Asian J. Exp. Sci., Vol. 36, No. 2, 2022 & Vol. 37, No. 1, 2023; 11-27

Online available on : www.ajesjournal.com, ISSSN : 0971-5444

2d QSAR Modeling Study of ACAT Inhibitors as Potent Anti-Hyperlipidemic Agents

Shagufta Khan[1,2]*, Asim Y. Najmi[2]

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

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

E mail: aanajmi@jazanu.edu.sa, khan.shagufta2@gmail.com

Received: Apri17, 2022; Revised: June 5, 2022; Accepted: June 17, 2022

Abstract : In order to interpret the evidence encrypted by the molecular structure of the compounds, standard physicochemical descriptors centered Quantitative Structure Activity Relationship (QSAR) approach was implemented on a data set of indoline derivatives reported to be acyl-coenzyme A (CoA): cholesterol acyltransferase ACAT inhibitors. Chemo metric models were designed by inserting a battery of statistical techniques in the current study that demonstrate the linear approaches of analysis including multiple linear regression (MLR), partial least square PLS and non-linear method such as artificial neural networks (ANN). The activity contributions of these molecules were analyzed through regression equation and the best QSAR model was created with an excellent correlative and predictive ability. Significant statistical values **S = 0.35, F = 60.30, r = 0.92, r^2 = 0.85, r^2 (CV) = 0.82** of the designed model were obtained using stepwise MLR and a comparable PLS and FFNN model with **r^2 (CV) = 0.82, 0.88 and 0.86** respectively. The model reveals that inertia moment 1 size, Kier Chiv4 (cluster) index, Kier Chiv6 (ring) index are prerequisite descriptors for determining further promising ACAT antagonist with high and liable potency against the target. The acquired physicochemical properties (electronic, topological, and steric) showed the important structural features required for activity against ACAT. Therefore, these features can be effectively employed for the modeling and screening of active agents as novel ACAT inhibitors.

Key Words: QSAR, ACAT, FFNN, MLR, PLS

Introduction

Hypercholesterolemia is an autonomous risk element ultimately associated to coronary heart disease, which is the prominent source of demise in industrialized nations and its burden on health facilities is growing worldwide. Cholesterol in the body is derived from two sources: endogenous biosynthesis and absorption from the diet. Inhibition of either process represents an attractive approach to lowering plasma cholesterol. Agents controlling total plasma cholesterol levels are expected to serve as an effective therapeutic method for atherosclerosis, since lowering plasma cholesterol levels has been proven to reduce mortality from myocardial infarction (Badimon *et al.*, 1993; Karam *et al.*, 2017; The Lancet, 1994; Shephard *et al.*, 1995).

Acyl-CoA: cholesterol O-acyltransferase (ACAT, EC 2.3.1.26) is an intracellular enzyme responsible for catalyzing the esterification of free cholesterol with fatty acyl-CoA to produce cholesteryl esters Brown *et al.* (1975). In mammalian species, two isoforms have been identified: ACAT-1 is predominant in human liver, macrophages, and adrenal gland, while ACAT-2 is present exclusively in hepatocytes and intestinal cells (Ta-Yuan Changa *et al.,* 2001). This enzyme plays important role in the absorption of dietary cholesterol from the small intestine, the secretion of very low density lipoprotein (VLDL) from the liver, and the accumulation of cholesteryl esters in atherosclerotic lesions. Inhibition of ACAT should reduce the absorption of cholesterol, lower plasma cholesterol levels (Krause *et al*., 1993; Largis*et al.*, 1989; Carr *et al.*, 1992; Burrier *et al.*, 1994; Brown *et al.*, 1983) and should arrest the progression and promote the regression of atherosclerotic plaque (Gillies *et al.,* 1990 and Sliskovic *et al.,* 1991). Therefore, ACAT inhibitors are a prime objective in the development of new therapeutic agents for hypercholesterolemia and atherosclerosis.

Quantitative Structure-Activity Relationship (QSAR) is one of the promising areas of research in medicinal chemistry and chemo metrics arena. It aims to derive relationship between the structural features or descriptors of the chemical entities and their own biological activity through linear or nonlinear mathematical equation. Thus, QSAR

studies provide useful information that how the structural features of a chemical or drug molecule influences the biological activity. In silico QSAR has become one of the advantageous approaches for bioactivity evaluation as compared to experimental testing (Kubinyi *et al.* 1997). The success of QSAR model depends on the quality of the input data, selection of appropriate descriptors and statistical methods to validate the developed model (Patankar *et al*., 2000 and Li *et al*. 2015). In this study, we report the development of a 2D-QSAR model by implementing diverse chemometric techniques to achieve improved inhibitory action towards ACAT enzyme as hypolipidemic agents which have not been explored so far. QSAR analysis has become an economic necessity to reduce the empiricism in drug design to ensure that every molecule designed and tested should be as meaningful as possible.

Materials and Methods

Outlining of primary structures and data set provision

The drawing of all designated compound structures from the reported series of in doline derivative Shoji *et al.* (2009) (Table-1), as ACAT inhibitors, was achieved using Chemdraw 8.0 followed by smoothing of their geometries via CLEAN STRUCTURE OPTION of chemdraw and the assemblies of all compounds were imported on TSAR worksheet. TSAR (TOOLS FOR STRUCTURE ACTIVITY RELATIONSHIP) is a unified compendium, which is used to analyze the types of interactions of QSARs. Importing of the IC_{50} values of the structures derived experimentally and reported in the series was done into TSAR work sheet as negative logarithm, which is used for QSAR model development. TSAR application used in this study is an assimilated analysis platform frequently used in therapeutic and pharmacological research. Upon utilizing the structural features of the molecule, TSAR creates a scientific model that revealed comprehensive analysis of correlation between the biological activity and chemical moieties.

Defining substituents and an enhanced 3D arranged structure building

A molecule template was chosen on the basis of its substitution pattern representing least obscurity and provided perfect picture of the structural design of the molecule. The present study defined five substituents (**R^1, R^2, R^3, R^4**and **R^5**) (Figure 1 and Table 2) surrounding a central nucleus via DEFINED SUBSTITUENTS OPTION present in TSAR worksheet's toolbar (version 3.3; Accelrys Inc., Oxford, England). Structures of the molecule were converted into 3D molecular structures of excellent features using CORNIA OPTION since some molecular properties are well defined by their 3D orientations (Dalby *et al.,*1992). CORNIA program results in an automatic generation of 3D molecule's atomic coordinates by applying information through connection table or linear string (Sadowski *et al.*, 1993). Energy calculation of the molecules was done through COSMIC FORCE FIELD, which is used to evaluate the total energy of a molecule via totaling up van der Waals, columbic, bond length, bond angle and torsion angle terms for all appropriate group of atoms. These calculations also include valence electrons from the molecular atoms and were observed terminated when the energy gradient became reduced to 1 × 10-5 and 1 × 10-10 kcal/mol, respectively (Kovatcheva *et al.,* 2003).

