

PREDICTION OF pEC₅₀(M) AND MOLECULAR DOCKING STUDY FOR THE SELECTIVE INHIBITION OF ARACHIDONATE 5-LIPOXYGENASE

N. R. DAS¹, P. G. R. ACHARY²✉

¹Department of CSIT, Siksha 'O' Anusandhan deemed to be University, Bhubaneswar, Odisha, India;

²Department of Chemistry, Faculty of Engineering and Technology (ITER),
Siksha 'O' Anusandhan deemed to be University, Bhubaneswar, Odisha, India;

✉e-mail: pgrachary@soa.ac.in

Received: 28 April 2021; **Accepted:** 12 November 2021

Arachidonate 5-lipoxygenase (ALOX5) is considered a prime target for drug discovery in the area of liver fibrosis, rheumatoid arthritis, atherosclerosis, cancer and asthma. To date, the lead rate in the discovery of drugs that inhibit ALOX5 for the treatment of the above diseases is not satisfactory. So, the development of powerful and effective ALOX5-targeted drugs is desired. In this regard, Quantitative Structure-Activity Relationship (QSAR) and molecular docking can have a major role in screening and designing drugs. In this work, 3D-QSAR models were proposed, which were built using the techniques like Multiple Linear Regression (MLR), and Partial Least Squares (PLS) for the pEC₅₀(M) taking a diverse dataset of 112 molecules. The technique of the 'Index of Ideality of Correlation (IIC)' was also investigated to generate an optimal descriptor derived from the SMILES molecular structure. The effect of the number and nature of descriptors on the model were analyzed. The models can be helpful in providing better directions for the development of novel drug targets for 5-lipoxygenase. A significant improvement in the stability of the model was observed by the incorporation of the optimal descriptor. The molecular docking results showed that the ALOX5 receptor was well inhibited by the 112 ligands showing the least binding energy (-10.8 Kcal/mol). In order to validate the binding mode of the ligands docked with AutoDock Vina software, the top-scored compounds were re-docked using DockThor online docking server. The results obtained from docking suggest that the ligands with IDs 18, 20, 24, 30 and 44 are some of the potential inhibitors for ALOX5.

Key words: 5-lipoxygenase, QSAR, SMILES, molecular docking.

5-Lipoxygenase (5-LO or ALOX5) is an important enzyme that helps in producing pro-inflammatory mediators, like leukotriene B₄ and cysteinylleukotrienes. Inhibition of ALOX5 can be a potential remedial approach for inflammation. At present, the orally active inhibitor for the ALOX5 is Zileuton brought by Abbott Laboratories(1996). The other brand names associated with Zileuton are Zylflo and Zylflo CR. However, it is reported this drug is known to cause liver diseases or liver toxicity [1]. The discovery of potent drugs that inhibit ALOX5 drug targets, without any harmful side effects like liver toxicity is a challenging task.

QSAR is an important chemometric method that is widely used in virtual screening to discover new leads [2-6]. A number of high-quality research papers based on QSAR study are available in literature [7-12]. An inclusive study on the present system of ALOX5 is necessary for the discovery of potent drugs without any side/or toxic effects. However, the literature review revealed, it is now gaining the right attention and is all set to achieve the momentum. Five good QSAR models for developing benzoquinone derivatives as ALOX5 inhibitors using CoMFA, RF, MLR, and SVM were described in [13]. Another group of authors has used QSAR

models to indicate the significance of the chemical characteristics for the ALOX5 inhibition for a sequence of coumarin derivatives [14]. The binding affinity and interactions with the active sites of human 5-LOX were estimated by the molecular docking study. Computational methods were implemented to get 5-LOX inhibitors and to screen chalcone and flavones derivatives [15]. The potential hits from grid-based ligand docking with energetics were re-docked using Genetic Optimization. 3D QSAR models were developed using IC_{50} values of 51 compounds for analyzing the biological activities of the inhibitors of 5-LOX with $R^2 > 0.75$ [16]. In the literature, QSAR study on EC_{50} values for the ALOX5 inhibition has not been reported until now.

In this communication, QSAR models were built using MLR and simplified molecular input line entry system (SMILES). A good combination of 2D, 3D and optimal descriptor is evaluated for the prediction of pEC_{50} values for the ALOX5 receptor. In our recent publication the prediction of pEC_{50} values for Adenosine A2A Receptor was reported, where the incorporation of the optimal descriptor showed better performance [17]. In this article, the combination of 2D, 3D and optimal descriptor is evaluated for the prediction of pEC_{50} values for the ALOX5 receptor. Molecular docking has also been performed to find the active binding sites available in ALOX5 receptor.

Molecules, Software Codes and methods

Data set preparation and data reduction. The dataset contained EC_{50} (nM) values of 112 different inhibitor compounds for the ALOX5 receptor derived from the popular Binding database [18]. The EC_{50} values converted to their pEC_{50} equivalent (negative decimal logarithm of EC_{50}). Spartan-10 and OpenBabel were used to generate the SMILES and MDL (.mol) chemical structures for these molecules from the SDF (structure data file).

Descriptors. PaDEL_2.18 was used to obtain more than 900 molecular 2D and 3D descriptors. CORAL software [19-21] was used to generate the optimal descriptors based on SMILES. Preliminary scans were performed using 100 descriptors each time. The preliminary scan was performed to identify high correlation descriptors for ALOX5.

Optimal descriptor and Index of Ideality of Correlation (IIC). The application of 'Index of Ideality of Correlation (IIC)' in QSAR/QSPR is elaborated in the literature [27-31]. The established principle of IIC is obtained from the distribution of

points in the graph plotted between experimental and observed pEC_{50} values [32]. The optimal descriptor (DCW) can be obtained by using the following mathematical relationships:

$$DCW(T^*, N^*) = DCW_s(T^*, N^*) \quad (1)$$

$$DCW_s(T^*, N^*) = CW(HARD) + \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) + \sum_{k=1}^{NA-2} CW(SSS_k) \quad (2)$$

$$TGF = T - F_1 + IIC_{CLB} \times 0.1, \quad (3)$$

where HARD, S_k , SS_k and SSS_k are parts of SMILES code of each molecule [31, 33, 34] and the calculation of IIC_{cb} is represented elsewhere [28-31]:

$$IIC_{cb} = r_{cb} \frac{\min(-MAE_{cb}, +MAE_{cb})}{\max(-MAE_{cb}, +MAE_{cb})} \quad (4)$$

$$-MAE_{cb} = \frac{1}{-N} \sum_{k=1}^{-N} |\Delta_k|, \quad \Delta_k < 0; \text{ } ^-N \text{ is the number of } \Delta_k < 0 \quad (5)$$

$$+MAE_{cb} = \frac{1}{+N} \sum_{k=1}^{+N} |\Delta_k|, \quad \Delta_k \geq 0; \text{ } ^+N \text{ is the number of } \Delta_k \geq 0 \quad (6)$$

$$\Delta_k = observed_k - calculated_k \quad (7)$$

The calculation of DCW for the SMILES molecular structure: Clc4cc(CCC3c(c(=O)OCC)c1c(c2ccccc2c(O)c1)[nH]3)ccc4 (ID-1) is supplied in the Table 6S of the Supplementary file. The individual correlation weight (CWs) of each smiles attribute (SA_k) such as $Cl... = 0.3084$, $c.. = -0.0554$, similarly (SA_{kk}) $c...Cl..... = 0.3995$ and so on. Therefore for the full SMILES code, the DCW was calculated using the eq. 1, which became 4.54963. The extrapolative QSAR model on the pEC_{50} for ALOX5 could be modelled by the following simple mathematical linear equation:

$$pEC_{50} = C_0 + C_1 * DCW(T^*, N^*) \quad (8)$$

The above eq. 8 has been used to develop the QSAR model in this work which is described in section 3.2. The optimal parameter DCW has also been used to develop hybrid QSAR model detailed in section 3.3.

Molecular Docking. Molecular docking is one of the most outstanding techniques used in recent days for the realistic design of drugs. Docking deals with the process of binding drugs with the protein by determining active binding sites. Since molecular docking facilitates the realization of the biological activities of the drug targets, these days it has become a widely used technique in discovering effective drugs for a variety of diseases. In this article, the development of the therapeutic activities of the above-mentioned 112 different inhibitors against ALOX5 has been analyzed through the docking process. Molecular docking was carried out to estimate the binding energies and site interactions to evaluate the inhibition potential of the ALOX5 main protease.

The .pdb files for the ligands were prepared using OpenBabel (ver. 2.3.2). The crystallographic structure of ALOX5 was taken from the Protein Data Bank. The PDB code of the protein structure was 3O8y. For docking purposes, the protein structure derived from the data bank and then was prepared by removing all water molecules, adding polar hydrogen and Kolman charges. Molecular docking was executed with the help of Autodock vina [38] and some other Auto Dock Tools. Docked structures and interaction of the ligand with the protein residues in the active sites were analyzed by Discovery Studio Visualizer.

QSAR Models for ALOX5 Inhibition

QSARINS (QSAR-Insubria) software [22] code developed at the University of Insubria was applied to build the desired robust QSAR models. Since more than 900 different descriptors were there and taking all descriptors simultaneously to build the model requires heavy computational time, preliminary scans were performed using 100 descriptors in each case to identify potential descriptors for the 5-LOX receptor. Finally, a descriptor set of 33 molecules with their pEC50(m) was chosen to construct QSAR models using GA. The 112 ligands dataset for the ALOX5 receptor was randomly distributed into four different sets such as the training, invisible training, calibration and validation sets as the robustness of a model basically depends on the quality of the training set and during the training process, the decrease in performance for a model due to over-training can be significantly removed by the calibration set [23-25]. Once, these processes of the model building were over, the performance of the QSAR models was further verified by an external validation set.

