



Received on 24 June 2022; received in revised form, 17 August 2022; accepted 30 August 2022; published 01 March 2023

## QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP AND GROUP-BASED QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP: A REVIEW

Sanket B. Bhatshankar and Amit S. Tapkir \*

Department of Pharmaceutical Chemistry, PES Modern College of Pharmacy, Nigdi, Pune - 411044, Maharashtra, India.

### Keywords:

QSAR, Physicochemical property, QSAR descriptors, GQSAR, QSAR Model, QSAR Method

### Correspondence to Author:

**Dr. Amit Suryakant Tapkir**

Assistant Professor,  
Department of Pharmaceutical  
Chemistry, PES Modern College of  
Pharmacy, Nigdi, Pune - 411044,  
Maharashtra, India.

**E-mail:** amittapkir.8@gmail.com

**ABSTRACT:** Structure-Activity relationship (SAR) only identifies the chemical group responsible for producing the target biological effect in the organism. It was first present in 1865, structural activity relationship was refined dates back to the nineteenth century and now it has advanced from QSAR to 3D QSAR. Quantitative Structure-Activity Relationships (QSAR) are attempted to correlate the structural and biological properties of compounds or molecules. QSAR is used in drug design and medicinal chemistry. By using QSAR, scientists predefined toxicity related to organic molecules and determined the perfect potential biological activity that helps produce drug moiety. Group Based-Quantitative Structure-Activity Relationship (G-QSAR) is fragment dependent method which bases on molecule descriptors. Statistical parameters validate the GQSAR method. The cross-term fragment descriptor is used in GQSAR to study the relation of molecular fragments and variation in biological-based response. It provides a clue for designing new molecules and predicting their activity. In this article, we mainly concentrate on the various QSAR model such as Hansch Analysis, Free Wilson Analysis, various physicochemical properties, QSAR development process, QSAR model, and methodology of the GQSAR model, implementation and working as well as some general aspects in QSAR and GQSAR study.

**INTRODUCTION:** QSAR is an attempt to quantitatively associate a compound's structure or property descriptor with biological activity. Hansch and Fujita first discovered QSAR to understand the structure-activity relationship of drugs for lead recognition and lead optimization. The computational approach has used to determine constitutional, thermodynamics, fragment constants, conformational, hydrophobicity, topology, electronic characteristics, HBD, HBA, and steric effect parameters.

Structure refers to the properties or descriptors of a QSAR molecule, and its function corresponds to biological/biochemical experiments. QSAR has made many advances in drug design and drug development. QSARS perform various activities, from protein binding affinities and toxicity measurements to rate constants. It also includes chemical measurements and biological assays.

When a biological property is determined, it is called a QSPR; unlike a biological property, toxicity determination is called a QSTR. The QSAR method does not clarify which part of the molecule needs to be replaced or altered to increase its activity<sup>1</sup>. Interpreting the model created by the QSAR approach is tricky since it uses separate substituent-based descriptors, which do not provide clear instructions about where to improve. GQSAR is an innovative approach based on the fragment

<p><b>QUICK RESPONSE CODE</b></p> 	<p><b>DOI:</b> 10.13040/IJPSR.0975-8232.14(3).1131-48</p>
<p>This article can be accessed online on <a href="http://www.ijpsr.com">www.ijpsr.com</a></p>	
<p>DOI link: <a href="http://dx.doi.org/10.13040/IJPSR.0975-8232.14(3).1131-48">http://dx.doi.org/10.13040/IJPSR.0975-8232.14(3).1131-48</a></p>	

that gives useful information on viable places of substitution, chemical composition, and the ultimate interface that consequence molecule behavior<sup>2-4</sup>. The modern QSAR approach<sup>5</sup> has presented, which includes descriptor access fragment molecule created by specified fragmentation guidelines for a given dataset. G-QSAR allows for the flexibility of defining and studying individual molecule sites and the ease of interpreting the resulting model, making it easier to build novel molecules<sup>6</sup>. This article looks at the many systematic aspects of QSAR and the new GQSAR approach in structured activity relationships.

**QSAR and GQSAR History:** In the 19<sup>th</sup>-century QSAR method has discovered. In 1868, Crum Brown and Fraser first printed the QSAR equation. Consider the first formulation of the QSAR. Richard *et. al* (1983) It has been described that the poisonousness of organic compounds is inversely proportional to their water solubilities or that their biological activity varies due to changes in chemical and physiological properties. Fujita and Ban clearly defined QSAR in 1970<sup>7, 8, 9, 10</sup>. GQSAR is a procedure created by V Life Sciences Technologies that streamlines the customary QSAR approaches by help in understanding difficulties.

### Objective of QSAR and GQSAR:

#### Objective of QSAR:

1. QSAR is a method for determining the association in the structure of a compound and its biological properties. It helps in determining the biological activity of the lead compound.
2. QSAR decide the toxicity of the lead compound and assist to keep away from the chemical impact of lead compound at the surroundings within the drug design and clinical trial study.
3. Improve existing leads to improve bioactivity.
4. QSAR reduces the time it takes to manufacture a drug.

#### Objective of GQSAR:

1. GQSAR eliminates QSAR interpretation issues.
2. GQSAR solves the inverse QSAR problem.

3. By combining fragments, GQSAR creates a new molecule.
4. Using a cross-term fragment descriptor and a fragment descriptor, GQSAR can investigate the relationship between fragment molecules and discover variability in biological response.

**QSAR Development Process:** The input of molecular structures and the building of 3D models are the initial steps in QSAR research. To calculate the geometric description, a 3D molecule model is needed. The second key stage in QSAR research is the creation of molecular structure descriptors. The third step is to choose descriptors which are done using feature assortment approaches. Developing a QSAR model using a descriptor set is the IV<sup>th</sup> major step in QSAR studies; the fifth and final step validates the model by forecasting the activities of the molecule from an external forecasting set. To rapidly determine the best fitting model, the results obtained by the estimation are compared with those obtained by training set and cross-validation set<sup>11, 12</sup>.

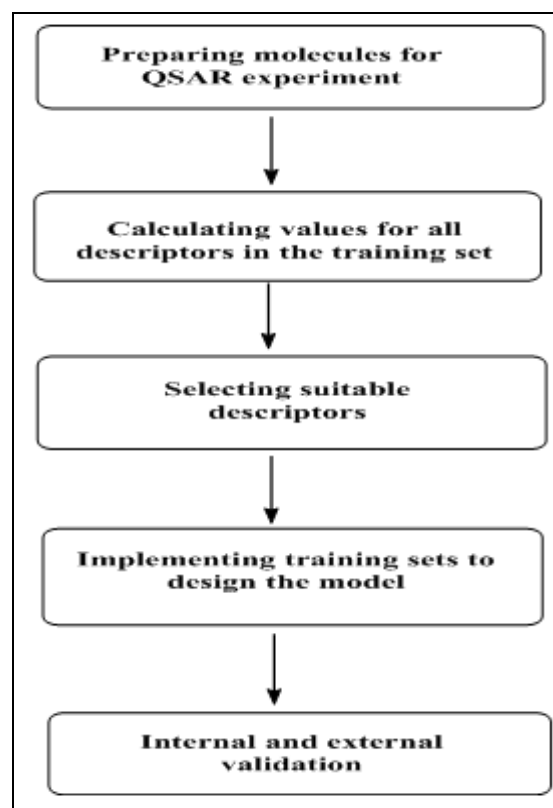


FIG. 1: QSAR DEVELOPMENT PROCESS<sup>11</sup>

**QSAR Models:** Since, the generation of QSAR technique various models in QSAR introduced:

**Hansch Analysis:** There are two types of linear free-related energy approaches:

**Linear Models Corwin:** Hansch determined the fundamental lipophilicity, broadly defined as the partition coefficient of octanol-water. (P), on biologically derived activity in 1969. This property calculates compound bioavailability, which decides compound reaches the target. The equation is:

$$\text{Log}(1/C) = a \log P + b$$

In those equations, 'C' is the molar level of concentration of the compound that produces a general response (e.g., LD<sub>50</sub>, ED<sub>50</sub>, IC<sub>50</sub>, EC<sub>50</sub> etc.) The correlation progressed with the aid of using combining Hammett's electronic parameter and Hansch's degree of lipophilicity with the aid of using the usage of the equation as:

$$\text{Log}(1/C) = k_1\pi + k_2\sigma + k_3$$

Thus,  $\sigma$  = Hammett substitution factor,  $\pi$  is similar to  $\sigma$ <sup>7</sup>.

**Non-Linear Models:** The failure of linear equations in the wide range of hydrophobic ties has led to the evolution of the Hansch parabolic equations, which include (log P) elements in the QSAR equations. This can be explained in one of two ways a word that refers to the fact that multiple membranes must be crossed for compounds to pass through together desired location, and those that have the most hydrophobicity of the membranes will become localized initially come across<sup>7, 13</sup>. Hansh's approach correlates differences in chemical structure with differences in the property, including lipophilic, electronic, and possibly steric substituents in biological reactions. It is showed in a mathematical way as:

$$\log(1/C) = \Delta G_h + \Delta G_e + \Delta G_s + \text{constant}$$

$$\log(1/C) = a \log P - b (\log P)^2 + c\sigma + dE_s + \text{constant}$$

Although logarithm in partition coefficient is Log P, the Hammett electronic constant is  $\sigma$ , and the Taft steric constant is  $E_s$ . The coefficients derived by multiple regression analysis suit the biological data a, b, c and d<sup>7</sup>.

