WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A REVIEW: AN OVERVIEW ON QSAR DEVELOPMENT AND 2 DQSAR IN CHYMOTRYPSIN LIKE PROTEASE.

Mali Asmita Jagannath, Shalgaonkar Akshada Pradip

Department of pharmaceutical chemistry pharmacy , Late NarayandasBhavandasChhabada Institute of pharmacy, Raigaon,Satara,Shivaji university,Kolhapur Maharashtra,India
Telephone no: 8767504312, +918999816238
Gmail: asmitamali279@gmail.com akshupradipshalgaonkar@gmail.com

ABSTRACT:

Over the last few decades, QSAR has been used extensively and successfully to create predictive models for the activity of bioactive compounds. By developing mathematical correlations relating chemical structures and pharmacological activity in quantitative matter of series of compounds, QSAR has revolutionised the drug discovery process. The mathematical relationship between molecular descriptors and activity is used to identify the parameters that impact biological activity and to estimate the properties of other molecules. Mathematical and statistical analysis have been used to describe molecule structure, electronic orbital reactivity, and the function of structural and steric components.(1)

Keywords: QSAR, structure, molecular,model

Introduction:

Quantitative structure-activity relationships (QSARs) are empirical statistical models and are a mature part of the computational chemistry toolkit. While there are many successful and predictive QSAR models, equally there are many models in the literature that encode little more than noise. This is because QSAR derivation and validation is fraught with complexity and pitfalls. The machine learning algorithms used are extremely powerful tools to fit any data to a signal, even random noise if they are not used with care. This chapter highlights some examples of good practice in model building and use and gives examples of influential QSAR models. The chapter also provides a glimpse of how automation is being used to remove the subjectivity in QSAR model generation and testing, to provide more robust QSAR models. QSAR model building is more likely to be the domain of a skilled computational chemist, or even an automated QSAR model building system.[1]

History

Introduction fundamental observation in chemistry is that changes in structure lead to changes in a measured experimental property. The changes in the structure lead to variations in the shape and electronics of the compound, which in turn can be approximated by numerical descriptors. Quantitative structure-activity relationship (QSAR) analysis seeks to relate the changes in the observed property to these numerical descriptors, with the goal of predicting novel compounds with the desired properties. QSAR analysis was first described by Hansch.[2]

What Is 2D QSAR:

2D QSAR is a powerful tool for explaining the relationships between chemical structure and experimental observations. Key elements of the method are the numerical descriptors used to translate a chemical structure into mathematical variables, the quality of the observed data, and the statistical methods used to derive the relationships between the observations and the descriptors. There are some caveats to what is essentially a simple procedure: overfitting of the data, domain applicability to new structures, and making good error estimates for each prediction. 2D QSAR models are used routinely during the process of optimization of a chemical series toward a candidate for clinical trials.[3]

The most interesting feature of predictive QSAR models is that the behavior of any new or even hypothesized molecule can be predicted by the use of mathematical equations. The phrase “2D-QSAR” signifies the development of QSAR models using 2D descriptors. Such predictor variables are the most widely practiced ones because of their simple and direct mathematical algorithmic nature involving no time-consuming energy computations and having reproducible operability.[3]

II.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 and 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 employing them for virtual screening followed by experimental validation. Quantum chemistry remains a powerful tool for exploring fundamental reactivity determinants in QSAR. The construction of the 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.[4,5]

Data Understanding:

It is a crucial step because it helps the researcher to know the nature of the data prior to model construction thereby 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, and data types refer to the characteristics of the data which is quantitative or qualitative.

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 the further construction of a QSAR model. Some of the properties explored in QSAR.

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.

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.

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.

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 a 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 used 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 a biological system. These are model-free mapping devices that are capable of approaching QSAR.