A linear regression model was formed between the response variable pEC50 and the descriptors using the ordinary least squares (OLS) method. The models were arranged in accordance with their R^2 , Q^2 , R^2-Q^2 and RMSE values. Internal and external validation methods and principal component analysis were explored to verify the robustness and predictability of the constructed models. The model with the minimum value of R^2-Q^2 was considered more stable.

2D-QSAR models. The semi-empirical (AM1) quantum chemical calculation was used for the geometrical optimization of the molecules. After the optimization process, mathematical models have been developed with a good collinearity between the descriptors with the endpoint.

The 2D-descriptors like MATS3c, SpMax1_Bhp and ATSC6s showed excellent correlation with the experimental pEC50 values for ALOX5. The model built with these descriptors mathematically described in the following equation.

$$pEC50 = 17.1019 + 1.8165(ATSC6s) + 4.7179(MATS3c) - 4.9211(SpMax1_Bhp) \quad (9)$$

Internal and external validation parameters for these models and performance towards the validation set are presented in Table 1, 2 and 3. The R^2 for the training set is 0.7340 (Eq. 9) and for the validation set is 0.9481 (Eq. 9). For a robust model Average R^2_m should be greater than 0.5 and ΔR^2_m should be lower than 0.2 [26], where as k and k' should be in the range of 0.85 and 1.15. For the above model (Eq. 9), the values of Average R^2_m and ΔR^2_m , k and k' are within the required ranges. The model qualified the required internal and external validation characteristics to justify that this is a robust model.

Single Optimal Descriptor Based QSAR Models. The model based on the optimal descriptor defined in section 2.3 is described in Eq. 10. For this model the optimal descriptor DCW was first determined from their SMILES Attributes (SAs). The calculation of these type of descriptors is described in literature [23, 34-36]. This model displayed very good statistical parameters (Table 1-3). However, to describe any model as robust only these statistical parameters such as R^2 (test set), Q^2 (test set), etc. are not sufficient. Another validation characteristic like $^cR^2p$ is also needed [37]. For good models, the value of $^cR^2p$ should be greater 0.5. For the model defined by Eq. 10 the $^cR^2p$ was found to be 0.8718 (training), 0.9443 (validation) and 0.7971 (test set).

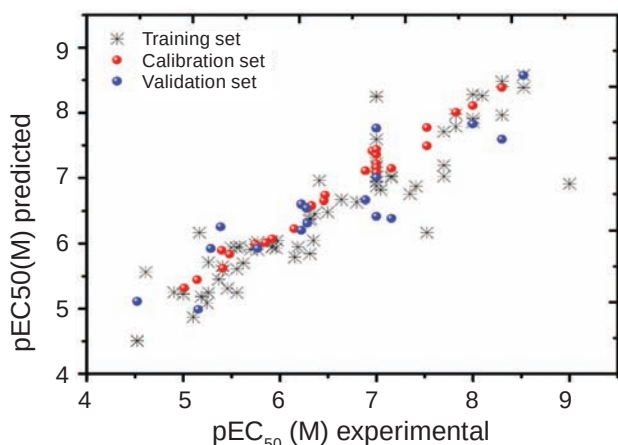


Fig. 1. Experimental VS predicted pEC_{50} values by Eq. 10

$$pEC_{50} = 6.7556863 (\pm 0.0067206) + 0.0755563 (\pm 0.0003592) * DCW(1,7) \quad (10)$$

The plot of the experimental and predicted pEC_{50} obtained by Eq. 10 is given in the Fig. 1.

Hybrid QSAR models(2D,3D& DCW). Some models were built using 2D descriptors, 3D descriptors and the optimal descriptor introduced in the previous section. Fig. 2(a) shows R^2 and Q^2 values for the training dataset without the optimal descriptor. From this figure, it is clear that as the number of variables is increased, the R^2 and Q^2 values rise up to a maximum of six variables then a decrement in Q^2 is detected. The maximum R^2 (training set) is 0.7340 for the three variable QSAR model (Eq. 9) with R^2_{ext} 0.9481 (Eq. 9). Increasing the number of descriptors (more than 3) shows a marginal increase in the R^2 (training set) with a significant reduction in the internal and external validation characteris-

tics. Such a type of performance occurs due to the overtraining of the dataset. Therefore the optimal descriptor has been introduced in the model. The graph for R^2 and Q^2 with the optimal descriptor is presented in Fig. 2(b) which shows a significant enhancement in the R^2 and Q^2 values.

Principal Component Analysis (PCA) for the descriptors E_{aq} (kJ/mol), XLogP, E HOMO (kJ/mol), MATS6c and DCW(1,7) was studied by means of the score plots and loading plots. Fig. 3 is the score plot for the descriptors E_{aq} (kJ/mol), DCW(1,7), XLogP and MATS6c which form the hybrid model (Eq. 11). It clearly shows that molecule number (ID = 77) is an outlier. Similarly, PCA Loading plot for the descriptors Eq. 11 is given in Fig. 4 for the above four descriptors.

One of the robust QSAR models built with DCW(1,7) as one of the descriptors is defined as follows:

$$pEC_{50} = 5.7771 + 0.0001 (E_{aq} \text{ (kJ/mol)}) + 0.0727 (DCW(1,7)) - 0.0642 (XLogP) + 0.8800 (MATS6c) \quad (11)$$

For the model presented in this section (Eq. 11) the R^2 for the training set is 0.8681. The R^2 for the validation set is 0.9762. Tables 1, 2 and 3 show the internal and external validation parameter values for this model. The values of average R^2_m , ΔR^2_m , k and k' are within the required ranges confirming the robustness of this hybrid model. The value of R^2 shows a good fit for modelling ALOX5 inhibition. The LOF is very small which makes sure that there is no overfitting. The low value of Kxx specifies that the correlation between the model descriptors is very less resulting in a model having the least redundant in-

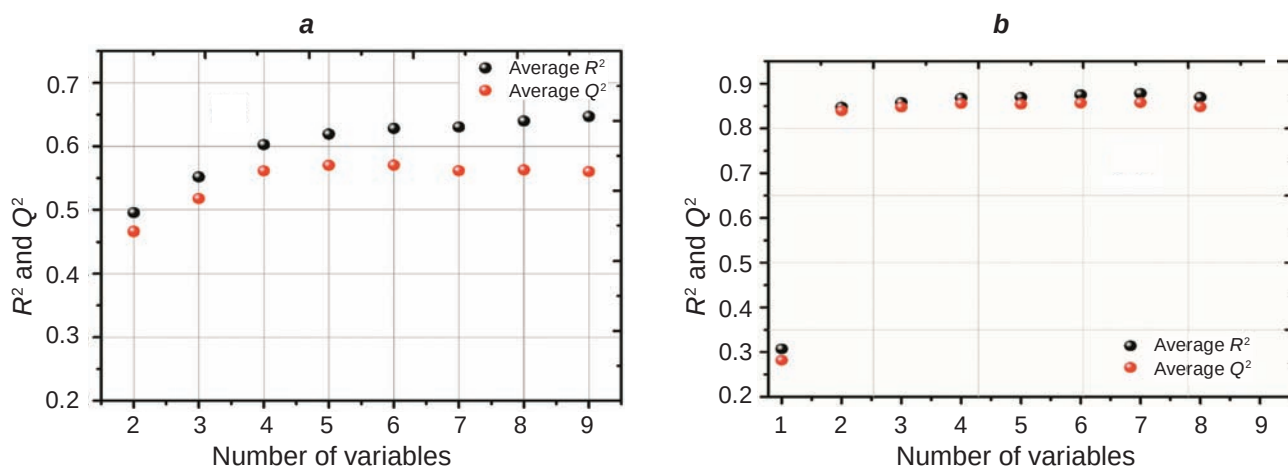


Fig. 2. **a:** R^2 and Q^2 for the models without DCW, **b:** with DCW

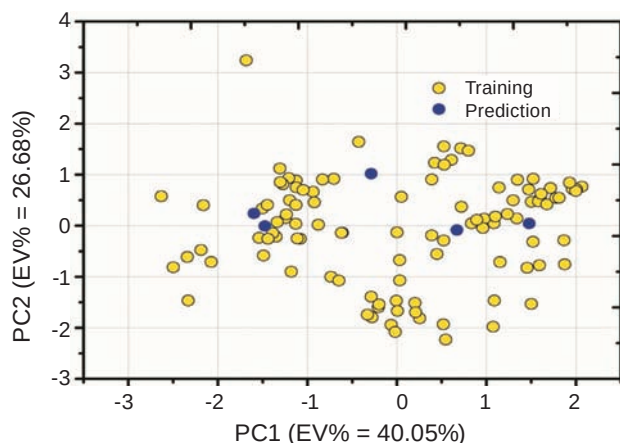


Fig. 3. Score plot for model 8 (Eq. 11)

formation in the descriptors. Fig. 5(a) demonstrates a comparison between the predicted values and the experimental values of pEC50 for the training dataset.

