

A Review of 3D-QSAR in Drug Design

Thirumurthy Madhavan[†]

Abstract

Quantitative structure-activity relationship (QSAR) methodologies have been applied for many years, to correlate the relationship between physicochemical properties of chemical substances and their biological activities to generate a statistical model for prediction of the activities of new chemical entities. The basic principle behind the QSAR models is that, how structural variation is responsible for the difference in biological activities of the compounds. 3D-QSAR has emerged as a natural extension to the classical Hansch and Free-Wilson approaches, which develops the 3D properties of the ligands to predict their biological activities using various chemometric techniques (PLS, G/PLS, ANN etc). It has served as a valuable predictive tool in the design of pharmaceuticals and agrochemicals. This review seeks to provide different 3D-QSAR approaches involved in drug designing process to develop structure-activity relationships and also discussed the fundamental limitations, as well as those that might be overcome with the improved methodologies.

Key words : 3D-QSAR, CoMFA, FAK

1. Introduction

Quantitative structure-activity relationship (QSAR) methodologies have been used to develop computational or mathematical models, in order to find statistical correlation between structure and function using chemometric techniques. Various QSAR approaches have been developed over a period of years and served as a valuable predictive tool, particularly in the process of drug designing techniques. The methods have evolved from Hansch and Free-Wilson's one or two dimensional linear free-energy relationships (classical methods), via Crammer's three -dimensional QSAR to Hopfinger's fourth and Vedani's fifth and sixth-dimensions^[1-5]. Classical QSAR methods are much simpler, faster and more amenable to automation than 3D-QSAR approaches, however they suffer series limitations in certain situation. A key difference between classical and 3D-QSAR is the form of the output, source of numerical descriptors, and unable to directly suggest new compounds to synthesize. Due to the limited scope for these classical methods, in this paper we focus only on

3D-QSAR approaches in drug design.

2. Experimental Section

2.1. Objectives of QSAR

Most of the QSAR methods focus on the following goals,

To optimize the existing lead compounds, in order to improve their biological activities.

QSAR approaches correlate the relationships between chemical structure alterations and respective changes in biological endpoint to understand which chemical properties are most likely determinants for their biological activities.

To predict the biological activities of untested and sometimes yet unavailable compounds.

2.2. Classification of QSAR Methodologies

The QSAR methods are classified based on the structural representation or the way by which the descriptor values are derived:

1D-QSAR correlating activity with global molecular properties like pKa, log P etc.

2D-QSAR correlating activity with structural patterns like connectivity indices, 2D-pharmacophores etc., without taking into account the 3D-representation of these properties

Centre for Bioinformatics, Department of Biochemistry, School of life sciences, University of Madras, Guindy campus, Chennai-600025, India.

[†]Corresponding author : thiru.murthyunom@gmail.com
(Received : February 5, 2012, Revised : March 25, 2012,
Accepted : March 27, 2012)

3D-QSAR correlating activity with non-covalent interaction fields surrounding the molecules

4D-QSAR additionally including ensemble of ligand configurations in 3D-QSAR

5D-QSAR explicitly representing different induced-fit models in 4D-QSAR

6D-QSAR further incorporating different salvation models in 5D-QSAR

2.3. Assumption in 3D-QSAR Methods

3D-QSAR correlate macroscopic target properties with compound atom-based descriptors derived from the three dimensional representation of the molecular structures^[6-8]. All 3D-QSAR methods are based on various assumptions in order to develop a model as close to the real one which is described in given below.

Receptor binding is directly proportional to the biological activity. Differential effects on second messengers or other signaling steps which transpire between receptor binding and experimentally observed response, are not taken into consideration.

There is an underlying relationship between molecular structure and biological activity.

Molecular structure can be measured and represented with a set of numbers usually called descriptors, which encode all physical, chemical and biological properties of the molecule.

The lowest energy conformation of the ligand is its bioactive conformation, and it is this single conformation of the ligand which exerts the binding effects.