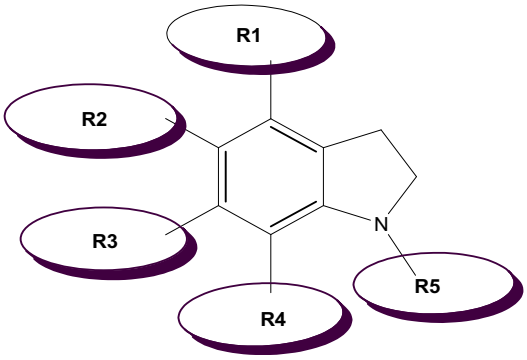


Fig. 1: Depicting substitution pattern around indoline nucleus.

Generation of Descriptors and data reduction

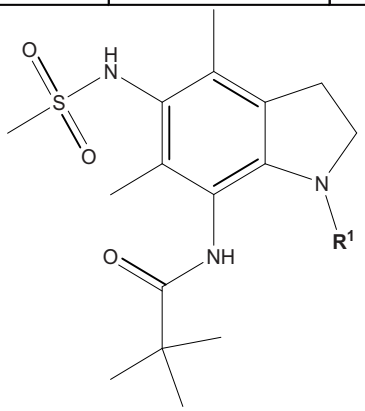
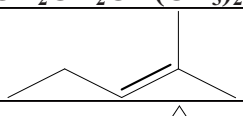
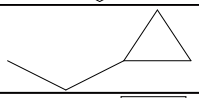
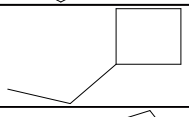
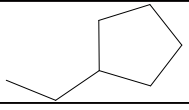
Generation of TSAR model desires descriptors to be portrayed numerically and exposed to correlation statistically. Molecular descriptor calculation for the whole molecule and their substituents was done, which differ in common positions of the standard moiety. Vast variety of descriptors such as electronic, topological, stearic, shape indices, connectivity, hydrophobic descriptors were obtained to display their physicochemical properties. While choosing only pertinent and noteworthy arrays of descriptors, data reduction was carried out to eliminate the risk of over fitting of the data and its redundancy. Descriptors possessing zero values for all the molecules were rejected. A correlation matrix strategy Paliwal *et al.* (2011) was opted to analyze the data pattern and also to eliminate the large set of descriptors pool. Therefore, the extremely inter-correlated descriptors possessing high correlation with biological activity Paliwal *et al.* (2009) were retained for the development of QSAR model, while others were rejected. Finally, the QSAR model comprised of three descriptors without any inter-correlation. The selected descriptors including, Inertia moment 1 size (Whole molecule)-X1, Kier

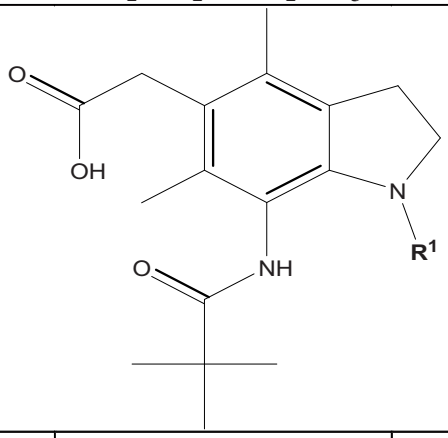
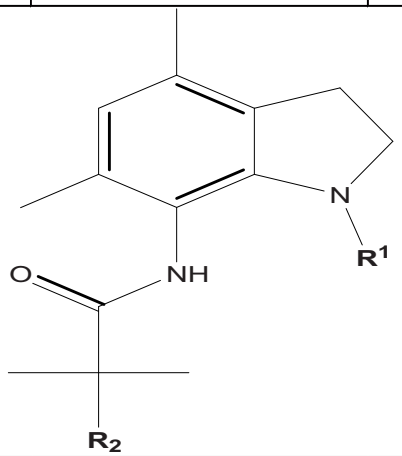
Chiv4 (cluster) index (Whole molecule)-X2 and Kier Chiv6 (ring) index (Whole molecule)-X3 were observed statistically significant and mainly correlated with the biological activity.

Training and Test set Congregation

Subsequently, a series of 36 compounds was arbitrarily segregated into the training and test set. It was split in such a way that both enclosed the compounds with various chemical structures and biological activity. A training set holding 22 compounds (Table-3) was engaged in the construction of the QSAR model for establishing a perfect relationship between the structural features and the biological activity. While, the test set enclosing 14 compounds (Table 3) was implemented to validate the predictive power of the developed model. Therefore, the selected compounds of the training and test set were engaged in the validation and development process of the proposed model. Few molecules may appear as outliers and they are not included in the process of model development as they represent the higher residual (beyond two orders of magnitude) values.

Table 1: Data set used for the development of the proposed QSAR model.

S. No.		R^1	R^2	$-LogIC_{50}$ (nm)
				
1	1a	CH_3-	-	8800
2	1b	CH_2CH_3-	-	1200
3	1c	$CH_2CH_2CH_3$-	-	240
4	1d	$CH_2(CH_2)_2CH3$-	-	69
5	1e	$CH_2(CH_2)_3CH3$-	-	49
6	1f	$CH_2(CH_2)_4CH3$-	-	42
7	1g	$CH_2(CH_2)_6CH3$-	-	13
8	1h	$CH(CH_3)_2$-	-	3700
9	1i	$CH_2CH(CH_3)_2$-	-	71
10	1j	$CH_2C(CH_3)_3$-	-	61
11	1k	$CH_2CH_2CH(CH_3)_2$-	-	40
12	1l		-	68
13	1m		-	160
14	1n		-	36
15	1o		-	95
16	1p	$CH_2CH_2OCH_3$-	-	270

17	**1q**	**$CH_2CH_2OCH_2CH_3$-**	**-**	**310**
				
18	**2a**	**$CH_2(CH_2)_4CH3$-**	**-**	**4300**
19	**2b**	**$CH_2(CH_2)_5CH3$-**	**-**	**5800**
20	**2c**	**$CH_2(CH_2)_6CH3$-**	**-**	**2300**
21	**2d**	**$CH_2(CH_2)_7CH3$-**	**-**	**1200**
22	**2e**	**$CH_2(CH_2)_8CH3$-**	**-**	**910**
23	**2f**	**$CH_2(CH_2)_9CH3$-**	**-**	**960**
24	**2g**	**$CH_2(CH_2)_{10}CH3$-**	**-**	**1100**
				
25	**3a**	**CH_3-**	**CH_3-**	**31000**
26	**3b**	**CH_2CH_3-**	**CH_3-**	**4000**
27	**3c**	**$CH_2CH_2CH_3$-**	**CH_3-**	**570**
28	**3d**	**$CH_2(CH_2)_2CH3$-**	**CH_3-**	**220**
29	**3e**	**$CH_2(CH_2)_3CH3$-**	**CH_3-**	**110**
30	**3f**	**$CH_2(CH_2)_4CH3$-**	**CH_3-**	**67**

31	**3g**	**$CH_2(CH_2)_5CH3$-**	**CH_3-**	**30**
32	**3h**	**$CH_2(CH_2)_6CH3$-**	**CH_3-**	**24**
33	**3i**	**$CH_2(CH_2)_8CH3$-**	**CH_3-**	**27**
34	**3j**	**$CH_2(CH_2)_{10}CH3$-**	**CH_3-**	**58**
35	**4a**	**$CH_2(CH_2)_2CH3$-**	**H-**	**3100**
36	**4b**	**$CH_2(CH_2)_2CH3$-**	**$-CH_2CH_3$**	**53**

