# Course 07: Virtual Screening

**Introduction:**

Virtual screening, also known as in silico screening, is a method used in pharmaceutical chemistry and structural biology to identify compounds that are potentially active against a biological target.

The design of a new drug is a very long and costly process, from the identification of a biological target of therapeutic interest to the patient, in which clinical trials follow preclinical development. In the past, the classic method for designing a new drug was based on a quasi-experimental method, "high-throughput screening" ("HTS").

As such, virtual screening by molecular docking is a new approach aimed at simulating and predicting the affinity of a very large number of ligands for the active site of a given therapeutic target, which is quite easy to set up, faster, at a much lower cost than experimental screenings.

The fundamental advantage of virtual screening is that it yields a very short list of molecules exhibiting the desired activity. Only these molecules will be tested experimentally. Such optimization should shorten the time between defining a therapeutic need and delivering the corresponding medication to the patient, saving both time and money.

Virtual screening, an in silico analogue of HTS, can schematically be likened to a funnel into which a large number of compounds, constituting the chemical library to be screened, are poured, in order to obtain, using a screening algorithm, a smaller number of compounds which will then be tested experimentally.

**1. Context and Necessity of Virtual Screening:**

- With the explosion in the availability of biological data and technological advances, virtual screening has emerged as an effective way to accelerate the drug discovery process.
- It addresses the pressing need to reduce the costs and time associated with the search for new drugs.

**2. Differences between Virtual Screening and Experimental Screening:**

- While experimental screening involves actual laboratory tests, virtual screening uses computer simulations to predict the interaction between molecules.
- Virtual screening allows the exploration of a large number of compounds even before their synthesis or physical acquisition.

**3. Objectives and Advantages of Virtual Screening:**

- The main objectives include the rapid identification of potential drug candidates and the reduction of failures in the preclinical phase.
- The advantages include speed, cost savings, and the ability to target specific molecular interactions.

**4. History and Evolution of Virtual Screening:**

- The emergence of computer science and molecular modeling techniques paved the way for virtual screening in the 1970s.
- Recent advances in computer technologies and the increasing availability of molecular databases have significantly strengthened this approach.

**5. The role of Virtual Screening:**

The role of screening methods is therefore to eliminate compounds assumed to be inactive or undesirable molecules while prioritizing compounds most likely to be active.

All virtual screening can be broken down into three important steps:

- The development of the initial chemical library , which can be either a screening collection (available molecules) or a virtual collection (molecules to be synthesized)
- The screening itself, which consists of predicting both the conformation and the relative orientation of each molecule in the chemical library , with respect to the target of interest
- The selection of a list of virtual hits. After various treatments to eliminate "false positives", the final selection of molecules to be evaluated involves an individual examination of the 3D interaction of each virtual hit with the receptor.

**6. Virtual Screening Tools**

**6.1 Target**

The 3D structures of various proteins are available and freely accessible in the Protein Data Bank (PDB). This is the largest archive of structural data for biological macromolecules (proteins, DNA, RNA, etc.), with over 160,000 3D structures available. The vast majority of these structures were solved using X-ray crystallography.

### 6.2 Ligand

Currently, there are two ways to obtain the chemical structure of a given ligand: **The first method,** often commercial in nature, consists of databases of chemical structures called chemical libraries or chemical spaces. **The second method** involves using ligands from the literature that can be designed, optimized, and saved in various formats (pdb, mol, mol2, etc.) using molecular construction software such as ChemDraw , Arguslab , Titan, or Sybyl , etc. It should be noted that the Titan and ChemDraw programs were used in this work to construct the ligands of interest.

Regarding chemical libraries , there are two main types: real chemical libraries and virtual chemical libraries .

chemical libraries are often in the form of well plates, each containing a different product. These plates are therefore ready to be tested.

chemical libraries represent a collection of information, more or less organized and hierarchical, containing data on up to 2 million compounds. In these libraries , the coordinates of each compound are stored in files using SMI, SDF, MOL2, and PDB formats. Examples of virtual chemical libraries include: the French National Chemical Library , the Curie Institute Chemical Library , ZINC, PubChem , etc.

### 6.3 Programs

Over the past two decades, a wide variety of more than 60 different docking programs have been proposed (DOCK, AutoDock , FlexX , Surflex , OR, ICM, Glide, CDocker , LigandFit , MCDock , and many others). Although these programs are most often based on specific algorithms, their protocol consists of two essential steps:

**The first step:** called docking proper is the selection step, consisting of placing the ligand in the active site of the protein and sampling the possible conformations, positions and orientations (poses), retaining only the one that represents the most favorable modes of interaction.

