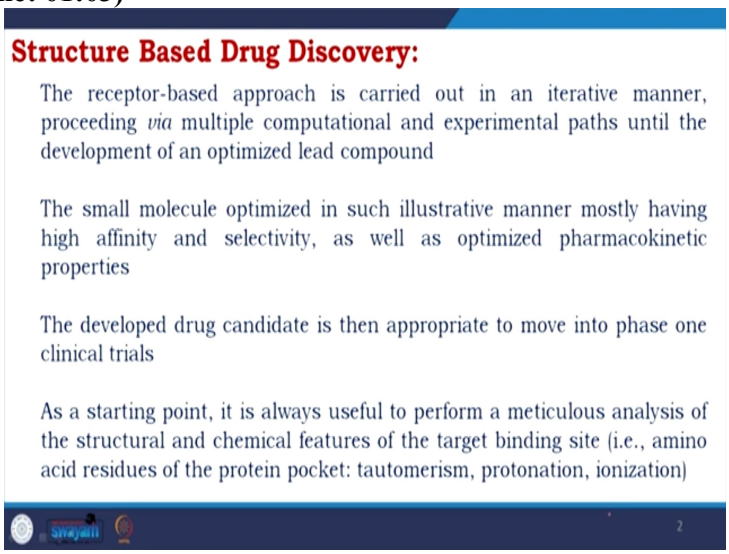
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology, Roorkee

Lecture - 58
Docking Based Virtual Screening: Progress, Challenges and Future perspective

Hi, everyone; welcome to the structural biology course. We have talked about traditional drug discovery history of drug discovery. Then we come into rational drug discovery where we are talking about pharmacophore. We also talked about de novo drug designing, where we design a drug from scratch.

(Refer Slide Time: 01:03)



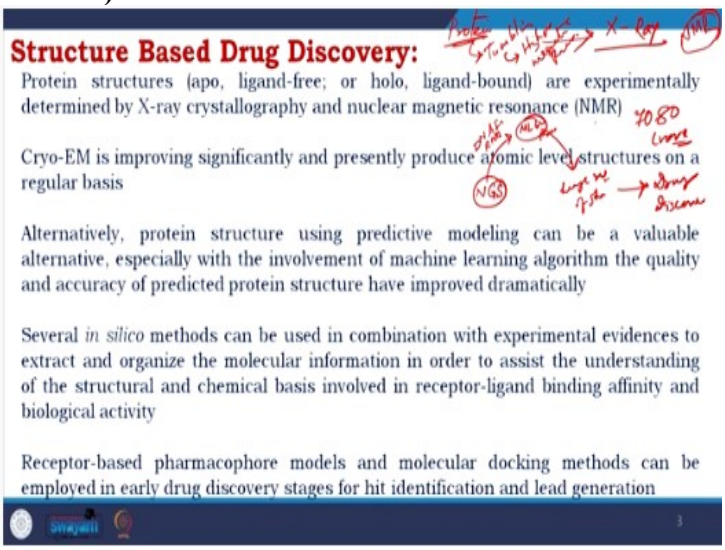
Today, we will emphasize structure-based drug discovery, continuing from the last class. So, the receptor-based approach is carried out iteratively, proceeding via multiple computational and experimental paths until an optimized lead compound is developed. So, in the receptor-based approach, the structure-based approach, we do an iteration by combining as I was talking about combining the structural information, high-resolution structural information, and the amazing advancement in computation.

So, you call it structure-based drug discovery, but the best way to call it is computer-assisted structure-based drug discovery. The small molecule is optimized in such an illustrative manner, mostly having high affinity and selectivity and optimized pharmacokinetic properties. Now, up to now, we are talking about starting from a small molecule that can be a drug to come to a position where it establishes itself as a drug.

But remember, I talked about that not all small molecules even could inhibit a critical enzyme. Let us say for a microorganism is not always a good drug. To be a good drug, there are certain criteria that we will discuss in the following lectures. The developed drug candidate is appropriate to move into phase one clinical trial. So, first, we do the screen, let us say, high throughput screening; maybe it is replaced now with virtual screening.

Then reduce the number, do the biochemical tests, the in vitro testing's, get interesting results, go to in vivo, get good results, then come to drug ability testing, the ADMETox. If you are successful, then you are going to clinical trials. As a starting point, it is always useful to perform a meticulous analysis of structural and chemical features of the target binding site that is the amino acid residue of the protein pocket which means the active site looking at the tautomerism, protonation, ionization, all sort of variation which could bring the possibility to modify or optimize.

(Refer Slide Time: 04:07)

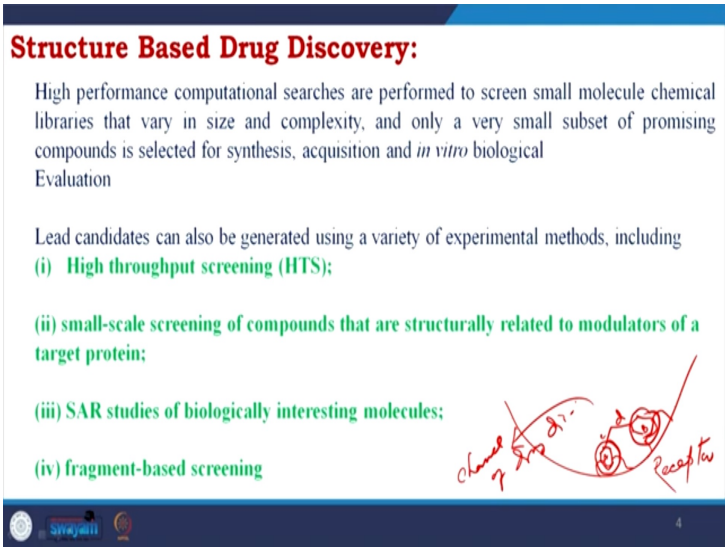


Protein structures, be it in that apostate that is ligand-free or in a holo state where the ligand is bound, are experimentally determined by X-ray crystallography and nuclear magnetic resonance. We used to say that, but now, Cryo-EM is improving significantly and presently producing atomic level structure regularly. So, yes, X-ray has been historically important for a long time. Even now, if you want a very high resolution, the atomic-level structure of the protein or the protein with the substrate inhibitor and all, your best bet is an X-ray.

So, now that we can get more and more predictive structure, it would help pull in developing model, be it a pharmacophore model, be it a huser model, be it what we are going to discuss today, docking based virtual screening. Receptor-based pharmacophore models and molecular docking methods can be employed in early drug discovery stages for heat

identification and lead generation. So, the major cut-off, the major selectivity, the major screening are required in the initial stage. Where you have millions of small molecules, and you have to cut a good number with that significant improvement in NGS, structural biology, and predictive techniques, it is possible now that these processes are coming up with a more improved model.

(Refer Slide Time: 10:40)

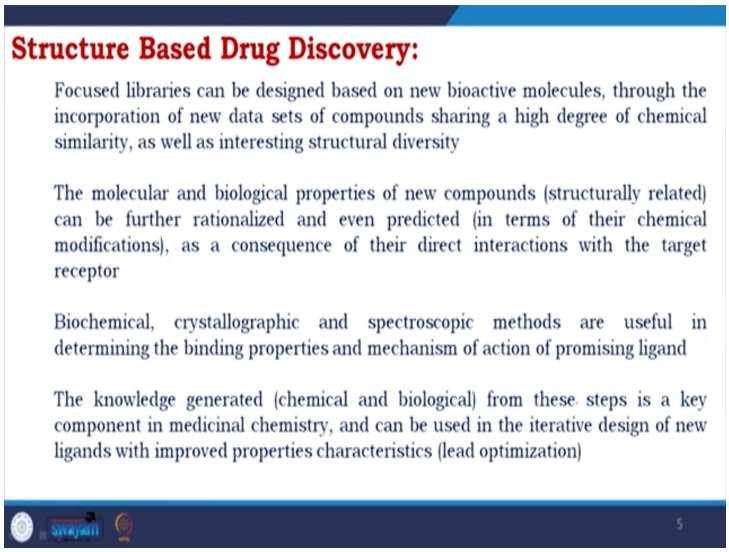


High-performance computational searches screen small-molecule chemical libraries that vary in size and complexity. Only a very small subset of promising compounds is selected for synthesis, acquisition, and in vitro biological evaluation. So, if you remember, when I started talking about rational drug designing, I was repeatedly talking that traditional drug designing was not a failure. Rather, we get a lot of drugs through rational drug design.

With improvement or replacing those high throughput experimental techniques more and more by using computation by combining experiments with computation. Less experiment you want to do, so that you could speed up the process and cut the cost, more computer involvement, and less cost.

So, lead candidates can also be generated using various experimental methods, including high throughput screening, small-scale screening of structurally related compounds to modulators of a target protein, structure-activity relationship studies of biologically interesting molecules, and fragment-based screening.

(Refer Slide Time: 14:08)

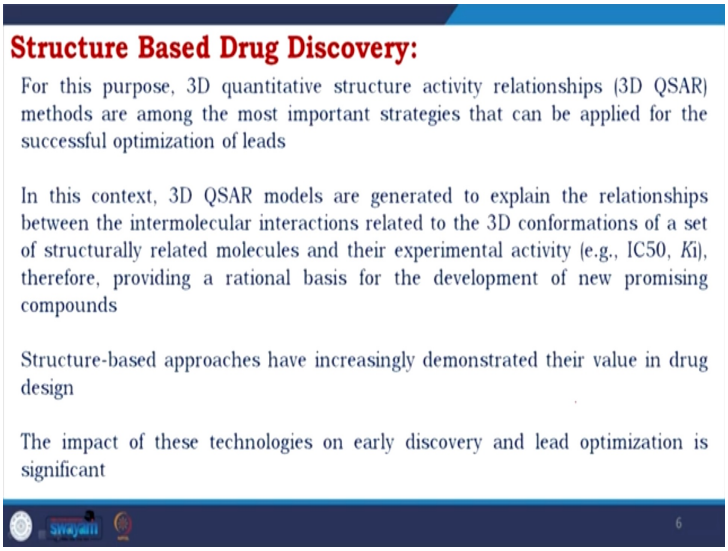


Focused libraries can be designed based on new bioactive molecules by incorporating new data sets of compounds sharing a high degree of chemical similarity and interesting structural diversity. The molecular and biological properties of new compounds related to the structure can be further rationalized and even predicted in terms of their chemical modification due to that direct interaction with the target receptors.

Biochemical, crystallographic, and spectroscopic methods are useful in determining the binding properties and mechanism of action of promising ligands.

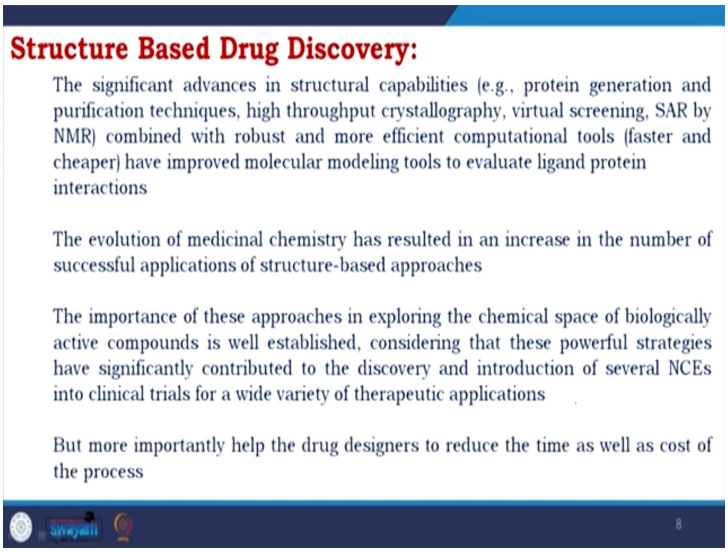
You could do a lot of screening and get many ideas about the receptor-ligand interaction. This step's knowledge generated from chemical-based knowledge and biology-based knowledge is a key component in medicinal chemistry. It can be used in the iterative design of new ligands with improved properties.

