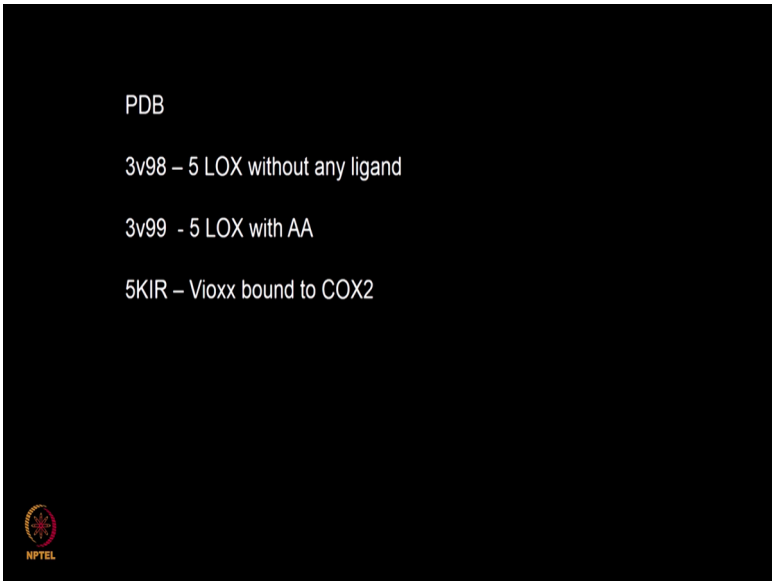
Computer Aided Drug Design
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology - Madras

Lecture - 34
Target Based Drug Design

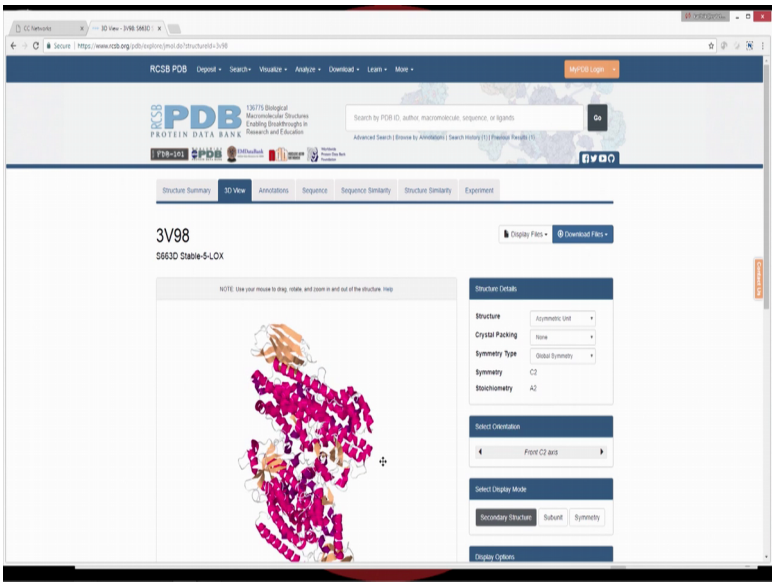
Hello everyone, welcome to the course on Computer aided drug design, we will continue on the topic of target based drug design are related to docking.

(Refer Slide Time: 00:30)



So let us again look at this protein databank that is called PDB, which contains the 3 dimensional structure of large number of proteins, it also has ligands also present in some cases, and in some cases ligands are also not present.

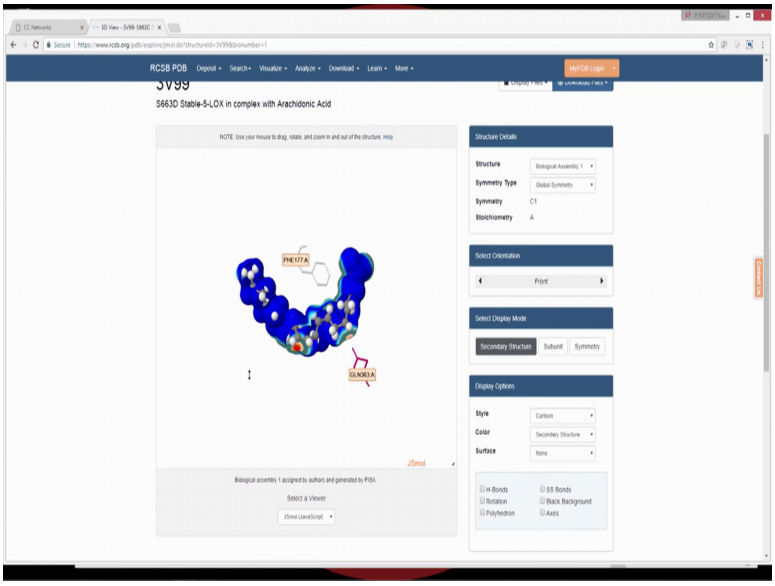
(Refer Slide Time: 00:45)



So just we will have a look at that, okay this is the PDB it is the protein databank, so we can a do a search the PDB if you know the PDB ID we can do a search for example 3v98 okay that is 3v983, that is a 5 lipoxygenase, that is an enzyme in the arachidonic acid pathway which converts arachidonic acid into HETE and so on. So we can view that under JSmol okay, so this is protein as you can see is got A and B okay.

You can see this protein it got both the A and B, now it is also got an ion okay, because lipoxygenase has an ion, so the ion is involved in the redox cycle that activates the protein and when arachidonic acid by institute and gets converted, and gets deactivated and so on actually okay. So this is the protein which does not have any ligand attached to it.

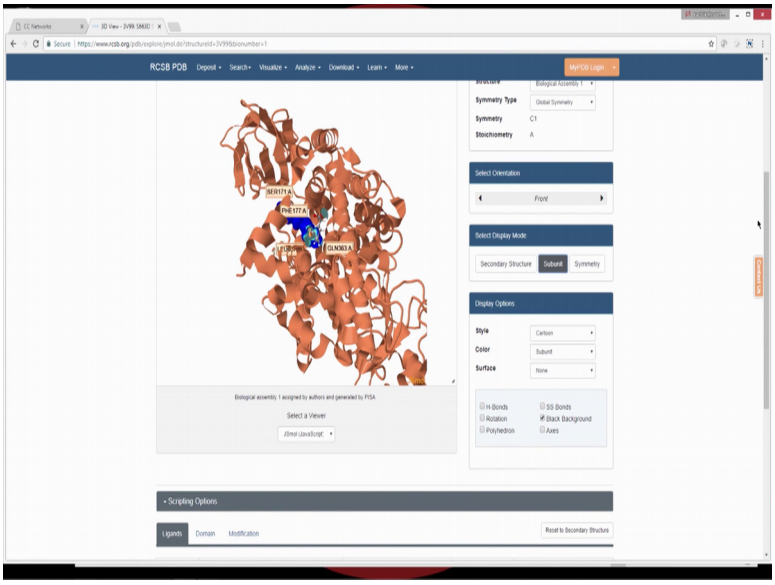
(Refer Slide Time: 02:31)



There is again 3V99 okay, that is again 5 lipoxygenase okay, it is got an arachidonic acid attached to this, so we can okay. So as you can see here, so this protein has arachidonic acid as well as the ion, so we can specifically look at where the arachidonic acid present okay, so this is the arachidonic acid. So we can see view the protein from the arachidonic acid pocket okay, so you can see that this is your arachidonic acid it is a long chain hydrocarbon.

