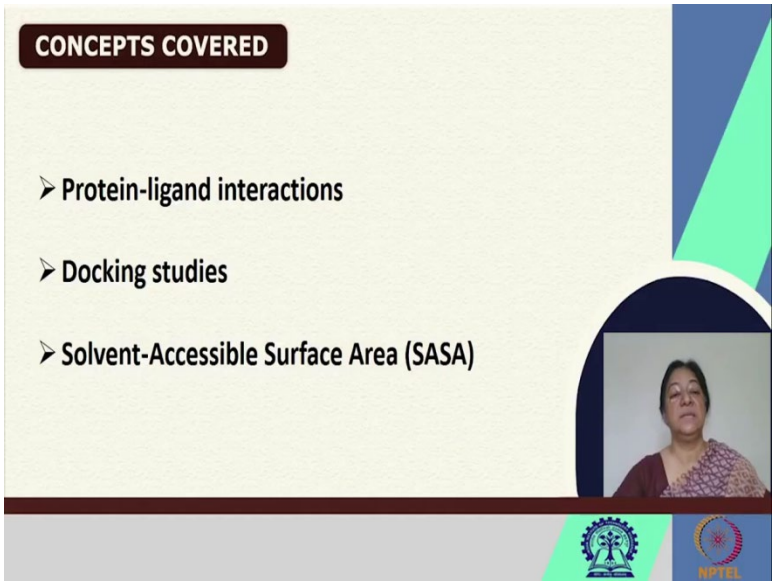
Fundamentals of Protein Chemistry
Prof. Swagata Dasgupta
Department of Chemistry
Indian Institute of Technology, Kharagpur

Module - 05
Protein ligand interactions
Lecture - 24
Protein Ligand Docking

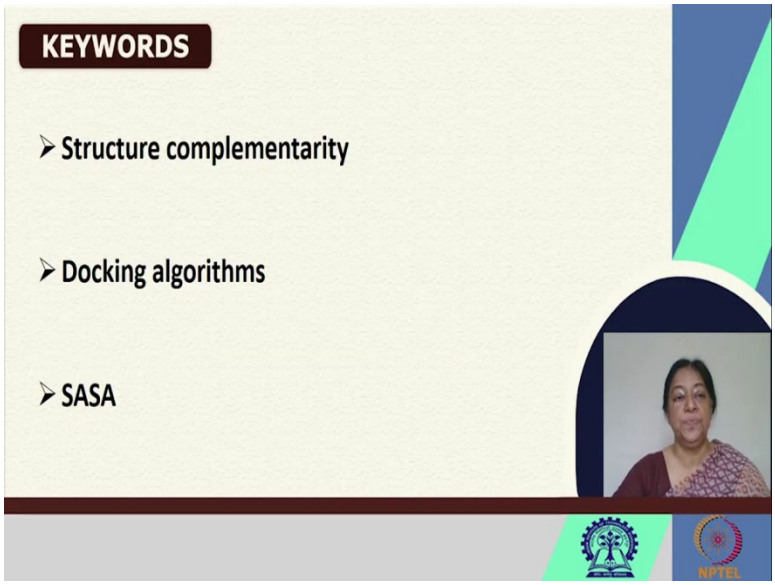
In this lecture we will be looking at protein ligand docking studies, as part of protein ligand interactions.

(Refer Slide Time: 00:23)



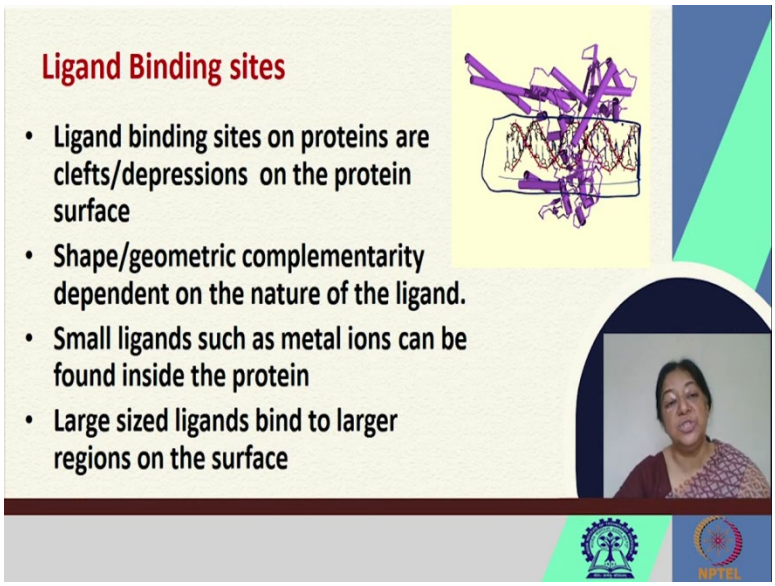
In our discussion here, we will be looking at protein ligand interactions from a different perspective, looking at docking studies and Solvent-Accessible Surface Area. We will understand what these terminologies mean as we go along in the lecture.

(Refer Slide Time: 00:39)



The keywords that we will look at here are structural complementarity, docking algorithms and what is known as SASA, as a Solvent Accessible Surface Area.

(Refer Slide Time: 00:51)

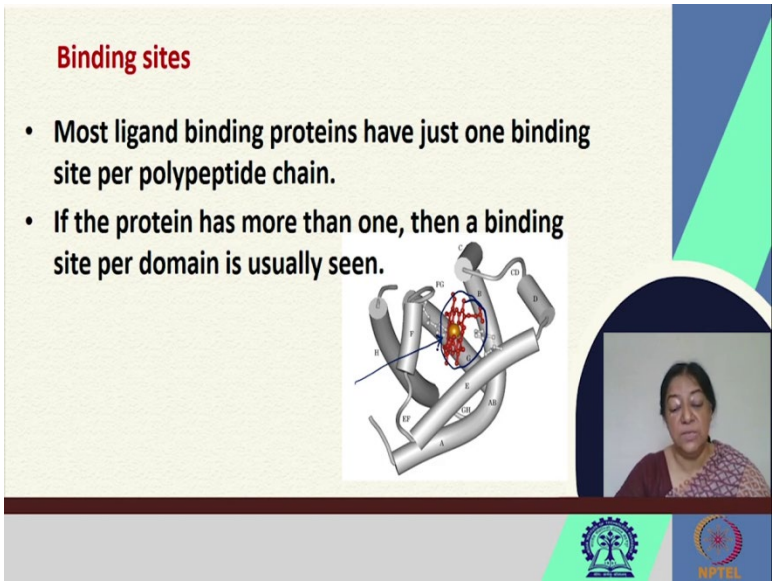


We have seen about protein ligand in the previous lectures. Now we will look at the ligand binding sites in a bit more detail, in the sense that we understand that the ligand binding sites on proteins are clefts or depressions on the protein surface to account for their binding.

The importance of the binding comes from the shape and the geometric complementarity as well as having small ligands such as small molecules, like oxygen or carbon dioxide or even metal ions that are found inside the protein, as they have a network or channel that takes them from the surface to the center of the protein or where their active site is.

Larger sized ligands on the other hand, tend to bind on larger regions on the surface. For example, if we look at a protein and we look at a DNA binding of the protein, we see that this binds through a larger surface on the protein molecule.

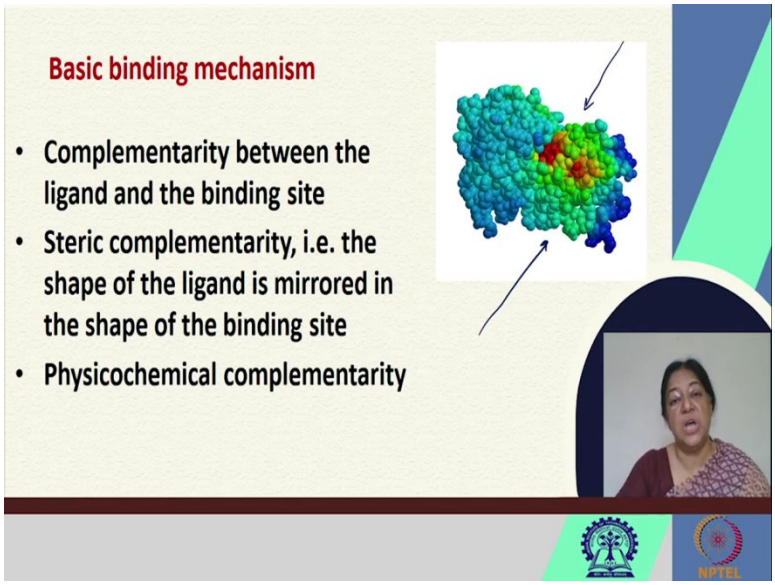
(Refer Slide Time: 01:58)



Most of the ligand binding proteins have just one binding site per polypeptide chain. And if the protein has more than one, then we have a binding site per domain. For example, if we look at the structure of myoglobin given here [refer to slide], which we know is a monomeric chain, we have the porphyrin ring the heme group rather here, that has the iron molecule that binds the oxygen.