Model validation according OECD principles

OECD principles were used to ascertain the efficiency of the QSAR models proposed in this work. According to these principles, the models should have a definite endpoint. The endpoint for the described models is pEC50. The second principle says that models should be represented using a definite algorithm that can derive a proper relationship between the descriptors and the endpoint. The algorithms used to obtain such a relationship here are MLR and OLS. The third principle states that the models can have reliable predictions with leverage values below the critical leverage with ± 3 standard deviations. William graph was used to represent the applicability domain (DA) of the models. According to the fourth principle, the difference between the experimental values and forecast values should be the minimum. The difference between the experimental values and the values predicted by the models was very low. The goodness of the fit of the models was measured with the coefficient of determination (R^2) and adjusted R^2 (R^2_{adj}). R^2 is used to compare between the predicted and experimental activities. The difference between the R^2 and R^2_{adj} value for the defined models were less than 0.3 which indicates that the number of descriptors involved in the QSAR model is acceptable. The value for R^2_{adj} indicates the ease of adding a new descriptor to the model. The fit of the QSAR models can be determined by root-mean-squared error (RMSE). This method is used to

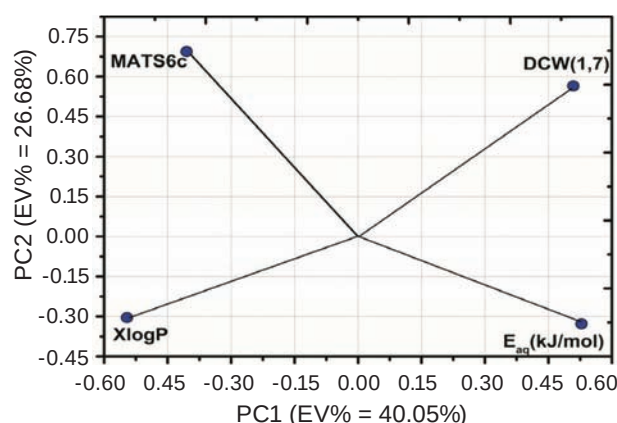


Fig. 4. PCA loading plot for model 8 (Eq. 11)

decide if the model possesses the predictive quality reflected in the R^2 . The use of RMSE shows the error between the mean of the experimental values and predicted activities. RMSE values greater than 1.0 reflect the model's poor ability to accurately predict the bioactivities. The models defined in this work are having RMSE less than 1.0 indicating higher predictive power of the models. Q^2_{LOO} and LMO internal validation methods were used to confirm the potency of the defined models. The Y-scrambling method was employed to prove that the models are not the result of a chance correlation.

The validity of the models was evaluated by the OECD principles and their regulatory values. The models were validated by the LOO and LMO internal validation methods. The obtained results authenticate the internal predictions as the value estimated by LOO (Q^2_{LOO}) is almost the same as the R^2 value signifying the reliability of the defined models. The error in the predictions is very low. Fig. 5(b) presents the similarity between the experimental values and values estimated by LOO (leave-one-out). The Leaving-Many-Out (LMO) method that leaves out thirty percent of the dataset to evaluate the performance of the models is very helpful in the validation process as each deviation of data is treated as important, unlike LOO.

The statistical validation parameters for the defined models are presented in Tables 1, 2 and 3. For good predictability R^2 - Q^2 value should not exceed 0.3. It can be seen in Table 2 that, the difference between R^2 and Q^2_{LMO} is very less than 0.3, authenticating the models as robust.

From the statistical measures, it is clear that the QSAR models defined in this work satisfy both the

Table 1. Fitting parameters

Model	R^2	R^2_{adj}	$R^2-R^2_{\text{adj}}$	LOF	RMSE _{tr}	MAE _{tr}	RSS _{tr}	CCC _{tr}	s	F	Kxx
Eq. 9	0.7340	0.7261	0.0080	0.3255	0.5376	0.4409	30.0553	0.8466	0.5482	91.99	0.3366
Eq. 10	0.8434	0.8419	0.0015	0.1855	0.4227	0.2794	19.1186	0.9151	0.4267	565.65	0
Eq. 11	0.8681	0.8629	0.0052	0.1759	0.388	0.2611	16.1121	0.9294	0.3974	167.76	0.2216

Table 2. Cross validation parameters

Model	Q^2_{LOO}	$R^2-Q^2_{\text{LOO}}$	RMSE _{cv}	PRESS _{cv}	MAE _{cv}	CCC _{cv}	Q^2_{LMO}	$R^2_{Y\text{-ser}}$	RMSE _{AVY-ser}	$Q^2_{Y\text{-ser}}$
Eq. 9	0.7109	0.0231	0.5604	32.6649	0.4592	0.8335	0.7049	0.0284	1.0274	-0.0514
Eq. 10	0.8385	0.0050	0.4294	19.7261	0.2842	0.9124	0.8355	0.0095	1.0632	-0.0286
Eq. 11	0.8567	0.0114	0.4044	17.5027	0.2735	0.9231	0.8552	0.0379	1.0477	-0.0611

Table 3. External validation parameters

Model	RMSE _{ext}	MAE _{ext}	PRESS _{ext}	R^2_{ext}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC _{ext}	$R^2_m(\text{Avg})$	ΔR^2_m
Eq. 9	0.3338	0.3289	0.5572	0.9481	0.9270	0.8886	0.8974	0.9279	0.7109	0.0977
Eq. 10	0.1875	0.1299	0.1757	0.9906	0.9436	0.9419	0.9692	0.9639	0.7383	0.0722
Eq. 11	0.1932	0.1426	0.1867	0.9762	0.9401	0.9382	0.9673	0.9621	0.7482	0.0755

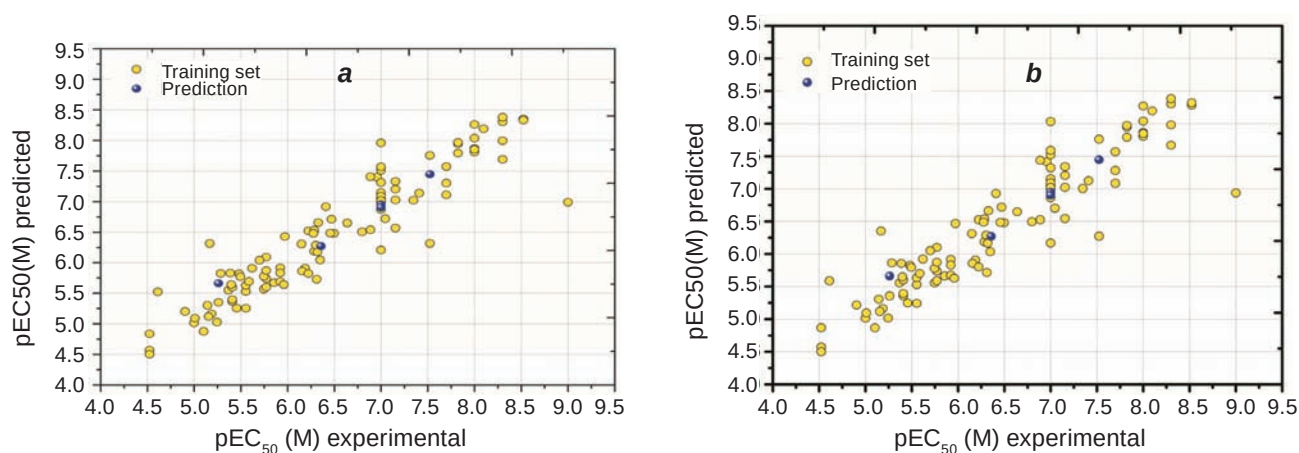


Fig. 5. **a:** Predicted values versus the experimental values for pEC_{50} (Eq. 11). **b:** Values calculated by LOO versus the actual experimental values of pEC_{50} (Eq. 11)

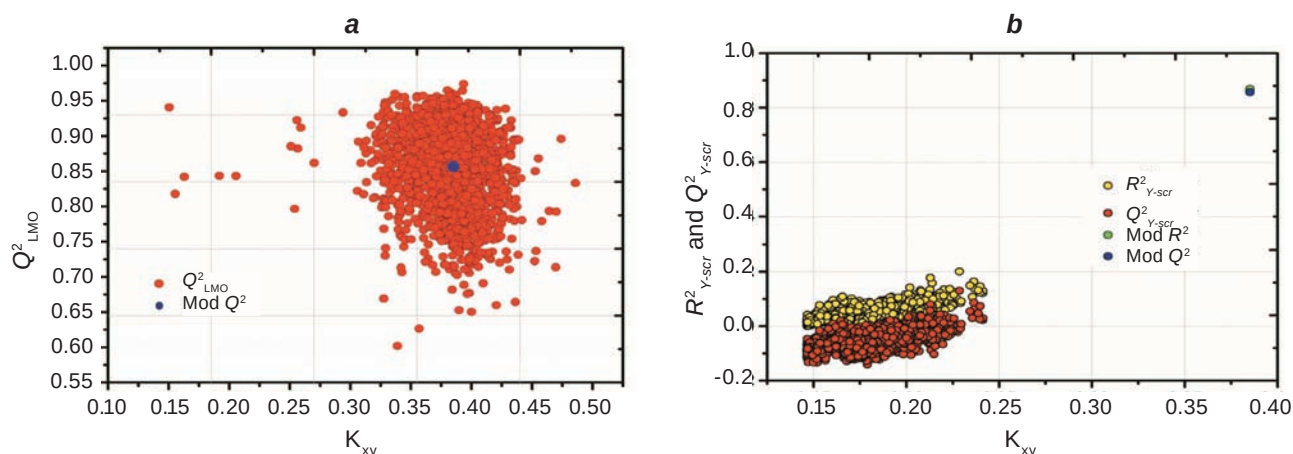


Fig. 6. **a:** Q^2_{LMO} values against $mod Q^2$ (Eq. 11). **b:** R^2_{Y-scr} and Q^2_{Y-scr} values against R^2 and Q^2 (Eq. 11)

internal and external validation criteria as required. Moreover, the models with the optimal descriptor are showing better results and therefore they can be regarded as robust and can be considered for further applications in drug discovery.

