Development of QSAR Model using MLR, PLS and NN Approach to Elucidate the Physicochemical Properties Responsible for Antihyperlipidemic Potential of ACAT inhibitors

Shagufta Khan[a&b], Asim Y. Najmi[b], Seema Kesar[a], Kirtika Madan[a], Richa Arya[a]

[a]*Department of Pharmacy, Banasthali Vidyapith, Banasthali-304022, Rajasthan, India*

[b]*Department of Pharmaceutical Chemistry, College of Pharmacy, Jazan University, Jazan-45142, Kingdom of Saudi Arabia*

***Corresponding author**

Shagufta Khan

Department of Pharmaceutical Chemistry, College of Pharmacy, Jazan University, Jazan-45142, Kingdom of Saudi Arabia

Email id: khan.shagufta2@gmail.com, shaguftak@jazanu.edu.sa

Abstract

In an endeavor to discover novel therapeutics targeting ACAT enzyme against numerous diseases, a QSAR model was developed. The present study involves software based drug design, a 2D QSAR model validated for its ideal statistical quality using Multiple linear regression (MLR) and Partial least Square (PLS) methods. The study resulted in significant statistical values such as S value=0.44, F value=62.34, r=0.91, r^2=0.83 and r^2cv=0.79. The mode of dependencies of chosen descriptors on the pharmacological activity was carried out by building neural networks using the descriptors and the inhibitory activities. The 2D QSAR model provided insights for the structural optimization designed for improved selectivity of ACAT inhibitors. The exhaustive study of the proposed model gives the perception to formulate optimized compounds exhibiting improved selectivity profile.

Keywords: QSAR, ACAT, MLR, PLS, FFNN

Introduction

Majority of deaths worldwide is directly associated with coronary heart diseases and the growing risk factor for this is atherosclerosis. Hyperlipidemia is characterized by an increased concentration of lipids in the systemic circulation and can be regarded as an alarming situation for atherosclerosis. An eye-catching target for hyperlipidemia is inhibition of ACAT (Acyl CoA- Cholesterol acyl transferase), an intracellular enzyme portrait as a key for esterification of cholesterol into long chain fatty acid cholesteryl esters. It also plays essential role in the intestinal absorption of the dietary cholesterol, lipoprotein cholesterol production and secretion in liver and the accumulation of cholesteryl ester in foam cells as atheromatous plaque.[1] There are two isoforms of ACAT found in mammalian species, ACAT1 and ACAT2. In humans, both are present in the liver and intestine but differ in cell locations, metabolic functions and membrane orientation. In macrophages, ACAT 1 exhibits crucial role in the formation of foam cells; whereas the overall cholesterol absorption process is carried out by ACAT 2 in the intestinal mucosal cells.[2] Therefore, inhibition of ACAT[4] enzyme symbolizes a good therapeutic approach for the discovery of newer antihyperlipidemic agents.[3]

At this developing phase of technology, computational approach could be an urgent necessity for the expansion of drug-oriented discovery[5]. Chemo metrics is one of the most reliable techniques (Figure 1) for exploring the quantitative relationship between chemical structures and pharmacological activity for a series of compounds. This relationship can be subjugated for the design and optimization of novel compounds with desired selectivity and potency. It saves resources with reduced time and cost and expedites the probability of success in drug discovery process. Considering these parameters, we decided to develop a model from classical QSAR descriptors using multiple linear regressions, partial least square and neural networks[6, 7].

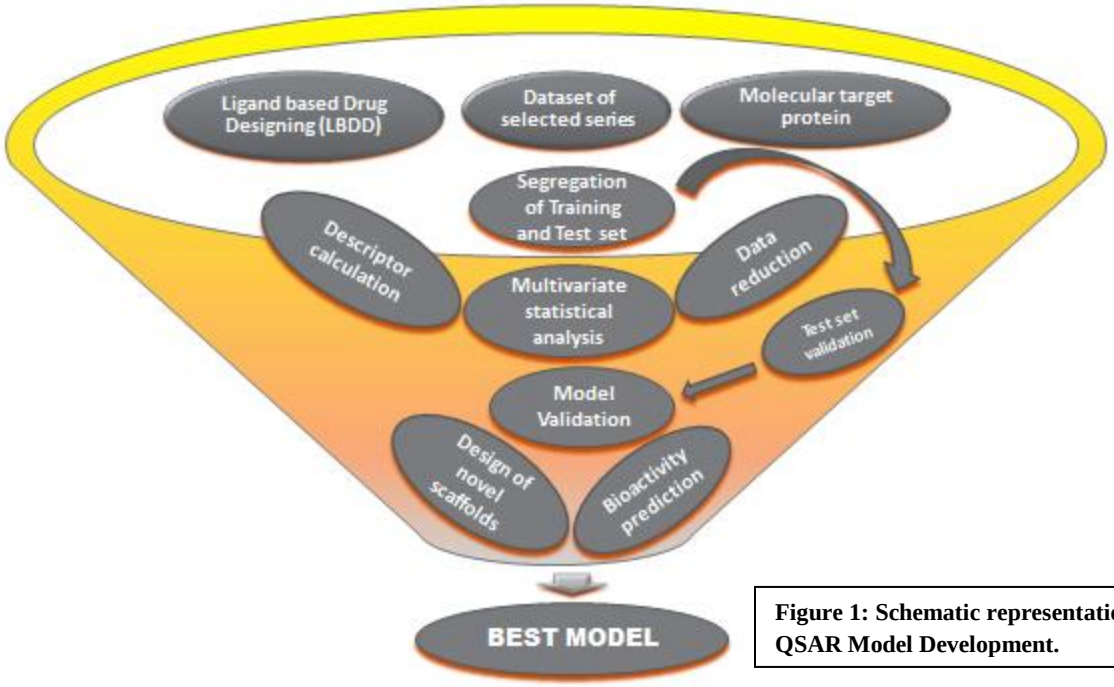


Figure 1: Schematic representation of Best QSAR Model Development.

