**Course: Advanced Bioinformatics**

**Module title: VHT in Silico-Screening**

**Module no. : 165**

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 target, typically a protein receptor or enzyme.

Virtual screening has been defined as the "automatically evaluating very large libraries of compounds" using computer programs. As this definition suggests, VS has largely been a numbers game focusing on how the enormous chemical space of over 1060 conceivable compounds can be filtered to a manageable number that can be synthesized, purchased, and tested. Although searching the entire chemical universe may be a theoretically interesting problem, more practical VS scenarios focus on designing and optimizing targeted combinatorial libraries and enriching libraries of available compounds from in-house compound repositories or vendor offerings. As the accuracy of the method has increased, virtual screening has become an integral part of the drug discovery process.

**Methods**

There are two broad categories of screening techniques: ligand-based and structure-based.

Ligand-based: Given a set of structurally diverse ligands that binds to a receptor, a model of the receptor can be built by exploiting the collective information contained in such set of ligands. These are known as pharmacophore models. A candidate ligand can then be compared to the pharmacophore model to determine whether it is compatible with it and therefore likely to bind.

A different strategy is to develop logic-based rules describing features of substructures and chemical properties related to activity using support vector inductive logic programming. The logic-based features provide insights into activity which can be understood by medicinal chemists. Support vector machine integrate the features to yield a quantitative QSAR, which is then used to screen a database of molecules. This approach is well suited to scaffold hopping to identify novel active molecules and is implemented in the package INDDEx.

Another approach to ligand-based virtual screening is to use 2D chemical similarity analysis methods to scan a database of molecules against one or more active ligand structure.

A popular approach to ligand-based virtual screening is based on searching molecules with shape similar to that of known actives, as such molecules will fit the target's binding site and hence will be likely to bind the target. There are a number of prospective applications of this class of techniques in the literature.

Ligand-based virtual screening methods have been extensively compared on the large ChEMBL database in several machine learning challenges, where Deep Learning emerged as the best performing technique.

Structure-based

Structure-based virtual screening involves docking of candidate ligands into a protein target followed by applying a scoring function to estimate the likelihood that the ligand will bind to the protein with high affinity.

**Virtual high throughput screening**

High throughput screening (HTS) is typically used at an early stage of the drug design process in order to test a large compound collection for potential activity against the chosen target. Unfortunately, HTS is time consuming and costly. For this reason, its computational corollary, the vHTS, has become an important tool to precede the large in vitro screening assays performed in pharmaceutical companies. vHTS aims at using computational tools to estimate a priori, from an entire database of existing compounds (or compounds that could be made), those that are the most likely to have some affinity for the target. There are basically two approaches to this topic: ligand- and structure-based vHTS.

**In silico fragment-based drug design:**

Since a few years, FBD has become an attractive alternative to experimental or virtual HTS. Contrarily to HTS, where complete molecules are screened for activity, FBD aims at building new ligands piece-by-piece by connecting small and well-chosen compounds that bind into separate binding pockets, close enough to be chemically linked in their relative favorable positions. When tested experimentally, hit molecular fragments exhibit generally only weak affinities, with IC50 in the order of 1 mM to 30 μM. However, they provide interesting starting points for follow-up strategies trying to connect several of them to give new efficient lead compounds. Fragment-based design can be performed in silico or experimentally using nuclear magnetic resonance (NMR) or X-ray crystallography. This review will focus on in silico approaches.