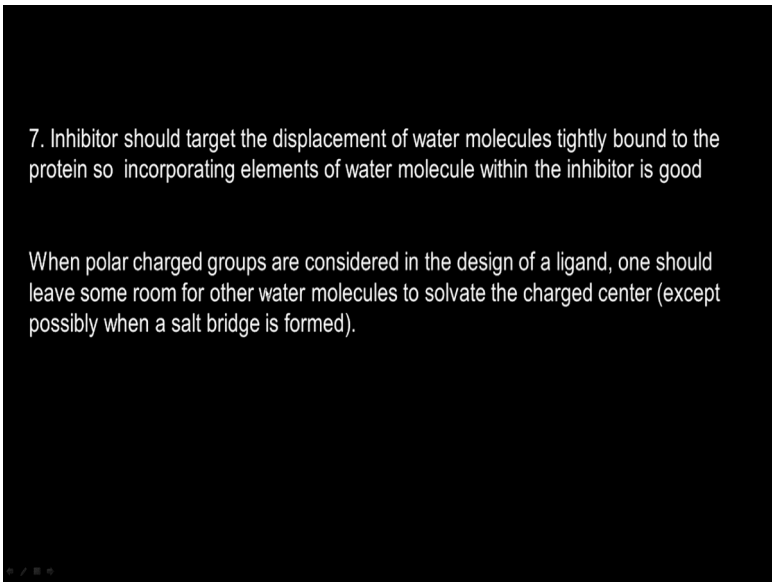
(Refer Slide Time: 18:26)



Ligand protein electrostatics can be achieved by placing a positive charge in close vicinity to an enzyme negative charge. If you are designing ligands you always find that the ligand takes bioactive conformation while binding to the active site inside the protein okay. Van der waal interactions occur over a short distance, and the attraction decreases as 1/r raised to the power 6 okay, so you need to keep that point in mind.

Van der waal interactions are a short distance they are not long distance, so you need to keep that point also in your mind actually. Attractive Van der Waals interactions occurs as the shape of the protein binding site and the shape of the ligand match well, so that means the shape comes into the picture, shape of the binding site and the shape of the ligand they should match nicely.

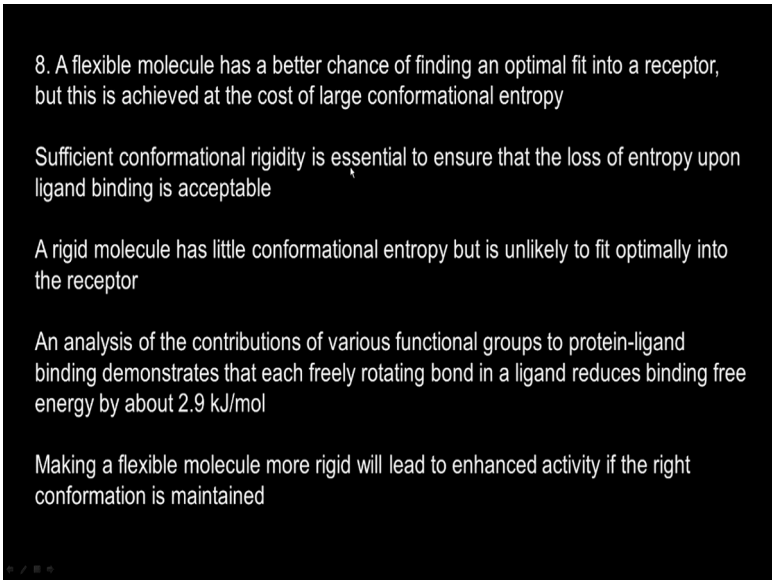
(Refer Slide Time: 19:37)



The inhibitor should target the displacement of water molecules tightly bound to the protein, so incorporating elements of water molecule within the inhibitor is good okay. So that means the inhibitors should have the capability of displacing water molecules that are quiet tightly bound okay, so ideally you should have some grooves which can displace that water.

When polar charged groups are considered in the design of ligand one should leave some room for other water molecules to solvate the charge centre except possibly when a salt bridge is formed okay.