Fig. 6(a) shows the correlation between descriptors and pEC_{50} inhibition (K_{xy}). From the figure, it can be observed that the values of Q^2_{LMO} are very alike authenticating the models as a good fit. The Y-scrambling method has been carried out to exhibit that the models are not the result of casual correlations. The low values of R^2_{Y-scr} and Q^2_{Y-scr} indicates the robustness of the developed models. The R^2_{Y-scr} and Q^2_{Y-scr} values against R^2 and Q^2 are presented in Fig. 6(b). The R^2 and Q^2 values are far away from the R^2_{Y-scr} and Q^2_{Y-scr} values confirming the nonexistence of random correlation in the model. The extrapolative capability of the models was evaluated with

some external prediction tools such as R^2_{ext} [37], RMSEext (RMSEext is the Root Mean Square Error in external prediction), MAEext (Mean Absolute Error in external prediction), PRESSext (Predictive Residual Sum of Squares in external validation), Q^2_{F1} , Q^2_{F2} , Q^2_{F3} , CCC_{ext} (CCC in external prediction), average R^2_m and ΔR^2_m . These values are similar to the values calculated by the training set. Since the predictions that are within the applicability domain (AD) are considered reliable the approach of leverage (h) and standardized residuals were also applied here to present the AD of the models. The leverage value for the defined hybrid model is calculated as 0.140. Fig. 7(a) presents William's graph for Eq. 11 which shows that the majority of the compounds are within the AD of the model. Fig. 7(b) is the William's graph calculated by LOO for the same model. In both graphs, the molecule with ID = 77 is an outlier. The

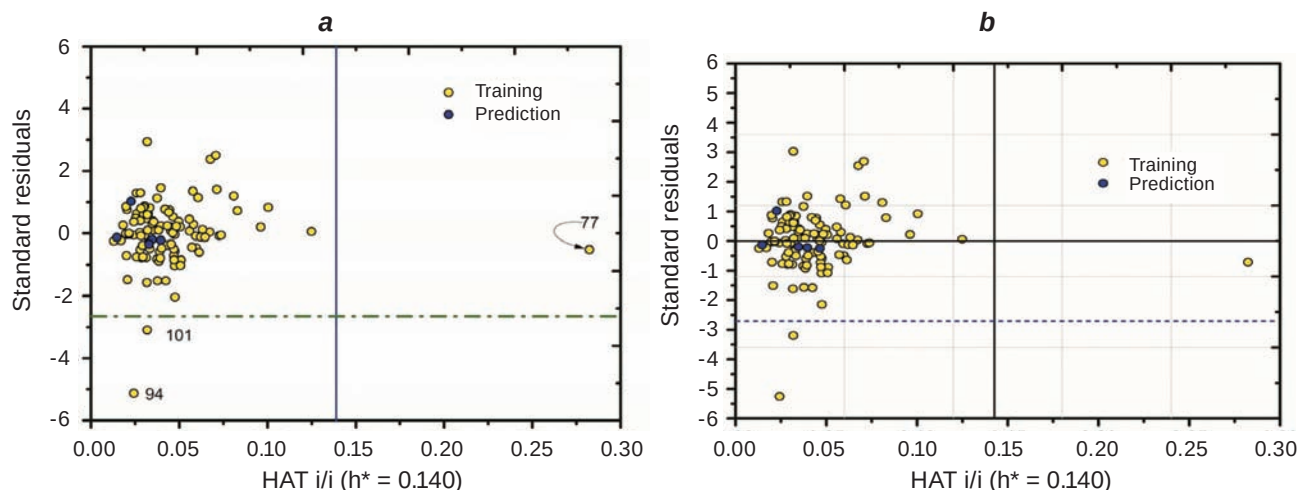


Fig. 7. **a:** William's graph with outlier (Eq. 11). **b:** William's graph by LOO (Eq. 11)

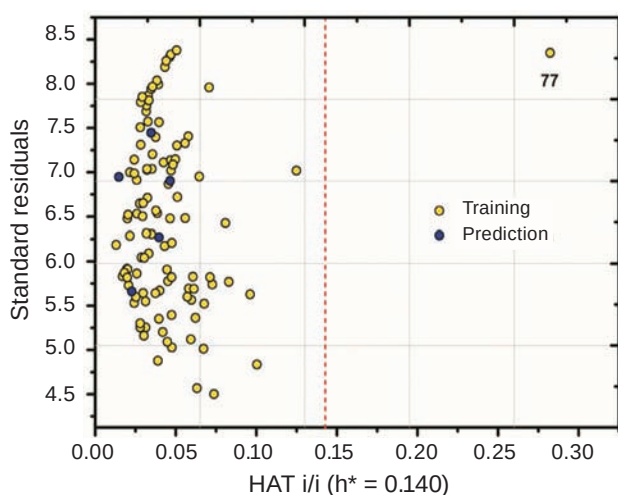


Fig. 8. Graph of Insubria (Eq. 11)

prediction of the outlier by William's plot justifies and gives a second confirmation after the prediction by the score plot obtained by the PCA study (Fig. 3). The graph of Insubria (Fig. 8) of QSARINS facilitates visualization of the model's AD. It can find out the molecules lacking experimental response. Here, it is quite similar to William's graph. The graph of Insubria is also indicating that molecule number 77 is an outlier.

Docking Results

The results of the docking process described in this section include the docking scores for different compounds with different EC values. The highest negative binding energy indicates the best docking score.

Table 4 shows the interaction between the protein and different ligands in 3D. It provides the docking scores of the interaction of the protein with different ligands. To validate the docking results of Autodockvina the protein receptor was re-docked with the ligands which scored high ranks in the docking process performed with Autodock Vina. The re-docking was performed using DockThor online docking server. The ordering of the ligands on the basis of their scores is the same for both the docking protocols, which enhances the validity of the docking results obtained from the first attempt using AutodockVina. Table 5 shows docking scores for different conformations of the ligand 30. The scores are quite similar indicating favourable interactions between the binding sites and the ligand.

The 2D figures of the ligand's interaction with the active site residues of the protein target are presented in this section (Fig. 9-12). PHE450, GLN549, TYR470, ALA453, SER447, ARG370, ALA456, ARG457, VAL243, ARG246, LEU244, VAL361, LEU288, ASP285, and GLU287 are found to be the active site residues of the receptor. Hydrogen bonds are a primary contributor factor in supporting the binding affinity of drugs with the receptor. Strong hydrogen bonding interaction represents a high binding capability between the ligand and the protein. The ligands 30, 20, 18, 24 and 44 have shown a strong binding affinity towards the receptor by forming hydrogen bonds. Strong hydrogen bonds have bond angles close to 170 or 180 degrees. Some characteristics of the h-bonds, such as distance, the bond angles between the donor atom and the acceptor atom, the name of the donor and acceptor atom

Table 4. Docking Scores of different compounds

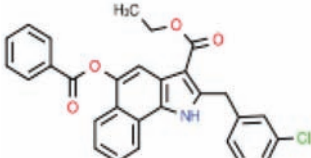
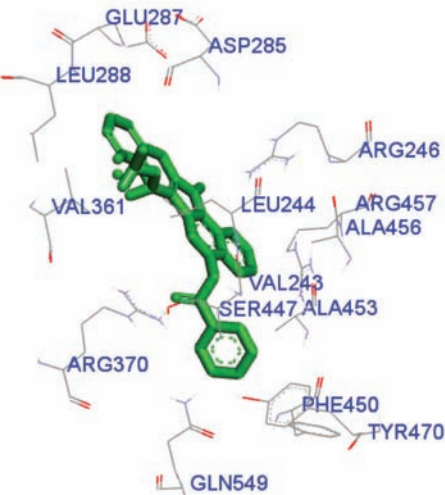
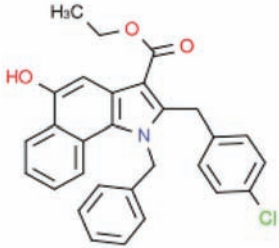
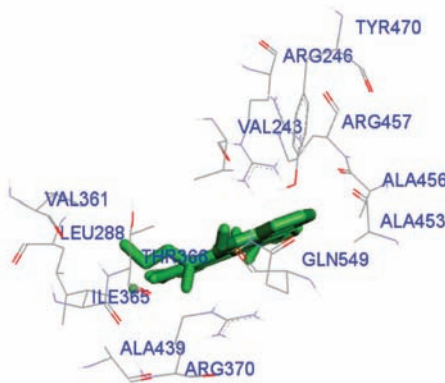
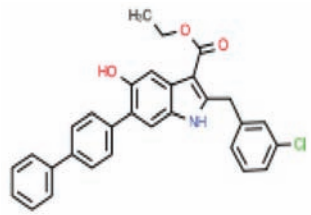
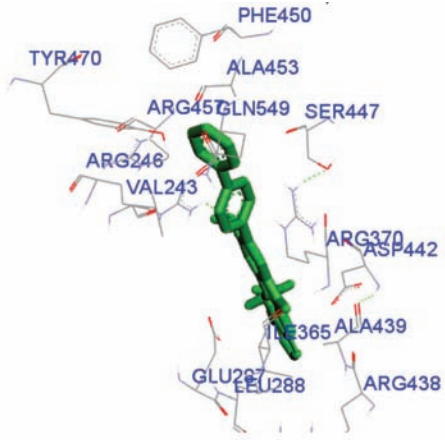
ID	2D structure of the ligand	Ligand position in Protein-ligand Interaction	EC50 (nM)	Score (kcal/mol) Autodock Vina	Score (kcal/mol) Dock Thor
30			>10000	-10.8	-7.9
22			1800	-10.6	-7.5
20			2800	-10.3	-7.5