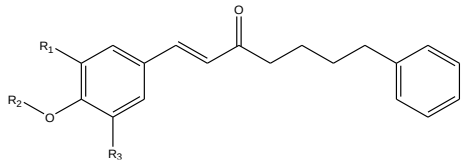
Materials and Methods

Data set Preparation

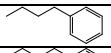
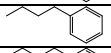
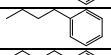
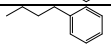
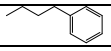
The dataset of 43 chemically diverse structures was chosen from the literature for the desired 2D QSAR studies. The computational studies were performed by using Tools for structure activity relationship (TSAR) Software version 3.3. The structures were drawn using ChemDraw Ultra 8.0 and were imported to TSAR 3.3 worksheet. Experimentally acquired IC_{50} values were represented as the negative logarithm of the IC_{50} and inserted in separate columns of the TSAR sheet due to skewness of the authentic data that process the free energy of binding. The substitution for each compound was defined separately using the protocol "define substituent" as R^1, R^2, R^3 and R^4 **(Table 1**). Numerical allotment of each substituent was done on the basis of their position in the molecules, and an individual structure owns its defined set of substituents attached to the lead by a single bond. Descriptor calculation needs prior conversion of 2D molecules into their respective 3D molecules through CORNIA option, followed by geometry and energy optimization via COSMIC selection route, which signifies the stability of the conformer.

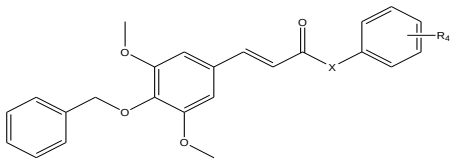
Table 1: Data set used for QSAR model development

N-(3, 5-Dimethoxy-4-n-octyloxycinnamoyl)-N'-(3, 4-dimethylphenyl) piperazine analogues

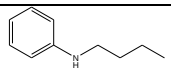
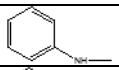
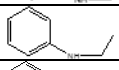
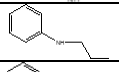
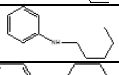
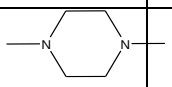
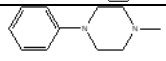
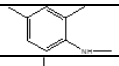
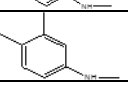


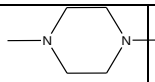
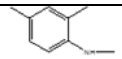
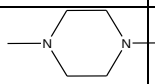
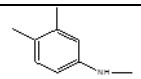
(Basis Structure of serial no: 1-7)

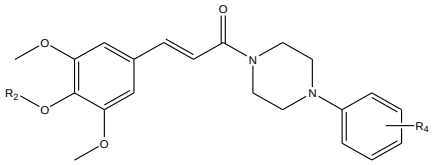
S. No.	Compound Name	Structure of Compound			Substituent's Defined For TSAR				IC_{50} nM
		R_1	R_2	R_3	Subs 1	Subs 2	Subs 3	Subs 4	
1	1	OMe	H	H	R_1	OR_2	H		20600
2	2	OMe	OMe	H	R_1	OR_2	H		9800
3	3	OMe	$C_6H_5CH_2$-	H	R_1	OR_2	H		2700
4	4	OMe	C_6H_5CO-	H	R_1	OR_2	H		38000
5	5	H	$C_6H_5CH_2$-	H	R_1	OR_2	H		21600
6	6	C_2H_5O-	$C_6H_5CH_2$-	H	R_1	OR_2	H		3500
7	7	OMe	$C_6H_5CH_2$-	OMe	R_1	OR_2	OMe		1700



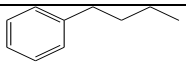
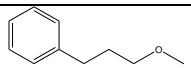
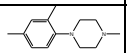
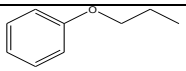
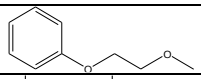
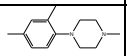
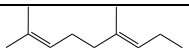
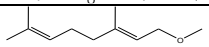
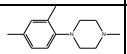
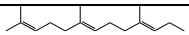
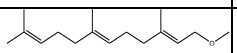
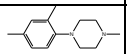
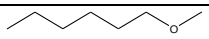
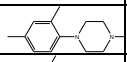
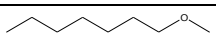
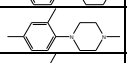
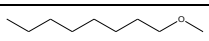
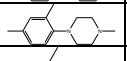
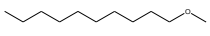
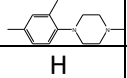
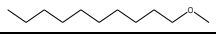
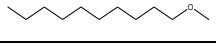
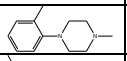
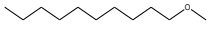
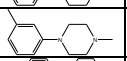
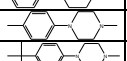
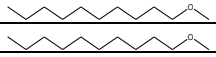
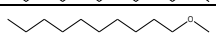
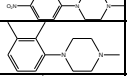
(Basis Structure of serial no: 11-20)

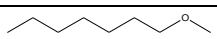
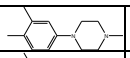
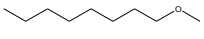
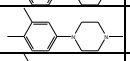
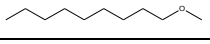
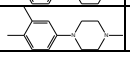
S. No.	Compound Name	Structure of Compound		Substituent's Defined For TSAR				IC_{50} nM
		X	R_4	Subs 1	Subs 2	Subs 3	Subs 4	
1	11	-NH$(CH_2)_3$-	H	OMe	-$OCH_2C_6H_5$	OMe		840
2	12	-NH-	H	OMe	-$OCH_2C_6H_5$	OMe		260
3	13	-$NHCH_2$-	H	OMe	-$OCH_2C_6H_5$	OMe		3000
4	14	-NH$(CH_2)_2$-	H	OMe	-$OCH_2C_6H_5$	OMe		1700
5	15	-NH$(CH_2)_4$-	H	OMe	-$OCH_2C_6H_5$	OMe		900
6	16		H	OMe	-$OCH_2C_6H_5$	OMe		550
7	17	-NH-	2,4-$(CH3)_2$-	OMe	-$OCH_2C_6H_5$	OMe		52
8	18	-NH-	3,4-$(CH3)_2$-	OMe	-$OCH_2C_6H_5$	OMe		47

