# ORIGINAL RESEARCH ARTICLES

## QSAR STUDIES ON HUMAN CARBONIC ANHYDRASE II INHIBITORS

**Tanvi V. Wani[a] and Mrunmayee P. Toraskar[a]***

(Received 28 February 2020) (Accepted 05 October 2021)

**ABSTRACT**

Carbonic anhydrase II is one of the forms of human α carbonic anhydrases which are ubiquitous metalloenzymes that catalyze inter-conversion of carbon dioxide and water to bicarbonate and proton, overexpression of which leads to disorders such as glaucoma. 2D and 3D Quantitative Structure Activity Relationship studies were carried out on previously synthesized series of sulfanilamide derivatives by VLife MDS software using stepwise variable, multi-linear regression and k-nearest neighbor molecular field analysis methods. 2D-QSAR model depicts contribution of halogens (such as chlorine and fluorine), methylene and oxygen atoms to inhibition of human carbonic anhydrases II activity. Using k-nearest neighbor molecular field analysis method two 3D-QSAR models (model A and B) were generated from which model A was found to be the best validated model with $q^2$ (0.9494), pred_$r^2$ (0.7367) and $q^2$_se (0.2037). It displayed the fact that the inhibitory action of sulfanilamide derivatives against human carbonic anhydrases II is influenced by hydrophobicity and electro positivity.

**Keywords:** Human carbonic anhydrase II inhibitor, sulfanilamide, 2D-QSAR, 3D-QSAR, Schiff's base

### INTRODUCTION

The carbonic anhydrases (CA) are a family of ubiquitous metalloenzymes consisting metal atom like $Zn^{2+}$ that catalyze the inter-conversion of carbon dioxide ($CO_2$) and water to bicarbonate ($HCO^-_3$) and a proton; this is reversible reaction. This enzyme plays a major role in transportation of products such as $CO_2$ and protons across the biological membranes such as intercellular, intracellular and extracellular spaces[1]. Carbonic anhydrase isoforms are involved in critical physiological processes such as respiration and acid–base regulation, electrolyte (like sodium and potassium) secretion, bone remodeling, calcification and biosynthetic reactions includes lipogenesis, gluconeogenesis, and ureagenesis, which require bicarbonate as a substrate[2]. There are 7 families of carbonic anhydrases, which are α, β, γ, δ, ζ, η, and Θ[3]. Studies related to carbonic anhydrase inhibition have been carried out in our lab since the past few years. Found in 1940, human carbonic anhydrase II (hCAII) isoenzyme of α carbonic anhydrase family was the first zinc metallo enzyme; it has 29.3-kDa protein with 260 amino residues. It efficiently carries $CO_2$ hydration with $k_{cat}$ of about $1.6x10^6 s^{-1}$ at 25 °C and pH 9. The interaction mechanism of hCAII and its inhibitors would be greatly helpful in discovering its novel inhibitors[4]. The sulfonamides play an important role in chemotherapy, alone or in combination with other drugs.

The sulfonamides that inhibit the zinc enzyme carbonic anhydrase possess many applications as diuretic, anti-glaucoma, anti-cancer or anti-epileptic drugs[5]. Series of previously synthesized sulfanilamide derivatives which showed potent human carbonic anhydrase II inhibitory activities with their $K_i$ values, were used for QSAR analysis[6-8]. In this study, 2D and 3D-QSAR models were developed based on these sulfanilamide inhibitors. With these models, it is possible to further understand structural features of hCAII and to find out the interactions between the structural information of these sulfanilamide inhibitors and their potential activities. And it can also provide quantitative estimation of activity and insights into key structural elements to design new hCAII potential lead candidates[4].

[a]Department of Pharmaceutical Chemistry, Bharati Vidyapeeth's College of Pharmacy, Sector 8, C.B.D, Belapur, Navi Mumbai - 400 614, Maharashtra, India

*For Correspondence: E-mail: mrunmayeetoraskar@gmail.com

https://doi.org/10.53879/id.58.11.12350

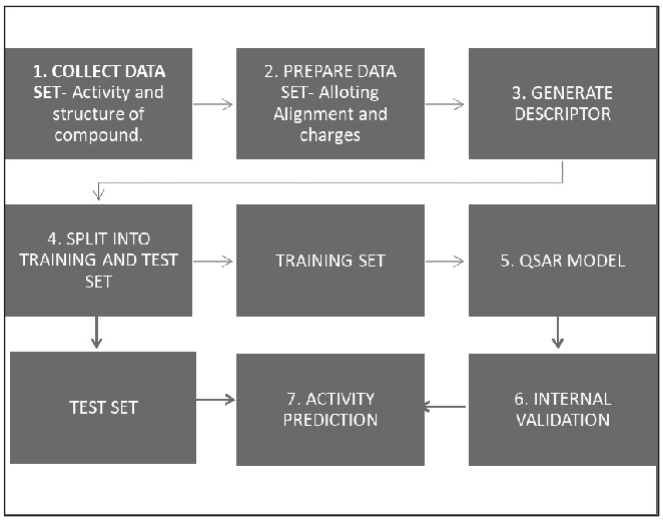



**Fig. 1: Flowchart of QSAR Method**

## MATERIALS AND METHODS

The method followed for QSAR study is stepwise demonstrated in Fig. 1.

## DATA SELECTION

The large set of 32 sulfanilamides, with varied structures such as sulfonyl semicarbazides (compound codes 1-10)[7]; Schiff's bases with iminoureido (compound codes 11-22)[8]; sulphanilamide connected to iminoureido Schiff base through methylene-bridge carbon (compound codes 23-32)[6], were chosen for the present study. *In vitro* hCAII inhibitory constant (Ki) in nM units of the series previously synthesized and evaluated compounds are reported in Table I. The molecules have high structural as well as activity diversity. Hence, this data set becomes suitable for our QSAR studies. To reduce the skewness of data set, the $K_i$ values were converted to a negative logarithmic scale (p $K_i$ = -log ($K_i$)) in molar units and subsequently used as the dependent variable for the QSAR analysis. Common template[9] used during the study is as shown in Fig. 2.