Defining substituents and an enhanced 3D arranged structure building

A molecule template was chosen on the basis of its substitution pattern representing least obscurity and provided perfect picture of the structural design of the molecule. The present study defined five substituents (**R^1, R^2, R^3, R^4**and **R^5**) (Figure 1 and Table 2) surrounding a central nucleus via DEFINED SUBSTITUENTS OPTION present in TSAR worksheet's toolbar (version 3.3; Accelrys Inc., Oxford, England). Structures of the molecule were converted into 3D molecular structures of excellent features using CORNIA OPTION since some molecular properties are well defined by their 3D orientations (Dalby *et al.,*1992). CORNIA program results in an automatic generation of 3D molecule's atomic coordinates by applying information through connection table or linear string (Sadowski *et al.*, 1993). Energy calculation of the molecules was done through COSMIC FORCE FIELD, which is used to evaluate the total energy of a molecule via to taling up van der Waals, columbic, bond length, bond angle and torsion angle terms for all appropriate group of atoms. These calculations also include valence electrons from the molecular atoms and were observed terminated when the energy gradient became reduced to 1 × 10-5 and 1 × 10-10 kcal/mol, respectively (Kovatcheva *et al.,* 2003).

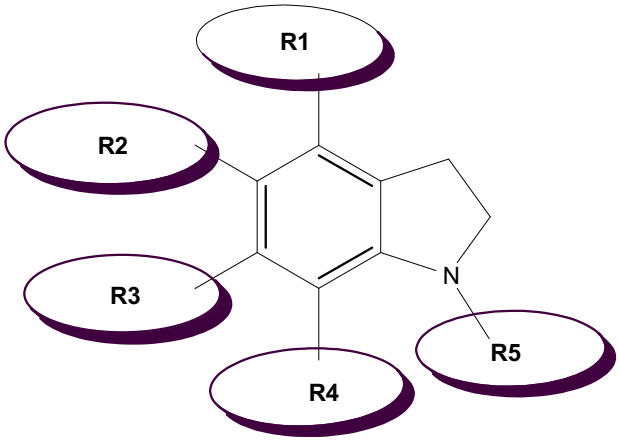
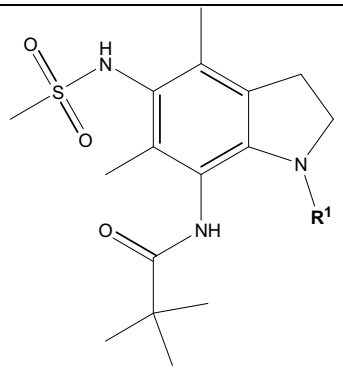


Fig. 1: Depicting substitution pattern around indoline nucleus.

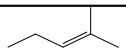
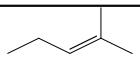
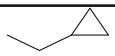
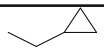
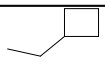
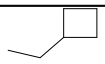
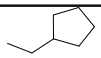
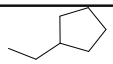
Table 2- : Data set with their defined substituents used for the development of the proposed QSAR model.

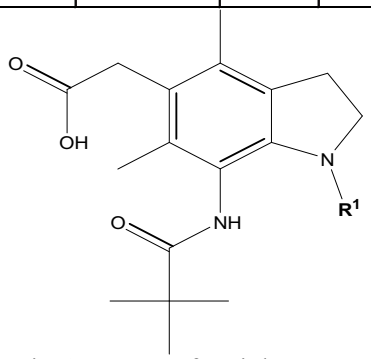
Indolidine derivatives



(Basis Structure of serial no: 1-17)

S. No.	Compound Name	Structure of Compound		Substituent's Defined For TSAR					IC$_{50}$ nM
		R_1	R_2	Subs 1	Subs 2	Subs 3	Subs 4	Subs 5	
1	1a	CH_3-	-	Me	-$NHSO_2CH_3$	Me	-$NHCOC(CH_3)_3$	CH_3-	8800
2	1b	CH_2CH_3-	-	Me	-$NHSO_2CH_3$	Me	-$NHCOC(CH_3)_3$	CH_2CH_3-	1200
3	1c	$CH_2CH_2CH_3$-	-	Me	-$NHSO_2CH_3$	Me	-$NHCOC(CH_3)_3$	$CH_2CH_2CH_3$-	240
4	1d	$CH_2(CH_2)_2CH_3$-	-	Me	-$NHSO_2CH_3$	Me	-$NHCOC(CH_3)_3$	$CH_2(CH_2)_2CH3$-	69
5	1e	$CH_2(CH_2)_3CH_3$-	-	Me	-$NHSO_2CH_3$	Me	-$NHCOC(CH_3)_3$	$CH_2(CH_2)_3CH3$-	49
6	1f	$CH_2(CH_2)_4CH_3$-	-	Me	-$NHSO_2CH_3$	Me	-$NHCOC(CH_3)_3$	$CH_2(CH_2)_4CH3$-	42
7	1g	$CH_2(CH_2)_6CH_3$-	-	Me	-$NHSO_2C$	Me	-$NHCOC(CH_3)_3$	$CH_2(CH_2)_6CH3$-	13

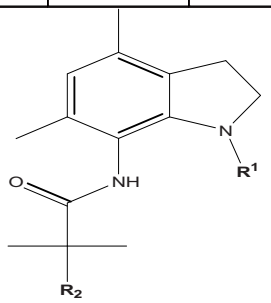
8	**1h**	**$CH(CH_3)_2$-**	**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**	**$CH(CH_3)_2$-**	**3700**
9	**1i**	**$CH_2CH(CH_3)_2$-**	**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2CH(CH_3)_2$-**	**71**
10	**1j**	**$CH_2C(CH_3)_3$-**	**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2C(CH_3)_3$-**	**61**
11	**1k**	**$CH_2CH_2CH(CH_3)_2$-**	**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2CH_2CH(CH_3)_2$-**	**40**
12	**1l**		**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**		**68**
13	**1m**		**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**		**160**
14	**1n**		**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**		**36**
15	**1o**		**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**		**95**
16	**1p**	**$CH_2CH_2OCH_3$-**	**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2CH_2OCH_3$-**	**270**
17	**1q**	**$CH_2CH_2OCH_2CH_3$-**	**-**	Me	**-$NHSO_2CH_3$**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2CH_2OCH_2CH_3$-**	**310**



(Basis Structure of serial no: 18-24)

18	**2a**	**$CH_2(CH_2)_4CH_3$-**	**-**	Me	**-CH_2COOH**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_4CH3$-**	**4300**

19	**2b**	**$CH_2(CH_2)_5CH_3$-**	**-**	Me	**-CH_2COOH**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_5CH3$-**	**5800**
20	**2c**	**$CH_2(CH_2)_6CH_3$-**	**-**	Me	**-CH_2COOH**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_6CH3$-**	**2300**
21	**2d**	**$CH_2(CH_2)_7CH_3$-**	**-**	Me	**-CH_2COOH**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_7CH3$-**	**1200**
22	**2e**	**$CH_2(CH_2)_8CH_3$-**	**-**	Me	**-CH_2COOH**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_8CH3$-**	**910**
23	**2f**	**$CH_2(CH_2)_9CH_3$-**	**-**	Me	**-CH_2COOH**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_9CH3$-**	**960**
24	**2g**	**$CH_2(CH_2)_{10}CH_3$-**	**-**	Me	**-CH_2COOH**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_{10}CH3$-**	**1100**



