MANTECH Publications

Journal of Pharmaceutical Chemistry and Drug formulation
Volume 4, Issue 1, January-April, 2022

Quantitative Structure-Activity Relationship (QSAR) Study Applications in Drug Design

Pooja Srivastava[1], Alka Goyal[2]
Assistant Professor[1], Student[2]
Department of Pharmaceutical Chemistry
Tatyasaheb Kore College of Pharmacy
Corresponding Author's Email: - *ggoyalalka123@gmail.com[2]*

Abstract

In silico strategies for rational drug design include quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) research. The goal of these strategies is to enhance the biological activity and physicochemical qualities of current leads. Predicting the biological actions of unproven and often yet unavailable substances is another goal.

This article offers a broad review of many QSAR/QSPR investigations from past research. In certain research, the R2 and Q2 parameters are utilised to predict the predictability and resilience of the created models. QSAR studies were used in all of the papers listed to investigate pharmacological activity or binding mode on individual receptors.

Keywords: *Squared Correlation Coefficient Q2, QSAR/QSPR, physicochemical qualities, Molecular Descriptor, Drug design*

INTRODUCTION

Drug research and development is a process that tries to create safe and effective treatments in order to improve life quality and minimize suffering to the bare minimum. However, the process is very complicated, time-consuming, and resource intensive, necessitating multidisciplinary knowledge and novel techniques. According to recent estimates, it can take up to 13.5 years and 1.8 billion

US dollars to get a new medicine to market.

Medicine and health care technology have advanced quickly in recent decades. Biomedical engineering development is an important role in the solution of medical issues.

Over the last ten to twenty years, there has been a greater emphasis on applying computational abilities to the combined chemical and biological space in order to simplify drug discovery and process design.

In compared to traditional drug discovery approaches, rational drug design methods reduce the time and expense required in the drug design process. QSAR/QSPR research may be used to develop and find novel inhibitors from scratch, as well as to improve the absorption, distribution, metabolism, excretion, and toxicity profiles of known compounds from multiple sources. Advances in computational techniques and technology have simplified the use of in silico methodologies in the design process. Structure-based drug design (SBDD) and Ligand-based drug design (LBDD) are the two types of drug design. SBDD is a method of developing an inhibitor based on the structural knowledge of the therapeutic target. In the lack of receptor 3D information, LBDD depends on compounds that bind to the biological target of interest. Figure I depicts the many categories and types of drug creation strategies.