(Refer Slide Time: 20:09)

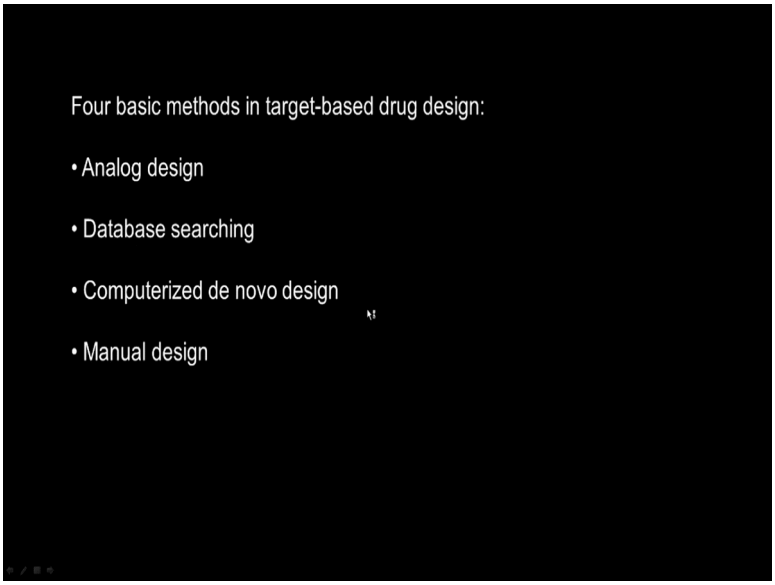


A flexible molecule has a better chance of finding an optimal fit into a receptor okay, but then if it is too much flexible it is going to be at the cost of large conformational entropy. So sufficient conformational rigidity is essential, so you need to balance this. We ensure that the loss of entropy upon ligand binding is acceptable. A rigid molecule has little conformational entropy, so but it is unlikely to fit optimally into the receptor.

So you see that if it is too much rigid, it might not fit optimally but if it is too much flexible yes it can take different conformation, but it will have too much of conformational entropy okay. Analysis of contribution of various functional groups to protein-ligand binding demonstrates that each freely rotating bond in a ligand reduces binding energy by 2.9 kilo joule per mole okay. So if you have 1, 2, 3, 4 flexible bond, it keeps on reducing by 2.9 kilo joule per mole.

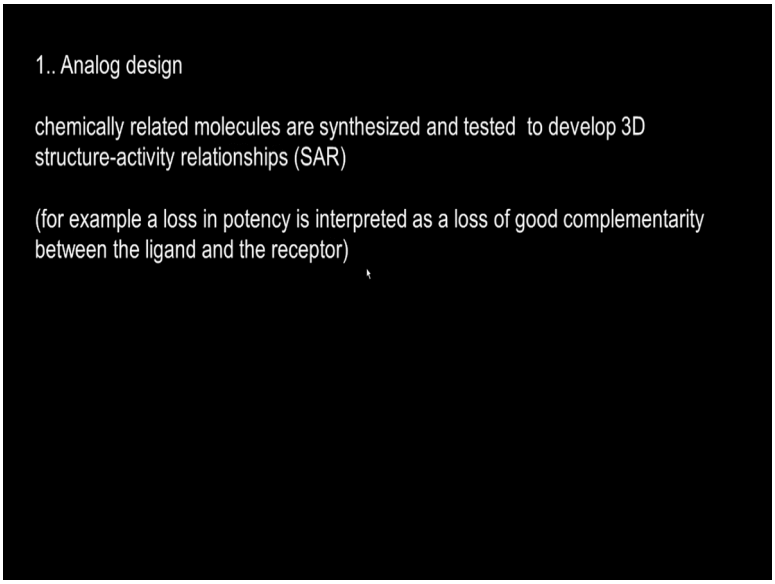
So making a flexible molecule more rigid, it will lead to enhanced activity, if the right conformation is maintained. So like I said it is good to make it rigid because you can reduce the conformational entropy increases the binding energy, but then it conformation should match with the biological active conformation as well as optimally binding to the active site of the protein. So you need to balance all these if you want to achieve a good ligand structure and shape.