9	19		2,4-$(CH3)_2$-	OMe	$-OCH_2C_6H_5$	OMe		62
10	20		3,4-$(CH3)_2$-	OMe	$-OCH_2C_6H_5$	OMe		60



(Basis Structure of serial no: 21-46)

S. No.	Compound Name	Structure of Compound		Substituent's Defined For TSAR				IC_{50} nM
		R_2	R_4	Subs 1	Subs 2	Subs 3	Subs 4	
1	21		2,4- $(CH_3)_2$	OMe		OMe		100
2	22		2,4- $(CH_3)_2$	OMe		OMe		790
3	23		2,4- $(CH_3)_2$	OMe		OMe		6.4
4	24		2,4- $(CH_3)_2$	OMe		OMe		4.5
5	25	n- C_6H_{13}	2,4- $(CH_3)_2$	OMe		OMe		79
6	26	n- C_7H_{15}	2,4- $(CH_3)_2$	OMe		OMe		15
7	27	n- C_8H_{17}	2,4- $(CH_3)_2$	OMe		OMe		15
8	28	n- C_9H_{19}	2,4- $(CH_3)_2$	OMe		OMe		19
9	29	n- $C_{10}H_{21}$	2,4- $(CH_3)_2$	OMe		OMe		22
10	30	n- $C_{10}H_{21}$	H	OMe		OMe	H	52
11	31	n- $C_{10}H_{21}$	2- CH_3	OMe		OMe		18
12	32	n- $C_{10}H_{21}$	3- CH_3	OMe		OMe		38
13	33	n- $C_{10}H_{21}$	4- CH_3	OMe		OMe		23
14	34	n- $C_{10}H_{21}$	4- $C(CH_3)_3$	OMe		OMe		35
15	35	n- $C_{10}H_{21}$	4- OCH_3	OMe		OMe		37
16	36	n- $C_{10}H_{21}$	4- F	OMe		OMe		47
17	37	n- $C_{10}H_{21}$	4- NO_2	OMe		OMe		180
18	38	n- $C_{10}H_{21}$	2,3- $(CH_3)_2$	OMe		OMe		24
19	39	n- $C_{10}H_{21}$	2,5- $(CH_3)_2$	OMe		OMe		316
20	40	n- $C_{10}H_{21}$	2,6- $(CH_3)_2$	OMe		OMe		650
21	41	n- $C_{10}H_{21}$	3,5- $(CH_3)_2$	OMe		OMe		93
22	42	n- $C_{10}H_{21}$	3,4- $(CH_3)_2$	OMe		OMe		26
23	43	n- C_6H_{13}	3,4- $(CH_3)_2$	OMe		OMe		54

24	**44**	**n- C_7H_{15}**	**3,4- $(CH_3)_2$**	**OMe**		**OMe**		**38**
25	**45**	**n- C_8H_{17}**	**3,4- $(CH_3)_2$**	**OMe**		**OMe**		**11**
26	**46**	**n- C_9H_{19}**	**3,4- $(CH_3)_2$**	**OMe**		**OMe**		**20**

Generation of Descriptors and data reduction

TSAR model development needs descriptors to be depicted numerically and subjected to correlation statistically. TSAR can be defined as a hub of descriptors and affords vast variety of descriptor calculations such as molecular mass, molecular connectivity indices, ellipsoidal volume, dipole moments, lipole moments, molecular mass, Wiener index, molecular connectivity indices, nuclear repulsion energy, heat of formation, etc.

In order to interpret the data concerning physicochemical parameters of each compound, descriptor calculation was carried out. The present study involved more than 150 descriptors, which were not regarded consistent for building a best model. Therefore, to avoid data redundancy and coextensive correlation due to large data pool, data reduction was implemented and merely the appropriate set of descriptors was adopted. By employing correlation matrix, descriptors showing high correlation with biological activity were selected while others with poor correlation were discarded from the dataset.

Ultimately, only three descriptors i.e. Dipole moment X component (Whole molecule), Kier Chi 6 (path) index (Whole molecule) and Balaban Topological index (Whole molecule) were found to be statistically significant and observed extremely correlated with the biological activity.

Training and Test set Congregation

After data reduction, the dataset of 43 compounds was randomly distributed into the training set of 30 compounds and the test set of 13 molecules. In order to get the preferred division of the entire dataset, it entails to be adjusted by hit and trial method. The best QSAR model can be achieved if the values of certain statistical parameters are satisfying to the set criteria. These parameters include Fischer test value (F), Standard deviation (s), Coefficient of correlation (r), Squared correlation coefficient (r^2) and Cross validated coefficient of determination (r^2 CV) Figure 2. After the accomplishment of these parametric measures, the concerned model can be regarded as consistent, robust and

predictable. Some of the exceptional compounds, regarded as outliers, were excluded from the model development as they represent the higher residual (beyond two orders of magnitude) values.

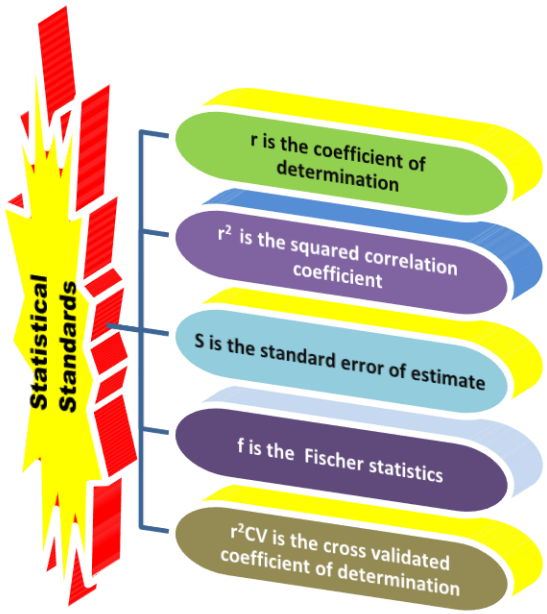


Figure 2: A set of statistical standards representing the QSAR models.