(Refer Slide Time: 16:45)



For this purpose, 3D quantitative structure-activity relationships 3D QSARs are among the most important strategies that can be applied to optimize leads. In this context, 3D QSAR models are generated to explain the relationship between the intermolecular interaction related to the 3D conformation of a set of structurally related molecules and their experimental activity, therefore, providing a rational basis for developing new promising compounds. Structure-based approaches have increasingly demonstrated their value in drug designing. The impact of these technologies on early discovery and lead optimization is very significant.

(Refer Slide Time: 17:43)



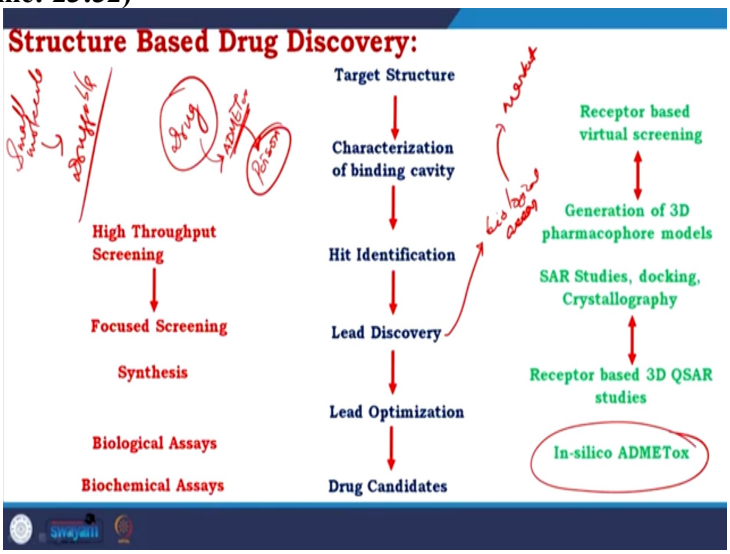
Although there is a diversity of different approaches being employed in the early stages of drug discovery, structure-based drug discovery or SBDD is one of the most powerful techniques and has been used quite frequently by scientists in the pharmaceutical industry and academic laboratories over the past 50 years. Structure-based drug discovery approaches continue to drive important advances in drug design, integrating traditional and modern technologies from the field of medicinal chemistry.

Computational chemistry, informatics, biology, biochemistry, structural biology, and structure-based methods bring the 3D structure of protein to light and greatly enable many drug discovery efforts to identify novel, small molecule drug candidates that selectively and potentially modulate the right biological target. Making knowledge-based decisions during the early phases of drug discovery is the key to decreasing hit to lead and lead optimization cycle times.

There are many different processes in traditional drug discovery, and in rational drug discovery, we say ligand-based drug discovery structure-based drug discovery. In the ligand-based company, there is pharmacophore modeling, huser. Similarly, now, for better drug designing, we are not restricted to a certain bit of process. We are merging all of them and using what is needed for that step. The evolution of medicinal chemistry has increased the successful application of structure-based approaches. The importance of these approaches in exploring the chemical space of biologically active compounds is well established, considering that these powerful strategies have significantly contributed to the discovery and introduction of several new chemical entities in NCEs or New Chemical Entities into clinical trials for a wide variety of therapeutic applications.

But more importantly, it helps drug designers reduce the time and cost of the process. So, we are talking about different things. Still, in one word, if we want to understand that, it is not very different, traditionally, we collect the Chinese traditional medicine, in Ayurveda, we look at animals, we look at birds because people who are culturing those used to be living in the forest. So, have close relation between like them with animals.

(Refer Slide Time: 23:32)



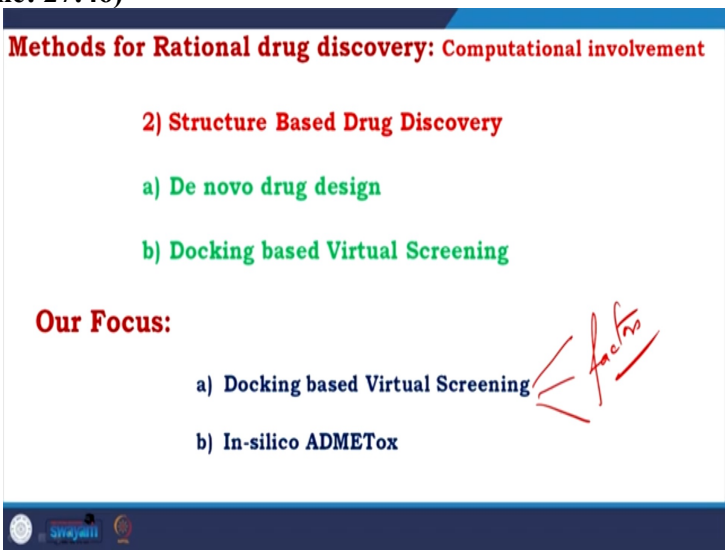
In structure-based drug discovery, you have a target structure. And when you get the target structure, at high resolution, you would be able to characterize the binding cavity, let us say, the active site of an enzyme that will help you to design interaction, go for hit identification, and ultimately towards lead discovery. Now, once you discover a lead, you are not limited there. You might go directly to biological assays, market but a much better way is to go from lead discovery to lead optimization and ultimately get the right candidates going for clinical trials. So, high throughput screening, as I told now converted to focus screening, which means we have to perform less synthesis and less biological assays that are costly and the biochemical assays. We do receptor-based virtual screening, which generates 3D pharmacophore models.

So, you see the virtual screening is connected to the pharmacophore model and then structure-activity studies, docking, high-resolution crystallography with receptor-based 3D QSAR studies.

The drug is something which you should consume, it should not have side effects, or at least it should not have a lot of side effects, a lot of drugs have some side effects, but it should not kill you. So that is a challenge, you do everything, and then you get a drug. You are happy because you could make money out of that, but then you do ADMETox. And it says no, it is a

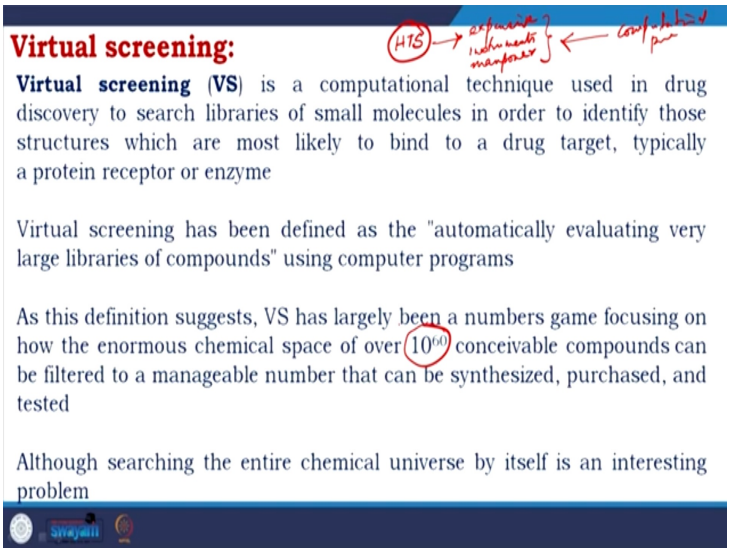
poison. And that is why people start thinking, is it possible that we could develop something the ADMETox what we are performing experimentally, can we do something so that we could do at least a set of the screen at the initial level. So that a small molecule, we could at least say that it is a druggable molecule, the druggability is answered by in silico ADMETox, saves a lot of money.

(Refer Slide Time: 27:46)



Methods for rational drug discovery: Structure-based, de novo designing, and docking-based virtual screening. Here our focus is docking-based virtual screening, all the factors influencing them, and then in silico ADMETox. So, today we will discuss docking-based virtual screening. Next class, we are going to discuss in silico ADMETox.

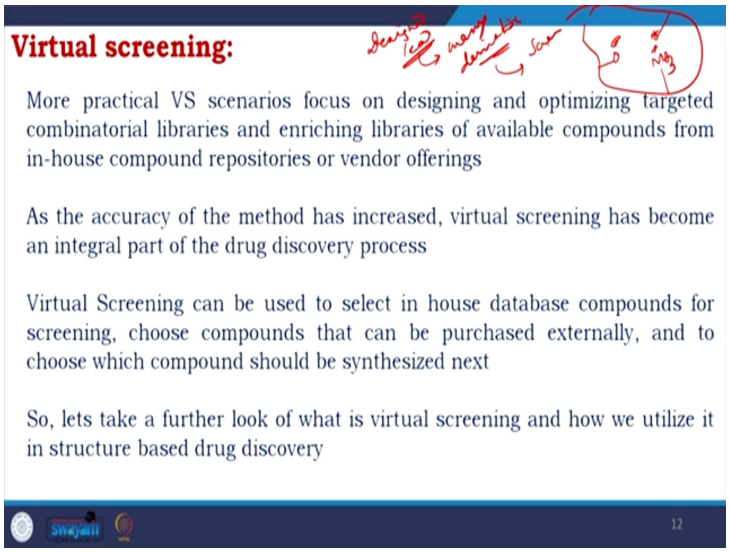
(Refer Slide Time: 28:17)



What is virtual screening? Virtual screening is a computational technique used in drug discovery to search libraries of a small molecule to identify structures that are most likely to bind to a drug target, typically a protein receptor or enzymes. In traditional drug discovery, there was something called high throughput screening. Now high throughput screening is expensive. It needs instruments, and it needs manpower. So, peoples start understanding the process of high throughput screening and try to replace it with a computational process that is called virtual. So, you do the screening in a virtual world. That is why it is called virtual screening. Virtual screening has been defined as automatically evaluating very large libraries of compounds using computer programs. As this definition suggests, factual screening has largely been a numbers game focusing on how the enormous chemical space of over 10^{60} conceivable compounds can be filtered to a manageable number that can be synthesized, purchased, and tested. So, let us say you want to work in a drug designing project, if I give you, let us say, not even very large 1 lakh compounds and you have to synthesize all this 1 lakh compound, it is very difficult, it will take years for you, and that is where virtual screening is coming.

It helps you screen initially making a computational model. When the number cutoff to an achievable one, you synthesize and do further in vitro, ex vivo, in vivo testings.

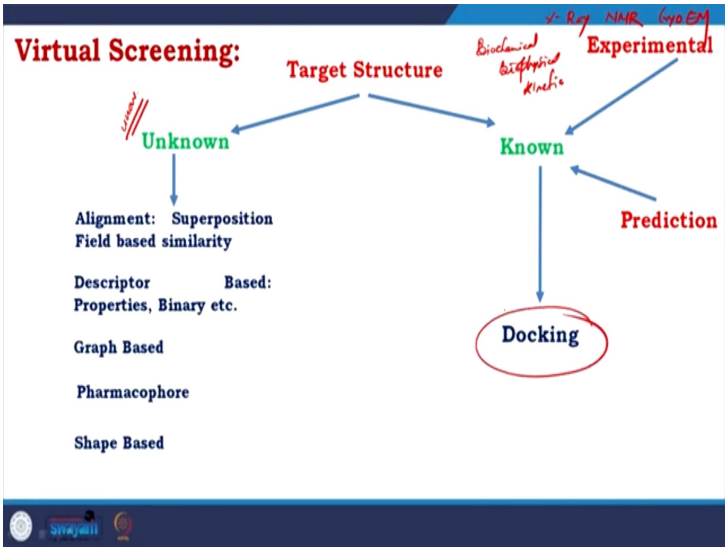
(Refer Slide Time: 31:20)



More practical virtual screening scenarios focus on designing and optimizing targeted combinatorial libraries and enriching libraries of available compounds from in-house compound repositories or vendor offerings. So, what does it talk about? So, you know its structure, you know the proteins, you know their charges, based on them you start to design, so, you get a designed lead. Then you make many derivatives, screen them, and then see that available libraries are there are not. If not, you synthesize them in-house or ask if some pure derivatives are available or not for testing. As the accuracy of the method has increased, the virtual screening with the, like, information coming with more and more compounds coming in the database, with more and more structure is solved, with more and more biochemical, biophysical in formations are there with more and more predictive modeling happened, the accuracy of virtual screening increased, it becomes an integral part of the drug discovery process because it is what we dreamed about.