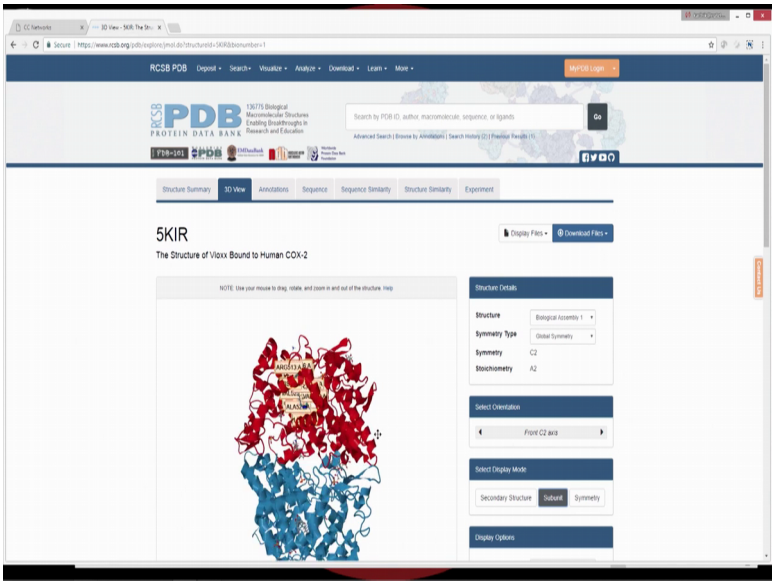
You can see that protein okay, and if you want you can even look at the ion that is present as you can see here, this is the ion that is present in the protein and with the corresponding amino acids that are surrounding there. So this picture shows how the arachidonic acid is interacting with okay with the protein okay what are they amino acids that are involved which interacts this is the arachidonic acid, this is the entire protein that can be viewed.

(Refer Slide Time: 04:23)



We can look at another protein of interest and that is called, you can see this, so you can see the arachidonic acid here and here, and the protein okay, so you can see the amino acids with which it is interacting.

(Refer Slide Time: 04:45)

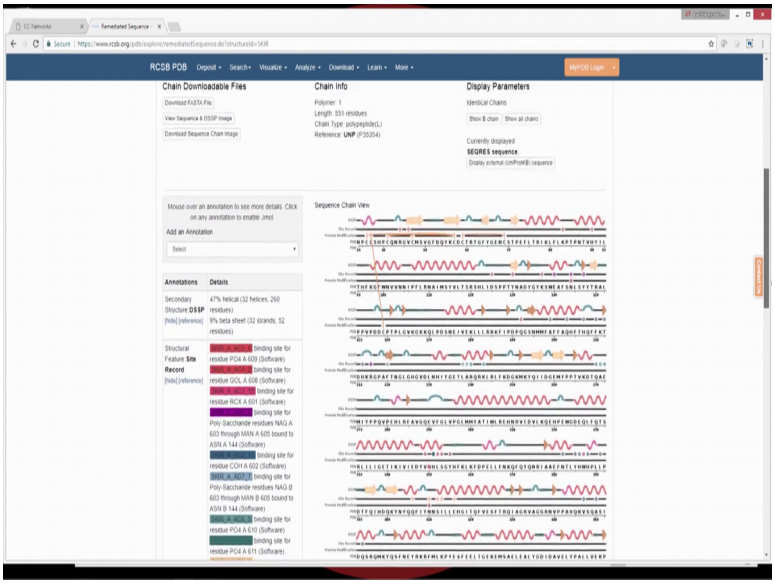


Now we can look at another protein and that is called 5KIR, this cyclooxygenase 2 cox-2, 5KIR this is cox-2 cyclooxygenase 2, and we have a vioxx this is the selective cyclooxygenase 2 drug which is bound to that protein okay, so when they crystallize this protein they put in this molecule and then they perform a crystallization. So the advantage is when we crystalize we will know where this particular molecule is present, so you will know the active site pocket and so on.

Okay, so this is the protein call the cyclooxygenase 2, so you can see there a lot of ligands which are there present in this protein. So vioxx is also called trophic oxide, so we can look at how they are present okay, so we can look at how vioxx is present as you can see here, vioxx is present inside this particular cyclooxygenase 2 protein you can see there here. So PDB is a quiet a useful protein databank, it contains large number of proteins.

It also gives you the sequence for the amino acids okay, sequence is presented in PDB 5KIR okay.

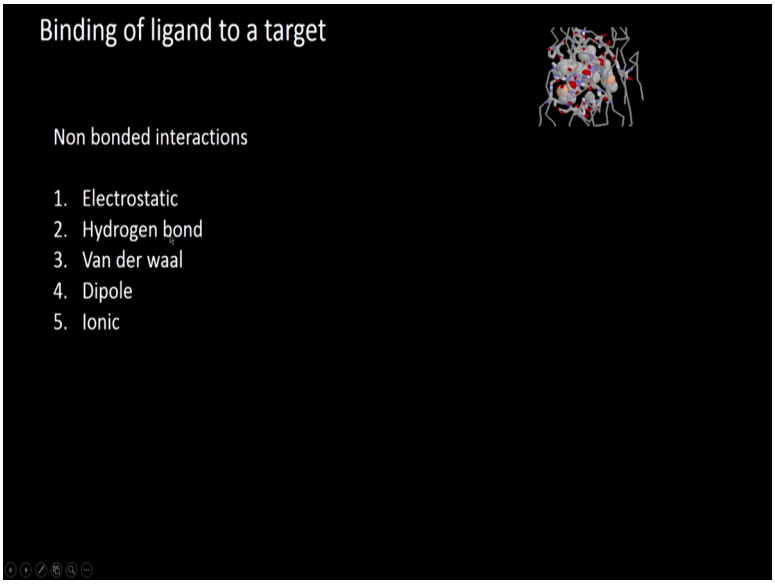
(Refer Slide Time: 06:41)



So as you can see here it gives you the sequence of this, we can also do the sequence similarity and structure similarity, so we can be a sequence similarity also using this as you can see, sequence similarity we can also do structural similarities, increasing PDB 5KIR, so we can look at the structural similarity. So this database gives you structures, so when we tried to perform docking later on we may have to download this one of the protein of interest.

