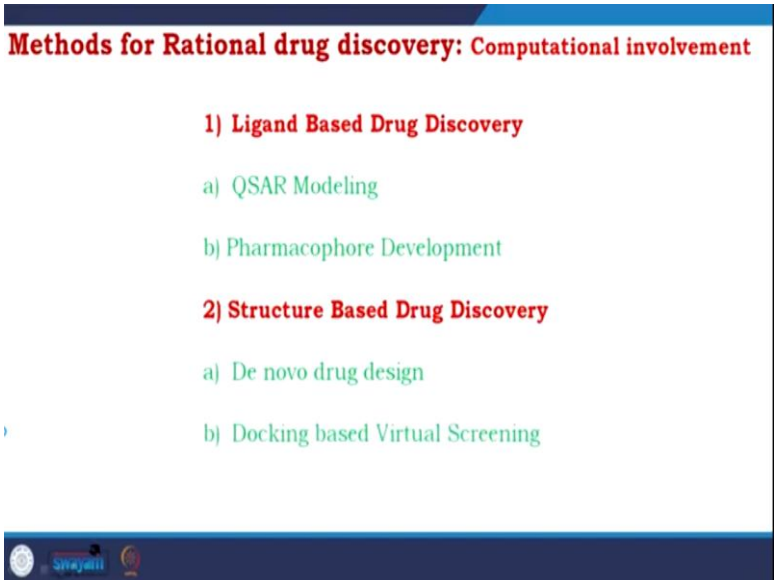
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology – Roorkee

Lecture – 57
Rational Drug Discovery

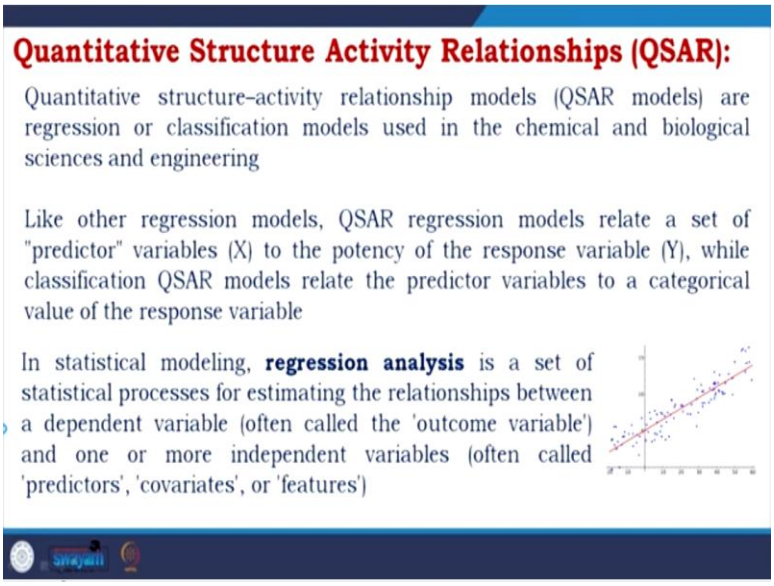
Hi everyone, welcome again to the course on structural biology. We are continuing with our last module, structure-based drug discovery. In the last class, the introductory class of structure-based drug discovery, I talked about history, how drug designing developed, and about methodologies. Today, my discussion would be on logic base or rational drug discovery.

(Refer Slide Time: 01:38)



So, I will talk about the methods for rational drug discoveries, especially the enhancement in the speed of those innovations with the incorporation or advancement of computation. So, as I told earlier, rational drug discovery has been divided into two major parts, ligand-based drug discovery, where two major methodologies are QSAR modeling and Pharmacophore development. At the same time, another is structure-based drug discovery, where De novo drug designing, where the drug designing is happening from scratch without any prior information, and Docking-based virtual screening.

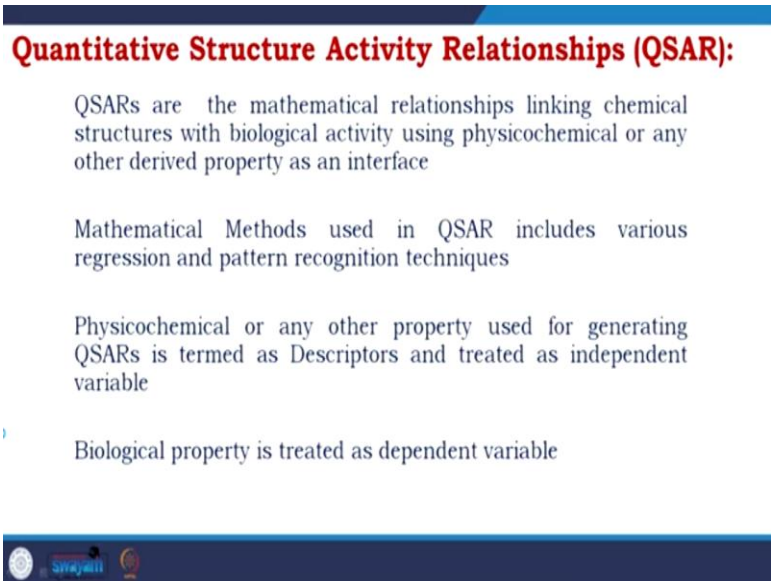
(Refer Slide Time: 03:37)



Quantitative structure-activity relationship (QSAR) models are regression or classification models used in chemical, biological, and engineering. Like other regression models, QSAR regression models relate a set of predictor variables X to the potency of the response variable Y. In contrast, classification QSAR models relate the predictor variables to a definite value of the response variable. So, when I am talking about my definition of QSAR is based on regression.

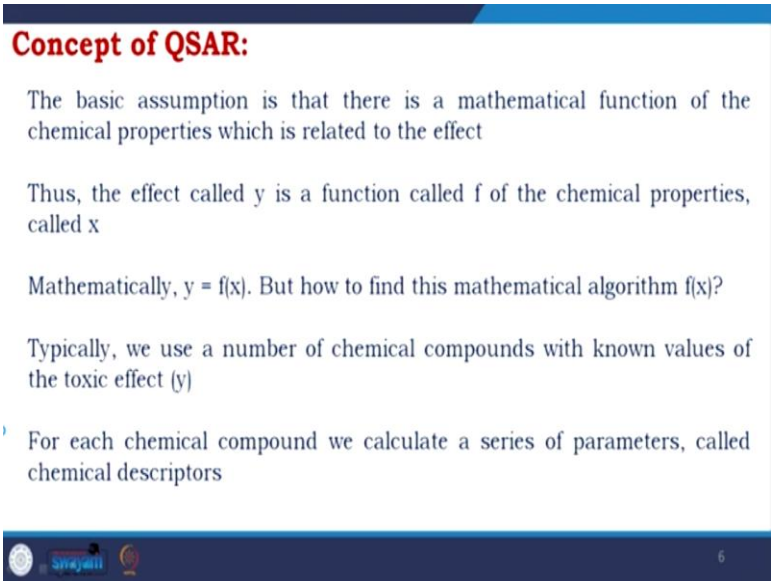
And if you do not know about regression, regression is used in statistical modeling. Regression analysis is a set of statistical processes for estimating the relationship between a dependable variable called outcome variable and one or more independent variables, often called predictors, covariates, or features.

(Refer Slide Time: 05:55)



QSARs are the mathematical relationships linking chemical structure with a biological activity using physical, chemical, or other derived property as an interface. Mathematical methods used in QSAR include various regression and pattern recognition techniques, physicochemical or any other property used for generating QSAR is termed as Descriptors and treated as an independent variable. Biological properties are treated as a dependent variable. So, your descriptors could be coming from the property of the small molecule, which could be related to the biological factors, biological properties that the descriptors are independent whereas, biological properties are dependent.

(Refer Slide Time: 06:53)

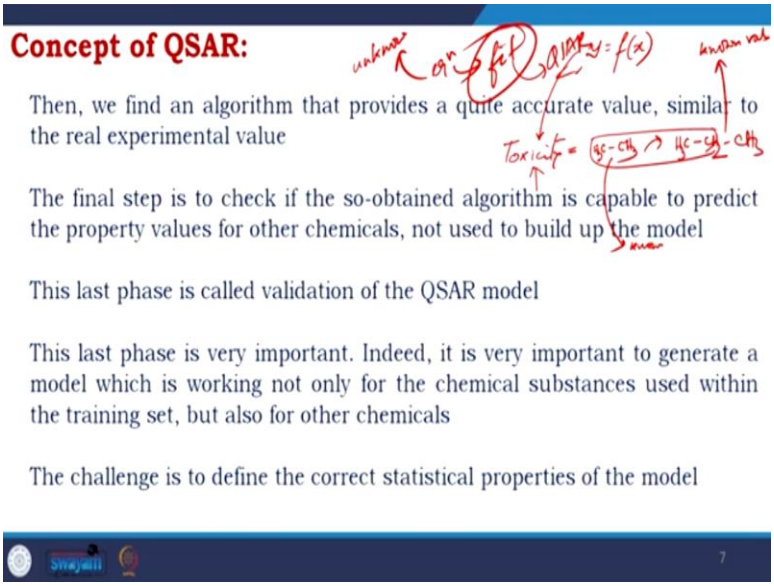


The basic assumption is that there is a mathematical function of the chemical properties related to the effect in terms of concept. The effect is called y is a function called f of the chemical property called x

mathematically y = f[x], but how to find this mathematical algorithm and make that equation y equal to fx for a particular compound and its biological property.

Typically, we use several chemical compounds with the known values of that toxic effect. We calculate a series of a parameter called chemical descriptors,

(Refer Slide Time: 07:46)

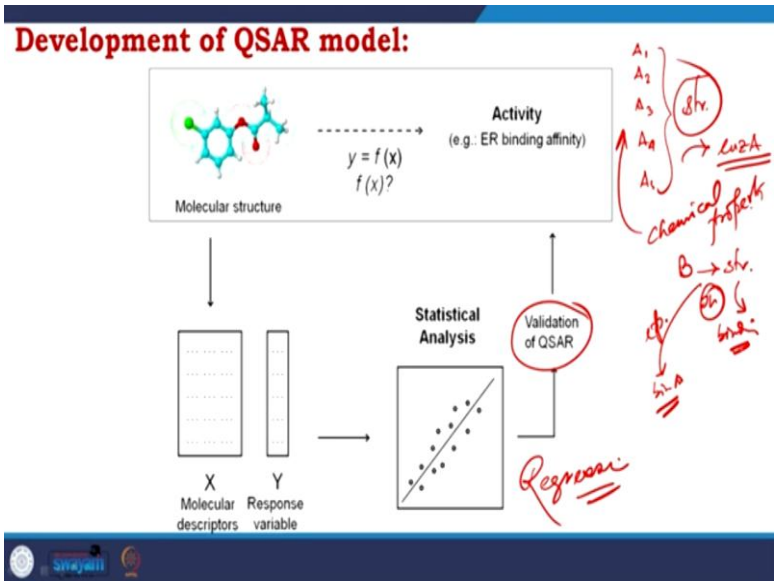


Then, we find an algorithm that provides a quite accurate value similar to the real experimental values to replace the experimental performance. The final step is to check if the so obtained algorithm can predict the property values of other chemicals not used to build up the model. So, you have y = f [x]. Now, you take y as a property like the toxicity of a small molecule.

