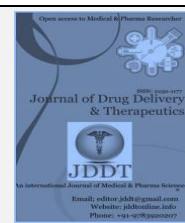
Available online on 15.07.2020 at <http://jddtonline.info>

Journal of Drug Delivery and Therapeutics

Open Access to Pharmaceutical and Medical Research

© 2011-20, publisher and licensee JDDT, This is an Open Access article which permits unrestricted non-commercial use, provided the original work is properly cited



Open Access

Review Article

Virtual Screening, Molecular Docking and QSAR Studies in Drug Discovery and Development Programme

Mithun Rudrapal*, Dipak Chetia

Department of Pharmaceutical Sciences, Dibrugarh University, Dibrugarh, Assam, India

ABSTRACT

Structure-based drug design (SBDD) and ligand-based drug design (LBDD) are the two basic approaches of computer-aided drug design (CADD) used in modern drug discovery and development programme. Virtual screening (or *in silico* screening) has been used in drug discovery program as a complementary tool to high throughput screening (HTS) to identify bioactive compounds. It is a preliminary tool of CADD that has gained considerable interest in the pharmaceutical research as a productive and cost-effective technology in search for novel molecules of medicinal interest. Docking is also used for virtual screening of new ligands on the basis of biological structures for identification of hits and generation of leads or optimization (potency/ property) of leads in drug discovery program. Hence, docking is approach of SBDD which plays an important role in rational designing of new drug molecules. Quantitative structure-activity relationship (QSAR) is an important chemometric tool in computational drug design. It is a common practice of LBDD. The study of QSAR gives information related to structural features and/or physicochemical properties of structurally similar molecules to their biological activity. In this paper, a comprehensive review on several computational tools of SBDD and LBDD such as virtual screening, molecular docking and QSAR methods of and their applications in the drug discovery and development programme have been summarized.

Keywords: Virtual screening, Molecular docking, QSAR, Drug discovery, Lead molecule

Article Info: Received 29 April 2020; Review Completed 16 June 2020; Accepted 27 June 2020; Available online 15 July 2020



Cite this article as:

Rudrapal M, Chetia D, Virtual Screening, Molecular Docking and QSAR Studies in Drug Discovery and Development Programme, Journal of Drug Delivery and Therapeutics. 2020; 10(4):225-233
<http://dx.doi.org/10.22270/jddt.v10i4.4218>

*Address for Correspondence:

Mithun Rudrapal, Department of Pharmaceutical Sciences, Dibrugarh University, Dibrugarh, Assam, India

INTRODUCTION

Drug-like properties are defined as molecule's intrinsic properties that remain an integral element in drug discovery research. The physicochemical properties of a drug molecule have an important impact on its pharmacokinetic (PK), metabolic fate and toxicity in the body, and so a good understanding of these properties, coupled with their prediction, are crucial for a successful drug discovery programme¹. Virtual screening is a predictive *in silico* high throughput screening (HTS) tool used in computer-aided drug design (CADD) which evaluates the drug-likeness (including ADME-Tox and toxicity properties) of ligands/ molecules that have potential problems related to PK and toxicities, independent of their intended drug target. Physicochemical properties that affect the PK and toxicity of drug molecules are required to be constantly re-evaluated (or optimized) in order to bring success in drug discovery by enhancing speed, efficiency and quality of research²⁻⁴. Inefficient research, attrition, and costs are reduced if compounds have good drug-like properties. The assessment of drug-likeness, ADME-Tox properties and toxicity

parameters are, therefore, important aspects in the development of good clinical candidates that have a balance of activity and molecular properties. A drug lead molecule with desired drug-like properties can be transformed into quality drug candidate^{5,6}. Further, Good activity and drug-like properties are complementary, and both are necessary even for a good drug product. A less potent compound with better properties may produce a better therapeutic response and thus be a better drug candidate⁷. Molecular docking is a computational tool used extensively in Structure-Based Drug Design (SBDD) for the determination of binding affinity and relative orientation between a protein and a ligand when they are bound to each other. This technique simulates the process of molecular recognition by predicting the binding free energy as well as interaction geometry of a bound protein-ligand complex. Docking thus predicts the preferred orientation (i.e., 'best fit' orientation in 3D space) of a ligand that binds to a particular protein of interest. Structure-Activity Relationship (SAR) is the study of relationship between the chemical or structural properties of a molecule and its biological activity. The analysis of SAR determines the chemical groups responsible for evoking the biological

response a drug molecule. This is a common practice of medicinal chemistry which allows modification of the effect (i.e., activity) of a bioactive compound (or drug molecule) by changing its chemical structure.

VIRTUAL SCREENING AND DRUG-LIKENESS

Drug-likeness

Virtual screening (or *in silico* screening) has been used in drug discovery program as a complementary tool to high throughput screening (HTS) to identify bioactive compounds. It is a new approach of Computer-Aided Drug Design (CADD) that has gained considerable interest in the pharmaceutical research as a productive and cost-effective technology in the search for novel molecules of medicinal interest. Apart from physical HTS, virtual screening refers to a range of *in silico* techniques that perform computational screening of large libraries of chemical compounds (chemical databases/compound collections) to identify a smaller number of possible bioactive compounds in order to prioritize them for synthesis or biological assay. The virtual screening is done mainly by two methods, one is on the basis of known biological structures (structure-based screening or docking) or the other is by means of molecular similarity analyses using active compounds as templates (ligand-based screening). However, the purpose of virtual screening is, therefore, the reduction of a huge virtual chemical space (of small organic molecules), to synthesize and/or test against a specific target protein, to a relatively smaller number of compounds that have a good chance to lead to a drug candidate. Virtual screening of compound libraries has

recently gained significant interest to early identification of drug leads in drug discovery programs.

Many drug candidates fail at the pre-clinical and clinical stages of development due to their lack of efficacy and/or due to the reasons unrelated to the efficacy against intended drug target. Failure of drug molecules due to the lack of optimal drug-like properties is dominating. Pharmacokinetic (PK) and toxicity issues are accounted for more than half of the failure in the clinical trials. It results in increased time and costs per new drug product development. Therefore, preliminary study of virtual screening evaluates the drug-likeness (including ADME-Tox and toxicity properties) of the ligands/molecules that have potential problems related to PK and toxicities, independent of their intended drug target. In later stage, screening of drug-like molecules is done by molecular docking method based on the structure of intended drug target. Molecules (ligands) predicted to bind well with the complement target molecule are selected for the further experimental study¹⁻⁴.