#### Advantages:

1. Descriptors for tiny organic molecules ( $\sigma$ ,  $\pi$ ,  $E_s$ , etc.) used to characterize biological systems.

2. Predictions are statistically quantifiable and measurable.
3. It's quick and simple.
4. Extrapolation possibilities

#### Disadvantages:

- 1) The compounds are required in large numbers.
- 2) The use of small molecule descriptors on biological systems are mostly limited.
- 3) In biological systems, steric factors have a restricted application.
- 4) Drug partial protonation in physiological conditions<sup>7</sup>.

#### Free Wilson Analysis / De Novo Approach:

Wilson analysis is a simple process. This tool is useful in the primary stages of lead structure optimization. The least number of compounds desired for Free Wilson analysis is a type of regression equation analysis that looks at each parameter<sup>10, 14, 15</sup>.

Linear regression analysis is used to solve a set of linear equations that have been formulated. The free-wilson approach was a fact-based SAR model. For each structural trait that differs from a set of randomly chosen compounds, an indicator variable is created<sup>7</sup>.

$$\text{Log}(1/C) = \sum a_j X_{ij} + \mu$$

"This de novo technique is like a standard QSAR which assumes that substituent effects are additive and constant"<sup>7</sup>. Log (1/C) has indicated physiological activity. The value of the third substituent  $X_j$  is 1 if it is present (a specific substituent or structural feature) and 0 if it is not.

This signifies the importance of the Jth substituent to physiological activity, thus representing a total average action. The sum of all action counts in each level is equal to zero<sup>7</sup>.

#### Advantages:

1. It is simple to create a table for regression analysis.

- The insertion and removal of compounds are straightforward and have little impact on the values of other regression coefficients.
- As a reference compound, any compound can be chosen.
- A pseudo substituent that always occurs together in two distinct locations of the molecule is composed of two substituents
- Problems with singularity are usually avoided.

#### Limitations:

- Structure variation is necessary at two separate substitution positions, first and foremost. Otherwise, a useless group contribution would arise, one for each component.
- It has the disadvantage of not providing a solid foundation for evaluating results in the form of drug-receptor interactions.
- The related group input includes the actual error in the experiment of such a particular physiological data set, at least to single point confirmation, for each substituent that only occurs in the data set.
- A large amount of parameters is required in most circumstances to characterize a small number of compounds, resulting in statistically insignificant equations.

**Mixed Approach:** Based on the idea that the Free Wilson model and the linear Hansch analysis are each efficient, they appear extraordinarily diverse and essentially related. Both strategies start with group additivity contributions to the biological process. Free and Wilson have been simplest interested in assigning incremental values to everything. The Hansch model interprets various groups and substituents in physicochemical terms, such as activity contributions<sup>7, 14</sup>. Hansch analysis link found with the Free-Wilson analysis<sup>16</sup>. Because the method used in Hansch analysis is so similar to Free-Wilson analysis, they can both be employed in the same manner," however, due to their hypothetical consistency and numerical activity contribution equivalencies, the mixed approach is the name given to this development, which has shown by the equation below<sup>7</sup>.

$$\text{Log}(1/C) = \sum a_j + \sum c_j \theta_j + \text{Constant}$$

The word  $a_j$  denoted the contributions of each  $i$  h substituent, where  $j$  denotes any physiological or chemical property of  $X_j$  that is a substitute. Depending on the following assumption, a mixed strategy was developed:

- The parent structure of all the compounds in the study is the same.
- Distinct derivatives have to be identical to different substitution patterns.
- The additive role of substitution in biological activity.
- Unaffected by occurrence or absence of other factors substitution.

**Other Approaches:** Pattern recognition techniques have received much attention in the last two decades. In concept, they are similar to the traditional QSAR method. The number of variables in a pattern recognition system is the only thing that matters. The study is significantly higher than the Hansch analysis. Consistent Multivariate approaches, like principal component analysis, are used to achieve outcomes. Two options are techniques such as component analysis or soft modeling, such as SIMCA or PLS analysis. Many different but essentially interconnected QSAR technologies begin with hyperstructures and virtual molecules. In the stepwise optimization approach, the presence or absence of specific hyper-structured atoms or groups within a single molecule is associated with biological activity. For example, LOCON and LOGANA<sup>7, 17</sup>.

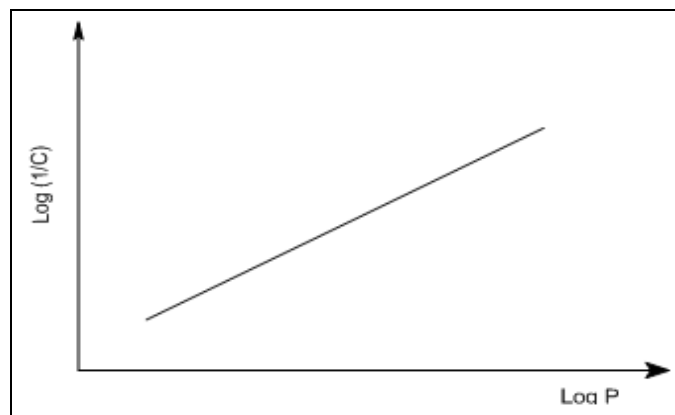
**Physicochemical Properties:** The physicochemical properties of QSAR include lipophilic parameters, electronic parameters, and steric factors or effects.

**Lipophilic Parameter:** Lipophilicity is the most researched physicochemical property. Lipophilicity tests have been there for a long time. *In-silico* lipophilicity technologies that are reliable and affordable are frequently utilized in drug development<sup>18, 19</sup>.

**Partition Coefficient:** The partition coefficient determines the lipophilic nature of a drug and indicates its potential to penetrate cell membranes.



This has stated as the ratio between nonionized drugs distributes in equilibrium in organic and aqueous layers. Drugs along with higher partitions coefficient may cross biological membranes. The diffusion of drug molecules through the velocity control membrane highly depends on the partition coefficient. Sustained-release oral formulations are undesirable for drugs with a low partition coefficient and inadequate for drugs with a high partition coefficient. It is a drug that reaches the site of action and passes into a series of biological membranes. P is a measurement used to move the amount of drug across these membranes<sup>18</sup>. This is because the type of relationship formed is known from the substance used. If the range of possible P-values is narrow, you can use regression analysis. A linear equation is used to represent the result **Fig. 2**.



**FIG. 2: THE DATA ACCESSIBILITY POINTS TO THE BEST LINE**

The equation represents a linear association between a drug's partitioning coefficients and its activities.

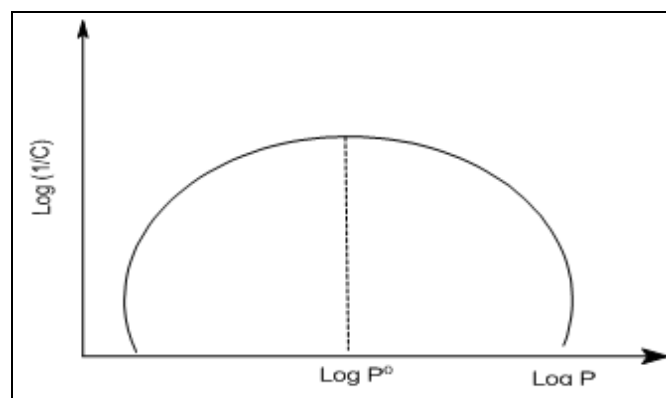
$$\text{Log } (1/C) = k_1 \log P + k_2$$

Where, C stands for the drug concentration required to generate a conventional action at a specific time.

The logarithmic of the material's partitioning coefficient between 1-octanol and water is called logP. k<sub>1</sub>, k<sub>2</sub> are constant.

**Regression Analysis:** Regression analysis has been a useful technique when developing models. Statistics establish a connection among biological activities and molecular descriptors. The data can be analyzed in combination with appropriate statistical and alternative selection methods to build a QSAR

model containing a subset of the statistically most important descriptors for identifying biological activity<sup>7</sup>. This is a mathematical procedure for obtaining mathematical equations that incorporate various datasets calculated based on theoretical consideration and experimental work into appropriate computer programs. **Fig. 2** shows a linear relationship between the partition coefficient and the activities of numerous related substances, and these data may be written as linear equations ( $y = mx + c$ ). These values of m and c give the line corresponding to data are calculated by regression analysis<sup>18</sup>.



**FIG. 3: THE LOG [1/C] VS LOG P CURVES CALCULATED**

A broader range of P-values **Fig. 3** and log P's log (1 / C) graphs have maximum values and parabolic shapes (logP<sub>0</sub>). This development has an optimal balance between lipids and water solubility, showing maximum bioactivity for maximum values that correlate with bioactivity.

**Lipophilic Substituent Constant (π):** These would be referred to as hydrophobic substituent constants. The parameter, which represents a substituent's relative hydrophobicity, is defined as

$$\pi = \log P_X - \log P_H$$

P<sub>X</sub> and P<sub>H</sub> has represented the partitioning coefficients of a derivative and its parent molecules. The difference in hydrophobicities among a parent chemical and substitute homologue represent by substituent constant<sup>18</sup>.

Often exchanged by the additional common molecular term log for log K<sub>ow</sub> (logP), the 1-octanol / water partitioning coefficient. You can use a lipophilic substituent constant instead of the partition coefficient. The value of π depends on the

solvents system to calculate the partitioning coefficient. The octanol/water system has been used to determine most values. Negative numbers indicate that substituents are less lipophilic than hydrogen<sup>18, 20</sup>.