(Refer Slide Time: 21:46)



So there are 4 basic methods in target based drug design: analog design, database searching, computerized de novo design, manual design. So there are 4 different approaches by which we can do this, we will look at each one of them a little bit in more detail.

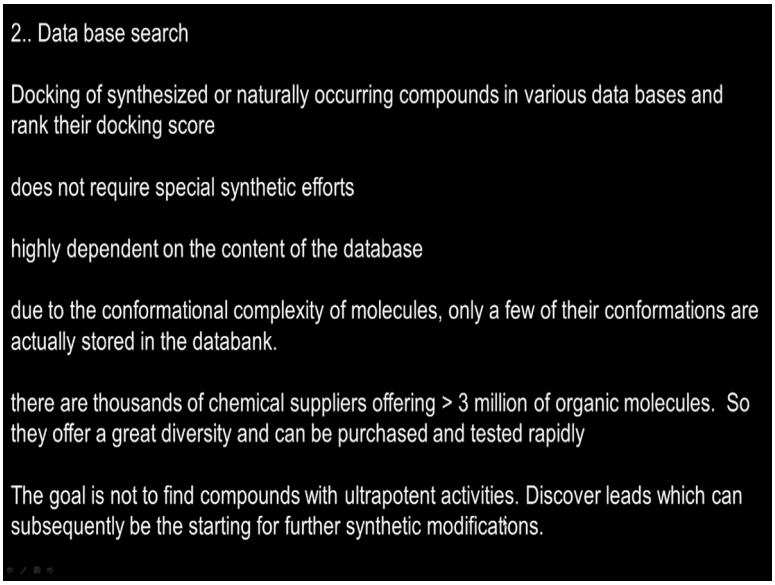
(Refer Slide Time: 22:05)



So what is this analog design, chemically related molecules are synthesized and tested to develop 3D structure-activity relationship, that is what is called analog. So you make analog of parent compound and then we develop a structure a 3D structure-activity relationship based on their structural features and activity like we said. So a loss in potency is interpreted as a loss of good complementarity between the ligand and a receptor okay.

So if the potency is going down because you removed a hydrogen bond donor, then you know that that complimenting feature that is available in the active site is not made use of okay that is called the.

(Refer Slide Time: 22:49)



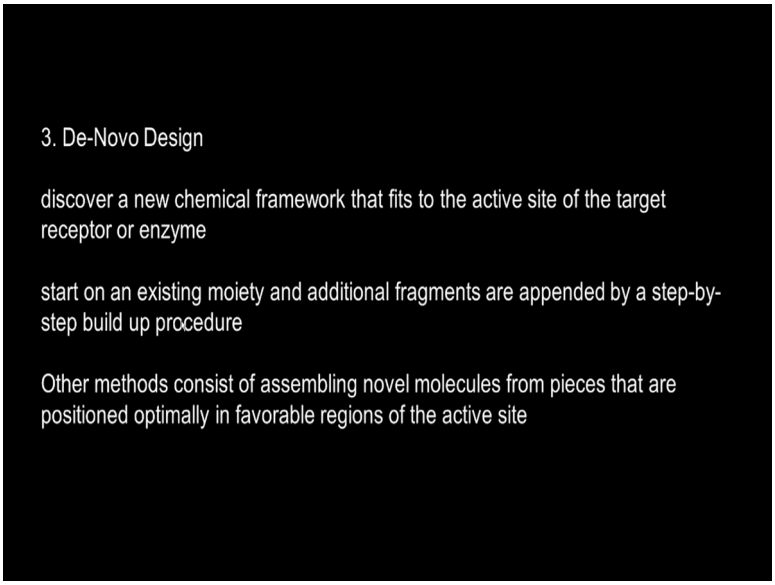
Database search, so docking of synthesized or naturally occurring compounds in various databases and the rank their docking score, so this is what we generally follow okay. So we select large number of molecules based on say Lipinski's rule or other rules, and then start docking them, and then rank them based on the docking scores. So it does not require any synthesis as of the previous analog design.