Drug-likeness screening predicts the likelihood that a molecule has desired drug-like characteristics and physicochemical properties, called drug-like properties. Drug-like properties are defined as molecule's intrinsic properties that remain an integral element in drug discovery research. The physicochemical properties of a drug molecule have an important impact on its PK, metabolic fate and toxicity in the body, and so a good understanding of these properties, coupled with their prediction, are crucial for a successful drug discovery programme^{5,6}. A list of the drug-like properties of interest is presented in Table 1.

Table 1: Drug-like properties⁶

Structural	Physicochemical	Biochemical	PK and Toxicity
Lipophilicity (LogP)	Solubility	Metabolism (Phases I and II)	Bioavailability
Dissociation constant (pKa)	Permeability	Protein and tissue binding	Half-life
Molecular weight (MW)	Chemical stability	Transport (uptake, efflux)	Clearance
H-bonding		Target affinity	Drug-drug interaction
Polar surface area (PSA)			LD ₅₀

Drug-likeness screening basically performs calculation of fundamental molecular properties and Lipinski's parameters along with other physicochemical properties, which assesses the acceptability of compounds as drug-like molecules. They represent the combined physicochemical, PK and pharmacodynamic (PD) properties (collectively termed as drug-like properties) of molecules to exhibit good drug-likeness behavior in human body. In fact, molecular properties are the fundamental structural parameters that determine the physicochemical and biochemical properties, which ultimately determine molecule's pharmacokinetics, toxicity and pharmacodynamics (affinity and efficacy) properties^{7,8} (Figure 1).

According to Lipinski's rule of five (1997), compounds are more likely to be drug-like and orally bioavailable if they obey the following criteria: LogP_{o/w} (octanol/water partition coefficient) ≤ 5, MW (molecular weight) ≤ 500, HBAs (hydrogen bond acceptors) ≤ 10 and HBD (hydrogen bond donors) ≤ 5. To further substantiate Verber et al. stated that compounds with ≤ 10 RotB (rotatable bonds) and TPSA (total polar surface area) of ≤ 140 Å² are more likely to show membrane permeability and good bioavailability.

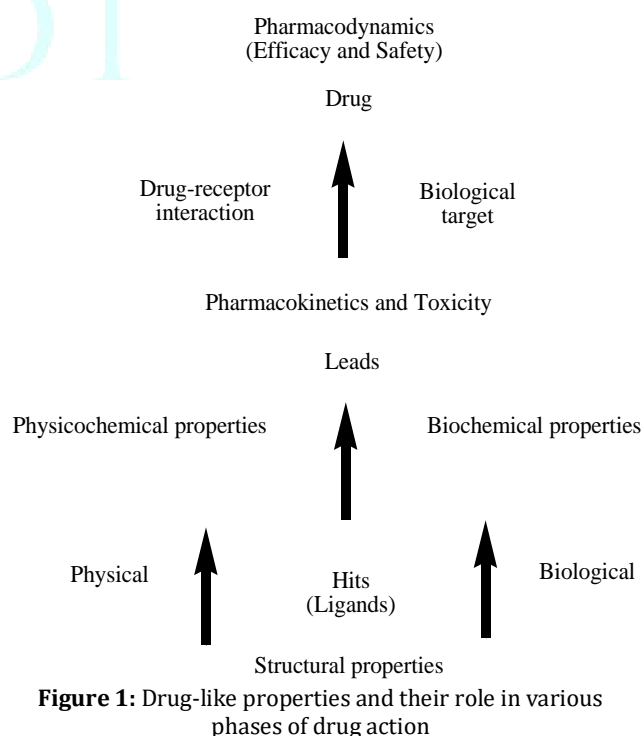


Figure 1: Drug-like properties and their role in various phases of drug action

Poor absorption or permeation is more likely when the MW is over 500, the calculated logP (clogP) is higher than 5, when there are more than 10 HBAs (expressed as sum of Ns and Os), more than 5 HBDs (expressed as sum of OHs and NHs) and greater than 140 Å² molecular TPSA (due to polar atoms, N and O plus attached hydrogens)⁷⁻⁹. Typical ranges of properties related to drug-likeness are depicted in Table 2. Lipinski's rule of five is considered predictive guideline for oral bioavailability; however, 16% of oral drugs violate at least one of the criteria and 6% fail in two or more. About 70% of drug-like molecules have between 2 and 9 HBAs, 0 and 2 HBDs and between 2 and 8 RotB. Lipinski's rule is therefore used as a filter to decrease the library size before docking or to eliminate some ligands before synthesis in the design of new drug molecules or optimization of drug leads^{9,10}.

Table 2: Typical ranges of properties related to drug-likeness⁹

Parameter	Minimum	Maximum
LogP	-2	5
MW	200	500
HBAs	0	10
HBDs	0	5
TPSA	0	140
RotB	0	8

ADME-Tox prediction

ADME-Tox (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties have a predictable influence on PK, metabolic fate and PD effects of drug molecules. A majority of candidate drugs fail at the pre-clinical stage of drug

development which increases the costs of per new drug development. It has been reported that about 63% of the candidate drugs fail due to their poor or unsatisfactory ADME-Tox properties. The calculation of ADME-Tox properties is therefore intended as the first step towards optimizing new drug molecules, whether to check the failure of lead molecules which may cause toxicity and are unable to cross the intestinal membranes or metabolized by the body into an inactive form. The prediction of ADME-Tox properties reduce failure rate of compounds in late stage of drug discovery. It focuses on predictions of chemical modifications of compounds that will likely improve PK and PD properties.

The optimal approach for performing the ADME-Tox prediction is one that uses both *in vitro* and *in silico* studies in a complementary way ensuring that ADME-Tox is considered at almost every stage of the drug discovery process, from hit identification to lead optimization. In the hit identification stage, the goal of the *in silico* ADME-Tox prediction is to identify compounds with acceptable drug-like properties. In the lead identification stage, the objective is to identify a small number of compounds with the activity, selectivity and drug-like properties required for a potential drug candidate. *In vitro* assays are used to measure the ADME-Tox properties of the newly synthesized compounds. However, this information is valuable for the refinement of the *in silico* ADME-Tox models (Table 3). There are two key aims of ADME-Tox prediction: first, to design new compounds and compound libraries so as to reduce the risk of attrition at late stage; and second, to optimize the screening and testing prioritizing only the most promising compounds¹¹⁻¹³. The ADMET properties that can be evaluated using *in silico* models are intestinal permeability, aqueous solubility, lipophilicity, human oral bioavailability, blood-brain barrier permeation, plasma protein binding, metabolic stability, interactions with cytochrome P450s and hepatotoxicity.