If you have CH 3 – CH 3 and it changed to CH 3 – CH 2 - CH 3, how toxicity would be changed. So, you have known value, and you see the change and try to fit an equation. Now, when this relation is fixed by using many searches, you get an unknown one to apply that equation and see if it fits. If it fits, you have a good QSAR. The last phase is to check if the algorithm can predict the property values of other chemicals. This is called the validation of the QSAR model. So, here based on the known data, you develop a model and then validate it with the unknown chemical. The last phase, the validation phase, is very important. It is very important to generate a model that works not only for the chemical substances used within the training set but also for the other chemicals, the unknown set the not included training set. The challenge is to define the correct statistical properties of the model.

So, you could take QSAR in a way, you have the chemical compound you could have made, or you could have described many properties of the molecule, let us say solubility, let us say it is reactivity, let us say it is chirality all of these are properties, which could correlate to the biological activity. But instead of doing that, if you get numbers, that could be a much more understandable way to relate them. That is what QSAR is doing.

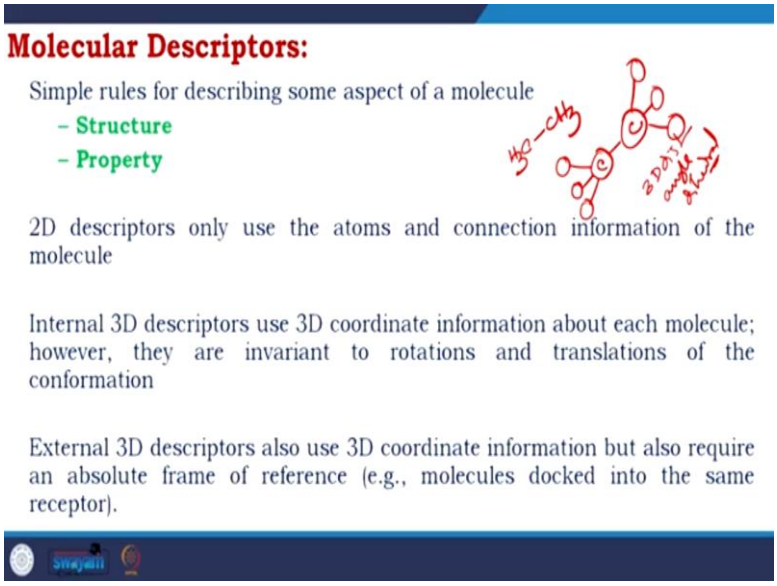
(Refer Slide Time: 11:02)



So, this is the user model, and you have the molecule or structure, you have the relation of y equal to f[x], you have to set effects, you know the activity like there is a structure, and you know it is binding activity. So, you have molecules A1, A2, A3, A4, A5. You know their structures and binding affinity towards enzyme A. So, now you have the molecular descriptors, they describe the chemical property of this group of compounds. Then, there would be the binding affinity you develop a relationship between them, which would come through linear regression.

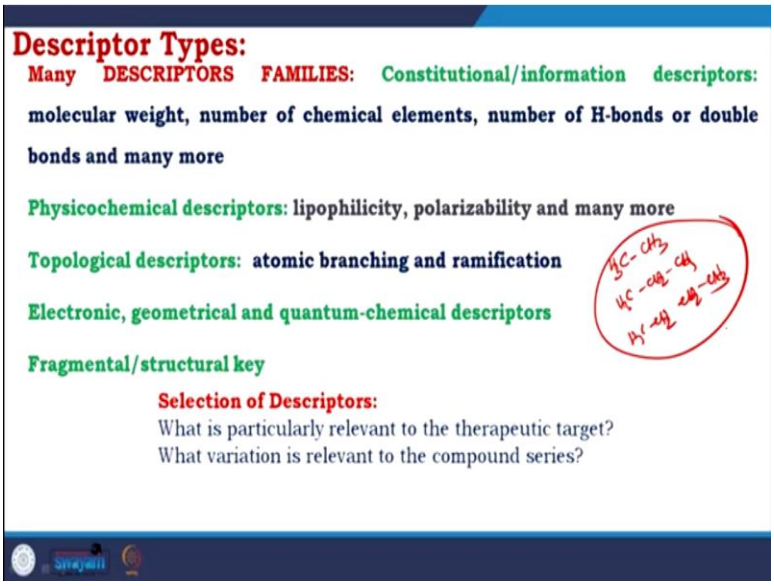
And then, when you set up the model, you validate it by bringing B. For B, the structure, you determine the binding energy theoretically using this equation, and then you do it experimentally. It is binding affinity, and see these two are coming closer or not. If these two come closer, then your model is good.

(Refer Slide Time: 12:37)



Molecular descriptors are simple rules for describing some aspect of a molecule, as you could talk about the structure and the property. 2D descriptors only use the atoms and connection information of the molecule. Internal 3D descriptors use 3D coordinate information about each molecule. So, it will show the 3D distance angle, dihedral, and all that information. However, they are invariant to rotation and translation of the conformation. External 3D descriptors also use 3D coordinate information and require an absolute frame of reference molecule docked into the same receptor.

(Refer Slide Time: 14:22)



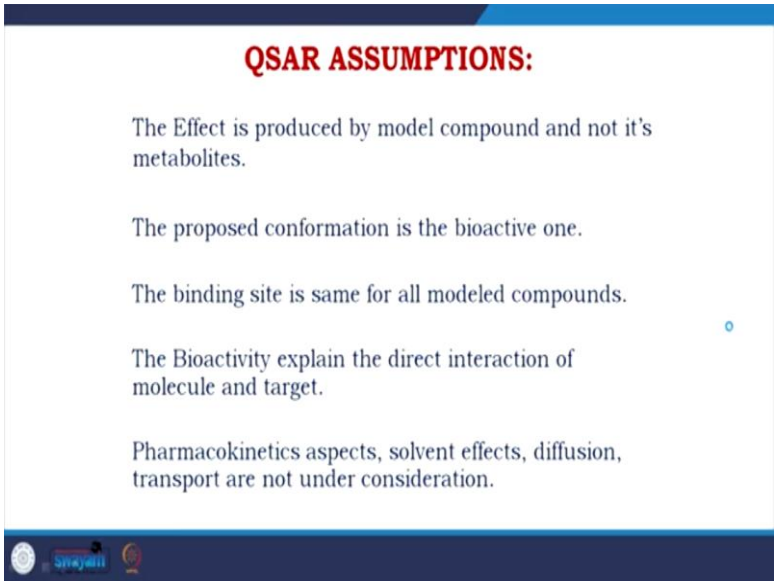
Descriptor types: descriptors you already understand have a very important role in the QSAR model. You understand the chemical molecule by the definitions or descriptions from the descriptors. So, there are many descriptive families constitutional or information descriptors that talk about molecular weight, number of chemical elements, number of hydrogen bonds or double bonds, and many more.

Physicochemical descriptors talk about lipophilicity, polarizability, and many others, and topological descriptors talk about atomic branching and ramification. Electronic geometrical and quantum chemical descriptors are there. When you go for fragment-based rock designing, the fragmental structural key is needed because fragmentation means you fragment eyes a molecule and how the properties are changed.

So, it is critical to select the descriptors. What is particularly relevant to the therapeutic target? If you ask somebody, is this person honest? And the guy replies he is very good at academics. So, the question did not get answered. So, you have to choose proper questions.

So, how the descriptors would relate to the biological activity, your question selection should depend on that. What variation is relevant to the compound series? So, you have to understand those questions when you select the descriptors.

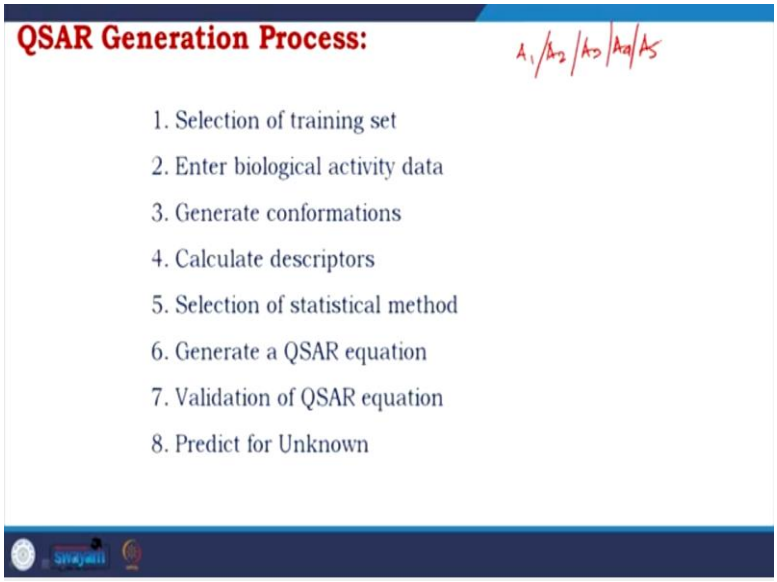
(Refer Slide Time: 17:23)



There are a few assumptions when you perform QSAR model development. The effect is produced by the model compound and not its metabolites. So, if we think about a biological molecule involved in a metabolic pathway, we should know the compound we are targeting and not the transformed version. The proposed conformation is the bioactive one. The structure you are selecting should have the optimized conformation related to the bioactivity we are looking for.