And then we can now perform the docking okay, we look at docking at a later point of time. So let us go back to our presentation okay.

(Refer Slide Time: 07:45)

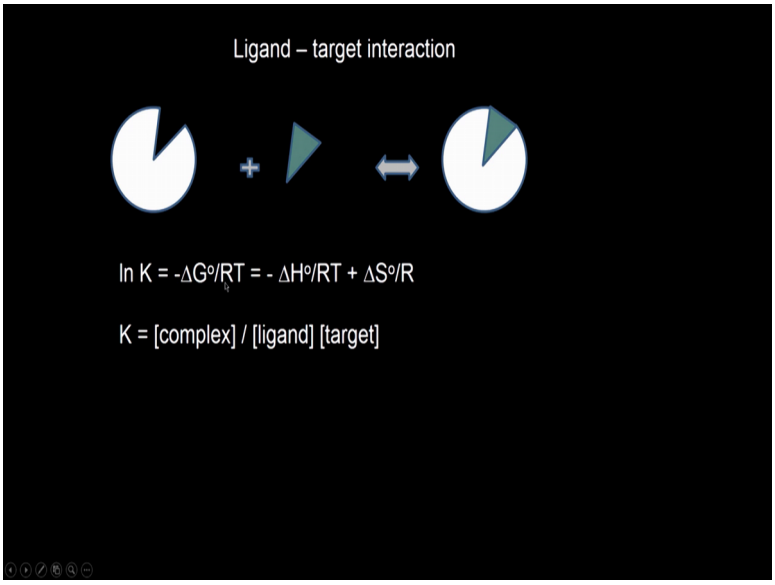


So a protein and ligand interacts generally through non-bonded interaction okay, so we have looked at what those non-bounded interactions could be long time back, electrostatic interactions okay, because you have charged so there could be some interactions electrostatically. Hydrogen bonds okay, so there could be a hydrogen bond donor, acceptors in the protein there could be hydrogen bond on that is the amino acids in the active site of the protein.

There could be hydrogen acceptors, donors in the ligand, so there could be a hydrogen bond interaction. There could be a van der waal interaction okay between the protein and ligand. There could be dipole interaction, because you have water so many functional groups which form them. And of course ionic interaction, the protein there could be here anion cationic in the active site, similarly, the ligand also could have a cation anion, so there could be ionic interactions.

So all these are called non-bonded interactions, unlike the normal covalent bonds that generally the bond energies of these non-bonded interactions are very low when compared to that of the covalent bond. But these bonds keep the ligand in place and keep it in the proper confirmation for inhibition or conversion to happen okay. If it is a light base catalyzed reaction or enzyme catalyzed reaction you are going to have a conversion okay.

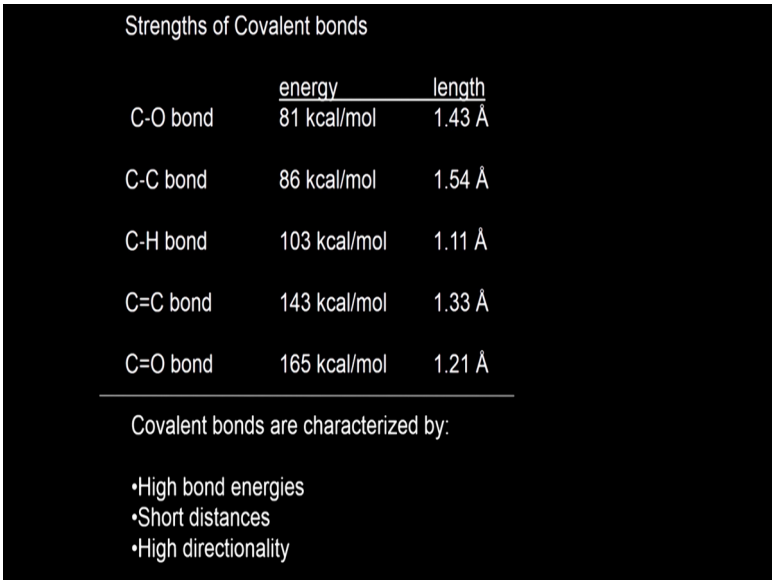
(Refer Slide Time: 09:26)



So basically what happens we have the ligand this is just a pictorial, there is something called an active site or there would be some other secondary site, allosteric site, minor grooves, major grooves, so the ligand goes and binds to that and forms ligand protein, ligand target complex. So if you look at thermodynamics ln K=-delta G0/RT, delta G0 can be written as delta H0 okay and T delta S delta H-T delta S, so that T here goes away, you have these 2 terms.

Now this K equilibrium constant is complex/ligand*target okay concentration okay, this you must have studied long time back in thermodynamics, so this is very standard okay. So from delta G we should be able to calculate K from this equation okay.

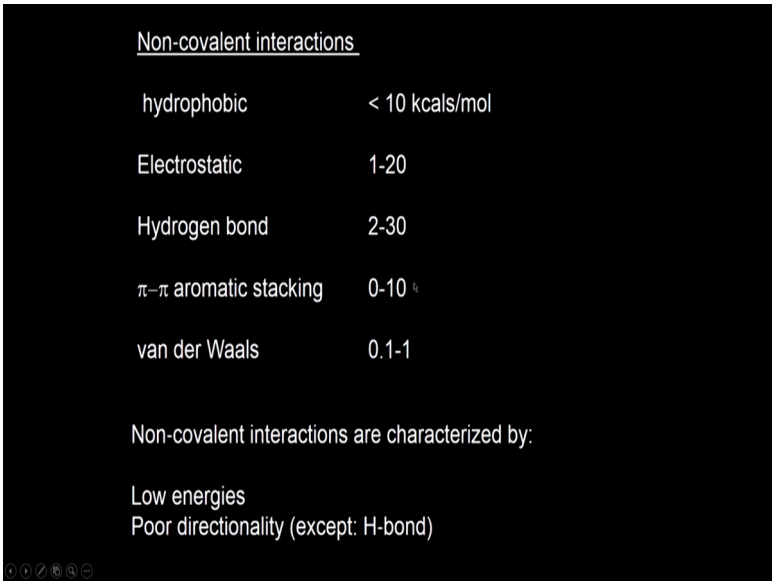
(Refer Slide Time: 10:29)



So if you look at covalent bond the strengths of covalent bond as I said they are quite large C O single bond C O double bond look at these energies, and again look at the distances, so if C double bond O, the distance is going to be much smaller than C single bound O. Similarly, C single bond C, you have 86 kilocalories, length is 1.54 angstroms, C double bond C 143 kilocalories big jump 1.33 angstroms they come closer and closer okay.