Model Development and Validation

Multiple Linear Regression (MLR) Study

MLR helps to design a model that defines how the chosen set of descriptors is dependent on its biological activity. Analysis via linear regression gives information to establish correlation between the dependent (log $1/IC_{50}$) and independent (descriptors) variables of the selected series. The model is said to be fit if the sum-of-squares of the differences of predicted and observed values is reduced. Least squares curve fitting method is expedient for estimating the values of regression coefficient by MLR. The regression model can be built by using training set compounds and is said to be statistically significant if it satisfies the set values of F, s, r, r^2 and r^2 CV. The test set is used to evaluate the predictive power of the suggested model and its correlation value (r^2) should be less than that of the training set.

Partial Least Squares (PLS) Study

An advantage of PLS is to establish models predicting more than one dependent variable. It determines the robustness of the model derived from the MLR study. This technique circumvents the possibility of over fitting and gives advanced information about the proposed model. It involves the validation of MLR results as the same dataset was utilized in PLS technique. In order to determine the eminence of the developed PLS model, the correlation values of both training and test set were assessed.

Forward Feed Neural Network (FFNN) Study

FFNN is an artificial multilayered neural network that delivers the information regarding dependencies of the selected descriptors on the biological activity. These are represented in graphical form and at times they represent better accuracy in the prediction as compared to linear regression model. Artificial neural network (ANN) is build up of three layers: input, hidden and output layers. After receiving signals from input data, the hidden layer is organized with input vectors and each neuron carries out the nonlinear transfer functions on them. FFNN model was constructed by the finest set of descriptors preferred on the basis of linear regression analysis. This analysis is considered advantageous over linear regression, since it carries more flexible parameters bearing same number of variables.

Model Validation

Validation of the proposed model can be achieved through cross-validation (r^2CV) and test set predictions. Cross validation technique is required to judge the consistency and reliability of the statistically significant model. Leave-one-out mode was opted for swaying the molecules from training set to test set and vice-versa. Correlation value (r^2CV) should be achieved within the set range i.e. not lower than 0.60. Prediction of the activity via test set compounds signifies the prognostic ability of the proposed model. Generally, it is achieved on the basis of the r^2 value which should not be lower than 0.6.

RESULTS AND DISCUSSION

The QSAR model based on the classical descriptor centered analysis was built by using huge dataset of several analogues. Multivariate models including MLR, PLS and FFNN were applied for the construction of these models. The present QSAR study revealed total 43 compounds with their structures and pharmacological activity are shown in table 1. After surplus data reduction, the compounds were segregated into the training set of 30 compounds and the test set of 13 compounds to estimate the prognostic power of the 2D QSAR model generated by means of linear and nonlinear statistical approaches[8].

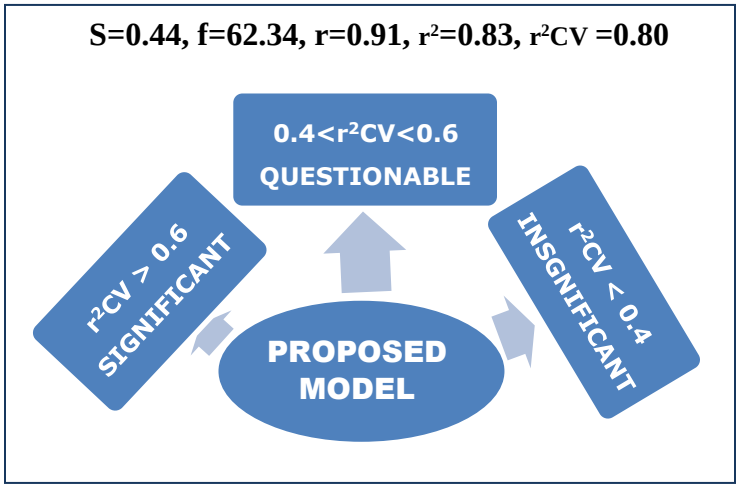
Analytical Statistics via Linear methods

Entire descriptor pool was subjected to the calculation of linear regression analysis i.e.

MLR and PLS, which showed insignificant results specifying reduced predictability of the proposed model. An urgent demand for consistent and instructive descriptor set, possessing excellent correlation with biological activity and no inter-correlation, resulted in correlation matrix application. Ultimately, final set of three descriptors was taken into consideration, which was independent of each other and observed significant in model development. The final selected descriptors were Dipole moment X component (Whole molecule), Kier Chi 6 (path) index (Whole molecule) and Balaban Topological index (Whole molecule) and their correlation matrixes are represented in the table 2. Prediction via test set compounds resulted in the validation of the activity of training set compounds chose to construct the desired model. Evaluation methods opted were similar for both training and test set compounds. For the enrichment of predictability and statistical significance of the proposed model, it was analyzed for the outliers. But this model was observed free of outliers. The final equation produced by MLR analysis is shown in equation 1.

$$\mathbf{Y = 0.2288584\ X1 + 1.0356268\ X2 + 3.2055991\ X3 - 10.944022} \qquad \mathbf{(1)}$$

Where, X1= Dipole moment X component (Whole molecule), X2= Kier Chi 6 (path) index (Whole molecule) and X3= Balaban Topological index (Whole molecule)



In view of the diverse statistical values such as Fischer test value (F), Standard deviation (s), Coefficient of correlation (r), Squared correlation coefficient (r^2), Cross validated coefficient of determination (r^2 CV) and the test for statistical significance (t), the anticipated QSAR model was evaluated. Correlation coefficient represented by the above equation is observed excellent between descriptors and ACAT inhibitors, **r** = 0.91, which shows the amount of fitting of the created model. High $\mathbf{r^2}$ value signifies the percent data within a particular equation and it was found out to be 0.83, which means 83% variance in biological activity of the applied data set. The reliability of the designed model was further analyzed by cross-validated squared correlation coefficient ($\mathbf{r^2CV}$) and it was 0.80, which clearly proved its significance that the acquired regression equation is excellent in its prediction. The lower value of the standard deviation, **s** = 0.44 states that how precisely the model is capable of fitting statistically. The lesser is the value of s the better will be the chance of getting statistically significant QSAR model. F-test defines the degree of variance (**F**= 62.34) expressed by the generated model. The higher is its value, the greater will be the statistical reliability of the model.

Table 2: Correlation matrix of three most important descriptors used to develop the model.