(Basis Structure of serial no: 25-36)

25	**3a**	**CH_3-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**CH_3-**	**31000**
26	**3b**	**CH_2CH_3-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**CH_2CH_3-**	**4000**
27	**3c**	**$CH_2CH_2CH_3$-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2CH_2CH_3$-**	**570**
28	**3d**	**$CH_2(CH_2)_2CH_3$-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_2CH3$-**	**220**

29	**3e**	**$CH_2(CH_2)_3CH_3$-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_3CH_3$-**	**110**
30	**3f**	**$CH_2(CH_2)_4CH_3$-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_4CH_3$-**	**67**
31	**3g**	**$CH_2(CH_2)_5CH_3$-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_5CH_3$-**	**30**
32	**3h**	**$CH_2(CH_2)_6CH_3$-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_6CH_3$-**	**24**
33	**3i**	**$CH_2(CH_2)_8CH_3$-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_8CH_3$-**	**27**
34	**3j**	**$CH_2(CH_2)_{10}CH_3$-**	Me	Me	**H**	Me	**-$NHCOC(CH_3)_3$**	**$CH_2(CH_2)_{10}CH3$-**	**58**
35	**4a**	**$CH_2(CH_2)_2CH_3$-**	H	Me	**H**	Me	**-$NHCOCH(CH_3)_2$**	**$CH_2(CH_2)_2CH_3$-**	**3100**
36	**4b**	**$CH_2(CH_2)_2CH_3$-**	Et	Me	**H**	Me	**-$NHCOCH(CH_3)_2CH_2CH_3$**	**$CH_2(CH_2)_2CH_3$-**	**53**

Generation of Descriptors and data reduction

Generation of TSAR model desires descriptors to be portrayed numerically and exposed to correlation statistically. Molecular descriptor calculation for the whole molecule and their substituents was done, which differ in common positions of the standard moiety. Vast variety of descriptors such as electronic, topological, stearic, shape indices, connectivity, hydrophobic descriptors were obtained to display their physicochemical properties. While choosing only pertinent and noteworthy arrays of descriptors, data reduction was carried out to eliminate the risk of over fitting of the data and its redundancy. Descriptors possessing zero values for all the molecules were rejected. A correlation matrix strategy Paliwal *et al.* (2011) was opted to analyze the data pattern and also to eliminate the large set of descriptors pool. Therefore, the extremely inter-correlated descriptors possessing high correlation with biological activity Paliwal *et al.* (2009) were retained for the development of QSAR model, while others were rejected. Finally, the QSAR model comprised of three descriptors without any inter-correlation. The selected descriptors including, Inertia moment 1 size (Whole molecule)-X1, Kier Chiv4 (cluster) index (Whole molecule)-X2 and Kier

Chiv6 (ring) index (Whole molecule)-X3 were observed statistically significant and mainly correlated with the biological activity.

Training and Test set Congregation

Subsequently, a series of 36 compounds was arbitrarily segregated into the training and test set. It was split in such a way that both enclosed the compounds with various chemical structures and biological activity. A training set holding 22 compounds (Table-3) was engaged in the construction of the QSAR model for establishing a perfect relationship between the structural features and the biological activity. While, the test set enclosing 14 compounds (Table 3) was implemented to validate the predictive power of the developed model. Therefore, the selected compounds of the training and test set were engaged in the validation and development process of the proposed model. Few molecules may appear as outliers and they are not included in the process of model development as they represent the higher residual (beyond two orders of magnitude) values.

Table - 3: Training and test set compounds for building QSAR model with their observed and predicted activities.

S. No.	Compound Name	Actual Activity	Predicted Activity		
			MLR	PLS	FFNN
Training Set Compounds					
1	1b	-3.07918	-3.38735	-3.32724	-3.26692
2	1d	-1.83885	-2.16615	-2.29113	-1.97524
3	1g	-1.11394	-1.59009	-1.80237	-1.44492
4	1h	-3.5682	-3.12522	-3.10484	-2.99905
5	1j	-1.78533	-1.74056	-1.93469	-1.78664
6	1k	-1.60206	-1.97691	-2.13056	-1.78906
7	1m	-2.20412	-2.31869	-2.42054	-2.13211
8	1n	-1.5563	-	-	-1.74528
9	1p	-2.43136	-2.05801	-2.19937	-1.86758
10	1q	-2.49136	-1.8063	-1.98581	-1.63078
11	2a	-3.63347	-3.73723	-3.61944	-3.6999
12	2b	-3.76343	-3.39141	-3.32603	-3.4871
13	2c	-3.36173	-3.66244	-3.55599	-3.65474
14	2d	-3.07918	-3.39211	-3.32663	-3.48754
15	2f	-2.98227	-2.63702	-2.68598	-3.00718
16	3a	-4.49136	-4.4078	-3.95854	-4.04417

17	**3d**	**-2.34242**	**-2.13839**	**-2.03308**	**-2.66763**
18	**3f**	**-1.82607**	**-1.54593**	**-1.5304**	**-1.42961**
19	**3g**	**-1.47712**	**-1.87599**	**-1.81044**	**-1.67971**
20	**3h**	**-1.38021**	**-1.48267**	**-1.47673**	**-1.41256**
21	**3i**	**-1.43136**	**-1.49778**	**-1.48955**	**-1.41639**
22	**4a**	**-3.49136**	**-3.43637**	**-3.36504**	**-3.56985**
Test set Compounds					
1	**1a**	**-3.94448**	**-3.41208**	**-3.58036**	**-2.47076**
2	**1c**	**-2.38021**	**-2.59713**	**-2.6899**	**-1.979**
3	**1e**	**-1.6902**	**-2.02592**	**-2.06575**	**-1.81536**
4	**1f**	**-1.62325**	**-1.80919**	**-1.82894**	**-1.77529**
5	**1i**	**-1.85126**	**-2.30527**	**-2.37099**	**-1.8827**
6	**1l**	**-1.83251**	**-1.60677**	**-1.60776**	**-1.74502**
7	**1o**	**-1.97772**	**-1.76977**	**-1.78587**	**-1.76891**
8	**2e**	**-2.95904**	**-2.85522**	**-2.87755**	**-2.64947**
9	**2g**	**-3.04139**	**-3.21641**	**-3.27221**	**-3.0074**
10	**3b**	**-3.60206**	**-3.47932**	**-3.45547**	**-3.46265**
11	**3c**	**-2.75587**	**-2.83582**	**-2.75234**	**-2.77481**
12	**3e**	**-2.04139**	**-2.31477**	**-2.18301**	**-2.19218**
13	**3j**	**-1.76343**	**-1.53285**	**-1.32863**	**-1.66838**
14	**4b**	**-1.72428**	**-1.42658**	**-1.38832**	**-1.75118**

Model development and statistical analysis

MLR, PLS and NN were used as statistical gears to develop the quantitative models representing vital structural features of the moieties and their purpose in analyzing the biological activity. Luco *et al*. (1997) In order to generate the BEST QSAR model, MLR utilizes the biological activity as dependent and the attained descriptors as independent variables. The proposed model truly figures out the association between X and Y variables. Least Inter-correlated variables and leave-one-out (LOO) approach was implemented for cross validation in MLR technique, where one compound is removed at a time and the remaining data set was used for calculations to obtain the best model. The generated model was revealed by means of definite statistical limitations such as r^2 (correlation coefficient), r^2 CV (cross validated coefficient, which defines the vital measure of the prognostic power of the model); lowest S value (that implies the standard error of the regression model) and high F value Mittal *et al.* (2016) (Fig.- 2). In order to validate the outcomes of MLR approach, the same data group was exposed to PLS study and the results were evaluated. Partial least square analysis was also performed to check the stability, precision and high predictivity of the generated models. This regression method is particularly beneficial when the independent variables (descriptors) are comparable to or greater than the data points (compounds) Mittal *et al.* (2016). After the accomplishment of these parametric measures, the concerned model can be regarded as consistent, robust and predictable.