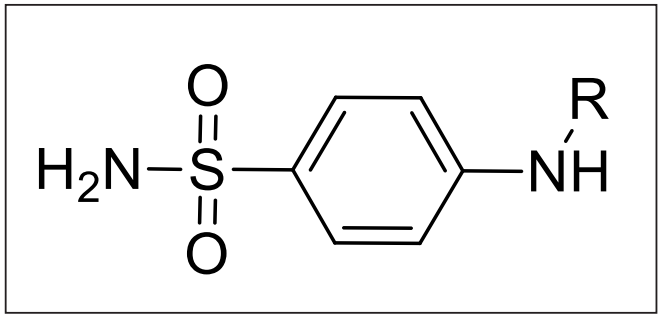
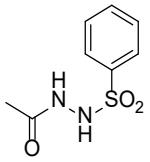
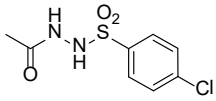
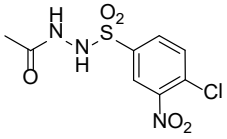
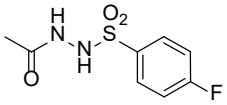
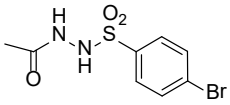
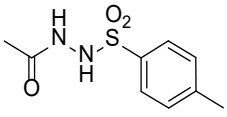
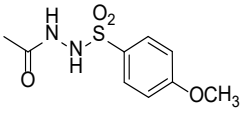
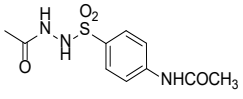
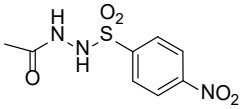
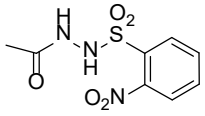
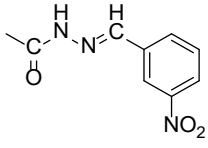
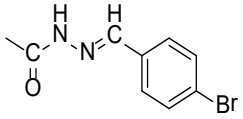
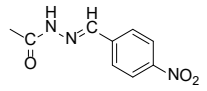
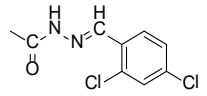
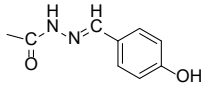
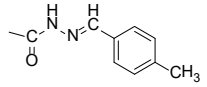
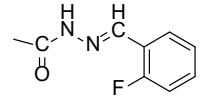
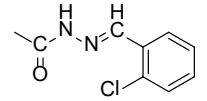
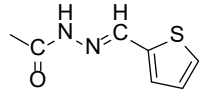
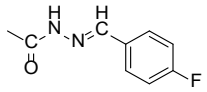
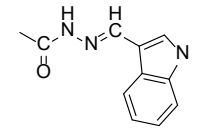
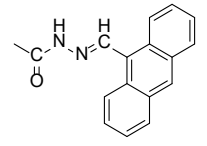
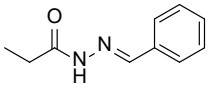
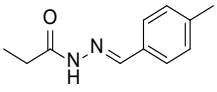
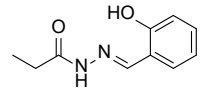
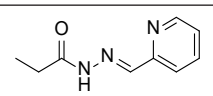
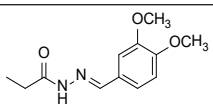
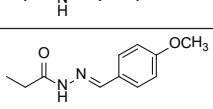
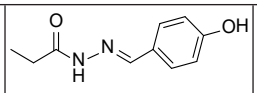
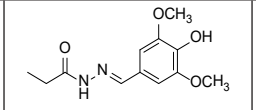
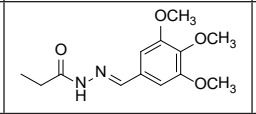
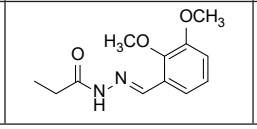



**Fig. 2: Template**

**Table I: Structures of dataset used for QSAR analysis**

| CODE | R* | $K_i$ (nM)* | -log($k_i$) |
|---|---|---|---|
| 1 |  | 4.3 | -0.63 |
| 2 |  | 3.5 | -0.54 |
| 3 |  | 71.8 | -1.85 |
| 4 |  | 14.4 | -1.15 |
| 5 |  | 3.9 | -0.59 |
| 6 |  | 6.1 | -0.78 |
| 7 |  | 3.8 | -0.57 |
| 8 |  | 3.6 | -0.55 |
| 9 |  | 6.0 | -0.77 |
| 10 |  | 9.1 | -0.95 |
| 11 |  | 5.6 | -0.74 |
| 12 |  | 6.2 | -0.79 |

| | | | |
|---|---|---|---|
| 13 |  | 7.4 | -0.86 |
| 14 |  | 75.0 | -1.87 |
| 15 |  | 6.1 | -0.78 |
| 16 |  | 6.7 | -0.82 |
| 17 |  | 4.0 | -0.60 |
| 18 |  | 4.2 | -0.62 |
| 19 |  | 4.6 | -0.66 |
| 20 |  | 19.5 | -1.29 |
| 21 |  | 51.6 | -1.71 |
| 22 |  | 4.7 | -0.67 |
| 23 |  | 549 | -2.73 |
| 24 |  | 914 | -2.96 |
| 25 |  | 32.1 | -1.50 |
| 26 |  | 418 | -2.62 |
| 27 |  | 758 | -2.87 |
| 28 |  | 703 | -2.84 |
| 29 |  | 255 | -2.40 |
| 30 |  | 483 | -2.68 |
| 31 |  | 446 | -2.64 |
| 32 |  | 240 | -2.38 |

**Structure and values taken from Wani et al[6], Pichake et al[7] and Singasane et al[8]*

## MOLECULAR MODELING

All computational experiments were performed using Samsung computer having genuine Intel Core Processor and Windows 7 Professional operating system using the software Molecular Design Suite (VLifeMDS 4.6.1). Structures were drawn using the ChemDraw Ultra 12 application and converted to 3D structures and subjected to an energy minimization and geometry optimization using parameter Merck Molecular Force Field for force field and charges with 1000 as maximum number of cycles, 0.01 as convergence criteria (root mean square gradient) and 1.0 as constant in dielectric properties. Analytical Gradient Type was selected for energy minimization[9].