The loss of translational and rotational degrees of freedom (entropy) upon binding is believed to follow a similar pattern for all the molecules.

For all the modeled ligands, the protein binding site is assumed to be same. Total number of rotatable bonds is the only method most frequently used to estimate the entropic cost for freezing non-terminal single-bond rotors.

2.4. Classification of 3D-QSAR Approaches

CoMFA

In 1987, Cramer developed a powerful 3D-QSAR methodology, Comparative Molecular Field Analysis (CoMFA)^[9], and it has become a prototype of 3D-QSAR methods. Standard CoMFA procedures have the following sequential steps

- Determine the bioactive conformations of each molecule.

- Automated or manual methods are used to superimpose or align all the molecules, the alignment of molecules indirectly represent the bioactive mode of receptor structure.

- The aligned molecules are placed in the grid box with spacing of 2 Å.

- The algorithm compares, in three-dimensions, the steric and electrostatic fields calculated around the molecules with different probe groups positioned at all intersections of the lattice.

- The interaction energy or field values are correlated with the biological activity data using PLS technique, which identifies and extracts the quantitative influence of specific chemical features of molecules on their biological activity.

- For visual understanding, the PLS output is represented in the form of an interactive graphics consisting of colored contour plots of coefficients of the corresponding field.

Compared to classical methods CoMFA have many advantages, but it has several pitfalls and limitations as given below^[10],

- Hydrophobicity not well-quantified
- Cut-off limits used
- Imperfections in potential energy functions
- Applicable only to in vitro data
- Uncertainty in selection of compounds and variables

2.5. GRID

GRID is an alternative method to the original CoMFA approach. It calculates the interaction energy fields in molecular field analysis and determines the energetically favorable binding sites on molecules of known structure. However this approach is similar to CoMFA, but it can able to computes explicit non-bonded (or non-covalent) interactions between a molecule of known 3D structure and a probe (i.e., a small chemical group with user-defined properties) located at the sample positions on a lattice throughout and around the macromolecule. GRID has two advantages, (i) using of a 6-4 potential function for calculating the interaction energies, which is smoother than the 6-12 form of the Lennard-Jones type in CoMFA, (ii) the availability of different types of probes. The GRID software is supplied by Molecular Discovery Ltd.

2.6. CoMSIA

Comparative Molecular Similarity Indices Analysis (CoMSIA) was developed to overcome certain limitations of CoMFA. In CoMSIA, molecular similarity indices calculated from modified SEAL similarity fields are employed as descriptors to simultaneously consider steric, electrostatic, hydrophobic and hydrogen bonding properties. For computing similarity at all grid points, the mutual distances between the probe atom and the atoms of the molecules in the aligned dataset are also taken into account. To describe this distance-dependence and calculate the molecular properties, Gaussian-type functions are employed. Since the underlying Gaussian-type functional forms are 'smooth' with no singularities, their slopes are not as steep as the Coulombic and Lennard-Jones potentials in CoMFA; therefore, no arbitrary cut-off limits are required to be defined.

Despite the fact that CoMSIA also suffer from most of the limitations of CoMFA, but it has following distinctive advantages:

- Use of the Gaussian distribution of similarity indices, which avoids the abrupt changes in grid-based probe-atom interactions.
- The choice of similarity probe, is not limited to either steric or electrostatic potential fields but also include hydrophobic and hydrogen bonding (hydrogen bond acceptors and donors) fields
- Effect of the solvent entropic terms can also be included by using a hydrophobic probe
- The standard CoMFA contours highlights those regions in space where the aligned molecules would favorably or unfavorably interact with a possible receptor environment. On the other hand, the CoMSIA contours indicate those areas within the region occupied by the ligands that "favor" or "dislike" the presence of a group with a particular physicochemical property.
- CoMSIA is provided by Tripos Inc. in the Sybyl software, along with CoMFA.