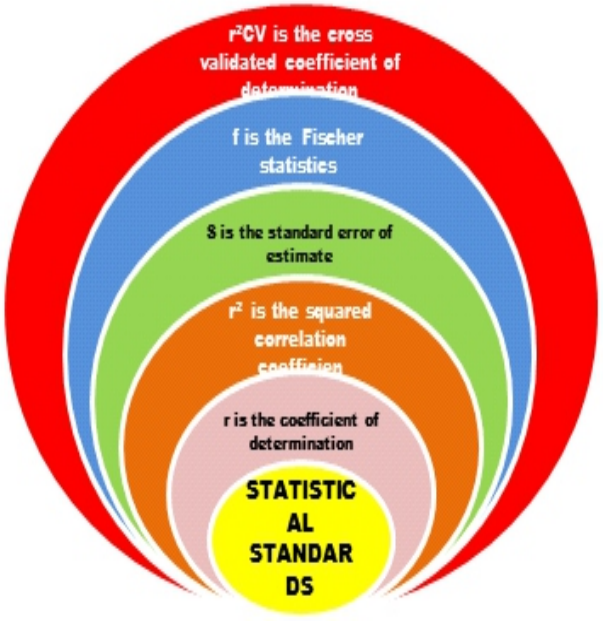


Fig.- 2: A set of statistical standards representing the QSAR models.

Forward Feed Neural Network (FFNN) Study

FFNN is also called as Neural network analysis, which is classified as a computer based platform representing the interconnections of simulated neurons in a mesh-like configuration that forms "layers" Rinaldi *et al.* (1973). Each neuron carries a definite variable value and an input layer constitutes the collection of values obtained from independent variables via input neurons. Results received from hidden neurons by accumulating values from other neurons were transferred to descendant neurons. Signals acquired from other units by output neurons correlates with dependent variables. Therefore, a complete neural network is projected as i-h-o, where i, h, and o represent input, hidden, and output layers, respectively. Similar set of descriptors that were used in MLR and PLS techniques were employed in FFNN model to judge its predictive power. 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 more 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

QSAR study is among the best suitable practices for optimizing lead moieties to propose a novel compound. Modeling via QSAR approach leads to finest equation that should exploit least number of descriptors to attain a BEST fit. Huge collection of data with diverse analogues was employed for the current study. Their biological activity (Log/IC_{50}) and chemical structures are represented in Table 1 and 2. In view of which key molecular features are significantly affecting the ACAT inhibition, three chemo metric approaches, MLR, PLS, and FFNN were employed to hypothesize conventional descriptor-based QSAR models. The structural scheme of 36 molecules was segregated as the training set of 22 compounds and the test set of 14 compounds to assess the predictive power of the generated 2D QSAR model via linear and nonlinear statistical techniques.

Primarily, around 200 classical descriptors

(geometric, electronic, topological, constitutive, etc.) were obtained through regression approach. Because of the profuse and redundant data, it results in an insignificant value of r2 CV (0.301), suggesting unsatisfactory prognostic power of the generated model. Therefore, there was a vital necessity for reliable and instructive descriptor set, owning the perfect correlation with biological activity with no inter-correlation. Subsequent removal of unwanted set of descriptors, the correlation matrix was applied and eventually an absolute set of three distinct descriptors was obtained, which were independent of each other and found momentous in model expansion. The ultimate chosen descriptors were Inertia moment 1 size (Whole molecule)-X1, Kier Chiv4 (cluster) index (Whole molecule)-X2 and Kier Chiv6 (ring) index (Whole molecule)-X3.

The eminence of the created model is typically projected in terms of statistical feature r, wherer^2 signifies the data percentage obtained via linear regression equation. Its value for the developed model was 0.92, which infers 92% variance in experimental values. High value of r^2CV defines the prognostic power of the model which indicates the outstanding quality of the developed model. The standard error s in the derived regression equation must be extremely low (0.35) which revealed that the chance of error in the created model is very less. F test value indicates the ratio of variance shown by the generated model and it is the variance due to error in the regression equation. Therefore, its greater values (60.31) signify the superior statistical quality of the developed model, which is highly consistent, robust and with exceptional prognostic power.

The final equation produced via MLR analysis is given in equation 1.

Y = 0.014418198*X1 + 1.5526917*X2 + 200.23163*X3 – 15.558502(1)

Where, X1= Inertia moment 1 size (Whole molecule), X2= Kier Chiv4 (cluster) index (Whole molecule) X3= Kier Chiv6 (ring) index (Whole molecule)and Y denotes the biological activity.

Subsequently, the similar training set was utilized for PLS evaluation, which further confirmed the statistical supremacy of the developed QSAR model because this assessment practice essentially cross validated the results of the MLR and showed minimal variation. The equation retrieved from PLS analysis is as follows:

Y = 0.013121762*X1 + 1.0500141*X2 + 202.58466*X3 -15.10587 (2)

Where, X1= Inertia moment 1 size (Whole molecule), X2= Kier Chiv4 (cluster) index (Whole molecule) X3= Kier Chiv6 (ring) index (Whole molecule) and Y denotes the biological activity.

Preferably, the values of r^2 and r^2 CV must be >0.6 and >0.5 respectively so as to obtain a perfect and a flawless 2D QSAR model. The current study therefore represented an extremely reliable and highly predictable 2D QSAR model that showed an outstanding statistical relevance with r^2 = 0.85and r^2CV= 0.82 respectively. Furthermore, the data for other statistical parameters retrieved through MLR regression equation including Fischer test value (F), Standard deviation (s), Coefficient of correlation (r), Squared correlation coefficient (r^2) and Cross validated coefficient of determination (r^2CV) are given in the figure 3.

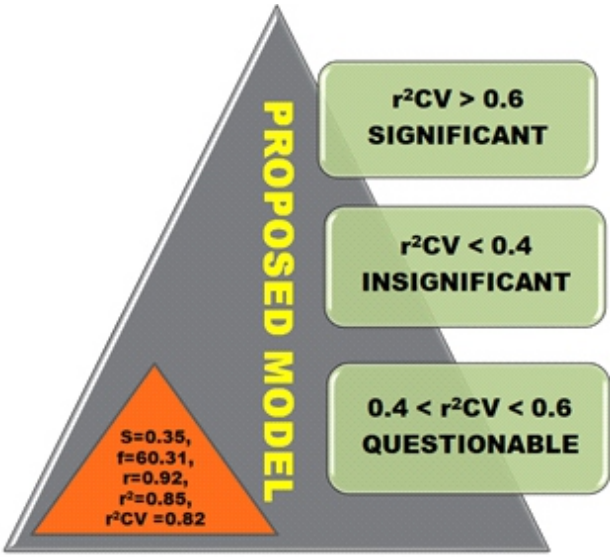


Fig. 3: A highly predictable 2D QSAR model with an outstanding statistical relevance.

