Indian Journal of Chemistry
Vol. 63, February 2024, pp. 190-202
DOI: 10.56042/ijc.v63i2.6141

Quantitative structure activity relationship studies of potent Endothelin-A receptor antagonist for the treatment of pulmonary arterial hypertension

Pragya Sharma*, Sarvesh Paliwal, Suman Sharma, Neha Chauhan & Smita Jain

Department of Pharmacy, Banasthali Vidyapith, Banasthali, Rajasthan 304 022, India

E-mail: pragya19977@gmail.com, paliwalsarvesh@yahoo.com, suman.pharmacy91@gmail.com, nehachauhan151@gmail.com, smitajain1994@gmail.com

Received 11 October 2023; accepted (revised) 25 January 2024

A traditional physicochemical descriptors-based QSAR analysis has been conducted on a data-set of Endothelin-A receptor antagonists for the treatment of pulmonary arterial hypertension. A variety of statistical techniques, including non-linear techniques like artificial neural networks and linear analytical techniques *i.e.*, 'Multiple Linear Regression' and 'Partial Least Squares' are implicated in current research. The development models have then been put through a validation process including the leave one out, which supported their high predictability and accuracy. A few statistical parameters have been used to build the model's predictive power and the resulting model has been found to have good statistical values, such as s=0.40, f=48.75, r=0.87, r^2=0.77, r^2CV=0.71 for training set. Three descriptors, including logP (whole molecule), total lipole (whole molecule), VAMP LUMO (whole molecule) are made relevant by the general model, which offers insightful information. As new results, these traits may be successfully used for the modelling and screening of new endothelin-A receptor antagonists that are active hypertensive drugs.

Keywords: Endothelin-A receptor antagonists, Quantitative structure activity relationship, Statistical analysis, Pulmonary arterial hypertension

Hypertension, also called as increased blood pressure, is a condition in which an arteriole has an abnormal high pressure for an expand duration. The capillary circulates blood through heart to each and every part of the body. Perpetually the heart beats; blood is transported within the veins. The impact of blood pressing on the walls of arteries as it pumped by the heart caused blood pressure. According to data released in 2017, hypertension or high blood pressure affects three in ten Indians and is responsible for 17.5% of all deaths in India[1].

Pulmonary hypertension is a substantial global health issue[2]. 'Pulmonary Arterial Hypertension' (PAH), chronic situation marked by unusually high B.P. in the pulmonary artery, which transports blood from the heart into lungs. PAH is a kind of pulmonary hypertension (PH) and dyspnea *i.e.*, shortness of breath (after exertion), angina, or lose consciousness are all symptoms of PAH. The specific aetiology of PAH is imprecise, and while it may be treated, there is no known cure. PAH is more common in women between the ages of 30 and 60. About 15-20% of individuals with PAH have genetic variants of the disease. Person with genetic PAH have either: (1)an autosomal dominant heritable disorder linked to mutations in BMPR2 gene and other newly discovered gene's linked to HPAH or other forms of PAH, and related complications like "pulmonary capillary hem angiomatosis" or "pulmonary veno-occlusive disease", (2)members of a family where PAH has been linked with genetic disorders.

There are three basic mechanisms implicated in aberrant proliferation, contraction of the smooth muscle's cells of the pulmonary artery. Currently, targets available for therapy in pulmonary arterial hypertension *i.e.*, nitric oxide, prostacycline pathway and endothelin pathway[3]. A novel target which is used for PAH is Endothelin receptor. Endothelin [ET]-potent vasoconstrictor peptide which was early isolated from the cultured procine endothelial cells. It contains 21 amino acids residue. There are three isoforms of ET that is ET-1, ET-2 and ET-3. ET receptor belongs to the family of GPCR's and there are two types of ET receptor that is ET_A and ET_B[4,5]. ET_A receptor mainly presents in smooth muscles cells and cardiac myocytes where ET_B receptor are found in both endothelial as well as smooth muscle cells. The mechanism of action on which ET_A receptor works as shown in Fig. 1.