So covalent bonds are characterized by high bond energies short distance high direction, the direction is very, very focused and fixed okay, so it is not very random and as you feel like. Whereas, hydrogen bond for example, is not so directional so that is the difference between a non-bonded interaction and a bonded interaction.

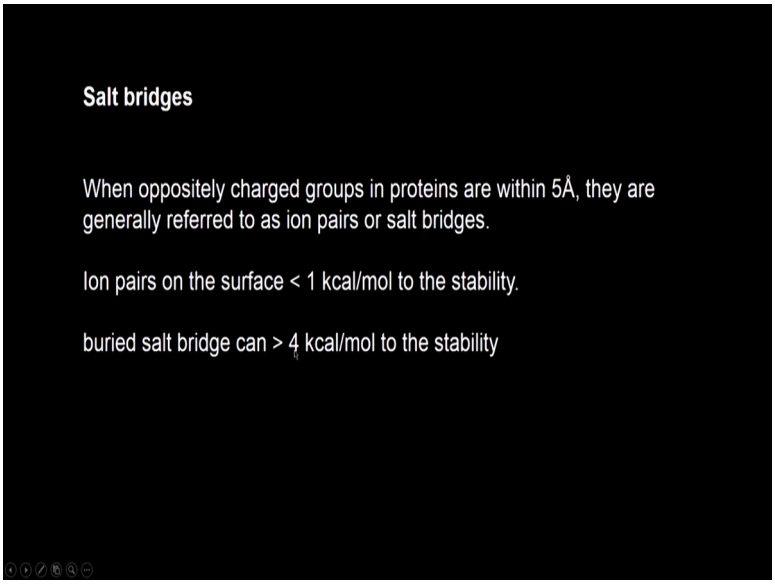
(Refer Slide Time: 11:35)



So look at the non-covalent interactions, we looked at covalent interaction, this is non-covalent hydrophobic interactions okay, that is hydrophobic groups are there right CH3, C2H5 the energies are very little <10 kilocalories look at these numbers 1 to 20 is electrostatic. Hydrogen bond 2 to:30, pi-pi aromatic stacking 0 to 10. Van der waal 0.1 to 1. So whereas look at these numbers big numbers 80 to 160, whereas these numbers are 1 to 20.

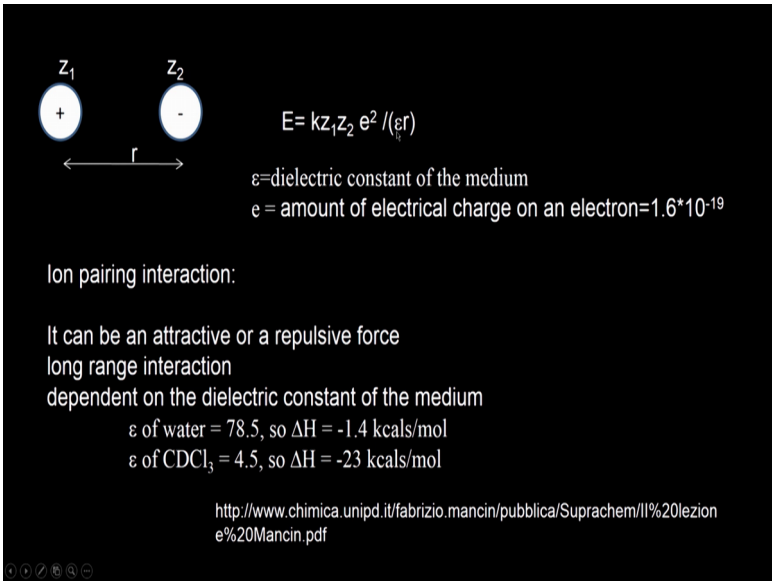
So non-covalent interactions are characterized by low energies poor directionality, direction is not very fixed unlike the covalent. Of course the hydrogen bond has some directionality, because you need to have generally AO or N forming a bond with HFO or OH or HFNH, so it is got a directionality attached to. Whereas other bonds like electrostatic or aromatic stacking it does not have much directionality.

(Refer Slide Time: 12:43)



Salt bridges when oppositely charged groups in proteins are within 5 angstroms okay, anion cation, cation anion they are generally referred to as ion pairs or salt bridges. So look at the energies there are quite low, ion pairs on the surface <1 kilocalorie per mole on to this for stability, buried salt bridge >4 kilocalories per mole to stable, so numbers are quite low okay very, very low. So ion pairs form salt bridges.

(Refer Slide Time: 13:17)

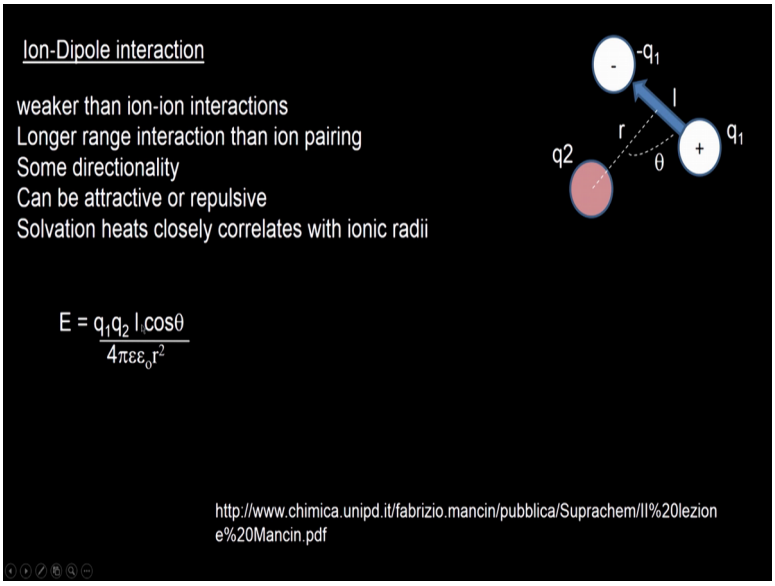


Let us once again look at these non-bonded interactions, although we looked at them in molecular modeling once again we look at. So the energy is given by some constant okay z1 z2 charges e is the amount of electrical charge on an electron that is 1.6*10 power-19, epsilon is dielectric constant of the medium that comes in the denominator r is the distance okay. The ion pairing interaction it can be attractive or repulsive force; they are long range interaction.