**Distribution Coefficient:** Lipophilicity is a sort of pharmacological activity mathematical analysis. The value of their distribution coefficients (D) is often employed to highlight the lipophilicity of ionisable compounds. D referred to the proportion of unionizing and ionizing compound amounts among an organic solvent and an aqueous medium. P values do not account for ionization among many compounds in aqueous solutions, but ionization significantly impacts absorption and dispersion. As a result, these medications are being distributed in Hansch and elsewhere.

E.g., The acid HA's distribution coefficient is usually provided:

$$D = [\text{HA}]_{\text{organic}} / [\text{H}^+_{\text{(aq)}}/\text{A}^-_{\text{(aq)}}]$$

So the pH of an aqueous medium determines the ionization of acids and bases,

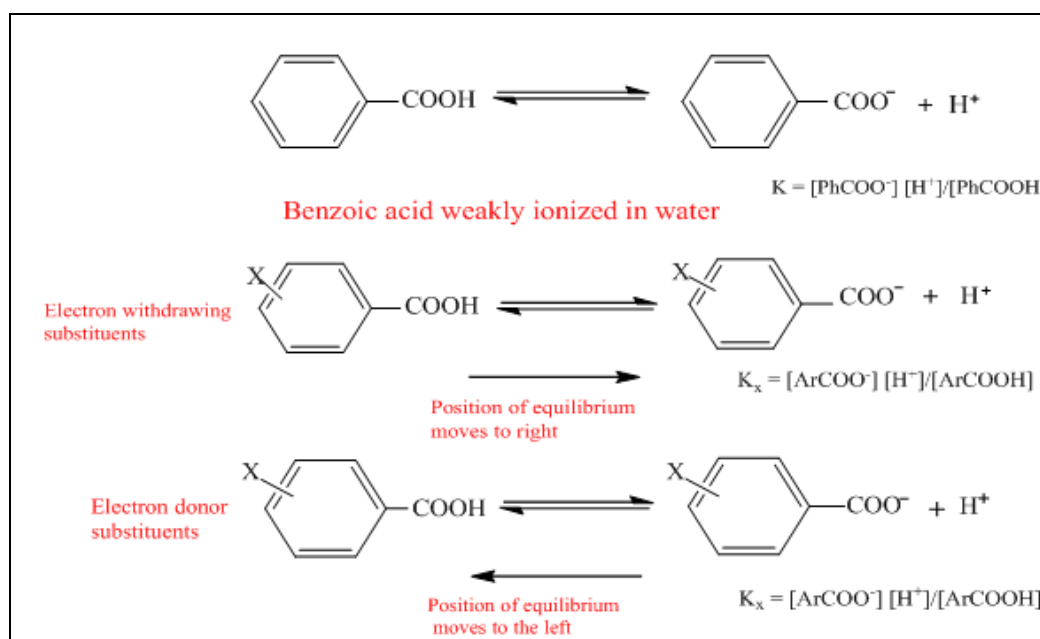
$$\text{Acids: } \text{Log}(P/D-1) = \text{pH} - \text{pKa}$$

$$\text{Bases: } \text{Log}(P/D-1) = \text{pKa} - \text{pH}$$

If the pKa and a value of P for a similar solvents system are known, this equation can be used to compute the effective lipophilicities of a chemical by any pH. Distribution coefficients are commonly used to calculate log D values<sup>18, 21</sup>.

**Electronic Parameter:** The transfer of electrons within a therapeutic molecule profoundly affects scheduled drug activity and delivery. The drug usually reaches the target through a series of biological membranes. When it reaches the site of action, the electron distribution within a drug structure regulates the nature of the bonds it exhibits with the target and determines biological activities.

**The Hamett Constant ( $\sigma$ ):** The Hamett equation connects observable stability or reaction rate changes to systematic substituent changes that alter electron donation and removal capabilities. The molecular structure determines the electron donor and removing groups and electronic distribution within the structure. To measure the influence of substituents on any reactions, Hammett utilized an empirical electronic substitution parameter ( $\sigma$ ) generated from the acidic constants Kx of substituted benzoic acid **Fig. 4**.



**FIG. 4: THE IMPACT OF AN ELECTRON WITHDRAWAL AND DONOR GROUP ON THE POSITION OF BALANCED SUBSTITUTED BENZOIC ACID<sup>18</sup>**

The carboxylate anion is stabilized and the carboxyl group's O-H bond is weakened whenever

an electron is removing substituents (-X), like the nitro group, replacing ring hydrogen. These

equilibrium position moves to the right, indicating that these substituting molecules are more powerful acid than benzene carboxylic acid ( $K_x > K$ ).

Adding an electron-donated substituent (-X) to the ring, as a methyl group, on the other hand, increases the acidic OH group while lowering carboxylate anions' stabilities. As a result, the equilibrium shifts to the left, showing that it's a weaker acid than benzoic acids ( $K > K_x$ ). Hammett studied the relationship between acid strength and aromatic acid structure using equilibrium constants. Hammett constants, also considered Hammett substitution constants ( $\sigma_x$ ), are derived the same for a variety of benzoic acid ring substituents (X) <sup>18</sup>.

The Hammett constants ( $\sigma_x$ ) are as follows:

$$\sigma_x = \log K_x / K$$

$$\text{i.e., } \sigma_x = \log K_x - \log K$$

$$\sigma_x = pK - pK_x \text{ [as } pK_a = -\log K_a]$$

A negative value for  $\sigma_x$  implies that the substituent was working as an electron donor group because  $K >> K_x$ . On the other hand, a positive value for

$K < K_x$  indicates that these substituents are operating as an electrons removing group. These value of  $\sigma_x$  changes depending on the substituent's position in molecule. Typically, the position indicates by a subscript o, m, and p. The substitution exhibits opposite sign depend on where it is on the ring, indicating that it works as both an electron withdrawing and electron donor group. Thus, Hammett constant involve each resonance and inductive components to an electron distribution.

**Steric Factor:** Steric factor is harder to quantify than the electronic or hydrophobic properties. A few techniques are utilized to decide steric factor are as per the following:

**Taft's Steric Factor ( $E_s$ ):** Taft defined the 1956 steric parameters using the relative rate constants of acid catalyse hydrolysis of substitution as methyl ethanoates **Fig. 5**. This was discovered that a rate of this hydrolysis must be controlled almost entirely by steric factors. Use of methyl acetate as a solvent. He defined it as a standard <sup>19</sup>.

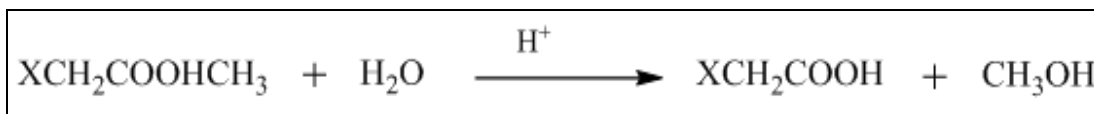


FIG. 5: A- SUBSTITUTED METHYL ETHANOATES HYDROLYSIS

$$E_s = \log K_x - \log K_o$$

Where,  $K_o$  showed the rate of hydrolysis of the starting ester.  $K_x$  showed the rate of hydrolysis of the substituted ester.  $E_s$  values get from a group by refer hydrolysis data can be applied to another structured including that group.

**Molar Refractivity (MR):** It is an estimation of a compound polarization as well as its volume. The refractive index has a proportion of the polarizability while the  $M/\rho$  are a proportion of the molar volume of the compound <sup>18</sup>.

$$\text{MR} = (n^2 - 1) M / (n^2 + 2) \rho$$

Where,  $n$  known as the refractive index.  $M$  defines as the relative mass.  $\rho$  defines as the density of the compound.

**Verloop Steric Parameter:** It utilizes a computer programme called sterimol, which uses

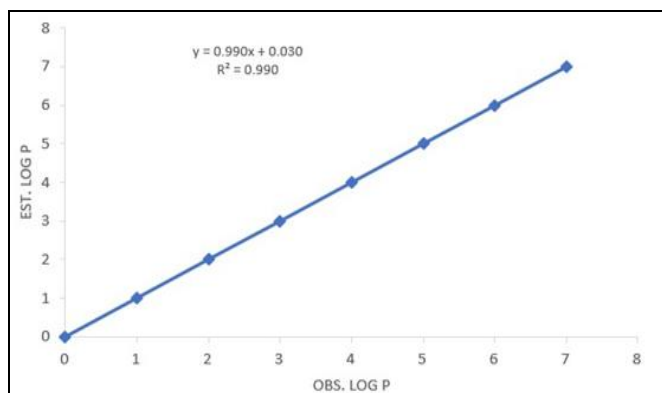
conventional Vander wals radii, bond length, bond angles, and possible confirmations is for a substituent to determine steric substituent values (Verloop steric parameters). The Verloop steric parameters, unlike the  $E_s$ , can be determined for any substituent <sup>18, 22</sup>.

**Statistical Methods used in QSAR:** The two kinds of the statistical method employed in QSAR. The correlation methodology utilised to create a link between structural features and biological activities will determine this.

**Simple Methods:** It consists of multiple linear regression (MLR) and partial least-squares (PLS)

**Multiple Linear Regression (MLR):** The multiple regression approach is used to exclude appropriate descriptors from a various-descriptors <sup>15</sup>. MLR identifies a linear relationship between the input descriptor and the activities.

MLR (Multiple Regression) is a modeling technique that expresses the relationship between two variables by applying linear equations to empirical data using two and many explanatory variables along one response variable. This method was used to correlate the binding affinity with the molecular descriptor<sup>23, 24</sup>. **Fig. 5** shows a graphical representation of each data set observed and calculated activities using multiple linear regression analysis<sup>25</sup>.