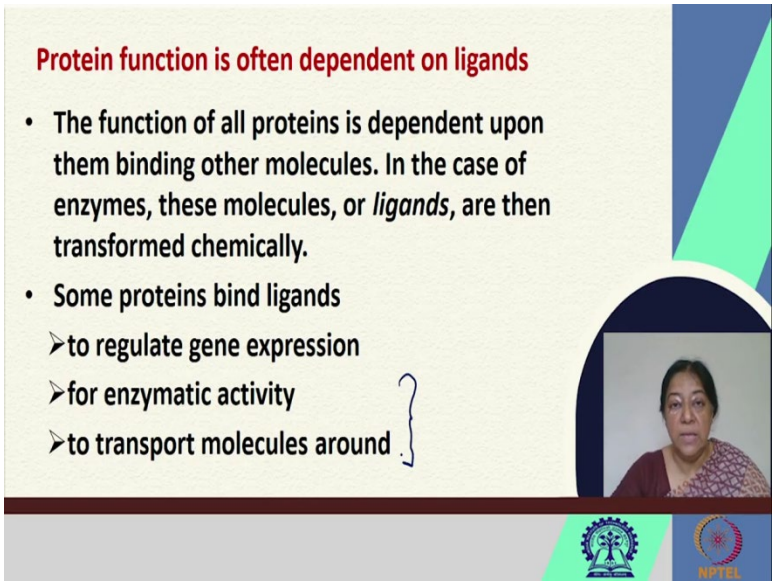
If we look at the hemoglobin that is a tetrameric molecule, similar to myoglobin it has four of these immigrants bound to the hemoglobin.

(Refer Slide Time: 02:38)



The basic binding mechanism therefore, requires a complementarity between the ligand and the binding site, there is a geometric complementarity. How do we expect a molecule or a ligand to bind to a specific region on a protein? So whether we have the molecule bound here [refer to slide] or we have the molecule bound here, there is a geometric complementarity that we have to look at, that is the shape of the ligand is actually mirrored on the shape of the binding site.

(Refer Slide Time: 03:19)

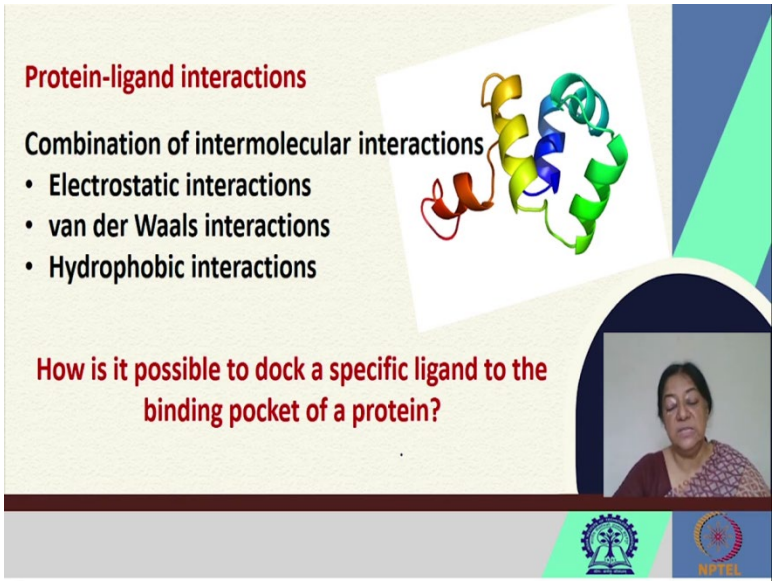


We also have physicochemical complementarity, where we are looking at the chemical structure. Do we have a favorable interaction? What kind of interactions are we looking at? Now protein function is often dependent on the ligands. And what kind of function do we have? The function of all proteins is dependent upon them binding other molecules.

In this case of enzymes, these molecules or ligands have transformed chemically, which we will see in the next module, where we will be looking at enzyme substrate complexes, which is a subset of protein ligand complexes. Some proteins bind ligands for enzymatic activity, others bind to regular gene expression or as we looked at human serum albumin in the previous lecture, they bind to transport the molecules around.

Regulation of gene expression is the most common reason why a protein would bind the ligand. The others being, for an enzymatic activity and also to transport molecules around. Whether they are drug molecules or whether they are compounds that have to be taken from one region to another or we are even binding oxygen, where hemoglobin and myoglobin are working in the body.

(Refer Slide Time: 04:37)



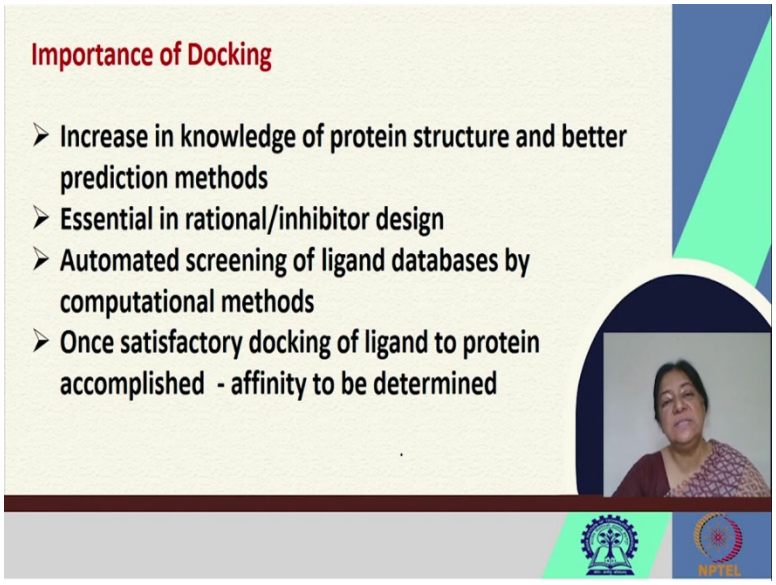
When we look at protein ligand interactions, we understand that there is a combination of intermolecular interactions. So we have a geometric complementarity, we have a chemical complementarity. We have electrostatic interactions that need to be favourable, van der Waals interactions and hydrophobic interactions that are going to bring our protein and ligand together to form a complex.

And we looked at the equilibrium constant of this complex, we also looked at the free energy of binding of the complex and we realized why we would need to have a spontaneous binding to bring about a protein ligand interaction. Now the question is, that if I have a multitude of protein structures available, as we know there are available in the protein data bank and I want to design an inhibitor for an enzyme or I want to design a molecule that is going to bind to a specific protein in a stronger manner. Then, given that I know a part of the active site, given I know a substrate molecule that binds to an enzyme, I can design a molecule. I go back to the writing board, where I draw the molecule.

I draw the molecule and with the computational facilities available now, we can have what we call the three dimensional structure of the molecule. And since we already have the three

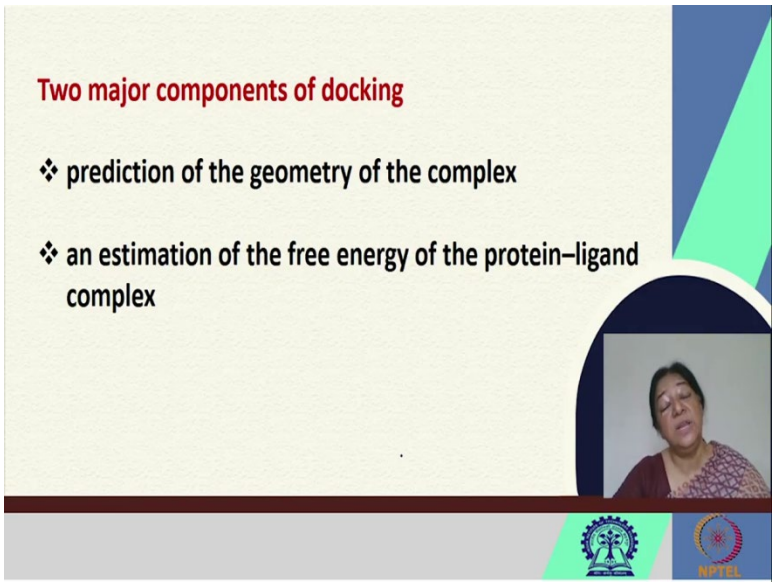
dimensional structure of the protein of interest, say it is available in the protein data bank, we can now use this to dock our specific ligand to the binding pocket of a protein. In the rest of the lecture we will see how this is possible.