Highly dependent on the content of the database, so you have to watch out, if I select one database maybe it has got only one set of molecules, so you may miss out other set of molecules. Due to conformational complexity of molecules, only a few of the conformations are actually stored in the databank, that is also important. So a molecule can take several conformations but only few conformations maybe stored, so only those conformations will be docked to the active site.

So you may miss out some other conformations of the same molecule like another. There are thousands of chemical suppliers offering >3 million of organic molecules, so they offer a great diversity and can be purchased and tested rapidly. So database search is a good idea to follow, we can purchase them or even if you can synthesize them later on, if they are found suitable in our in-silico, and then we can test it rapidly.

The goal is not to find compounds with ultra-potent activity, but discover subsequently because they could be the starting for further synthetic modifications, that would be a sort of a lead and then we could modify the structure, so that we can improve their activity.

(Refer Slide Time: 24:30)



Third approach is called the de novo design okay, here discover a new chemical framework that fits to the active site of the target receptor or enzyme. Then basically you start on an existing moiety and we keep adding fragments step-by-step, so that maybe when you add a particular group it increases the electrostatic interaction or when you add another particular group maybe it improves the hydrophobic interaction, or it creates a good hydrogen bond donor.

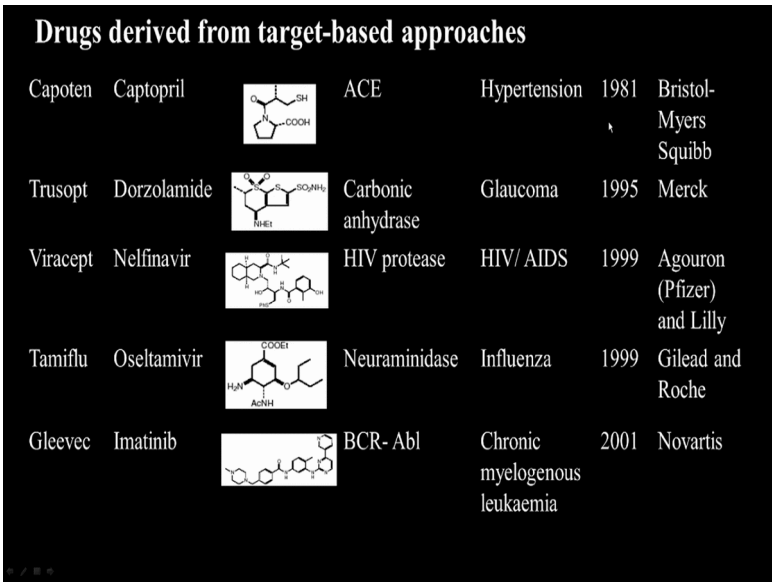
So that is called de novo design, so you start with an existing moiety and keep on improving so that you get a better interaction between ligand new design ligand and the active site. Other methods consisting of assembling novel molecules from pieces that are positioned optimally in favorable regions of the active site. So you take 3, 4 pieces and place them in different places inside the active site, so that they create a very good interaction.

And then you create by linking those pieces together that is called de novo design. So you may come up with new structures, and of course those structures whether I can synthesize them or not that is a different story we may have to check it out later with the help of medicinal chemist or organic synthetic organic chemist.

(Refer Slide Time: 25:50)

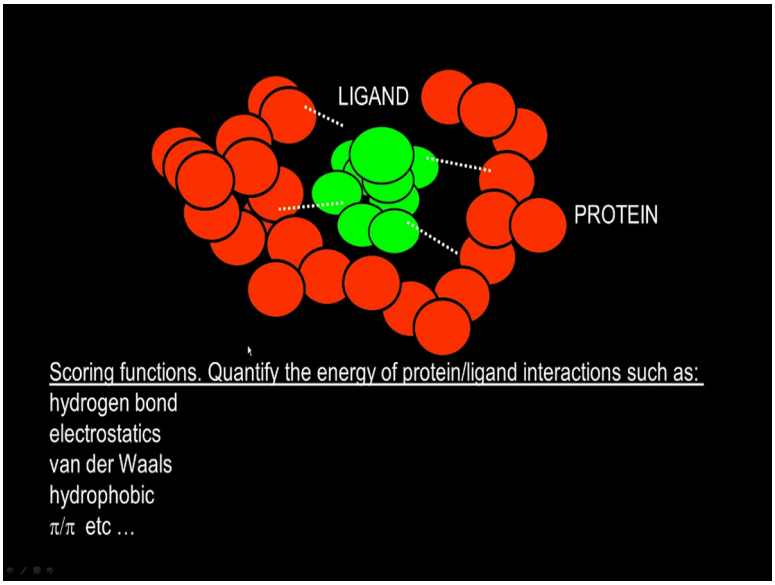
Manual design, so start from an existing solution it is consisting of the visualization of a complex okay, so you look at existing solution and then I keep modifying. This procedure is currently undertaken either by the medicinal chemist or the molecular modeler. So we have different approaches of course you can combine some of them together, so that ultimately you end up with the most favorable ligand, which has a best bioactive conformation which binds very optimally to the active site of the target okay.