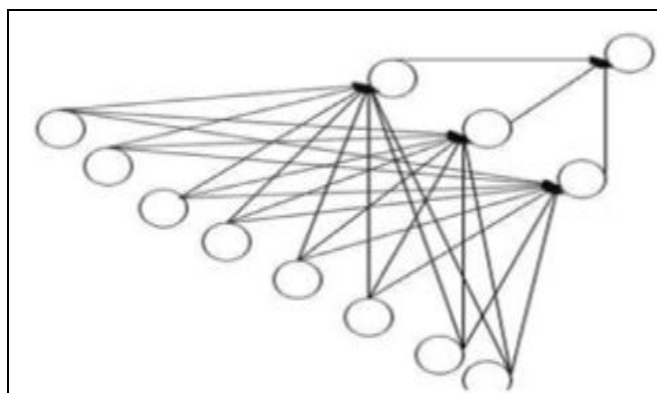
**FIG. 6: MLR GRAPH SHOWING THE DIFFERENCE BETWEEN OBSERVED AND ESTIMATED BIOLOGICAL ACTIVITY**

**Partial Least-Squares (PLS):** PLS may be a modification of MLR that converts the input descriptors and activity-containing space employing Principal Component Analysis (PCA) before performing linear regression<sup>26</sup>. This allows PLS to manage high correlating input descriptors and less likely to identify random relationships<sup>26</sup>. PLS is a method for building predictive models using many highly co-aligned components. It is used in various applied sciences. Traditional algorithms are commonly referred to as PLS. The algorithm preference depends on the shape of the data matrix. This updated method or orthogonalization method using a small update matrix is one of the calculation methods for solving a new algorithm. PLS has been used to monitor and control industrial processes with hundreds of adjustable variables and dozens of outputs. However, PLS is effective when predictions are needed, and there is no physical limitation on the number of items that could be measured. It is also a common method of soft modeling in industrial applications<sup>27</sup>.

**Non-linear Methods:** It includes artificial neural networks (ANN) and Random Forest Method.

**Artificial Neural Network (ANN):** ANN has been more popular in recent years. "Artificial neural networks (ANNs) were originally designed to mimic the structure of neurons in the brain, but the latest implementation has been somewhat removed from this original idea"<sup>28</sup>. Node is organized into layers in the standard ANN feed-forward architecture and has input, Layer, hidden layers, and output layers are most common<sup>28</sup>. The QSAR inset of chemical descriptors derived from the MLR and an observed activity can be predicted using a neural network (ANN). This is a generalization of the model of biological systems in mathematics. The ability to build is ANN's most important feature, using data from experimental measurements in the problem area to model the problem using drug design. ANN has been used to address a variety of issues related to pharmaceutical processes and product development. The neural network has become a model-free mapping device can capture the complex non-linear relationships of the underlying data, often not found in standard QSAR techniques.

**Fig. 7** shows a schematic presentation of the ANN<sup>24</sup>.



**FIG. 7: SCHEMATIC REPRESENTATION OF ANN**<sup>24</sup>

**Random Forest Method (RFM):** The RFM is a revolutionary machine intelligence technique that has quickly established itself as the industry standard for constructing global statistical models based on QSARs. Random forest models contain a huge set of impartial decision trees or regression trees (typically 100-500). Sagging is a term used to describe methods. Every tree in the forest is built with this procedure using separate bootstrapping samples of training data. N compounds have been chosen for the sample substitution for the original dataset. By only evaluating a portion of each tree's



descriptors, a second source of randomization is introduced, split node; as a result of these two sources of unpredictability, each tree represents a distinct aspect of the average predictions across the forest based on input data. Trees consistently offer accurate predictions<sup>28, 29, 30, 31</sup>. Random forests are a good approach to determining the relative value of incoming descriptors, and the variation of predictions across trees gives a decent indication of the predicted random variable<sup>32</sup>.

**Molecular Descriptors:** Molecular descriptors convert a compound structure into numerical values set that indicate numerous molecular attributes considered significant for the compound function describing the activity. Two major descriptors are separated based on the requirement on information regarding the molecule's 3D orientation and conformation.

**2D QSAR Descriptors:** The various descriptor employed in 2D QSAR have similar properties of being autonomous of the 3D direction of the connection. The descriptor measures molecular units based on topology characteristics and calculates geometric characteristics, electrostatics and quantum chemistry descriptor, and innovative fragment counting methods.

**Constitutional Descriptors:** Constitutional-descriptor show a molecule property about the elements that make up the structure. Determining this descriptor is quick and simple. The molecule mass, the number of atoms contained in the molecule, and the atoms of distinct identities are all examples of constitutional descriptors. Bond properties such as the number of singlets, doublet, triplets and aromatic bonds, and an aromatic ring are also considered<sup>33</sup>.

**Electrostatic and Quantum-Chemical Descriptors:** Electrostatic descriptors are used to describe the electrical nature of a molecule. Descriptors describe atomic net and partial charges. The negative, positive, and molecular polarizability descriptions with the highest negative, positive and molecular polarizability are the most informative. Solvent-accessible atomic surface areas, either negatively or positively charged, have also been used as a data source. Intermolecular electrostatic descriptors bonding of hydrogen solvent-accessible

negatively or positively charged atomic surface regions have also been used as a data source. As derivative quantities like absolute hardness, the energies of the highest occupied and lowest empty molecule orbitals form important quantum chemical descriptors<sup>33-38</sup>.

**Topological Descriptors:** Topological descriptors treat the structure of a compound as a graph, atoms act as vertices, and covalent bonds act as edges. Based on this method, many indexes have been developed to measure the connectivity of molecules. It starts with the Wiener index<sup>39, 40</sup>, which counts the total number of shortest path bonds between all pairs of non-hydrogen atoms. Other topological descriptors include a random index  $x$ , a Balaban's  $J$  index, and a Schultz index, clearly defined as the sum of the geometric mean edges of atoms in a path of a given length<sup>41, 42</sup>. Descriptors, eg Kier and Hall index  $xv$  or Galvez topological charge index<sup>33, 43, 44</sup>. The Topological Sub-Structural Molecular Design (TOSS-MODE/TOPS-MODE)<sup>45, 46</sup> rely on spectral moments of bond adjacency matrix amended with information on for *e.g.*, bond polarizability. The atom type electro-topological (E-state) indices<sup>47, 48</sup> use electronic and topological organization to define the intrinsic atom state and the perturbations of this state induced by other atoms<sup>33</sup>.

**Geometrical Descriptors:** The structural distribution of the atoms that make up a molecule is the basis for geometrical descriptors. The surface information acquired from atomic Vander Waals regions and their intersections is one of these descriptors<sup>49</sup>. Molecular volume can be calculated using atoms van-der waal mass<sup>50</sup>. Gravitational indices and principal moments of inertia record data on the arrangement of spatial kind atoms in a molecule<sup>51</sup>. Shadow areas were also created via projecting the molecules along there two primary axes. Another descriptor included is the total solvents-accessible surface area<sup>33, 52, 53, 54</sup>.

**Fragment-Based Descriptors and Molecule Fingerprints:** The descriptor based on substructure ideas is widely employed, particularly for fast-showing very large databases. Bits is used to create BCI fingerprints that denote the occurrence or absence of specific elements within molecule fragment, such as atoms and their surroundings,

ring-based fragments, atom pairs and sequences<sup>33, 55</sup>. The basic set of 166 MDL keys uses a similar method. Other MDL key variations<sup>56</sup>. However, it can also be accessed using an extended or compact keyset. The latter results from special pruning methods or removal processes such as FRED / S-KEYS (fast random removal of descriptor/substructure keys). The newly announced hologram-QSAR (H-QSAR) technology relies on calculating the occurrence of substructured pathways for specific functional groups<sup>57, 58</sup>. It no longer relies on a predefined list of substructure motifs<sup>59, 60</sup>. A natural development of fragment-based descriptors is the Daylight fingerprint. Each molecule's fingerprint has a collection of bits<sup>61</sup>.

On the other hand, a structural idea in molecules does not equate to a single bit but rather from a sequence of bits that is added to the fingerprint using a logical hashing function. Bits in diverse patterns might overlap due to the wide variety of probable patterns and the determinate length of a bits string. As a result, the presence of single bit or many bits in a fingerprint does not indicate that the pattern is there. The pattern is also required to be removed from the molecule if one of the bits relating to it is not set. This makes it possible to quickly identify molecules that lack specific structural patterns<sup>33, 62</sup>.

**3D-QSAR Descriptors:** The three dimensional-QSAR method significantly extra difficult to compute than the two dimensional-QSAR method. Obtaining the complex structure's numerical descriptors involves several procedures in general. The chemical's conformation must first be determined using empirical values or molecular mechanics and then adjusted by applying energy minimization. The dataset's conformer must be evenly aligned in space. Finally, many descriptors for the space with submerged conformer are computed. Some techniques have also been developed that are not dependent on compound alignment<sup>33, 63, 64</sup>.

**Alignment-Dependent 3D QSAR Descriptors:** The group containing methods that require molecular alignments before calculating descriptors is entirely based on receptor knowledge for the modelled ligand. The receptor-ligand complexes

are examined in the alignment. Computational approaches can be used to superimpose structures in spaces<sup>33, 65, 66</sup>.