Descriptor	Dipole moment X component (Whole molecule)	Kier Chi 6 (path) index (Whole molecule)	Balaban Topological index (Whole molecule)	Biological Activity
Dipole moment X component (Whole molecule)	**1**	**0.0692**	**-0.4511**	**-0.4695**
Kier Chi 6 (path) index (Whole molecule)	**0.0692**	**1**	**-0.1378**	**-0.7919**
Balaban Topological index (Whole molecule)	**-0.4511**	**-0.1378**	**1**	**0.8401**
Biological Activity	**-0.4695**	**-0.7919**	**0.8401**	**1**

Furthermore, PLS analysis was carried out on the chosen set of descriptors to obtain model interpretability. This method further validates the MLR results and display comparison of the two approaches with marginal variation. Interpretation of the PLS outcomes via statistical parameter, r^2 further ascertains the precision of the created model. The final regression equation produced by PLS analysis is shown in equation 2.

Y = 0.28671637 X1 + 0.98600513 X2 + 3.717665 X3 - 11.248522 (2)

Where, X1= Dipole moment X component (Whole molecule), X2= Kier Chi 6 (path) index (Whole molecule) and X3= Balaban Topological index (Whole molecule)

The $\mathbf{r^2}$ value of the training and test set of MLR and PLS is highly comparable (MLR, PLS $_{\text{Training set}}$) r^2 0.85 & 0.84 and (MLR, PLS $_{\text{Test set}}$) r^2 0.81 & 0.83 respectively. The smaller is the difference, the higher will be the robustness and prognostic ability of the model. The r^2CV= 0.80, was also evaluated.[9] Likewise, the intimacy perceived between the observed and predicted values of the training and test set (Table 3) specifies the significance of the retrieved MLR and PLS equations for statistical assessment. Figure 3 and Figure 4 portrays the graphical representation of the actual and predicted values of training and test set compounds obtained via MLR and PLS techniques respectively[10-15].

In order to portray a validated 2D-QSAR model demonstrating stability and worthiness, Artificial Neural Network (ANN) approach was applied. It further assess the firmness of the generated model via MLR and PLS methods. The present study showed inputs as descriptors and the output was log $1/IC_{50}$, while the numbers of hidden neurons were computed by the TSAR software automatically in the form of training and test patterns. Statistical analysis revealed from the FFNN ACAT inhibitors data comprised net configuration of 3-2-1 and the $\mathbf{r^2}$ for training and test set was 0.87 and 0.86 respectively. A plot representing the relationship established by ANN between the observed and predicted activity is shown in figure 5, which further proved that the proposed model is of commendable statistical quality[16-19].

Table 3: Structures of the training and test set compounds for building QSAR model with their observed and predicted activities.

Compound Name	Actual Activity	Predicted Activity		
		MLR	PLS	FFNN
Training Set Compounds				
1	-3.23045	-3.07031	-3.91139	-3.92433
3	-2.41497	-2.36631	-3.57097	-3.7625
4	-3.23045	-3.0282	-3.84454	-3.92772
5	-2.74036	-2.66201	-4.98637	-4.47633
7	-1.716	-1.84945	-3.06224	-3.21228
12	-1.6721	-1.85408	-2.16933	-2.17708
14	-1.77815	-2.19911	-3.03403	-3.16343
16	-2	-2.08142	-2.6794	-2.85004
17	-2.89763	-2.05035	-1.64714	-1.61051
18	-0.80618	-0.93759	-1.65801	-1.6258
20	-0.65321	-0.34996	-2.24487	-2.31902
21	-1.17609	-1.62475	-2.10821	-2.17068
22	-1.27875	-1.49082	-2.06785	-2.1269
23	-1.716	-2.02049	-0.92039	-1.02649
24	-1.25527	-1.79653	-0.2921	-0.75271
26	-1.57978	-1.77165	-1.67766	-1.58498
28	-1.36173	-1.70677	-1.53529	-1.46723
30	-1.54407	-1.27257	-2.04083	-1.99738
31	-1.6721	-1.73313	-1.82184	-1.75342
32	-2.25527	-1.66354	-1.80649	-1.74182
33	-1.38021	-1.65135	-1.74668	-1.67905
34	-2.49969	-1.58386	-1.3351	-1.30584
36	-1.96848	-1.48167	-1.7809	-1.71097
37	-1.41497	-1.5103	-1.69053	-1.62776
38	-1.04139	-1.60887	-1.70231	-1.62578
39	-1.30103	-1.50819	-1.62173	-1.55132
41	-3.23045	-3.07031	-1.53631	-1.47559
42	-2.41497	-2.36631	-1.55719	-1.49701
45	-3.23045	-3.0282	-1.65442	-1.57651
46	-2.74036	-2.66201	-1.53971	-1.48036
Test set Compounds				
2	-3.99123	-4.18863	-3.83557	-4.07651
6	-3.54407	-3.28303	-3.64598	-3.56457
11	-2.92428	-3.07854	-3.14585	-2.9065
13	-3.47712	-3.18708	-3.21474	-3.47287
15	-2.95424	-2.96431	-2.98521	-2.43687
19	-1.79239	-1.57761	-1.36181	-1.75326
25	-1.89763	-1.95217	-1.90659	-1.67836
27	-1.17609	-1.78015	-1.52496	-1.09105
29	-1.34242	-1.60813	-1.53149	-1.6976
35	-1.5682	-1.61877	-1.96182	-1.95658
40	-2.81291	-1.73353	-1.81245	-1.86723
43	-1.73239	-1.95341	-1.97081	-1.80733
44	-1.57978	-1.8674	-1.89545	-1.82076