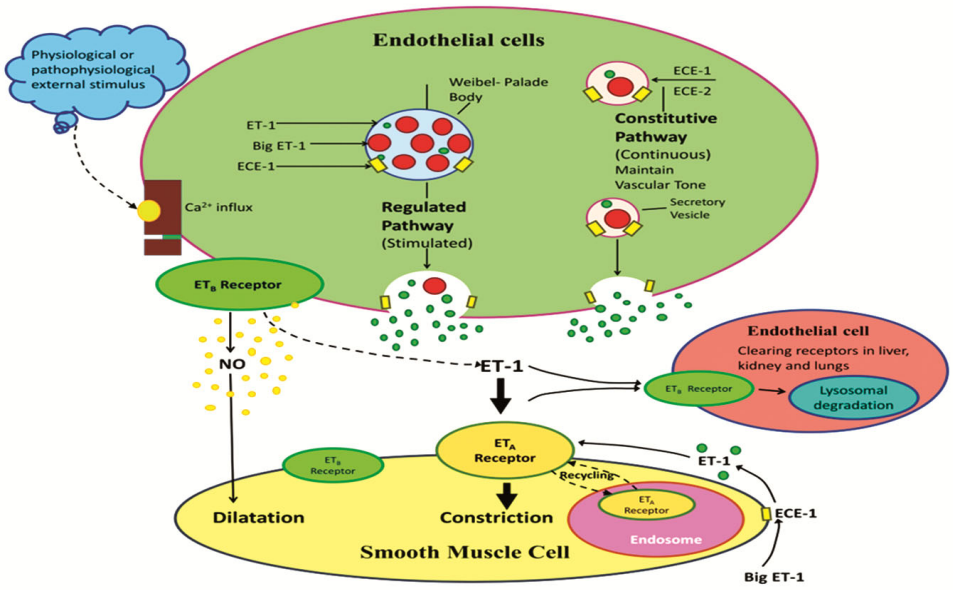


Fig. 1 — Mechanism of action of ET_A receptor

In this figure, ET-1 is produced *via* a dual secretory route in human endothelial cells. ET-1 is constantly produced from the intrinsic pathway's tiny vesicles to collaborate with ET receptors and contributing to vasomotor tone. ET-1 is secreted *via* a control pathway in response to extrinsic stimuli from the Weibel-Palade body of endothelial cells. ET-1 produced from endothelial cells primarily synergizes with ET_A receptors located beneath the smooth muscle of human blood vessels, with a limited number of ET_B receptors causing contraction of some, but not for all, arteries. Intrinsic inhibition of the ET-1/ET_A complex in the endosperm precedes cell surface receptor recycling, thus providing a mechanism for the ET_A antagonist to reverse the response of ET-1. Some ET-1 may also have auto-secretory interactions with endothelial ET_B receptors, limiting the spasmodic response by releasing vasodilators *i.e.*, nitric oxide. The ET_B antagonist cannot replace the receptor-binding ligand because the ET-1/ET_B complex is integrated and destroyed in the lysosome. Big ET-1 can also be converted to mature peptides by smooth muscle ECE[6–9].

Materials and Methods

Data size and variety

The selected series must have more than 30 compounds and the difference between the biological activity of most active and least active compounds[10] should be more than 3.2 orders of magnitude.

Data set preparation

The structures of selected series were drawn using ChemDraw and import in Tools for structure-activity relationship(TSAR) window. Prior to the molecular descriptor generation, the structure of all series was subjected to CORINA make 3D to make 3D structure of small, medium sized typically drug-like molecules. COSMIC force field was modified and extended and calculated molecular energies like bond length,bond angle,Vander-walls for all compounds.

Descriptor calculation and Data Reduction

Calculating descriptors is primarily done to decode information about the physicochemical characteristics of each molecule responsible for a certain biological action[11]. More than 300 descriptors were calculated in the TSAR sheet. Out of these descriptors only few shows significant correlation with the activity for the final model. Then,data reduction was done to get rid of data duplication and the incidence of accidental correlations in order to identify meaningful sets of descriptors. Firstly, those descriptors having '0' value for all compounds were deleted. For further reducing the data, correlation matrix was generated between two descriptors. The level of co-linearity is indicated by the correlation terms included in the correlation matrix. Out of the two descriptors having inner-correlation coefficient above 0.5 was kept and remained was deleted. This process is repeated for each set of two consecutive parameters. At last, only 3 descriptor's *i.e.*, logP (WM), total lipole (WM)

and VAMP LUMO (WM) were found for the final model.

Training and test set preparation

The preparation of training[12] and test set[13] is important step in determining the quality of the model. The whole compounds were separate into training set and test set in such a way that both contains structurally diverse,active and inactive compounds. Training set was used for model development, some of the molecules behaved as outliers and were deleted for model development.

Model validation and development

Linear Regression Analysis

Multiple Linear Regression was performed in order to quantify relationship between descriptor's and its biological activity. The significance of model was determined by various statistical values[14,15] such as correlation coefficient(r^2) value, cross validation (r^2CV)should be more than 0.8, f-value, s-value (standard error of estimate) which should be small.

