**Course: Advanced Bioinformatics**

**Module title: QSAR Methods**

**Module no. : 169**

Quantitative structure–activity relationship models (QSAR models) are regression or classification models used in the chemical and biological sciences and engineering. Like other regression models, QSAR regression models relate a set of "predictor" variables (X) to the potency of the response variable (Y), while classification QSAR models relate the predictor variables to a categorical value of the response variable.

In QSAR modeling, the predictors consist of physico-chemical properties or theoretical molecular descriptors of chemicals; the QSAR response-variable could be a biological activity of the chemicals. QSAR models first summarize a supposed relationship between chemical structures and biological activity in a data-set of chemicals. Second, QSAR models predict the activities of new chemicals.

Related terms include quantitative structure–property relationships (QSPR) when a chemical property is modeled as the response variable.

As an example, biological activity can be expressed quantitatively as the concentration of a substance required to give a certain biological response. Additionally, when physicochemical properties or structures are expressed by numbers, one can find a mathematical relationship, or quantitative structure-activity relationship, between the two. The mathematical expression, if carefully validated can then be used to predict the modeled response of other chemical structures.

A QSAR has the form of a mathematical model:

Activity = f(physiochemical properties and/or structural properties) + error

The error includes model error (bias) and observational variability, that is, the variability in observations even on a correct model.

Types

* Fragment based (group contribution)

Analogously, the "partition coefficient"—a measurement of differential solubility and itself a component of QSAR predictions—can be predicted either by atomic methods (known as "XLogP" or "ALogP") or by chemical fragment methods (known as "CLogP" and other variations). It has been shown that the logP of compound can be determined by the sum of its fragments; fragment-based methods are generally accepted as better predictors than atomic-based methods. Fragmentary values have been determined statistically, based on empirical data for known logP values. This method gives mixed results and is generally not trusted to have accuracy of more than ±0.1 units.

Group or Fragment based QSAR is also known as GQSAR. GQSAR allows flexibility to study various molecular fragments of interest in relation to the variation in biological response. The molecular fragments could be substituents at various substitution sites in congeneric set of molecules or could be on the basis of pre-defined chemical rules in case of non-congeneric sets. GQSAR also considers cross-terms fragment descriptors, which could be helpful in identification of key fragment interactions in determining variation of activity.[10] Lead discovery using Fragnomics is an emerging paradigm. In this context FB-QSAR proves to be a promising strategy for fragment library design and in fragment-to-lead identification endeavors.

An advanced approach on fragment or group-based QSAR based on the concept of pharmacophore-similarity is developed.[12] This method, pharmacophore-similarity-based QSAR (PS-QSAR) uses topological pharmacophoric descriptors to develop QSAR models. This activity prediction may assist the contribution of certain pharmacophore features encoded by respective fragments toward activity improvement and/or detrimental effects.

* 3D-QSAR

3D-QSAR refers to the application of force field calculations requiring three-dimensional structures, e.g. based on protein crystallography or molecule superimposition. It uses computed potentials, e.g. the Lennard-Jones potential, rather than experimental constants and is concerned with the overall molecule rather than a single substituent. It examines the steric fields (shape of the molecule), the hydrophobic regions (water-soluble surfaces), and the electrostatic fields.

The created data space is then usually reduced by a following feature extraction (see also dimensionality reduction). The following learning method can be any of the already mentioned machine learning methods, e.g. support vector machines. An alternative approach uses multiple-instance learning by encoding molecules as sets of data instances, each of which represents a possible molecular conformation. A label or response is assigned to each set corresponding to the activity of the molecule, which is assumed to be determined by at least one instance in the set (i.e. some conformation of the molecule).

3-D QSAutogrid-R MPGRS example image

On June 18, 2011 the Comparative Molecular Field Analysis (CoMFA) patent has dropped any restriction on the use of GRID and partial least-squares (PLS) technologies and the Rome Center for Molecular Design (RCMD) team (www.rcmd.it) has opened a 3D QSAR web server (www.3d-qsar.com) based on the 3-D QSAutogrid/R engine.

GOLPE stands for Generating Optimal Linear PLS Estimations. 3-D QSAutogrid/R covers all the main features of CoMFA and GRID/GOLPE with implementation by multiprobe/multiregion variable selection (MPGRS) that improves the simplification of interpretation of the 3-D QSAR map. The methodology is based on the integration of the molecular interaction fields as calculated by AutoGrid and the R statistical environment that can be easily coupled with many free graphical molecular interfaces such as UCSF-Chimera, AutoDock Tools, JMol and others.

* Chemical descriptor based

In this approach, descriptors quantifying various electronic, geometric, or steric properties of a molecule are computed and used to develop a QSAR. This approach is different from the fragment (or group contribution) approach in that the descriptors are computed for the system as whole rather than from the properties of individual fragments. This approach is different from the 3D-QSAR approach in that the descriptors are computed from scalar quantities (e.g., energies, geometric parameters) rather than from 3D fields.

An example of this approach is the QSARs developed for olefin polymerization by half sandwich compounds.