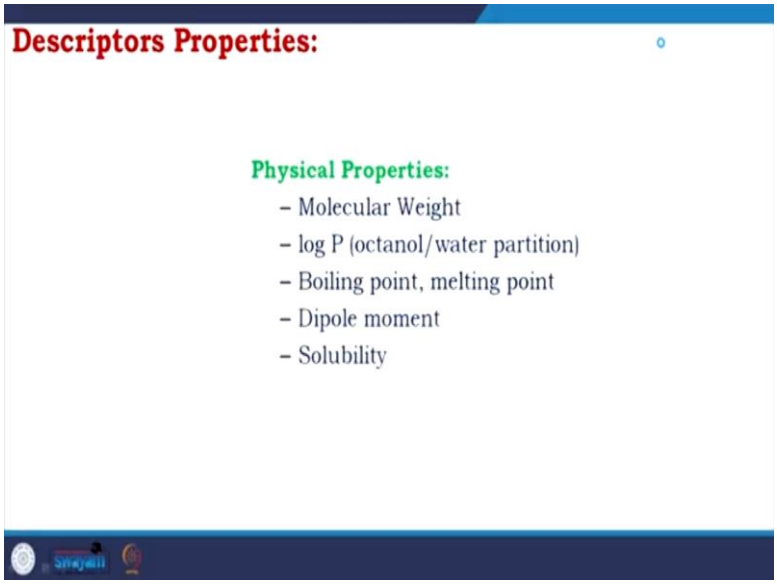
The binding site is the same for all model compounds. If you consider a series of molecules towards binding to an enzyme or protein; it should bind to the same biological macromolecules. The bioactivity explained the direct interaction of molecule and target. Pharmacokinetics aspects solvent effects diffuse and transport are not under consideration. So, how you make that model simplified depends on those assumptions.

(Refer Slide Time: 18:57)



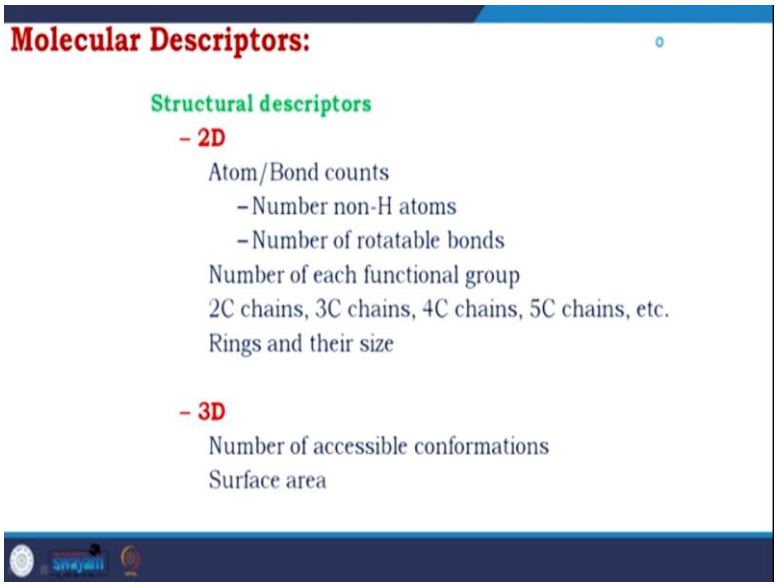
So, how you go for model development, first selection up training set you will take molecules A 1, A 2, A 3, A 4, A 5 you have to select them first. Enter biological activity data, Generate confirmations, Calculate the descriptors, select the statistical method, Generate QSAR equation, Validate the QSAR equation, predict it for the unknown data set. So, that is the process for generating or working with a QSAR model development.

(Refer Slide Time: 19:45)



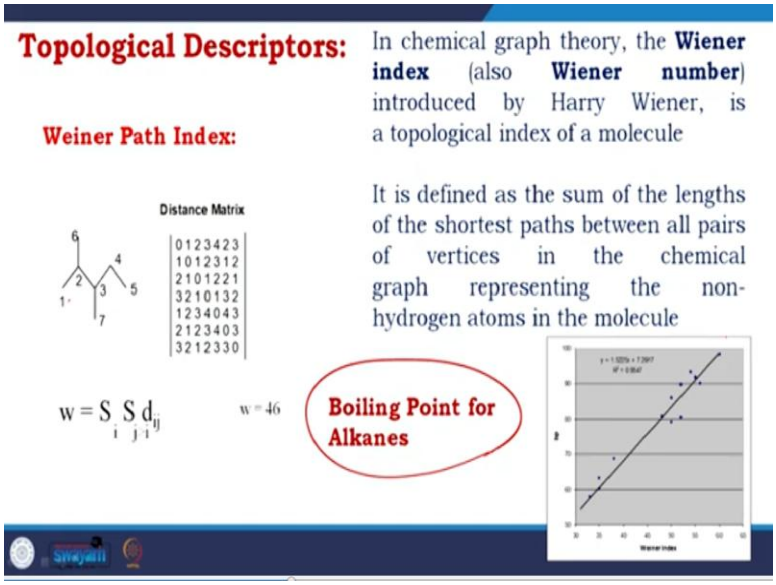
Descriptor properties could have physical property like molecular weight, log P, log P, is the permeability of a compound in octanol and water. Because octanol is an organic solvent and water is an aqueous solvent. Boiling point, melting point, dipole moment, and solubility, so, all the physical properties.

(Refer Slide Time: 20:21)



Structural descriptors are divided into two parts 2D and 3D. In 2D, as we have explained earlier, you have to look at atom bond counts, number of non-hydrogen atoms, number of rotatable bonds, number of each functional group, two carbon chains, three-carbon chains, four-carbon chains, five-carbon chains etc, Rings and their size. In a 3D number of accessible confirmations which are biologically active, and surface area.

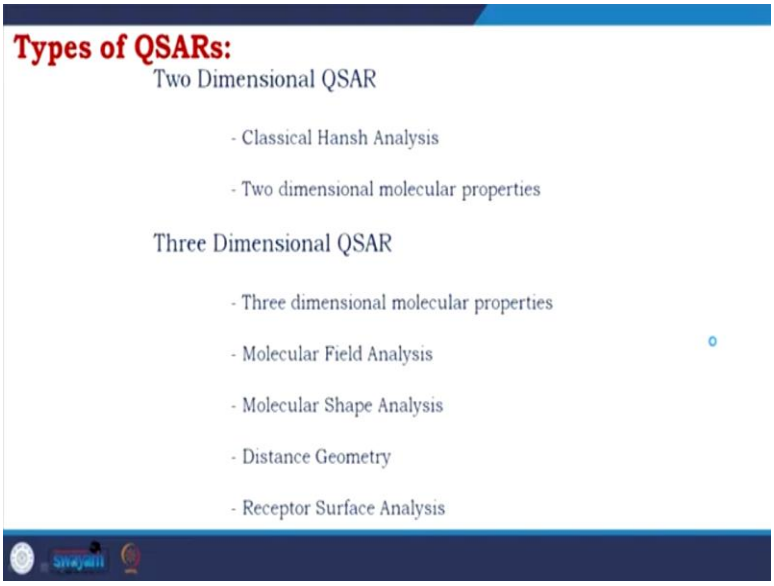
(Refer Slide Time: 21:04)



Topological descriptors: topological descriptors are defined by the Wiener path index. What is the Weiner path index? In chemical graph theory, the Wiener index, also called the Wiener number introduced by Harry Wiener, is a topological index of a molecule. It is defined as the sum of the lengths of the shortest paths between all pairs of vertices in the chemical graph representing the non-hydrogen atoms in the molecule.

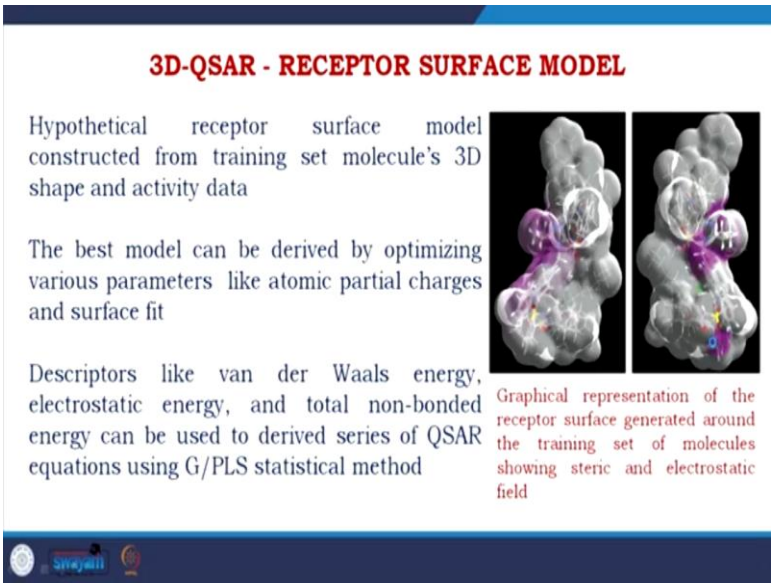
This relation could be calculated when the boiling point for alkanes is considered. So, the y-axis is the boiling point, and the x-axis is the Wiener index. You see that there is a very nice correlation. So, you could successfully use topological descriptors using Wiener number with the molecule's biological activity.

(Refer Slide Time: 22:26)



Type of QSARs: 2-dimensional QSAR like classical Hansh analysis, and 2-dimensional molecular properties. 3-dimensional QSAR like 3-dimensional molecular properties, molecular field analysis, molecular shape analysis, distance geometry, receptor surface analysis.

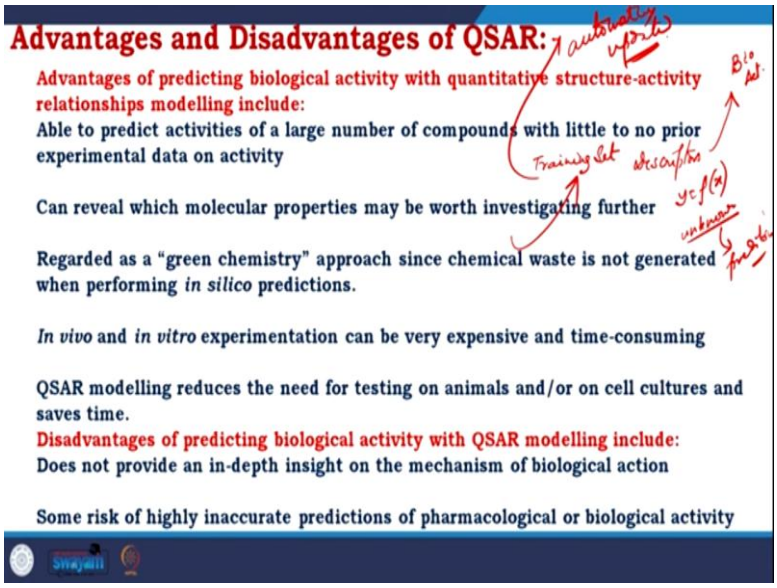
(Refer Slide Time: 22:48)



So, let us look at a case study of the receptor surface model. Hypothetical receptor surface model constructed from training set molecules 3D shape and activity data. The best model can be derived by optimizing various parameters like atomic partial charges and surface fit. Descriptors like Van der Waals energy, electrostatic energy, and total non-bonded energy can derive a series of QSAR equations using the G-PLS statistical method.