**Comparative Molecular Field Analysis:** The electrostatic *i.e.*, coulombic and three-dimensional (van der Waals) energy fields defined by the chemicals under study, are used in Comparative Molecular Field Analysis (CoMFA). Next, place the aligned molecules on a 3D grid. Probe atoms with a unit charge are placed at each grid position to determine the energy field's potential (Coulomb and Lennard Jones). They are then used as descriptors in subsequent analysis. This is usually done with partial least squares regression. This analysis identifies structural areas that are associated with the advantages and disadvantages of the activity at hand<sup>33, 67</sup>.

**Comparative Molecular Similarity Indices Analysis:** It includes CoMSIA and CoMFA. The sensor atom's resemblance to the investigated molecule is calculated. CoMSIA, unlike CoMFA, uses a different possible function (the Gaussian-type function). The probe atom's steric, electrostatic, and hydrophobic properties are then calculated, yielding a property of unit hydrophobicity. Perhaps of Lennard Jones or coulombic functions, a Gaussian-type possible function permits more accurate information on grid places inside the molecule. Excessively huge value is attained in these points because possible functions and arbitrary cuts must be employed in CoMFA<sup>33, 68</sup>.

**Alignment-Independent 3D QSAR Descriptors:** This is an additional type of three-dimensional descriptor corresponding to the rotation and translation of molecules in space; as a result, no compound determination is necessary.

**Comparative Molecular Moment Analysis (CoMMA):** The second moments of mass and charge distribution is used in CoMMA. The moment is related to the center of gravity and the center of dipole. The principal moment of inertia, the magnitude of the dipole moment, and the principal quadrupoles moment belong to the CoMMA descriptor. Descriptors that relate the charge to the mass distribution, such as the magnitude of the dipole projection at the first

moment of inertia and the distance in the centre of gravity and centre of the dipole, are also defined<sup>33, 69</sup>.

**Weighted Holistic Invariants Molecular Descriptors (WHIM):** The robust information is provided by a weighted holistic invariant molecule (WHIM)<sup>70, 71</sup> and a molecular surface WHIM descriptor<sup>72</sup> that use principal components analysis. The centre coordinates of atoms that make up the molecules. This convert the molecule into a space that captures the greatest changes. Some statistics such as variance, ratio, symmetry, kurtosis, etc. are calculated and used as the direction descriptor for this space. Undirected descriptors are generated by combining directed descriptors. Chemical properties can be weighted to the contribution of each atom, resulting in various principal components that represent the variance within the property. Mass, van der Waals volume, electro-negativity of an atom, Polarizability of an atom, keel and hole exponents of an electrical topology and electrostatic potential of a molecule can all be used to measure the weight of an atom<sup>33</sup>.

**VolSurf:** The VolSurf technique is predicated on unique probes investigating the grid across the molecule, together with hydrophobic interactions, H-bond acceptor, and donor groups. The descriptors primarily based on volumes of three-D contours, decided through the same cost of the probe molecule interplay electricity, are computed using the lattice bins that result. Different molecular traits may be quantified using numerous probes and electricity cut-off values. Molecule quantity and floor, in addition to hydrophobic and hydrophilic areas, are examples. It is likewise viable to compute spinoff values together with molecules globularity, elements linking floor of hydrophobic and hydrophilic areas to the floor of complete molecules<sup>33, 73, 74</sup>.

**Grid-Independent Descriptors (GRIND):** GRIND was created to address the interpretability issue that plagues alignment-independent descriptors. It works like VolSurf by probing the grid with a special probe. The locations with the most favorable interaction energies are selected, assuming that the distance between them is quite large. Probe-based energy is presented in a molecular structure-independent manner. To

achieve this, the distance in the node of the lattice is discretized into a series of the bin. Nodes with high product energy are kept per spacer bin, and the product value acts as a numeric descriptor<sup>33, 75</sup>.

**Application of QSAR:** Qualitative Structure Activity Relationship (QSAR) application in drug design and medicinal chemistry are as follows<sup>15</sup>:

- ❖ To improve the prevailing leads to recover their biological activities.
- ❖ Before synthesis, identify the hazardous compounds and toxicity of the therapeutic molecule. The toxicity of environmental species and other biological systems will be reduced due to this.
- ❖ Pharmacological and pesticidal activity optimization
- ❖ This identifies and selects molecules to achieve the best biological response and pharmacokinetic properties.
- ❖ To determine the role of numerous qualities in creating a therapeutic molecule and which properties are better for improving biological activity.

**Limitation of QSAR:** The QSAR contains well-defined physicochemical descriptors for selecting numerous compounds and computational screening of molecular databases. However, daily challenges in drug design show that this has some limitations<sup>15</sup>.

- ❖ Biomolecules are mostly found in elaborate three-dimensional forms, whereas traditional QSAR exclusively deals with two-dimensional structures.
- ❖ Only a smaller number of descriptors are considered when employing 2D descriptors, which is a shortcoming of the old method.
- ❖ Despite its abundance, there is no illustration of the stereochemistry and 3-D structure of the molecule.
- ❖ Because the resultant model lacks predictability, a synthesis on behalf of the 2D model is difficult.

- ❖ The random association, rather than the actual prediction, is better in 2D QSAR models.
- ❖ Given that the standard QSAR equations do not directly suggest new compounds for synthesis, constructing a molecule requires much knowledge about substituent constants in physical organic chemistry.

**Methodology of GQSAR:** The current methods for generating QSAR using fragment descriptors include the Free-Wilson approach, HQSAR, and two-dimension topological QSAR<sup>76-78</sup>. The recommended novel approach G-QSAR, on the other hand, varies from it in two ideas: (i) In the G-QSAR method, every molecule in the database has fragmented using a group of predetermined criteria prior to the fragmentation descriptors being calculated. This differs from earlier methods, which analyze the molecule for a specified fragment (or group) and then utilize this even as a descriptor, such as an indicator variable, a count and related index, such as molecular connectivity indices. (ii) The G-QSAR technique uses cross-connection terms to account for fragment connection in the QSAR model; however other methods do not use these descriptors.

**Preparation of the Dataset:** Marvin Sketch was used to create the structures of the congeneric derivative dataset. The VLifeEngine module of VLifeMDS was used to generate the 2D structures into 3D structures<sup>79</sup>. The force field batch minimizes methods of V-LifeEngine were utilized to accomplish energy minimization of 3D compounds. This procedure is used to optimize the molecules until they reach their lowest stable energy levels. Marvin Sketch was also used to create the template, which maintained similar structures moiety in the congeneric database.

**Calculation of Fragment Descriptors:** The GQSAR module of VLifeMDS<sup>5, 79</sup> is used for this step. The compounds' pIC<sub>50</sub> values were then entered manually into VLifeMDS, and several 2-D physicochemical descriptors for such functional groups occurring in a distinct area of substitution of the molecules were calculated<sup>1</sup>.

1. For multiple fragments observed within every molecule in the series, recognized 2D-descriptor such as chi-indices, valence-based

chi-indices, electron topological indices, Baumann alignment independence topological descriptors<sup>80</sup>. HBA, HBD, rotatable bonds, and/or other 3-D alignment non-dependent descriptors such as dipole moment, the radius of gyration, volume, Polar Surface Area (PSA) *etc.*

2. Cross-interaction indifferent fragments were generated and used to construct various QSAR models as a descriptor; however, the numerous 2-D and 3-D descriptors were derived for different fragments included in the molecule.

**Creation of Training Set and Test Set:** There was a total of 37 compounds in the dataset used for the GQSAR study. These compounds were then manually split into training and test sets to verify that both sets had a uniform distribution of existing and dormant compounds. To maintain a balance ratio, the 37 compounds were randomly separated into a test set (30 % of the dataset) and a training set (70 % of the dataset) as in previous GQSAR research<sup>81, 82, 83</sup>. Molecules 2, 11, 14, 15, 20, 23, 28, 29, and 33 had contained in the test set, whereas the remaining had in the training set.

**Building of the GQSAR Model:** The variable Selection and Model Building methods like Stepwise Forward, Backward, and Forward-Backward, Simulated Annealing, Genetic Algorithm methods for Variable Selection and Multiple Regression, Partial Least Square, Principal Component Regression methods for Model Building are utilized and applied in the GQSAR model. The Stepwise Forward variables selection method had used in this study to generate a subset of descriptors from a pool of descriptors. The variable selection method as stepwise forward start by building a trial model with only one independent variable one step at a time. The solo variables are introduced one by one at each step, and the model has changed appropriately. If the regression coefficient of the final variables entered the model is negligible, or if all of the variables in the model contain insignificant regression coefficients, the procedure ends<sup>84</sup>. The Partial Least Square method was mostly utilized to construct the models. This method connects a Y matrix of dependent factors (such as a molecule's biological activity) to an X matrix of solo variables (like physicochemical



descriptors). This method has two main goals: to estimate the two matrices and to minimize the correlation between them. Matrix X is divided into many latent variables that best correlate with the molecule's activity<sup>85</sup>.

**Validation of the Developed GQSAR Model:** G-QSAR model is constructed by taking into account a number of crucial statistical factors.  $R^2$ ,  $q^2$ , pred  $r^2$ , F-test, and standard deviation are among them. A statistical approach of comparing two separate models is the  $r^2$ , correlation coefficient, and F-test. Lower pred  $r^2$ ,  $q^2$  and  $r^2$  values, as well as a higher F test result, indicated a good model. Validation represents a crucial phase in the creation of QSAR models. Validation of models in a QSAR

investigation includes greater than statistical fitting, relevance, and predictability with cross-validations. Validation now involves data quality assessment and applicational and mechanical interpretations. Validation procedures need to determine the reliability of a QSAR model on unknown data and the difficulty of QSAR models that justify data under examination. For extensive validation of QSAR models, a variety of approaches published, including least squares fit ( $r^2$ ), cross-validation ( $q^2$ ), adjusted  $r^2$  ( $r^2_{adj}$ ), chi-squared test ( $\chi^2$ ), Root Mean Squared Error (RMSE), bootstrapping, and scrambling (Y-Randomization). The confirmed QSAR models were validated using the usual leave-one-out procedure<sup>86</sup>.

