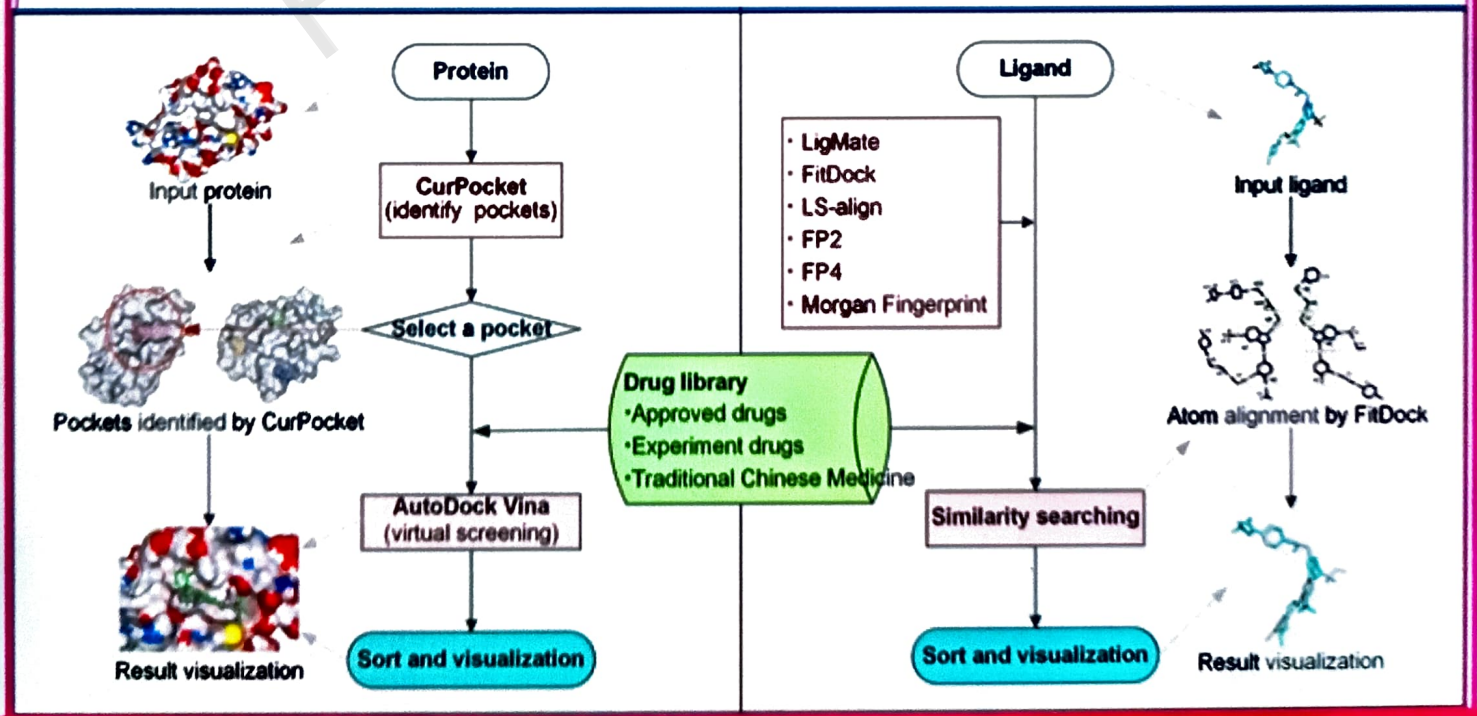
UNIT-III

Molecular modeling and virtual screening techniques

Points to be covered in this topic

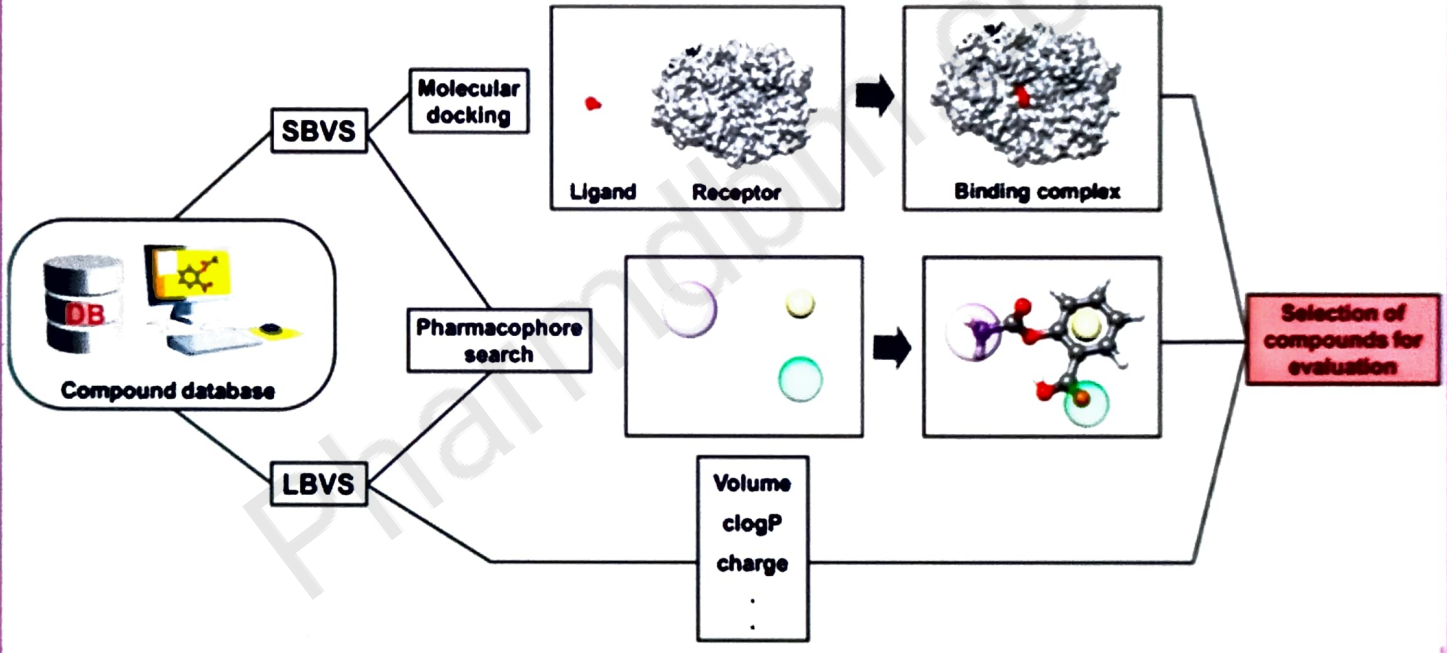
- ❑ INTRODUCTION
- ❑ DRUG LIKENESS SCREENING
- ❑ PHARMACOPHORE MAPPING
- ❑ PHARMACOPHORE SCREENING
- ❑ MOLECULAR DOCKING
- ❑ DOCKING BASED SCREENING
- ❑ *De novo* DRUG DESIGN