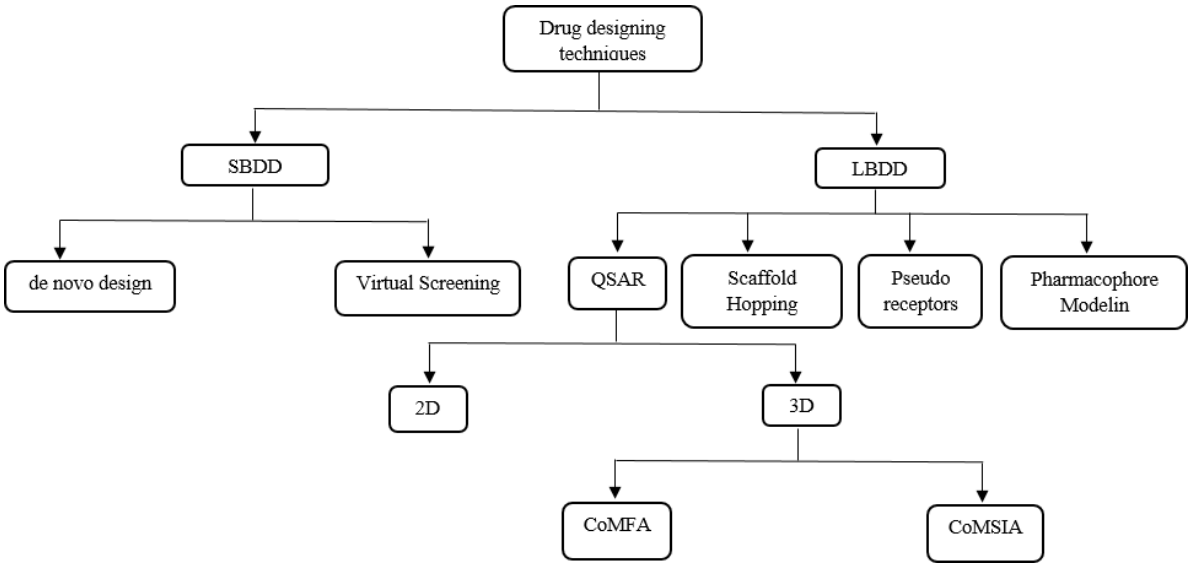

Fig I: Different Groups and Types of Drug Designing Technique

These days, quantitative structure-activity relationships (QSAR) play an important part in the drug discovery process since they are a less expensive alternative to medium throughput in vitro and low throughput in vivo experiments.

In addition, QSAR models are currently considered as a scientifically reliable technique for predicting and categorizing the biological activities of untested chemicals, drug resistance, toxicity prediction, and physicochemical property prediction in drug development and environmental toxicology.

The QSAR technique is founded on the idea that variations in a series of chemicals' biological activity may be quantitatively associated with changes in their molecular structure. As a result, all biological activities and functions of molecules are associated with unique chemical descriptors, and specialised regression techniques may be used to quantify the relative contributions of those descriptors to the biological impact.

METHODS

A. QSAR Definition and Development

Quantitative structure activity relationship (QSAR) is one of the widely used approaches in ligand based drug designing processes.

In QSAR/QSPR studies quantitatively correlate and recapitulate the relationships between trends in chemical structure alterations and respective changes in biological endpoint for comprehending which chemical properties are most likely determinants for their biological activities or physicochemical properties.

Quantitative Structure Activity Relationships (QSARs) mean computerized statistical method which helps to explain the observed variance in the structure changes caused by the substitution. In this concept it is assumed that the biological activity exhibited by a series of congeneric compounds is a function of various physio-chemical analysis is performed it shows that certain physio-chemical properties are favorable to the concern activity, the latter can be optimized by choosing such substituent’s which would enhance such physiochemical properties. A major goal of Quantitative Structure Activity Relationship (QSAR)/ Quantitative Structure Property Relationship (QSPR) studies is to find a mathematical relationship between the activity or

property under investigation, and one or more descriptive parameters or descriptors related to the structure of the molecule.

In QSAR, the structure of a molecule must contain the features and properties responsible for its physical, chemical, and biological activities [14]. Figure 2 describes different stages the development QSAR model process.

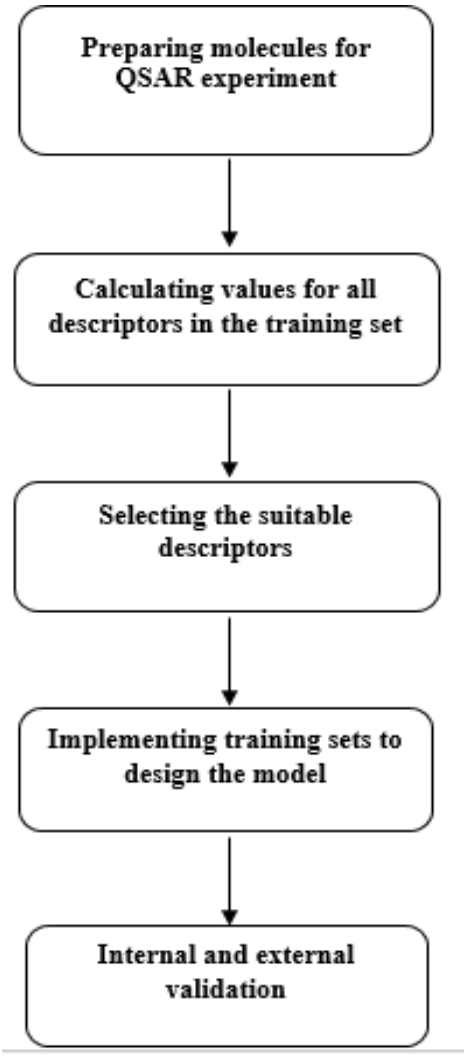

Figure 2: QSAR Development process.