The **r^2** value of the training and test set of MLR and PLS is greatly comparable (MLR, PLS $_{\text{Training set}}$) r^2 0.88 & 0.86 and (MLR, PLS $_{\text{Test set}}$) r^2 0.86& 0.85 respectively. The minor is the difference; the greater will be the stoutness and predictive ability of the model. Similarly, the apparent relationship between the observed and predicted values of the training and test set (Table-3) stipulates the importance of the derived MLR and PLS equations for statistical evaluation. Figure 4 and Figure 5 portrays the graphical representation of the actual and predicted values of training and test set compounds obtained via MLR and PLS techniques respectively. The greater correlation coefficient (0.86) value observed for the test set compounds further demonstrate the robustness of the produced model.

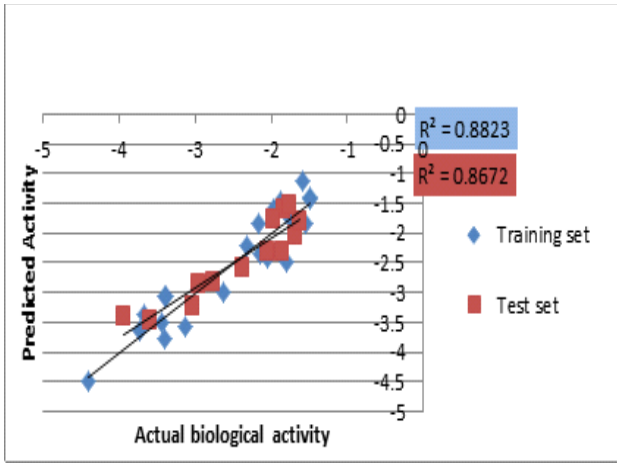


Fig. 4 - MLR Graph representing plot between actual verses Predicted activity.

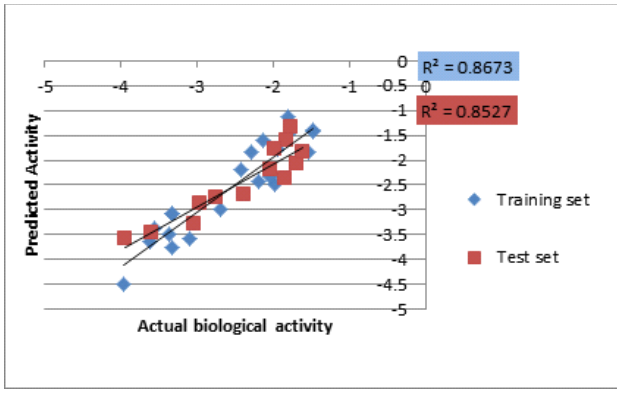


Fig. 5 - PLS Graph representing plot between actual verses predicted activity

In order to accomplish the aim of this study, an Artificial Neural Network (ANN) methodology was implemented, which confirmed the dependency of the obtained descriptors with the biological activity, and hence it ascertained the effectiveness and firmness of the created 2D-QSAR model. The current study covered three hidden neurons with output level displaying experimentally evaluated logIC_{50} values of the biological activity. An absolute structural design of the produced ANN model indicated the net configuration of (3-3-1) and the $\mathbf{r}^2$ for training and test set was 0.87 and 0.72 respectively. An ANN graph of observed and predicted activity (Fig.- 6) further authenticates that the developed model indicated a commendable statistical quality. The dependency plots (Figs. 7, 8, & 9) achieved through FFNN (Forward Feed Neural Network) finally predicted that the dependence of biological activity on the structural architecture even more consistent.

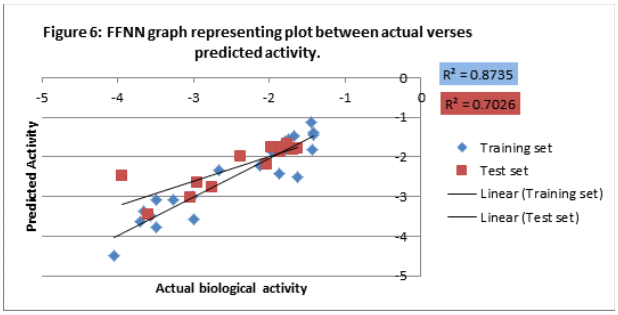


Fig. 6 - FFNN Graph representing plot between actual verses predicted activity.

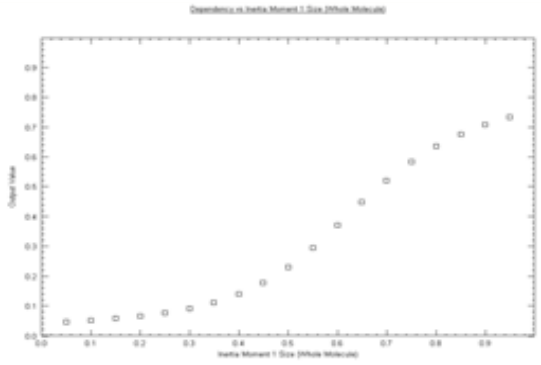


Fig. 7: Dependency graph of neural analysis demonstrating correlation between Inertia Moment 1 Size (whole molecule) and actual activity data.

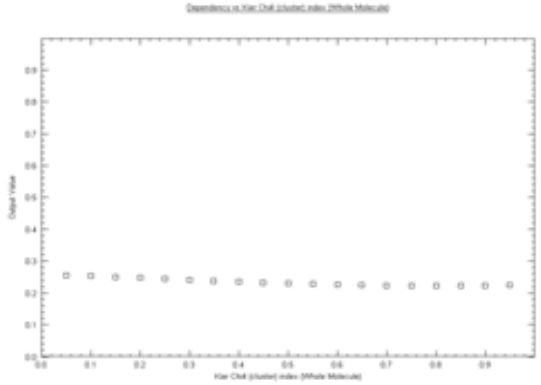


Fig. 8: Dependency graph of neural analysis demonstrating correlation between Kier Chi4 (cluster) index (whole molecule) and actual activity data.

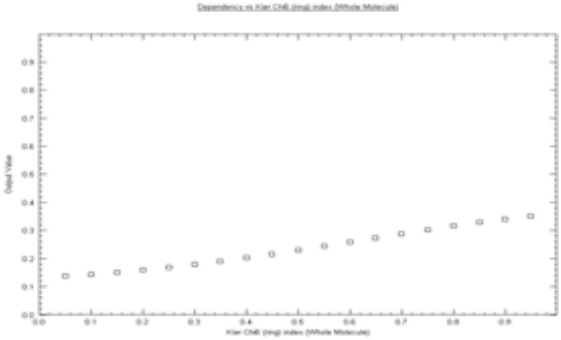


Fig. 9: Dependency graph of neural analysis demonstrating correlation between Kier Chi6 (ring) index (whole molecule) and actual activity data.