Virtual screening can be used to select in-house database compounds for screening, choose compounds that can be purchased externally, and choose which compounds should be synthesized next.

(Refer Slide Time: 33:22)

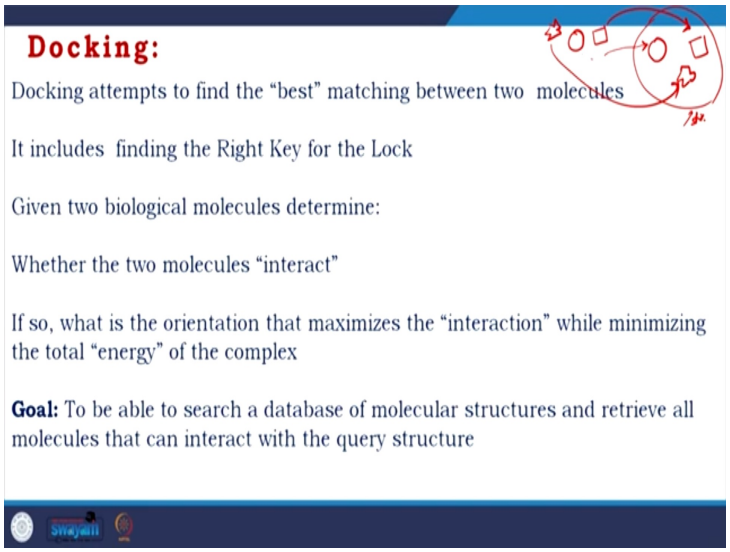


So, as I told you in structure-based drug discovery, there is a target structure. Still, as I told you, you will mix, so virtual screening is possible with the unavailability of structure where the structure is unavailable. Like alignment, which is superposition field-based similarity that could be a process, descriptor-based, we have already talked about descriptors while we are talking about QSAR, graph-based we talk about the graph theory, we have a binding site, we get the points, and then we connect them, we get the pattern, and that is where graph theory works.

Pharmacophore, we have already talked about a lot and shape-based. Whereas known structure, how could the structure be known? We all know now that there is X-ray, NMR, and Cryo-EM. But many biochemical, biophysical, kinetic methods include information about how the interaction between the receptor and the ligand happens.

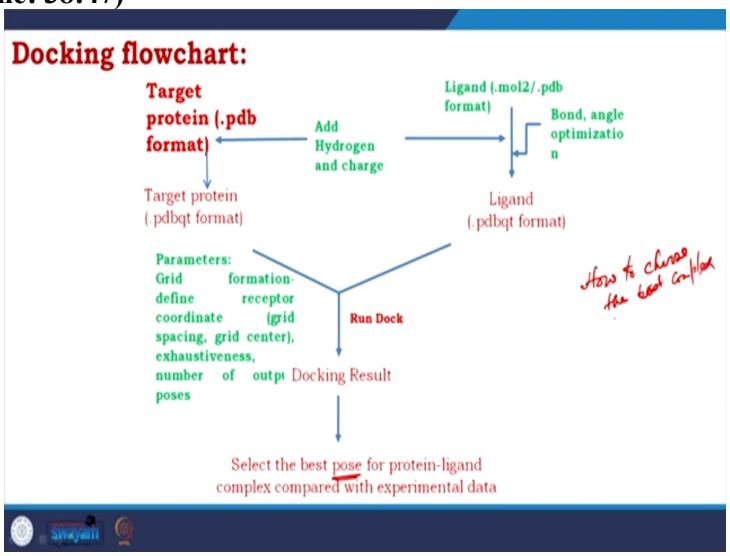
But now we know that the prediction of the model structure is also getting more and more accuracy contributing hugely to the known structure-based analysis. And here it is docking, docking-based virtual screening.

(Refer Slide Time: 36:41)



Docking attempts to find the best matching between 2 molecules. It includes finding the right key for the lock given to biological molecules to determine whether the two molecules interact. If they interact, what is the orientation that maximizes the interaction while minimizing the total energy of the complex? The goal of docking is to search a database of molecular structures and retrieve all molecules interacting with the query structure.

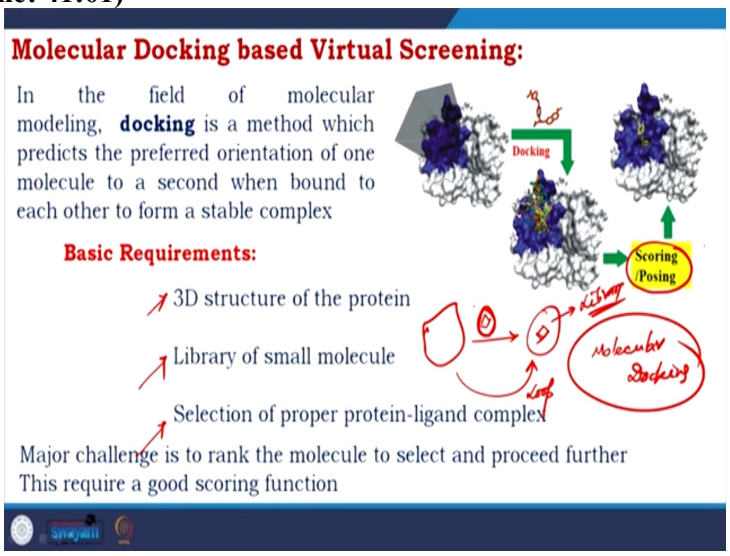
(Refer Slide Time: 38:47)



This is the flow chart to give you a simplistic idea of docking. You have the target protein present in .pdb format, and I am sure now you understand what .pdb? The ligand should be in .mol2 or .pdb format. First, you have to add hydrogen and charge. When you do that, when you add hydrogen and charge both of them, protein and ligand would come to .pdbqt.

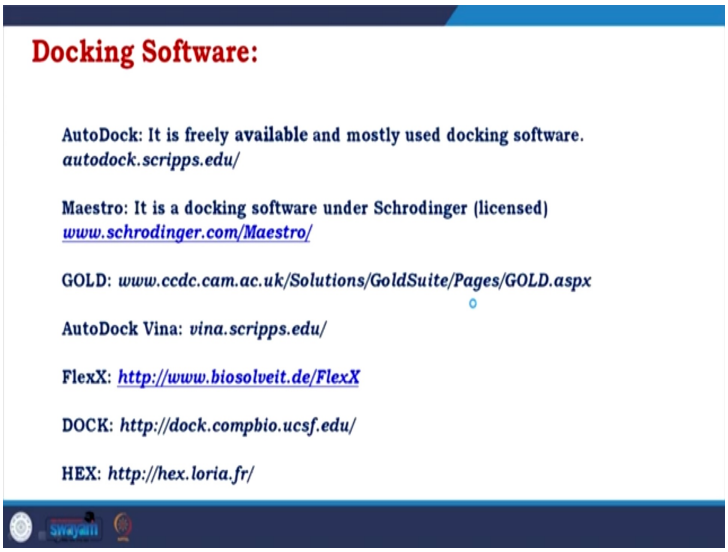
This is very important because docking works only in .pdbqt format. .pdbqt is docking specific format. Now, you could optimize bonds angles, especially if you remember; in the case of ligand, I talked about the rotatable bond. Because if you think a molecule of more rotatable bonds, it could change conformation, so, it comes flexibility issues and all, we will talk about them. You have them, and then you set the grid formation parameter, define the receptor coordinate, grid spacing, grid center, exhaustiveness is the redundancy and number of output poses. Then you run the docking, and you get the docking result. This is not the end. You have to select the best pose for protein-ligand complex compared with experimental data.

(Refer Slide Time: 41:01)



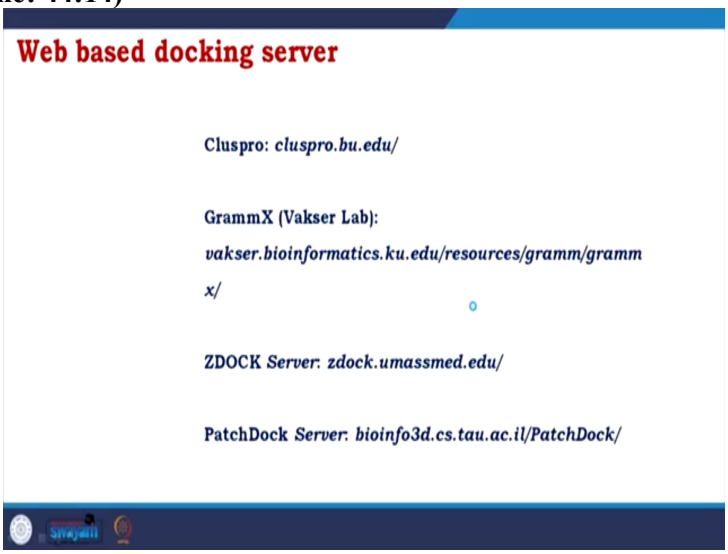
In molecular modeling, docking is a method that predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. So, when you want to do drug designing, you do docking. So, this is the process of docking, and then you have to choose by getting the pose and scoring it. So, the basic requirements in docking are the 3D structure of the protein and the library of a small molecule because you have to run them and select the proper protein-ligand complex.

(Refer Slide Time: 43:45)



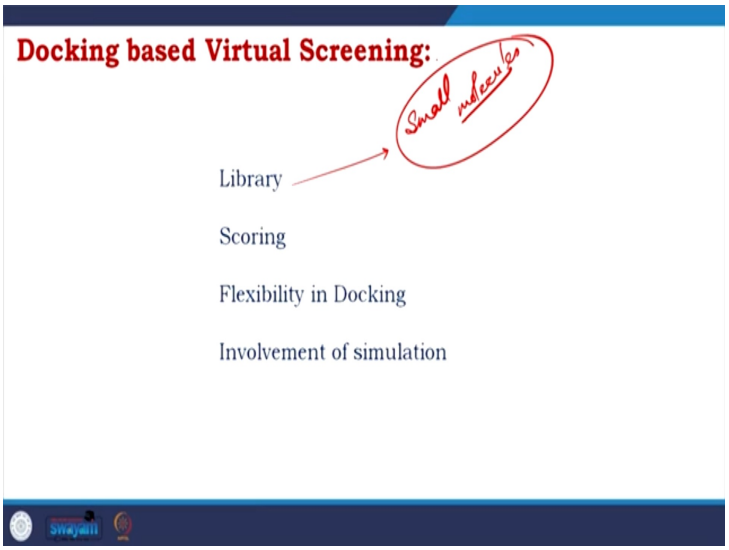
I have given a lot of docking software; AutoDock is the software most people use. You could use Maestro. Maestro is a commercial one, GOLD, AutoDock Vina again it is AutoDock, and AutoDock Vina is from Scripps, they are freely available, FlexX, DOCK, HEX.

(Refer Slide Time: 44:14)



Other ones are web-based docking servers. I have given the links Cluspro, GrammX, ZDOCK, PatchDock. You go there, provide your protein, and provide a small molecule, and they will automatically do the docking and providing.