And when you do that, you see this is a graphical representation of the receptor surface generated around the training set of molecules showing steric and electrostatic fields. So, by generating these, you could describe you could develop a much better model from the original structure, which would be beneficial towards its further interaction with biological macromolecules.

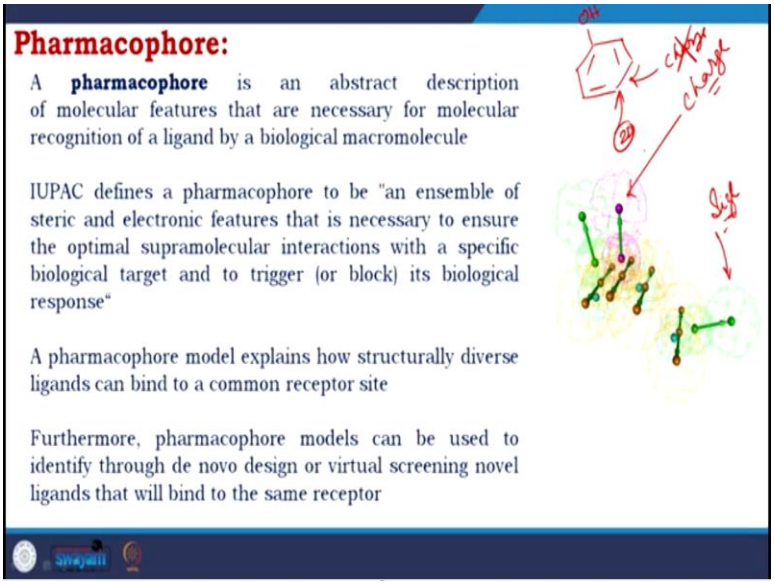
(Refer Slide Time: 23:53)



Advantages and disadvantages of QSAR: Advantages of predicting biological activity with quantitative structure-activity relationship modeling include predicting activities of a large number of compounds with little to no prior experimental data on activity. So, using QSAR, it is possible to predict activities of many compounds with little to no prior experimental data on activity is extremely beneficial towards drug development. It can reveal which molecular properties may be worth investigating further. It is a green chemistry approach since chemical waste is not generated when performing in silico predictions. In vivo and in vitro experimentation can be very expensive and time-consuming. Also, QSAR modeling reduces the need to test animals and cell cultures and saves time. So, you already get a correlation so, you get selectivity.

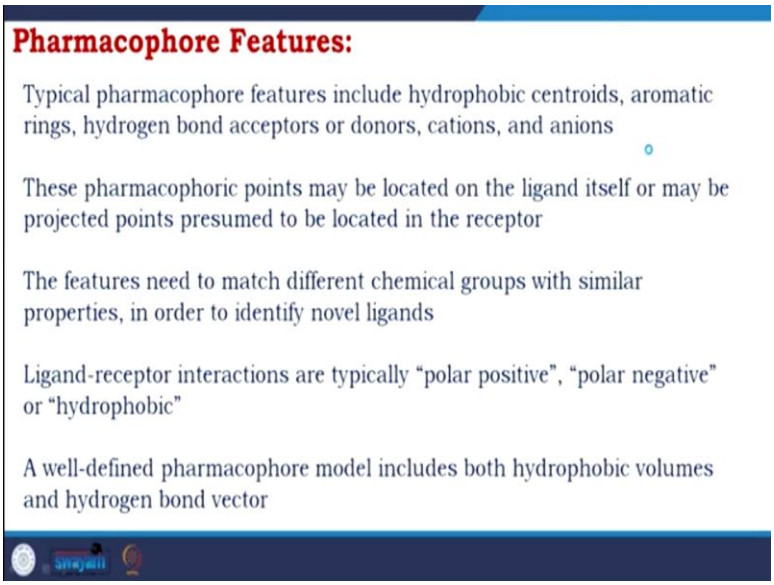
The disadvantages of QASR do not provide in-depth insight into the mechanism of biological action. Some risk of highly inaccurate prediction of pharmacological or biological activity is always there.

(Refer Slide Time: 27:42)



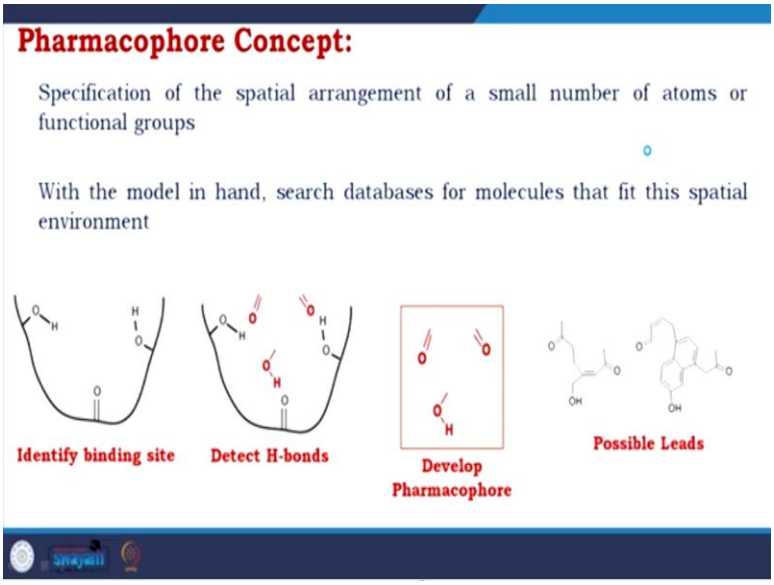
A pharmacophore is an abstract description of molecular features necessary for molecular recognition of a ligand by biological macromolecules. IUPAC defines a pharmacophore as an ensemble of steric and electronic features necessary to ensure the optimal supramolecular interactions with a specific biological target and trigger its biological response. So, size and charge are the major factor of a small molecule to predict how it interacts with the biological molecule. And that is what pharmacophore is doing. A pharmacophore model explains how structurally diverse ligands bind to a common receptor site. Furthermore, pharmacophore models can be used to identify through de novo designed or virtual screening novel ligands that will bind to the same receptor.

(Refer Slide Time: 29:59)



Pharmacophore Features: Typical pharmacophore features include hydrophobic centroids, aromatic rings, hydrogen bond acceptors or donors, cations, and anions. These pharmacophoric points may be located on the ligand itself or projected points presumed to be located in the receptor. The features need to match different chemical groups with similar properties to identify novel ligands. Ligand receptor interactions are typically polar positive, polar negative, and hydrophobic. The pharmacophore model will define both hydrophobic volumes and hydrogen bond vectors.

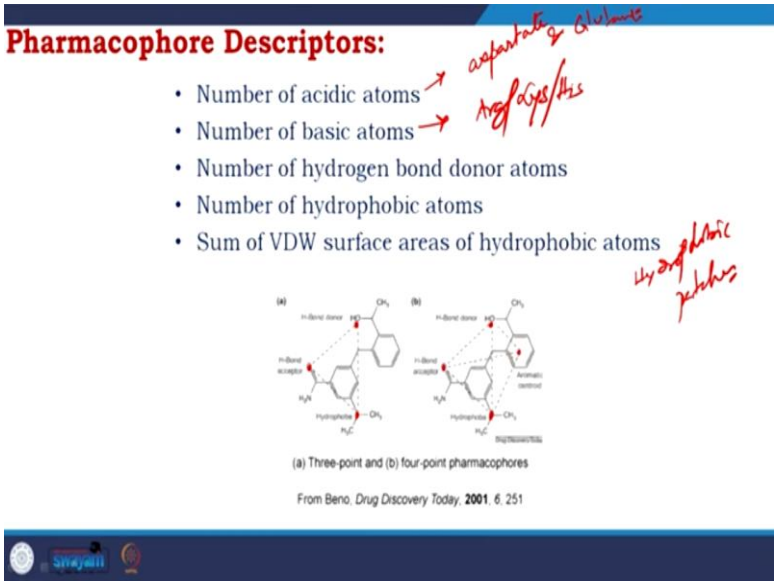
(Refer Slide Time: 31:12)



Pharmacophore concept: Specification of a small number of atoms or functional groups' spatial arrangement. You have a small molecule, and you are not bothered about that architecture. You are bothered about the representative charges, how they are presented, and the volume so that you could fit. So, you are pointing out the important parts of a small molecule, which might play a critical role in the interaction.

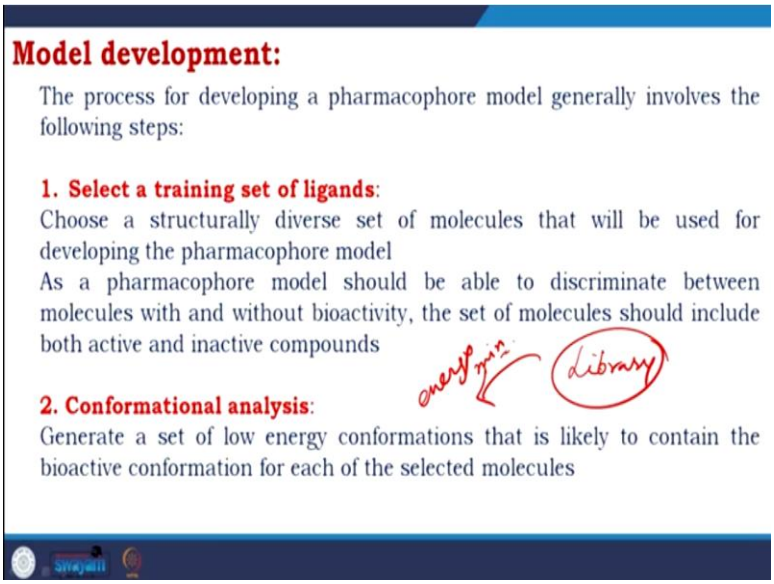
With the model in hand, search databases for molecules that fit this spatial environment. So, instead of taking a complete molecule, you will take the pharmacophore model you get to identify the similar type of interactor similar type of 3-dimensional volume containing molecules.

