Course 08: Modelling protein structures using molecular docking

1. Molecular docking

docking is an in silico approach that involves predicting and simulating the most favorable position of a ligand within its receptor (often a protein target). Currently, there are more than 40 molecular docking programs (software) . Although these programs are most often based on specific algorithms, their protocol consists of two complementary steps:

The first step, called docking, allows the ligand to adopt several conformations and positions at its receptor in order to retain the most favorable one.

The second step, called scoring , evaluates the affinity between the ligand and the protein and assigns a score to the poses obtained during the docking phase . This score allows the best pose to be selected from all those proposed.

2. Applications

There are at least two main applications of the docking program . The older one is the prediction of the interaction mode, which consists of determining the correct positioning of the ligand relative to its receptor.

The second use lies in searching, through virtual screening, for new molecules active against a given therapeutic target.

Virtual in silico screening is an approach that allows the simulation and prediction of the interaction mode of a very large number of molecules (collected from chemical libraries) with a targeted active site. In other words, virtual screening is a molecular docking of several thousand molecules toward a single receptor.

3. Molecular docking type

- **Rigid docking**

docking methods , finding the optimal pose is limited to positioning. This operation consists of an exhaustive search within the discretized space of the six degrees of freedom. Some programs, while not belonging to the family of rigid docking techniques , use several successive optimization steps, the first of which may resemble rigid docking . For example, the Glide program initially uses, in its multi-step approach , a systematic search to approximate the ligand's position within the protein's active site.

- **Flexible Docking**

docking methods take ligand flexibility into account, two steps are performed sequentially throughout the docking process . The first step involves exploring the conformational space to identify the bioactive conformation among the available options. During the second step, a scoring function evaluates these conformations. Several types of algorithms exist for handling ligand flexibility: systematic methods (fragmentation/reconstruction), random methods, and simulation methods (molecular dynamics).

- **Semi-flexible docking**

conformational space of ligands, the number of degrees of freedom in the search space can be substantial, especially for highly flexible molecules. In such cases, the use of exhaustive search methods often proves inappropriate, as they require significant simplifications in the sampling process. Other algorithms, known as fragmentation algorithms, are employed to incrementally construct the ligand within the protein's active site. The ligand's conformational space is then restricted to the neighborhood of an initial set of simplified states. This construction-based search strategy, which exists in various forms, is notably adopted by the DOCK, FLExX , and Hammerhead programs .

4. Molecular docking tools

➢ The receiver

The first major approach to studying and designing bioactive molecules through molecular modeling is based on receptor structure. This approach relies on exploiting the three-dimensional molecular structure of the target protein. Three experimental methods currently

allow for the determination of protein structure: nuclear magnetic resonance (NMR), electron microscopy, and X-ray crystallography. The latter technique is responsible for the majority of structures in a database called the Protein Data Bank (PDB).

The PDB is a global repository of information on the three-dimensional structure of proteins and nucleic acids. These molecules originate from all biological kingdoms. The PDB is freely accessible via the internet (http://www.rcsb.org/pdb/). It contains several thousand protein structures obtained either by X-ray crystallography or by NMR. If the target protein is not yet deposited in the database, and the database contains a protein with similar sequences, homology modeling is used to construct the 3D structure of the desired target.

- **The ligand**

In molecular docking , ligand selection is a crucial step. This choice must be appropriate due to the specificity of the target's active site, avoiding unnecessary testing of molecules.

For molecular docking , the ligand must also be in 3D form. Currently, there are two ways to obtain the chemical structure of a given ligand: The first, often commercial in nature, consists of databases of chemical structures called chemical libraries or chemical spaces. The second method involves using ligands from the PDB or literature that can be drawn, optimized, and saved in various formats (pdb, mol, mol2, etc.) using molecular construction software such as ChemDraw , Arguslab , Titan, or Sybyl , etc.

- **molecular docking program**

docking is accomplished in two complementary steps. The first consists of searching for ligand conformations capable of establishing ideal interactions with the receptor. The second is a scoring function that allows these conformations to be evaluated by a rapid calculation of their interaction energy with the receptor.

5. Filtering by ADME-TOX

The abbreviation ADMET represents the following pharmacokinetic concepts: absorption, distribution, metabolism, and excretion. These criteria describe the disposition of a bioactive molecule within an organism. They express the concentrations of the product in the body's various tissues and circulatory system. The results obtained using these criteria allow for the analysis of a product's performance and efficacy, determining its potential to become a drug and its suitability for further development in clinical trials.

Tox (Absorption, Distribution, Metabolism, Elimination, and Toxicity) filters have rapidly gained popularity. They are based on several criteria that determine the potential pharmacokinetic properties of molecules and are now widely used to reduce the number of compounds in a chemical library by selecting those most suitable as drug candidates, prior to any screening process. The use of these filters has yielded very good results.

Absorption: This involves analyzing a molecule's ability to penetrate the body after administration. Low solubility (for example, due to excessive hydrophobicity) or high polarity has a drastic impact on the intestinal absorption of a compound.

Distribution: This criterion measures a molecule's ability to diffuse, for example via the bloodstream, throughout the body. A molecule must be able to move from one compartment to another in order to ultimately reach its target. Strong binding to plasma proteins negatively impacts a molecule's distribution.

Metabolism: The metabolism filter aims to detect (a) the stability of the molecule in the body, which impacts its duration of action, and (b) the metabolites of the initial molecule, namely the compounds resulting from its degradation or enzymatic modifications occurring within the body. In humans, liver cytochrome P450 enzymes are the main enzymes that modify xenobiotics .

These compounds are made more hydrophilic by the introduction of oxygen atoms. The metabolites can be inactive, more active than the original compound, and of course potentially toxic, hence the need to characterize and study them.

Excretion (elimination): In order to avoid accumulation phenomena, often synonymous with toxicity, care must be taken to ensure that the administered compounds, as well as their metabolites, are properly excreted from the body, for example via urine or feces.

Toxicity : As its name suggests, this filter is used to measure the toxicity of a compound and its metabolites. Toxicity and lack of efficacy of drug candidates are now the two biggest causes of failure in drug development. Different types of toxicity are evaluated, including carcinogenicity.

A major contributor to the rapid and large-scale identification of drug-like molecules is commonly known as the " Lipinski rules " or "rule of 5," which allows for the estimation of a compound's oral bioavailability based on its two-dimensional (2D) structure. These rules

concerning physicochemical properties were defined after the analysis of 2,245 marketed drugs or drugs in the final stages of development.

 The molecular weight of the compound must not exceed 500 daltons (Da).

 The decimal logarithm of the water / 1-octanol partition coefficient, denoted logP , must be less than 5.

 The number of hydrogen bond donors must be less than 5.

 The number of hydrogen bond acceptors must be less than 10.

Other criteria have been put in place to complement and adjust Lipinski 's rules in the selection of " drug-like " compounds.

Thus, Veber chooses to use the following criteria: the polar surface area (PSA) of the compound must be less than 140 Å2 and the number of rotatable bonds must be less than 10. These criteria are often used in addition to the "rule of 5". These criteria were established by the study of the oral bioavailability of drug candidates.