Table 3: Prediction of ADME-Tox properties at various stages of drug discovery³

Hit identification	Lead identification	Lead optimization
Identify compounds or series with acceptable drug-like properties	Prediction of chemical modification that improves the ADME-Tox properties	Similar to the process of lead identification Refinement of compounds using experimental data

The prediction of ADMET properties estimates the following requirements for a drug that

- must bind tightly to the biological target *in vivo*
- must pass through one or more physiological barriers (cell membrane or blood-brain barrier)
- must remain long enough to produce the effect
- must be removed from the body by metabolism, excretion, or other means^{13,14}

Toxicity prediction

The prediction of toxicity parameters are used for early identification of potential problems related to toxicity of drug-like molecules. *In silico* toxicity parameters include

rodent toxicity, rodent carcinogenicity, skin sensitivity, Ames mutagenicity and so on^{9,15}.

Significance of drug-likeness screening

The structural and physicochemical properties that affect the PK and toxicity of drug molecules are required to be constantly reevaluated (or optimized) in order to bring success in drug discovery by enhancing speed, efficiency, and quality of research. Inefficient research, attrition, and costs are reduced if compounds have good drug-like properties. The assessment of drug-likeness, ADME-Tox properties and toxicity parameters are therefore important aspects in the development of good clinical candidates that have a balance of activity and molecular properties. A drug lead molecule with desired drug-like properties can be transformed into quality drug candidate (Figure 2)

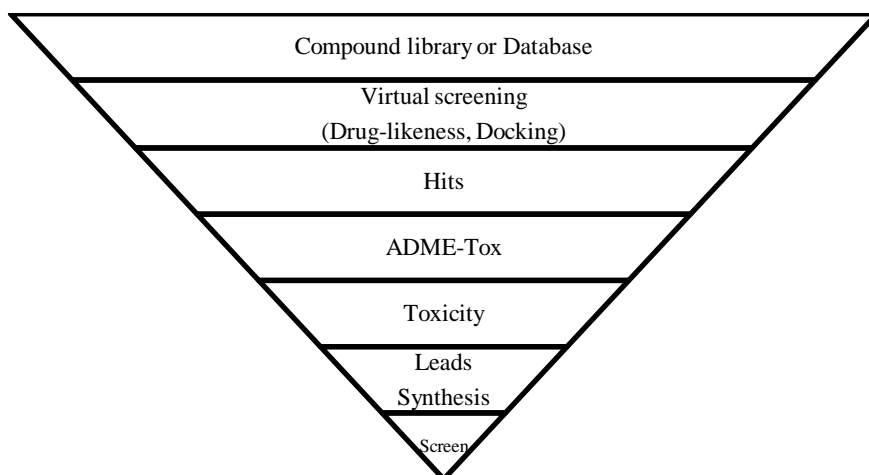


Figure 2: Pyramid of virtual screening

Thus, in virtual screening, lead identification and lead optimization consist of an iterative process, starting from *in silico* prediction, chemical synthesis to experimental testing and confirmation^{6,16}.

Good activity and drug-like properties are complementary, and both are necessary even for a good drug product. The most active or selective compound may not make the best drug product because of property limitations that cause poor PK or safety profile. A less potent compound with better properties may produce a better therapeutic response and thus be a better drug candidate (Figure 3)¹⁷⁻²¹.

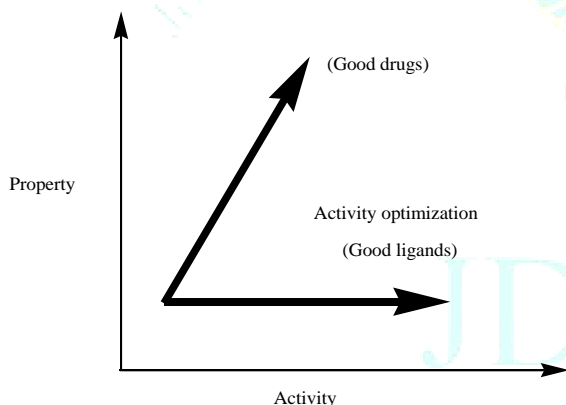


Figure 3: Property-activity relationship

MOLECULAR DOCKING

Molecular docking is a computational tool used extensively in Structure-Based Drug Design (SBDD) for the determination of binding affinity and relative orientation between a protein and a ligand when they are bound to each other. This technique simulates the process of molecular recognition by predicting the binding free energy as well as interaction geometry of a bound protein-ligand complex. Docking thus predicts the preferred orientation (i.e., 'best fit' orientation in 3D space) of a ligand that binds to a particular protein of interest. The aim of molecular docking is to achieve an energetically favorable conformation for both the protein and ligand such that the ligand molecule undergoes optimal interaction with the binding site of protein (or receptor) molecule to obtain a stable protein-ligand complex^{22,23}.

Docking components

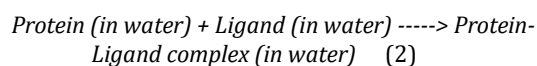
The molecular docking is divided into two main components, viz., search algorithm and scoring function.

Search algorithm determines all possible optimal conformations for a given protein-ligand complex i.e., the position and orientation of both molecules relative to each other. They also calculate the energy of the resulting complex and of each individual interaction. More clearly, a search algorithm generates 'poses' i.e., orientations of particular conformations of the ligand in the active (binding) site of protein molecule. It is followed by finding the best conformation of the ligand appropriate for optimal interaction with the protein structure which is measured by scoring function. The different types of algorithms that can be used for docking searches are molecular dynamics, Monte Carlo methods, genetic algorithms and fragment-based methods²²⁻²⁵.