❑ INTRODUCTION

❖ Virtual screening techniques

- Virtual screening (VS) is a computational technique **used in drug discovery** to search libraries of small molecules in **order to identify those structures** which are most likely to bind to a drug a drug target typically a protein receptor or enzyme.
- Virtual screening has been **defined as "automatically evaluating very large libraries of compounds"** using computer programs.
- The technique applied depends on **the amount of information** available about the **particular disease target.**
- Virtual screening has **become an integral part of the drug discovery.**



➢ These are two broad categories of screening technique:

a) **Structure based:-** A computational approach used in **the early-stage drug discovery campaign** to search a chemical compound library for novel bioactive molecules against a certain drug target. **i.e. molecular docking and scoring.**

b) **Ligand based:-** The information present in known, active ligands rather than the structure of a target protein for **both lead identification and optimization. i.e. Chemical similarity, pharmacophore and QSAR.**

❑ DRUG LIKENESS SCREENING

- Drug likeness is **defined** as a composite balance of various structure and molecular properties **features** which **determine whether particular molecule is similar to the known drugs.**
- The fastest method **for evaluating the drug-like properties of a compound** is to apply **"rules."**
- **Rules are a set of guidelines for the structural properties of compounds that have a higher probability of being well absorbed after oral dosing.**
- **"Lead-like" or "Drug-like" hits** derived from HTS **(High throughput screening)** campaigns that provide **good starting points for lead Optimization.**

❖ ADMET Properties and Lipinski's rule of 5

MW <500	• **Better absorption and low level of allergic reactions**
Hydrogen bond donors and acceptors <5 and 10	• **Circumvent non-specific binding**
logP value <5	• **Low level of toxicity, non-specific binding and possible oral administration**
logD pH (7.4) > 0	• **An indicator of lipophilicity of a drug;** • **high level of metabolic clearance by P450 enzymes of liver were expected**
Topological polar surface area (TPSA) >60 Å2 and < 140 Å2	• **A high possibility of complete absorption**

➢ The drug likeness can be assessed by following methods:

A. Simple counting method:- Database collections of known drug are typically used to extract knowledge about structure properties of potential drug molecules. **Molecular weight, lipophilicity, charge are profiled to the relevant description of the ADMET related parameter.**

B. Functional group filters:- Reactive, toxics, or unsuitable compounds, such as **natural product derivatives are removed using specific filters.** Typical reactive functional groups include, **for example**, **reactive alkyl halide peroxide and carbazide.**

C. Topological filter:- **It is generally assumed that compound having the structure similarity with known drug may exhibit drug like properties such as oral bioavailability, low toxicity, membrane permeability and metabolic stability.**

D. Pharmacophore filter:- It is based on the assumption that drug like molecules should contain **at least two distinct pharmacophore groups,** functional motifs that guarantee hydrogen bonding capability that are essential for the **specific interaction of the drug molecules with its biological target.**

❑ CONCEPT OF PHARMACOPHORE

- The concept of pharmacophore was first established **by Ehrlich in 1909.**
- **IUPAC defines a pharmacophore to be** "an ensemble of steric and electronic features that is necessary to **ensure the optimal supramolecular interactions with a specific biological target** and to trigger (or block) its biological response".
- Pharmacophore approaches have been used in **virtual screening, de novo design** and other applications **such as lead optimization.**

❖ Ligand-based pharmacophore modeling:-

- It is usually carried out by extracting common chemical features from 3D structures of a set of known ligands representative of essential **interactions between the ligands and a specific large-scale molecular target.**

❖ Structure-based pharmacophore modeling

- Structure-based pharmacophore modeling works directly with the 3D structure of a large scale **molecular target or a macromolecule-ligand complex.**

❑ PHARMACOPHORE MAPPING:-

- Pharmacophore Mapping is the definition and placement of **pharmacophoric features** and the alignment techniques used to overlay 3D.
- Pharmacophore mapping **attempts to find features important for receptor binding.**
- Pharmacophore mapping may be used for ***de novo* compound design.**
- The goal of Pharmacophore mapping is **to establish the bioactive conformations of the ligand and how to superimpose the mapping,** one needs structure-activity relationships of structurally diverse and conformationally informative molecules.