Table 4. Docking Scores of different compounds (Continued)

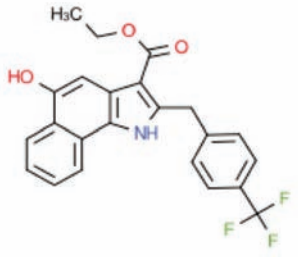
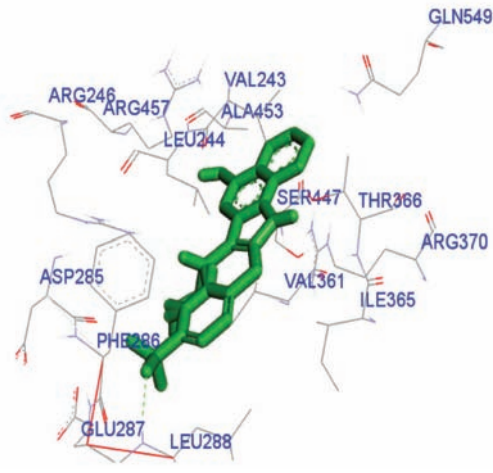
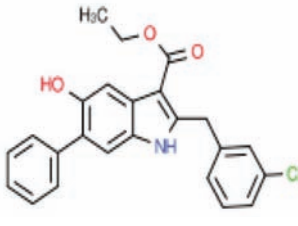
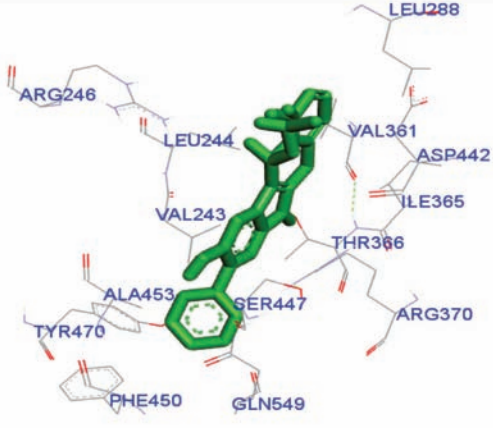
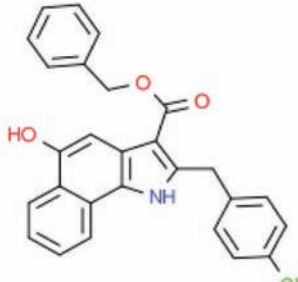
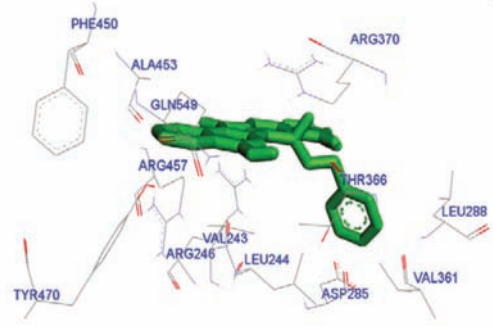
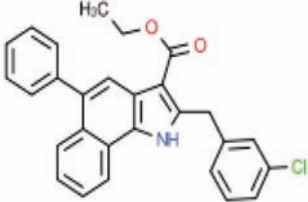
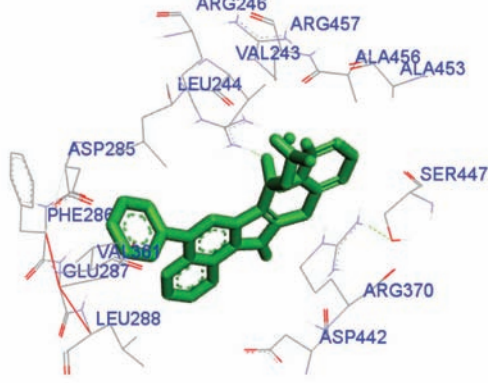
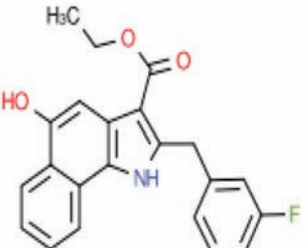
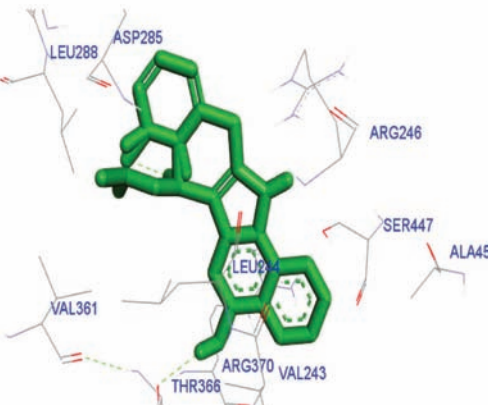
ID	2D structure of the ligand	Ligand position in Protein-ligand Interaction	EC50 (nM)	Score (kcal/mol) Autodock Vina	Score (kcal/mol) Dock Thor
18			1700	-10.2	-7.4
24			5700	-10.2	-7.4
16			480	-10.1	-7.4

Table 4. Docking Scores of different compounds (Continued)

ID	2D structure of the ligand	Ligand position in Protein-ligand Interaction	EC50 (nM)	Score (kcal/mol) Autodock Vina	Score (kcal/mol) Dock Thor
44			>30000	-10	-7.4
13			340	-9.8	-7.1

are presented in Table 6. Almost all the bond angles are above 120 degrees and close to 170 degrees confirming the strength of the bonds formed between the receptor and the ligands. From all the active residues the aminoacids ASP442, ARG246, THR366, ARG370 and LEU288 are forming hydrogen bonds with the interacting ligands. The qualitative aspect of this interaction is that hydrogen bonds are distributed over the sides and centre of the molecule, which represents high inhibition efficiency to bind the receptor-binding domain. The formed hydrogen bonds were in the categories of strong and moderate hydrogen bonds (1.76-2.60 Å) showing the high binding potential of the ligands for the receptor.

Conclusion. Computational techniques for estimating the activities of ALOX5 inhibitors can smooth the progress of the drug design process by reducing cost and time. In the present communication,

successful three sets of QSAR models were presented. The first model was built using 2D-Descriptors (MATS3c, ATSC6s and SpMax1_Bhp); the second model was built using a single optimal descriptor (DCW) and the third model was built using some 3D-descriptors [E HOMO (kJ/mol)], XLogP, Eq (kJ/mol)] along with the DCW descriptor and one of the above discussed 2D descriptors. The models fulfil all regulatory principles established by OECD; the robustness of the model was tested through internal validation techniques (LOO, LMO and Y-scrambling), and the predictability of the models was determined with an external prediction set. The presented MLR based QSAR models provide an added mode of control to screen, check and develop better drug candidates. The study of PCA, William's Plot, graph of Insubria (AD) were helpful in identifying the outliers in the dataset. The incorporation of the