## 2D-QSAR ANALYSIS

## CALCULATION OF DESCRIPTORS

Numbers of descriptors were calculated after optimization or minimization of the energy of the data set molecules. Descriptors calculated were: Individual (Molecular weight, Polarizability, H-Acceptor, H-Donor count, slog P, X log P, SMR, etc.), atomic valence connectivity index (ChiV), ChiV chain, retention index (Chi), Path count, Chain Path count, Chi chain, Cluster, Kappa, Path cluster, Element count (H, C, N, S count etc.), Distance based topological (DistTopo, Wiener Index, Connectivity Index, Balaban Index), Estate numbers (SsCH3count, SssCH2count, SdCH2count, StCH count, etc.), Estate contribution (SsCH3- index., SssCH2-index, SdCH2- index, StCH index), Information theory based (IPC, ID etc.) and Polar surface area. Numbers of alignment independent descriptors were also calculated using the following attributes. A few examples are T_Cl_Cl_4, T_F_S_5, T_O_Cl_7, T_O_S_4, T_O_O_6, T_N_F_6, T_C_N_6, etc. The invariable descriptors (the descriptors

that are constant for all the molecules) were removed, as they do not contribute to QSAR[10].

## DATA SELECTION

Biological activity was assigned as the dependent variable and all the calculated descriptors as independent variable. In order to evaluate the QSAR model, data set was split into training and test set using selection methods like sphere exclusion, random selection and manual selection method. Training set is selected based on compounds whose biological activity data are known. Test set are compounds which were not included in training set, it is used to verify the QSAR model developed based on the training set to assess the predictive power of the model[10].

## VARIABLE SELECTION AND MODEL BUILDING

Variable selection is one of the important steps in a QSAR study, stepwise, genetic algorithm (GA), and simulated annealing (SA) are the variable selection methods. In the forward stepwise (SW) variable selection method was used and the cross correlation limit was set at 0.7, the number of variables at 4 and the term selection criteria at $r^2$. The variable selection methods were used with the multiple linear regression (MLR) as the regression methods for model building. Variable selection and regressions were done by Vlife MDS software. Two dimensional quantitative structure activity relationship (2D QSAR) studies by means of multiple linear regression (MLR) method were performed on a series of sulfanilamide derivatives using software QSAR pro (VLife MDS).MLR is a method used for modeling linear relationship between a dependent variable Y (Activity) and independent variable X (2D/3D descriptors). MLR is based on least squares. The model is fit such that sum-of squares of differences of observed and a predicted value is minimized. MLR estimates values of regression coefficients ($r^2$) by applying least squares curve fitting method[9].

## ANALYSIS OF 2D – QSAR MODEL

2D QSAR Model was analyzed considering various parameters such as:

n is number of molecules ( > 20 molecules), k is number of descriptors in a model (statistically n/5 descriptors in a model), df is degree of freedom (n-k-1) (higher is better), $r^2$ is coefficient of determination (> 0.7), $q^2$ is cross-validated $r^2$ (>0.5), pred_$r^2$is $r^2$ for external test set (>0.5), SEE is standard error of estimate (smaller is better), F-test is F-test for statistical significance of the model (higher is better, for same set of descriptors and compounds), Z score is Z score calculated by the randomization test (higher is better), best_ran_$q^2$ is highest $q^2$ value in the randomization test (as low as compared to $q^2$), best_ran_$r^2$ is highest $r^2$ value in the randomization test (as low as compared to $r^2$), alpha statistical significance parameter by randomization test (<0.01).

## 3D-QSAR

## ALIGNMENT

The most important requirement for 3D-QSAR studies is that the 3D structure has to be aligned according to a conformational template, which is assumed to be one 'bioactive' conformation. Compounds were aligned on template by using template alignment method in VLifeMDS 4.6.1 software. The molecules were superimposed based on minimizing RMS deviation as shown in Fig. 3. The resulting alignments of molecules were used for kNN 3D-QSAR models building[11].

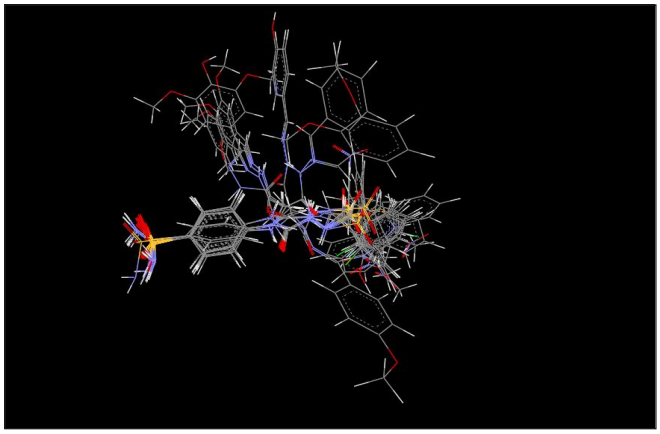


**Fig. 3: 3D view of aligned molecules**

## CALCULATION OF MOLECULAR FIELD DESCRIPTORS FOR 3D-QSAR ANALYSIS

Molecular field analysis (MFA) was employed to derive the 3D-QSAR model in this study. Molecular field analysis (MFA) model is predictive and sufficiently reliable to guide the chemist in designing novel compounds. After superimposition, the overlaid set of molecules is positioned in the center of a lattice or grid box, to calculate interaction energies between the ligands and different probe atoms placed at each intersection of the lattice. The aligned biologically active conformations of sulfanilamide derivatives were used for the calculation of molecular fields. Molecular fields are the electrostatic, steric and hydrophobic interaction energies which were computed at the lattice points of the grid using a methyl probe of charge +1 considering Del Re charges. 10.0 kcal $mol^{-1}$ electrostatic cut-off and 30.0 kcal $mol^{-1}$ steric cut-off were employed to perform descriptor calculations. The term

descriptor is utilized to indicate field values at the lattice points. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. A value of 1.0 is assigned to the distance-dependent dielectric constant. Numbers of three dimensional descriptors were calculated using VLife MDS software. These included electrostatic, steric, and hydrophobic field descriptors for all the compounds in separate columns[11].

## DATA SELECTION

Aligned data set was divided into training and test set using sphere exclusion, random and manual selection method, as performed in 2D-QSAR.The invariable descriptors (the descriptors that are constant for all the molecules) were removed, as they do not contribute to QSAR[12].

## VARIABLE SELECTION AND MODEL BUILDING

SW as the variable selection methods were used with the kNN as the regression methods, with number of maximum and minimum neighbors 4 and 2, respectively, positive as most active and prediction method since distance based weighted average k-Nearest neighbor molecular field analysis (kNN–MFA) is the 3D-QSAR method which has been used to produce the 3D models to indicate the regions where changes with respect to substitution can be made to alter biological activity. In this method, an unknown pattern is classified according to the majority of the class memberships of its k nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g. a molecular similarity measure, calculated using field interactions of molecular structures). The standard kNN method is implemented simply as follows: (i) calculate distances between an unknown object (u) and all the objects in the training set; (ii) select k objects from the training set most similar to object u, according to the calculated distances, (iii) classify object u with the group to which a majority of the k objects belong. An optimal k value is selected by the optimization through the classification of a test set of samples or by the leave-one out cross-validation. The variables and optimal k values are chosen using different variable selection methods as described below[10].