(Refer Slide Time: 06:31)



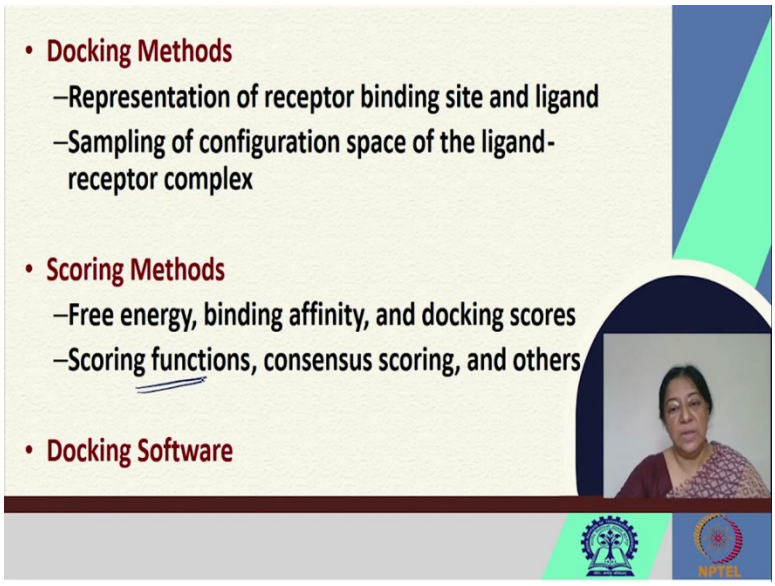
The importance of docking lies because there is an increase in the knowledge of protein structure and better prediction methods from protein structures. It is essential in rational inhibitor design, we have the automated screening of ligand databases by computational methods and we can find the affinity for a specific ligand to a protein without going for its synthesis right away.

(Refer Slide Time: 07:02)



The two major components of docking are the prediction of the geometry of the complex, now the protein ligand complex. And an estimation of the free energy of the protein ligand complex. There are several docking algorithms or several docking software that we will just briefly discuss and in the estimation of the free energy, some software give you a score. This score is a measure of how good your ligand is fitting into the protein site, protein active site or any allosteric site as well.

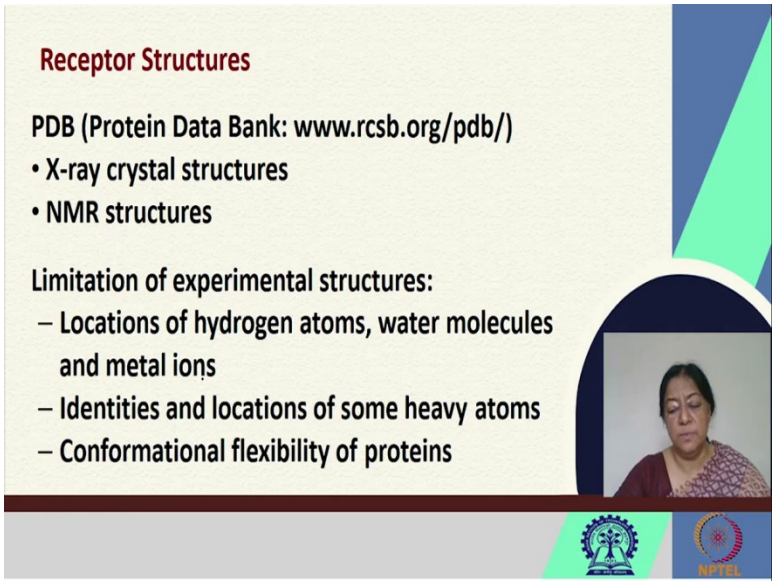
(Refer Slide Time: 07:44)



The docking methods look at a representation of the receptor binding site and the ligand. It samples the configurational space available and we have scoring methods. The scoring methods look at free energy, the binding affinity and docking scores. So there are specific scoring functions.

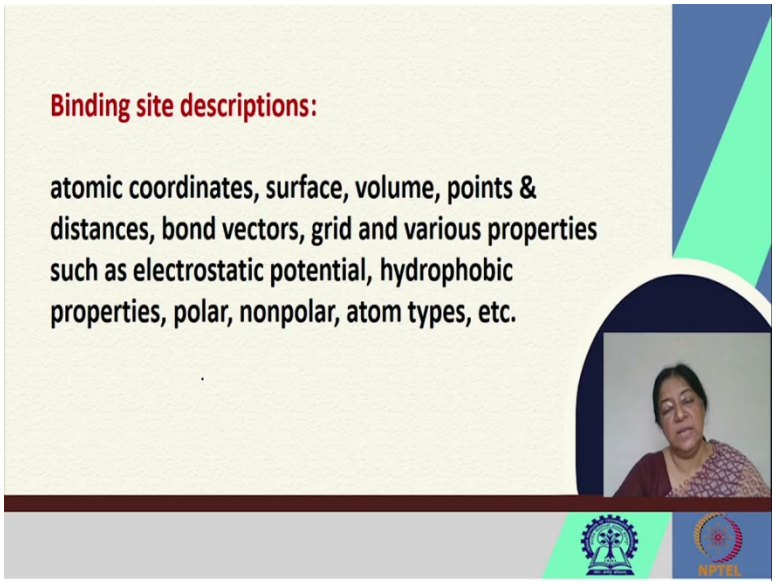
From the scoring functions, we can even bind the substrate or the ligand that we know has a high affinity to the protein of interest and look at the docking score or the scoring function. Based on that designer molecule, that would also have a similar or even better scoring function and the docking software that are going to be available to do all this.

(Refer Slide Time: 08:33)



The receptor structures are available in the protein data bank, where we look at X-ray crystal structures or NMR structures. But the limitations there lie in the locations of the hydrogen atoms, the water molecules and the metal ions. When we are going to do a docking of our ligand molecule and see whether there is possible hydrogen bonding, we would like the location of the hydrogen atoms and also the overall charge on the residues in the protein, in the way they would exist in a solution. The identities and locations of some heavy atoms and the conformational flexibility of proteins, that is always not taken into consideration in the computational software, because of the fact that it gets extremely computationally intensive.

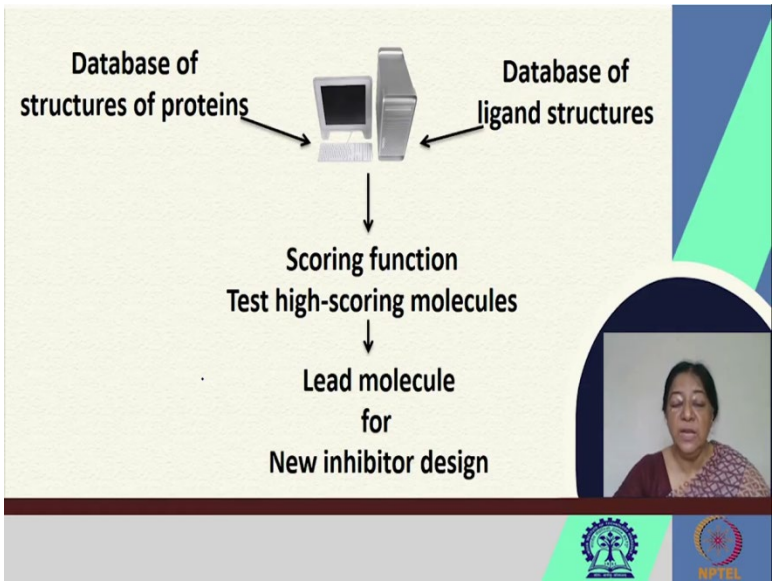
(Refer Slide Time: 09:21)



In the binding site descriptions, we see the atomic coordinates, a surface volume, the points and distances, bond vectors, grids, and various properties related to not only the binding site, but also to the ligand. So we are looking at the electrostatic potential of the protein at that particular site

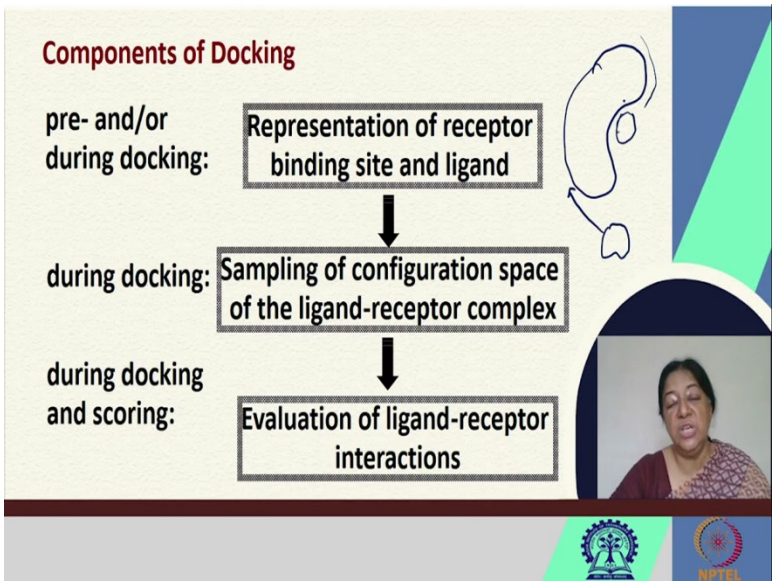
to bind a molecule, to bind a ligand, hydrophobic properties, polar, nonpolar, the atom types. We want one molecule that is our receptor, our protein to bind a ligand in a favorable manner.