I mentioned this term long range when compared to short range, because you have r in the denominator. Whereas if you look at van der waal hydrogen bond you will have r raised to the power 6, r raised to the power 12, r raised to the power 10 in the denominator, so they are called short range. Of course this is dependent on the dielectric constant of the medium, so dielectric constant can vary dramatically.

For example, if you look at water 78.5, if you look at CDCL3 it is 4.5, so big difference okay almost 20 times difference. So this is taken from this particular reference, so the dielectric constant of the medium can change the energy of attraction or repulsion between 2 charged molecule.

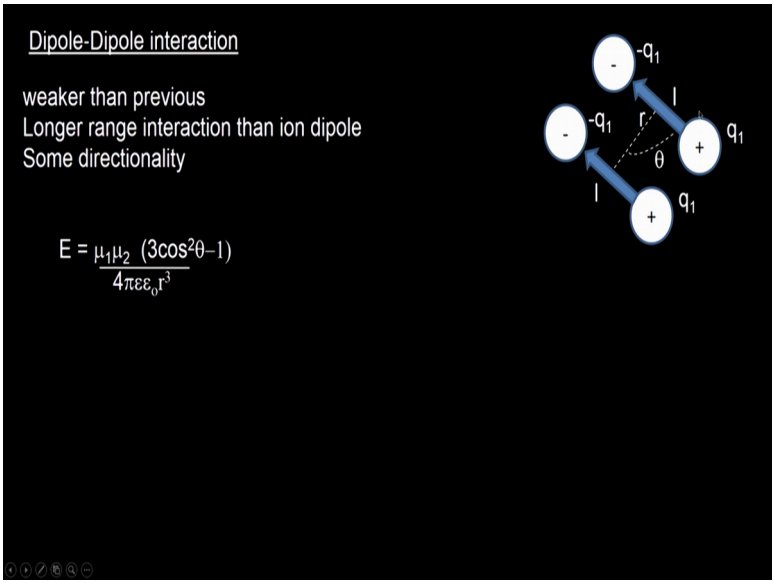
(Refer Slide Time: 14:50)



Ion-dipole interaction, so you know what is dipole for example, water HOH, so the HS are all forming + like a magnet small magnet O is forming -, so it is like a small magnet with - and +, so that is the dipole and then you have the ion of certain charge. So these are weaker than ion-ion interactions, long range interactions then ion pairing, some directionality is there because the interaction it can be attractive or repulsive.

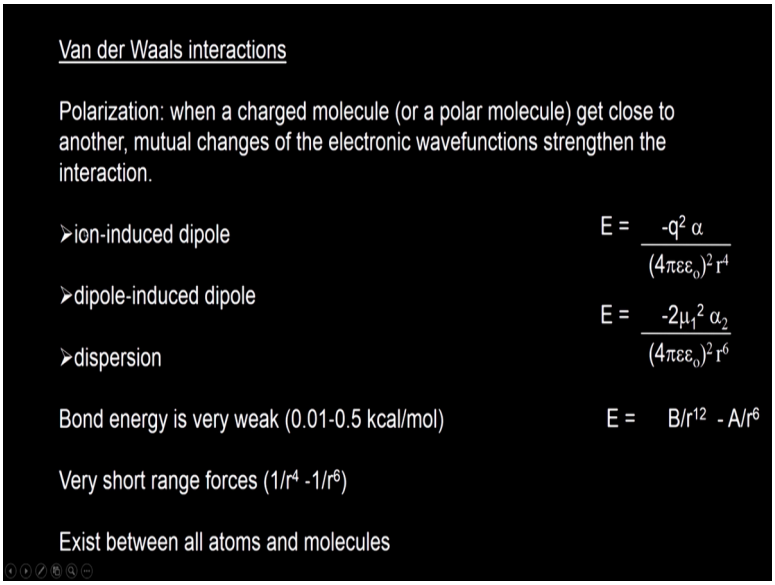
Solvation heats closely correlates with ionic radii, so this is the equation for that. So we have a suppose you have q1 and q1 okay which is the dipole + and this is- and then we have a ionic okay. So the q1*q2 l is the distance cos theta/4 pi epsilon epsilon0 r squared, this is called the ion-dipole interaction.

(Refer Slide Time: 15:53)



Dipole-dipole interaction, so you can have 2 dipoles they are very weak, they are long range interaction, now you have r raise to the power 3. Some directionality is there again okay called the dipole-dipole interaction.

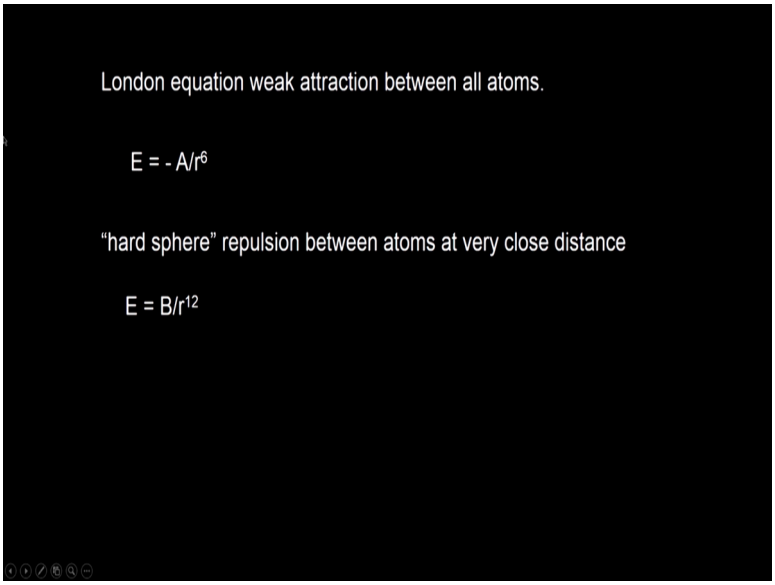
(Refer Slide Time: 16:07)



Of course van der waal interaction you all know right. Polarization: when a charged molecule or a polar molecule gets close to another, mutual changes of electronic wavefunctions strengthen the interaction. So we can have ion-induced dipole, dipole-induced dipole, dispersion, all these. So they are very short range because you have 1/r power 4-1/8 power 6, of course A will be there, B will be there, exist between all atoms and molecules.