## ANALYSIS OF 3D QSAR MODEL

The test set compounds were computed by internal statistical validation (denoted with $r^2$ and $q^2$) and external statistical validation metrics (denoted with pred_r2) to confirm predictive ability of the model. 3D-QSAR models are considered reliable and acceptable if internal validation, cross-validated correlation coefficient $q^2$ and correlation coefficient $r^2$ is greater than 0.50 and 0.60 respectively[13-15]. 3DQSAR model was analyzed considering various parameters such as: n (> 20 molecules), df (higher is better), q2 (> 0.7), pred_$r^2$ (>0.5), $q^2$ se (smaller is better), pred_$r^2$ se (smaller is better).

## RESULTS AND DISCUSSION

Different sets of QSAR models were generated. The obtained statistical parameters and predicted values in different methodologies have been shown in Table II and Table III respectively. Some statistically significance QSAR models based on the statistical parameters were selected for discussion.

**Table II: Statistical evaluation of 2D-QSAR and 3D-QSAR models**

| Statistics | 2D-QSAR Model | 3D-QSAR Model A | 3D-QSAR Model B |
|---|---|---|---|
| kNN | - | 2 | 2 |
| N | 22 | 22 | 22 |
| Degree of freedom | 17 | 17 | 18 |
| $r^2$ | 0.9915 | - | - |
| $q^2$ | 0.9870 | 0.9494 | 0.9211 |
| F_test | 495.8376 | - | - |
| $r^2$_se | 0.0958 | - | - |
| $q^2$_se | 0.1186 | 0.2037 | 0.2626 |
| pred_$r^2$ | 0.8870 | 0.7367 | 0.6055 |
| pred_$r^2$se | 0.3069 | 0.4931 | 0.5585 |
| Z Score R^2 | 6.95762 | - | - |
| Z Score Q^2 | 3.75500 | - | - |
| Best R and R^2 | 0.56071 | - | - |
| Alpha R and R^2 | 0.00000 | - | - |

**Table III: Actual and predicted activities for compounds based on the 2D/3D-QSAR models**

| CODE | Actual $pk_i$ value | Predicted 2D-QSAR Model | Residual For 2D-QSAR Model | Predicted 3D-QSAR Model A | Residual For 3D-QSAR Model A | Predicted 3D-QSAR Model B | Residual For 3D-QSAR Model B |
|---|---|---|---|---|---|---|---|
| 1 | -0.63 | -0.64 | -0.01 | -1.68 | -1.05 | -0.61 | 0.02 |
| 2 | -0.54 | -1.85 | -1.31 | -0.67* | -0.13 | -0.6* | -0.06 |
| 3 | -1.85 | -1.81 | 0.04 | -0.67 | 1.18 | -0.67* | 1.18 |
| 4 | -1.15 | -1.18 | -0.03 | -1.68 | -0.53 | -0.67* | 0.48 |
| 5 | -0.59 | -0.64 | -0.05 | -0.68 | -0.09 | -0.68 | -0.09 |
| 6 | -0.78 | -0.64 | 0.14 | -0.67* | 0.11 | -0.67 | 0.11 |
| 7 | -0.57 | -0.58 | -0.01 | -0.67* | -0.1 | -0.67 | -0.1 |
| 8 | -0.55 | -0.58 | -0.03 | -0.68 | -0.13 | -0.75 | -0.2 |
| 9 | -0.77 | -0.43* | 0.34 | -1.69 | -0.92 | -0.68 | 0.09 |
| 10 | -0.95 | -0.58* | 0.37 | -1.69* | -0.74 | -0.75 | 0.2 |
| 11 | -0.74 | -0.68 | 0.06 | -0.68 | 0.06 | -0.82* | -0.08 |
| 12 | -0.79 | -0.72 | 0.07 | -0.67 | 0.12 | -1.5* | -0.71 |
| 13 | -0.86 | -0.68* | 0.18 | -0.68 | 0.18 | -0.82 | 0.04 |
| 14 | -1.87 | -1.91 | -0.04 | -0.67 | 1.2 | -1.8 | 0.07 |
| 15 | -0.78 | -0.7* | 0.08 | -0.68 | 0.1 | -0.82 | -0.04 |
| 16 | -0.82 | -0.72 | 0.1 | -0.76* | 0.06 | -0.97* | -0.15 |
| 17 | -0.6 | -0.7 | -0.1 | -0.68 | -0.08 | -0.64* | -0.04 |
| 18 | -0.62 | -0.7 | -0.08 | -0.67* | -0.05 | -0.64 | -0.02 |
| 19 | -0.66 | -0.72 | -0.06 | -0.67* | -0.01 | -0.96 | -0.3 |
| 20 | -1.29 | -1.26 | 0.03 | -0.67 | 0.62 | -0.98 | 0.31 |
| 21 | -1.71 | -0.7 | 1.01 | -0.67 | 1.04 | -1.5 | 0.21 |
| 22 | -0.67 | -0.68 | -0.01 | -0.63 | 0.04 | -1.17* | -0.5 |
| 23 | -2.73 | -2.8* | -0.07 | -1.68 | 1.05 | -2.69 | 0.04 |
| 24 | -2.96 | -2.8 | 0.16 | -1.67 | 1.29 | -2.92 | 0.04 |
| 25 | -1.5 | -2.74 | -1.24 | -2.63 | -1.13 | -2.57 | -1.07 |
| 26 | -2.62 | -2.8 | -0.18 | -1.67 | 0.95 | -2.51 | 0.11 |
| 27 | -2.87 | -2.76 | 0.11 | -1.68 | 1.19 | -2.91 | -0.04 |
| 28 | -2.84 | -2.78 | 0.06 | -1.71* | 1.13 | -2.56* | 0.28 |
| 29 | -2.4 | -2.78* | -0.38 | -1.65* | 0.75 | -2.57 | -0.17 |
| 30 | -2.68 | -2.74 | -0.06 | -1.65 | 1.03 | -2.79 | -0.11 |
| 31 | -2.64 | -2.74 | -0.1 | -1.63 | 1.01 | -2.51 | 0.13 |
| 32 | -2.38 | -2.72* | -0.34 | -1.7 | 0.68 | -2.51 | -0.13 |