Scoring function is a mathematical method used to predict the strength of intermolecular interaction (non-covalent) between protein and ligand after they have been docked. The prediction of the binding free energy (binding affinity) is called scoring. A molecule with a good docking score is potentially a good binder. The scoring function assigns a priority order to a set of structurally diverse ligands docked to the same protein and estimates binding (or interaction) energy. The ligands are ranked according to the interaction energy based on scoring functions. There are different types of scoring functions, namely, force field-based, empirical-based, knowledge-based, and consensus scoring. More accurate a scoring functions, more expensive it is. Most scoring functions are physics-based molecular mechanics force fields that estimate the energy of binding interaction of ligand molecules. The empirical scoring function of any docking program is expressed as follows (Equation 1):

$$\text{Fitness} = \text{Elec} + \text{vdW} + \text{H-bond} \quad (1)$$

The binding energy refers to that change in free energy of the system as depicted below (Equation 2):



The free energy of binding is calculated as the sum of the electrostatic energy (due to ionic interaction, G_{elec}), van der Waals energy (G_{vdW}), solvation effects (due to H-bonding, G_{sol}) and internal energy changes due to conformational (G_{conf}), rotational (G_{rot}) and torsional (G_{tor}) energies. It is represented as ΔG_{bind} (kcalmol^{-1}) and is calculated as follows (Equation 3):

$$\Delta G_{\text{bind}} = \Delta G_{\text{elec}} + \Delta G_{\text{vdW}} + \Delta G_{\text{sol}} + \Delta G_{\text{conf}} + \Delta G_{\text{rot}} + G_{\text{tor}} \quad (3)$$

The ΔG bind value has to be low for a structure to be stable. A low (negative) binding energy indicates a stable complex and thus a more likely binding interaction^{22,24,25}.

Assessment of docking

The results of docking are analyzed by a statistical scoring function which converts interacting energy into numerical values, called the docking score. The 3D pose (binding mode) of the docked ligand can be visualized using different visualizing tools like Chimera, Pymol, Rasmol etc. which could help in to identify the best fit orientation of ligand considering both docking score (or binding energy) and various non-covalent interactions. Docking methods are usually assessed by their ability to reproduce the binding mode of experimentally resolved protein-ligand complexes. The ligand is removed from the complex, a search area is defined around the actual binding site, the ligand is redocked into the protein, and the binding mode generated is compared with the experimental positions usually in terms of a root mean square deviation (RMSD). RMSD refers to the distance between the initial and final position of the ligand in the binding cavity of protein molecule. RMSD has often been used to measure the quality of reproduction of a known (i.e., crystallographic complex) binding pose by a docking method. Lower the value of RMSD, higher is the accuracy of docking. If the docking protocol is able to produce similar docking pose of a ligand with respect to the biological configuration of the same ligand in the crystal structure of complex protein then it means that the docking method is validated. If the RMSD is below 2 Å, it is generally

considered a successful prediction. This represents good reproduction of the correct pose (or true binding pose)²⁶⁻³⁰.

Docking methodology

Molecular docking studies are usually performed on free like AutoDock (Art Olsen, David Goodsell, Scripps), and commercial software packages such as Discovery Studio (Accelrys), Glide (Schrodinger), GOLD (CCDC) and FlexX (BiosolveIT). The crystal structures of protein molecules of interest are retrieved from the RCSB Protein Data Bank (<https://www.rcsb.org>)^{31, 32}. The methodology involved in molecular docking is depicted in Figure 4.

Applications of docking

Docking has two important roles: the search for the conformation and configuration of the ligand in the binding site, which rank poses for a given small molecule on a given target (docking), and the evaluation of the interaction energy between the target and ligand, which rank different ligands according to their relative affinity for a given target (compounds selection). Molecular docking studies are thus frequently used to predict the binding orientation (or binding mode/pose) of small molecules to their protein targets (or receptor) in order to predict the affinity and activity of the small molecules. A binding interaction between a small molecule ligand and an enzyme protein may result in activation or inhibition of the protein. If the protein is a receptor, ligand binding may result in agonistic or antagonistic effect^{22,24,33-36}.

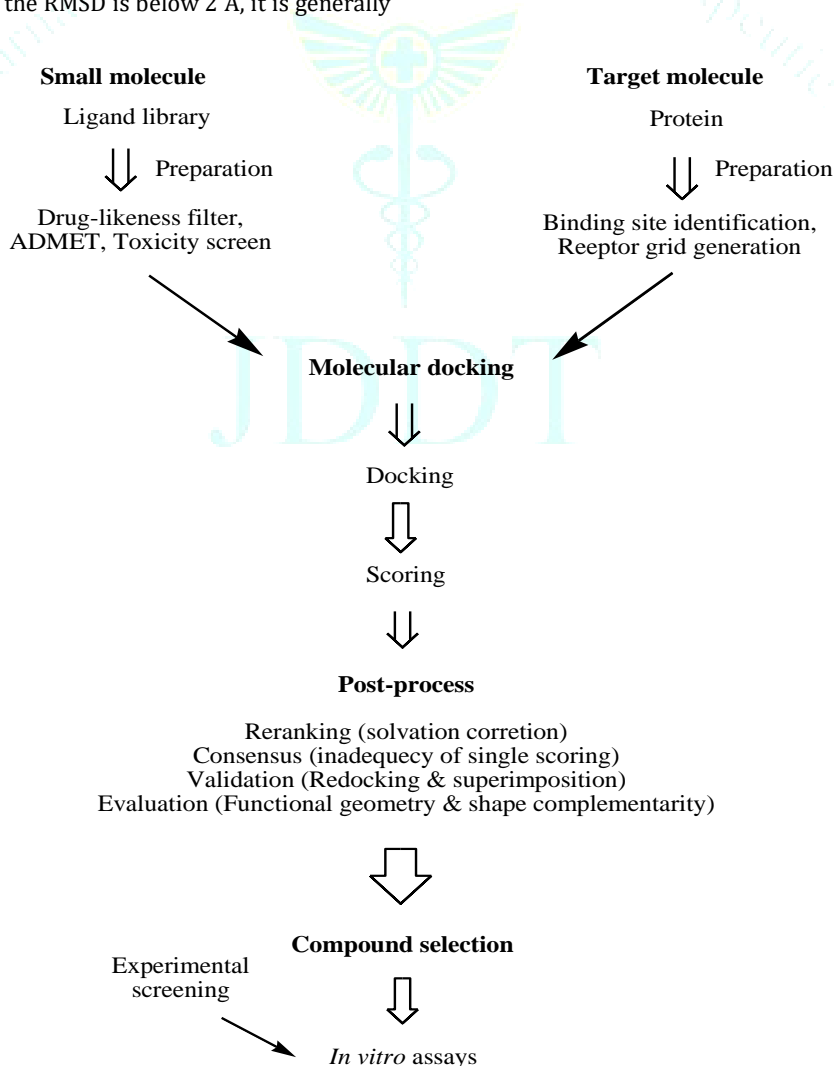


Figure 4: Methodology of molecular docking

Predicting the mode of protein-ligand interaction can assume the active site of the protein or receptor molecule, and further help in protein annotation. Docking is also used for virtual screening of new ligands on the basis of biological structures for identification of hits and generation of leads or optimization (potency/property) of leads in drug discovery program. Hence, docking plays an important role in rational design of new drug molecules³⁷⁻⁴¹.