(Refer Slide Time: 10:00)



So, we have the database of the structures of proteins. We have a database of ligand structures. Our next attempt is to get the scoring function, look for high scoring molecules, which will lead us to new inhibitor design or what we know is called a lead molecule.

(Refer Slide Time: 10:21)

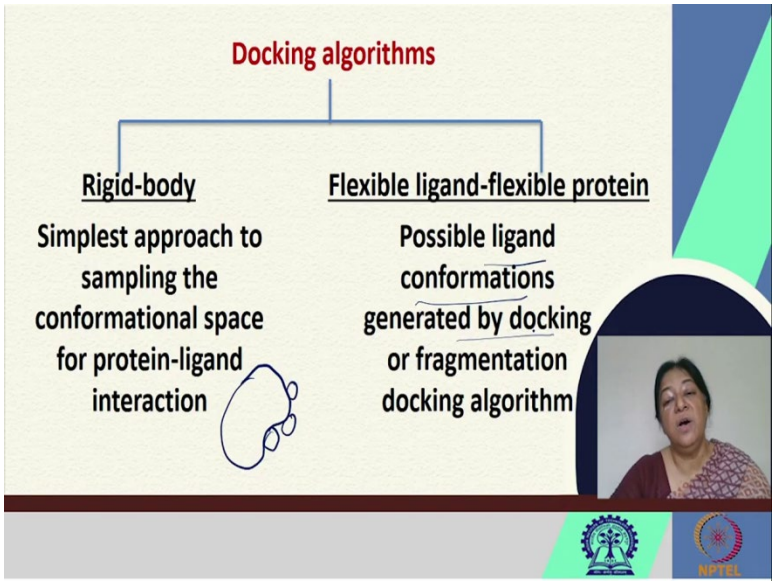


If we look at the components of docking now, we have pre and some components during docking. Here we are looking at a representation of the receptor binding site and the ligand. Then during the docking what we look at is, we look at a specific area of the configurational space that

is going to be sampled in the sense, that if we have a protein molecule and we would see whether our particular ligand would be so, we look at all the space available to it and then get proper scoring functions that are going to tell us about the ligand receptor interactions.

From that we evaluate those interactions to see, which is favorable in the terms of binding and affinity.

(Refer Slide Time: 11:15)

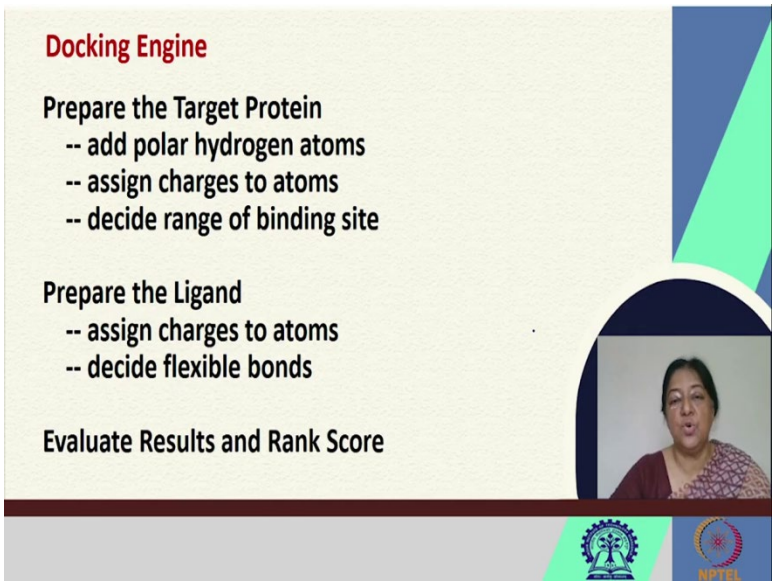


The docking algorithms that are available, are rigid body and flexible ligand or flexible protein. In the rigid body case, this is the simplest approach to sampling the conformational space for the protein and ligand interaction. Meaning that we have the protein that is a rigid receptor model, where we have the ligand that is sampled over three dimensional space.

In a flexible ligand or a flexible protein setup, we allow for ligand conformations generated by docking or also fragmentation. In a flexible ligand approach, we look at possibilities of movements of the ligand atoms or the ligand bonds, the rotation around the bonds that allows the molecule to adapt to the active site.

This is common in most docking algorithms. The flexible protein allows for torsion angles; we studied about torsion angles in our protein structure module, where we have the rotation about the backbone atoms that would allow a flexible protein receptor model.

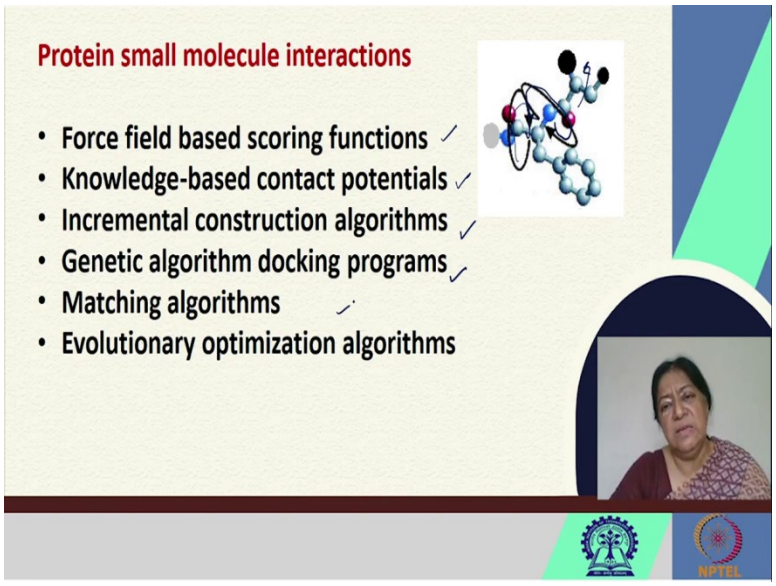
(Refer Slide Time: 12:32)



The docking engines therefore, prepare the target protein, they add the polar hydrogen atoms, they assign charges to the atoms and decide the range of binding site, if you do not know where your active site is. For example, if you are preparing an inhibitor for an enzyme, knowing that it is a substrate analog, it is likely that it is going to bind to the active site. So we can choose the region of the active site only, for the docking experiment or the docking studies.

We prepare the ligand molecule where we assign the charges to the atoms and we decide which bonds we will allow for rotation. And we evaluate the results and then rank the score accordingly.

(Refer Slide Time: 13:23)

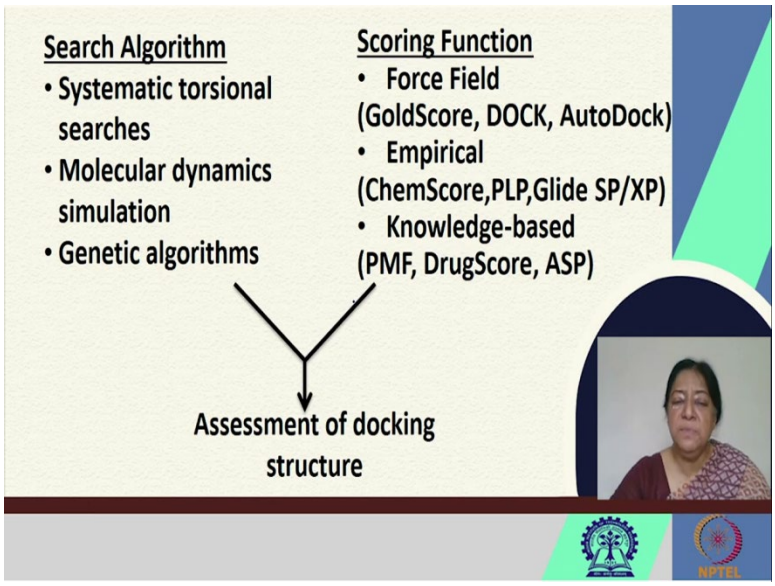


This would be where we would have the rotations possible for a ligand. So, we could have specific rotations that would allow different orientations, sampling of the different orientations,

to see which scoring function could give us a good value for a fit. We could have force field based storing functions, knowledge based contact potentials, incremental construction algorithms, general genetic algorithm docking programs, matching ones and evolutionary optimization algorithms.

These are all the possibilities that are there, there are endless possibilities to look at the ligand molecules and look at how they bind.