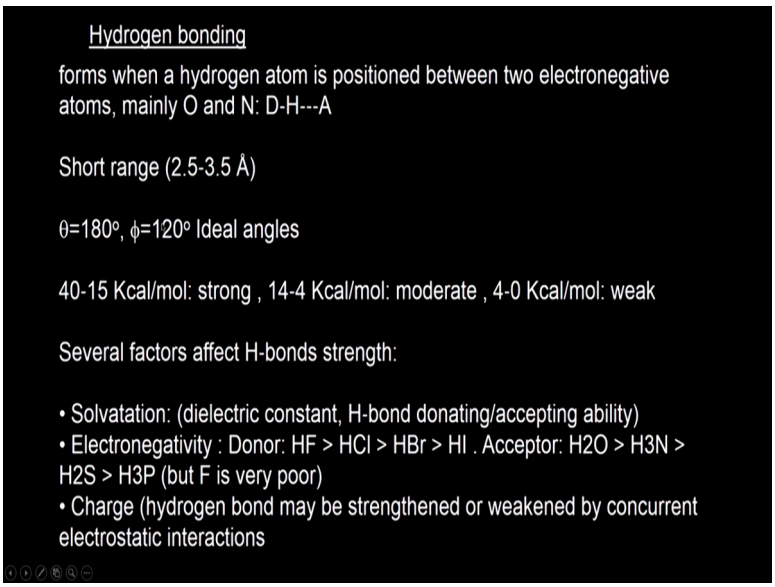
So like this know B raise to the r power 12-A raise to the r power 6 and so on. So ion induced dipole will be like this, so you have r power 4 coming here, the dipole dispersion forces will be like this r power 6, so this is called the van der waal interaction.

(Refer Slide Time: 16:56)



Then we have the London equation that is weak attraction r power 6, and this is the repulsion between atoms are very close distance r power 12.

(Refer Slide Time: 17:07)



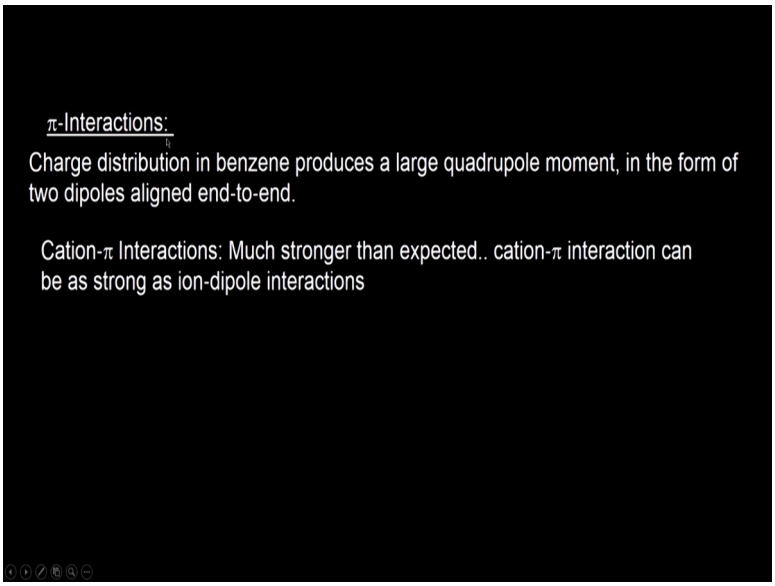
So we have different types of effect based on r okay for an electrostatic we have divided by r, for ion-dipole interaction we have divided by r squared, for dipole-dipole interaction we have divided by r power 3, for van der waal, London dispersion and then and repulsion we have a power 6 and power 12 in the denominator. Then comes hydrogen bonding this is a very

important parameter which comes forms when hydrogen atom is position between 2 electronegative atoms mainly oxygen and nitrogen okay.

So we can have like this you know a donor with H sorry a donor with the H and then an acceptor. These short range distances we are talking in terms of 2.5 to 3.5 angstroms, theta=180, pi=120 ideal angles, 40 to15 kilocalories per mole that is strong, 14 to 4 kilocalories per mole when moderate, 4 to 0 kilocalories per mole for very weak. Several factors affect hydrogen bond: solvation, dielectric constant of the medium, hydrogen bond donating accepting ability.

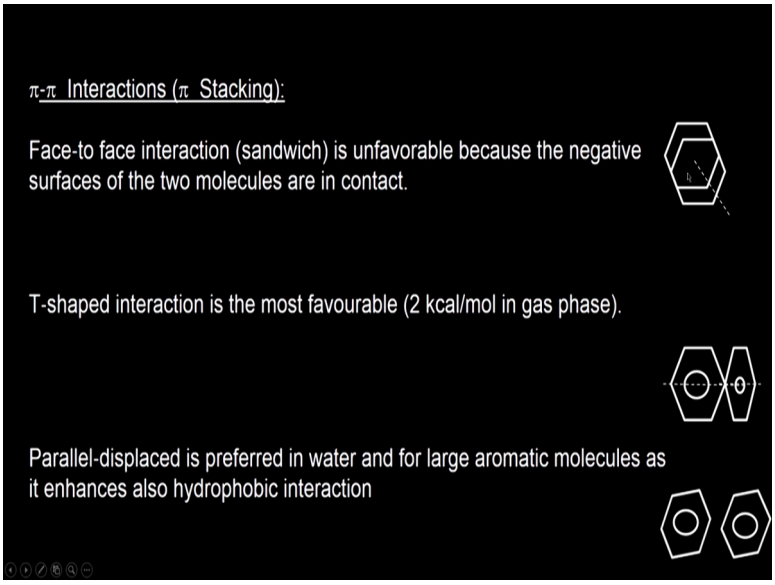
Electronegativity of the donor okay HF > HCL > HBr > HI, acceptor water > H3N and that is an ammonia H2S > H3P, but F is very poor. Charge, hydrogen bond may be strengthened or weakened by concurrent electrostatic interaction okay, so if the concurrent electrostatic interactions are there the hydrogen bond interactions can become weaker or stronger.

(Refer Slide Time: 18:50)



Then we have pi interactions, this is that benzene rings you know we can have 2 benzene rings charge distribution in benzene produces a large quadrupole moment, in the form of 2 dipoles aligned end-to-end. So we can have cation pi interactions, this is much stronger cation pi interactions may as strong as ion-dipole interaction okay.