Table 5. Scores of different confirmations for the ligand-30

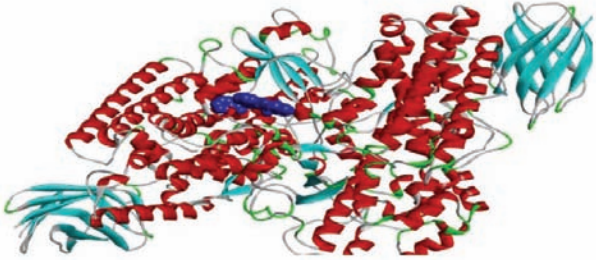
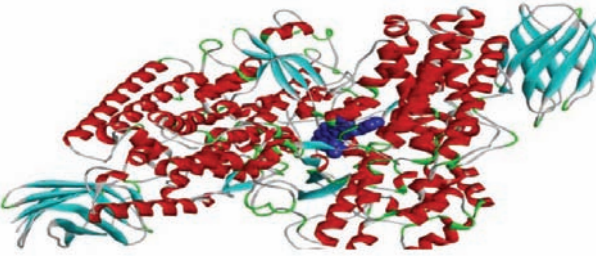
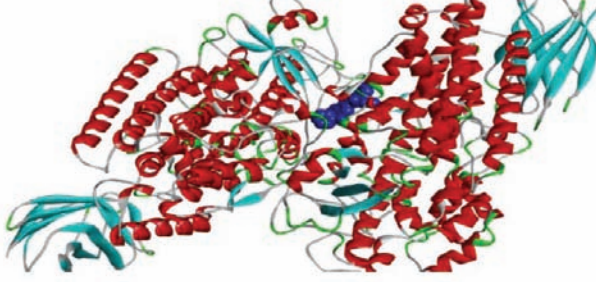
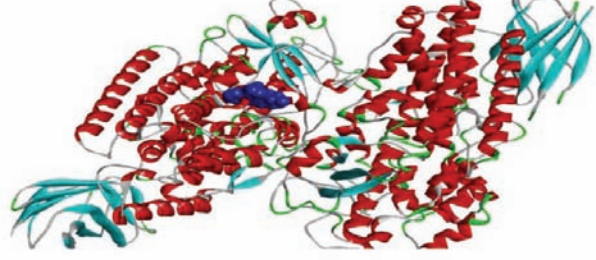
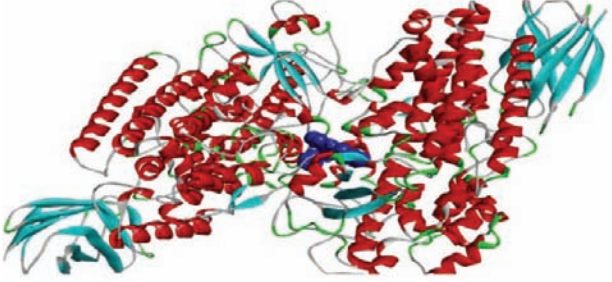
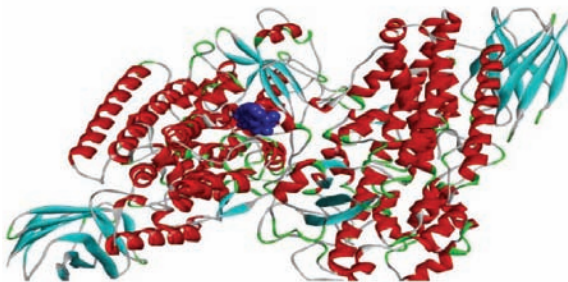
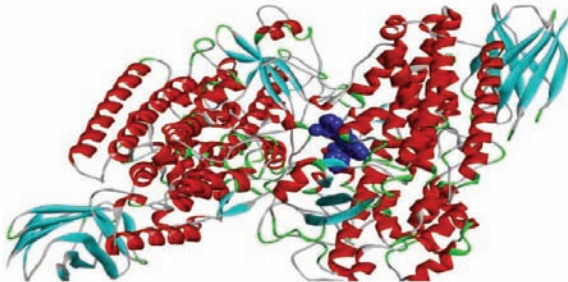
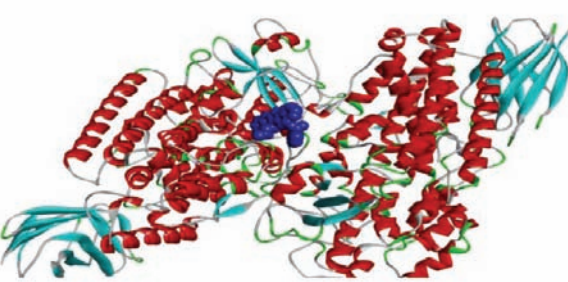
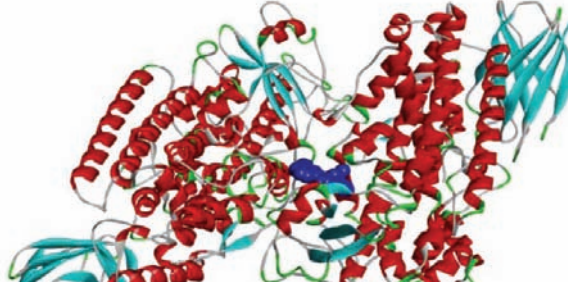
Confirmations	Position of the ligand	Score
1		-10.8
2		-10.8
3		-10.5
4		-10.4
5		-10.2

Table 5. Scores of different confirmations for the ligand-30 (Continued)

Confirmations	Position of the ligand	Score
6		-10.2
7		-10.2
8		-10.0
9		-10.0

DCW descriptor in the model building decreased the number of descriptors, registered impressive statistical parameters, internal validation characteristics and external validation characteristics. The proposed models can help to screen large databases to generate leads for the ALOX5 receptor. From the statistical parameters presented in Tables 1, 2 and 3 it can

be concluded that the use of the DCW descriptor along with the 2D, 3D descriptors is a good move in designing stable QSAR models, which in turn can find applications in drug design, screening and virtual screening, etc. The molecular docking computations provide evidence that compound 30 has the minimum binding energy while interacting with the

Table 6. Distances of the h-bonds formed between the receptor and the ligands

Ligand	HBondName	Distance (Å)	Donor atom	Acceptor atom	Angle DHA (Donor hydrogen acceptor)
30	A:ARG370:HH22 - A:SER447:OG	2.05485	HH22	OG	163.645
20	B:ARG246:HH22-:UNL1:O	2.35526	HH22	O	120.457
	B:ARG370:HH22 - B:SER447:OG	2.06678	HH22	OG	163.094
	B:ASP442:HN- B:ALA439:O	2.24637	HN	O	159.556
18	B:LEU288:HN - :UNL1:F	2.40638	HN	F	142.639
	B:THR366:HG1 - B:ASN241:OD1	2.08457	HG1	OD1	154.742
24	:UNL1:H - A:THR366:OG1	2.49049	H	OG1	152.245
	A:THR366:HN - A:VAL361:O	2.24918	HN	O	161.915
	A:THR366:HG1 - A:ASN241:OD1	2.07258	HG1	OD1	155.233
	A:ARG370:HH22 - A:SER447:OG	2.05485	HH22	OG	163.645
44	B:ARG246:HH22 - :UNL1:O	2.2936	HH22	O	119.086
	B:ARG370:HH22 - B:SER447:OG	2.06678	HH22	OG	163.094