There are several software available for QSAR development, both commercial and free. These include specialist software for sketching chemical structures, converting chemical file formats, producing 3D structures, computing chemical descriptors, and constructing QSAR models, as well as general-purpose software that contains all of the components required for QSAR creation. ChemDraw, ACD/ChemSketch, and Open Babel software are the most often used tools for structure drawing and file conversion. CORINA, Concord, Frog, and smi23d are 3D structure generation software. Dragon, Molconn-Z, and PaDEL-Descriptor software can be used to calculate descriptors.

The input of molecular structures and development of 3-D models is the first important stage in a QSPR/QSAR investigation. Geometric descriptor computations need 3-D molecular models. The development of molecular structure descriptors is the second main stage in a QSPR/QSAR investigation. The third stage is to choose the most significant descriptors, which may be accomplished using feature selection methods. The development of QSPR/QSAR models using descriptor sets is the fourth main stage in a QSPR/QSAR investigation. The

model is validated in the fifth and final stage by predicting the activity of substances in the external prediction set. To simply evaluate the model's fitness level, the predictions' results should be compared to those produced for the training and cross validation sets.

- **Molecular Descriptors**

Molecule descriptors are the end result of mathematical techniques that convert chemical information recorded in a molecular structure to a numerical representation. Dimensionality of molecular descriptors, as stated below, helps determine QSAR model type:

0D QSAR- These are descriptors derived from molecular formula e.g. molecular weight, number and type of atoms etc.

1D QSAR- A substructure list representation of a molecule can be considered as a one-dimensional (1D) molecular representation and consists of a list of molecular fragments (e.g. functional groups, rings, bonds, substituents etc.).

2D QSAR- A molecular graph contains topological or two dimensional (2D) information. It describes how the atoms are bonded in a molecule, both the type of bonding and the interaction of particular atoms (e.g. total path count, molecular connectivity indices etc.).

3D QSAR- These are calculated starting from a geometrical or 3D representation of a molecule. These descriptors include molecular surface, molecular volume and other geometrical properties. There are different types of 3D descriptors e.g. electronic, steric, shape etc.

4D QSAR-Four Dimensional Information is described in this type of models, and the fourth dimension is an ensemble of conformation of each ligand [16].

5D-QSAR– Five Dimensional information is described in this type of models, and the fifth dimension is the possibility to represent an ensemble of up to six different induced-fit models.

The descriptors are fall into 4 classes: Topological, Geometrical, Electronic and Hybrid.

Topological descriptors in chemistry are graph invariants generated by applying the theorems of graph theory. Examples of topological descriptors are: atom counts, ring counts, molecular weight, weighted paths, molecular connectivity indices, substructure counts, molecular distance edge descriptors, kappa indices, electro-

topological state indices, and some other invariants.

Aspects of the structures related to the electrons are encoded by calculating electronic descriptors. Examples of electronic descriptors are: partial atomic charges, HOMO or LUMO energies, dipole moment.

Geometric descriptors are used to encode the 3-D aspects of the molecular structure such as moments of inertia, solvent accessible surface area, length-to-breadth ratios, shadow areas, gravitational index.

A class of hybrid descriptors called charged partial surface area descriptors encode the propensity of compounds to engage in polar interactions. The set of cpsa descriptors is based on the partial atomic charges and the partial surface area of each atom. The two attributes lists are mixed and a set of approximately 25 cpsa descriptors can be generated by matching the two mixed lists with different weighting schemes. Examples of cpsa descriptors can include: fractional positive surface area, charged weighted negative surface area.