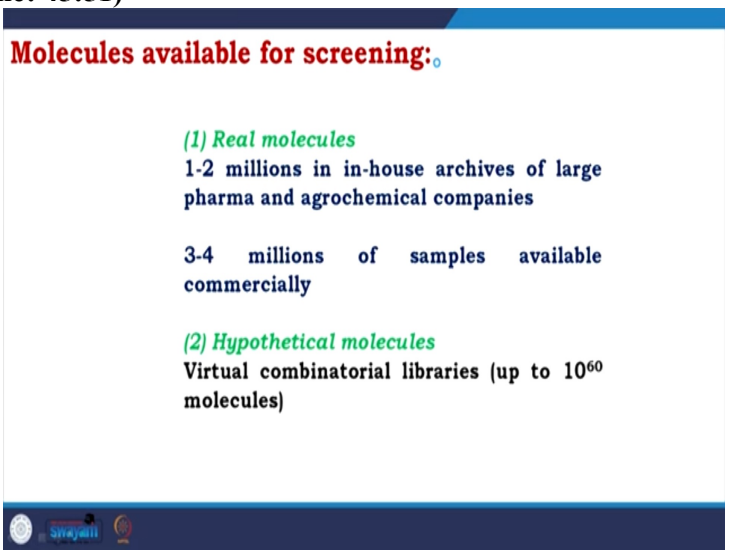
(Refer Slide Time: 44:31)



A few things are very important. One is the library, a library of small molecules. How should the small molecules be, from where you will get them, how we could separate them, what type of molecules we could pick up, how we could relate our lead with the library? All these you get an answer in getting a library. As I told posing and scoring is very important.

But then, most of our consideration of docking is based on rigid docking. With the advancement in science, with the advancement in computation, we want to explore the flexible area. So, flexibility in docking is a very interesting area. And when we talk about flexibility, we cannot talk without involving the molecular dynamic simulation.

(Refer Slide Time: 45:51)

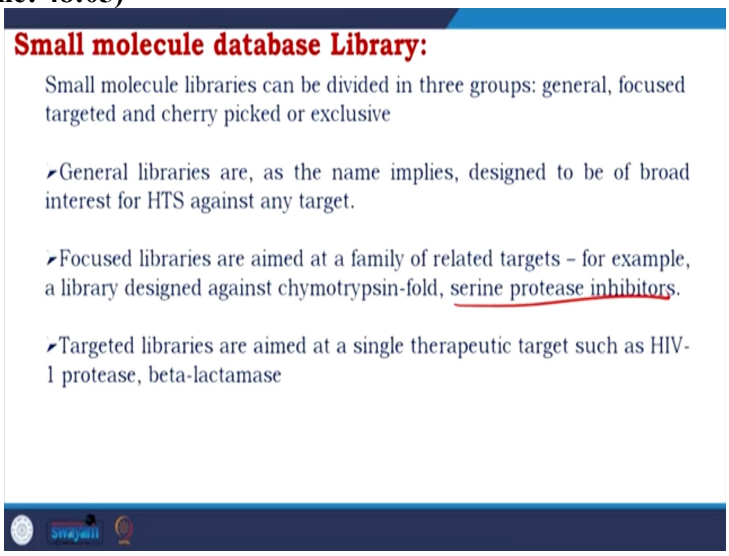


A lead compound is a small molecule that serves as the starting point for optimization, involving many small molecules closely related in structure to the lead compound. Many organizations maintain a database of chemical compounds.

Now, when you are thinking seriously about molecules, let me tell you that the whole world of molecules is divided into two major parts. One is called a real molecule. How does the real molecule come? Synthetic labs, companies, and pharmaceutical industries keep synthesizing compounds. Many synthetic labs optimize synthetic protocols; they work on the new synthetic process. So, whatever they synthesize and publish it and all these compounds actually exist in the world, they are called real molecules.

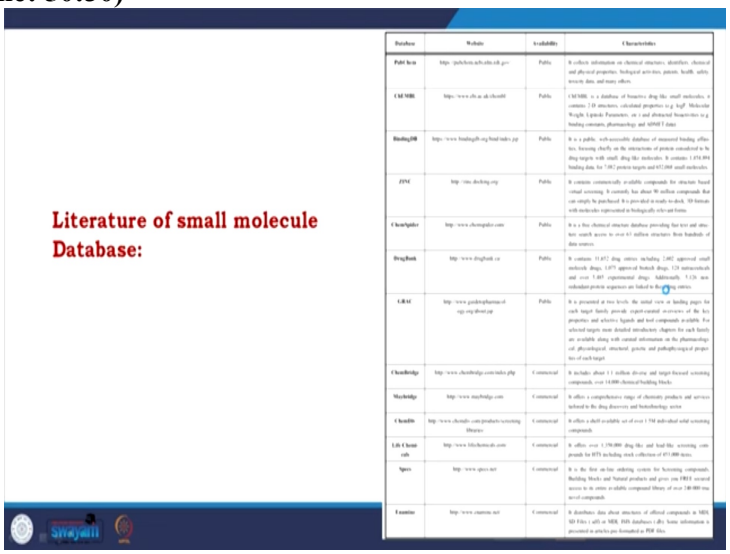
Hypothetical molecules are, let us say, you get a lead compound, then you do all types of modifications and design many molecules libraries. This library does not exist, but at least it has existence theoretically. They are called hypothetical molecules.

(Refer Slide Time: 48:05)



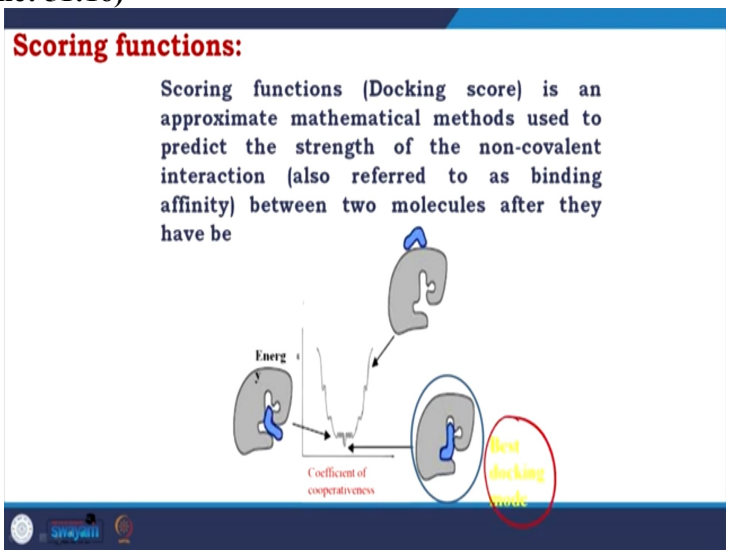
Small molecule libraries can be divided into three groups: general, focused, and targeted, but there is a special one, which is cherry-picked or exclusive library. As the name says, general libraries are generally designed to be of broad interest for high throughput screening against any target. Suppose a pharma industry has identified a lot of medicinally important plants, and they isolate the small molecules. So, they do not know what exactly they are destined for, but they think that it could be tested with many diseases. They are called the general library. The focused libraries are focused on a family of related targets. They are not specific for one protein, targeting a family. For example, a library designed against the chymotrypsin fold. You know that chymotrypsin is a protease. So, there is trypsin, and there is chymotrypsin. There are many such proteases, and because all those proteases work through serine as a catalytic residue, if you could inhibit them with a library of compounds, they are called serine protease inhibitors. Targeted libraries are aimed at a single therapeutic target such as HIV 1 protease. Cherry-picked library is the exclusive library of a set of compounds.

(Refer Slide Time: 50:30)



These are the database of different smaller molecular libraries, like PubChem, ChEMBL, BindingDB, ZINC, ChemSpider, DrugBank, G R A C GRAC, ChemBridge, Maybridge, ChemDiv, Life Chemicals, Specs, Enamine.

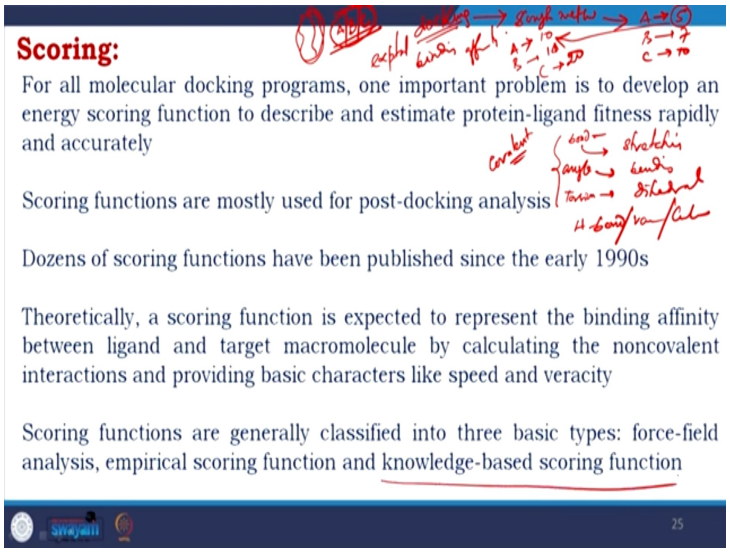
(Refer Slide Time: 51:16)



The scoring function is an approximate mathematical method used to predict the strength of the non-covalent interaction between two molecules after interacting. So, if you see it as an energy profile, you get, analyze those, and get the best docking mode.

To succeed in drug design, we have to optimize it, which needs comparison. The comparison needs scoring. That is why perfect scoring is a very critical and challenging thing.

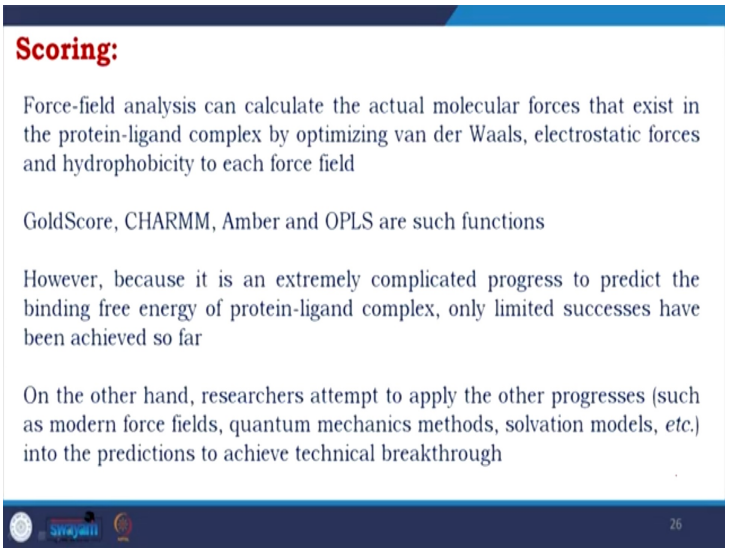
(Refer Slide Time: 53:25)



For all molecular docking programs, one important problem is developing an energy scoring function to rapidly and accurately describe and estimate protein-ligand fitness rapidly and accurately. The better the scoring function you have, the quicker the drug designing. The more accurate the scoring function you have, the better your success in drug designing.

Scoring functions are mostly used for post docking analysis. Dozens of scoring functions have been published since the early 1990s. Theoretically, a scoring function is expected to represent the binding affinity between ligand and target macromolecule by calculating the non-covalent interactions, the hydrogen bond, the salt bridge, the hydrophobic interaction, base stacking, metal binding, anything and providing basic characters like speed and veracity. Scoring functions are generally classified into three basic types: force field analysis, empirical scoring, and knowledge-based.