** Predicted values for test set molecule*

## 2D-QSAR

The QSAR analysis generated various models out of which only one model in which data set was divided randomly; variable selection by forward stepwise (SW) variable selection method and build by multiple linear regression (MLR) method was selected based on correlation coefficient and different statistical parameters. Further, based on comparison with actual and predicted activity, further model was built by manual selection which gave following parameters:

- 2.0762(± 0.0478) SssCH2count
- 1.2106(± 0.0730) T_2_Cl_5
- 0.5373(± 0.0731) T_2_F_5
+ 0.0212(± 0.0055) T_T_O_7
- 0.8355

## INTERPRETATION OF THE MODEL

From above equation, model explains 99.15 % ($r^2$= 0.9915) of the total variance in the training set as well as it has internal ($q^2$) and external (pred_$r^2$) predictive ability of 98.70 % and 88.70 %, respectively. The F test shows higher value, hence its shows significance of the model. In addition, the randomization test shows Alpha R and Pred R^2 = 0.00000 that the generated model is not random and hence chosen as the QSAR model. From 2D- QSAR model, the descriptor $SssCH_2$ count plays a pivotal role in determining activity. This descriptor $SssCH_2$ count signifies of $–CH_2$ group connected with two single bonds. This is a negative contributing descriptor toward inhibitory activity against hCAII and its contribution is approx 63 % (Fig. 4). The negative coefficient of $SssCH_2$ count showed that more values leads to less inhibitory activity against hCAII enzyme (like in compounds C1-C10). The second important descriptor in determining activity is T_2_Cl_5. This descriptor T_2_Cl_5 signifies the count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from chlorine atom by 5 bonds in a molecule. This is a negative contributing descriptor toward inhibitory activity against hCAII and its contribution is approx 25 % (Fig. 4). The negative coefficient of T_2_Cl_5 showed that increase in the values of this descriptor is not beneficial for the inhibitory activity against hCAII (like in compounds 3, 14). Another important descriptor in determining activity is T_2_F_5. This descriptor T_2_F_5 signifies the count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from fluorine atom by 5 bonds in a molecule. This is a negative contributing descriptor toward inhibitory activity against hCAII and its contribution is approx 11 % (Fig. 4). The negative coefficient of T_2_F_5 showed that increase in the values of this descriptor is not beneficial for the inhibitory activity against hCAII (like in compounds 4, 20). The last descriptor is T_T_O_7. This descriptor T_T_O_7 signifies the count of any atom (represented as T) separated from oxygen atom by 7 bonds. This is a positive contributing descriptor toward inhibitory activity against hCAII and its contribution is approx 3 % (Fig. 4). The positive coefficient of T_T_O_7 showed that higher value leads to better inhibitory activity whereas lower value leads to decrease inhibitory activity against hCAII (like in compounds 1, 6, 7, 8 etc)[10, 11].

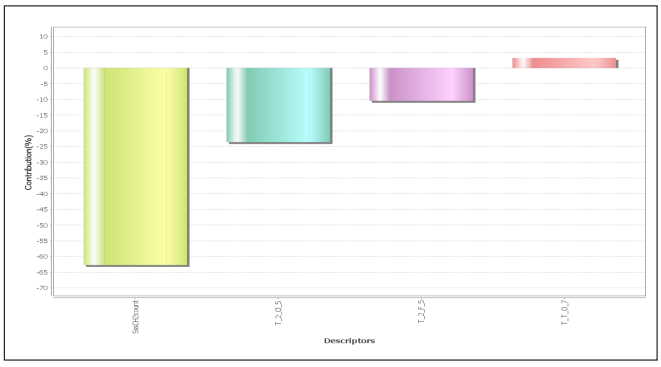



**Fig. 4: Contribution chart for 2D-QSAR model showing contribution of different descriptors**

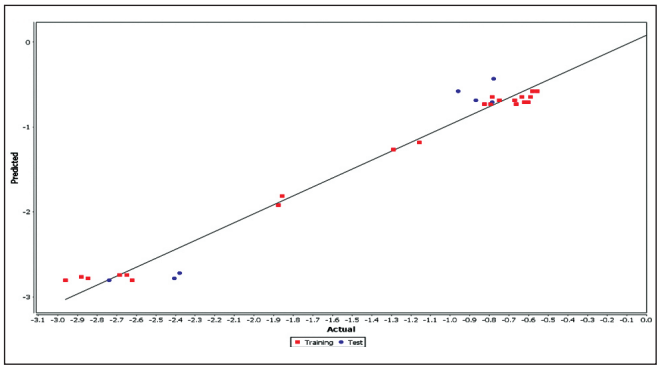



**Fig. 5: Data fitness plot for 2D-QSAR model**

Data fitness plot for model is shown in Fig. 5. The plot of observed vs predicted activity provides an idea about how well the model was trained and how well it predicts the activity of external test set.

The graph of observed vs. predicted activity of training and test sets for model is shown in Figs. 6 and 7 and the model is able to predict the activity of training set quite well as well as external test set, providing confidence of model. Result of the observed and predicted inhibitory are shown in Table II.

## 3D-QSAR

Using different descriptors, training test, test set and variance selection method has provided number of different models which are described as below:

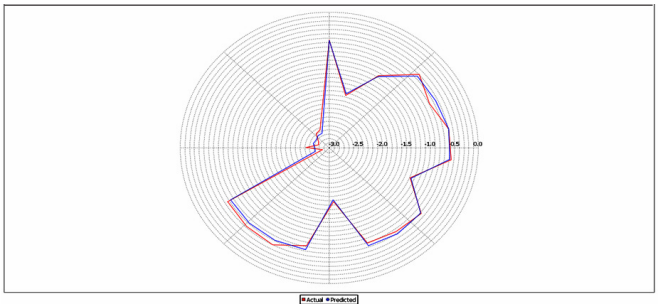


**Fig. 6: Graph between actual and predicted biological activity of training set for 2D-QSAR model**

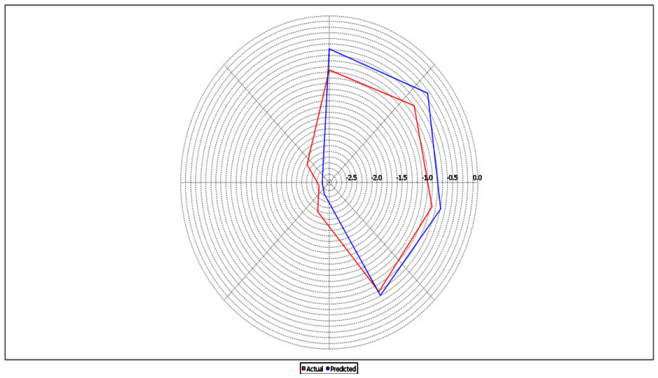


**Fig. 7: Graph between actual and predicted biological activity of training set for 2D-QSAR model**

## Model-A

pKi= H_2358 (0.1517, 0.2368);
E_1436 (2.5077, 3.1458);
E_2807 (0.0120, 3.0520);
E_3064 (0.0014, 0.9232);

H_2358, E_1436, E_2807, and E_3064 are the hydrophobic and electrostatic field energy of interactions.