(Refer Slide Time: 26:27)



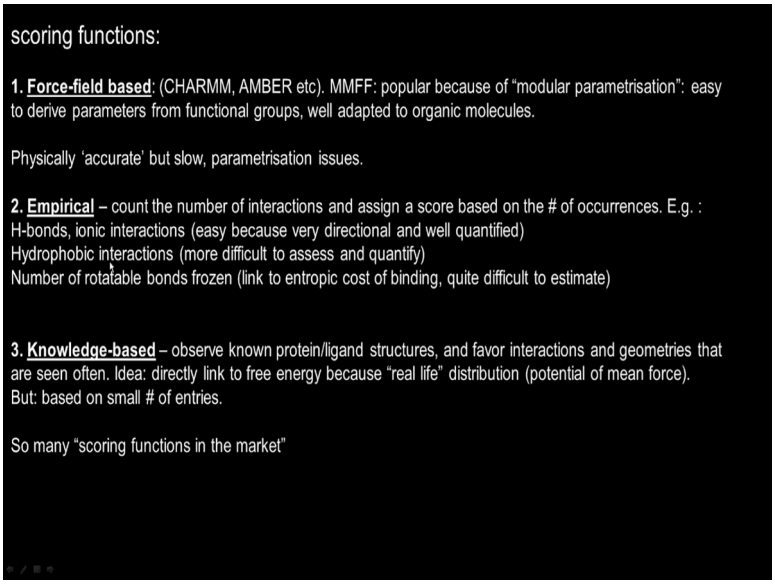
So there are many drugs derived from target based approaches. So Captopril this is a ACE okay, angiotensin converting enzyme, and it is you know Hypertension still may then Dorzolamide, it is a Carbonic anhydrase, Glaucoma Merck okay, HIV protease, Nelfinavir, Pfizer and Lilly, Oseltamivir, Influenza, Imatinib, Chronic myelogenous leukaemia, Novartis. So these are some examples there could be many more, but I am saying that it is possible to look at target based approaches okay.

(Refer Slide Time: 27:14)



So when a ligand is bound to the active site of the protein okay, you get the binding energy or interaction energy in non-bonded interaction okay. So software makes use of something called a scoring function, they calculate something called the scoring function okay based on the hydrogen bond, electrostatics, van der Waals, hydrophobic pi-pi interaction and so on. And then they report the scoring function. So higher is the scoring function, higher is the binding of interaction that is what is called actually, this is how it is performed actually okay.

(Refer Slide Time: 27:49)



The force field based as you know we talked quite a lot about force field CHARMM, AMBER different MMFF and all that. So we can have based on these force field and the results, we can calculate a scoring function. Empirical scoring function count the number of interactions and assign a score based on the number of occurrence, for example number of hydrogen bonds, number of ionic interactions, number of hydrophobic bonds, number of rotatable bonds which got frozen because of binding.

So then you calculate a scoring function. Knowledge-based, observe known protein ligand structures and favor interactions and geometries that are seen often okay, and then if you see your ligand also bind in that form give a score okay, directly linked to free energy because real life distribution, but based on small number of entries. So there are different scoring functions available in the market, each software has its own scoring function.

So the results from one scoring function and another scoring function need not match at all okay. So we will continue more on this topic of target based drug design in the next class as well, thank you very much for your time.