Partial least square method has been used to check[16] the robustness of the generated MLR model. On the basis of the r^2, r^2CV of the training and test sets, the PLS results were assessed.

Non-linear regression analysis[17]

Artificial neural networks act similarly to neural networks in imitating learning. The same MLR analysis descriptors were subjected to neural network analysis. Input, hidden and output layers were all present in the NNA. The output layer produced dependent variable, on the other hand input layer receives and processes data. The hidden layer linked layers indicated previously.

Model Validation

Cross-validation (r^2CV)

Cross validation analysis utilizing the leave-one-out (LOO) approach, in which compounds are eliminated one at a time and the activity of each deleted compound is predicted using a QSAR model, was carried out to explain the reliability of the suggested model. Cross-validation test (r^2CV) value for very robust model should be greater[18] than 0.6.

Prediction of test set

The QSAR models' predictive[19] ability was also assessed using a group of compounds that were divided up during the development of the model. The test sets' structure construction, charge derivation, optimization, and all other stages were carried out in the same manner as the training set compounds' activities, and the model's performance was evaluated based on the r^2 value. The test set's r^2 score of at least 0.6 indicates that the model has high predictive capacity.

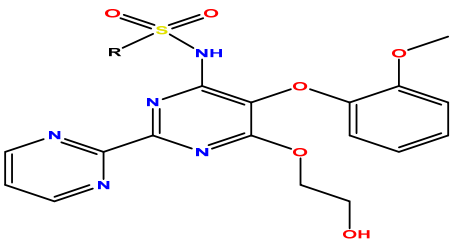
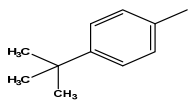
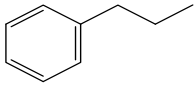
Outliers

Outliers are data points that the QSAR equation cannot adequately describe[20]. Higher residual value compounds shifted away from the regression line were eliminated from the model as outliers.

Results and Discussion

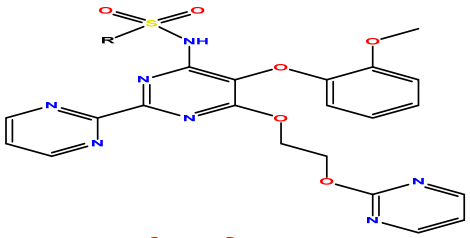
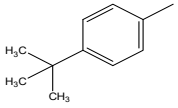
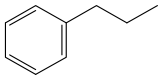
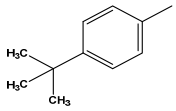
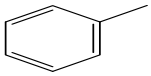
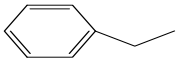
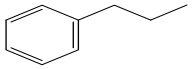
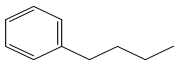
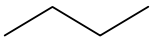
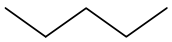
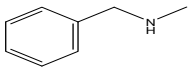
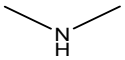
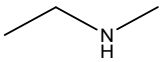
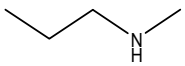
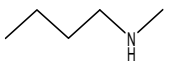
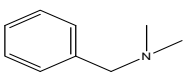
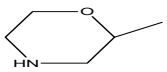
Data set of Alkyl sulfamide substituted pyrimidines derivatives attained from the literature was employed to generate the QSAR model's. The data set consists of 62 compounds (Table 1) which distributed into training set (47 compounds) for model development and a test set (15 compounds) for model validation.