(Refer Slide Time: 19:14)

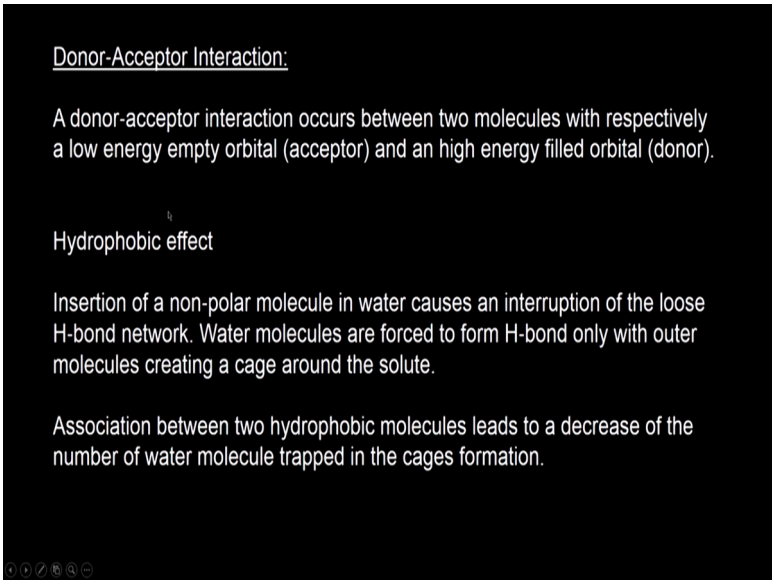


And you can have different types of benzene rings, face to face interaction this is very unfavourable, because the surface is generally negative. So one benzene ring has a negative surface, another benzene ring has negative surface, so the interaction is very poor, this pi-pi interaction is also called pi stacking. T-shaped interaction is a most favourable, that means one benzene ring is flat, another benzene ring is perpendicular like T energy is 2 kilocalories.

And of course this is the most favor parallel displaced, that means one benzene is flat and other benzene is on not exactly on top displaced, but parallel and water for large aromatic molecules as it enhances also hydrophobic interaction okay, this is the most liked actually is a especially prepared in water, and the T type of interactions is also favored, but this generally we may think one on top of another, that is the least favored because we have the negative charge.

There could be a negative charge, charge repulsion between 2 benzene ring. So this is also a quite an important non-bonded interaction need to consider.

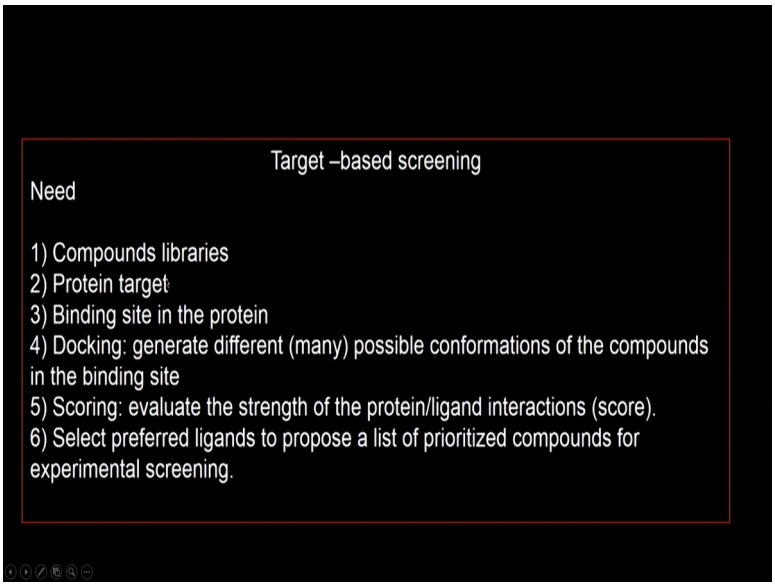
(Refer Slide Time: 20:27)



Then we have donor-acceptor interaction occurs between 2 molecules with respectively a low energy empty orbital acceptor and a high energy filled orbital that is donor okay, so it gives you the electron into the acceptor. Insertion of a non-polar molecule in water causes an interruption of the loose hydrogen bond ne2rk, then the water molecules are forced to form hydrogen bond only with outer molecules creating a cage around the solute right.

If you put in a small hydrophobic molecule, there is a cage like leaving the hydrophobic molecules and the other water molecules start interacting okay. Association between 2 hydrophobic molecules lead to a decrease of the number of water molecules trapped in the cage formation okay, so if you have a 2 hydrophobic molecules they start associating with each other so water molecules gets moved out actually okay, they are not trapped.

(Refer Slide Time: 21:26)



So when we do the target based drug design, and we need to know, they have a compound library, we need to get the protein target, protein target we can get from the PDB, compound library I told you so many compound libraries we can get them all to file or SDF file of the molecule, we can get the confirmation energy confirmation. We need to know something about the binding site of the protein.

So because a protein can have many sites EA substrate based site, active site, allosteric site, minor groove, so we need to know which site the ligand is going to bind to, so if you are doing a competitive inhibition, so obviously then the ligand has to be competing for the substrate of the protein. If you are looking at allosteric non-competitive, then you look at other sites. So deciding on the site also tells you what types of drugs you are planning to design.

Then you dock, dock means the ligand goes and binds to the active site, it forms an non-bonded interaction, we have the non-bonded interaction energy. This depends upon the various confirmation the ligand can take, so if the ligand is very flexible it can take different confirmation into that active site, so it can take the most favored confirmation, so that it goes and fits nicely into the active site, based on the various non-bonded interactions I talked about.

So a very flexible molecule can take different confirmation, so it can go and fit into nicely the site okay. Then many software's perform a scoring it calculates a number okay based on just like your marks, it calculates based on the energy of the non-bonded interaction that is the binding energy okay or interaction energy. So many software's have different ways of calculating the scoring function, so higher the scoring function then binding we can say is better.

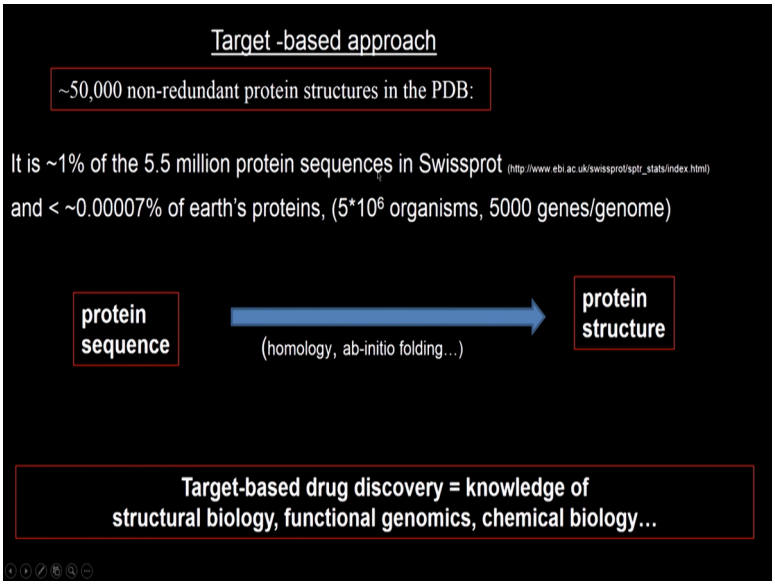
So if you have large number of molecules you keep binding, binding, and then look at the scoring function, and then say these molecules have very good scoring functions, this do not have good scoring function. Or if the software does not produce scoring function, it also gives you the binding energy, so more negative is the binding energy, we can say better is the interaction, we can also look at other interactions how many hydrogen bonds it has formed and so on.