STRUCTURE-ACTIVITY RELATIONSHIP (SAR) AND QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR) STUDIES

SAR/SPR

Structure-Activity Relationship (SAR) is the study of relationship between the chemical or structural properties of a molecule and its biological activity. The analysis of SAR determines the chemical groups responsible for evoking the biological response a drug molecule. This is a common practice of medicinal chemistry which allows modification of the effect (i.e., activity) of a bioactive compound (or drug molecule) by changing its chemical structure. Medicinal chemists use the techniques of chemical synthesis and computational drug design to insert new chemical groups/change existing groups in the bioactive molecule and test the modifications for their altered (increased or decreased) biological response. The information on SAR is obtained from empirical correlation of structural variation on the common template (or structural scaffold) of a congeneric series of compounds with the experimentally determined biological activity. The basic assumption in SAR hypothesis is that molecules with similar structures have similar activities. This is because of the fact that similar compounds may have similar chemical or physicochemical properties. Thus, there is a relationship between structure and activity of a bioactive molecule, and this is the basis of SAR study⁴².

The technique of SAR uses molecular properties of bioactive compounds to predict biological activity (efficacy, receptor binding, toxicity, and bioavailability). The biological effects of a new chemical compound can often be predicted from its molecular structure using SAR data about other similar compounds. Thus, the ability to draw conclusions about the biological property of an unknown compound mainly depends on the structural features that can be characterized. When appropriate judgment is made, SAR can be a powerful tool for understanding functional implications when structural similarities are found. SAR can be used to develop new molecules and optimization of existing compounds (or lead molecules) by appropriate assessment of biological (therapeutic or toxic) effects^{43,44}.

Structure-Property Relationship (SPR) defines how structure affects properties. This is complementary to SAR. SPR allows medicinal chemists to understand how structural modifications improve properties for their basic structural scaffold. SPR can be developed for a series of structurally similar molecules in the same manner as SARs. SAR defines how structure modification at one moiety in the molecule affects activity, while, SPR defines how structure modification affects properties. Medicinal chemists need to understand the relationship between chemical structure and the physicochemical properties of molecules and how this

translates into compound's stability, transport and the interaction receptor molecule. Interaction of drug molecules with biological targets implicates the PKs (ADME) and toxicities of a compound in biological systems which thereby modulate biological function^{45,46}.

QSAR/QSPR

The method of SAR can be refined to build mathematical relationships between a chemical structure and its biological activity, which is known as Quantitative Structure-Activity Relationship (QSAR). The fundamental of SAR hypothesis is that molecules with similar structures will have similar functions (activities), while QSAR hypothesizes that biological activity (similar activity) of different molecules can be compared quantitatively based on the characteristics of their structural components. This study is often referred as Quantitative Structure-Activity/Structure-Property/-Structure-Toxicity Relationship (QSAR/QSPR/QSTR)⁴⁷⁻⁴⁹.

Principle of QSAR analysis

QSAR is an important chemometric tool in computational drug design. The study of QSAR gives information related to structural features and/or physicochemical properties of structurally similar molecules to their biological activity. It attempts to establish a correlation between the experimental activity of a series of compounds and the structural information of compounds defined by molecular descriptors. QSAR/QSPR models represent mathematical equations correlating the response of bioactive compounds (activity/property) with their structural and physicochemical properties. Similar molecules with minor variation in their structures can have quite improved biological activities (Figure 5)^{50,51}.

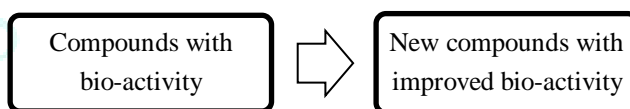


Figure 5: Basis of QSAR study

This kind of relationship between molecular structure and changes in biological activity is the center of focus of QSAR. QSAR methods are based upon two basic assumptions, first is to derive a quantitative measure from the structural properties significant to the biological activity of a compound, and the other assumption is to mathematically describe the relationship between biological activity to be optimized and the molecular property calculated from the structure. For a QSAR to be valid and reliable, the activity of all of the compounds of a congeneric series covered must be elicited by a common biochemical mechanism⁵².

Components of QSAR analysis

The steps involved in QSAR analysis are data set preparation, structural optimization, calculation and selection of molecular descriptor, correlation model development, and finally model evaluation and validation (Figure 6). Good QSAR modeling depends on the correct analysis and selection of computed descriptors as independent variables against biological activity as dependant variable in QSAR equations⁵³.

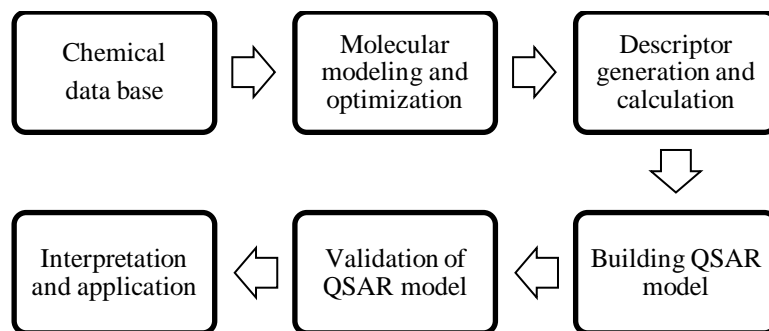


Figure 6: Steps involved in QSAR analysis

Molecular descriptors are numerical representations of chemical information encoded within a molecular structure via mathematical procedure. They include constitutional, hydrophobic, electronic, steric, geometrical, topological and quantum chemical (semi-empirical) properties of molecules.