❖ Pharmacophore mapping consists of three steps

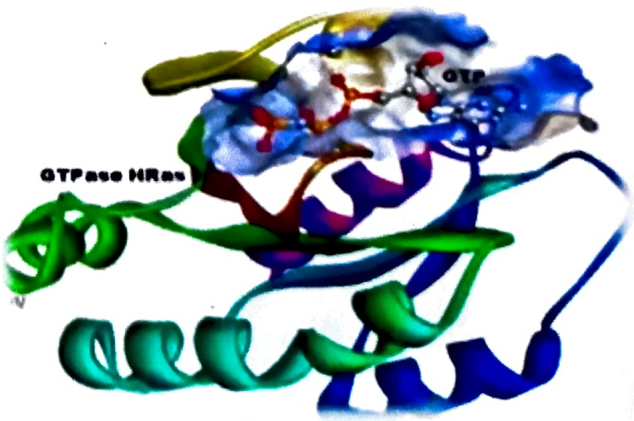
- Identifying common binding element that are **responsible for the biological activity.**
- Generating potential conformations that **active compound may adopt.**
- Determining the 3D relationship **between pharmacophore element in each conformation generated.**

➢ Software use for Pharmacophore mapping:

- **Discovery studio**
- **Hip-hop**
- **Hypogen**
- **Apex Gasp**
- **ROCS**

❖ Application:

- Pharmacophore mapping is used to understand the **biological activity observed in series of compounds.**
- So that we can design **new and more potent compound.**



❑ PHARMACOPHORE BASED SCREENING:-

- It is the process of matching atoms or **functional group and the geometric relations between** them to the pharmacophore in the query.

➢ **Usually pharmacophore based search are done in two steps.**

a) First the software checks whether the **compound has the atom types or functional groups required by the pharmacophore,**

b) Than it checks whether the **spatial arrangement** of this element matches the query.

- Flexible 3D searches identified a higher number of hits than rigid searches do.
- However flexible searches are more **time consuming** than rigid ones.

➢ **There are two main approaches for including conformational flexibility into the search.**

a) One is top generate a user defined **number of representative conformation for each molecules** when the database is to created,

b) The other is to **generate conformation** during the search.

- Pharmacophore filters are **much faster than docking approaches**, and therefore, **design greatly reduce the number of compounds** subjected to the more expensive docking application.
- Once a pharmacophore model is generated by either the ligand-based or the structure- based approach, it can be used for querying the 3D chemical database **to search for potential ligands,** which is **called pharmacophore-based virtual screening (VS).**
- Pharmacophore-based VS and docking-based VS represent the **mainstream of VS tools at the present time.**
- Pharmacophore-based VS **reduces the problems arising from insufficient consideration of protein flexibility** or the use of insufficiently designed and make **the best scoring functions by introducing a tolerance radius for each pharmacophoric feature.**

❖ Applications of pharmacophore-based VS

- In the VS, a pharmacophore model is **screened against large chemical libraries,** and molecules mapping the representation are collected in a virtual hit list.
- These molecules **fulfill the requirements of the model** and therefore have a high likelihood to be active in the experimental testing.

➢ Drug Discovery

- Pharmacophore-based VS is widely applied in different steps of the drug discovery process and facilitates the **initial selection of compound classes** as well as the **optimization of compound properties.**

➢ Lead Identification

- The ultimate aim is the **identification of novel lead compounds** for a specific disease-related target, which can be developed into drug candidates for the treatment of the intended disease.
- Virtual screening is often deployed in these projects **to prioritize molecules for testing and minimizing the number of compounds** to be investigated in biological screens.

➢ Structure-Activity Relationships

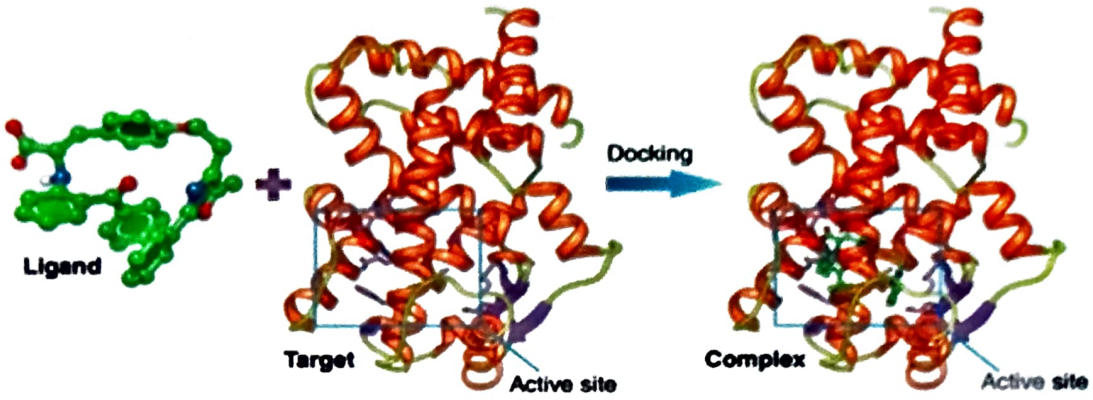
- It describes the **critical functionalities** required for a compound's activity.
- A pharmacophore model **differentiates between active and inactive molecules,** which makes **it highly valuable for establishing structure-activity relationships (SARS).**