Basic Concepts In QSAR:

Descriptors are one of the most essential components of predictive Quantitative Structure- Activity/Property/Toxicity Relationship (QSAR/QSPR/QSTR) modeling analysis, as they encode chemical information of molecules in the form of quantitative numbers, which are used to develop mathematical correlation models. The quality of a predictive mode l not only depends on good modeling statistics but also on the extraction of chemical features . A significant amount of research since the beginning of the QSAR analysis paradigm has led to the introduction of a large number of predictor variables or descriptors. [6,7]

I.Importance of Applicability Domain of QSAR

Model- Quantitative Structure-

Activity Relationship (QSAR) models have manifold applications in drug discovery, environmental fate modeling, risk assessment, and property prediction of chemicals and and pharmaceuticals. One of the principles recommended by the Organization of Economic Co- operation and

Development (OECD) for model validation requires defining the Applicability Domain (AD) for QSAR models.

II.Computational Techniques Application in Environmental Exposure Assessment

In this chapter,the application of computational techniques in environmental exposure assessment was described. The most important groups of the setechniquesareMulti media Mass-balance (MM) modelling and Quantitative Structure-Activity/Structure-Property Relationships (QSAR/QSPR) modelling.

Automated QSAR (Quantitative Self-Assessment Research):

In recent years, automating or semi-automating the model generation process has become popular.

Breaking down the model generation process into many parts, such as descriptor computation, feature selection, and model development, was pioneered by the Discovery Bus.

The components were then automatically blended to produce new models.

New approaches, such as a new descriptor set or modelling algorithm, may be added to the system in the future, and the system would generate new models.

In essence, the Discovery Bus employed a competitive mechanism to conduct a comprehensive search across model and descriptor space, identifying good approach combinations automatically.

AstraZeneca created (4)AutoQSAR, a system that semiautomates the model development process and keeps models up to date with the most recent observed data(.3)

Furthermore, the technology enables for the semi-automated creation of project- or series-specific models locally.(7)Software solutions for semi-automated model construction have been created.GlaxoSmithKline developed the QSAR Workbench in conjunction with Accelrys, which assists users through the model building, validation, and selection process(6).

2D QSAR in SARS-Cov 3 chymotrypsin-like Protease:

Developing broad-spectrum anti-coronavirus drugs is greatly important since the novel SARS-CoV-2 has rapidly become a threat to the public health and the economy worldwide. SARS- CoV 3-chymotrypsin-like protease (3CLpro), as highly conserved in beta coronavirus, is a viable target for anti-SARS drugs. A quantitative structure-activity relationship (QSAR) for inhibitory constants (PKI) of 89 compounds against the SARS-CoV 3CLpro enzyme was developed by using a support vector machine (SVM) and genetic algorithm.[8]

Quantitative structure-activity relationship (QSAR) in SARS-COV 3 Chymotrypsin-like Protease:

Models can be used for drug screening and mechanistic understanding of drug action. This technique has many advantages, such as lower cost and higher speed, and even can be used to evaluate drug candidates that have not been synthesized. But only a few researchers have carried out QSAR studies for inhibitor activities against SARS-CoV.

Methods-

1.Experimental data-

Table S1 in Supplemental Information shows the SMILES notations and inhibitory constants (Ki) of 89 molecules against SARS-CoV 3CLpro, which were taken from the binding database [11] and references [12–14]. The experimental Ki values varied from 3 to 56,000 nM, and their pKi (=logKi) values were in the range of 8.523–4.252 by converting to negative logarithm of Ki. A larger pKi value means a higher activity for the inhibitor. Inhibition constants, Ki, were obtained through measuring the apparent kinetic parameters at a constant substrate concentration (10 mM) and different inhibitor concentrations (0–200 mM) at 25 C [12]. These experimental data were randomly split into a training set (n = 65 inhibitors) and a test set (n = 24 inhibitors). QSAR models were developed with the training set and evaluated with test set.