Descriptors are used to study the quantitative effects of the molecular structure of bioactive compounds on their biological activities including inhibitory activities⁵⁴ (Figure 7).

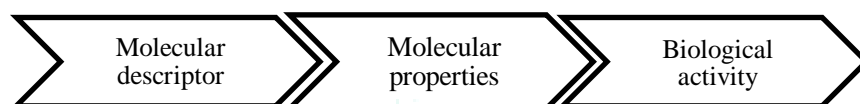


Figure 7: Relationship between descriptors and biological activity

2D-QSAR

QSAR models are generally developed with a database of specified chemical domain and easily calculable 2D molecular descriptors using simple regression-based statistical methods such as Multiple Linear Regression (MLR) and Partial Least Squares (PLS). Such quantitative

relationship between the biological activity of a set of compounds and the 2D molecular/structural descriptors of compounds is known as 2D-QSAR analysis⁵⁵. Various methods of 2D-QSAR analysis are depicted in Table 4.

Table 4: Methods of 2D-QSAR^{55,56}

Free energy models	Mathematical models	Statistical methods	Other methods
Hansch analysis (Linear Free Energy Relationship)	Free Wilson analysis Fujita-Ban modification	Linear Regression Analysis -Multiple Linear Regression (MLR) -Partial Least Squares (PLS) -Discriminant Analysis (DA) -Principle Component Analysis (PCA) -Principle Component Regression (PCR) -Genetic Function Approximation (GFA) -Genetic Partial Least Squares (G/PLS)	Pattern recognition -Cluster Analysis (CA) -Artificial Neural Network (ANN) - <i>k</i> -Nearest Neighbor (<i>k</i> NN) Topological methods Quantum mechanical methods

The 2D-QSAR analysis describes the activity (response property) of a data set and predicts (and/or classifies) activities for further sets of untested compounds. After its development, a QSAR model is usually verified by employing multiple statistical validation tools giving an estimation of its statistical quality and predictivity. According to the OECD guidelines, the development of a QSAR model should comply with unambiguous algorithm strategies and the model should pass various testes model fitness, robustness and predictivity. A reliable QSAR model once established and validated can be useful to predict the activities of molecules, and to know which structural features play an important role in biological processes^{57,58}.

Genetic Function Approximation (GFA) algorithm is a new artificial intelligence-based method used to build 2D-QSAR/QSPR/QSTR models. In GFA, the QSAR model is obtained by combining both genetic algorithm and statistical modelling. GFA is genetic based method which combines Holland's genetic algorithm and Friedman's Multivariate Adaptive Regression Spines (MARS). Application of GFA algorithm may construct higher quality predictive models and make available additional information not provided by standard regression techniques. The GA algorithm can handle a huge number of descriptors and generate models competitive with or superior to the results of standard regression analysis. GFA followed by MLR (GFA-MLR) or

PLR (G/PLS) regression is frequently used in research for the development of predictive QSAR/QSPR models^{55,58,59,60}.

3D-QSAR

In recent days, 3D-QSAR have been applied which involve the analysis of the quantitative relationship between the biological activity of a set of compounds and their three-

dimensional properties using statistical correlation methods. 3D-QSAR uses probe-based sampling within a molecular lattice to determine three-dimensional properties of molecules (particularly steric and electrostatic values) and can then correlate these 3D descriptors with biological activity^{60,61}. Various 3D-QSAR methods in use (on the basis of chemometric techniques used) are summarized in Table 5.

Table 5: Methods of 3D-QSAR⁶¹

Linear 3D-QSAR	Non-linear 3D-QSAR
-Comparative Molecular Field Analysis (CoMFA) -Comparative Molecular Similarity Indices Analysis (COMSIA) -Comparative Molecular Movement Analysis (COMMA) -Adaptation of Fields for Molecular Comparison (AFMoC) -Genetically Evolved Receptor Models (GERM) -Self-Organizing Molecular Field Analysis (SoMFA)	-COMPASS -COMFA-PLS (QPLS)

QSAR models help to design new molecules and also optimization of lead molecules with improved activity profile. QSAR is also an important tool in virtual screening of bioactive compounds used to eliminate (compounds lacking drug-like properties or compounds predicted to elicit a toxic response) from further development in early phase of drug discovery and development. It also acts as an informative tool by extracting significant patterns in descriptors related to the measured biological activity for understanding of mechanisms (drug-receptor interactions) of given biological response. QSAR has therefore gained significant momentum in ligand-based, receptor-based, or other methods of drug design^{62,63}.

CONCLUSION

In this review, various computational tools of structure-based/ ligand-based drug design (SBDD/LBDD) including virtual screening, molecular docking and QSAR have been summarized. Some advanced tools of drug design strategy that are widely used in modern drug discovery programme include pharmacophore modelling, fragment-based drug design and structure similarity search and molecular dynamics simulation. These computational tools are successfully used in the drug discovery and development programme along with bioinformatics and cheminformatics tools. In recent days, artificial intelligent technology has also been evolved which can successfully be used along with omics tools (such as genomics, proteomics, metabolomics etc.) and computational biology techniques in modern drug discovery and development programme.