## Stastistics

The model A explains values of k (2), $q^2$ (0.9494), pred_$r^2$ (0.7367), $q^2$_se (0.2037), and pred_$r^2$se (0.4931) prove that QSAR equation so obtained is statistically significant and shows the predictive power of the model is 94.94% (internal validation). Table II represents the predicted inhibitory activity by the model A for training and test set. The plot of contributions of electrostatic and hydrophobic field interactions (Fig. 8) indicates relative regions of the local fields (electrostatic and hydrophobic field) around the aligned molecules. Blue and yellow balls represent electrostatic and hydrophobic field effects, respectively. In the QSAR model, electrostatic field descriptors with positive coefficients represent regions such as phenyl rings substitutions and ortho position of sulphanilamide ring of compounds where electropositive groups are favorable. Electropositive groups like

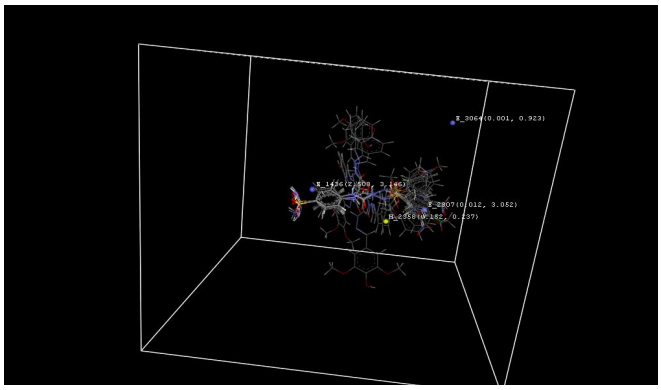


**Fig. 8: Contribution plot of 3D-QSAR model A**

alkyl, sulfo group may increase biological activity and electronegative groups may decrease activity. This can be seen with compounds (code 3 and 14) where presence of p-chlorophenyl substitution of sulfonyl semicarbazides and Schiff's bases with iminoureido, respectively, has shown less inhibitory activity. Hydrophobic field descriptors with positive coefficients represent regions near - $SO_2$ group of compounds codes 1-10 where hydrophobic groups are favorable. Hydrophobic groups like $C_6H_{11}$, Cl, -$C_6H_5$, -$C(CH_3)_3$, -$CF_3$, -$CH_3$ can increase biological activity. This can be seen with compound (code 1) where presence of sulfonyl semicarbazides with benzene substitution has shown good inhibitory activity. Thus, the contribution plot arising out of 3D-QSAR studies provide some useful insights for better understanding of the structural features of these compounds responsible for producing significant hCAII inhibitory activities[12].

Data fitness plot for model A is shown in Fig. 9. The plot of observed vs predicted activity provides an idea about how well the model was trained and how well it predicts the activity of external test set.

The graph of observed vs. predicted activity of training and test sets for model A are shown in Fig. 10 and 11.The model is able to predict the activity of training set quite well as well as external test set, providing confidence of model.

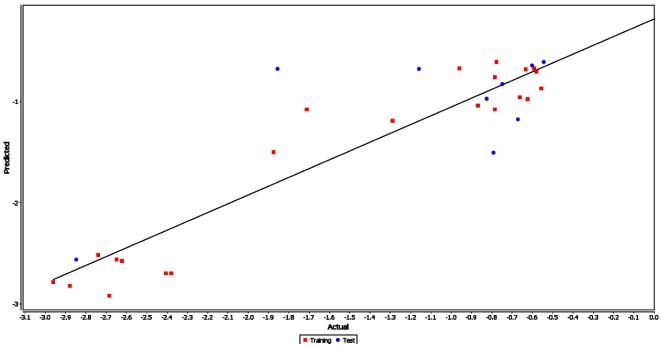


**Fig. 9: Data fitness plot for 3D-QSAR model A**

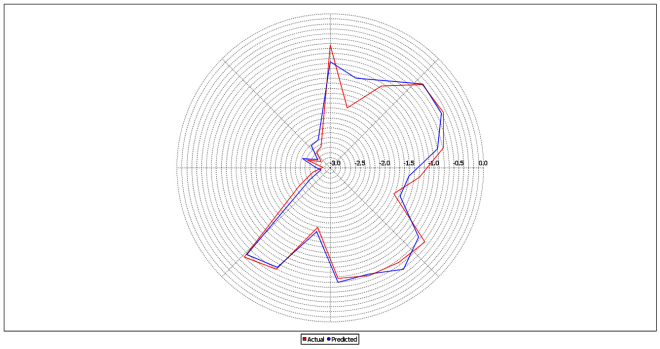


**Fig. 10: Graph between actual and predicted biological activity of training set for 3D-QSAR model A**

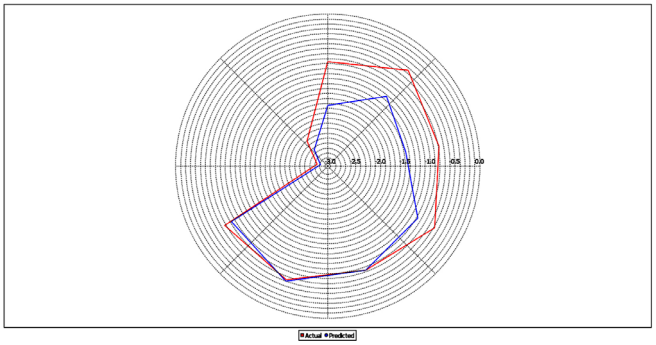


**Fig. 11: Graph between actual and predicted biological activity of test set for 3D-QSAR model A**

**Model-B**

pKi = H_2597 (0.1003, 0.1091);

S_1451 (-0.7007, -0.5335);

S_2820 (-0.0302, -0.0141).