(Refer Slide Time: 33:05)



Pharmacophore descriptor: Number of acidic atoms means negatively charged atoms, number of basic atoms means positively charged atoms. The number of hydrogen bond donor atoms, number of hydrophobic atoms, the sum of van dar Waal surface area of a hydrophobic atom means hydrophobic patches.

(Refer Slide Time: 34:22)



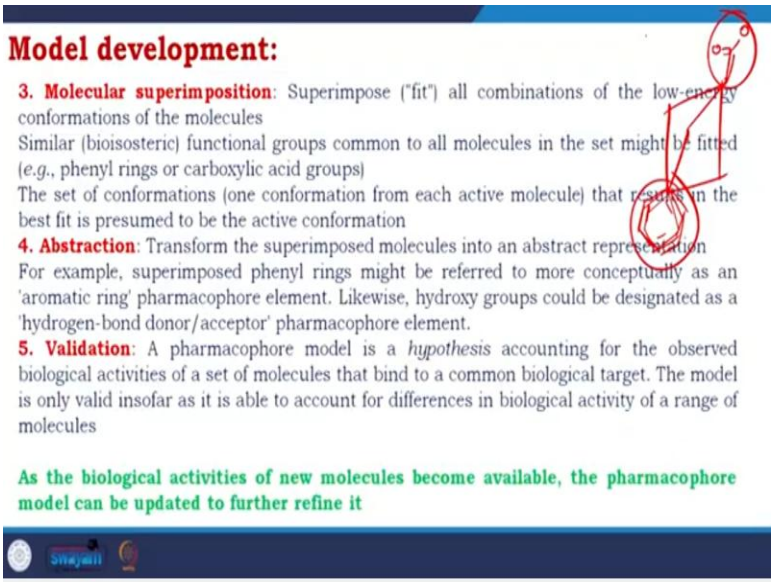
How do you develop a pharmacophore model? The process for developing a pharmacophore model generally involves the following steps.

Select a training set of ligands: Choose a structurally diverse set of molecules that will be used for developing the pharmacophore model. As a pharmacophore model should discriminate

between molecules with and without bioactivity, the set should include both active and inactive compounds.

Conformational analysis: Generates a set of low-energy conformations likely to contain each selected molecule's bioactive conformation.

(Refer Slide Time: 35:32)



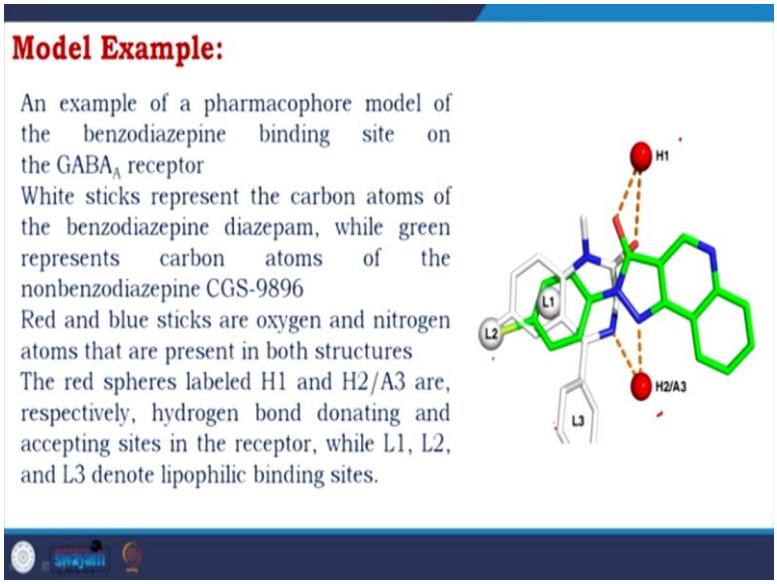
Molecular Superimposition: Superimpose all combinations of the low energy conformations of the molecule. Similar (bioisosteric) functional groups common to all molecules in the set might be fitted. For example, if they had a phenyl ring, all the phenyl rings common, be there close to each other, or all that acid groups might be the compounds are different, but those points would be fitted together. The set of conformations, one conformation from each active molecule that results in the best fit, is presumed to be the active conformation.

Abstraction: Transform the superimposed molecule into an abstract representation. For example, superimposed phenyl rings might be referred to more conceptually as an aromatic ring pharmacophore element; likewise, hydroxyl groups could be designated as hydrogen bond donor-acceptor pharmacophore elements in that way.

Validation: A pharmacophore model is a hypothesis accounting for the observed biological activity of a set of molecules that bind to a common biological target. The model is only valid as it can account for differences in the biological activity of a range of molecules. So, there are so many ligands they are diverse, but they represent their functional group, which could have the ability to bind to the receptor.

And then, if it correlates with the biological activity, the model is developed. As the biological activities of new molecules become available, the pharmacophore model can be updated to refine it further. More and more biological activity correlating to small molecules will be available updated better model could be obtained.

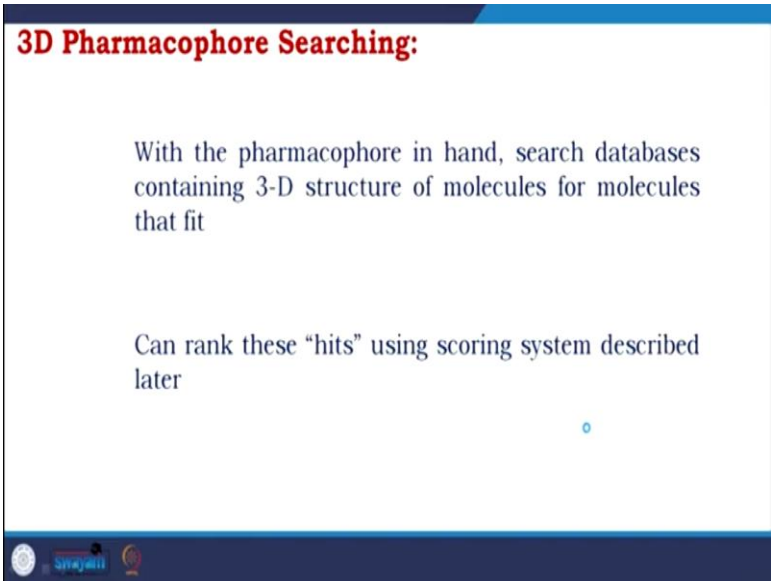
(Refer Slide Time: 38:05)



An example of a pharmacophore model of the benzodiazepine binding site on the GABA receptor. The white sticks represent the carbon atoms of the benzodiazepine diazepam, while green represents the carbon atom of the non-benzodiazepine CGS-9896. Red and blue sticks are oxygen and nitrogen.

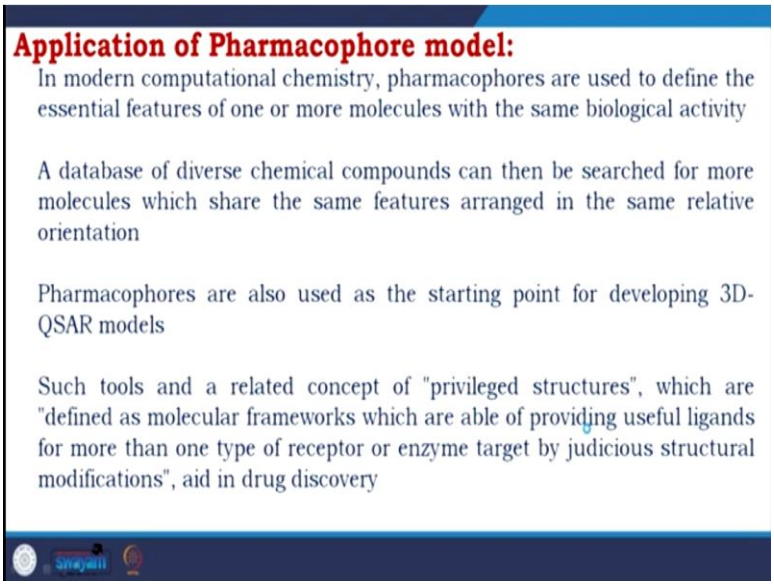
The red Spheres labeled H1 and H2/A3 respectively hydrogen bond donating and accepting site in the receptor, while L1, L2, and L3 denote lipophilic or hydrophobic binding.

(Refer Slide Time: 39:21)



How do you search the 3D Pharmacophore? With the pharmacophore in hand, search databases containing the 3D structure of molecules for molecules that fit. You could rank these hits using a scoring system.

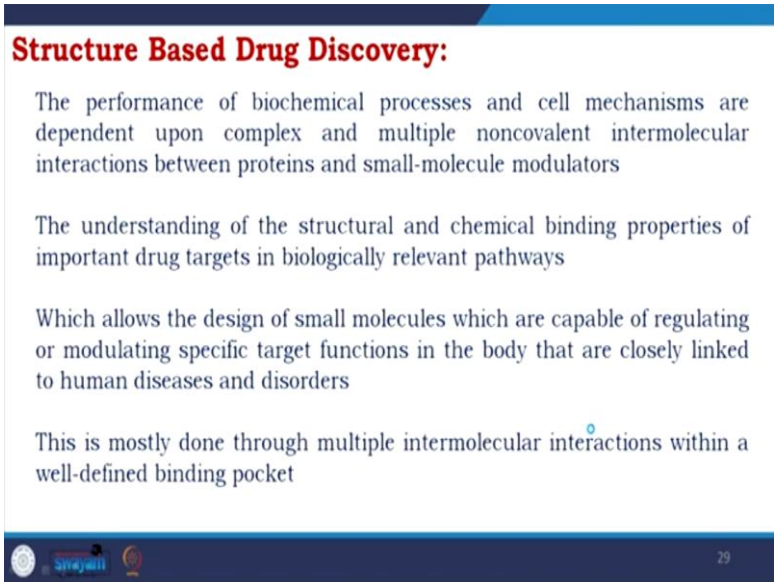
(Refer Slide Time: 39:47)



Application of pharmacophore model: In modern computational chemistry, pharmacophores define the essential features of one or more molecules with the same biological activity. A database of diverse chemical compounds can then be searched for more molecules, which share the same features arranged in the same relative orientation. Pharmacophores are also the starting point for developing a 3D QSAR model because they define the properties.