- **QSAR models validation**

Validation process aims to provide a model which is statistically reliable with selected descriptors as a consequence of the cause-effect and not only of pure numerical relationship obtained by chance. However, non-statistical validations such as verification of the model in terms of the known mechanism of action or other chemical knowledge are necessary; it is not acceptable to rely on statistics only in validation process. Actually, this is somehow a hard procedure for cases where no mechanism of action is known or where data sets are small.

Validation methods are needed to establish the productiveness of a model. There are two types of validation methods: Internal and external. Internal methods depend on training datasets like Q2 (squared correlation coefficient), R2 (coefficient of determination or the coefficient of multiple determination for multiple regression), chi-squared (X2), and root-mean squared error (RMSE). The major disadvantage of this approach is the lack of predictability of the model when it is applied to a new data set. However, external methods depend on the testing set and it is considered as best validation method.

It was reported that, in general, there is no relationship between internal and external predictivity: high internal predictivity may result in low external

predictivity and vice versa. In many cases, comparable models are obtained where some models show comparatively better internal validation parameters and some other models show comparatively superior external validation parameters. This may create a problem in selecting the final model. Therefore, it is must to develop some good validation techniques to overcome the entire above mentioned disputes.

B. QSAR in Drug design

QSAR is involved in drug discovery and designing to identify chemical structures with good inhibitory effects on specific targets and with low toxicity levels.

The implementation of QSAR in designing different types of drugs as antimicrobial and antitumor compounds by numerous works is a strong evidence of its efficiency in drug designing. Previous research in this field has been undertaken by different researchers.

Researchers investigated QSAR study on a series of 8- substituted xanthines as adenosine antagonists have been carried out. The chemical structure was described with parameters effect the receptors affinity.

In, two multilayer feed forward neural networks and docking studies were developed to investigate the hypothetical binding mode of the target compounds.

Two 3D-QSAR models for a series of non-purine xanthine oxidase inhibitors were designed to study different factors affect the oxidase inhibitors.

QSAR model of xanthine oxidase inhibitory flavylium salts was implemented to predict the inhibitory potency of anthocyanidins as a function of their molecular properties.

A three-dimensional QSAR study has been implemented to study epothilones– tubulin depolymerization inhibitors.

QSAR models are established for the toxicity of polycylic aromatic hydrocarbons (PAHs).

Four dimensional QSAR models is used to study a set of 18 structurally diverse antifolates including pyrimethamine, cycloguanil, methotrexate, aminopterin and trimethoprim, and 13 pyrrolo [2,3-d] pyrimidines.

The utility of Topological polar surface area (TPSA) was demonstrated in 2D

QSAR for 14 sets of diverse pharmacological activity data.

QSAR of Hydrazones of N-Amino-N'-hydroxyguanidine as Electron Acceptors for Xanthine Oxidase was built.

Antiviral QSAR models are implemented to predict by the first time an mt-QSAR model for 500 drugs tested in the literature against 40 viral species. The Markov Chain theory is used to calculate new multi-target entropy that fits a QSAR model.

RESULTS & DISCUSSION

A. QSAR Implementing in Drug Designing Results

It is very important to validate the model's performance to conclude whether the results satisfy researcher's expectations or not. R2 and Q2 are two statistical measures used for this purpose [35-41]. Values of R2 and Q2 obtained from different previous researches are listed in Table I.

R2 (called as coefficient of determination or the coefficient of multiple determinations for multiple regression.) is a statistical measure of how close the data are to the fitted regression line; High R-squared indicates that the model has a good fit. According to previous research, R2 should be ≥
0.6 to consider the model fits well. As it is shown in Table I, all QSAR models developed have a higher R2 value than 0.6. Q2 is squared correlation coefficient and it is used as a criterion of both robustness and predictive ability of the model. It can be considered as an indicator of the high predictive power of the QSAR model. However, high Q2 value is not enough to conclude that the model has acceptable predictive ability; models should be tested for their ability to predict the activity of compounds of an external test set also.