H_2597, S_1451, and S_2820 are the hydrophobic and steric field energy of interactions.

## Statistics

[kNN= 2; n= 22; Degree of freedom= 18; q2=0.9211;

q2_se=0.2626; pred_r2=0.6055; pred_r2se=0.5585]

The model B explains values of k (2), $q^2$ (0.9211), pred_$r^2$ (0.6055), $q^2$_se (0.2626), and pred_$r^2$se (0.5585) which prove that QSAR equation so obtained is statistically significant and shows the predictive power of the model is 92.11 % (internal validation). Table II represents the predicted inhibitory activity by the model B for training and test set. The plot of contributions of hydrophobic and steric field interactions (Fig. 12) indicates relative regions of the local fields (hydrophobic and steric field) around the aligned molecules. Yellow and green balls represent hydrophobic and steric field effects, respectively. In the QSAR model, hydrophobic field descriptors with positive coefficients represent regions near - $SO_2$ group of compounds codes 1-10 where hydrophobic groups are favorable. Hydrophobic groups like $C_6H_{11}$, Cl, $-C_6H_5$, $-C(CH_3)_3$, $-CF_3$, $-CH_3$ can increase biological activity. This can be seen with the compound (code 1) where presence of sulfonyl semicarbazides with benzene substitution has shown good inhibitory activity. Steric descriptors with negative coefficients indicate regions near phenyl rings substitution on chain R and ortho position of sulphanilamide ring where the bulky substituent is disfavored. Bulky substituent like t-butyl, $-C(CH_3)_3$ may lead to decrease in biological activity. This can be seen with compound code 30 where 3,5-dimethoxy-4-hydroxy substitution on sulphanilamide connected to iminoureido Schiff base through methylene-bridge carbon has shown less inhibitory activity. Thus, the contribution plot arising out of 3D-QSAR studies provides some useful insights for better understanding of the structural features of these compounds responsible for producing significant hCAII inhibitory activities[12].

It is seen that modifications in the structure of sulfanilamide based on the information obtained from the above studies could lead to new sulfanilamides with potent hCAII inhibitory activity. Supplementary *in silico* tests for a better understanding of the mechanism of action can be carried out using molecular docking, kinetic and dynamic studies. The field is also further open for designing, synthesis, biological evaluation, pharmacokinetic studies, and clinical studies to establish those molecules as drugs.

Data fitness plot for model B is shown in Fig. 13. The plot of observed vs predicted activity provides an idea about how well the model was trained and how well it predicts the activity of external test set.

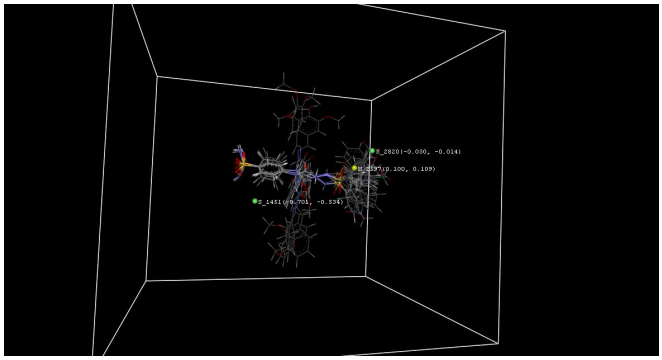


**Fig. 12: Contribution plot of 3D-QSAR model B**

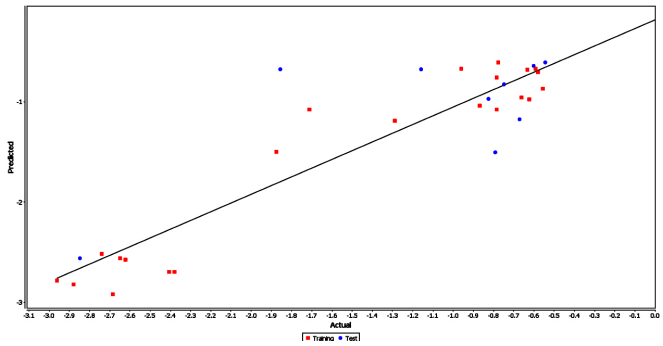


**Fig. 13: Data fitness plot for 3D-QSAR model B**

The graph of observed vs. predicted activity of training and test sets for model B is shown in Figs. 14 and 15. The model is able to predict the activity of training set quite well as well as external test set, providing confidence of model.