Table 1 — Structure of Alkyl sulfamide substituted pyrimidines derivatives

Main Structure	Compd	R	IC_{50} [μM]
	**1**		45.0
	2		25.7
	3	CH3	4534

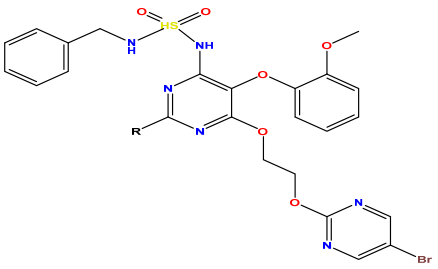
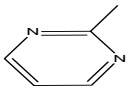
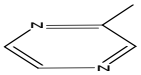
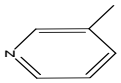
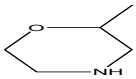
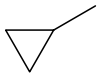
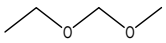
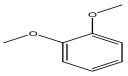
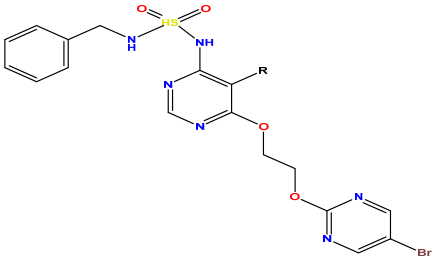
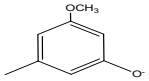
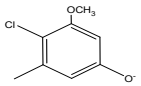
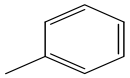
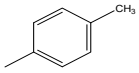
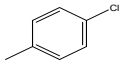
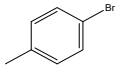
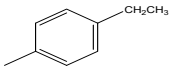
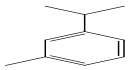
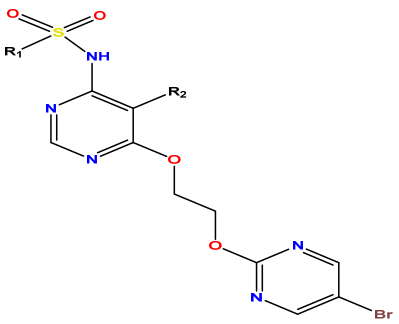
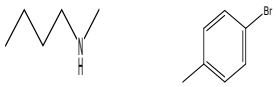
(*Contd.*)

Table 1 — Structure of Alkyl sulfamide substituted pyrimidines derivatives (*Contd.*)

Main Structure	Compd	R	IC_{50} [μM]
	**4**		4.3
	5		8.0
	6	CH_3	739
	7	H	8.0
	8	5-CH_3	17.8
	9	5-Cl	2.8
	10	5-Br	2.2
	11	5-OCH_3	3.2
	12	5-SCH_3	1.1
	13	5-CF_3	3.0
	14	5-cyclopropyl	7.7
	15	4,6-OCH_3	4775
	16		1.5
	17		10.4
	18		73
	19		2.2
	20		6.7
	21	CH_3	113
	22	CH_2CH_3	14
	23		9.4
	24		4.3
	25		1.9
	26		7.9
	27		9.9
	28		3.8
	29		1.7
	30		7.0
	31		187

(*Contd.*)

Table 1 — Structure of Alkyl sulfamide substituted pyrimidines derivatives (*Contd*.)

Main Structure	Compd	R	IC_{50} [μM]
	**32**		1.9
	33		2.1
	34		3.4
	35		1.4
	36		14
	37	OCH_3	40
	38		6.9
	39	H	6.9
	40		1.9
	**41**		2.1
	42		3.4
	43		1.4
	44		14
	45		40
	46		6.9
	47		6.9
	48		1.9
	**49**		1.6
	50		1.7
	51		0.8
	52		0.5
	53		0.3

(*Contd*.)

Table 1 — Structure of Alkyl sulfamide substituted pyrimidines derivatives (*Contd*.)

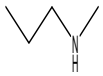
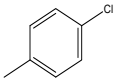
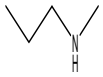
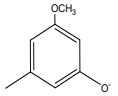
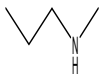
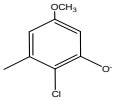
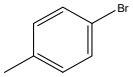
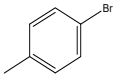
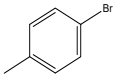
Main Structure	Compd	R	IC_{50} [μM]	Main Structure
	54			4.4
	55			4.3
	56			2.8
	57			1.2
	58			1.9
	59			0.4
	60			2.2
	61			1.1
	62	NH_2		3.4

Table 2 — Correlation matrix of the independent variable used in the construction of the final model

	Log value	Log P (WM)	Total lipole (WM)	VAMP LUMO(WM)
Log value	1	0.56715	0.289419	0.21396
Log P (WM)	0.56715	1	-0.13692	-0.25009
Total lipole (WM)	0.289419	-0.13692	1	-0.11855
VAMP LUMO (WM)	0.21396	-0.25009	-0.11855	1

Table 3 — Equation and statistical values used for the development of MLR model

MLR model	r	r^2	r^2CV	s-value	f-value	outlier
Equ.1	0.8790	0.7728	0.7153	0.4013	48.757	01

Linear regression models were developed using endothelin receptor inhibitory activity (dependent variable) and the three descriptor's (independent variables) namely log P(WM), total lipole(WM) and VAMP LUMO(WM) were retained after data reduction (Table 2).

