Research Article

A Review on QSAR Studies

P. Sree Mahalakshmi[*], Y. Jahnavi[1]*

[*]Sri Padmavathi Mahila Viswavidyalayam, Tirupati, Andhra Pradesh, India.
[1]G. Pulla Reddy college of Pharmacy, Hyderabad, Andhra Pradesh, India. .

ARTICLE INFO

ABSTRACT

QSAR has been applied extensively and successfully over several decades to find predictive models for activity of bioactive agents. QSAR have brought revolution in drug discovery process by development of the mathematical relationships linking chemical structures and pharmacological activity in quantitative matter of series of compound. The mathematical relationship between molecular descriptors and activity are used to find the parameters affecting the biological activity and estimate the property of other molecules. Description of the molecular structure, electronic orbital reactivity and the role of structural and steric components have been the subject of mathematical and statistical analysis.

1. Introduction

Quantitative Structure-activity relationship (QSAR) is a method for building computational or mathematical models which attempts to find a statistically significant correlation technique between structure and function using a chemo metric technique. The objectives of QSAR include: To quantitatively correlate the relationship between trends in chemical structure alterations and changes in biological end point for their biological activities. To optimize the existing leads to improve their biological activities. To predict the biological activities of untested and yet unavailable compounds.

Purpose of QSAR:
QSAR should not be seen as an academic tool to allow for the post-rationalization of data. We wish to derive the relationships between molecular structure, Chemistry and Biology for good reason. From these relationships we can develop models, and with luck, good judgment and expertise these will be predictive. There are many practical purposes of a QSAR and these techniques are utilized widely in many situations. The purpose of in Silico studies, therefore, includes the following: To predict biological activity and physic-chemical properties by rational means. To

comprehend and rationalize the mechanisms of action within a series of chemicals.

Underlying these aims, the reasons for wishing to develop these models include: Savings in the cost of product development (e.g. in the pharmaceutical, pesticide, etc. areas). Predictions could reduce the requirement for lengthy and expensive animal tests. Other areas of promoting green and greener chemistry to increase efficiency and eliminate waste by not following leads unlikely to be successful.

History of QSAR:
A QSAR, is a mathematical relationship between a biological activity of a molecular system and its geometric and chemical characteristics. 1868, A. Crum-Brown and T.R Fraser-formulated a suggestion that physiological activity of molecules depends on their constitution:

$$Activity = F (Structure)$$

They studied a series of quaternized strychnine derivatives, some of which possess activity similar to curare in paralyzing muscle. In 1869, B.J. Richardson- narcotic effect of primary alcohol varies in proportion to their molecular weights. In 1893, C. Richet has shown that toxicities of some simple organic compounds (ethers, alcohols, ketones) were inversely related their solubility in water. In 1899, H. Meyer and 1901, E. Overton have found variation of the potencies of narcotic compounds with LogP. In 1904, J. Traube found a linear relation between narcosis and surface

---

tension. In 1937, L.P. Hammett studied chemical reactivity of substituted benzenes:
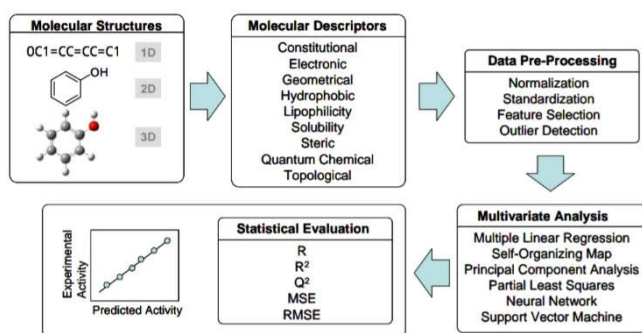
Hammett equation

Linear Free Energy Relationship (LFER)
1939, J. Fergusson formulated a concept linking narcotic activity, logP and thermodynamics. In 1952-1956, R.W. Taft devised a procedure for separating polar, steric and resonance effects. In 1964, C. Hansch and T. Fujita: the biologist's Hammett equation. In 1964, Free and Wilson, QSAR on fragments. In 1970-1980s-developmentof 2D QSAR (descriptors, mathematical formalism). In 1980s - 1990s, development of 3D QSAR (pharmacophore, CoMFA, docking). In 1990s- present, virtual screening. Requirements for a good QSAR model: A defined end point, An unambiguous algorithm, A defined domain of applicability and Appropriate measure of goodness of fit.