Linear regression study illustrates that Inertia moment 1 size (Whole molecule), Kier Chiv4 (cluster) index (Whole molecule) and Kier Chiv6 (ring) index (Whole molecule) are the main autonomous physicochemical (descriptors) features wrapping the entire model and expressing clear association with the biological activity. The negative coefficient of the Kier Chiv4 (cluster) index (Whole molecule) directed that the declining value of it may result in an improved biological activity, however the positive coefficients of the Inertia moment 1 size (Whole molecule) and Kier Chiv6 (ring) index (Whole molecule) showed that their increasing values can augment the ACAT inhibitory activity.

Interpretation of descriptors attained: Upon studying the complexity of the structural architecture, of the chosen set of molecules, suggested interesting evidences pertaining to impact of the retrieved descriptors on ACAT inhibitory activity. Importance of molecular descriptors presented in the recent study 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. Remarkably, the final model is observed well interrelated with the obtained three descriptors and showed excellent correlation to the change in pharmacological activity upon varying the substitution around the principal nucleus. (Fig.- 10).

Physical magnitudes signifying principal moment of inertia (PMI) are associated with the molecular rotational dynamics. The moment of inertia about an axis (equation 1) is:

$PMI = \sum_{1}^{A} mi.ri^{2}$ (1)

Moment of inertia is a steric factor and its value depends on the overall molecular mass, molecular distribution and a point of molecular rotation about its axis. It is a highly advanced spatial descriptor that signifies an optimal bulk, shape, orientation and the susceptibility of a molecule to attain various rotational transitions. The greater positive correlation coefficient of inertia moment proposes that the alignment performance pertaining to the size of the whole molecule is of extreme significance in improved molecular interaction with its receptor binding site and hence, the binding affinity. Therefore, it indicates that the substitution that increases the shape as well as the mass distribution of the whole molecule is expected to make a significant influence in enhancing the ACAT inhibitory activity. This is additionally supported by the FFNN dependency graph (fig.- 7) in which log IC^{50} increases with the increasing values of the inertia moment 1 size (WH). The increasing patterns are observed consistent with our MLR and PLS models (the coefficients are positive for the descriptors).

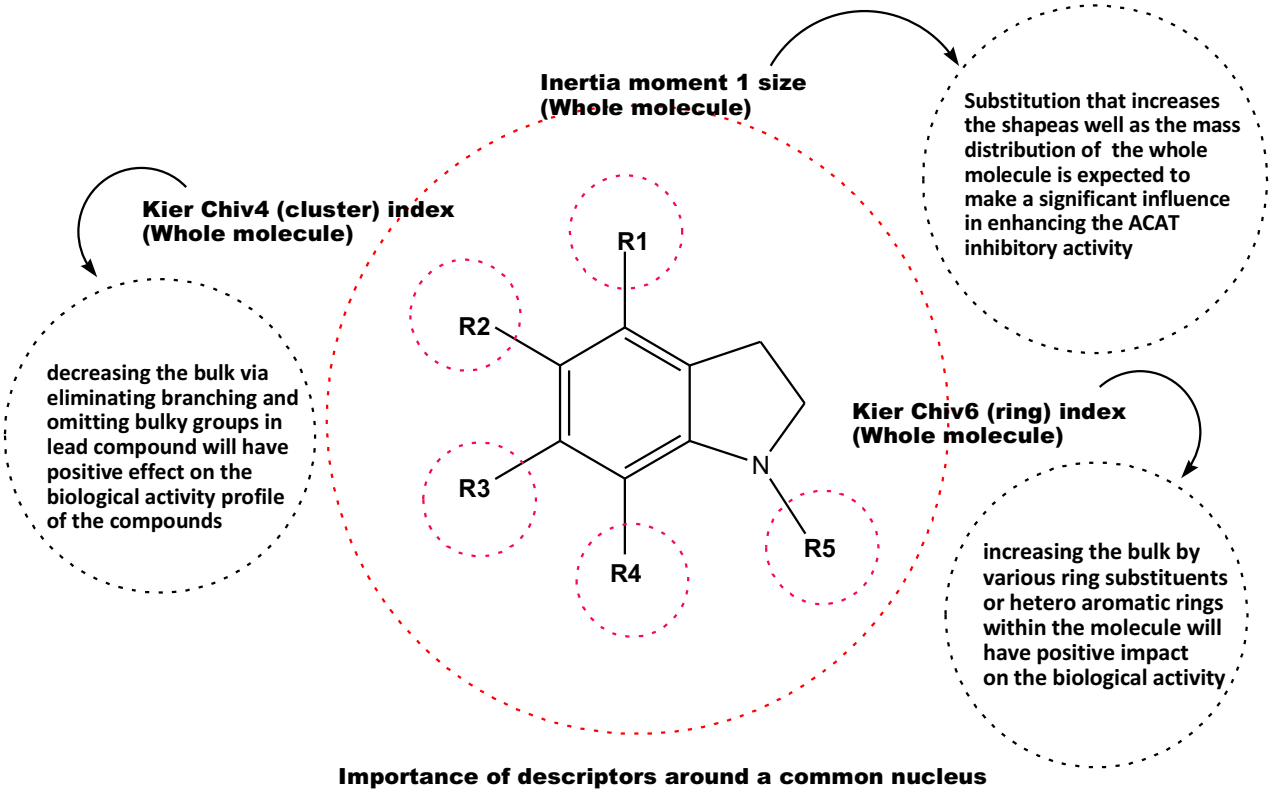


Fig. 10: Effect of descriptors on different substituent's of Indoline derivatives.

For a drug to interact effectively with its binding site, its orientation plays a major role, as it provides a perfect shape to the entire molecule, and therefore the best binding affinity. In contrast, the toxicity and the ADME profile is dependent on the mass of the molecule. Hence, a drug is found to be safe and effective if it possesses an optimal shape and mass.

A topological descriptor, Kier chi v4 (cluster) index (whole molecule), is the fourth order cluster valence connectivity index encrypts steric properties and structural intricacy of the whole molecule. Scientists named Kier and Hall established and well-defined this descriptor, which actually emphasize on the degree of branching in a molecule that provides information regarding bulk, size and the overall shape of the molecule. This supports in understanding the features that a molecule as a whole must possess in order to exhibit resilient and active interaction with the preferred target. It is often seen that the steric hindrance is undesirable, as it is triggered by the bulky groups present in the lead molecule, which further exert negative impact in bonding process of a molecule with its receptor. Therefore, an optimum extent of bulk and branching is required to assists in better alignment of the compound which augments the strong bonding interactions between the molecule and its receptor. As, Kier chi v4 (cluster index) (whole molecule) is observed negatively correlated with the bioactivity in the regression equation, decreasing the bulk via eliminating branching and omitting bulky groups in lead compound will have positive effect on the biological activity profile of the compounds. In the FFNN dependence graphs of the pentaparametric model (Figure 8), log IC_{50} decreases with the total Kier chi v4 (cluster index) (WM). This is consistent with the negative coefficients of final MLR and PLS models.

