QSAR Modeling approach in Virtual Screening of Active Compounds

Feroz Khan

Computational Biology Unit, CSIR-Central Institute of Medicinal & Aromatic Plants, P.O.-CIMAP, Lucknow-226015 (U.P.), INDIA

E-mail: f.khan@cimap.res.in

Abstract:

Quantitative structure activity relationships (QSAR) have emerged as a rational alternative in order to find new active molecules including anticancer compounds. Many topological molecular descriptors can be used to describe organic molecular structure with QSAR aims. Developing a model for predicting anticancer activity of any classes of organic compounds based on molecular structure is very important goal for medicinal chemist. Different molecular descriptors can be used to solve this problem. Molecular descriptors and biological activity based multiple linear regression analysis and 3D QSAR approaches are shown to be very successful in drug design. Besides, the structural similarity of compounds is so much that we may need linear machine learning models such as PLS, MLR etc. instead of non-linear machine learning methods e.g., artificial neural networks (ANN), support vector machine (SVM), etc. QSAR models are nowadays regarded as a scientifically credible tool for predicting and classifying biological activities of untested compounds. QSAR has become inevitably embedded as an essential tool in the pharmaceutical industry, from lead discovery, optimization to lead development and computer-aided drug designing & discovery. A growing trend is to use QSAR early in the drug discovery process as a screening and enrichment tool, so that to estimate those compounds which are lacking drug like properties or those chemicals predicted to elicit a toxic response. The basic assumption for all molecule-based hypotheses is that similar molecules have similar activities. This principle is also called Structure–Activity Relationship (SAR). The underlying problem is therefore how to define a small difference on a molecular level, since each kind of activity, e.g., reaction ability, biotransformation ability, solubility, target activity, and so on, might depend on another difference. QSAR models have been also used for risk management. QSARS are suggested by regulatory authorities e.g., US FDA and European Union by the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) regulation.

Keywords: QSAR, Machine Learning, ANN, SVM, PLS, MLR