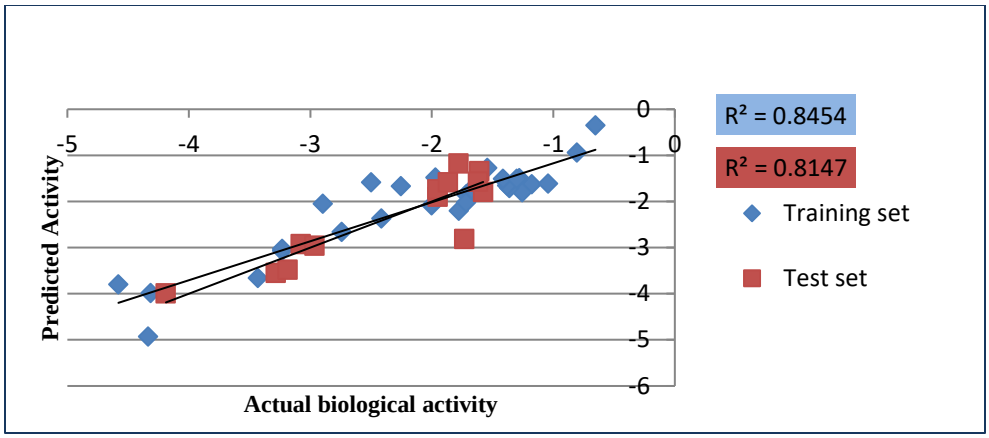


Figure 3: MLR graph representing plot between actual verses predicted activity

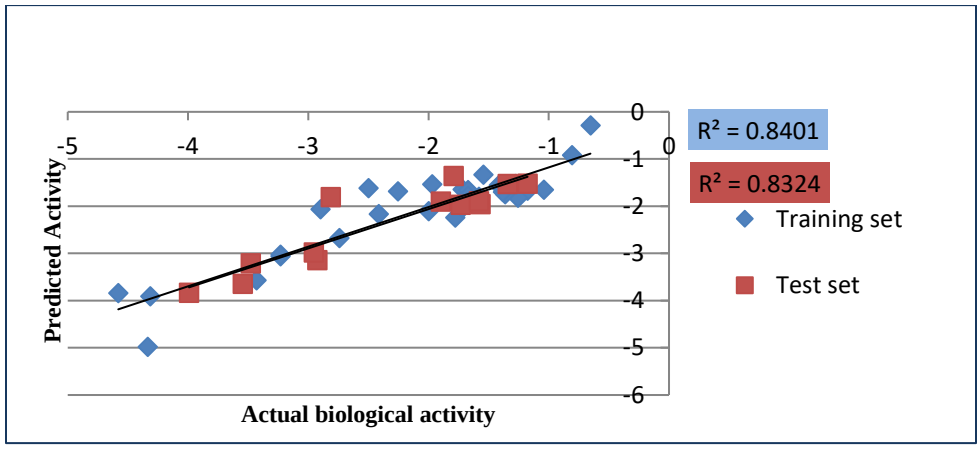


Figure 4: PLS graph representing plot between actual verses predicted activity

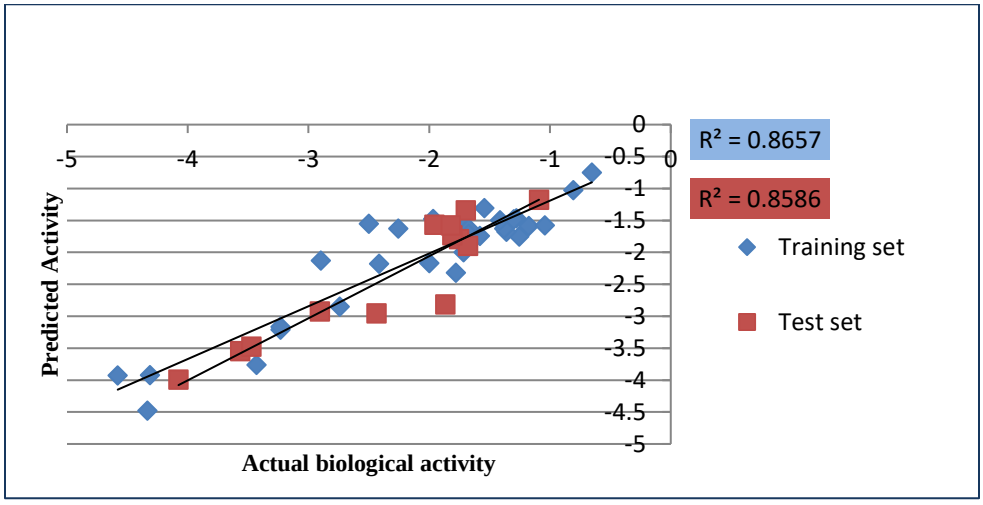


Figure 5: FFNN graph representing plot between actual verses predicted activity

Importance of descriptors entered: Importance of molecular descriptors presented in this section (figure 6) clearly suggest that, in the body, molecular structures of drugs impose their functions, and any changes of their chemical structure may induce changes in their functions[20,21].

An electronic parameter, which is molecular dipole moment, is due to the amount of charge distribution in a molecule and observed significant when electronic interactions are elaborated in drug-receptor interfaces. Dipole moment X component describes the moments using the substituent point of attachment as an origin with this bond placed along the X-axis. The X-axis (bond of attachment) and the components of μ are added to produce bond dipole as in Debyes [21]. It is an index of molecular polarity and its value can be defined as a function of the differences in the electronegativity of connected atoms and distance between them, which is considered as an essential factor for a molecule to bind with its biological target (Cartier and Rivail, 1987). Molecular dipole moment (X-component) (dx) was the first descriptor encountered in the model, which is positively correlated with the biological activity that clearly explains why more polar compounds show high potency and are considered as a prerequisite for the desired activity. It further indicates that upon increasing the polarity of the lead moiety by substituting such groups that increase the polarity of the molecule as a whole will account for an increase in the biological activity. FFNN dependence graphical representation of the proposed model has been shown in the figure 7, which showed linear relationship between $LogIC_{50}$ and dipole moment i.e., the higher is the polarity the better will be the potency.

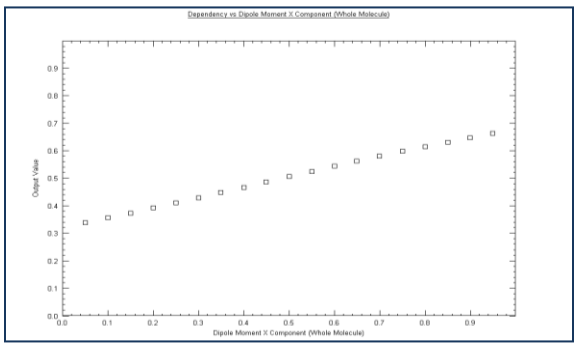


Figure 7: Dependency graph of neural analysis demonstrating correlation between dipole moment X component (whole molecule) and actual activity data.