❑ MOLECULAR DOCKING

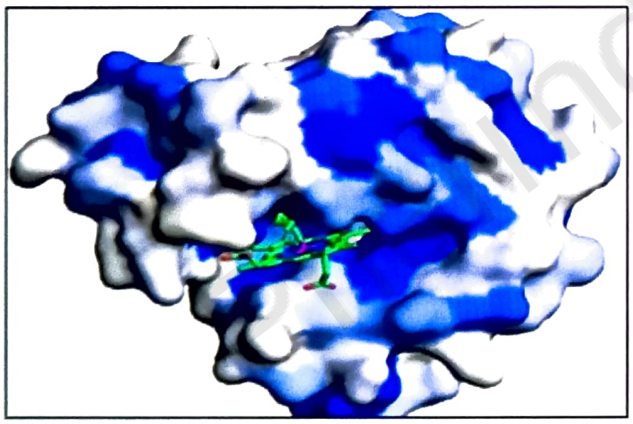
- Molecular docking is the process that **includes placing molecules** in suitable configurations to **interact with receptor.**
- Docking is a method which predicts the **preferred orientation of one molecule to other** when bound to each other to form a stable complex.
- They are able to **generate large number of possible structures.**

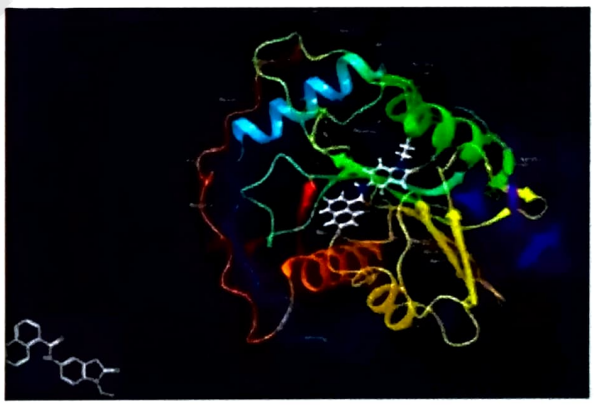
Use force field based strategy **to carry out docking.**

- It is one of the most frequently used methods in structure-based drug design, **due to its ability to predict the binding-conformation of small molecule ligands** to the appropriate target binding site.



- In many drug discovery applications, docking is done **between a small molecule and a macromolecule i.e.**
  1) **Protein-small molecule (ligand) docking**
  2) **Protein nucleic acid docking**
  3) **Protein-protein docking**





❖ POSE Vs. BINDING SITE

➢ Binding site (or "active site")

- The part of the protein where the **ligand binds proteins,**
- Generally a cavity on the protein surface can be **identified by looking at the crystal structure** of the protein bound with a known inhibitor.

➢ Pose (or "binding mode")

- The geometry of the ligand in the binding site.
- **Geometry = location, orientation and conformation**

❖ Docking approaches:-

- There are **two major perspectives**, particularly popular within the molecular docking community.

1. **Shape Complementarity:**

- Describe the protein and ligand as **a set of characteristics** that make them dock.
- The complementarity between the two surfaces with **shape matching description** assist discovering the complementary pose of docking **the target and the ligand molecules.**

2. **Simulation:**

- The docking process is **more complicated.**
- The protein and the ligand are isolated by some physical space and **the ligand finds its position into the protein's active site** after a certain number of "moves" in its conformational space.