Now, you use those properties, develop descriptors, correlate them, and now get a new QSAR model. Such tools and a related concept of privilege structures, defined as molecular frameworks, which provide useful ligands for more than one type of receptor or enzyme target by judicious structural modification, would aid in drug discovery.

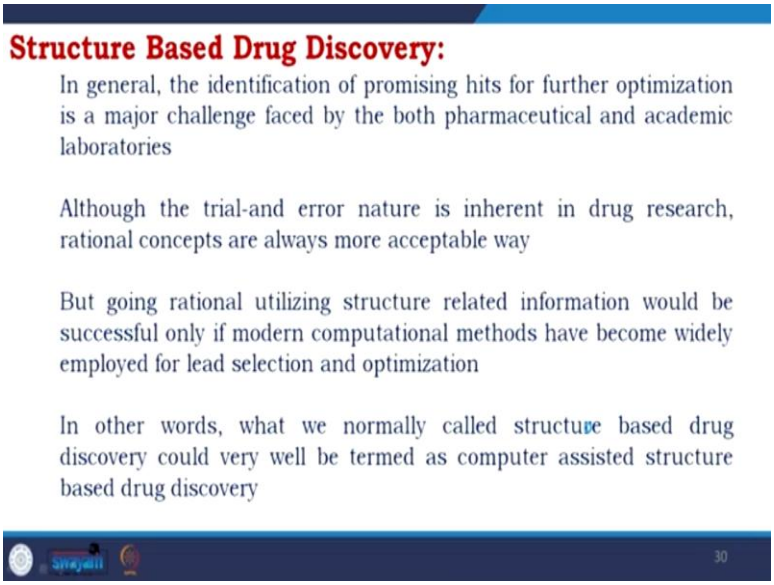
(Refer Slide Time: 40:50)



Structure-based drug discovery: The performance of biochemical processes and cellular mechanisms depends on complex and multiple non-covalent intermolecular interactions between proteins and small molecule modulators. Understanding the structural and chemical binding properties of important drug targets in biologically relevant pathways allows the design of small molecules capable of regulating or modulating specific target functions in the body.

That is closely linked to human diseases and disorders, which is our target. This is mostly done through multiple intermolecular interactions within a well-defined binding pocket. So, we look at a biological macromolecule, a target, and we generally try to get the binding pocket, which is the active site pocket in the case of an enzyme.

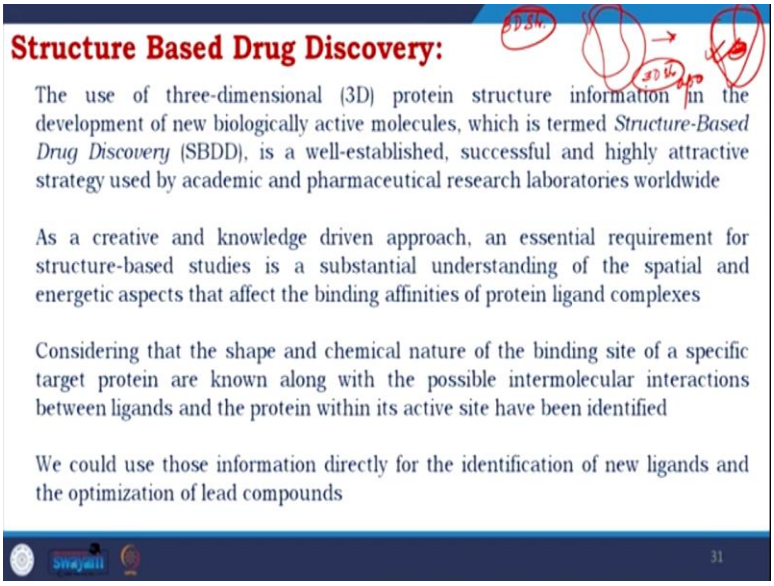
(Refer Slide Time: 41:56)



In general, identifying promising hits for further optimization is a major challenge faced by both pharmaceutical and academic laboratories. Although trial and error are inherent in drug research, rational concepts are always more acceptable. But going rational utilizing structure-related information would be successful. Only modern computational methods have become widely employed for lead selection and optimization.

So, you use the structure you get a high-resolution structure, but you have to use it to calculate by fitting with many that need the computation, the algorithms, and the ability to compute. In other words, what we normally call structure-based drug discovery could be very well termed as computer-assisted structure-based drug discovery.

(Refer Slide Time: 43:02)

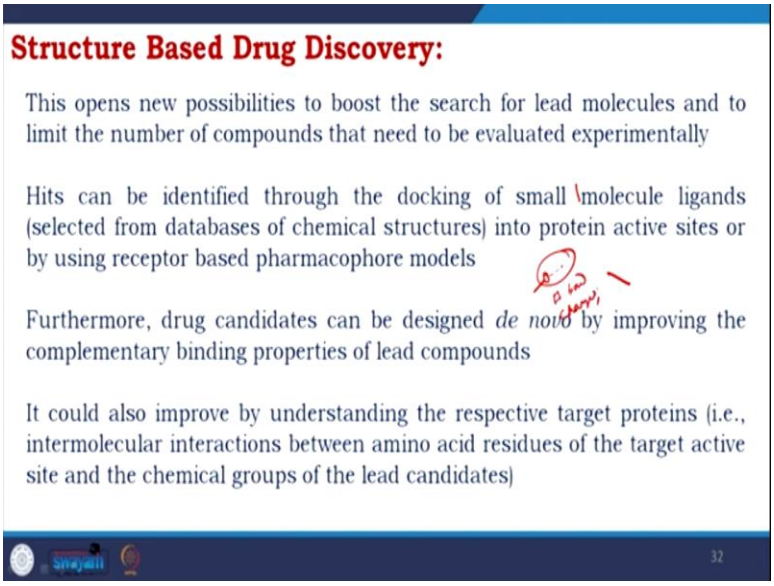


Structure-based drug discovery uses 3-dimensional protein structure information to develop new biologically active molecules. SBDD is a well-established, successful, and highly attractive strategy used by academic and pharmaceutical research laboratories worldwide. A creative and knowledge-driven approach as an essential requirement; for structured studies is a substantial understanding of the spatial and energetic aspects that affect the binding affinities of a protein-ligand complex.

So, as we started and talked continuously, drug designing is about the macromolecule and its interaction with small-molecule peptides or macromolecule. But we are now mostly focusing on the small molecule and peptides. Considering that the shape and chemical nature of the binding site of a specific target protein is known, along with the possible intermolecular interaction between ligands and the protein within its active site have been identified.

So, it is a very good point you want to the structure. But if you get a 3D structure of the protein in apo form, it is good, but it is not as good if you get it bound to the ligand. So, this will give you the set of interactions between the target molecule and the potential substrate or inhibitor, making your life more facile to work further from this existing foundation. We could use that information directly to identify new ligands and the optimization of lead compounds.

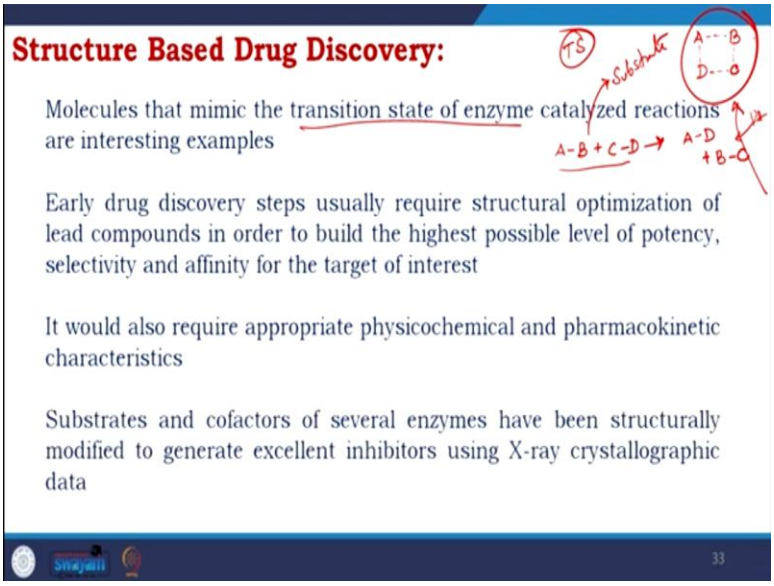
(Refer Slide Time: 45:10)



This opens new possibilities to boost the search for lead molecules and limit the number of compounds that need to be evaluated experimentally. Hits can be identified by docking small molecule ligands selected from a database of chemical structures into protein active sites or by using receptor-based pharmacophore models. Furthermore, drug candidates can be designed de novo by improving the complementary binding properties of the lead molecules.

It could also improve by understanding the respective target proteins, the intermolecular interaction between the amino acid residues of the target active site, and the chemical groups. So, if you have these, you know that there is an oxygen show how that interaction will happen, is it hydrogen bond, is it charge interaction and all those things.

(Refer Slide Time: 46:13)



Molecules that mimic the transition state of enzyme-catalyzed reactions are an interesting example; I will discuss these in the case study, but to explain a little bit. Transition state what is a transient state I talked about this, but like A-B, when reacting with C-D will form A-D + B-C. So, this is called the substrate. These are called product, but in between, there is a situation where the bond between A and B will be broken and the bond between C and D will also be broken.

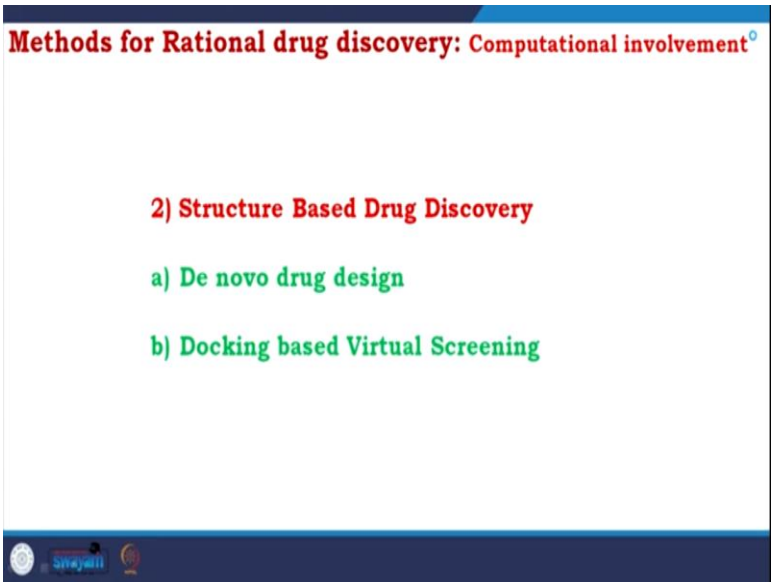
At the same time, they will bond between A and D going to form, the bond between B and C going to form. This situation which exists theoretically but does not exist practically, is called a transient state. Now, why the transient state of the enzyme is important? Because in the transition state, the enzyme generally shows its intrinsic property. And you remember it is very difficult or nearly impossible to change your intrinsic property.