CONFLICT OF INTEREST

No conflict of interest exists

REFERENCES

- Murumkar PR, Sharma MK, Giridhar R, Yadav MR. Virtual screening-based identification of lead molecules as selective TACE inhibitors. *Med Chem Res.* 2015; 24:226-244.
- Ooms F. Molecular Modeling and Computer Aided Drug Design. Examples of their Applications in Medicinal Chemistry. *Curr Med Chem.* 2000; 7:141-158.
- Shaikh SA, Jain G, Sandhu LN, Jayram B. From Drug targets to Lead Sketching A Physicochemical Pathway for Lead Molecule Design In Silico. *Curr Pharm Des.* 2017; 13:3454-3470.
- Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr Top Med Chem.* 2014; 14:1923-1938.
- Singh J, Kumar M, Mansuri R, et al. Inhibitor designing, virtual screening, and docking studies for methyltransferase: A potential target against dengue virus. *J Pharm Bioallied Sci.* 2016; 8(3):188-194.
- Kerns EH, Di L. *Drug-like Properties: Concepts, Structure Design and Methods from ADME to Toxicity Optimization.* New York: Academic Press, 2008:6-14.
- Lipinski CA, Lombardo F, Dominy BW, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 1997; 23:3-26.
- Shukla S, Srivastava RS, Shrivastava SK, et al. Synthesis, cytotoxic evaluation, docking and in silico pharmacokinetic prediction of 4-arylideneamino/cycloalkylideneamino 1,2-naphthoquinone thiosemicarbazones. *J Enzyme Inhib Med Chem.* 2013; 28(6):1192-1198.
- Alam S, Khan F. QSAR and docking studies on xanthone derivatives for anticancer activity targeting DNA topoisomerase II α . *Drug Des Dev Ther.* 2014; 8:183-195.
- Anderson AC, Wright DL. The Design and Docking of Virtual Compound Libraries to Structures of Drug Targets. *Curr Comput Aided Drug Des.* 2005; 1:103-127.
- Waterbeemd HVD, Gifford E. ADMET in silico modelling: Towards prediction paradise? *Nat Rev Drug Discov.* 2003; 2:192-204.
- Nuez ADL, Rodríguez R. Current methodology for the assessment of ADME-Tox properties on drug candidate molecules. *Biotechnol Apl.* 2008; 25:97-110.
- Gleeson MP, Hersey A, Hannongbua S. In silico ADME models: a general assessment of their utility in drug discovery applications. *Curr Top Med Chem.* 2011; 11:358-381.
- Ganguly S, Debnath B. Molecular Docking Studies and ADME Prediction of Novel Isatin Analogs with Potent Anti-EGFR Activity. *Med Chem.* 2014; 4:558-568.
- Roncaglioni A, Toropov AA, Toropova AP, Benfenati E. In silico methods to predict drug toxicity. *Curr Opin Pharmacol.* 2013; 13:802-806.
- Borchardt RT, Kerns EH, Hageman MJ, Thakker DR, Stevens JL. Optimizing the 'Drug-Like' Properties of Leads in Drug Discovery. In: Borchardt RT, Middaugh CR, editors. *Biotechnology: Pharmaceutical Aspects.* New York: Springer; 2006; 171-189.
- Sharma D, Chetia D, Rudrapal M. Design, Synthesis and Antimalarial Activity of Some New 2-Hydroxy-1,4-naphthoquinone-4-hydroxyaniline Hybrid Mannich Bases. *Asian J Chem.* 2016; 28(4):782-788.
- Kashyap A, Chetia D, Rudrapal M. Synthesis, Antimalarial Activity Evaluation and Drug-likeness Study of Some New Quinoline-Lawsone Hybrids. *Indian J Pharm Sci.* 2016; 78(6):892-911.
- Patowary P, Chetia D, Kalita J, Rudrapal M. Design, Synthesis and Antimalarial Activity of Flavonoid Derivatives. *Indian J Heterocycl Chem.* 2019; 29(1):53-58.
- Medhi A, Chetia D, Rudrapal M. Synthesis and Antimalarial Activity of Lawsone Mannich Base Derivatives. *Indian J Pharm Edu Res.* 2018; 52(3):472-479.