➢ Advantages

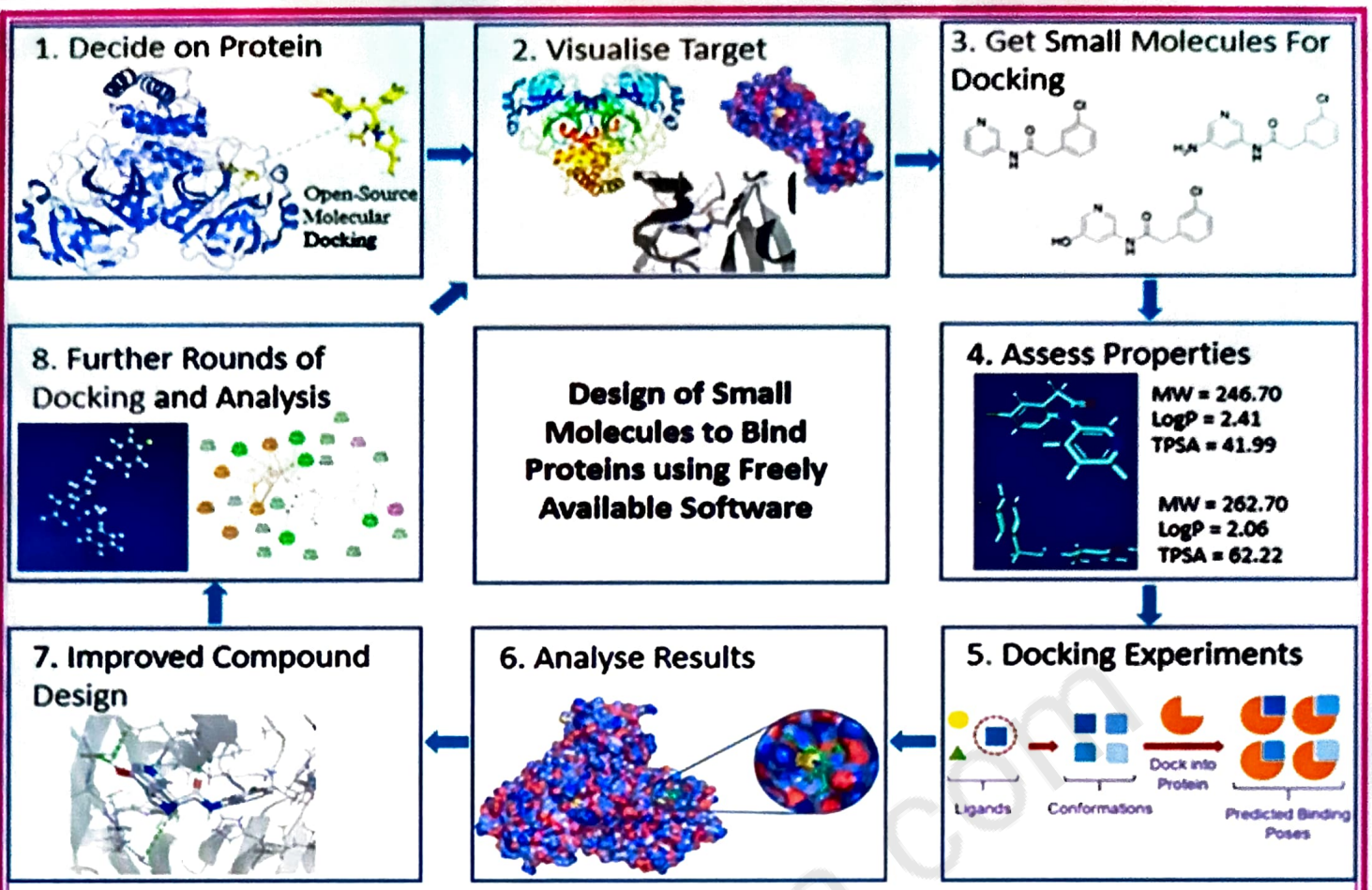
- It is more compatible **to accept ligand flexibility.**
- It is more real to assess the molecular recognition **between ligand and target.**

➢ Disadvantages

- Longer duration to estimate optimal docked conformer **due to the large energy** dissipating for each conformation.
- **Fast optimization method and grid-based tools** have dominantly revolutionized, this drawback to make simulation approach more user friendly.

❖ Steps involved in molecular docking:-

a) Start with crystal co-ordinates with target receptor.
b) Generate molecular surface for receptor.
c) **Generate spheres** to fill the active site of the receptor, the sphere become potential locations for ligand atoms.
d) Sphere centres are matched with the ligand atoms, to determine possible orientation for the ligand.
e) Find the top scoring or the best ranking.



❖ Applications of molecular docking:-

a) Hit identification

- Quickly screen large databases of potential drugs in silico to identify molecules that are likely **to bind to protein targets of interest.**

b) Lead optimization

- Docking can be used to **predict in where and in which relative orientation a ligand attach to a protein** (also referred to as the binding mode or pose).

c) Bioremediation

- Protein ligand docking can also be **used to predict pollutants that can be degraded by enzymes.**

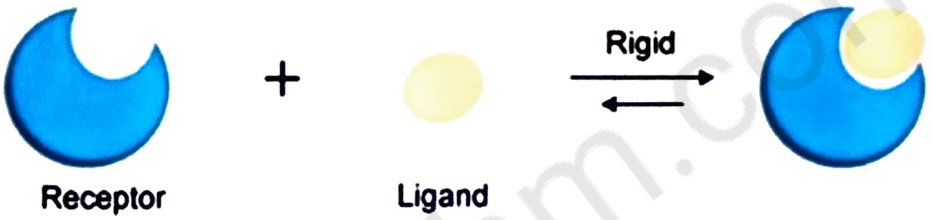
d) Drug-DNA interaction

- Molecular docking plays a prominent role in the **initial prediction of drug's binding properties to nucleic acid.**
- Medicinal chemists are constantly putting their efforts to elucidate the underlying anticancer mechanism of drugs at molecular level by investigating the interaction mode between nucleic acid and drugs.

❖ Types of Docking

I. Rigid Docking:-

- Basically, it is the **first approaches.**
- The ligand and protein are as rigid objects that **cannot change their spatial shape** during the docking process.
- A large number of conformations of each ligand are **generated in advance and each is docked separately.**
- It is Protein-Protein Docking where **Protein and ligand are fixed.**
- First apply steric constraints to limit search space and the **examine energetics of possible binding conformations.**
- Both molecules usually **considered rigid.**



❖ There are 3 major stages of algorithm:

1. Molecular Shape Representation

- Compute the **scattered molecular** surface of the molecule.
- Apply a surface segmentation algorithm that **partitions the surface** according to local shape curvature into concave, convex, and flat surface patches.