It was proven that for good predictability R2 – Q2 value should not be larger than 0.3 [42]. R2 – Q2 values are calculated for researches [35-41] and added in Table I.

As it can be noticed from Table 1, only in the R2 – Q2 value exceeds 0.3, while in all other works the values are very small (lower than 0.3), which indicates a good predictability of the constructed models in these works.

However, QSAR predictability and robustness levels cannot be proved by R2 and Q2 values only; more parameters should be involved to obtain a strong conclusion, as: chi- squared (□2), root-

mean squared error (RMSE), correlation coefficient R between the predicted and observed activities, slopes k and k' of the regression lines through the origin. Also, the chemical space of training and test sets has to be discussed and studied; real outliers, with respect to character and structure similarities, have to be found and removed. Only a small number of reported QSAR studies were implementing numerous different validation characteristics in their QSAR validation processes.

Table 1: R^2 and Q^2 Values Obtained By Applying Qsar Models in Different Works $R^2 – Q^2$ Values Are Calculated For Each Work.

Reference	Paper Title	Descriptors	R^2 value	Q^2 value	$R^2 – Q^2$
[35]	More Effective DPP4 Inhibitors as Antidiabetics Based on Sitagliptin Applied QSAR and Clinical Methods	Hydrophobicity, counts of rotatable bonds, hydrogen bond donor and acceptor atoms, and topological polar surface area.	0.85	0.77	0.08
[36]	Molecular modelling studies of 3,5-dipyridyl-1,2,4-triazole derivatives as xanthine oxidoreductase inhibitors using 3D-QSAR, Topomer CoMFA, molecular docking and molecular dynamic simulations	Steric, electrostatic, and hydrophobic fields.	0.988	0.578	0.41
[37]	Prediction of caspase-3 inhibitory activity	HOMO, LUMO energies	0.955	0.885	0.07

	of 1,3-dioxo-4-methyl-2,3-dihydro-1h-pyrrolo[3,4-c] quinolines: QSAR study				
[38]	Predictive QSAR modeling on tetrahydropyrimidine-2-one derivatives as HIV-1 protease enzyme inhibitors	Radial Distribution Function (RDF)	0.824	0.773	0.05
[39]	Development of an in Silico Model of DPPH• Free Radical Scavenging Capacity: Prediction of Antioxidant Activity of Coumarin Type Compounds	van der Waals volume	0.713	0.654	0.06
[40]	Comparative Molecular Field Analysis (CoMFA) of a Series of Selective Adenosine Receptor A2A Antagonists	Electrostatic and steric field	0.970	0.840	0.13
[41]	QSAR and docking studies on xanthone derivatives for anticancer activity targeting DNA topoisomerase II α	Dielectric energy, group count, LogP, shape index basic (order 3), solvent-accessible surface area	0.840	Not detected	/

Some applications of QSAR study in drug design are described in table 1. QSAR study was a predictive tool for investigations anti-diabetic drugs based on sitagliptin as potential antioxidant agents. Hydrophobicity, counts of rotatable bonds, hydrogen bond donor and acceptor atoms, and topological polar surface area were used as descriptors in this research. Based on the established QSAR equations, new

sitagliptin derivatives with possibly improved pharmacological effect as DPP4 inhibitors are proposed to investigate.

Also, by using QSAR study can be predict Antioxidant Activity of Coumarin Type Compounds. In this study the best correlation between activity and structure has shown van der Waals volume that was used as molecular descriptor.

In silico methods can be good predictive tool for evaluation inhibitory activity of molecules. In these investigations, QSAR studies often combine with other methods as docking studies and neural network.
For predict 3,5-dipyridyl-1,2,4-triazole derivatives as xanthine oxidoreductase inhibitors, QSAR study was used. The results suggested that the steric, electrostatic, and hydrophobic fields played an important role in the models.