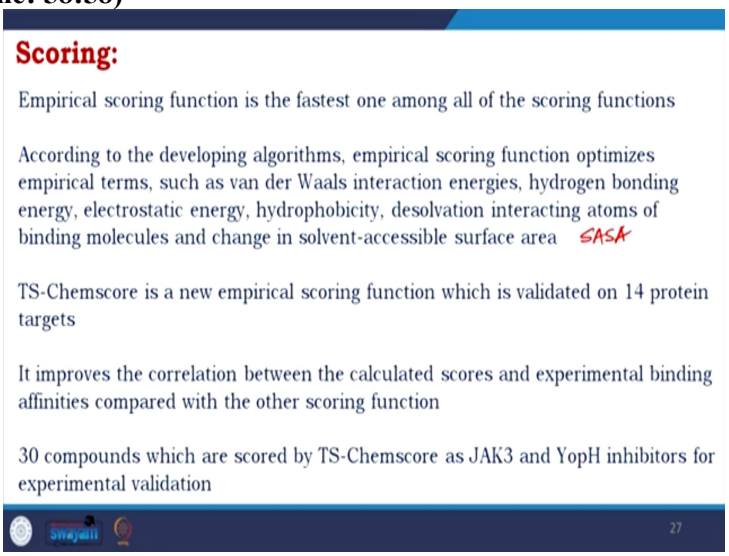
(Refer Slide Time: 57:53)



Force field analysis can calculate the actual molecular forces in the protein-ligand complex by optimizing, as I already talked about, Van der Waals electrostatic and hydrophobicity to each force field. GoldScore, CHARMM, Amber, OPLS are functions I have talked about. If you go back in the MD simulation, you will see those already. However, it is extremely complicated to predict the binding free energy of the protein-ligand complex.

If you remember, I talked about how complicated and computer-intensive they are. Only limited success has been achieved so far in this approach. On the other hand, researchers attempt to apply other progress such as modern force field, quantum mechanics methods, solvation models into the predictions to achieve technological breakthroughs.

(Refer Slide Time: 58:58)

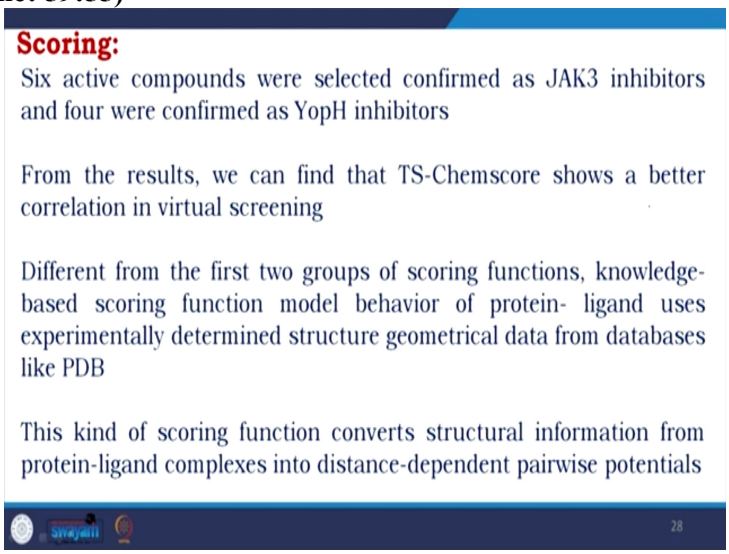


The empirical scoring function is the fastest among all scoring functions. According to the developing algorithms, the empirical scoring function optimizes empirical terms, such as Van

der Waals interaction energies, hydrogen bonding energy, electrostatic energy, hydrophobicity, desolvation interacting atoms of binding molecule, and change in solvent accessible surface area or SASA.

TS Chemscore is a new empirical scoring function validated on 14 protein targets. It improves the correlation between the calculated scores and experimental binding affinities compared with the other scoring function—thirty compounds scored by TS Chemscore as JAK3 and YopH inhibitors for experimental validation.

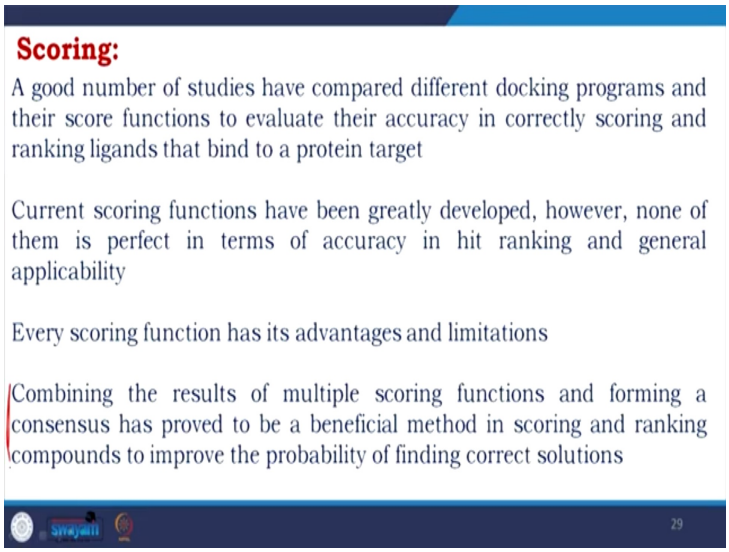
(Refer Slide Time: 59:55)



Six active compounds were selected, confirmed as JAK3 inhibitors, and four were confirmed as YopH inhibitors. So, empirical ones are rough and quick, but generally their values do not make much sense. Whereas, as you see in the Chemscore, the modern developed empirical scoring function is working, which is a good sign. From the results, we can find that TS Chemscore shows a better correlation in virtual screening than the existing scoring functions that we have worked on.

Unlike the first to a group of scoring functions, the knowledge-based scoring function model behavior of protein-ligand uses experimentally determined geometrical structure data from databases like PDB, as I have explained earlier. This scoring function converts structural information from protein-ligand complexes into distance-dependent pairwise potential.

(Refer Slide Time: 01:01:06)

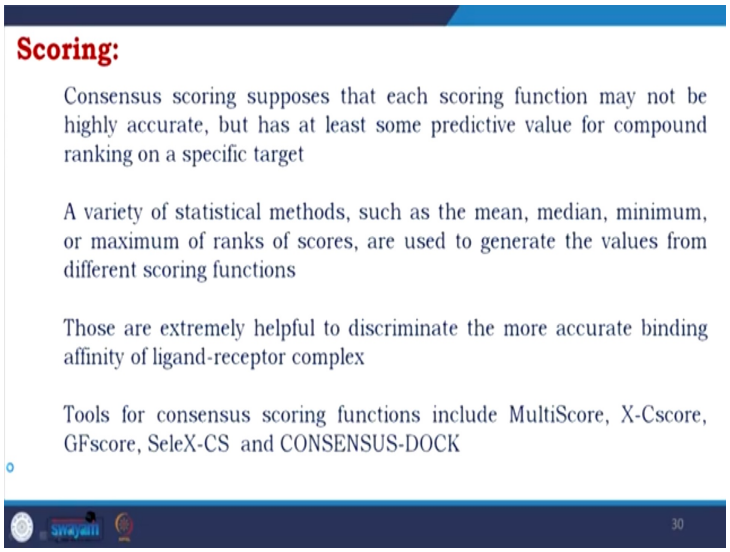


A good number of studies have compared different docking programs, and this score functions to evaluate their accuracy in correctly scoring and ranking ligands that bind to a protein target. Current scoring functions have been greatly developed. However, none of them is perfect in terms of accuracy in hit ranking and general applicability. This is a huge research font. If we could make it perfect, it would be possible to invent. It will be possible to discover a lot of drugs in a very short time.

Every scoring function has its advantages as well as limitations. Combining the results of multiple scoring functions and forming a consensus has proved to be beneficial in scoring and ranking compounds to improve the probability of finding correct solutions. This is a very important thing, you know, I always tell my students, telling you. Whenever you are doing computational in silico methods, never be restricted yourself to one procedure.

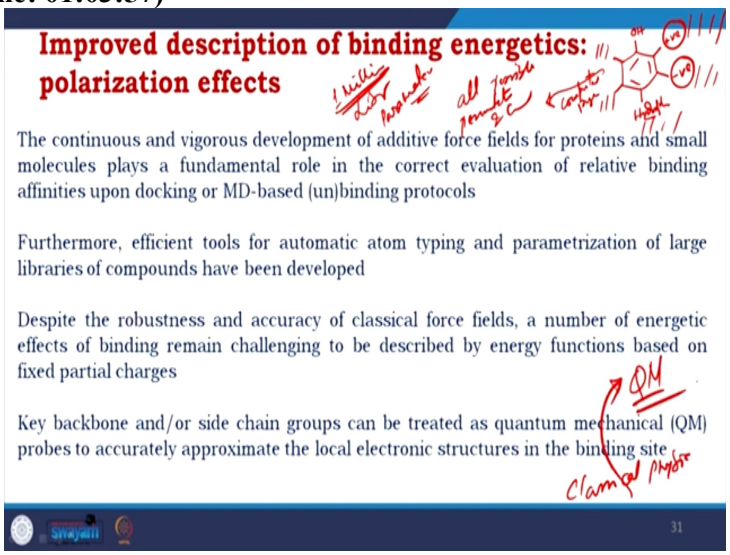
Because as I have told you from 1990, many pupils are working, a lot of pupils are coming up with different force fields, modified force fields, different processes, and all these things. And when you are going to be a user, in most cases, you have no idea what the physics is behind. So, what is my suggestion? My suggestion is if you are taking help of those software's for initial studies, take 4, 5 such functions, apply and see the correlations. If you see that there is a data consistency, then proceed. That would be extremely beneficial.

(Refer Slide Time: 01:03:15)



Consensus scoring supposes that each scoring function may not be highly accurate but has at least some predictive value for compound ranking on a specific target. Various statistical methods such as mean, median, minimum, or maximum of ranks of scores are used to generate the values from different scoring functions. Those are extremely helpful to discriminate the more accurate binding affinity of ligand-receptor complexes. Tools for consensus scoring functions include MultiScore, X-Cscore, GFscore, SeleX-CS, and CONSENSUS-DOCK.

(Refer Slide Time: 01:03:57)



So, going for an improved description of binding energetics is coming with the polarization effect. The continuous and vigorous development of additive force fields for protein and small molecules plays a fundamental role in correctly evaluating relative binding affinities

upon docking or MD-based binding or unbinding protocols. So, as you now understand, scoring is the most critical factor to get success in the field of drug discovery.

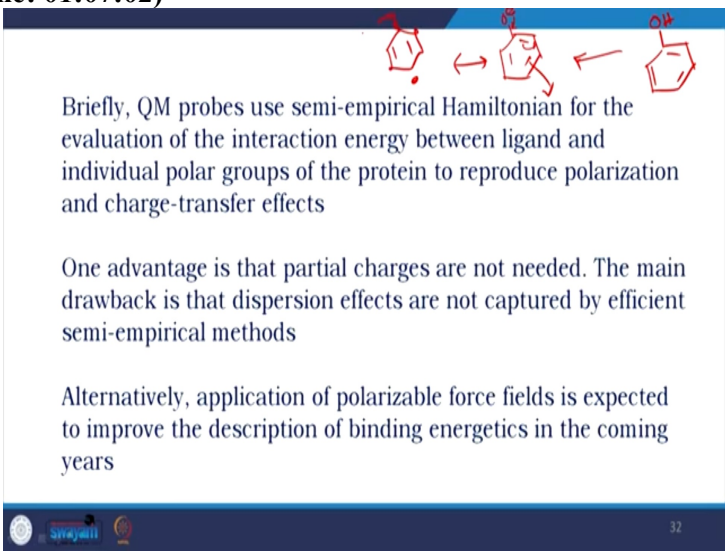
So, a lot of research is going on here. Already there are developments; furthermore, efficient tools for automatic atom typing and parameterization of large libraries of compounds have been developed. So, this is very interesting. Suppose this is your lead compound. Your lead compound is phenol. With the research, you know that a positive group is good here, a negative group is good here, and a hydrophobic one is good.