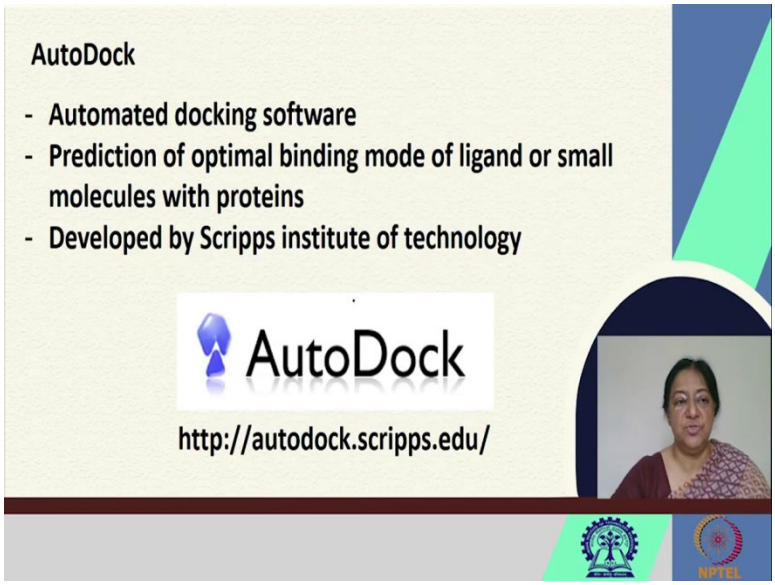
(Refer Slide Time: 14:10)



In the search algorithm therefore, we can look at systematic torsional searches or MD simulations that actually give us a much better idea of how the ligand is binding to the protein. But that is not always computationally feasible in the sense that we just need to understand whether our ligand is going to bind to the protein.

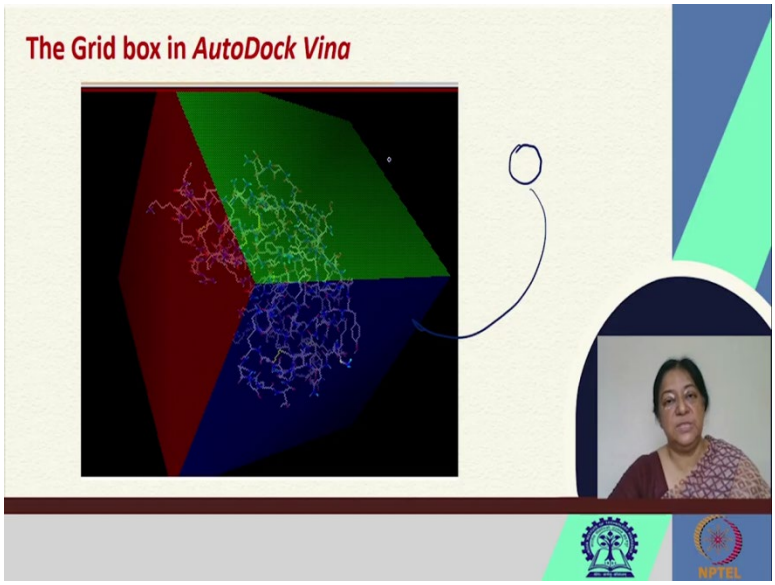
Instead of going for anything computationally intensive, we have a typical simple docking study, where we just look at the location of the molecule and then for further understanding we do MD simulation. In the scoring functions we these are the different software, docking software that are used force field type, empirical type, and knowledge based type. Then from this we go for an assessment of the docking structure.

(Refer Slide Time: 15:00)



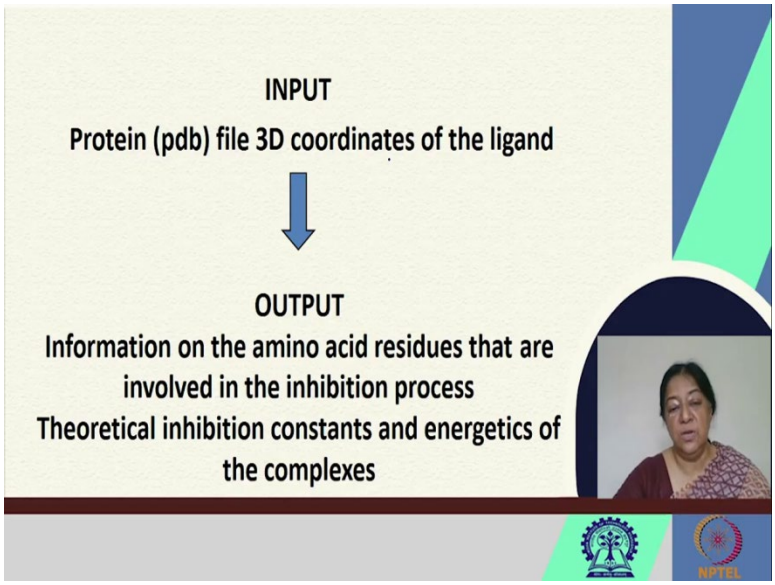
The AutoDock software, an automated docking software that can predict the optimal binding mode of a ligand or small molecules with proteins.

(Refer Slide Time: 15:11)



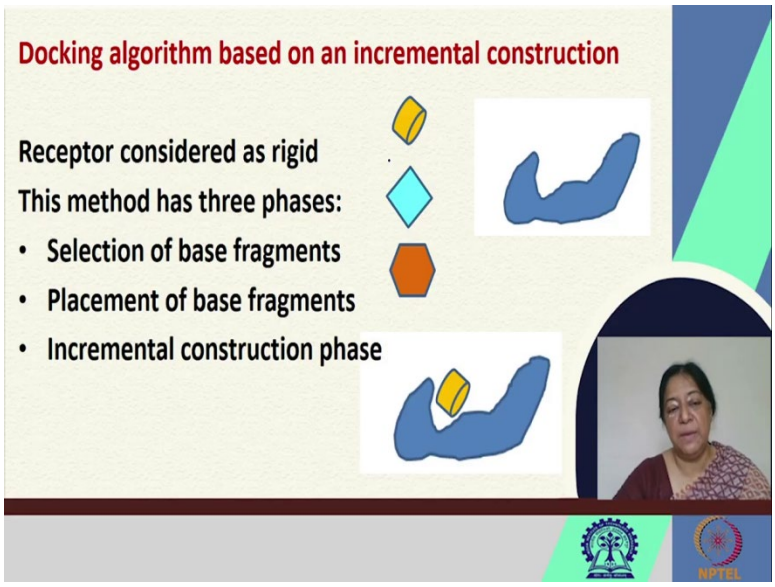
This [refer to slide] is a grid box where what you do is you place the whole receptor molecule, the protein molecule in this grid. And then the smaller ligand molecule that you have, allows for rotations about the bonds that we saw, sampling the whole space available to it, to look for clefts and nooks and crannies in the protein and bind to that specific site, give a specific score.

(Refer Slide Time: 15:39)



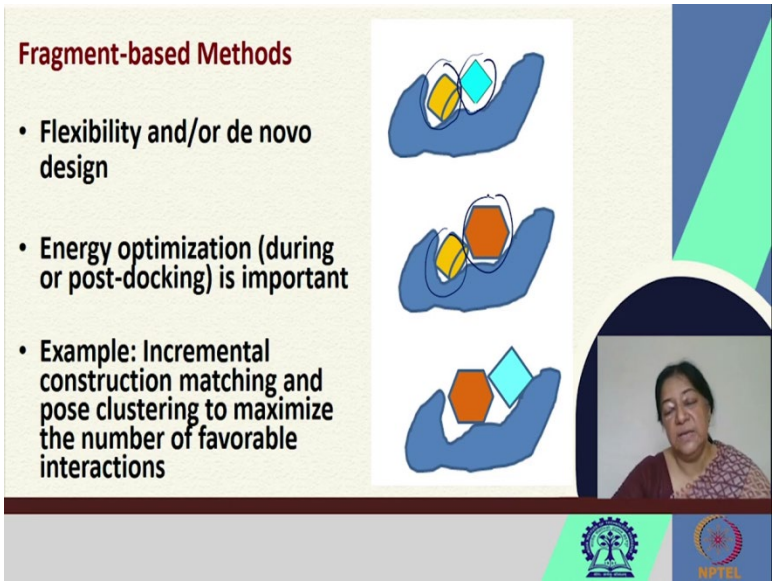
The input would be the protein three dimensional coordinates and the output would correspond to information about the amino acid residues that are involved in the ligand process or in which amino acids are actually bound to the ligand. In terms of enzyme inhibition, they would be involved in an inhibition process and from that we can get theoretical inhibition constants and the energetics of the complexes.

(Refer Slide Time: 16:05)



Let's take a look at the docking algorithm based on incremental construction.