And that is why if you somehow make it the stable molecule which mimics this transient state, they are always could drug I will talk about them. Early drug discovery steps usually require structural optimization of lead compounds to build the highest possible level of potency, selectivity, and affinity for the target of interest. It would also require appropriate physicochemical and pharmacokinetic characteristics.

Substrates and cofactors of several enzymes have been structurally modified to generate excellent inhibitors using X-ray crystallographic data because that gives you the highest resolution though, cryoelectron microscopy is surprisingly going higher to a higher level of

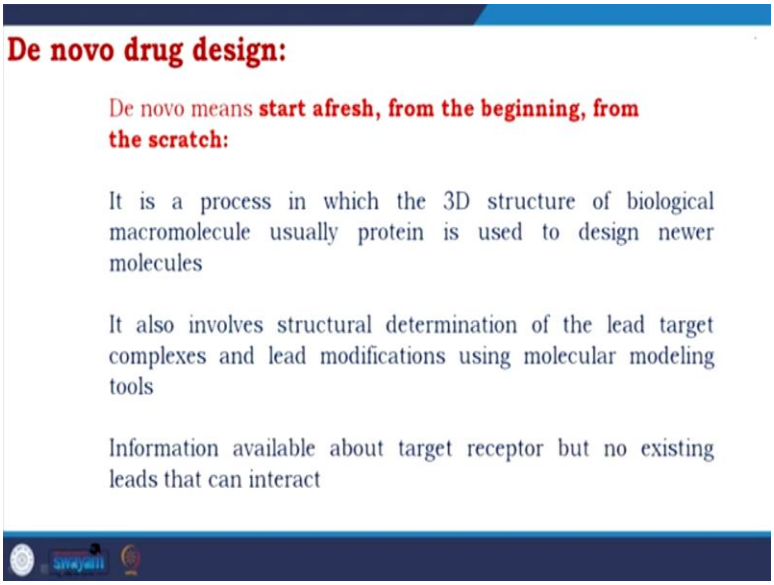
resolution, and now they are providing atomic-level resolution in a regular basis, which I have discussed earlier.

(Refer Slide Time: 49:28)



So, the methods of rational drug discovery, as I talked about the structure-based drug discovery, there is docking-based virtual screening and de novo drug designing.

(Refer Slide Time: 49:43)

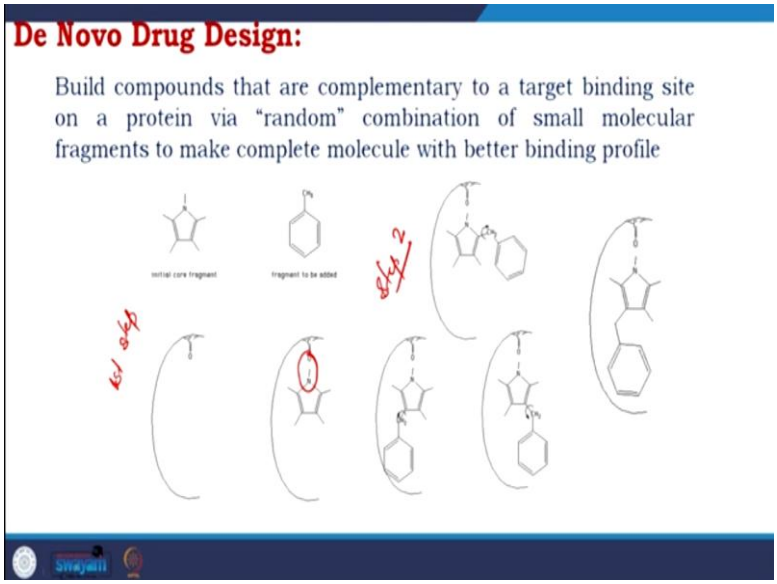


De novo drug designing means starting afresh, from the beginning, from scratch. So, if you start designing a drug from scratch, that unique process is called de novo. Now, if it is from scratch, you will ask me why it comes under structure-based drug discovery? Yes, I am going to give you

the answer. So, it is a process in which the 3D structure of biological macromolecules is usually used to design new molecules.

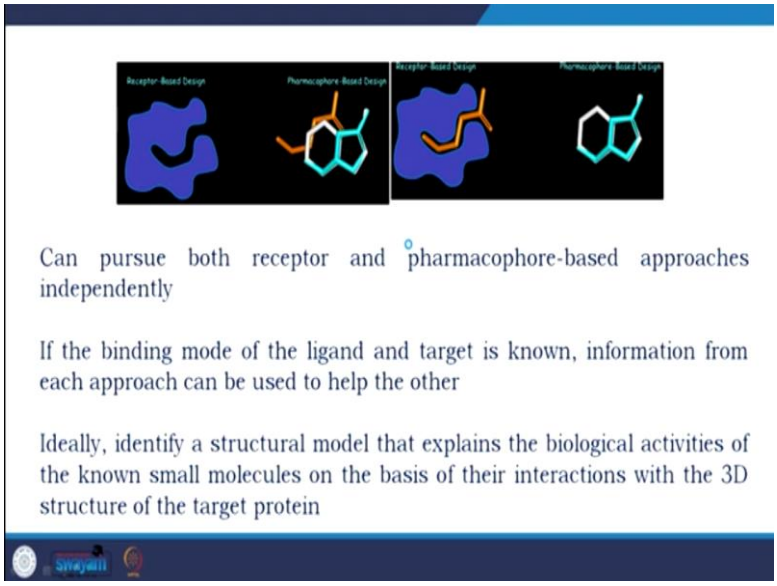
So, de novo drug designing is a scratch from the side of the small molecule, but there is much existence of the 3D structure of the biological target. It also involves the structural determination of the lead target complexes and lead modification using molecular modeling tools information about target receptors, but no existing leads can interact.

(Refer Slide Time: 51:06)



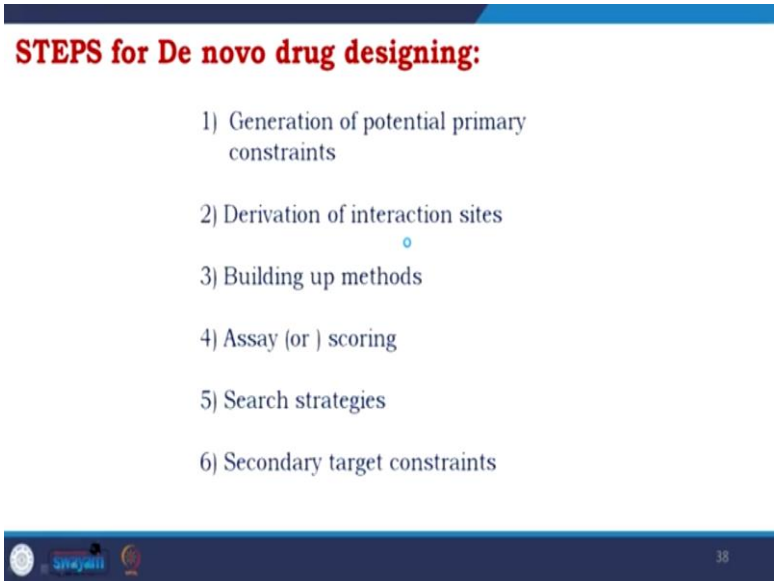
Build compounds complementary to a target binding site on a protein via a random combination of small molecule fragments to make a complete molecule with a better binding profile. So, you see, this is the initial core fragment, and this is the biological target the fragment to be added is this. So, initially, you see that there is an interaction possible this is the first step, then you include this, and you optimize. So, in step 2, you add a fragment and optimize its conformation; this is wrong this is right. And now, you have the lead compound.

(Refer Slide Time: 52:06)



So, you could have a receptor-based design or a pharmacophore-based design. Can pursue both receptor and pharmacophore-based approaches independently if the binding mode of the ligand and target is known. Information from each approach can be used to help the other. Ideally, identify a structural model that explains the biological activities of the known small molecules based on their interaction with the 3D structure of the target protein.

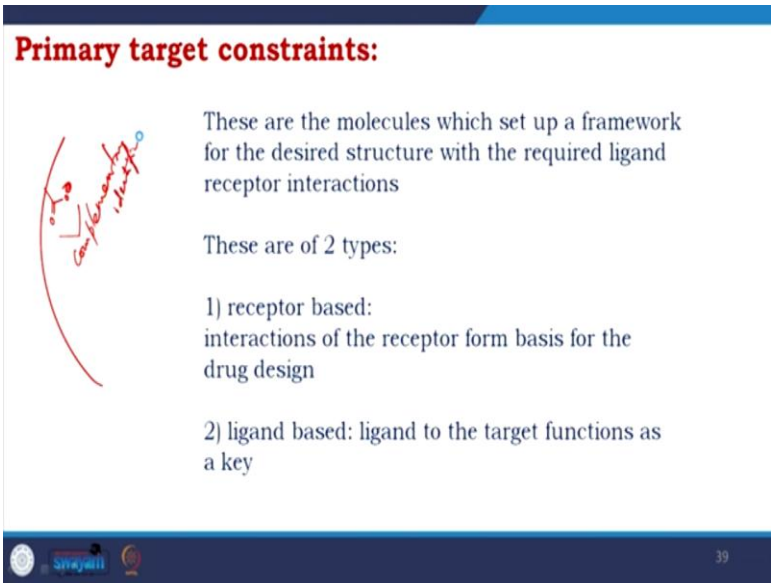
(Refer Slide Time: 52:49)



Steps of de novo drug designing, generation of potential primary constants was the first step, derivation of interaction sites, building up methods, assay or scoring, search strategies, secondary target constants these are the six steps. So, first, you do the generation of potentially primary constant, which I show you bring a fragment and see the interaction with the structure of the