2. Surface Patch Matching

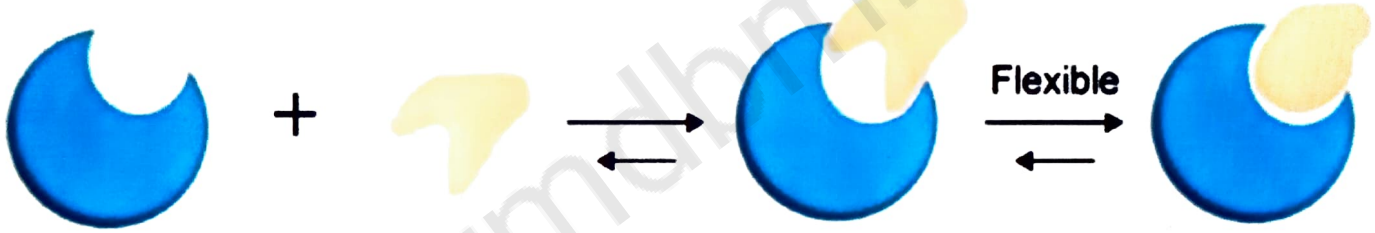
- Apply a hybrid of the geometric hashing and pose-clustering matching techniques **to match "critical" surface points of the patches** detected in the previous step.
- Concave patches are **matched with convex ones and flat patches** with any type of patches.

3. Filtering and Scoring

- Discard complex and remaining molecules are ranked **according to a geometric shape complementarity score**, where surface contact is scored positively and **"acceptable"** steric clashes are penalized.

II. Flexible Docking:-

- Now a days the **most common** form of docking.
- In flexible docking molecules are **flexible, confirmations of the receptor and the Ligand molecules,** as they appear in complex.
- Conformations of each molecule are **generated by the search algorithm** during the docking process.
- The algorithm can avoid considering conformations that **do not fit.**
- It is Protein-ligand docking where **ligand is Flexible and receptor is rigid.**
- Search space is **much larger** in flexible docking.
- An enumeration on the rotations of one of the molecules **(usually smaller one)** is performed. **Every rotation the energy is calculated; later the most optimum pose is selected.**



❖ Methods for handling ligand flexibility

- Many methods have been developed for incorporating **flexible small molecules** into docking software; **they include:**

1. Ligand-ensemble docking method:-

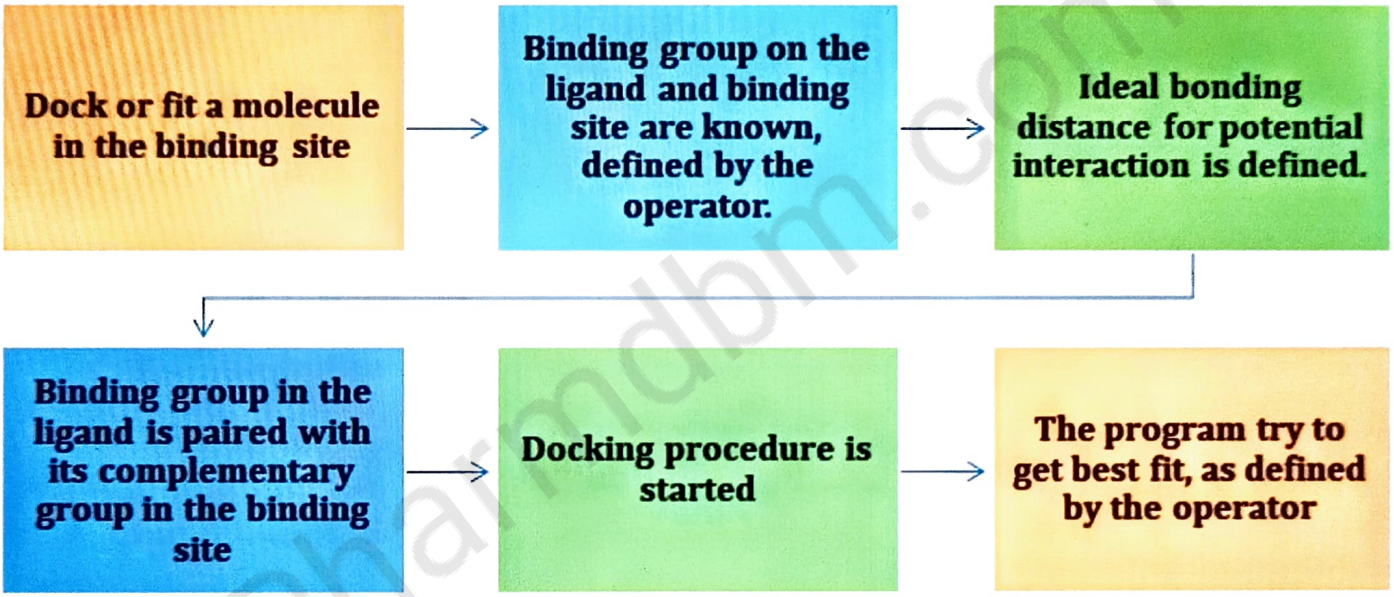
- In the **first step** low energy conformers of ligand are **generated by conformational analysis.**
- In the **second step,** rigid docking is applied for **each conformer independently** in order to find the most favourable small molecule-protein complex.