(Refer Slide Time: 16:30)

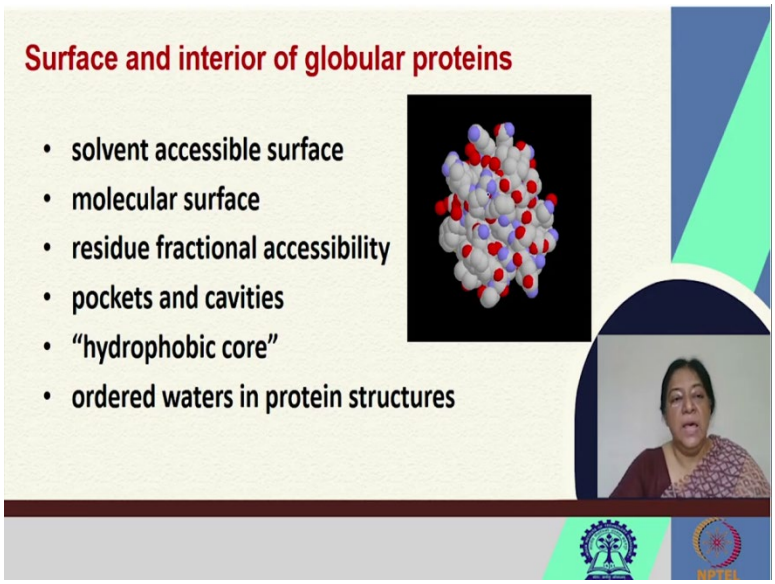


Here [refer to slide] is our receptor that we can consider to be rigid and this method has three phases. We have a selection of base fragments that are parts of the ligand molecule. We then place these fragments and look for an incremental construction.

In this case, when we are looking at fragment based methods, it has a flexibility, a de novo design, where we are looking at energy optimization, which is important and the incremental matching sometimes helps to give us a better match instead of looking at the ligand molecule together.

If these are our specific fragments, we can place them one by one and then see how they fit, depending upon the score that we get.

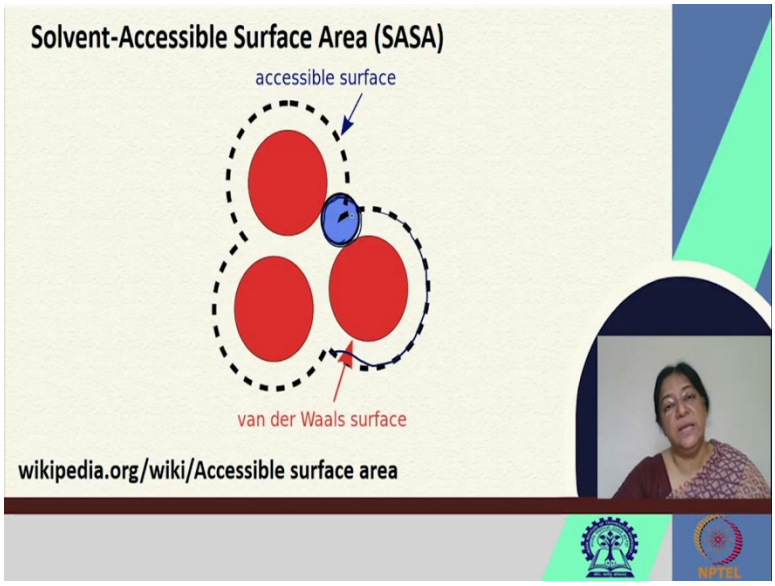
(Refer Slide Time: 17:04)



Now that we have our ligand bound to our receptor, our protein molecule, there is another aspect that we can look at, in looking at the solvent accessible surface. Now the reason why we want to look at the solvent accessible surface, is we want to get a knowledge of which residues are involved in our protein ligand interactions.

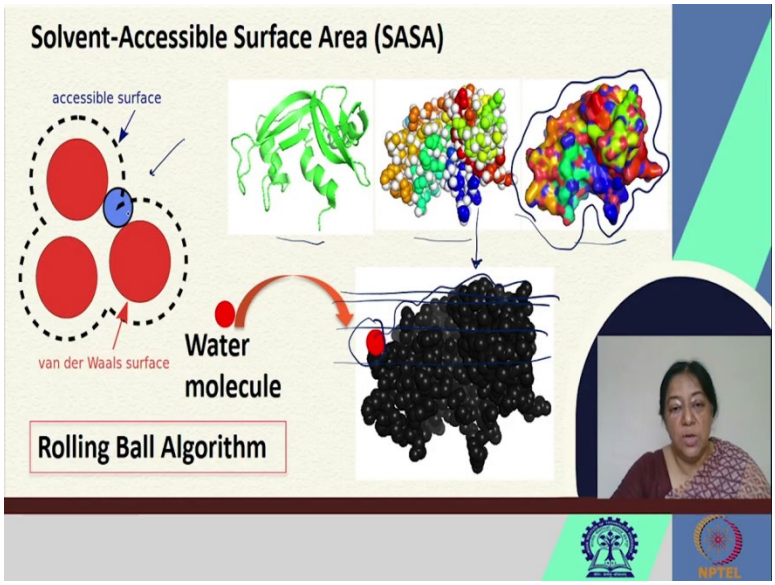
We would also like to know, how much of the surface of the protein has been lost in terms of the ligand binding. And this is more so for protein protein interactions, where we look at the loss of surface area on binding

(Refer Slide Time: 17:47)



Now in the calculation of the solvent accessible surface area, we have a specific radius probe that is usually the radius of a water molecule, that allows for a calculation of how much a ligand can penetrate a protein molecule. How much of the surface of the protein is available to the ligand molecule.

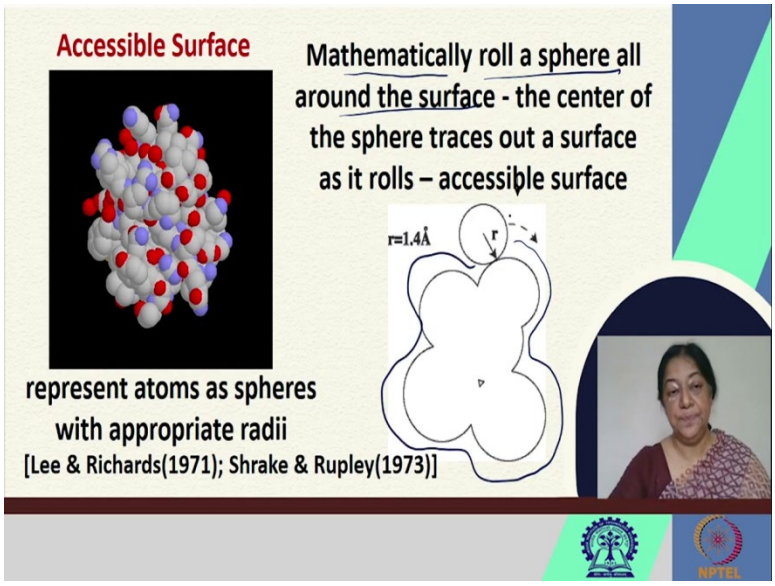
(Refer Slide Time: 18:10)



If we look at combinations like this, where we have our protein molecule and we are trying to sample the surface to see which residues are present on the surface. So, we have a water molecule and there is what is called a rolling ball algorithm, that takes slices of the protein.

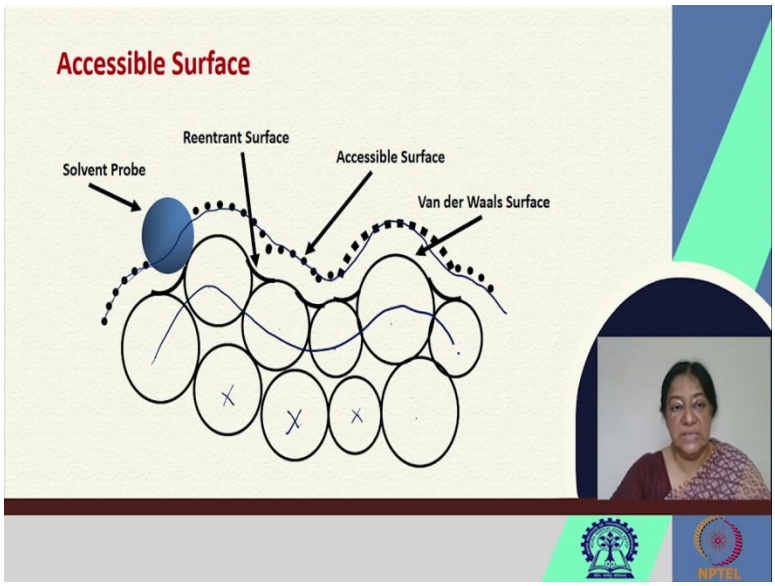
So if we look at specific slices, cross sectional areas and we are look at this [refer to slide] direction, we will see something like this on the left. And then we have the rolling ball that is going to look at the protein and see which residues it comes across. So, the water molecule then rolls around this protein.