A QSAR study was performed on a series of 1,3-dioxo-4-methyl-2,3-dihydro-1H-pyrrolo[3,4-c] quinolones in pursuit of better caspase-3 inhibitors. The study reveals that when increasing the conformational minimum energy while decreasing the lowest unoccupied molecular orbital energy (LUMO) and highest occupied molecular orbital energy (HOMO), the biological activity can be increased. On the basis of a selected QSAR model, a new series of 1,3-dioxo-4-methyl-2,3-dihydro-1H-pyrrolo[3,4-c] quinolines compounds, calculated their caspases inhibitory activity and found that the designed compounds were more potent than the existing compounds.

QSAR model was carried out to predict HIV-1 protease receptors inhibitors activity. In this study Radial Distribution Function (RDF) was used as molecular descriptor that has shown the best correlation with HIV-1 protease inhibition. The QSAR model also indicates that the descriptors (RDF010u, RDF010m, TPSA (NO), F04[C–N]) play an important role in enzyme binding.

The CoMFA approach to studies of 3D-QSAR for series of compounds has proven to be a valuable technique for building predictive model. In this study electrostatic and steric field were used as descriptors.

A QSAR model was developed to explore the anticancer compounds from xanthone derivatives by the multiple linear regression method. High activity descriptors relationship accuracy are obtained referred by regression coefficient and a high activity prediction accuracy. Molecular descriptors: dielectric energy,

group count (hydroxyl), LogP (the logarithm of the partition coefficient between n-octanol and water), shape index basic, and the solvent-accessible surface area were found to correlate with anticancer activity.

CONCLUSION

All of the described papers used QSAR studies to investigate pharmacological activity or binding mode on individual receptors. The descriptors with the highest correlation in this study provide information on crucial functional groups in the structures of examined substances. According to this, we can boost the pharmacological activity or physicochemical qualities of medications by modifying specific groups in their structure.

In general, experimental determinations are quite expensive, and QSPR investigations allow for a cost decrease. It is primarily used to investigate biological activities with diverse qualities connected with structures, which aids in explaining how structural aspects of a pharmacological molecule impact biological activities. QSPR/QSAR approaches can be used to create models that anticipate organic compound characteristics or activities. However, for accurate model building, an appropriate approach to encode the structures with estimated molecular structure descriptors is necessary. The descriptors used in model creation might give a chance to focus on specific traits that account for the compound's property or activity of interest. QSAR should not be used in place of experimental results, although it is a valuable forecasting tool that may be beneficial if no data were available.

REFERENCES

1. Sahigara F, Mansouri K, Ballabio D, Mauri A, Consonni V, Todeschini
2. R. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. Molecules. 2012; 17, 4791-4810.
3. Badnjevic A, Gurbeta L. Development and Perspectives of Biomedical Engineering in South East European Countries. IEEE 39th International convention on information and communication technology, electronics and microelectronics (MIPRO), 30. May to 03. June 2016. Opatija, Croatia
4. Kapetanovic IM. Drug Discovery and Development - Present and

Future. InTech. 2016; DOI: 10.5772/1179.