21. Gogoi J, Chetia D, Kumawat MK, Rudrapal M. Synthesis and Antimalarial Activity Evaluation of Some Mannich Bases of Tetraoxane-Phenol Conjugate. *Indian J Pharm Edu Res.* 2016; 50(4):591-597.
22. Kashid AM, Dube PN, Alkutkar PG, et al. Synthesis, biological screening and ADME prediction of benzylindole derivatives as novel anti-HIV-1, anti-fungal and anti-bacterial agents. *Med Chem Res.* 2013; 22:4633-4640.
23. Singh S, Srivastava P. Molecular Docking Studies of Myricetin and its Analogues against Human PDK-1 Kinase As Candidate Drugs for Cancer. *Comput Mol Biosci.* 2015; 5:20-33.
24. Segall, MD, Barber C. Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discov Today.* 2014; 19:688-693.
25. Tang Y, Zhu W, Chen K, et al. New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. *Drug Discov Today Technol.* 2006; 3:308-313.
26. Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr Top Med Chem.* 2014; 14:1923-1938.
27. Kalita J, Chetia D, Rudrapal M. Design, Synthesis, Antimalarial Activity and Docking Study of 7-Chloro-4-(2-(substituted benzylidene)hydrazineyl)quinolines. *Med Chem.* 2020; in press. DOI: 10.2174/1573406415666190806154722.
28. Kalita J, Chetia D, Rudrapal M. Molecular Docking, Drug-likeness Studies and ADMET Prediction of Quinoline Imines for Antimalarial Activity. *J Med Chem Drug Des.* 2019; 2(1):1-7.
29. Rudrapal M, Banu ZW, Chetia D. Newer series of trioxane derivatives as potent antimalarial agents. *Med Chem Res.* 2018;27(2):653-668.
30. Rudrapal M, Chetia D, Singh V. Novel series of 1,2,4-trioxane derivatives as antimalarial agents. *J Enzyme Inhib Med Chem.* 2017; 32(1):1159-1173.
31. Sharma R, Goswami A, Rudrapal M, Sharma D, Sharma HK, Chetia D. In vitro Evaluation of the Antimalarial Activity of a Designed Novel Quinuclidine Derivative. *Curr Sci.* 2016; 111(12):2028-2830.
32. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual Screening for drug discovery: Methods and applications. *Nat Rev Drug Discov.* 2004; 3:935-949.
33. Anderson AC, Wright DL. The Design and Docking of Virtual Compound Libraries to Structures of Drug Targets. *Curr Comput Aided Drug Des.* 2005; 1:103-127.
34. Usha T, Middha SK, Goyal AK, et al. Molecular docking studies of anti-cancerous candidates in *Hippophae rhamnoides* and *Hippophae salicifolia*. *The J Biomed Res.* 2014; 28(5):406-415.
35. Nikam MD, Mahajan PS, Damale MG, et al. Synthesis, molecular docking and biological evaluation of some novel tetrazolo[1,5-a]quinoline incorporated pyrazoline and isooxazoline derivatives. *Med Chem Res.* 2015; 24:3372-3386.
36. Sapre NS, Gupta S, Pancholi N, et al. Molecular docking studies on tetrahydroimidazo-[4,5,1-jk] [1,4]-benzodiazepinone (TIBO) derivatives as HIV-1 NNRT inhibitors. *J Comput Aided Mol Des.* 2008; 22:69-80.
37. Khoshneviszadeh M, Shahraki O, Khoshneviszadeh M, et al. Structure-based design, synthesis, molecular docking study and biological evaluation of 1,2,4-triazine derivatives acting as COX/15-LOX inhibitors with anti-oxidant activities. *J Enzyme Inhib Med Chem.* 2016; 31(6):1602-1611.
38. Faidallah HM, Al-Mohammadi MM, Alamry KA, et al. Synthesis and biological evaluation of fluoropyrazolesulfonyleurea and thiourea derivatives as possible antidiabetic agents. *J Enzyme Inhib Med Chem.* 2016; 31(sup1):157-163.
39. Singh SP, Konwar BK. Molecular docking studies of quercetin and its analogues against human inducible nitric oxide synthase. *SpringerPlus* 2012; 1:69.
40. Leopoldini M, Russo N, Toscano M. The molecular basis of working mechanism of natural polyphenolic antioxidants. *Food Chem.* 2011; 125:288-306.
41. Grover J, Kumar V, Singh V, et al. Synthesis, biological evaluation, molecular docking and theoretical evaluation of ADMET properties of nepodin and chrysophanol derivatives as potential cyclooxygenase (COX-1, COX-2) inhibitors. *Eur J Med Chem.* 2014; 80:47-56.
42. Alam S, Khan F. QSAR and docking studies on xanthone derivatives for anticancer activity targeting DNA topoisomerase II α . *Drug Des Dev Ther.* 2014; 8:183-195.
43. Selassie CD. History of Quantitative Structure-Activity Relationships. In: Abraham DJ, editor. *Burger's Medicinal Chemistry and Drug Discovery (Volume 1: Drug Discovery)*. New York: John Wiley & Sons, Inc.; 2003:3, 9-11.
44. Rudrapal M, Chetia D. QSAR Analysis of 7-Chloro-4-Aminoquinoline Derivatives as Antimalarial Agents. *Asian J Org Med Chem.* 2016; 1:51-54.
45. Taxak N, Bharatam PV. 2D QSAR Study for Gemfibrozil Glucuronide as the Mechanism-based Inhibitor of CYP2C8. *Indian J Pharm Sci.* 2013; 75:680-687.
46. Jain SV, Ghate M, Bhadoriya KS, et al. 2D, 3D-QSAR and docking studies of 1,2, 3-thiadiazole thioacetanilides analogues as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. *Org Med Chem Lett.* 2012; 2:22.
47. Dai Y, Wang Q, Zhang X, et al. Molecular docking and QSAR study on steroidal compounds as aromatase inhibitors. *Eur J Med Chem.* 2010; 45:5612-5620.
48. Liton MAK, Salma U, Bhowmick AC. Cytotoxicity and 2D-QSAR study of some heterocyclic compounds. *Arabian J Chem.* 2014; 7:639-646.
49. Muriuki B, Midiwo JW, Mbugua PM, Keriko JM. The antimycobacterial MICs, SARs, and QSARs of some ethnobotanically selected phytocompounds. *Med Chem Res.* 2013; 22:5141-5152.
50. Podunavac-Kuzmanovic SO, Cvetkovic DD, Barna DJ. QSAR Analysis of 2-Amino or 2-Methyl-1-Substituted Benzimidazoles against *Pseudomonas aeruginosa*. *Int J Mol Sci.* 2009; 10:1670.
51. Wei QL, Zhang SS, Gao J, Li WH, Xua LZ, Yub ZG. Synthesis and QSAR studies of novel triazole compounds containing thioamide as potential antifungal agents. *Bioorg. Med. Chem.* 2006; 14:7146-7153.
52. Yadav AK, Thakur J, Prakash O, Khan F, Saikia D, Gupta MM. Screening of flavonoids for antitubercular activity and their structure-activity relationships. *Med Chem Res.* 2013; 22:2706-2716.
53. Hansch C, Fujita T. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc.* 1964; 86:1616-1626.
54. Silakari O, Kohli DV, Chaturvedi S. Quantitative Structure Activity Relationship Analysis of Ester and Amide Derivatives of Indomethacin. *Indian J Pharm Sci.* 2003; 65(5):471-476.
55. Khatkar A, Nanda A, Kumar P, Narasimhan B. Synthesis, antimicrobial evaluation and QSAR studies of gallic acid derivatives. *Arabian J Chem.* 2017; 10:S2870-S2880.
56. Kumar P, Narasimhan B, Ramasamy K, Mani V, Mishra RK, Majeed ABA. Synthesis, antimicrobial, anticancer evaluation and QSAR studies of 2/3-bromo-N'-(substituted benzylidene/3-phenylallylidene)benzohydrazides. *Arabian J Chem.* 2017; 10:S3740-S3748.
57. Alam S, Khan F. QSAR and docking studies on xanthone derivatives for anticancer activity targeting DNA topoisomerase II α . *Drug Des Devel Ther.* 2014; 8:183-195.
58. Singh J, Grover P, Pathak DP. Synthesis, anticonvulsant activity and comparative QSAR study of some novel 1,2,5-trisubstituted benzimidazole derivatives. *Acta Pharma Sci.* 2010; 52:511-522.
59. Tugba E, Ilkay Y, Betul TG, et al. Synthesis, biological evaluation and 2D-QSAR analysis of benzoxazoles as antimicrobial agents. *Eur J Med Chem.* 2009; 44:501-510.
60. Rudrapal M, Chetia D. QSAR Study of Trioxane Derivatives as Antimalarial Agents. *Curr Trends Pharm Res.* 2016; 3(1):1-17.
61. Roy K, Kar S, Das RN. Statistical methods in QSAR/QSPR. In: Roy K, editor. *A Primer on QSAR/QSPR Modeling*. Springer International Publishing: Switzerland, 2015; 37-59.
62. Shahlaei M. Descriptor Selection Methods in Quantitative Structure-Activity Relationship Studies: A Review Study. *Chem Rev.* 2013; 113:8093-8103.
63. Verma J, Khedkar VM, Couthinho E. 3D-QSAR in Drug Design-a review. *Curr Top Med Chem.* 2010; 10:95-115.