2.6. MSA

Molecular Shape Analysis (MSA) is a ligand-based 3D-QSAR methodology, which try to combine conformational analysis with the classical Hansch approach. This method deals with the quantitative characterization, representation and manipulation of molecular shape in the construction of a QSAR model^[11]. In this method, each molecule in the data set to a fixed valence

geometry intramolecular conformational analysis with a scan at 30° increments for all torsion angles except for amide torsion which is scanned at 180° increment. For each compound, all apparent intramolecular energy minima are identified and recorded, each of which are then used as starting points in rigorous fixed valence geometry energy minimizations. To identify the active conformation of each analog, the LBA-LCS (loss in biological activity-loss in conformational stability) approach is used; this is based on the identification of stable low-energy intramolecular conformer states common to the active analogs, which is a high energy, unstable state for the inactive analogs. MSA is being provided in the Cerius2 software from Accelrys Inc.

2.7. COMBINE

Comparative Binding Energy Analysis (COMBINE) method was developed in order to use structural data from ligand-macromolecule complexes for 3D-QSAR model. It is based upon the hypothesis that the free energy of binding can be correlated with a subset of energy components calculated from the structures of receptors and ligands in bound and unbound forms^[12,13]. The ligands are divided into fragments and the same number of fragments is allocated to all the compounds, adding "dummy" fragments to the ligands lacking a particular fragment. Using a molecular mechanics force field, the non-bonded (van der Waals and electrostatic) interaction energies are computed between each residue of the receptor and every fragment of the ligand. The energies are also calculated between all pairs of residues/fragments for the complexes and for the free ligands and receptor. The electrostatic interactions are computed using a distance-dependent dielectric constant, and no cutoff limits are employed for the non-bonded interactions. The insignificant descriptors are then eliminated from the data using the variable selection utility in GOLPE program, and finally the biological activities of the molecules are correlated with the interaction energy values by employing PLS technique. Like all other interaction energy based 3D-QSAR approaches, COMBINE also suffers from the inherent errors involved in the computation of these energies.

2.8. Statistical Methods Used for Building QSAR Models

QSAR methods are also classified in two categories,

depending upon the type of correlation technique employed to establish a relationship between structural properties and biological activity:

Linear methods including linear regression (LR), multiple linear regression (MLR), partial least-squares (PLS), and principal component analysis/regression (PCA/PCR). Non-linear methods consisting of artificial neural networks (ANN), k-nearest neighbors (kNN), and Bayesian neural nets

2.9. Validation of 3D-QSAR Models

Validation is an important criterion of any QSAR analysis. The reliability of a 3D-QSAR model depends on how well the model can predict the activity of compounds outside the training set rather than how well the model reproduces the biological activity of compounds included in the model. The relationship between the structural parameters and the biological activities has been quantified by the PLS algorithm^[14,15]. To select the best model, the cross-validation procedure was performed using leave one out (LOO) method, in this procedure one compound was removed from the data set and its activity was predicted using the model build from rest of the data set. It gives cross-validation correlation coefficient (q^2) and the optimum number of components. Final analyses i.e. non-cross-validation was performed to calculate conventional r^2_{ncv} using optimum number of components obtained from cross validation method. The cross-validated coefficient, q^2 is calculated using below Equation 1.

$$q^2 = 1 - \frac{\sum(Y_{predicted} - Y_{observed})^2}{\sum(Y_{observed} - Y_{mean})^2} \quad (1)$$

$Y_{predicted}$, $Y_{observed}$, and Y_{mean} are predicted, actual, and mean values of the target property (pIC_{50}), respectively. $\sum(Y_{predicted} - Y_{observed})^2$ is the predictive sum of squares (PRESS) and the lowest PRESS value is used to derive the final PLS models.