2. Fragmentation method:-

- Fragmentation methods **break down the molecule into small rigid fragments,** the fragments are then reassembled in the binding pocket.

III. Manual Docking:-

- The ligand is placed in the interacting site and the association energy is calculated at each step.
- The user manually moves, rotates or translates the compound inside the protein cavity and docking assessment are recorded.
- It is still applicable if only small ligand modifications are explored.
- **Advantages:** Quick, Can be very efficient if the user knows well the interacting site.
- **Drawbacks:** Users dependant, It can really produce stupid results this rudimentary method surprisingly provided interesting results in the past.



❑ DOCKING BASED SCREENING

- Virtual screening is the computational or *in silico* analogue biological screening
- The aim is to score, rank or filter a set of chemical structures using one or more computational procedures.
- It can be used
  a) To help decide which compounds to screen **(experimentally)**
  b) Which libraries to synthesize
  c) Which compounds to purchase from an external company
  d) To analyse the results of an experiment, such as a HTS

- The docking based screening was performed in 3 screening protocol, starting with high throughput virtual screening (HTVS) followed by standard precision (SP) and extra precision (XP) methods.
- The high throughput virtual screening (HTVS) mode is designed to screen large libraries quickly with rough scoring functions, hence 8.5 million compounds were screened by this method.
- The top ranked hits (top 20%) were passed through standard precision (SP) mode, which is ten times slower and more precise than HTVS. The SP method is more exhaustive in conformational sampling and more precise than HTVS method with the expense of time.
- About 20,000 compounds obtained from SP method **(best 50% of the compounds)** were further evaluated with even more precise and more computationally intensive extra precision (XP) method.
- About 1000 compounds obtained from XP method were shortlisted based on docking score that are -9.0 and above.
- The high glide score indicated a high binding affinity towards the target.
- Finally checked for the following interactions, hydrogen bonds, salt bridges, halogen bonds, aromatic bonds, pi-cation and also pi-pi interactions all of which contribute towards the stability of the protein-ligand complexes.

❑ *De novo* DRUG DESIGN:-

- **De novo** means start afresh, from the beginning, from the scratch.
- It is a process in which the 3D structure of receptor is used to design newer molecules.
- It involves structure determination of the lead target complexes and the design of lead modifications using molecular modeling tools.
- Ligand optimization can be done by analysing protein active site properties that could be probable area of contact by the ligand.
- The analysed active site properties are described to negative image of protein **such as** hydrogen bond, hydrogen bond acceptor and hydrophobic contact region.

- It involves structural **determination of the lead target complexes and lead modifications** using molecular modeling tools.
- Information available about target receptor but no existing leads that can interact.

➢ Some important points to take into consideration in de novo design are the following:

- Flexible molecules are **better than rigid molecules** because the former are more likely to find an **alternative binding conformation.**
- It is pointless designing for molecules which are **difficult or impossible to synthesize.**
- Similarly, **it is pointless designing for molecules** which need to adopt an unstable conformation in order to bind.
- The **consideration of the energy losses** involved in water desolvation should be taken into account.
- This is significant if the structure of the binding site used for de novo design is based on a protein that is non-human in origin.

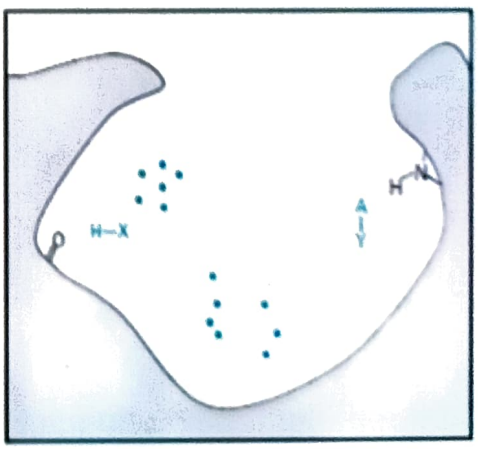
❖ Automated de novo drug design

- Several computer software programs have been written which automatically design novel structures to **fit known binding sites.**
- One of the **best known de novo software programs is called LUDI**, which works by fitting molecular fragments to different regions of the binding site, then linking the fragments together.

➢ There are three stages to the process.

a) Stage 1: Identification of interaction sites:-

- The atoms present in the binding site are analysed to identify those that can take **part in hydrogen bonding interactions**, and those **that can take part in van der Waals interactions.**

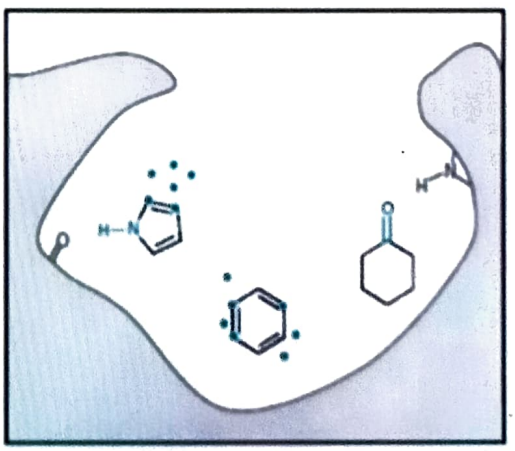


Interaction sites

- Oxygen atoms and tertiary nitrogen atoms are **identified as hydrogen bond acceptors.**
- Any hydrogen attached to oxygen or nitrogen is **identified as a hydrogen bond donor.**
- **Aromatic and aliphatic carbons are identified** as such, and are capable of taking part in van der Waals interactions.
- This can be done by **defining the hydrogen bond interaction site** as a vector involving two atoms.
- The position of these atoms is **determined by the ideal bond lengths and bond angles for a hydrogen bond.**