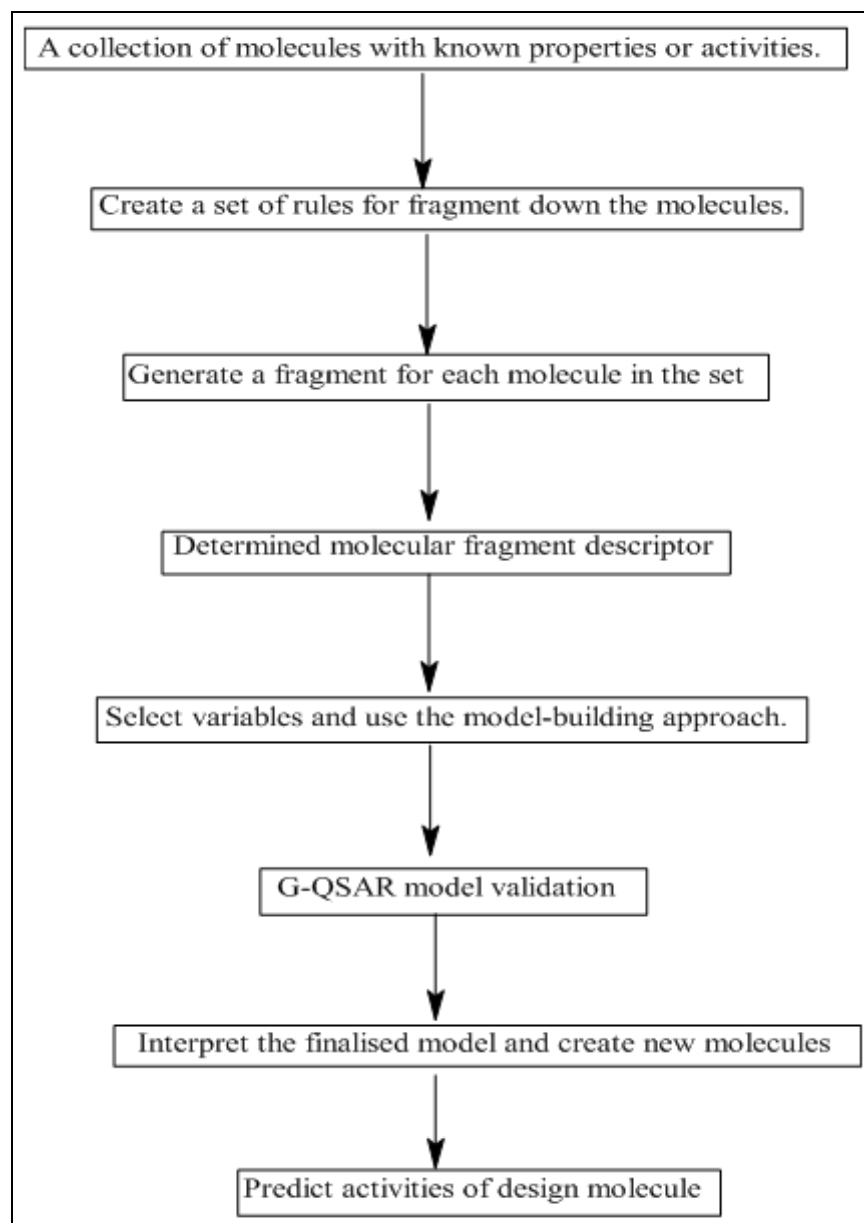


FIG. 8: FLOW CHART OF GQSAR METHODOLOGY

**Implementation of QQSAR in VLife MDS**

**Software:** VLife Technologies Pvt. Ltd. created QQSAR, a unique group (fragment) largely producing the QSAR method that considerably improves the capabilities of conventional QSAR. It looks into the connection between two molecular fragments of interest. And the variance in its biological response, taking into account the connection between fragments using cross-term fragmentation descriptors. The QQSAR provides site-specific insights for developing novel compounds and estimating overall activity quantitatively<sup>79</sup>.

**System Requirement:**

**Operating System:** Windows- XP, Windows-Vista, Windows - 7, Linux [Fedora, Ubuntu, CentOS]

**Required Hardware:**

1. The lowest free hard disk space require is 1 GB.
2. The minimum memory required is 2 GB.

**Graphic Cards:** The standard graphic card is required, which supports OpenGL.

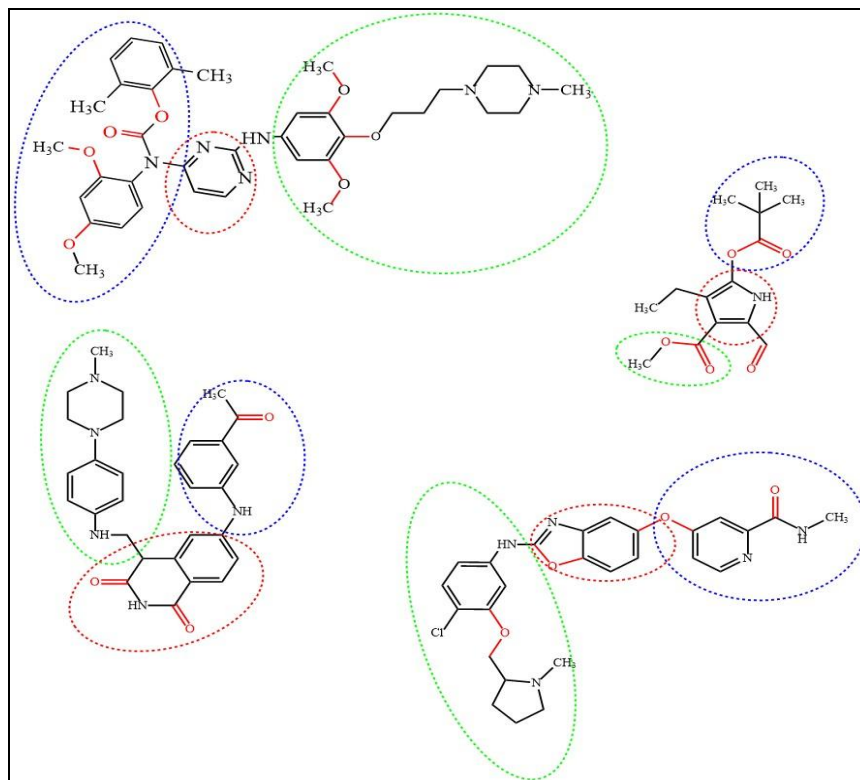
**Working:** In two ways, the QQSAR method differs from the traditional fragment-based QSAR method:

In the database, every molecule is represented by a single molecule and is fragmented using a set of predetermined criteria. QQSAR supports both automatic and manual molecular fragmentation.

**Automatic:** Approach dependent on templates, especially for congeneric series of molecules.

**Manually:** Operator-defined strategy, especially for non-congeneric molecular series. The participation of neighboring groups is included in the fragments produced by these techniques. The QQSAR approach uses cross/interaction words as descriptors to account for fragment interactions.

To use the QQSAR modules of the Vlife MDS, the common scaffold was employed as a template for the fragment-based QSAR model. Isoquinoline-1,3-dione, pyrimidine, 3-cyano-6-hydroxy quinoline, benzoxazole, pyrrole, and other chemical scaffolds were used to group the compounds.



**FIG. 9: FRAGMENTATION PATTERN. THE DATABASE GROUP INTO DIFFERENT CHEMICAL STRUCTURES DEPENDING ON SIMILAR CHEMICAL PARTS, MAINLY AROMATIC HETEROCYCLIC RING NAMED AS THE SCAFFOLD OR FRAGMENT R-2 (RED CIRCLE). THE 3D STRUCTURE FRAGMENTED INTO THREE FRAGMENTS R-1 (BLUE CIRCLE), R-2 (RED CIRCLE), R-3 (GREEN CIRCLES)**

All molecules then fragmented into three different fragments, as shown in **Fig. 9**.

1. Fragment R1: Aromatic substitution on the scaffold, i.e., Fragment R2.
2. Fragment R2 is formed by an aromatic heterocyclic ring (chemical scaffold) structure after Fragment R1.
3. Fragment R3: Substitution of the scaffold, i.e., Fragment R2<sup>2</sup>.

The optimized compounds were entered directly into the QSAR worksheet and the biological activity. Molecular characteristics are estimated using molecular descriptors, which are numerical representations of a molecule's chemical information. Molecular descriptors are calculated using logical and mathematical procedures on the equation. For the groups existing at the substitutional site in every molecule, GQSAR calculated 239 physicochemical descriptors from subclasses such as individual, chi, chiv, chain path count, cluster, path Cluster, kappa, estate numbers, and estate contributors.