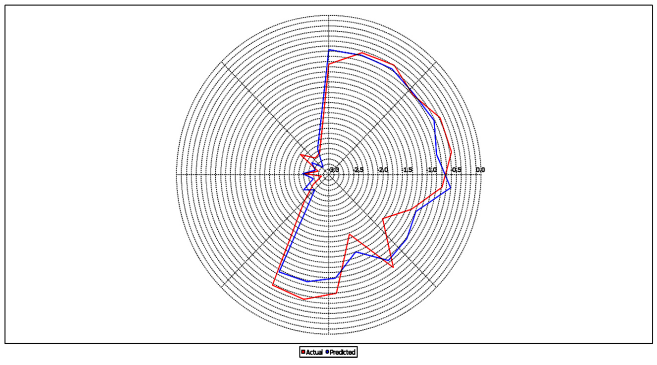


**Fig. 14: Graph between actual and predicted biological activity of training set for 3D-QSAR model B**

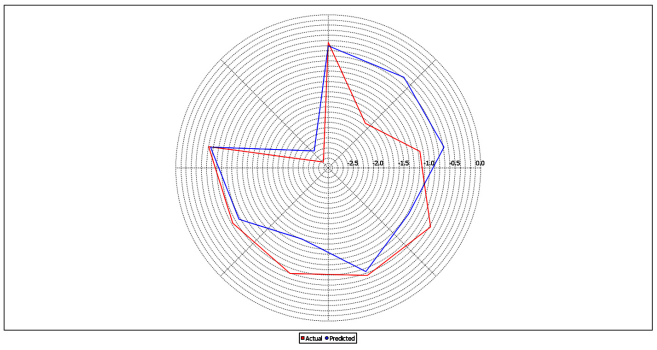


**Fig. 15: Graph between actual and predicted biological activity of test set for 3D-QSAR model B**

## CONCLUSION

2D and 3D-QSAR models were generated to understand and improve structural requirements so that inhibitory activities of some sulfanilamide derivatives against hCAII can be enhanced. The best 2D-QSAR models indicate that the descriptors of $SssCH_2$count, T_2_Cl_5, T_2_F_5, and T_T_O_7 influenced hCAII inhibition activity. With the help of 3D QSAR analysis, it becomes possible to relate chemical structures and binding affinity of ligands with respect to different bio-targets. The model A using both hydrophobic and electrostatic fields gave $q^2$= 0.9494; $q^2$_se= 0.2037; pred_$r^2$= 0.7367; pred_$r^2$se= 0.4931 values. The combination of the steric and hydrophobic model B gave q2=0.9211; q2_se=0.2626; pred_$r^2$=0.6055; pred_$r^2$se=0.5585 value. The best 3D-QSAR model was found to be model A leading to the highest $q^2$ (0.9494), pred_$r^2$ (0.7367) and the lowest $q^2$_se (0.2037). For these reasons, we considered model A (H+E) to be the best possible combination. Thus it provides a direct view of factors expressed in terms of molecular fields (electrostatic, steric, and hydrophobic) affecting the binding affinity. This in turn could give the reasonably good prediction of binding affinity. The location and range of function values at the field points selected by the models provide clues for the design of new molecules such as presence of sulfonyl semicarbazides (in compound codes 1-10)[7] showed overall best activity followed by schiff bases with iminoureido (compound codes 11-22)[8] whereas sulphanilamide connected to iminoureido schiff base through methylene-bridge carbon (compound codes 23-32)[6] results in decrease in activity. The present investigation will help to discover more potent hCAII inhibitors, with sulfonyl semicarbazides derivatives.


## ACKNOWLEDGEMENTS

The authors are thankful to Mr. Amit Bedi., Dr. Kundan Ingle and VLife MDS for helping in each and every aspect starting from software downloading to result interpretation.


## REFERENCES


1. Hassan M., Shajee B., Waheed A., Ahmad F. and Sly W.: Structure, function and applications of carbonic anhydrase isozymes. **Bioorg. Med. Chem.,** 2013, 21(6), 1570–1582.
2. Del Prete. S., Vul D., De Luca V., Carginale V., Osman S. M., Al Othman Z., Supuran C.T. and Capasso C.: Anion inhibition profiles of α-, β- and γ-carbonic anhydrases from the pathogenic bacterium Vibrio cholera. **Bioorg. Med. Chem.,** 2016, 24 (16), 1-15.
3. Vulo D., Del Prete S., Osman S., Al Othman Z., Capasso C., Donald W. and Supuran C.: *Burkholderia psedomallei* γ-carbonic anhydrase is strongly activated by amino acids and amines. **Bioorg. Med. Chem. Lett.,** 2017, 27(1), 77-80.
4. Hung H., Pan X., Tan N., Zeng G. and Ji C.: 3D-QSAR study of sulfonamide inhibitors of human carbonic anhydrase II. **Eur. J. Med. Chem.,** 2007, 42(3), 365-372.
5. Agrawal V., Bano S., Supuran C. and Khadikar P.: QSAR study on carbonic anhydrase inhibitors: aromatic/ heterocyclic sulfonamides containing 8-quinoline-sulfonyl

moieties, with topical activity as antiglaucoma agents. **Eur. J. Med. Chem.,** 2004, 39(7), 593-600.

6. Wani T., Bua S., Khude P., Chowdhary A., Supuran C. and Toraskar M.: Evaluation of sulfonamide derivatives acting as inhibitors of human carbonic anhydrase isoforms I, II and *Mycobacterium tuberculosis* β-class enzyme Rv3273. **J. Enzyme Inhib. Med. Chem.,** 2018, 33(1), 962-971.
7. Pichake J., Kharkar P., Ceruso M., Supuran C., and Toraskar M.: Carbonic anhydrase inhibitors: Design, synthesis and biological evaluation of novel sulfonyl semicarbazide derivatives. **ACS Med. Chem. Lett.,** 2014, 5(7), 793-796.
8. Singasane N., Kharkar P., Ceruso M., Supuran C., and Toraskar M.: Inhibition of carbonic anhydrase isoforms I, II, IX and XII with Schiff'S bases incorporating iminoureido moieties. **J. Enzyme Inhib. Med. Chem.,** 2015, 30(6), 901-907.
9. Khzaei A., Sarmasti N., Seyf J., Rostami Z., and Zolfigol M.: QSAR Study of the non-peptidic inhibitors of procollagen C-proteinase based on multiple linear regression, principle component regression and partial least squares. **Arab. J. Chem.,** 2017, 10 (6), 801-810.
10. Thombare S., Gonsalves S. and Shaikh A.: 2D and 3D QSAR using KNN-MFA method of pyrazolyl-thiazolinone derivatives as potential EGFR and HER-2 kinase inhibitors**. J. Comput. Methods Mol. Des.**, 2012, 2(3), 107-121.
11. Bhadoriya K., Sharma M., Sharma S., Jain S. and Avchar M.: An approach to design potent anti-alzheimer's agents by 3D-QSAR studies on fused 5,6-bicyclic heterocycles as Γ-Secretase modulators using KNN–MFA methodology. **Arab. J. Chem.,** 2014, 7(6), 924-935.
12. Sharma M. and Kohli D.: Insight into the structural requirement of substituted quinazolinone biphenyl acylsulfonamides derivatives as angiotensin II AT1 receptor antagonist: 2D and 3D QSAR approach. **J. Saudi Chem. Soc**., 2014, 18(1), 35-45.
13. Wang Y., Guo H., Tang G., He Q., Zhang Y., Hu Y., Wang Y. and Lin Z.: A selectivity study of benzenesulfonamide derivatives on human carbonic anhydrase II/IX by 3D-QSAR, molecular docking and molecular dynamics simulation. **Comput. Biol. Chem.,** 2019, 80, 234–243.
14. Baviskar B., Deore S. and Jadhav A.: 2D and 3D QSAR Studies of saponin analogues as antifungal agents against candida albicans. **J. Young Pharm.,** 2020, 12(1), 48-54.
15. Chitre T., Patil S., Sujalegaonkar A. and Dhirendra K.: Designing of thiazolidin-4-one pharmacophore using QSAR studies for Anti-HIV activity. **Indian J. Pharm. Educ. Res.,** 2021, 55(2), 581-589.