(Refer Slide Time: 18:56)



The accessible surface is where we represent the atoms with spheres. We mathematically roll the sphere all around the surface and the center of the sphere from a cross sectional area, tells us which residues are present on the surface.

(Refer Slide Time: 19:18)

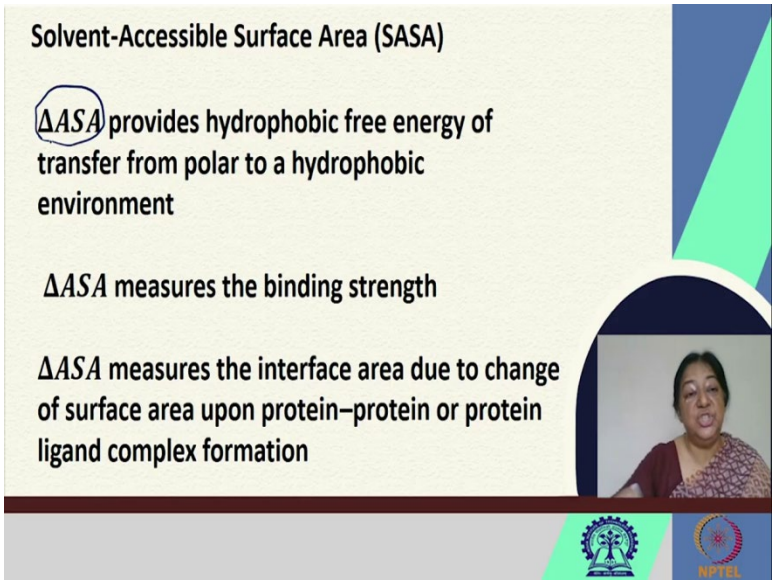


We have the solvent probe that we roll around the surface. And these [refer to slide] are the atoms of the protein that we have mentioned.

We realize that this atom or this atom or this atom or even depending upon whether what is placed here, these atoms are definitely at the core. So the solvent would not be able to access these, which would mean that they are buried in the protein, giving us a zero accessible surface.

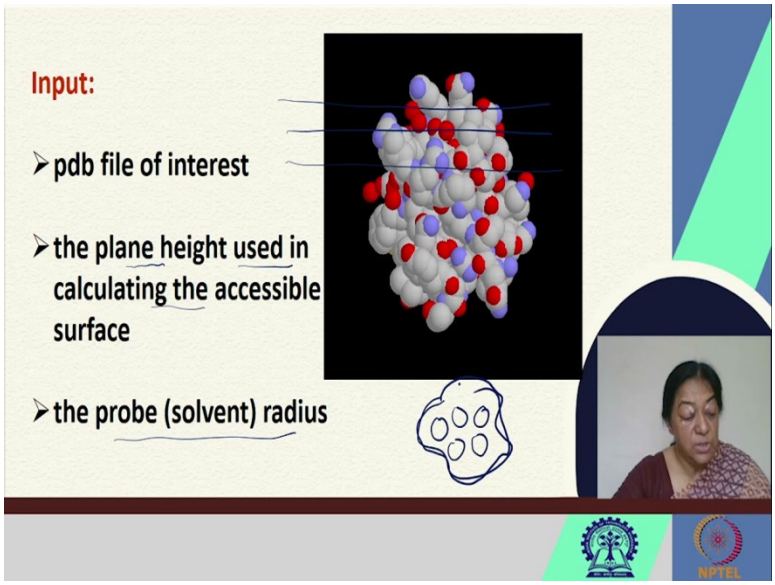
But those on the surface would have varied accessible surface and depending upon as we looked at the specific types of amino acids present on the surface, we would know where our ligand is binding and from the binding determine the loss in accessible surface.

(Refer Slide Time: 20:25)



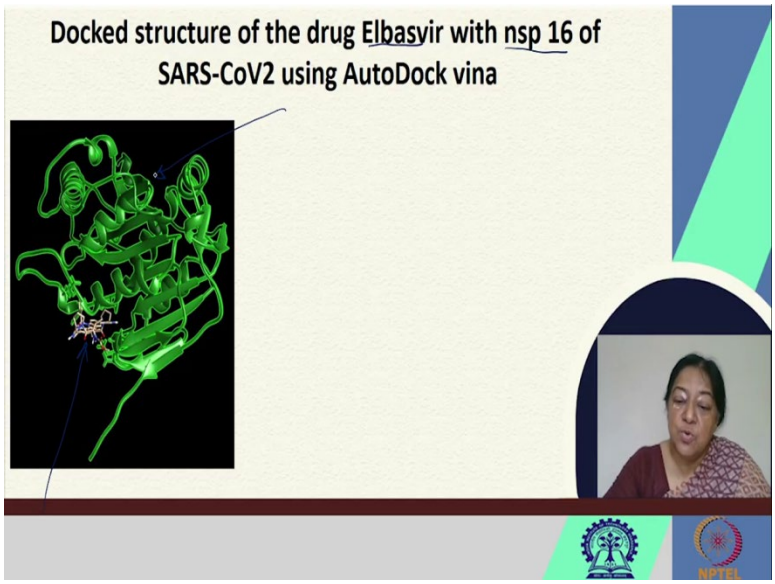
So what we have, is we have a measure of the delta ASA that is the change in the solvent accessible surface area. This provides us with some information about the binding strength, about the hydrophobic free energy and also measures the interface area, due to the change of the surface upon protein protein or protein ligand complex formation.

(Refer Slide Time: 20:53)



So, we know the coordinates of the atoms; we know the positions of the amino acid residues; we have the plane height that is used in calculating the accessible area, because we take slices of the protein and we roll our probe over each of these contours to see which amino acid residues are present on the surface of the protein.

(Refer Slide Time: 21:41)



This [refer to slide] is an example of a docked structure of the drug Elbasvir with nsp 16, which is a non structural protein 16 of the SARS- CoV2 using AutoDock vina. Now, as you can see from the diagram here, the protein structure is shown in green and we have the ligand shown in its elemental part here.

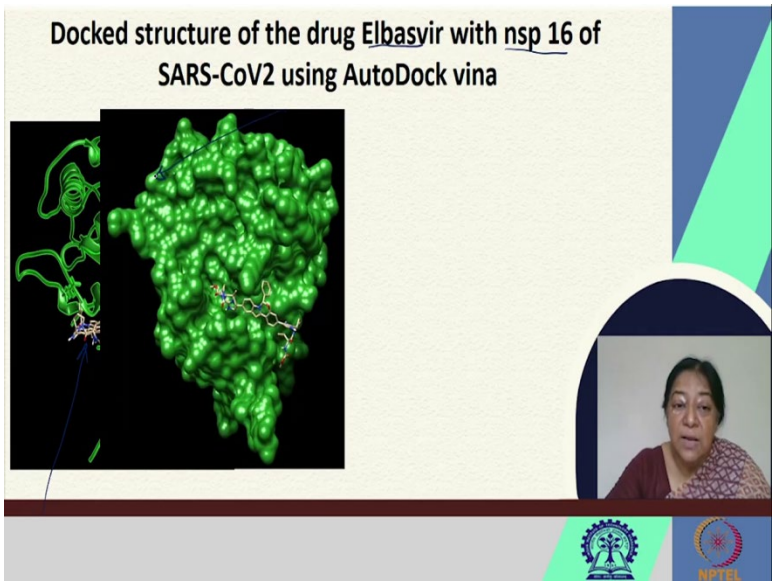
If we want to look at the surface of the protein to see which molecules or which amino acids are involved. Now that we have an idea of the structure of the protein, we can do a docking study. The docking study of this particular drug, with this protein, gives us this information.

If now we want to know which residues are involved, we can get that information from the dock structure. If we want to design a better molecule, we would look at the specific interactions involved, the specific interactions add up to give us this scoring function. So for the scoring function, there is a contribution from the electrostatic interactions, there is a contribution from the hydrophobic, the lipophilic type of interactions.

And all these put together give us our scoring function, which then in specific software are either the scoring function or given as a delta g, relative to the different types of options that we can see. We realize that we can have a rotation about the ligands, about the bonds in the ligand.