**The second step,** scoring , is the ranking stage, which consists of evaluating the affinity between the ligand and the protein, and assigning a score to the poses obtained during the docking phase . This score will allow the best pose to be selected from all those proposed.

In silico methods are grouped into two main families: structure- based virtual screening and ligand- based virtual screening .

- **based " virtual screening :**

based screening methods rely on prior knowledge of ligands with activity against the therapeutic target. These ligands can then be used as a starting point for identifying other similar compounds with activity characteristics common to known ligands of the target. Various types of molecular descriptors can be calculated to quantify the similarity between compounds. Depending on the number of known ligands of the therapeutic target, several methods can be employed: similarity searches, pharmacophore screening , or QSAR approaches.

- **Structure- based virtual screening :**

based virtual screening is considered an in silico equivalent of an experimental assay investigating ligand-target biomolecular binding. However, this screening is essentially dependent on the availability of the 3D structure of the biological target, which is obtained either through experimental methods (X-ray and NMR) in databases (such as PDB: Protein Database) or through 3D structure prediction methods using sequence homology . Several approaches can be used to perform this screening: the construction of 3D pharmacophoric models , the development of 3D-QSAR models, de novo design, and molecular docking methods , which are the most popular.

## 7. The different stages of virtual screening

Virtual screening is widely used to identify new bioactive substances . This approach often consists of several essential steps:

### a- Target preparation

Preparing the 3D structure of a therapeutic target begins by selecting the best available crystallographic structure in protein-derived polymerase (PDB). This selection is based on the resolution value, which defines the quality of the structure. It is important to note that 3D protein structures obtained from PDB require prior preparation before use in molecular docking experiments . Specifically, only one chain should be retained for proteins containing multiple identical subunits to facilitate docking . The reference inhibitor and other elements called heteroatoms (alpha-L- fucose , 1,2-Ethanediol, and N- Acetyl -D-glucosamine) used for crystallization must also be removed, leaving only the A chain structure of the protein. Subsequently, any missing hydrogens must be added, while all water molecules must be removed except those within the enzyme's catalytic cavity. The protonation states of each residue in the enzyme cavity must be checked and the intramolecular interaction energy of the protein must be minimized.

**b- Preparation of the chemical library**

In some cases, it is preferable to clean the chemical library before running virtual screening calculations. This step involves retaining only compounds with a suitable structure and good drug candidate potential. To do this, it is necessary to apply so-called "ADMET" filters to the chemical library in question. The compounds in the cleaned (or uncleaned) library must be prepared by generating the different ionized forms and tautomers because, under physiological conditions, a molecule can take on different forms by undergoing tautomerism or ionization.

**c- Virtual screening proper**

Virtual screening must be performed using one or more software programs. Despite the existence of numerous docking programs , none can be applied universally to all biological systems, as each has its own advantages and limitations. Therefore, using several docking programs followed by the application of a consensus method is recommended to combine and maximize the benefits of the docking software used within a single study. That said, this approach allows for the combination of information from different scoring functions to compensate for their individual imperfections, thereby improving the quality of the results obtained. The underlying assumption is that the probability of a molecule being active should increase if it is associated with good affinity scores according to several scoring functions. Similarly, the probability of a false/positive molecule being correctly classified should decrease because, with their different scoring functions, the software is less likely to make the same error.

**d- Visual Analysis and Selection of Compounds for Experimental Testing**

Although the application of a virtual screening protocol allows for the automatic elimination of a large number of undesirable compounds, human intervention cannot be replaced. A well-known bias, for example, is the high ranking of high molecular weight ligands, which, due to a greater number of atoms, are simply capable of creating a greater number of interactions, even at the entrance or outside of the cavity under study. Therefore, visual inspection of the interaction mode of the compounds that obtained the best scores is a crucial step in order to identify a reduced list of promising molecules for experimental testing. During this inspection, several selection criteria must be taken into consideration: - Positioning of the compounds within the cavity under study

-The mode of interaction (optimal number of H bonds, ionic interactions, hydrophobic, Pi-Pi…etc).

-Interaction with key residues of the target.

- Structural diversity

-Availability of real molecules in the chemical library .