3. Conclusion

3.1. Concluding Remarks

This review has briefly discussed about various 3D-QSAR approach. Even though, these methodologies have some pitfalls, but it's quite useful application in drug designing process. 3D-QSAR is an important medicinal chemistry tool for explaining the structure-

activity relationships quantitatively and retrospectively, for developing synthetic guidance leading to logical and experimentally testable hypotheses. Before using the predictive models in real-life situations, one has to look the fundamental limitations underlying in QSAR methodologies, in order to avoid inappropriate use and misinterpretation. Particularly the problems related with alignment dependency and conformational sensitivity must be taken into consideration. A comprehensive understanding and error-free practice of QSAR modeling certainly benefit the medicinal chemists to prioritize their experimental activities and considerably improve the experimental hit rates.

References

- [1] C. Hansch and T. Fujita, "p- p- α -II Analysis. A Method for the correlation of biological activity and chemical structure", *J. Am. Chem. Soc.*, Vol. 86, pp. 1616-1626, 1964.
- [2] C. Hansch, "Quantitative approach to biochemical structure-activity relationships", *Acc. Chem. Res.*, Vol. 2, pp. 232-239, 1969.
- [3] S. M. Free, J. W. Wilson, "A Mathematical contribution to structure-activity studies", *J. Med. Chem.*, Vol. 7, pp. 395-399, 1964.
- [4] T. Fujita and T. Ban, "Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters", *J. Med. Chem.*, Vol. 14, pp. 148-152, 1971.
- [5] H. Kubinyi, "Quantitative structure-activity relationships. IV. Nonlinear dependence of biological activity on hydrophobic character: a new model", *Arzneimittelforschung*, Vol. 26, pp. 1991-1997, 1976.
- [6] M. Akamatsu, "Current state and perspectives of 3D-QSAR", *Curr. Top. Med. Chem.*, Vol., 2, pp. 1381-1394, 2002.
- [7] A. J. Hopfinger and J. S. Tokarski, "Three-Dimensional Quantitative Structure-Activity Relationship Analysis. In: Practical Application of Computer-Aided Drug Design; Charifson", P.S., Ed. Marcel Dekker, Inc. New York, USA, pp. 105-164, 1997.
- [8] Y. C. Martin, "3D QSAR: Current State, Scope, and Limitations", *QSAR. in Drug Design*, Vol. 3, pp. 3-23, 1998.
- [9] R. D. Cramer, D. E. Patterson and J. D. Bunce, "Comparative molecular field analysis (comfa). i. effect of shape on binding of steroids to carrier proteins", *J. Am. Chem. Soc.*, vol. 110, pp. 5959- 5967, 1988.

- [10] K. H. Kim, "Comparative molecular field analysis (CoMFA). In: Molecular Similarity in Drug Design", Dean, P. M. Ed. Blackie Academic & Professional, Glasgow, UK, pp. 291-331, 1995.
- [11] A. J. Hopfinger, "A QSAR investigation of dihydrofolate reductase inhibition by Baker triazines based upon molecular shape analysis", *J. Am. Chem. Soc.*, Vol. 102, pp. 7196-7206, 1980.
- [12] A. R. Ortiz, M. T. Pisabarro, F. Gago and R. C. Wade, "Prediction of drug binding affinities by comparative binding energy analysis", *J. Med. Chem.*, Vol. 38, pp. 2681-2691, 1995.
- [13] G. H. Lushington, J. X. Guo and J. L. Wang, "Whither combine? New opportunities for receptor-based QSAR", *Curr. Med. Chem.*, Vol. 14, pp. 1863-1877, 2007.
- [14] S. J. Cho and A. Tropsha, "Cross-validated R2-guided region selection for comparative molecular field analysis: A simple method to achieve consistent results", *J. Med. Chem.*, Vol. 38, pp.1060-1066, 1995.
- [15] S. Wold, M. Sjostrom, and L. Eriksson, "PLS-regression: a basic tool of chemometrics", *Chemometrics and intell. lab. Sy.*, Vol. 58, pp. 109-130, 2001.