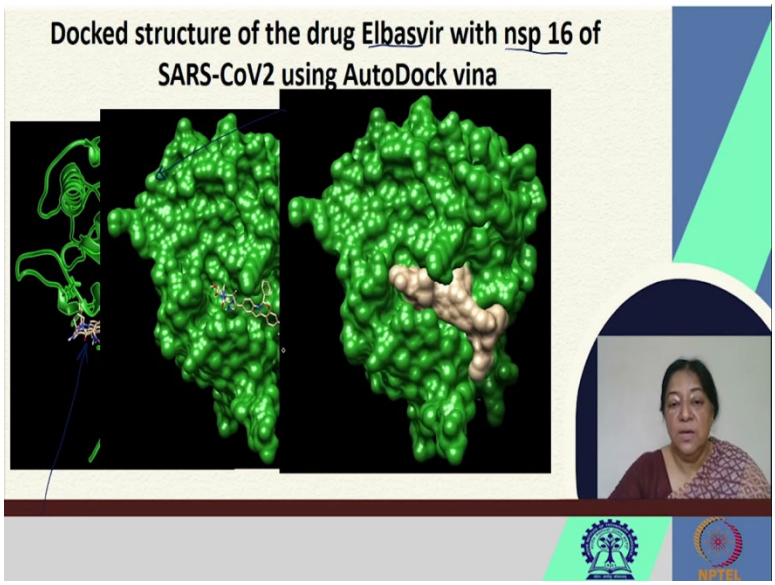
Each rotation would give us a new structure and a new scoring function or it may so happen that it could bind to a different region of the protein. We could have had the drug bound here, which would give us a different scoring function.

(Refer Slide Time: 23:53)



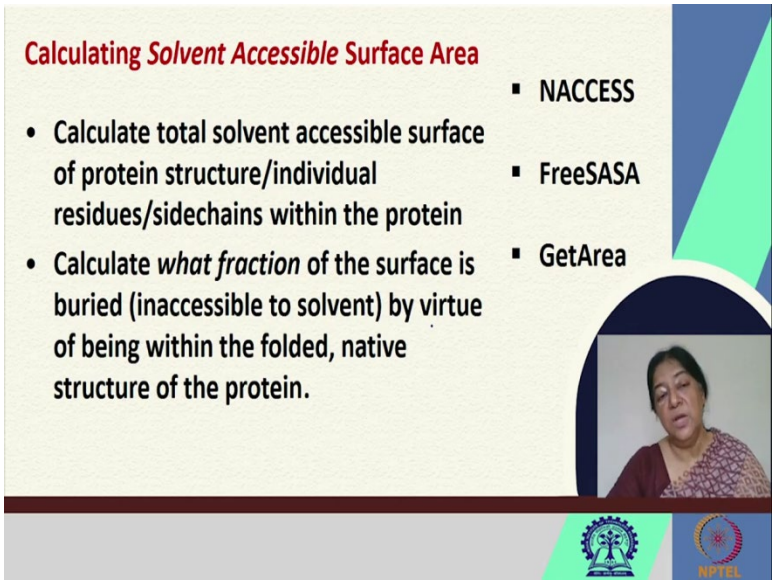
We look at the best scoring function, we then look at the cleft of the molecule. So our drug molecule has found a cleft for it to bind in.

(Refer Slide Time: 24:02)



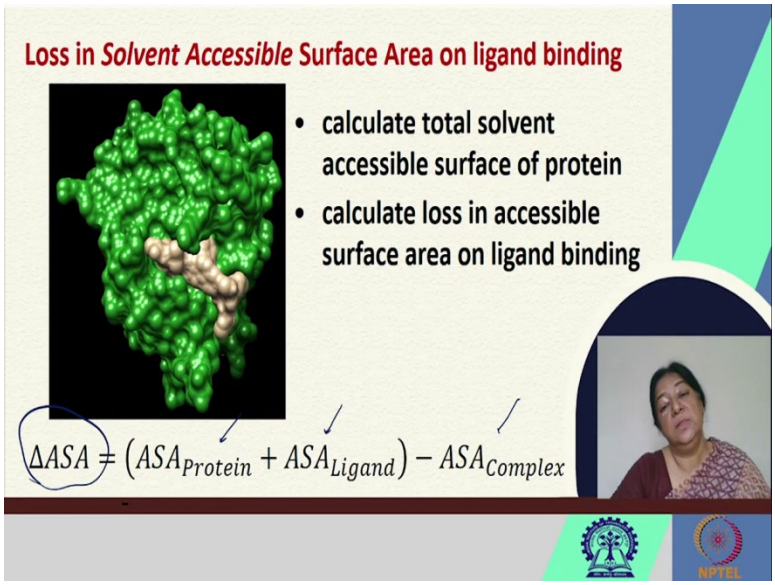
From that if we look at now the surface of our ligand molecule and the surface of the protein we see a geometric complementarity, a mirroring of the ligand surface with the receptor surface. From this protein ligand binding, from the docking algorithms, the docking structures, we get the information about which amino acids are present here.

(Refer Slide Time: 24:37)



The calculating the total solvent accessible area of the protein structure, the residues, the side chains, we can find out which particular residue is involved and we can calculate what fraction of the surface is buried. And here [refer to slide] there are specific software that are available to do these calculations.

(Refer Slide Time: 24:58)



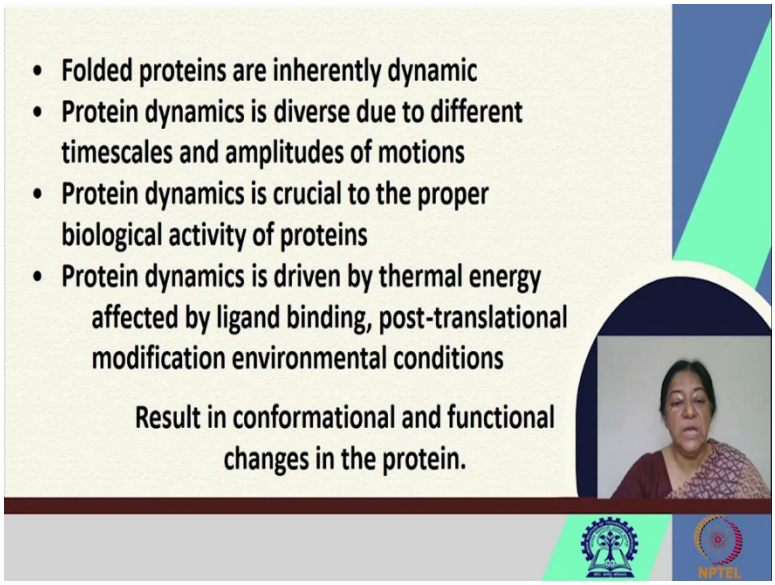
What we can then measure is a loss in solvent accessible surface area on ligand binding. What we have here is we calculate the total solvent accessible surface of the protein, without the ligand being present. Then, we calculate the loss in accessible surface area on ligand binding.

From that we can find out the change in the accessible surface area, by taking

$\Delta ASA = (ASA_{Protein} + ASA_{Ligand}) - ASA_{Complex}$

that is going to tell us the loss in accessible surface area on the ligand binding.

(Refer Slide Time: 25:38)

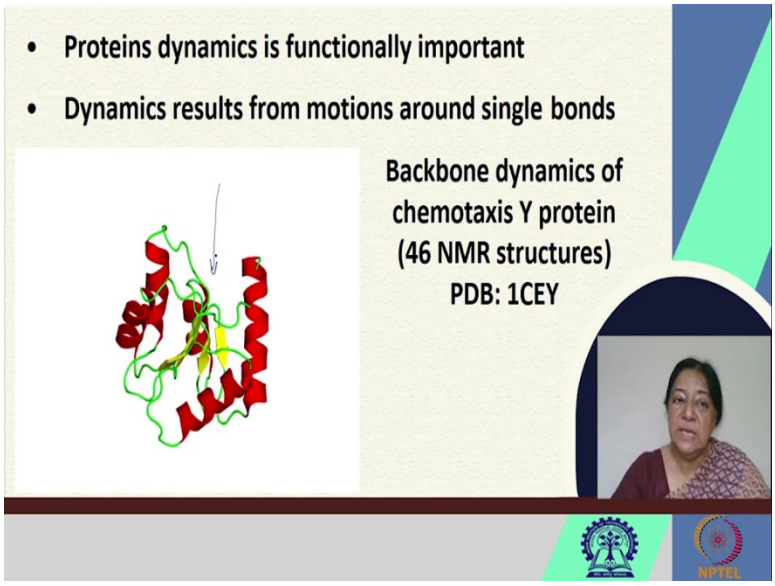


But folded proteins are inherently dynamic. And the protein dynamics is diverse due to the difference time scales and the amplitudes of motion. Protein dynamics is therefore very crucial

for a proper understanding of the biological activity or the function of the protein. The protein dynamics we know is driven by thermal energy that is going to affect the ligand binding, post-translational modifications, the environmental conditions.

This results in overall conformational functional changes in the protein when they are actually in solution.

(Refer Slide Time: 26:20)



Protein dynamics is important and from this dynamics we would like to see our ligand bound now, see the motion of the protein, how it binds the ligand and how the affinity changes on binding.

(Refer Slide Time: 26:41)

These [refer to slide] are the references.

Thank you.