Among the three parameters, Kier Chi v6 (ring) index (Whole molecule) was found out to be an essential descriptor in clarifying the biological activity of indoline derivatives as ACAT inhibitors. It was primarily elaborated by Randic and consequently by Kier and Hall, that revealed number of series through 'order' and 'subgraph' type. A chi index prefers calculation of a standard variety of sub graphs for example P (path), C (cluster), PC (path/ cluster), and CH (ring), which are analyzed through delta values as a function. This descriptor defines various features associated with atomic bonding within the molecular structure. It furthers assists in exploring the substituents arrangements in the phenyl ring and the amount of branching rings. (27) The present study showed that the Kierchiv6 (ring) index (Whole molecule) descriptor is positively associated with bioactivity (Fig.-9); which essentially demonstrates structural complexity including the heteroatom content as well as the size of the ring systems. As it is observed that by varying the size of a ring or altering it with different hetero aromatic rings, the molecule is expected to have an optimum level of bulk, which further aids in better orientation of the compound with enhanced bonding interactions of a molecule with is target. Therefore, upon increasing the bulk by various ring substituent's or hetero aromatic rings within the molecule will have positive impact on the biological activity. There is steady rise observed with MLR and PLS models, with the positive coefficients for the descriptors.

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.

Reference

Badimon J. J. Fuster V. Chesebro J. H. and Badimon L. (1993). Coronary Atherosclerosis. A multifactorial disease. *Circulation*., 87: (3 Suppl): II 3—16.

Brown M. S. and Goldstein J. L. (1983). Lipoprotein metabolism in the macrophage: implications for cholesterol deposition in atherosclerosis. *Annu. Rev. Biochem.,* 52: 223—261.

Brown M. S. Dana S. E. and Goldstein J. L. (1975). Cholesterol ester formation in cultured human fibroblast. Stimulation by oxygenated sterols. *J. Biol. Chem.,* 250: 4025–4027.

Burrier R. E. Deren S. McGregor D. G. Hoos L. M. Smith A. A. and Davis Jr. H. R. (1994). Demonstration of a direct effect on hepatic acyl Coa:cholesterol acyl transferase (ACAT) activity by an orally administered enzyme inhibitor in the hamster. *Biochem. Pharmacol*., 47: 1545—51.

Carr T. P. Parks J. S. and Rudel L. L. (1992). Hepatic ACAT activity in African green monkeys is highly correlated to plasma LDL cholesteryl ester enrichment and coronary artery atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.,* 12: 1274 –

1283.

Chang T. Y. Chang C. C. Y. Lin S. Yu C. Li B. L. and Miyazaki A. (2001). Roles of acyl-coenzyme A: cholesterol acyltransferase-1 and -2. *Curr. Opin. Lipidol.*, . **12**: 289-296.

Dalby A. Nourse J. G. Hounshell W. D. Gushurst A. K. I. Grier D. L. Leland B. A. and Laufer J. (1992). Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.,* 32(3): 244 – 255.

Gillies P. J. Robinson C. S. and Rathgeb K. A. (1990). Regulation of ACAT activity by a cholestrol substrate pool during the progression and regression phases of atherosclerosis: implications for drug discovery. *Atherosclerosis.*, 83(2-3): 177— 185.

Karam I. Yang Y. J. and Li J.Y. (2017). Hyperlipidemia Background and Progress. *SM J Cardiolog and Cardiovasc Disord.*, 3(2): 2-8.

Scandinavian Simvastatin Survival Study Group (1994). Randomized trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian simvastatin survival study (4S). *Lancet*., *344(8934):* 1383-1389.

Kovatcheva A. Buchbauer G. Golbraikh A. Wolschann P. (2003) QSAR Modeling of α-Campholenic Derivatives with Sandalwood Odor. *J. Chem. Inf. Comput. Sci.*, 43(1): 259 – 266.

Krause B. R. Anderson M. Bisgaier C. L. Bocan T. Bousley R. DeHart P. Essenburg A. Hamelehle K. Homan R. Kieft K. McNally W. Stanfield R. and Newton R. S. (1993). In vivo evidence that the lipid-regulating activity of the ACAT inhibitor Cl-976 in rats is due to inhibition of both intestinal and liver ACAT. *J. Lipid Res*. **34**: 279—294.

Kubinyi H. (1997). QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discov Today,* 2(11): 457 – 467.

Largis E. E. Wang C. H. DeVries V. G. and Schaffer S. A. (1989). CL 277, 082: a novel inhibitor of ACAT-catalyzed cholesterol esterification and cholesterol absorption. *J Lipid Res*., **30**: 681—690.

Li X. Zou Y. Zhao Q. Yang Y. Wu M. Huang T. Hu H. and Wu Q. (2015). Synthesis, biological evaluation, and molecular docking studies of xanthone sulfonamides as ACAT inhibitors. *Chem. Biol. Drug Des*., 85(3): 394–403.

Luco J. M. and Ferretti F. H. (1997). QSAR Based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives. *J. Chem. Inf. Comput. Sci.* 37(2): 392 – 401.

Mittal A. Sharma M. and Singh A. (2016). QSAR Modelling of PDE5 Inhibitory Activity of Tetracyclic Guanine Derivatives as Antihypertensive Agents. *Int. J. Drug Dev. & Res.*, 8(3): 043 – 051.

Paliwal S. Narayan A. Paliwal S. (2009). Quantitative Structure Activity Relationship Analysis of Dicationic Diphenylisoxazole as Potent AntiTrypanosomal Agents. *QSAR Comb. Sci.*, 28(11-12): 1367-1375.

Paliwal S. K. Pandey A. and Paliwal S. (2011). Quantitative Structure Activity Relationship Analysis of N-(mercaptoalkanoyl)- and [(acylthio)alkanoyl] Glycine Derivatives as ACE Inhibitors. *Am. J. Drug Discov. Dev.*, 1(2): 85-104.

Patankar S. J. and Jurs P. C. (2000). Prediction of IC_{50} Values for ACAT Inhibitors from Molecular Structure. *J. Chem. Inf. Comput. Sci.,* 40(3): 706-723.

Rinaldi D. and Rivail J. L. (1973). Molecular polarizability and dielectric effect of medium in the liquid phase. Theoretical study of the water molecule and its dimers. *Theor. Chim. Acta*., 32: 57-70.

Sadowski J. and Gasteiger J. (1993). From atoms and bonds to three-dimensional atomic coordinates: automatic model builders. *Chem. Rev*., 93(7): 2567 – 2581.

Shephard J. (1995). The West of Scotland Coronary Prevention Study: a trial of cholesterol reduction in Scottish men. *Am J Cardiol*., 76 (9): 113C-117C.

Shoji Y. Takahashi K. Ohta M. Kasai M. Kunishiro K. Kanda M. Yogai S. Takeuchi Y. and Shirahase H. (2009). Novel indoline-based acyl-CoA: cholesterol acyltransferase inhibitor: Effects of introducing a methanesulfonamide group on physicochemical properties and biological activities. *Bioorg. Med. Chem.*, 17(16): 6020–6031.

Sliskovic D. R. and White A. D. (1991). Therapeutic potential of ACAT inhibitors as lipid lowering and anti-atherosclerotic agents. *Trends Pharmacol. Sci.,* 12(5): 194—199.

The Lipid Research Clinics Coronary Primary Prevention Trial Results. II. The Relationship of Reduction in Incidence of Coronary Heart Disease to Cholesterol Lowering. *JAMA.,* 251(3): 365-74