2.Molecular descriptors-

Besides three-dimensional (3D) QSAR methods based on ligand-receptor interactions, the two-dimensional (2D) QSAR models derived only from the ligand molecules can be used for describing the activity of biologically active compounds. The structural and physicochemical features of active compounds become the critical factors determining inhibitory constant (Ki) when the inhibitors have the same biological target (e.g., SARS-CoV 3CLpro enzyme). In this study, the structures of inhibitors were used to derive molecular descriptors for 2D QSAR models of inhibitory constants (Ki). According to the SMILES notations in Table S1, molecular structures were drawn using Chem Bio Draw Ultra 12.0 in Chem Bio Office 2010. Subsequently, 3D-structures were generated using ChemBio3D Ultra 12.0 and optimized with semi-empirical AM1 method in Gaussian 09. Finally, the optimized molecules were used to calculate molecular descriptors with Dragon 6.0. Totally, 648 descriptors were obtained when those molecular descriptors with high co-linearity ($|R| > 0.90$) or being a constant were removed.

3.SVM principle-

For the nonlinear support vector regression machine, the low-dimensional data need to be mapped to the high-dimensional space, from which the linearly separable hyperplane would be found. Finally, the hyperplane in the high-dimensional space should be mapped back to the low-dimensional space, so as to realize SVM regression or classification. However, mapping low-dimensional data to high-dimensional space and then performing regression analysis involve a great number of computations. Especially for high-dimensional data, the problem of over-fitting can occur. Kernel function is introduced to solve this problem. Replacing the linear terms in linear equations with kernel function can make the original linear algorithm nonlinear, that is to say, it can do nonlinear regression. Thus, the introduction of kernel function can achieve the purpose of increasing dimension and effectively control ovover fitting Radial basis function was used in this work.

CONCLUSION:

When utilised appropriately, QSAR models are a tremendously powerful tool. Some of the difficulties have already been mentioned (bad data sets, over-fitting, domain applicability, model deterioration, error and confidence estimates), but if these issues are addressed, QSAR models can be very useful in the design process. QSAR models should not be viewed in isolation, as each model requires its own protocol and equipment to provide a result. They can be thought of as parts of a larger design process. Several organisations have begun to develop architectures in which models are delivered as standalone web services, allowing them to be captured as workflow nodes in Knime or Pipeline Pilot. Because drug design is fundamentally a multi-objective optimization, the results of multiple models, such as bioactivity, solubility, and hERG inhibition, can be integrated to produce a solution.

REFERENCES:

1. Chackalamannil S, Rotella D, Ward S. Comprehensive medicinal chemistry III. Elsevier; 2017 Jun
2. Free SM, Wilson JW. A mathematical contribution to structure-activity studies. Journal of medicinal chemistry. 1964 Jul;7(4):395-9.
3. Sathyamurthy B, Sushmitha HS. In Silico drug designing studies on Dengue Virus NS1 Protein. PharmaTutor. 2018 Oct 1;6(10):31-6.
4. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic acids research. 2016 Jan 4;44(D1): D1045-53.
5. Thanigaimalai P, Konno S, Yamamoto T, Koiwai Y, Taguchi A, Takayama K, Yakushiji F, Akaji K, Chen SE, Naser-Tavakolian A, Schön A. Development of potent dipeptide-type SARS-CoV 3CL protease inhibitors with novel P3 scaffolds: design, synthesis, biological evaluation, and docking studies. European journal of medicinal chemistry. 2013 Oct 1; 68:372-84.
6. Mahalakshmi PS, SreeMahalakshmi P, Jahnavi Y. A review on QSAR studies. Int J Adv Pharm Biotechnol. 2020;6:19-23.
7. Wood DJ, Buttar D, Cumming JG, Davis AM,Norinder U, Rodgers SL. Automated QSAR with a hierarchy of global and local models. Mol Inform 2011, 30:960–972.
8. Cox R, Green DVS, Luscombe CN, Malcolm N,Pickett SD. QSAR workbench: automating QSAR modeling to drive compound design. J Comput Aided Mol Des 2013, 27:321–33
9. Martin E, Mukherjee P, Sullivan D, Jansen J.Profile-QSAR: a novel meta-QSAR method that combines activities across the kinase family to accurately predict affinity, selectivity, and cellular activity. J ChemInf Model 2011, 51:1942–1956.