2. Development of QSAR Model

QSAR modeling was founded by Corwin Hansch. Now it is one of the commonly employed tools for modeling of physical and biological properties of chemicals in use today. QSAR models find broad applications for assessing the potential impacts of chemicals, nanomaterials on human health and ecological systems. The medicinal scientists worked together to discover novel molecules with unique biological activities, this was achieved by developing QSAR models and employed them for virtual screening followed by experimental validation. Quantum chemistry remains a powerful tool for exploring fundamental reactivity determinants in QSAR. The construction of QSAR/QSPR model comprises of mainly two steps: (i) description of molecular structure and (ii) multivariate analysis for correlating molecular descriptors with observed activities or properties. The preliminary step in this model is data understanding. Intermediate steps include data preprocessing and statistical evaluation.

I. Data Understanding: It is a crucial step because it helps the researcher to know the nature of the data prior to model construction there by reducing the errors. Such preliminary testing is useful for the association of the data. This is what the exploratory data analysis which often starts with simple observation of the data matrix particularly the variables (also known as attributes or fields), its corresponding data types and data samples. When it is applied to QSAR, variables represent molecular descriptors, data samples represent each unique compound, data types refer to the characteristics of the data which is quantitative or qualitative.

II. Molecular Descriptors:
Molecular descriptors can be defined as the essential information of a molecule in terms of physicochemical properties such as constitutional, electronic, geometric, hydrophobicity, lipophilicity, solubility, steric, quantum, mechanical and topological descriptors. These are the chemical information that is encoded within the molecular structures in which they are available for transformation. Once the molecular descriptors are calculated, they serve as an independent model for further construction of an QSAR model. Some of the properties explored in QSAR:

Biological properties	Chemical properties
Bioconcentration	Boiling point
Biodegradation	Chromatographic retention time
Carcinogenicity	Dielectric constant
Drug metabolism and clearance	Diffusion coefficient
Inhibitor constant	Reactivity
Mutagenicity	Solubility
Permeability	Stability

III. Data Pre-processing: It is the most important phase of data mining as it helps to ensure the integrity of the data before preceding further. To obtain reliable QSAR models it is important to handle the data with great care.

IV. Data cleaning: Data cleaning, as raw data often contain errors, inconsistencies such as missing data, incomplete data and invalid character values which may cause trouble for data mining software if left untreated.

V. Data transformation: Variability in the distribution and range of each variable in the dataset. This may create a problem in neural networks, which is handled by applying statistical techniques such as min-max normalization or Z-score standardization. In min-max, the minimum and maximum values are adjusted in between 0 & 1.

VI. Multivariate analysis: An approach to quantitatively describe the relationship between independent variables (molecular descriptors) and dependent variables (biological/chemical properties of interest).

Multiple linear regression: It is used to screen the appropriate descriptor from a large pool of descriptors. It is a method used to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation, to correlate the binding affinity with that of the molecular descriptors.

Partial least squares: Also referred to as classical algorithm. It is a method of constructing predictive models when the factors are many. This method will give the minimum number of variables, which is essential to generate the model and at the same time to gather the information about the molecule. It is also a popular method for soft modeling in industrial applications.

Artificial Neural Network: Neural networks can be use to generate predictive models of QSAR between a set of molecular descriptors obtained from multiple linear regression and observed activity. It is the generation of the mathematical models of biological system. These are model free mapping devices that are capable by approaching QSAR.

Classification of QSAR methodologies: Based on dimensionality

1D-QSAR	Molecular representations and molecular fragments i.e., pKa, log P with biological activity.
2D-QSAR	Contains topological information i.e., physico-chemical properties with biological activity.
3D-QSAR	Correlation of various 3D properties which surrounds the molecule.
4D-QSAR	Ligand receptor interactions of the drug molecule with the 3D properties.
5D-QSAR	Representing different induced-fit models in 4D-QSAR.
6D-QSAR	Incorporating different salvation models in 5D-QSAR.

3. Advancements of QSAR

In Drug Discovery & Design:

Now-a-days the structure activity studies are more important to design a drug and its development. There are so many diseases which are not having any drugs to cure, so adopting new techniques and methods is necessary. Drug discovery and design not only requires discovery or design, but also the synthesis of the drug, method of administration, safety assessment. Extraction of natural compounds such as morphine and cocaine used by the physicians were too safe. For these, the search to find fewer toxic substances, the development was based on the structures of known pharmacologically active compounds. These are referred to as leads.

SAR stands for Structure Activity Relationship while QSAR stands for Quantitative Structure Activity Relationship. SAR deals with the relationship of structure with biological activity while QSAR defines the relationship of magnitude of various structural properties with the biological activity. Compounds with similar structures to a pharmacologically active drug are often themselves biologically active. A study of SAR of a lead compound may be used to determine the structure of the lead compound responsible for both its beneficial biological activity i.e., pharmacophore and its unwanted side effects. SAR are simply to made different structural changes to make the molecule beneficial. These changes may be: The size, shape and branching of parent structure, Type of substitution and their nature, The stereochemistry of lead compound. QSAR/QSPR are of great importance in medicinal chemistry and biochemistry because they can accelerate the development of new compounds for use as drugs.