So, if you have a simple computer program that will allow you to introduce options here and then make all possible permutations and combinations from one small compound, you get around 1 million compounds containing the library where the lead compound is this. So, this is automatization, also with that the parameter of the molecule, if it could also be generated automatically they would be extremely beneficial.

Despite the robustness and accuracy of classical force fields, several energetic effects of binding remain challenging to be described by energy functions based on fixed partial charges. Key backbone and side-chain groups can be treated as quantum mechanical probes to approximate the local electronic structure in the binding site accurately. So, you all know when we are doing MD, we are doing classical physics.

But when we are considering the small molecules here, we do not need to go for classical physics, Newtonian physics. We could go to quantum mechanics, where the calculations would be pretty accurate; you do not need to do that experiment.

(Refer Slide Time: 01:07:02)

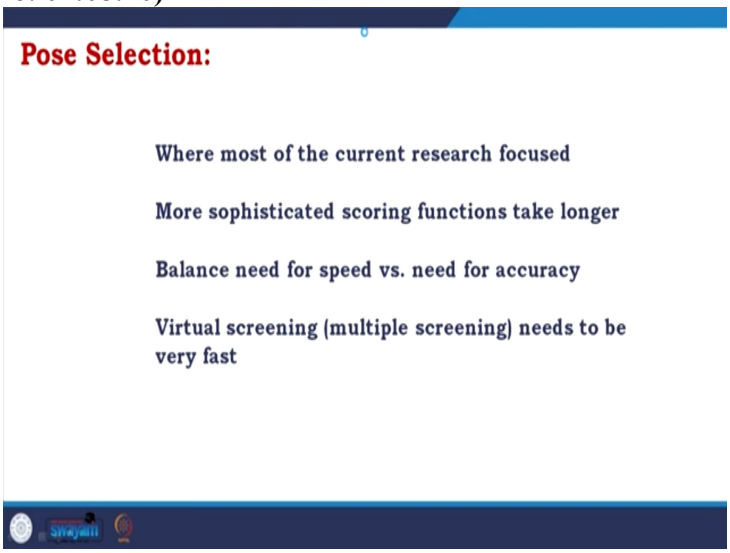


Briefly, quantum mechanics probes use a semi-empirical Hamiltonian to evaluate interaction energy between ligand and an individual polar group of the protein to reproduce polarization

and charge transfer effects. One advantage is that partial charges are not needed. The main drawback is that an efficient semi-empirical method does not capture dispersion effects. You have to go to a more classic quantum approach.

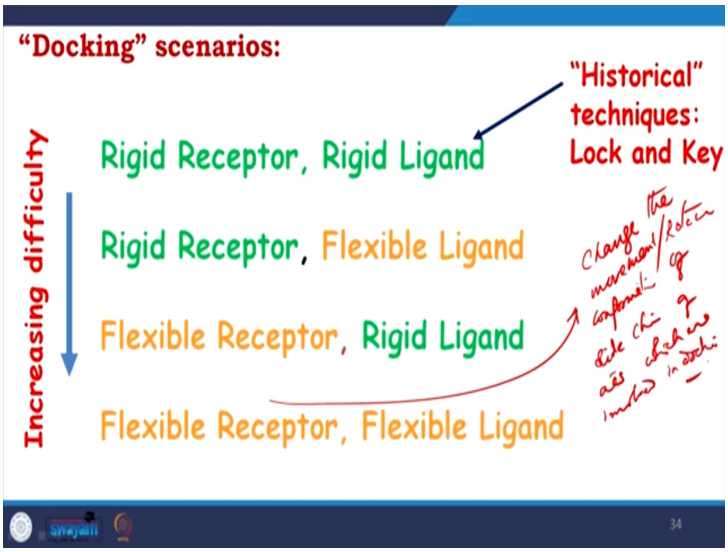
Alternatively, applying polarizable force fields is expected to improve the description of binding energy in the coming years. What is a polarization force field? So, once you are talking about. I am giving a very simple example, O H phenol. Now when it could be when it becomes O minus, it could resonate and develop a Quinone type structure. So, those change dynamicity is the polarization effect.

(Refer Slide Time: 01:08:20)



So, as I told already, post-selection, where most of the current research is focused, more sophisticated scoring functions take longer computation. Balance need for speed versus the need for accuracy. And virtual screening needs to be very fast because, with our ability to make it faster, more compounds would be included in the screen, and more numbers mean more chances to succeed.

(Refer Slide Time: 01:08:55)

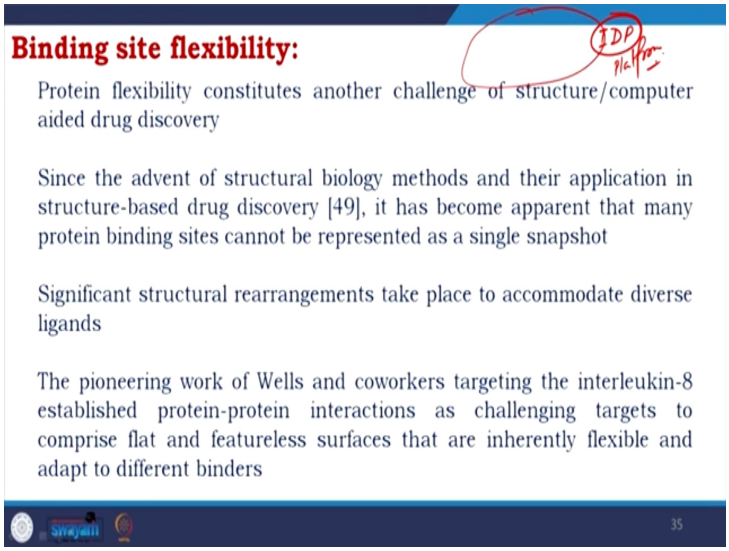


Now, we will move towards flexibility. If you look at the docking scenario, historically, it was a rigid receptor and rigid ligand. It shifts to the rigid receptor, and the protein is still rigid and flexible ligand. If you look at modern-day docking programs as I have listed, but if you have to ask me to name, if you go to AutoDock, you will see that you could fix or allow the rotatable bonds to rotate or not to rotate.

So, you could make the small molecule flexible, but you cannot make the protein flexible because it is big. Your calculation was way much more intense. Then you might have flexible receptors and rigid ligands. This is not a very common biological scenario. Still, if you consider chemical modifications in chemical engineering nanotechnology, you find this scenario where the receptor is flexible and the ligand is rigid.

And now, what is our dream? We want to go there flexible receptor and flexible ligands but, before ending this up, this rigid receptor, rigid ligand is historical where it started where we had the lock and key concept, but this is not far away. Now, if you look at AutoDock, you could change the movement or rotamer conformation of side chains of amino acids involved in docking, giving a certain amount. So, what I mean, you have the receptor. You could make these flexible.

(Refer Slide Time: 01:11:33)



So, protein flexibility constitutes another challenge of structure computer-aided drug discovery. Since the advent of structural biology methods and their application in structure-based drug discovery, it has become apparent that many protein binding sites cannot be represented as single snapshots. Once we started our journey as a drug designing historically, our base was the high-resolution X-ray crystallography structures because we started getting X-ray crystallography structures in the 1960s.

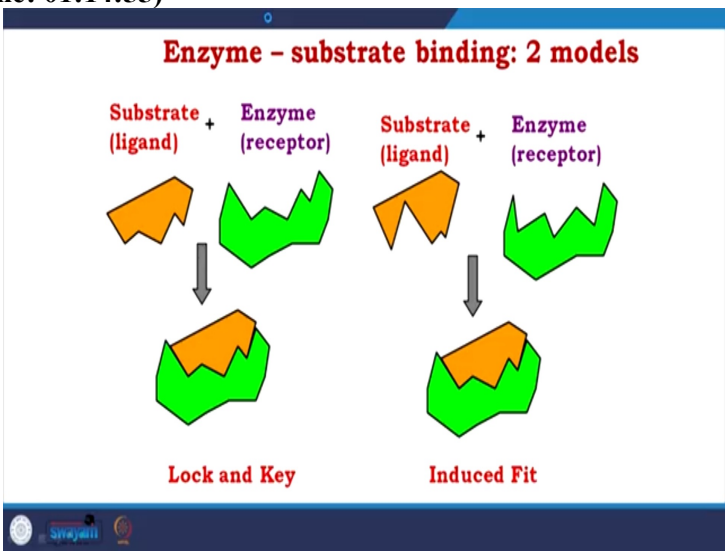
So, we thought that this was the representation of the protein. It is a static representation, but with more advancement, we understand that no protein moves, protein breaths, and those small changes in confirmations are extremely important in drug designing. Significant structural rearrangement takes place to accommodate diverse ligands. The pioneering work of Wells and coworkers targeting the interleukin 8 established Protein-Protein interactions as challenging targets.

To comprise flat and featureless surfaces that are inherently flexible and adapt to different binders. In these contexts, I would also introduce something called IDP. There are proteins, especially very important proteins, which use kind of as a platform. In a very critical biological event, let us say replication, transcription all in this type of process, we talk about 1 or 2 or 3 proteins.

But actually, hundreds of protein comes and takes part, and how do they switch? How do they specifically interact? It is because internally disordered proteins are like proteins, as I talked about in the entire course, are getting into the energetically stable state. So, generally, protein gets a stable fold, not with the internally disordered proteins. Internally disordered proteins take shape only when they interact with a specific partner.

In that way, they could interact with A in one way and interact with B in another way. So, as I told you, switching in and switching on continuously happens in a large biological macromolecular assembly. A set of proteins is coming. A set of proteins is leaving. All those are controlled by internally disordered proteins and are real challenges in drug design.

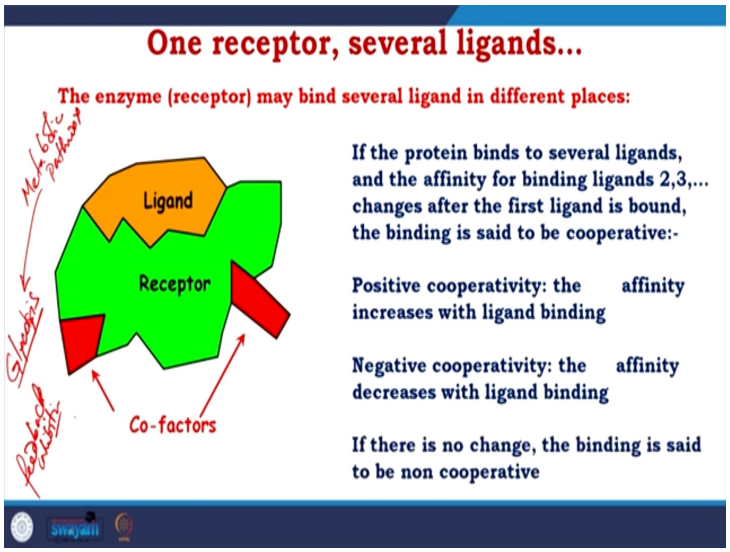
(Refer Slide Time: 01:14:55)



So, since the discovery, several other examples of protein flexibility upon ligand binding have been shown. Cryptic pockets could be discovered by accident or detailed analysis, often by more than one independent approach. As in the case of the polo box domain of polo-like kinase 1, a mediator of phosphorylation-dependent Protein-Protein interaction and cell cycle-dependent cancer target.