And then we can conclude that certain ligands have good binding to the active sites, certain ligands are poor. And then select the preferred ligands okay, so ligands which you think bind very well, then prepare a list of prioritized compounds maybe you start synthesizing them and then start studying the experimental activity. Of course in between you may also have to look at Lipinski's rule and other ADME rules to see whether those molecules satisfy the drug likeness property or biocompatibility property.

We looked at so many different rules, so we can also shortlist molecule. So if you have said thousands of molecules, we can either do the Lipinski's rule and ADME rules first and shortlist molecules, then perform the docking and then select the top 10 or 20% of molecules which have very good docking score or bind very well with high binding energy, and then take them for experimental synthesis as well as experimental biological evaluation using cell line or whatever biochemical assays which you have planned.

And this is how the target based screening is carried out actually, so we can combine both the structure based that is the ligand based screening and target based screening and ADME rules screening altogether when we talk about in silico drug design okay.

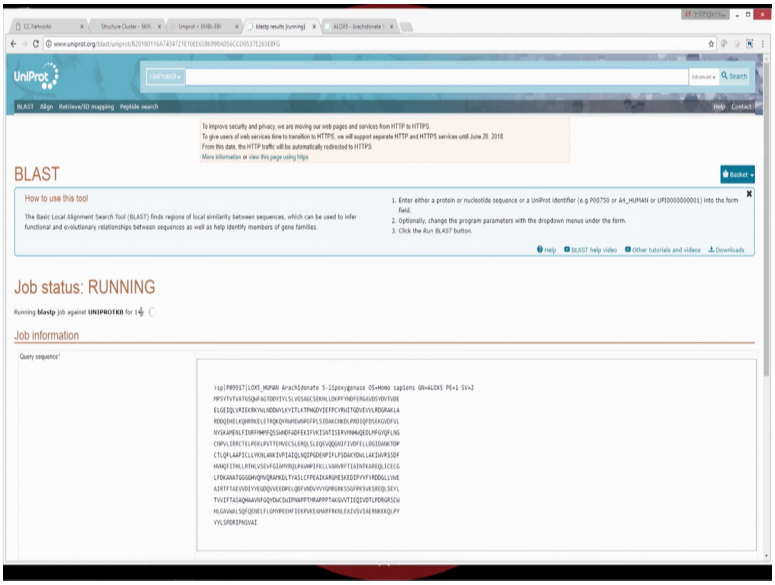
(Refer Slide Time: 25:54)



So what do we do we start with the look at the PDB protein databank like I showed you it contains more than 50000 non-redundant protein structures, but then it is only 1% of the 5.5 million protein sequence in swiss-prot okay, so that swiss-prot is web database which contains sequences of proteins, that means it is a 2-dimensional sequence. Whereas PDB contains 3 dimensional structure.

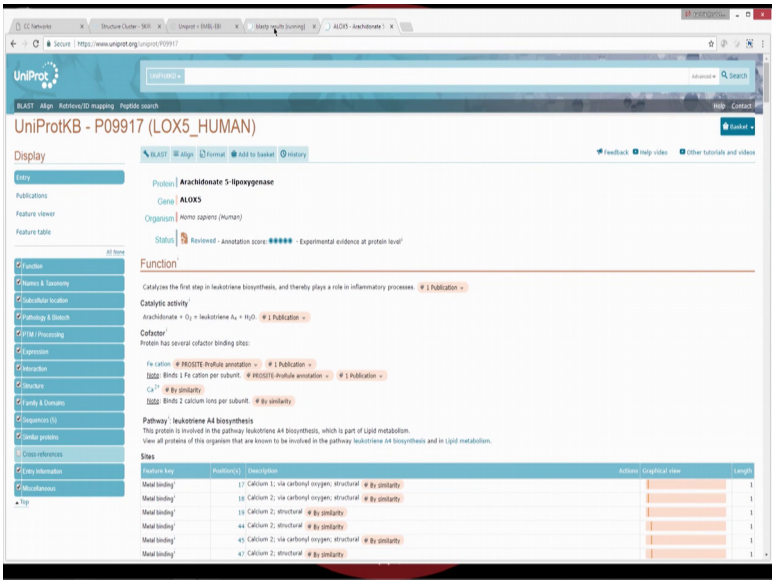
So that means a lot of proteins have still not been crystalized and they are 3D structures have not been obtained, so there is still good opportunity for us to do a crystallization study and get the details 3-dimensional details of proteins okay. So let us look at the swiss-prot okay.

(Refer Slide Time: 26:44)



So this is swiss-prot or uniprot which gives the sequence of a large number of proteins okay, and then we can summons them and compare them okay, so the 5 lipoxygenase is the sequence is given actually okay, sequence of 5 lipoxygenases is given. So it is trying to look at proteins different 5LOX proteins and trying to get their sequences that is the 2-dimensional sequences okay.

(Refer Slide Time: 29:05)



This I got the uniprot number that is called okay this is the uniprot number of lipoxygenase is human 5 lipoxygenase is called P09917, okay the job is still running. So we can look at various proteins 5LOX proteins which has sequence similarity of this particular protein actually, okay it is still running it will take a long time, because there could be a lot of proteins of 5 lipoxygenase, this is from human which we got, this is from human which we got okay.

There could be many proteins from different species okay. So what it means is there are only 50000 proteins structure 3 dimensional structure in the protein databank, but this swiss-prot has almost 5.5 billion protein sequences that is 2 dimensional protein sequences which is only 1%. So we still do not have 3 dimensional structure of many, many proteins okay, and that this number is only a very, very small amount of proteins on the earth.

That means there are about 5 million organisms each organism has 5000 genes, so that is a very small number that means we do not have information about large number of proteins, almost 99.99% of proteins found in the earth we do not have actually okay. So we will not find the 3 dimensional structure of many proteins involved in many diseases for humans, so we may find for other species, so we need to perform something called a homology modeling.

That means we from the protein sequence of say human related protein, we try to see whether any protein 3 dimensional structure of protein is available in PDB, and then we perform a homology model and get the structure okay. So target based drug discovery is knowledge of structural biology, functional genomics, chemical biology and so on actually. So, so many areas combined together, we need to have some knowledge to get the target based drug design okay.

So if you again go back, let us see whether the job has been done, the job is still running because there are too many proteins. So we will continue more on the target based design in the next class as well, thank you very much for your time.