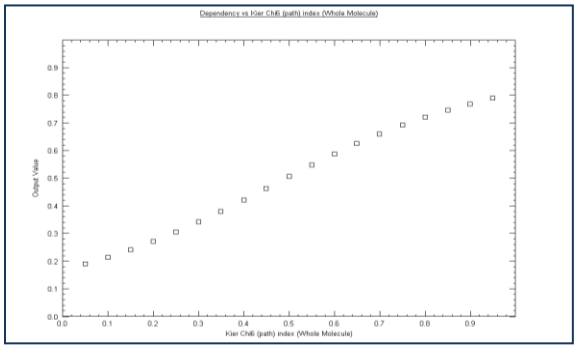


Figure 8: Dependency graph of neural analysis demonstrating correlation between Kier Chi6 (path) index (whole molecule) and actual activity data.

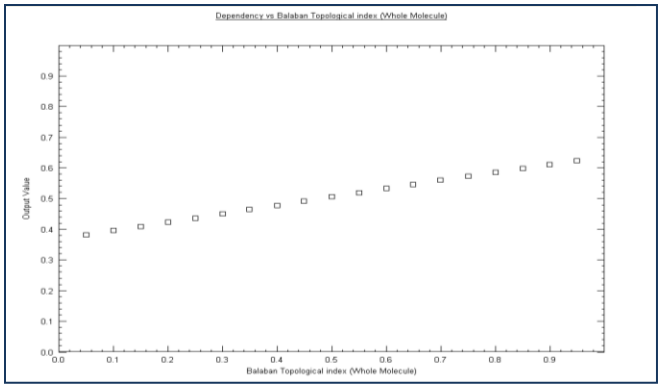


Figure 9: Dependency graph of neural analysis demonstrating correlation between Balaban Topological index (whole molecule) and actual activity data.

The second descriptor encountered was Kier Chi 6 (path) index (Whole molecule), which is an essential parameter well-defined by Randic and subsequently by Kier and Hall. It demonstrates numerical series characterized by "order" and "subgraph type."

It is defined as a Chi index used for the calculation of known type of sub graph: Path, Cluster, Path/Cluster, and ring Chain (P, C, P/C & CH), subjected by a function of delta values. These types put emphasis on diverse characteristics of atom connectivity within a molecule. It benefits in determining the patterns of substitutions in the ring and the extent to which branching occurs as ring structures with flexibility. Figure 8 represents Kier Chi 6 (path) index, which was observed positively correlated with the permeability as the present study revealed that the sixth order valence connectivity index (Kier Chi 6) encrypts structural complexity for example the size and the branching or substitution patterns of the ring. It is expected for a molecule to exhibit excellent interactions, only if a satisfactory substitution patterns accompanied by optimum branching at appropriate position is present. Consequently, upon increasing the size or bulk via replacements with bulky groups or through branching at this particular position will demonstrate a positive impact on the biological activity of the compounds.

Balaban Topological index (Whole molecule) was the third descriptor known for the representation of the number of rotatable bonds within the molecule. The present study showed positive correlation of Balaban index with biological activity figure 9; therefore, upon substituting those groups which lead to an increased value of it will definitely increase the biological activity. There is steady rise observed with MLR and PLS models, with the positive coefficients for the descriptors.

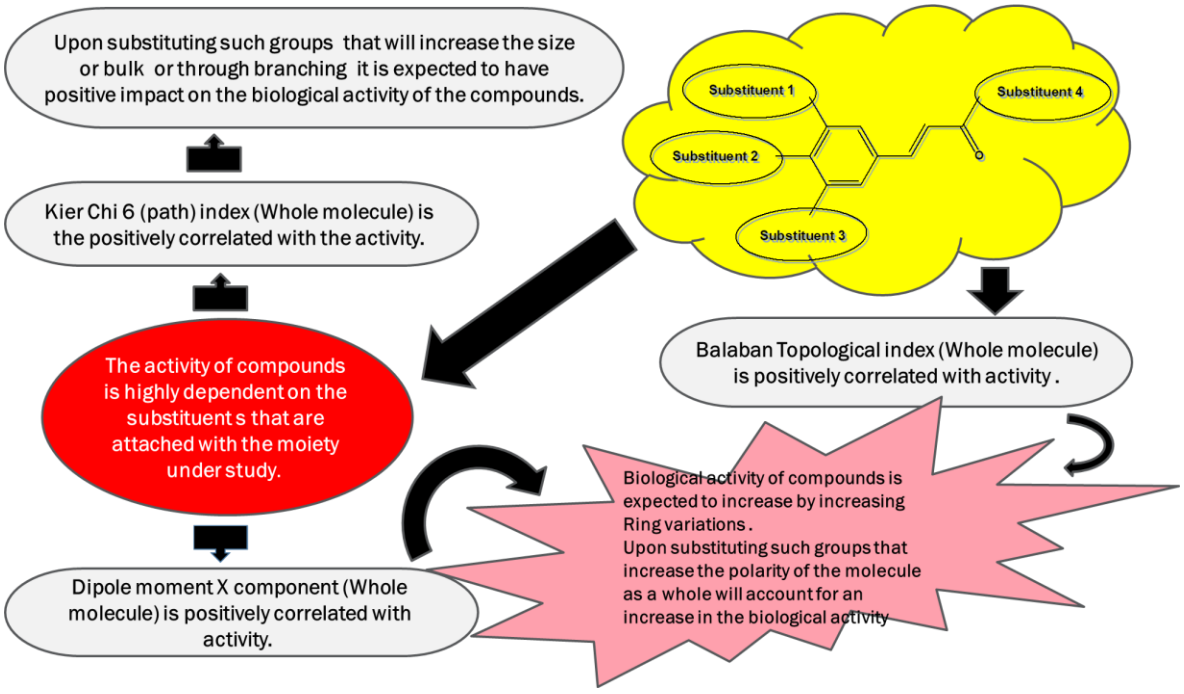


Figure 6: Effect of descriptors on different substituent's of piperazine derivatives.