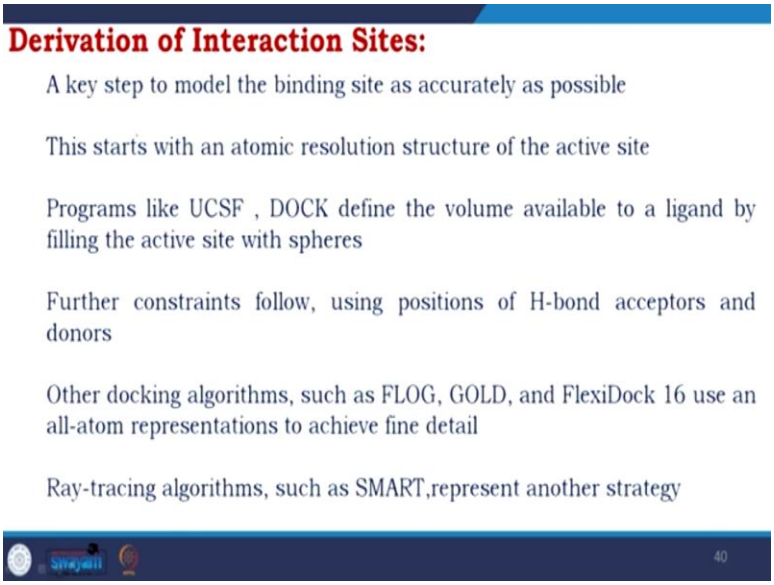
biological macromolecules. Then, you derive that interaction site. You build up by linking others, and then you do the assay, get the molecule used as a lead, search strategies, check for other means, and check for side effects.

(Refer Slide Time: 53:46)



Primary target constants are the molecules that set up a framework for the desired structure with the required ligand-receptor interaction. So, you have it, and you have a group here. Based on these, you do complementary identification. These are of 2 types: one receptor-based interaction from bases for the drug designing and ligand-based ligand to the target function as a key.

(Refer Slide Time: 54:29)



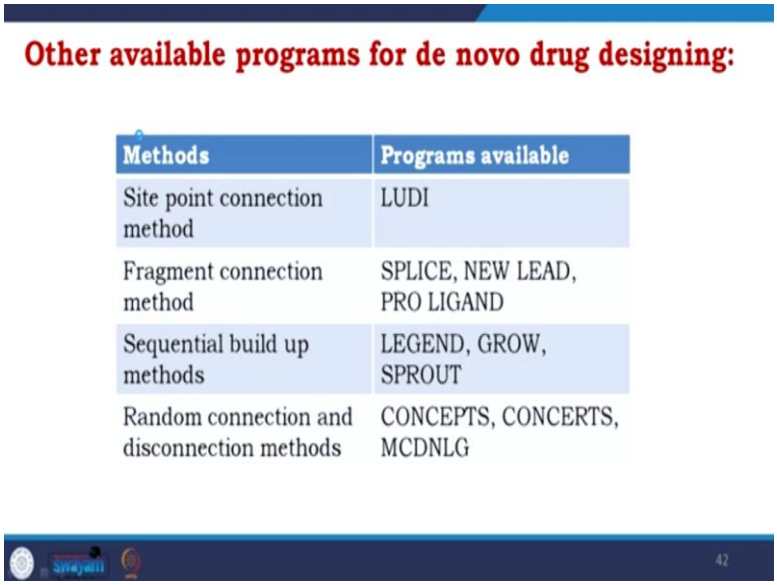
Derivation of interaction site: A key step to model the binding site as accurately as possible. This starts with an atomic resolution structure of the active site. Programs like UCSF DOCK define the volume available to a ligand by filling that active site with spheres. Further constraints follow using the position of hydrogen bond acceptor and donors. Other docking algorithms such as FLOG, GOLD and Flexi dock 16 use an all-atom representation to achieve fine detail. Ray tracing algorithms such as SMART represent another strategy to filling up.

(Refer Slide Time: 55:16)

Building up methods, there are four types. Growing, linking, lattice-based sampling and molecular dynamics-based methods. Growing means it would grow gradually by coming up with different ligands and how one fits. Linking: it comes with fragments; you have a library there are, let us say, 2, 3 binding sites, or there are part of the binding sites.

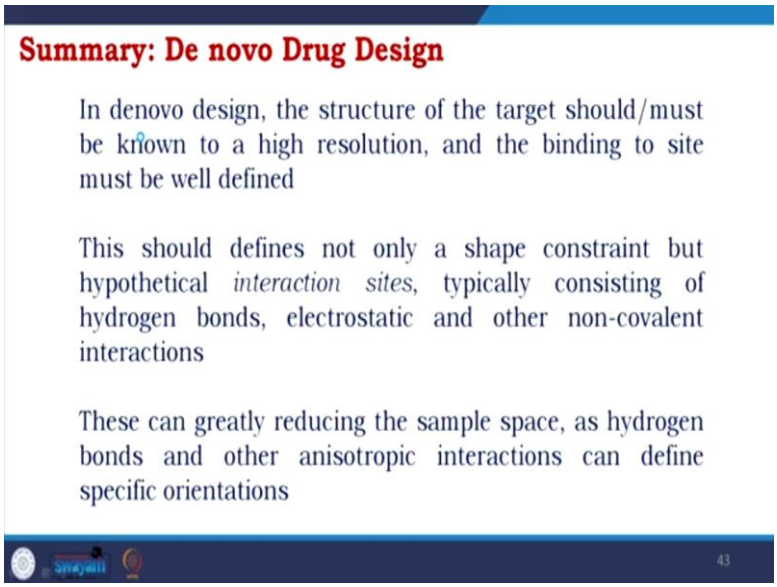
Lattice best sampling is another one. The whole lattice would be created, and after growing, a linking defeat to the lattice and lattice-based sampling would be performed. The molecular dynamics-based method is computationally intensive.

(Refer Slide Time: 56:48)



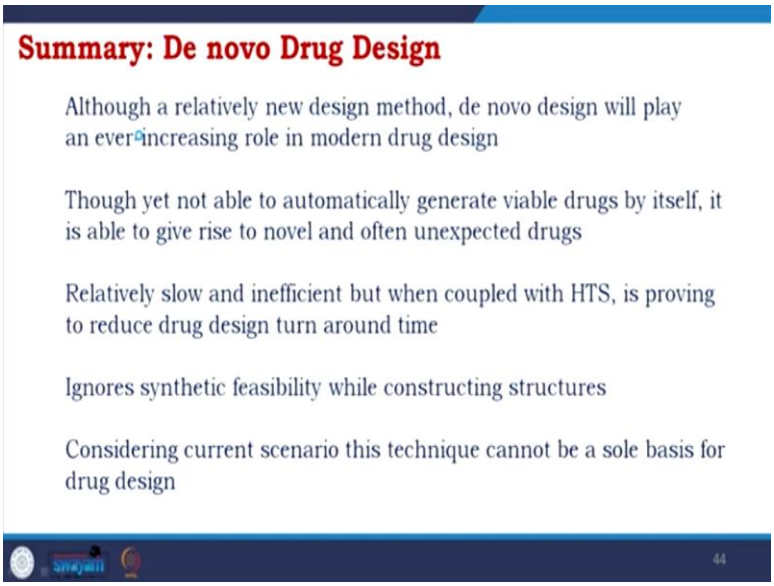
There are other programs available for de novo drug designing. For the sight point connection method, there is LUDI. The fragment connection method is SPLICE, NEW LEAD, PRO LIGAND, sequential buildup method LEGEND, GROW, SPROUT, random connection and disconnection method CONCEPTS, CONCERTS, MCDNLG.

(Refer Slide Time: 57:13)



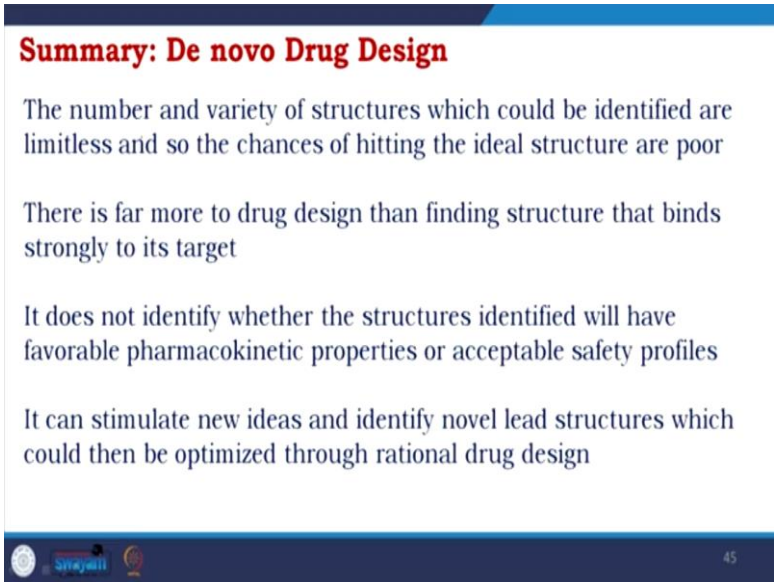
Coming to the summary in de novo design, the structure of the target should or must be known to the high resolution, and the binding to site must be well defined. This should define a shape constraint and hypothetical interaction sites typically consisting of hydrogen bond electrostatic and other non-covalent interactions. These can greatly reduce the sample space as hydrogen bonds and other anisotropic interactions can define specific orientations.

(Refer Slide Time: 57:48)



Although a relatively new design method, de novo design, will play an ever-increasing role in modern drug designing, they cannot automatically generate viable drugs by themselves. It can give rise to novel and often unexpected drugs. So, it is always good to go for this method if you want some magic. Relatively slow and inefficient, but high throughput screening is proving to reduce drug designing turnaround time. Ignores synthetic visibility while constructing structures, which is a very good thing. Considering the current scenario, this technique cannot be the sole basis for drug designing. It is dependent on the other techniques.

(Refer Slide Time: 58:33)



The number and variety of structures that could be identified are limitless. And so, the chances of hitting the ideal structure are normally poor. There is far more to drug designing than finding a structure that binds strongly to its target. It does not identify whether the identified structure will have favorable pharmacokinetics or an acceptable safety profile. So, the process of de novo is not at all connected to identifying a compound that did not have other biological side effects.

It can stimulate new ideas and identify novel lead structures, which could then be optimized through rational drug design.