5. Kubinyi, H., Hamprecht, FA.; Mietzner, T. Three-dimensional quantitative similarity activity relationships (3D QSiAR) from SEAL similarity. J. Med. Chem. 1998; 41, 2553-2564
6. Badnjevic A, Gurbeta L, Boskovic D, Dzemic Z. Medical devices in legal metrology. IEEE 4th Mediterranean Conference on Embedded Computing (MECO). pp: 365-367, 14 – 18 June 2015, Budva, Monténégro
7. Badnjevic A, Beganovic E, Music O. Facts about solution based and cartridge based devices for blood gas analyses. IEEE 18th International Conference on System, Signals and Image Processing. pp:1-5, 16-18 June 2011, Sarajevo, Bosnia and Herzegovina.
8. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR Models - Strategies and Importance. International Journal of Drug Design and Discovery. 2011; 2(3),
9. Doichinova IA, Natcheva RN, Mihailova DN. QSAR studies of 8-substituted xanthines as adenosine receptor antagonists, European Journal of Medicinal Chemistry. 1994; 29(2), 133-138, ISSN 0223-5234
10. Kubinyi H. QSAR: Hansch Analysis and Related Approaches. In Methods and Principles in Medicinal Chemistry; Mannhold R, Kroogsgard-Larsen P, Timmerman H, Eds.; Wiley-VCH: Weinheim, Germany, 1993; 1, 240.
11. Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in Drug Design - A Review. Current Topics in Medicinal Chemistry. 2010; 10, 95-115.
12. Badnjevic A, Gurbeta L, Boskovic D, Dzemic Z. Measurement in medicine – Past, present, future. Folia Medica Facultatis Medicinae Universitatis Saraeviensis Journal, 2015; 50(1): 43-46
13. Santos-Filho OA, Hopfinger AJ, Cherkasov A, de Alencastro RB. The receptor-dependent QSAR paradigm: an overview of the current state of the art. Med. Chem. (Shāriqah (United Arab Emirates)) 2009; 5, 359– 366.
14. Boskovic D, Badnjevic A. Opportunities and Challenges in Biomedical Engineering Education

for Growing Economies. IEEE 4th Mediterranean Conference on Embedded Computing (MECO), pp: 407- 410, 14 – 18 June 2015, Budva, Monténégro

15. Van de Waterbeemd H, Testa B, Folkers G. Computer-Assisted Lead Finding and Optimization. VHChA and VCH, Basel, Weinheim. 1997; 9-28.
16. Kiralj R, Ferreira MMC. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. Journal of the Brazilian Chemical Society. 2009; 20, ISSN 1678-4790.
17. Todeschini R, Gramatica P. 3D QSAR in Drug Design. H. Kubinyi, G. Folkers, Y. C. Martin (Eds.), Kluwer/ESCOM, Dordrecht (The Netherlands). 1998; 2, 355–380.
18. Badnjevic A, Beganovic E, Gvozdenovic V, Sehic G. Automated Closed Loop Controller of Insipred Oxygen System for Improved Mechanical Ventilation in Newborns. IEEE 34th International convention on information and communication technology, electronics and microelectronics (MIPRO), pp: 145-149, 23.-27. May 2011. Opatija, Croatia
19. Kapetanovic IM. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chem-Biol. Interact. 2008; 171(2), 165-176.
20. Aparoy P, Reddy K, Reddanna P. Structure and Ligand Based Drug Design Strategies in the Development of Novel 5-LOX Inhibitors, Current Medicinal Chemistry. 2012; 19(19), ISSN 3763-3778.
21. Pathan S, Ali SM, Shrivastava M. Quantitative structure activity relationship and drug design: A Review. Int. J. Res. Biosciences. 2016; 5(4), 1-5.
22. Eleni P, Dimitra HL. Review in quantitative structure activity relationships on lipoxygenase inhibitors. Mini Rev. Med. Chem. 2003; 3(5), 487-499.
23. Quantitative Structure Activity Relationship (QSAR). VLife Sciences Technologies Pvt. Ltd. Pride Purple Coronet, 1st floor, S No. 287, Baner Road, Pune 411 045, INDIA.
24. Agrawal VK, Dubey VK, Shaik B, Singh J, Singh K, Khadikar PV.

Modeling of Lipophilicity of Some Organic Compounds Using Structural and Topological Indices. J. Indian Chem. 2009; Soc., 86, 1-9.

25. Labute PA. Widely applicable set of descriptors. J Mol Graph Model. 2000; 18, 464-77.
26. Wold S. Cross-validation estimation of the number of components in factor and principal components models. Technometrics, 1978; 20, 397–405.511-519