Conclusion

The problem stemmed from the lack of selective ACAT inhibitors appealed to apply extensive efforts in designing innovative ACAT inhibitors. In order to achieve this goal, a validated, reliable and statistically significant QSAR model was established, and the information, conferred by the chosen descriptors, encouraged us to focus on the structural requirements of ACAT inhibitors with the standards providing enhanced selectivity. The knowledge bestowed was thus implemented to discover the expected novel entities. The MLR, PLS and FFNN techniques were employed to accomplish a highly predictive QSAR model provided with estimation of quality of prediction.

Acknowledgement

I am immensely gratified to the Vice Chancellor, Banasthali Vidyapith, my supervisor, my associates and my lab mates for providing incomparable support to me for my research work.

References

(1) Brown MS, Goldstein J. Lipoprotein Metabolism in the Macrophage: Implications for Cholesterol Deposition In Atherosclerosis. Annu. Rev. Biochem. **1983**; 52: 223-261.

(2) Nissen, SE, Tuzcu EM, Brewer HB, Sipahi I, Nicholls SJ, Ganz P, Schoenhagen P, Waters DD, Pepine CJ, Crowe TD, Davidson MH, Deanfield, JE, Wisniewski LM, Hanyok JJ and Kassalow LM. Effect of ACAT Inhibition on the Progression of Coronary Atherosclerosis. N Engl J Med. **2006**; 354: 1253-63.

(3) Gillies PJ, Robinson, CS, Rathgeb KA. Regulation of ACAT Activity by a Cholesterol Substrate Pool during the Progression and Regression Phases of Atherosclerosis: Implications for Drug Discovery. Atherosclerosis. **1990**; 83: 177-185.

(4) Miyazaki A, Kanome T and Watanabe T. Inhibitors of Acyl-Coenzyme A: Cholesterol Acyltransferase. Cardiovasc Hematol Disord Drug Targets. **2005**; 5: 463-469 463.

(5) Goodarzi M, Dejaegher B, and HeydenYV. Feature Selection Methods in QSAR Studies. J AOAC Int. **2012**; 95(3): 636-51.

(6) Chhabria MT, Brahmkshatriya PS, Mahajan BM, Darji UB and Shah GB. Discovery of Novel Acyl Coenzyme A: Cholesterol Acyltransferase Inhibitors: Pharmacophore-Based Virtual Screening, Synthesis and Pharmacology. Chem Biol Drug Des. **2012**; 80: 107–113.

(7) Patankar SJ and Jurs PC. Prediction of IC50 Values for ACAT Inhibitors from Molecular Structure. J. Chem. Inf. Comput. Sci. **2000**; 40(3): 706-723.

(8) Putz MV, Duda-Seiman C, Duda-Seiman D, Putz A-M, Alexandrescu I, Mernea M and Avram S. Chemical Structure-Biological Activity Models for Pharmacophores' 3D-Interactions. Int. J. Mol. Sci. **2016**; 17(7): 1087.

(9) Doreswamy and Vastrad CM. Predictive comparative QSAR analysis of sulfathiazole analogues as Mycobacterium Tuberculosis H37RV Inhibitors. **2012**; 13(3): 379-390.

(10) Jain S, Chauhan N, Sharma S, Reddy KR, Sadhu V and Kulkarni RV. The link between anxiety and Alzheimer's disease: critical facts. Nanomaterials in Diagnostic Tools and Devices. **2020**; 331-349.

(11) Jain S, Bisht A, Verma K, Negi S, Paliwal S, Sharma S. The role of fatty acid amide hydrolase enzyme inhibitors in Alzheimer's disease. Cell Biochem Funct. **2021**; 1–12.

(12) Jain S, Chauhan N, Bhardwaj A, Yadaw G, Singh MK, Mishra A. QSAR Modeling of α-Ketooxazole Motif Analogues as Potent Anti-Alzheimer Agents. YMER, **2022**, 21(5); 624-640.

(13) Nagpal A and Paliwal SK. QSAR Model Development using MLR, PLS and NN Approach to Elucidate the Physicochemical Properties Responsible for Neurologically Important JNK3 Inhibitory Activity. J. Pharm. Sci. & Res. Vol. **2017**; 9(10): 1831-1843.

(14) Mittal A, Sharma M and Singh A. QSAR Modelling of PDE5 Inhibitory Activity of Tetracyclic Guanine Derivatives as Antihypertensive Agents. Drug Dev & Res. **2016**; 8: 043-051.

(15) Ashima N and Sarvesh P. Design of novel JNK1 inhibitors using molecular modeling technique: An in silico approach. IJC-B; 58B (03): 403-415.

(16) Ojha P, Mishra P, Kesar S, Singh S. CHEMOMETRIC AND SIMILARITY BASED ANALYSIS OF DGAT-1 INHIBITORS. Int. j. adv. res. **2016**; 4(4): 780-806.

(17) Kesar S, Paliwal S, Sharma S, Mishra P, Chauhan M, Arya R, Madan K and Khan S. In-Silico QSAR Modelling of Predicted Rho Kinase Inhibitors Against Cardio Vascular Diseases. Curr. Comput.-Aided Drug Des. **2019;** 15(5): 421 – 432.

(18) Kesar S, Paliwal SK, Mishra P, Chauhan M. Quantitative Structure–Activity Relationship Analysis of Selective Rho Kinase Inhibitors as Neuro-regenerator Agents. Turk J Pharm Sci. **2019**; 16(2):141-154.

(19) Paliwal S and Sharma J. Quantitative structure activity relationship analysis of bisbenzofuran cations as antimalarial agents employing multivariate statistical approach. INDIAN J CHEM B. **2012**; 51(4):617-630.

(20) Singh S, Das S, Pandey A, Paliwal S and Singh R. Quantitative Structure Activity Relationship Studies of Topoisomerase I Inhibitors as Potent Antibreast Cancer Agents. J. Chem. **2013**; Article ID 849793: 9 pages.

(21) Khan S and Najmi AY. 2d Qsar Modeling Study of Acat Inhibitors as Potent Anti-Hyperlipidemic Agents. Asian J. Exp. Sci. **2022**; 36(2): 11-27.