Virtual Screening Approach in QSAR:

Virtual screening has emerged in drug discovery as a powerful computational approach to screen large libraries of small molecules for new hits with desired properties that can be tested experimentally. Among the virtual screening approaches, QSAR is the most powerful method due to its high and fast throughput. QSAR models are applied to predict the biological property of novel compounds. High throughput screening (HTS) technologies resulted in the explosion of amount of data suitable for QSAR modeling. The limitations in the above step are corrected as data curation as a first and mandatory step. Data curation procedures include removal of organometallics, mixtures, in organics etc. HTS can rapidly identify large subsets of molecules with desired activity from large screening collection of compounds using automated plate based experimental assays.

Here there are some applications of QSAR based virtual screening for the discovery of new hits and hit to lead optimization. QSAR models for malaria were built using descriptors (0D, 1D, 2D) and support vector machines (SVM) method. QSAR models were applied using chembridge database. After virtual screening, 176 anti-malarial compounds were identified and submitted to experimental validation. External validation are

necessary and is an essential step in any computer aided drug discovery study and wanted to reinforce the QSAR modeling.

Nano - QSAR Model:

A relatively newly developed type of QSAR methodology, QSAR for nano particles or nano-QSAR, is based on defining mathematical dependencies between the variance in molecular properties encoded by so-called **nano-descriptors** and the variance in biological activity for a set of nano material. During the last three decades a dramatic increase in research and applications involving the chemistry and technology of nanomaterials has been observed.

Nano-QSAR modeling of Physico-chemical Properties: In order to govern the unique physic-chemical properties of nano particles one needs a basic knowledge as well as a means of controlling their other characteristics such as solubility, stability, reactivity, surface properties. All these properties become extremely important when nanomaterials are produced in large quantities and one has to optimize them to reduce the cost of manufacturing. Unfortunately, such an approach, although perhaps cost-efficient, could negatively affect the environment directly by applied technology.

Nano-QSAR and Prediction of Toxicity: The biological activities, such as pharmacological activities and toxicity, are vital characteristics that relate to the peculiar physico-chemical properties of nanomaterials. It is known now that particle size, hydrophobicity, and protein identity all contribute to nano-protein association. For instance, fullerenes are capable of specific interactions with proteins as evidenced by the production of fullerene-specific antibodies. Fullerenes have been identified as inhibitors of Cysteine and serine proteinases as well as the protease specific for the human immunodeficiency virus (HIV) and are known to interact directly with the virus. The list of possible physico-chemical properties to which toxicity of nano particles can be related to. The list includes: Size, Size distribution, Agglomeration rate, Shape, Porosity, Surface area, Chemical composition, Surface chemistry, Structure-dependent electronic configuration, Surface charge and Crystal structure.

Automated Framework for QSAR Model Building:

The main focus of the current work is to present an implementation of a well-defined and efficient modeling procedure capable of building robust and reliable models validate them both internally and externally. To accomplish this, it was necessary to address two critical Issues in QSAR.

1. The first one is to know how to deal with high dimensional data by identifying and selecting the subset of descriptors sufficient to predict the desired biochemical property.

2. The second aspect in a modeling workflow is model validation, So that the model results can be unbiasedly assessed.

The key ideas behind structuring the presented automated QSAR modeling workflow platform are some discussed below:
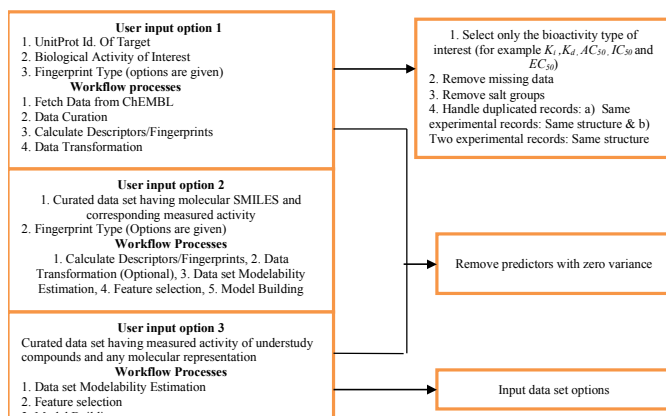
- It should be freely available and support any operating system with easy installation.
- Should be easily be applied for fully automated QSAR modeling by directly accessing up to date data from online molecules databases or by using private data sets.
- Reliably perform most critical tasks of QSAR modeling including Descriptors/fingerprints calculation, feature selection, model building, validation and prediction.
- Make a prior estimation of the feasibility of any given data set to produce a predictive QSAR model before the time-consuming process of feature selection, model building and validation.
- For different applications and reproduction of historical predictions, all outputs of intermediate tasks and each previous version of models must be stored on local machines.