It has also been shown that structural adaptations are not limited to Protein-Protein interfaces. The enzymes are also more plastic than originally predicted. As I talked about that enzyme, we develop drugs based on their substrates and products. But then understand the transition state theory that an enzyme and enzyme are not destined to bind its substrate or product perfectly. Rather its intrinsic property is to stabilize the transition state.

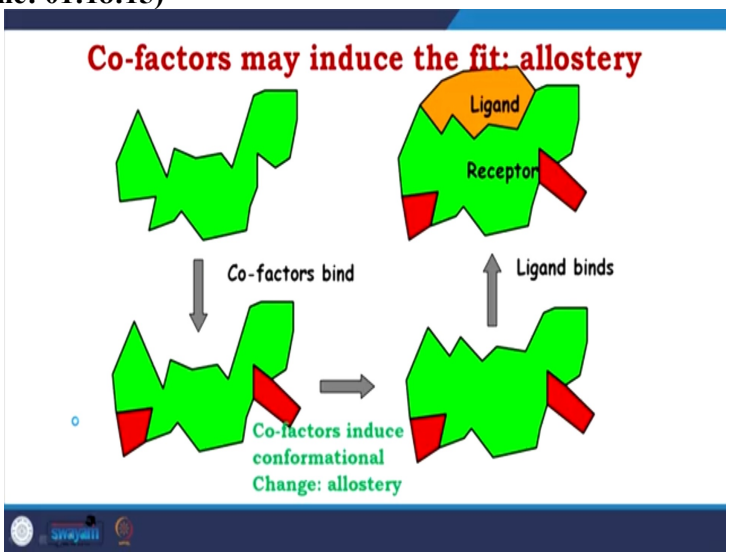
(Refer Slide Time: 01:16:43)



Also, there is the concept of cofactors 1, receptor 1, protein, and several ligands. The enzyme may bind several ligands in different places. If the protein binds to several ligands and the affinity for binding ligands 2, 3 changes, the binding is said to be cooperative after the first ligand is bound. I will give you the very known and common example of glycolysis if you look at the metabolic pathway.

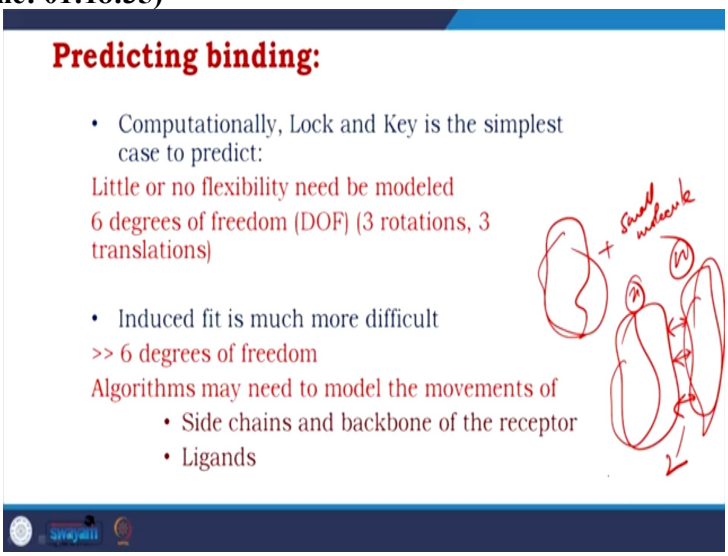
You see their feedback inhibition. So, the cooperative interaction might be good, like we see in the case of hemoglobin, as the monomer binds then comes the dimer, more oligomerization it is more speeding up. So, it is positive, but in feedback inhibition, it is negative. So, in positive cooperativity, the affinity increases with ligand binding. Negative cooperativity, the affinity decreases with ligand binding. And if there is no change, the binding is said to be non-cooperative.

(Refer Slide Time: 01:18:15)



So, as I told the cofactor may induce the fit, which is called allostery. As you can see that when the cofactor bind, the binding induces a conformational change, and then the ligand bind. This is called allostery.

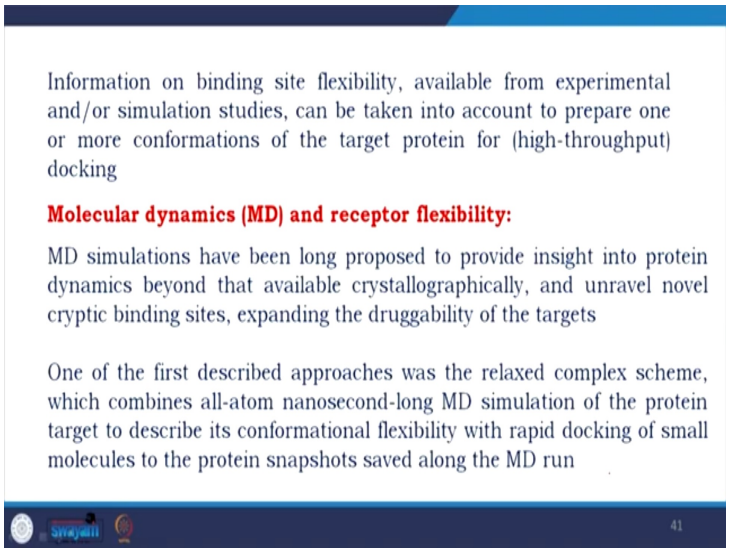
(Refer Slide Time: 01:18:35)



So, predicting binding computationally, as I told you, lock and key is the simplest case to predict. Little or no flexibility need to be modeled, 6 degrees of freedom, three rotation, three translations. Induced fit is much more difficult, much greater than 6 degrees of freedom. Algorithms may need to model the movements of the side chain and backbone of the receptor and ligands. Now, you see you are considering your protein, and then a small molecule is coming. So, you get the degrees of freedom.

Think about if both are macromolecules, so both have several possibilities. So, the possibility they could interact with each other is an absurd number. And that is why it is so difficult to study Protein-Protein theoretically.

(Refer Slide Time: 01:20:05)

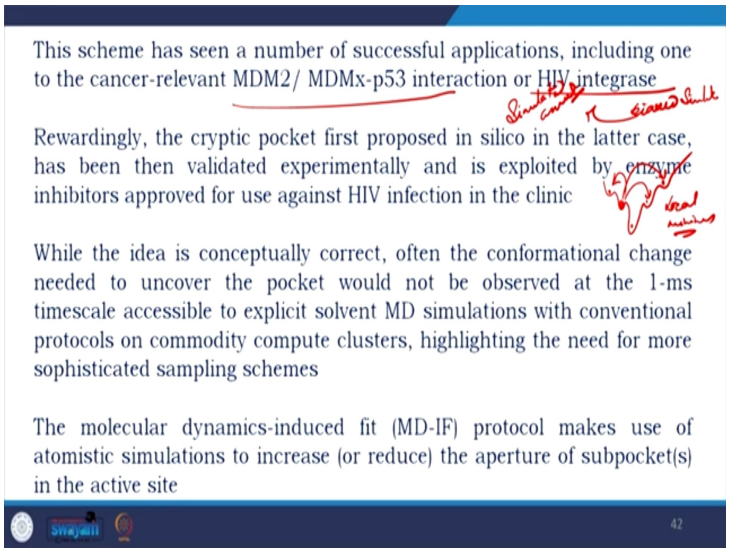


So, information on binding site flexibility, available from experimental or simulation studies, can be considered to prepare one or more confirmations of the target protein. So, you cannot model all possible confirmations, but you try to get more structures. If you get multiple confirmations, you include both of them in the docking ensemble, and in that case, molecular dynamics help get receptor flexibility.

So, MD simulations have been long proposed to provide insight into protein dynamics beyond that available crystallographically and unravel novel cryptic binding sites expanding the drug ability of the targets. So, I have already talked about MD simulation, and you include MD simulation. However, you remember that MD simulation is mathematically very intense, but introducing these would help you look at the dynamicity.

And we all know an even a low-resolution movie is much better or much truthful than a high-resolution picture. One of the first described approaches was the relaxed complex scheme which combines all-atom nanosecond long MD simulation of the protein target to describe its conformational flexibility with rapid docking of small molecules to the protein snapshots saved along the MD run. So, you do a classic MD run. You get the ensemble, look at the flexible conformational flexibility, take different conformers and do docking.

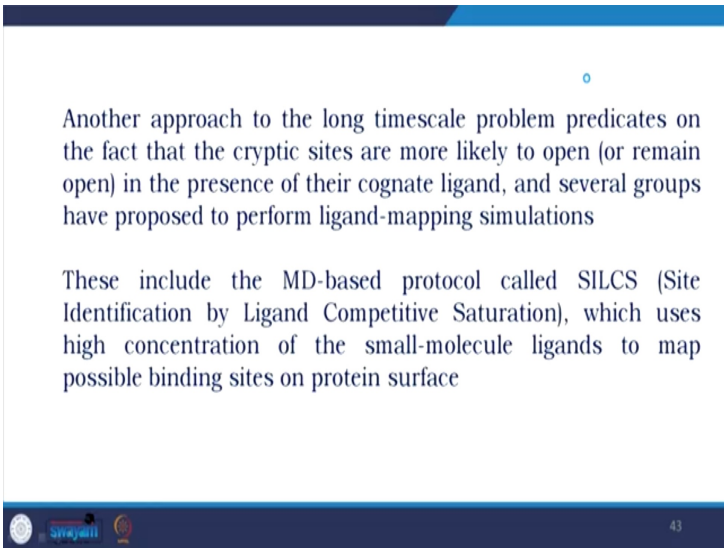
(Refer Slide Time: 01:22:11)



This scheme has been some successful applications, including one to the cancer-relevant MDM 2 P53 interaction of HIV, one is MDM2 p53 interaction, and the other is HIV integrase. Rewardingly, the cryptic pocket first proposed in silico in the latter case has been validated experimentally and exploited by enzyme inhibitors approved for use against HIV infection in the clinic.

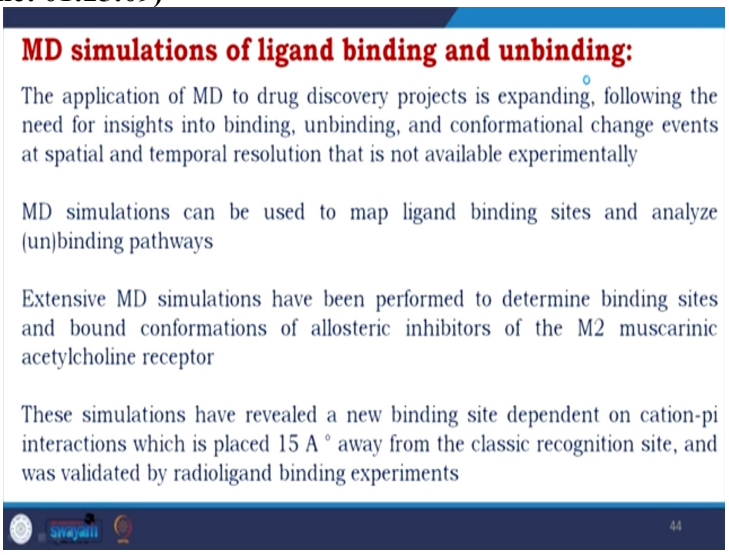
While the idea is conceptually correct, often, the conformational change needed to uncover the pocket would not be observed at the one-millisecond timescale accessible to explicit solvent MD simulations with conventional protocols on commodity compute clusters highlighting the need for more sophisticated sampling schemes. So, I could easily explain this. You know that protein has a lot of local minima. So, in most cases, it is trapped there, and you do not see the changes. You cannot push it from one minimum to the other or local minima to the global. Still, there are many other methods like replica-exchange like meta dynamics, so there are also pious methods. The molecular dynamics induced-fit MD IF protocol uses atomistic simulation to increase or reduce the aperture of the sub-pocket in the active site.