**CONCLUSION:** QSAR is the technique used in drug design and medicinal chemistry. Using QSAR, we can determine the toxicity and biological activity of the chemical compound. There is advancement in QSAR that is dependent on the descriptor use and their physicochemical property. The GQSAR is a fragment-based descriptor method used to interpret the QSAR model as it gives clear direction about the site for improvement. GQSAR comes with auto and manual molecule fragmentation methods to help predict the relation of molecular fragments and variation in biological response through cross-term fragment descriptors. Nowadays, these new GQSAR methods are widely used along with QSAR for structure-activity relationship in various QSAR models. In this review, we learn the general aspect of the QSAR and GQSAR that will help new researchers to study more advancement in various structure-activity relationship studies.

**ACKNOWLEDGEMENT:** The author acknowledges Dr. Amit Tapkir for their immense help in preparing the manuscript.

**CONFLICTS OF INTEREST:** The author declares no conflict of interest.

## REFERENCES:

1. Joshi K, Goyal S, Grover S, Jamal S, Singh A and Dhar P: Novel group-based QSAR and combinatorial design of CK-1δ inhibitors as neuroprotective agents. *BMC Bioinformatics* 2016 17.
2. Abdullahi AD, Abdulkader AM, Abdulsamat NB and Ingale K: Application of group-based qsar and molecular docking in the design of insulin-like growth factor antagonists. *Tropical Journal of Pharmaceutical Research* 2015; 14(6): 941–51.
3. Choudhari P, Kumbhar S, Phalle S, Choudhari S, Desai S and Khare S: Application of group-based QSAR on 2-thioxo-4-thiazolidinone for development of potent anti-diabetic compounds. *Journal of Molecular Structure* 2017; 1128: 355–60.
4. Singh A, Goyal S, Jamal S, Subramani B, Das M, Admane N and Grover A: Computational identification of novel piperidine derivatives as potential HDM2 inhibitors designed by fragment-based QSAR, molecular docking and molecular dynamics simulations. *Structural Chemistry* 2016; 27(3): 993–1003.
5. Ajmani S, Jadhav K and Kulkarni SA: Group-based QSAR (G-QSAR): Mitigating interpretation challenges in QSAR. *QSAR and Combinatorial Science* 2009; 28(1): 36–51.
6. Ajmani S, Jadhav K and Kulkarni SA: Group Based QSAR (G-QSAR): A Flexible and Focused Approach to Study Quantitative Structure Activity Relationship 287: 1–3.
7. Vaidya A and Jain S: Quantitative structure activity relationship: A novel approach of drug design and discovery. *Journal of Pharmaceutical Sciences and Pharmacology*, American Scientific Publishers 2014; 1: 219-32.
8. Crum BA and Fraser TR: On the connection between chemical constitution and physiological action; with special reference to the physiological action of the salts of the ammonium bases derived from strychnic, brucia, terbaia, codeia, morphia and nicotia. *Journal of Anatomy and Physiology* 1868; 2: 224–42.
9. Reichert D, Neudecker T, Spengler U and Henschler D: Mutagenicity of dichloroethylene and its degradation products trichloroacetyl chloride, trichloroacryl chloride and hexachlorobutadiene. *Mutation Research* 1983; 117: 21–29.
10. Fujita T and Ban T: Structure-activity study of phenylethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. *Journal of Medicinal Chemistry* 1971; 14: 148–52.
11. Illah AL, Veljovic E, Gurbeta L and Badnjevic A: Application of QSAR Study in Drug Design. *International Journal of Engineering Research and Technology (IJERT)* 2017; 6(06): ISSN: 2278-0181
12. Sahigara F, Mansouri K, Ballabio D, Mauri A, Consonni V and Todeschini R: Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules* 2012; 17: 4791-4810.
13. Penniston JT, Beckett L, Bentley DL and Hansch C: Passive permeation of organic compounds through biological tissue: a nonsteady-state theory. *Molecular Pharmacology* 1969; 5:333–36.

14. Free J and Wilson SM: Mathematical contribution to structure activity studies. *Journal of Medicinal Chemistry* 1964; 7: 395-99.
15. Kubinyi H: Free Wilson Analysis. Theory, Applications and its Relationship to Hansch Analysis. *Quantitative Structure-Activity Relationships* 1988; 7(3): 121-33.
16. Singer JA and Purcell WP: Relationships among current quantitative structure-activity models. *Journal of Medicinal Chemistry* 1967; 10: 1000-02.
17. Andrea IA and Kalayeh H: Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *Journal of Medicinal Chemistry* 1991; 34: 282-32.
18. Kumar K and Kapoor Y: Quantitative structure activity relationship in drug design: An overview. *SF Journal of Pharmaceutical and Analytical Chemistry* 2019; 2(2): 1017.
19. Hansch C, Leo A and Hoekman D: Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, Hydrophobic, Electronic and Steric Constants. *Journal of American Chemical Society* 1995; 1
20. Hansch C, Steward AR, Anderson SM and Bentley DL: Parabolic dependence of drug action upon lipophilic character as revealed by a study of hypnotics. *Journal of Medicinal Chemistry* 1968; 11: 1-11.
21. Manallack DT: The pKa Distribution of Drugs: Application to Drug Discovery. *Perspective in Medicinal Chemistry* 2007; 1: 25-38.
22. Tipker J and Verloop A: Use of STERIMOL, MTD and MTD\* Steric Parameters in Quantitative Structure-Activity Relationships. *Journal of American Chemical Society* 1984; 255: 279-96.
23. Ojha LK, Chaturvedi AM, Bhardwaj A, Thakur M and Thakur A: *Asian Journal of Research in Chemistry* 2012; 5(3): 377-82.
24. Ojha LK, Sharma R and Bhawsar MR: Modern drug design with advancement in QSAR: A review, *International Journal of Research in Bio Sciences* 2013; 2.
25. Pathan S, Ali S and Shrivastava M: Quantitative structure activity relationship and drug design: A Review, *International Journal of Research in Biosciences* 2016; 5(4): 1-5.
26. Clark M and Cramer RD: The probability of chance correlation using partial least-squares (PLS). *Quantitative Structure Activity Relationship* 1993; 12: 137-45.
27. Geladi P and Kowalski BR: Partial least-squares regression: a tutorial. *Analytica chimica Acta* 1986; 185: 1-17.
28. Richard AL and David W: Modern 2D QSAR for drug discovery. *WIREs Computational Molecular Science* 2014; doi: 10.1002/wcms.1187
29. Breiman L: Random forests. *Machine Learning* 2001; 45: 5-32.
30. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP and Feuston BP: Random Forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Scientists* 2003; 43: 1947-58
31. Breiman L: Bagging predictors. *Machine Learning* 1996; 24: 123-40.
32. Wood DJ, Carlsson L, Eklund M, Norinder U and Stalring J: QSAR with experimental and predictive distributions: an information theoretic approach for assessing model quality. *Journal of Computer Aided Molecular Design* 2013; 27: 203-19.
33. Arkadiusz ZD, Tomasz A and Jorge G: Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. *Combinatorial Chemistry & High Throughput Screening*, Bentham Science Publishers Ltd 2006; 9: 213-28.
34. Fei J, Mao Q, Peng L, Ye T, Yang Y and Luo S: The Internal Relation between Quantum Chemical Descriptors and Empirical Constants of Polychlorinated Compounds. *Molecules* 2018; 23(11): 2935.
35. Papp T, Kollar L, Kegl T: Employment of quantum chemical descriptors for Hammett constants: Revision Suggested for the acetoxy substituent. *Chemical Physics Letter* 2013; 588: 51-56.
36. Stanton DT, Egolf L M, Jurs PC and Hicks MG: *Journal of Chemical Information and Computer Scientists* 1992; 32: 306-16.
37. Bereket G and Hur E: Quantum chemical studies on some imidazole derivatives as corrosion inhibitors for iron in acidic medium. *Journal of Molecular Structure Theochem* 2002; 578: 79-88.
38. Oluwaseye A, Uzairu A, Shallangwa G and Stephen A: Quantum chemical descriptors in the QSAR studies of compounds active in maxima electroshock seizure test. *Journal of King Saud University – Science* 2018; 32(1).
39. Sun Q, Ikica B, Škrekovski R and Vukašinović V: Graphs with a given diameter that maximise the Wiener index. *Applied Mathematics and Computation* 2019; 356: 438-48.
40. Poulik S and Ghorai G: Determination of journeys order based on graph's Wiener absolute index with bipolar fuzzy information. *Information Sciences* 2021; 545: 608-19.
41. Agüero-Chapin G, Galpert D, Molina-Ruiz R, Ancedo-Gallardo E, Pérez-Machado G, De la Riva GA, Antunes A: Graph Theory-Based Sequence Descriptors as Remote Homology Predictors. *Biomolecules* 2020; 10(1): 26. <https://doi.org/10.3390/biom10010026>
42. Alijanabi HK: Schultz index, Modified Schultz index, Schultz polynomial and Modified Schultz polynomial of alkanes. *Global Journal of Pure and Applied Mathematics* 2017; 13(9): 5827-50.
43. Hall LH and Kier LB: Issues in representation of molecular structure: The development of molecular connectivity. *Journal of Molecular Graphics and Modelling* 2001; 20(1): 4-18.
44. Hall LH and Kier LB: The Meaning of Molecular Connectivity: A Bimolecular Accessibility Model. *Croatica Chemica Acta* 2002; 75(2): 371-82.
45. Saiz-Urra L, Teijeira M and Rivero-Buceta V: Topological sub-structural molecular design (TOPS-MODE): a useful tool to explore key fragments of human A3 adenosine receptor ligands. *Mol Divers* 2016; 20(1): 55-76. doi:10.1007/s11030-015-9617-z
46. Estrada E: On the topological sub-structural molecular design (TOSS-MODE) in QSPR/QSAR and drug design research. *SAR QSAR Environmental Research* 2000; 11(1):55-73. doi:10.1080/10629360008033229
47. Kier LB Hall LH: An electrotopological-state index for atoms in molecules. *Pharmaceutical Research* 1990; 7(8):801-7. DOI: 10.1023/a:1015952613760. PMID: 2235877.
48. Kier LB and Hall LH: Molecular structure description: The electrotopological state. San Diego: Academic Press 1999.
49. Gupta A, Kumar V and Aparoy P: Role of Topological, Electronic, Geometrical, Constitutional and Quantum Chemical Based Descriptors in QSAR: mPEGS-1 as a case study. *Current Topic in Medicinal Chemistry* 2018; 18(3): 1075-90.
50. Lu T, Chen Q: Van der Waals potential: an important complement to molecular electrostatic potential in



