**Course: Advanced Bioinformatics**

**Module title: Docking and Scoring**

**Module no. : 166**

**Basic Requirements for Drug Discovery:**

* A target protein structure with or without a ligand.
* A database for virtual compounds
* A computational framework to work out desired docking

In the fields of computational chemistry and molecular modelling, scoring functions are fast approximate mathematical methods used to predict the strength of the non-covalent interaction (also referred to as binding affinity) between two molecules after they have been docked. Most commonly one of the molecules is a small organic compound such as a drug and the second is the drug's biological target such as a protein receptor. Scoring functions have also been developed to predict the strength of other types of intermolecular interactions, for example between two proteins or between protein and DNA.

Utility:

Scoring functions are widely used in drug discovery and other molecular modelling applications. These include:

* Virtual screening of small molecule databases of candidate ligands to identify novel small molecules that bind to a protein target of interest and therefore are useful starting points for drug discovery
* De novo design (design "from scratch") of novel small molecules that bind to a protein target
* Lead optimization of screening hits to optimize their affinity and selectivity

A potentially more reliable but much more computationally demanding alternative to scoring functions are free energy perturbation calculations.

Prerequisites

Scoring functions are normally parameterized (or trained) against a data set consisting of experimentally determined binding affinities between molecular species similar to the species that one wishes to predict.

For currently used methods aiming to predict affinities of ligands for proteins the following must first be known or predicted:

* Protein tertiary structure – arrangement of the protein atoms in three-dimensional space. Protein structures may be determined by experimental techniques such as X-ray crystallography or solution phase NMR methods or predicted by homology modelling.
* Ligand active conformation – three-dimensional shape of the ligand when bound to the protein
* Binding-mode – orientation of the two binding partners relative to each other in the complex

The above information yields the three-dimensional structure of the complex. Based on this structure, the scoring function can then estimate the strength of the association between the two molecules in the complex using one of the methods outlined below. Finally the scoring function itself may be used to help predict both the binding mode and the active conformation of the small molecule in the complex, or alternatively a simpler and computationally faster function may be utilised within the docking run.

Classes

There are three general classes of scoring functions:

* Force field – affinities are estimated by summing the strength of intermolecular van der Waals and electrostatic interactions between all atoms of the two molecules in the complex. The intramolecular energies (also referred to as strain energy) of the two binding partners are also frequently included. Finally since the binding normally takes place in the presence of water, the desolvation energies of the ligand and of the protein are sometimes taken into account using implicit solvation methods such as GBSA or PBSA.
* Empirical – based on counting the number of various types of interactions between the two binding partners. Counting may be based on the number of ligand and receptor atoms in contact with each other or by calculating the change in solvent accessible surface area (ΔSASA) in the complex compared to the uncomplexed ligand and protein. The coefficients of the scoring function are usually fit using multiple linear regression methods. These interactions terms of the function may include for example:

hydrophobic — hydrophobic contacts (favorable),

hydrophobic — hydrophilic contacts (unfavorable),

number of hydrogen bonds (favorable contribution to affinity, especially if shielded from solvent, if solvent exposed no contribution), number of rotatable bonds immobilized in complex formation (unfavorable conformational entropy contribution).

* Knowledge-based (also known as statistical potentials) – based on statistical observations of intermolecular close contacts in large 3D databases (such as the Cambridge Structural Database or Protein Data Bank) which are used to derive "potentials of mean force". This method is founded on the assumption that close intermolecular interactions between certain types of atoms or functional groups that occur more frequently than one would expect by a random distribution are likely to be energetically favorable and therefore contribute favorably to binding affinity.

Finally hybrid scoring functions have also been developed in which the components from two or more of the above scoring functions are combined into one function.