- Regarding extensibility, the frame should provide useful starting points for performing customization to modify and further extend the existing workflow by domain specific interests.
- The application must follow the same protocol of training series to re-train and update models with new molecules as they become available and to make external predictions.
- It should adopt the best optimized feature selection methodology to select the adequate features for each problem. This is a critical task necessary to avoid over-fitting and to have better understanding of the data, the model and the factors involved.

For different applications and reproduction of historical predictions, all outputs of intermediate tasks each previous versions of models must be stored on local machines.

4. Fully automated mode

Fragment Based QSAR methods

Recently, Du et al. [19] introduced a 2D-QSAR method based on molecular fragments. The method uses a mixed Hansch-Fujita linear free energy equation and Free-Wilson equation. In particular, molecular fragments are first generated from ligands and the total binding free energy ΔG^0_i between the ligand i and the receptor is considered as the sum of contributions $\Delta g_{I,\alpha}$ from all fragments:

$$\Delta G^0_i = \sum_{\alpha=1}^{M} b_\alpha \Delta g_{i,\alpha}$$

Where $\Delta g_{i,\alpha}$ for each fragment. The binding free energy of a fragment, $\Delta g_{i,\alpha}$, is described by a set of physical and chemical properties of a fragment.

$$\Delta g_{i,\alpha} = \sum_{l=1}^{L} a_l p_{i,\alpha,l}$$

Where $p_{i,\alpha,l}$ is the l-th property of fragment $F_{i,\alpha}$ in the molecule mi and al is the coefficient of l-th property of the fragment.

Multi-Dimensional QSAR Methods:

Multi-dimensional (nD) QSAR methods are essentially extensions of characteristics or properties to tackle drawbacks of 3D-QSAR methods. One example is 4D-QSAR by Hopfinger et al. [51] which samples molecular conformations and alignments during the generation of a QSAR model. While incorporating some CoMFA features, it introduces the fourth dimension, which is the conformational Boltzman sampling, and enables

the method to be used as a receptor-independent (RI) method as well as receptor-dependent (RD) method in which the geometry of the receptor is known. It should be noted that their 4D-QSAR method does not solve the alignment problem but it allows a rapid evaluation of individual trial alignments. Such 4D-QSAR implementation can be found in XMAP program.

Conflict of Interest

The author(s) confirm that this article content has no conflict of interest.

References

1. Kyaw Zeyar Myint and Xiang-Qun Xie. Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods. International Journal of Molecular Sciences. 2010.
2. Drug designing : A Review by Pharmatutor.
3. Artem Cherkosav. QSAR modeling. Chemical Society Reviews.
4. Rakesh Bhatia. History in the revolution of QSAR : A Review.
5. Vinicius G Maltarollo et al. Current trends in quantitative structure-activity relationship validation and applications on drug discovery. Future Science. 3(4) : 2017.
6. Sunyoung Kwon et al. Comprehensive ensemble in QSAR prediction for drug discovery. BMC Bioinformatics. 2019.
7. Karolina Jagiello. Advantages and limitations of classic and 3D QSAR approaches in nano-QSAR studies based on biological activity of fullerene derivatives. Journal of nano particle Research. 18(9) : 2016.
8. Anil Kumar Sahdev. Role of QSAR : Significance and uses in molecular design. 5(1) : 2018.
9. Samina Kausar. An automated framework for QSAR model building. Journal of Cheminformatics. 2018.
10. Shanno Pathan et al. Quantitative structure activity relationship and drug design : A Review. International Journal of Research in Biosciences. 5(4) : 1-5, 2016.
11. Bruno J. Neves et al. QSAR-Based Virtual Screening : Advances and applications in drug discovery. Frontiers in Pharmacology. 2018.
12. Artem Cherkasov. QSAR Modeling. Journal of medicinal chemistry. 57(12) : 4977-5010, 2014.
13. Ojha Lokendra K et al. Modern drug design with advancement in QSAR : A Review. International Journal of research in Biosciences. 2(1) : 1-12, 2013.
14. Shirish Ambulgekar. QSAR models and guidance : A Review. Asian Journal of research in chemistry and pharmaceutical sciences. 4(1) : 1-10, 2016.
15. Umma Muhammad et al. Review on QSAR modeling. Journal of Analytical and Pharmaceutical research. 7(2) : 2018.
16. Chanin Nantasenamat et al. A practical review of Quantitative Structure-Activity Relationship. 2009.
17. Layla Abdel-IIah et al. Applications of QSAR in drug design. International Journal of Research and Technology. 6(6) : 2017.
18. Bakhtiyor Rasulev et al. Nano-QSAR : Advances and challenges. Nano science and Nanotechnology. 2012.