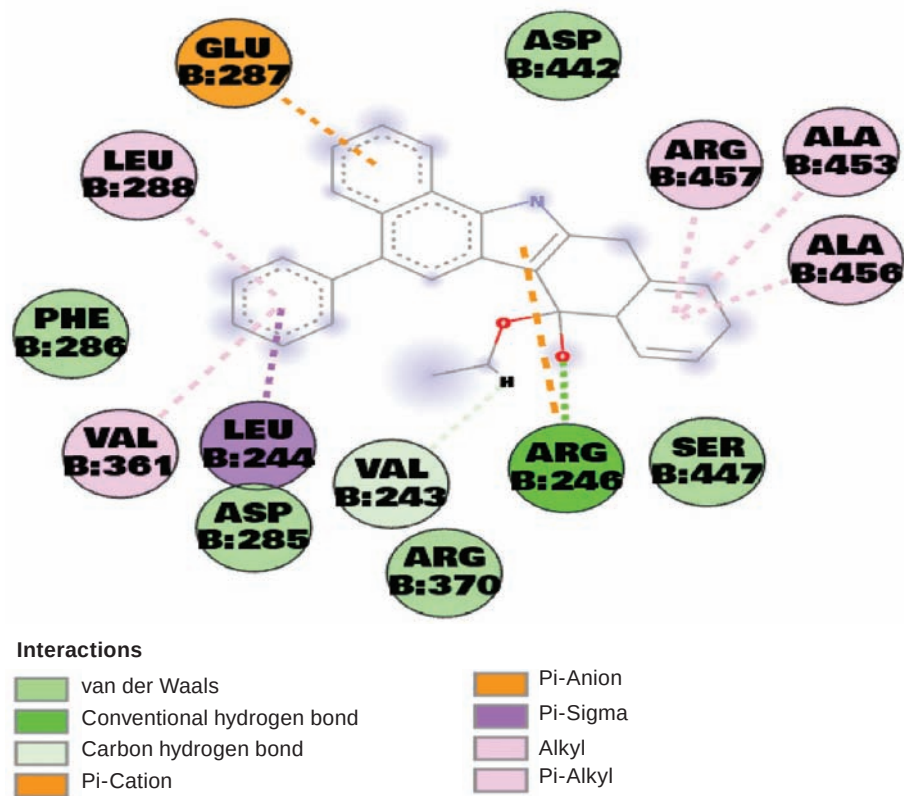


Fig. 9. Interaction of the protein with ligand 44

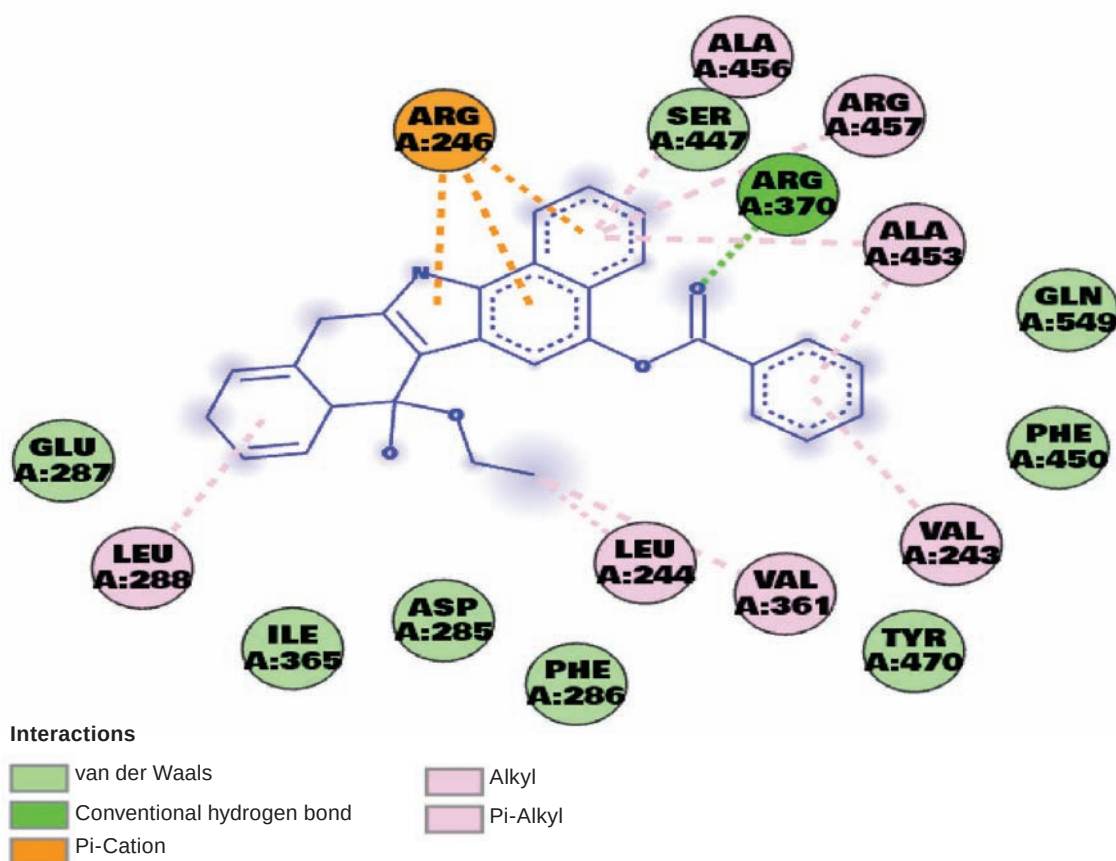


Fig. 10. Interaction of the protein with ligand 30

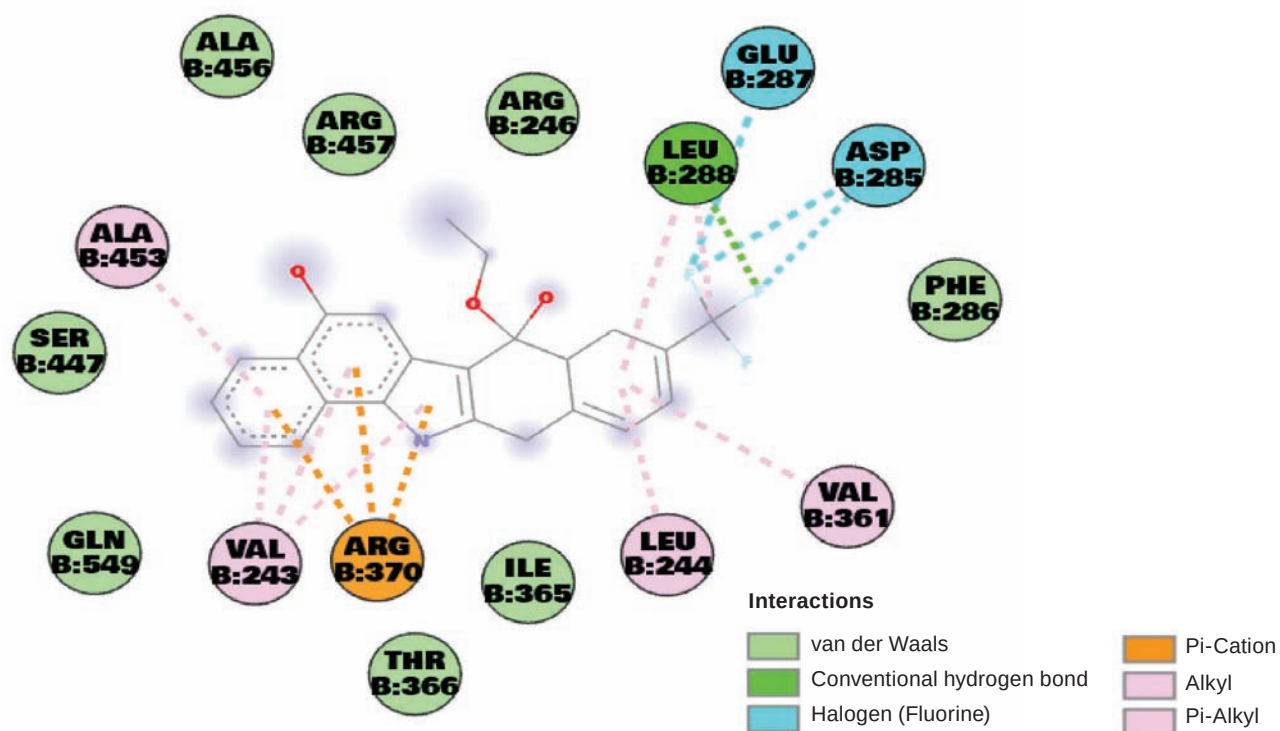


Fig. 11. Interaction of the protein with ligand 18

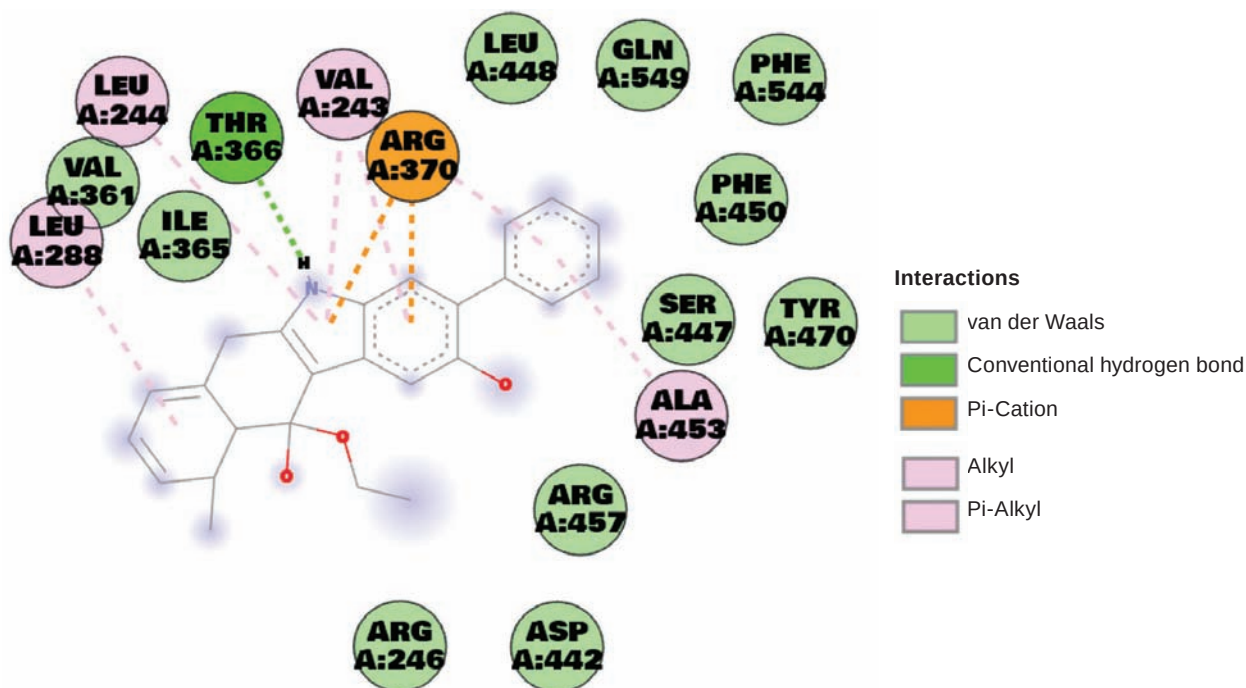


Fig. 12. Interaction of the protein with ligand 24

receptor. Strong hydrogen bonds were formed between the protein and the ligands indicating that the studied ligands have good potential as the inhibitors of ALOX5.

Conflict of interest. Authors have completed the Unified Conflicts of Interest form at http://ukr-biochemjournal.org/wp-content/uploads/2018/12/coi_disclosure.pdf and declare no conflict of interest.

Funding. The authors declare that there are no external funding sources available to carry out this research work.