- studying intermolecular interactions. *Journal of Molecular Modeling* 2020; 26: 315.
51. Uzan JP: Varying Constants, Gravitation and Cosmology. *Living Reviews in Relativity* 2011; 14(2).
  52. Abdulfatai U, Uzairu A and Uba S: Molecular docking and QSAR analysis of a few Gama amino butyric acid aminotransferase inhibitors, *Egyptian Journal of Basic and Applied Sciences* 2018; 5(1): 41-53.
  53. Ilah LA, Veljovic E, Gurbeta L and Badnjevic A: Applications of QSAR Study in Drug Design. *International Journal of Engineering Research & Technology (IJERT)* 2017; 6(6).
  54. El fadili M, Er-Rajy M, Kara M, Assouguem A, Belhassan A, Alotaibi A, Mrabti NN, Fidan H, Ullah R, Ercisli S, Zarougui S, Elhallaoui M: QSAR, ADMET *in-silico* Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia. *Pharmaceuticals* 2022; 15(6): 670. <https://doi.org/10.3390/ph15060670>
  55. Varnek A and Baskin I: Fragment Descriptors in SAR/QSAR/QSPR Studies, Molecular Similarity Analysis and in Virtual Screening 2008 <http://www.md1.com/>.
  57. Janela T, Takeuchi K and Bajorath J: Introducing a Chemically Intuitive Core-Substituent Fingerprint Designed to Explore Structural Requirements for Effective Similarity Searching and Machine Learning. *Molecules* 2022; 27: 2331. <https://doi.org/10.3390/molecules27072331>
  58. Muegge I and Mukherjee P: An Overview of Molecular Fingerprint Similarity Search in Virtual Screening. *Expert Opinion on Drug Discovery* 2016; 11: 137-48.
  59. Moll M, Bryant DH and Kavraki LE: The Label Hash algorithm for substructure matching. *BMC Bioinformatics* 2010; 11: 555. <https://doi.org/10.1186/1471-2105-11-555>
  60. Isayev O, Oses C, Toher C: Universal fragment descriptors for predicting properties of inorganic crystals. *Nature Communication* 2017; 8: 15679 <https://doi.org/10.1038/ncomms15679>
  61. <http://www.daylight.com/>.
  62. Deshpande M, Kuramochi M, Wale N and Karypis G: *IEEE Transactions on Knowledge and Data Engineering* 2005; 17: 1036-50.
  63. Bhonsle JB, Venugopal D, Huddler DP, Magill AJ and Hicks RP: Application of 3D-QSAR for Identification of Descriptors Defining Bioactivity of Antimicrobial Peptides. *Journal of Medicinal Chemistry* 2007 50(26): 6545-53.
  64. Akamatsu M: Current State and Perspectives of 3D QSAR. *Current Topics in Medicinal Chemistry* 2002; 2(12): 1381-94.
  65. Cruciani G, Carosati E and Clementi S: 25 - Three-Dimensional Quantitative Structure-Property Relationships. *The Practice of Medicinal Chemistry* 2003; 2: 405-16.
  66. Myint KZ and Xie XQ: Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods. *International Journal of Molecular Sciences* 2010; 11(10): 3846-66. <https://doi.org/10.3390/ijms11103846>
  67. Zhao X, Chen M, Huang B, Ji H and Yuan M: Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) studies on  $\alpha(1A)$ -adrenergic receptor antagonists based on pharmacophore molecular alignment. *International Journal of Molecular Science* 2011; 12(10): 7022-37.
  68. Klebe G, Abraham U and Mietzner T: Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of Medicinal Chemistry* 1994; 37(24): 4130-46. doi:10.1021/jm00050a010
  69. Damale MG and Harke SN: Recent Advances in Multidimensional QSAR (4D-6D): A Critical Review. *Mini-Reviews in Medicinal Chemistry* 2014; 14: 35-55.
  70. Kobayashi Y and Yoshida K: Quantitative structure-property relationships for the calculation of the soil adsorption coefficient using machine learning algorithms with calculated chemical properties from open-source software. *Environmental Research* 2021; 196: 110363. ISSN 0013-9351
  71. Mauri A, Consonni V and Todeschini R: Molecular Descriptors. In: Leszczynski, J., Kaczmarek-Kedziera, A., Puzyn, T., G. Papadopoulos, M., Reis, H., K. Shukla, M. (eds) *Handbook of Computational Chemistry*. Springer, Cham 2017; 2065-93.
  72. Todeschini R and Gramatica P: New 3D Molecular Descriptors: The WHIM theory and QSAR Applications. In: Kubinyi H, Folkers G, Martin YC. (eds) *3D QSAR in Drug Design. Three-Dimensional Quantitative Structure Activity Relationships*. Springer, Dordrecht 2002; 2: 335-80.
  73. Cruciani G, Pastor M and Guba W: VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *European Journal of Pharmaceutical Sciences* 2000; 11 2: 29-39. doi:10.1016/s0928-0987(00)00162-7
  74. Cruciani G, Crivori P, Carrupt PA and Testa B: Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. *Journal of Molecular Structure: THEOCHEM* 2000; 503(1-2): 17-30. ISSN 0166-1280
  75. Pastor M, Cruciani G, McLay I, Pickett S and Clementi S: Grid Independent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *Journal of Medicinal Chemistry* 2000; 43: 3233-43.
  76. Goswami D, Goyal S, Jamal S, Jain R, Wahi D and Grover A: QQSAR modeling and combinatorial library generation of 4-phenylquinazoline-2-carboxamide derivatives as antiproliferative agents in human Glioblastoma tumors. *Computational Biology and Chemistry* 2017; 69: 147-152. ISSN 1476-9271.
  77. Kumar SP, Jasrai YT, Pandya HA and Rawal RA: Pharmacophore-similarity-based QSAR (PS-QSAR) for group-specific biological activity predictions, *Journal of Biomolecular Structure and Dynamics* 2015; 56-69.
  78. Mitra I, Achintya S and Roy K: Quantification of contributions of different molecular fragments for antioxidant activity of coumarin derivatives based on QSAR analyses. *Canadian Journal of Chemistry* 2013; 91(6): 428-41.
  79. VLifeMDS: Molecular Design Suite, VLife Sciences Technologies Pvt. Ltd., Pune, India 2010
  80. Baumann K: *Journal of Chemical Information and Computer Scientist* 2002; 42: 26 - 35.
  81. Goyal S, Grover S, Dhanjal JK, Tyagi C, Goyal M and Grover A: Group-based QSAR and molecular dynamics mechanistic analysis revealing the mode of action of novel piperidinone+ derived protein-protein inhibitors of p 53-MDM2. *Journal of Molecular Graph Model* 2014; 51: 64-72.
  82. Singh A, Goyal S, Jamal S, Subramani B, Das M, Admane N and Grover A: Computational identification of novel piperidine derivatives as potential HDM2 inhibitors designed by fragment-based QSAR, molecular docking

- and molecular dynamics simulations. *Structural Chemistry* 2016; 27(3): 993–1003.
83. Goyal M, Dhanjal JK, Goyal S, Tyagi C, Hamid R and Grover A: Development of dual inhibitors against Alzheimer's disease using fragment-based QSAR and molecular docking. *BioMedicine Research Institute* 2014.
84. Xu L, Zhang WJ: Comparison of different methods for variable selection. *Analytica Chimica Acta* 2001; 446(1–2): 475–81. [http://dx.doi.org/10.1016/S0003-2670\(01\)01271-5](http://dx.doi.org/10.1016/S0003-2670(01)01271-5).
85. Ajmani S and Kulkarni SA: Application of QSAR for scaffold hopping and lead optimization in multitarget inhibitors. *Molecular Informatics* 2012; 31(6–7): 473–90. doi:10.1002/minf.201100160.
86. Choudhari P, Kumbhar S, Phalle S, Choudhari S, Desai S and Khare S: Application of group-based QSAR on 2-thioxo-4-thiazolidinone for development of potent anti-diabetic compounds. *Journal of Molecular Structure [Internet]*. 2017; 1128: 355–60. Available from: <http://dx.doi.org/10.1016/j.molstruc.2016.09.007>.

**How to cite this article:**

Bhatshankar SB and Tapkir AS: Quantitative structure activity relationship and group based quantitative structure activity relationship: a review. *Int J Pharm Sci & Res* 2023; 14(3): 1131-48. doi: 10.13040/IJPSR.0975-8232.14(3).1131-48.

All © 2023 are reserved by International Journal of Pharmaceutical Sciences and Research. This Journal licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

This article can be downloaded to **Android OS** based mobile. Scan QR Code using Code/Bar Scanner from your mobile. (Scanners are available on Google Playstore)