(Refer Slide Time: 01:24:23)



Another approach to the long timescale problem is predicted because cryptic sites are more likely to open in the presence of their cognate ligand, and several groups have proposed to perform ligand mapping simulation. So, you have the protein. You see the dynamics of the protein, the flexibility of the protein, then add the ligand and see where the flexibility is reduced. These include the MD-based protocol called SILCS, Site Identification by Ligand Competitive Saturation which uses a high concentration of the small molecule ligands to map possible binding sites on the protein surface.

(Refer Slide Time: 01:25:09)

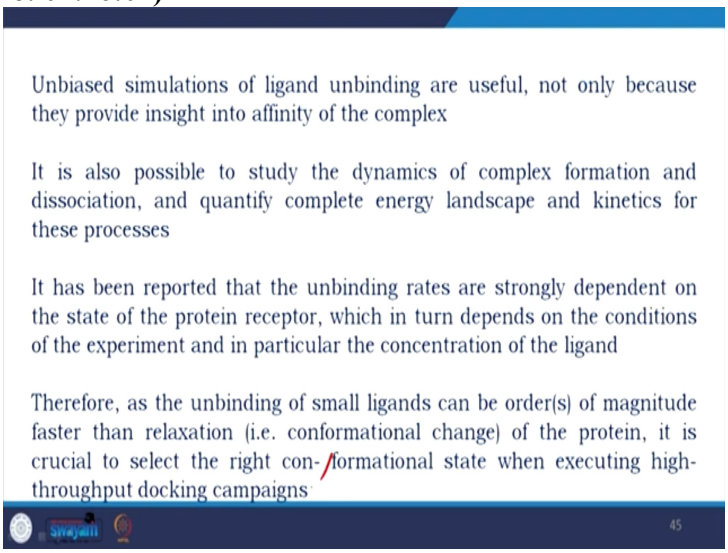


MD simulation of ligand binding and unbinding, the application of MD to drug discovery project is expanding following the need for insights into binding, unbinding and conformational change events at spatial and temporal resolution that is not available

experimentally. MD simulations can be used to map ligand binding sites and analyze unbinding pathways.

Extensive MD simulations have been performed to determine binding sites and bound conformation of allosteric inhibitors of the M2 muscarinic acetylcholine receptor. These simulations have revealed a new binding site dependent on cation pi interactions which is pleased 15 angstrom away from the classic recognition site and was validated by radio ligand binding experiments.

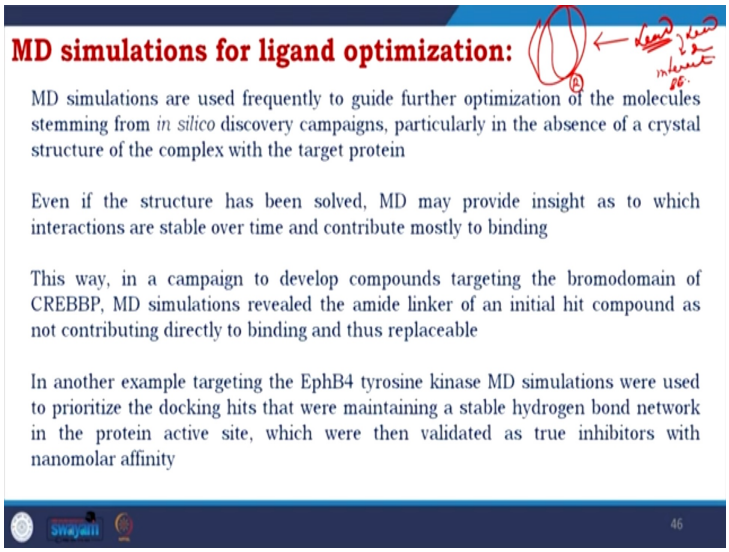
(Refer Slide Time: 01:26:07)



Unbiased simulation of ligand unbinding are useful not only because they provide insight into affinity of the complex, so, you have already known information is good, but then if you want to see a blind way of looking at the interaction of protein and ligand. It is also possible to study the dynamics of complex formation and dissociation and quantify complete energy landscape and kinetics for the processes.

It has been reported that the unbinding rates are strongly dependent on the state of the protein receptor which in turn depends on the condition of the experiment and in particular the concentration of the ligand. Therefore, as the unbinding of small ligands can be orders of magnitude faster than relaxation of the protein, it is crucial to select the right conformational state when executing high throughput docking campaigns.

(Refer Slide Time: 01:27:09)

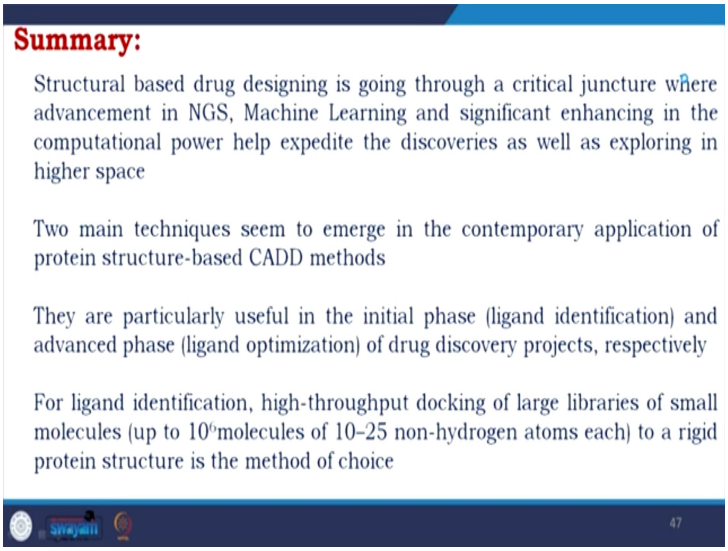


MD simulation for ligand optimization, MD simulations are used frequently to guide further optimization of the molecules stemming from in silico discovery campaigns, particularly in the absence of a crystal structure of the complex with the target protein. Even if the structure has been solved, MD may provide insight as to which interactions are stable over time and contribute mostly to binding. As I told you have the ligand, so, you have the protein structure high resolution, you have the receptor, you have the lead.

Now, you make lead derivatives and do simulations see the interactions as well as binding energy calculation. This way in a campaign to develop compounds targeting the bromodomain of CREBBP, MD simulation revealed the amide linker of an initial hit compound as not contributing directly to binding and thus replaceable. In other example, targeting EphB4 tyrosine kinase MD simulation were used to prioritize the docking hits that were maintaining a stable hydrogen bond network in the protein active site.

Which were then validated as true inhibitors with nanomolar affinity. So, there is a correlation coming up, there is a protocol developed now going through.

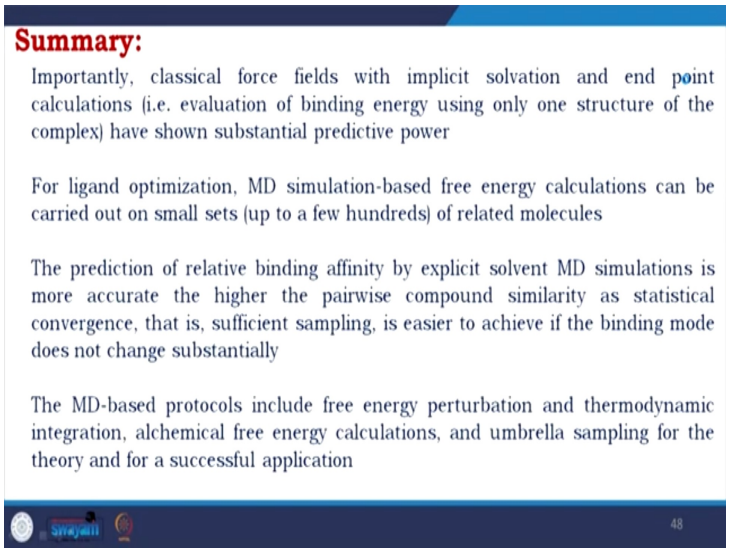
(Refer Slide Time: 01:28:42)



So, coming to the summary of the class structure based drug designing or drug discovery is going through a critical juncture a very prime time where advancement in NGS next generation sequencing, machine learning in biology and significant enhancing in the computational power help expedite the discoveries as well as exploring in higher space, higher numbers.

2 main techniques seem to emerge in the contemporary application of protein structure based computer assisted drug discovery methods. They are particularly useful in the initial phase which is ligand identification and also advanced phase ligand optimization of drug discovery projects. For ligand identification, high throughput docking of large libraries of small molecules up to 10 to the power 6 molecules of 10 to 25 non hydrogen atoms each to a rigid protein structure is a method of choice.

(Refer Slide Time: 01:29:45)

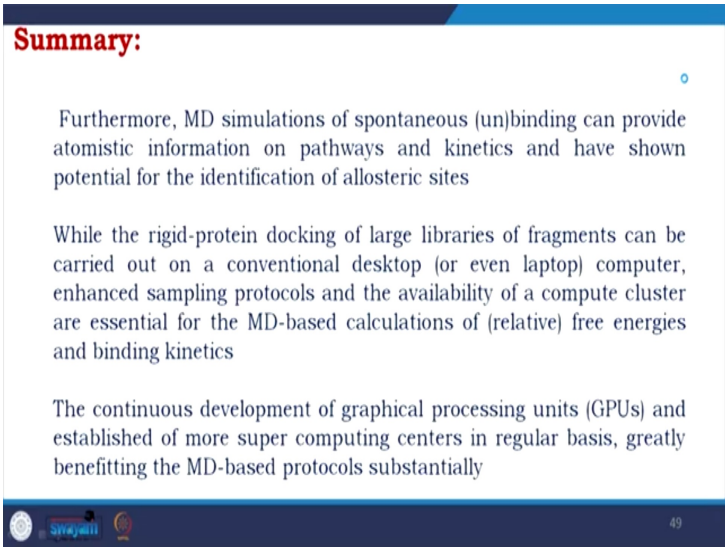


Importantly, classical force fields with implicit solvation and end point calculations that is the evaluation of binding energy using only one structure of complex have shown substantial predictive power. So, you need 4 structures, 5 structures, but when you are doing simulation, you are actually going through the ensembles, you are actually going through the structure space. For ligand optimization, MD simulation based free energy calculations can be carried out on small sets up to few 100 of related molecules.

So, once you make the initial cut, you do further studies coming in 100 and do that process. The prediction of relative binding affinity by explicit solvent MD simulation is more accurate, the higher the pair wise compound similarity as statistical convergence that is, sufficient sampling is easier to achieve if the binding mode does not change substantially. If it is very flexible, if after binding even it is not stable then it is a problem, otherwise, it is totally good.

The MD based protocols include free energy perturbation, FEP and thermodynamic integration, TI, alchemical free energy calculation, umbrella sampling for the theory and for successful application.

(Refer Slide Time: 01:31:14)



Furthermore, MD simulation of spontaneous unbinding or binding can provide atomistic information on pathways and kinetics and have shown potential for the identification of allosteric sites. So, if you remember, I was constantly comparing with the high throughput the traditional method, in the traditional method, if we get effects, we do not know what is happening.

But now, with high resolution structure we could see but also in the MD simulation, each and every change, their effects, their interactions, their energies, everything is in front of you, so, it is much easier for us to take accurate decision. While the rigid protein docking of large libraries of fragments can be carried out on a conventional desktop or even laptop computer enhanced sampling protocols and the availability of a compute cluster are essential for the MD based calculation of relative free energy and binding kinetics.

The continuous development of graphical processing units, the GPUs and establishment of more supercomputing centers in regular basis also benefiting the MDS protocol substantially because now, we have access to higher computational power. So, with that, I would make an end of this description of structure based drug designing. Mainly docking based virtual screening.