ПРОГНОЗУВАННЯ $pEC_{50}(M)$ ТА МОЛЕКУЛЯРНИЙ ДОКІНГ У ПОШУКУ СЕЛЕКТИВНОГО ІНГІБУВАННЯ АРАХІДОНАТ-5-ЛІПОКСИГЕНАЗИ

N. R. Das¹, P. G. R. Achary²✉

¹Department of CSIT, Siksha 'O' Anusandhan deemed to be University, Bhubaneswar, Odisha, India;

²Department of Chemistry, Faculty of Engineering and Technology (ITER), Siksha 'O' Anusandhan deemed to be University, Bhubaneswar, Odisha, India;

✉e-mail: pgrachary@soa.ac.in

Арахідонат-5-ліпоксигеназа (ALOX5) вважається головною мішенню дії лікарських

препаратів проти фіброзу печінки, ревматоїдного артриту, атеросклерозу, раку та астми. Наразі розробка препаратів для лікування цих захворювань, що інгібують активність ALOX5, не є достатньою. Тому актуальною задачею є винахід потужних та ефективних ALOX5-таргетних препаратів. Методи пошуку кількісних співвідношень структура-властивість (QSAR) та молекулярний докінг можуть відіграти важливу роль у скринінгу та створенні ліків. У цій роботі запропоновано 3D-QSAR моделі, побудовані з використанням множинної лінійної регресії (MLR) і методу часткових найменших квадратів (PLS) для вимірювання $pEC_{50}(M)$, на основі пакету даних 112 молекул. Також проаналізовано критерій Index of Ideality of Correlation (IIC) для створення оптимального дескриптора, представленого з використанням специфікації SMILES. Показано вплив кількості та природи дескрипторів на модель. Ці моделі можуть бути використані для розробки нових ALOX5-таргетних препаратів. Значне підвищення стабільності моделі спостерігалось за введення оптимального дескриптора. Розрахунки молекулярного докінгу показали, що рецептор ALOX5 добре інгібується 112 лігандами. Найменша енергія зв'язування становила -10,8 ккал/моль. Сполуки з найкращими показниками були повторно доковані за

допомогою докінг-серверу DockThor. Показано, що ліганди з ідентифікаторами 18, 20, 24, 30 і 44 мають високий потенціал як інгібітори ALOX5.

Ключові слова: 5-ліпоксигеназа, QSAR, SMILES, молекулярний докінг.

References

- Rossi A, Pergola C, Koeberle A, Hoffmann M, Dehm F, Bramanti P, Cuzzocrea S, Werz O, Sautebin L. The 5-lipoxygenase inhibitor, zileuton, suppresses prostaglandin biosynthesis by inhibition of arachidonic acid release in macrophages. *Br J Pharmacol*. 2010; 161(3): 555-570.
- Ragno R. www.3d-qsar.com: a web portal that brings 3-D QSAR to all electronic devices-the Py-CoMFA web application as tool to build models from pre-aligned datasets. *J Comput Aided Mol Des*. 2019; 33(9): 855-864.
- Ha H, Park K, Kang G, Lee S. QSAR study using acute toxicity of *Daphnia magna* and *Hyalella azteca* through exposure to polycyclic aromatic hydrocarbons (PAHs). *Ecotoxicology*. 2019; 28(3): 333-342.
- Klüver N, Bittermann K, Escher BI. QSAR for baseline toxicity and classification of specific modes of action of ionizable organic chemicals in the zebrafish embryo toxicity test. *Aquat Toxicol*. 2019; 207: 110-119.
- Ma W, Wang Y, Chu D, Yan H. 4D-QSAR and MIA-QSAR study on the Bruton's tyrosine kinase (Btk) inhibitors. *J Mol Graph Model*. 2019; 92: 357-362.
- Jacob B, Bisht LK, Chandy V. 3D QSAR Studies of 2,3-Disubstituted Quinazolin Phenyl Acetic Acid Derivatives as Antimicrobial Agents. *Res Rev A J Pharm Sci*. 2019; 8(2): 4-7.
- Patil R, Tambe S. Multi-dimensional QSAR Studies on Aurora-A kinase Inhibitors. Proc Int Conf Drug Discov (ICDD). 2020. Available at SSRN: <https://ssrn.com/abstract=3533692>.
- Gramatica P. Principles of QSAR Modeling: Comments and Suggestions From Personal Experience. *Int J Quant Struct Relationships*. 2020; 5(3): 61-97.
- Shanthakumar B, Kathiravan MK. Insights into structures of imidazo oxazines as potent polyketide synthase XIII inhibitors using molecular modeling techniques. *J Recept Signal Transduct Res*. 2020; 40(4): 313-323.
- Galimberti F, Moretto A, Papa E. Application of chemometric methods and QSAR models to support pesticide risk assessment starting from ecotoxicological datasets. *Water Res*. 2020; 174: 115583.
- Liu S, Jin L, Yu H, Lv L, Chen CE, Ying GG. Understanding and predicting the diffusivity of organic chemicals for diffusive gradients in thin-films using a QSPR model. *Sci Total Environ*. 2020; 706: 135691.
- Cañizares-Carmenate Y, Mena-Ulecia K, Perera-Sardiña Y, Torrens F, Castillo-Garit JA. An approach to identify new antihypertensive agents using Thermolysin as model: In silico study based on QSARINS and docking. *Arab J Chem*. 2019; 12(8): 4861-4877.
- Ahamed TKS, Rajan VK, Muraleedharan K. QSAR modeling of benzoquinone derivatives as 5-lipoxygenase inhibitors. *Food Sci Human Wellness*. 2019; 8(1): 53-62.
- Lončarić M, Strelec I, Pavić V, Šubarić D, Rastija V, Molnar M. Lipoxygenase Inhibition Activity of Coumarin Derivatives-QSAR and Molecular Docking Study. *Pharmaceuticals (Basel)*. 2020; 13(7): 154.
- Amin SNM, Idris MHM, Selvaraj M, Jamari H, Kek TL. Virtual screening, ADME study, and molecular dynamic simulation of chalcone and flavone derivatives as 5-Lipoxygenase (5-LO) inhibitor. *Mol Simul*. 2020; 46(6): 487-496.
- Sogawa S, Nihro Y, Ueda H, Izumi A, Miki T, Matsumoto H, Satoh T. 3,4-Dihydroxychalcones as potent 5-lipoxygenase and cyclooxygenase inhibitors. *J Med Chem*. 1993; 36(24): 3904-3909.
- Das NR, Mishra SP, Achary PG. Evaluation of Molecular Structure based Descriptors for the Prediction of pEC50 (M) for the Selective Adenosine A2A Receptor. *J Mol Struct*. 2021; 123: 130080.
- Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res*. 2016; 44(D1): D1045-D1053.
- Toropov AA, Toropova AP, Cappelli CI, Benfenati E. CORAL: Model for octanol/water partition coefficient. *Fluid Phase Equilib*. 2015; 397: 44-49.
- Toropova AP, Toropov AA. CORAL: QSAR models for carcinogenicity of organic compounds

- for male and female rats. *Comput Biol Chem.* 2018; 72: 26-32.
21. Toropova AP, Toropov AA. CORAL software: prediction of carcinogenicity of drugs by means of the Monte Carlo method. *Eur J Pharm Sci.* 2014; 52: 21-25.
 22. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J Comput Chem.* 2013; 34(24): 2121-2132.
 23. Achary PGR. Simplified molecular input line entry system-based optimal descriptors: QSAR modelling for voltage-gated potassium channel subunit Kv7.2. *SAR QSAR Environ Res.* 2014; 25(1): 73-90.
 24. Toropova AP, Toropov AA, Veselinović JB, Miljković FN, Veselinović AM. QSAR models for HEPT derivatives as NNRTI inhibitors based on Monte Carlo method. *Eur J Med Chem.* 2014; 77: 298-305.
 25. Toropov AA, Achary PGR, Toropova AP. Quasi-SMILES and nano-QFPR: The predictive model for zeta potentials of metal oxide nanoparticles. *Chem Phys Lett.* 2016; 660: 107-110.
 26. Ojha PK, Mitra I, Das RN, Roy K. Further exploring rm2 metrics for validation of QSPR models. *Chemom Intell Lab Syst.* 2011; 107(1): 194-205.
 27. Toropov AA, Toropova AP. Application of the Monte Carlo method for building up models for octanol-water partition coefficient of platinum complexes. *Chem Phys Lett.* 2018; 701: 137-146.
 28. Toropova AP, Toropov AA. Use of the index of ideality of correlation to improve models of ecotoxicity. *Environ Sci Pollut Res Int.* 2018; 25(31): 31771-31775.
 29. Toropova AP, Toropov AA. The index of ideality of correlation: A criterion of predictability of QSAR models for skin permeability? *Sci Total Environ.* 2017; 586: 466-472.
 30. Toropov AA, Carbó-Dorca R, Toropova AP. Index of Ideality of Correlation: new possibilities to validate QSAR: a case study. *Struct Chem.* 2018; 29(1): 33-38.
 31. Toropov AA, Toropova AP. The index of ideality of correlation: A criterion of predictive potential of QSPR/QSAR models? *Mutat Res Genet Toxicol Environ Mutagen.* 2017; 819: 31-37.
 32. Toropova AP, Toropov AA. Hybrid optimal descriptors as a tool to predict skin sensitization in accordance to OECD principles. *Toxicol Lett.* 2017; 275: 57-66.
 33. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* 1989; 45(1): 255-268.
 34. Achary PGR, Toropova AP, Toropov AA. Combinations of graph invariants and attributes of simplified molecular input-line entry system (SMILES) to build up models for sweetness. *Food Res Int.* 2019; 122: 40-46.
 35. Begum S, Achary PGR. Simplified molecular input line entry system-based: QSAR modelling for MAP kinase-interacting protein kinase (MNK1). *SAR QSAR Environ Res.* 2015; 26(5): 343-361.
 36. Achary PGR. QSPR modelling of dielectric constants of π -conjugated organic compounds by means of the CORAL software. *SAR QSAR Environ Res.* 2014; 25(6): 507-526.
 37. Golbraikh A, Tropsha A. Beware of q²! *J Mol Graph Model.* 2002; 20(4): 269-276.
 38. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010; 31(2): 455-461.