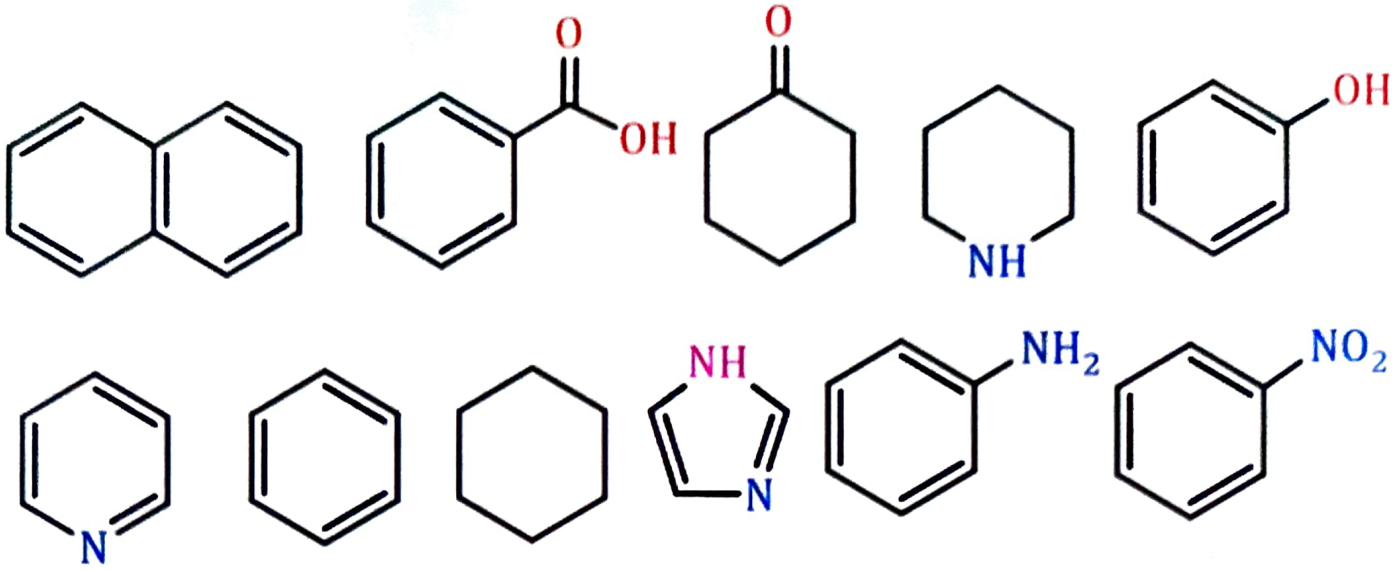
b) Stage 2: fitting molecular fragments

- The LUDI program accesses a library of several hundred molecular fragments.
- The molecules chosen are typically 5-30 atoms in size and are usually rigid in structure because **the fitting procedure assumes rigid fragments.**
- Some fragments are included which can **adopt different conformations.**
- The best fit will be the one that matches up the fragment with the **maximum number of interaction sites.**
- The program can 'try out' the various fragments in its library and **identify those that can be matched up or fitted to the available Interaction sites in the binding site.**



Fragment fitting

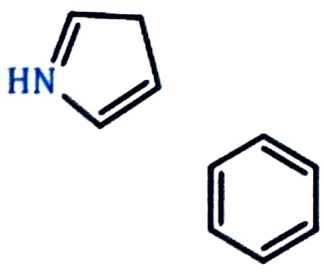
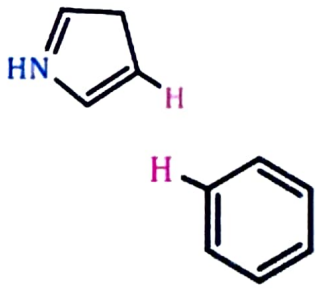
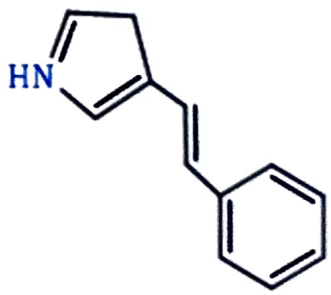
✓ **Examples of molecular fragments used by LUDI**



c) Stage 3: fragment bridging

- Fragments have been **identified and fitted to the binding site,** the final stage is to link them up.
- The program **first identifies the molecular fragments that closest to each other in the binding site, then identifies the closest hydrogen atoms.**
- These now **define the link sites** for the bridge.
- The program now tries out various molecular bridges from a stored library to find out which one fits best.
- A suitable bridge has been found, **a final molecule is created.**

✓ **The bridge process (LUDI)**

			
Identify closes fragments	**Identify closes hydrogen**	**Link points**	**Fit bridge**

✓ Examples of molecular bridges (LUDI)