Multiple Linear Regressions (MLR)

MLR is an earliest methodology and is employed to generate a linear[21] correlation among dependent variable and one or more independent variable[14,15]. The equation of the developed model by MLR is:

Y=0.50907332*X1+0.09950351*X2+0.75967538*X3-3.4923177 …equ.1

Where, Y=biological activity, X1=log P(WM), X2=total lipole(WM), X3=VAMP LUMO(WM). The final model exhibited r^2CV=0.7153 and r^2=0.7728 which shows excellent internal predictive power of the generated model (Table 3).

Partial Least Square(PLS)

PLS is an important model- development technique of specific interest in QSAR studies, noisy data with strong co-linearity and with many X variables can be examined. Thus, PLS is proficient to discover structure- activity problems, to examine data in more appropriate way, and to elucidate how molecular structure affects biological activity. In order to check

whether the variable selection by MLR results in loss of information or not, a partial least square(PLS) was carried out using all of the variables (equ.2). In this, dimension 7 was used for training and dimension 1 for test set.

Y=0.5629549*X1+0.13363029*X2+0.84975207*X3-4.0467372 …equ.2

Where, Y=biological activity, X1=log P (WM), X2=total lipole (WM), X3=VAMP LUMO (WM).

The r^2 value of training set was observed to be 0.763 and for test set it was 0.743.

Neural Network Analysis (NNA)

Many times neural network (NN) presents more accurate results than MLR, therefore NN analysis was also performed using TSAR 3.3, with similar descriptor's sets who were employed in case of regression analysis and the model generated was compared with the linear model. The major advantage of NNA is that it allows for the incorporation of non-linear[22,23] relationships between the independent and dependent variable's without the need of an explicit mathematical function. For the current study, the inputs for the NN were those descriptor's obtained after data reduction, while the output was the $\log 1/IC_{50}$ values. The number of hidden neurons, as well as the count of training and test set patterns, are automatically calculated by NNA functionality within the TSAR program. In our present study the NN run with 4 hidden nodes, 10% of training set and 25% of the test set generated best value of correlation coefficient (r^2). The r^2 of training set and test set was found to be 0.835 and 0.861 respectively. The actual and estimated value obtained from MLR, PLS and NNA analysis of the training set and test set [24] compounds are given in Table 4 and Table 5 and their related graphs are shown in Fig. 2, Fig. 3 and Fig. 4.

The test set compounds that were omitted in model development were used to check the external predictivity ability of NNA model. All of the compounds in the test set were handled similarly in the way they were handled in the training set. Generally, QSAR model is considered to have a high predictive power[25] only if the r^2 value is more than 0.6 for the test set. The r^2 value for test set compounds was 0.86, which indicates that the developed model is highly predictive and statistically significant.

Graphical representation of MLR, PLS and FFNN

Comparison between Linear and Non-Linear Methods

In comparison among the linear methods, both MLR & PLS methods were tremendously competitive QSAR analytical technique's applied to the selected data set because a comparable result was obtained from both the analysis. The study demonstrated that MLR and PLS are useful tools to understand the effects of various parameters in modeling. Generally neural networks provide more accurate estimates than linear regression in their ability to handle nonlinear correlations. But for the present study both the techniques can be used with equivalent efficiency to develop predictive models as evident by the r^2 values for training and test set. The r^2 value of training set are MLR=0.772, PLS=0.763 and NN= 0.835 and for test set values of r^2 for MLR=0.751, PLS=0.743 and NN= 0.861. It is abundantly obvious from the model's prediction ability that traditional MLR, PLS, and NN analysis produced a very reliable and predictive QSAR model, which is greatly capable of modeling the activity and structure data.

Description and importance of descriptors

The significance of Log P, Total Lipole, and VAMP LUMO is revealed by the findings of MLR, PLS, and NN. In fact, a substantial correlation has been found between ERAs and the three descriptors, with values for each descriptor displayed in Table 6.

Log P

Log P represents the measure of lipophilicity. The intermolecular interactions between a solvent and an organic molecule determine by log P. The bioavailability, permeability, and *in vivo* distribution of organic substances are significantly influenced by the physico-chemical property of lipophilicity. According to various literature surveys it is noted that the value of log P should be optimum for either drug penetration in body or drug formulation because a very high lipophilicity may cause adverse effect on protein binding and on drug absorption including solubility. For oral absorption or oral formulation, the value of log P should between the ranges of (0-3)[26]. According to the present model the equation obtained in linear regression, logP is positively correlated to permeability or compounds biological activity.

Total Lipole

Total lipole of a compound evaluate the lipophillic distribution. It is calculated from the sum of atomic

Table 4 — Actual and predictive activity data obtained from multivariate analysis of the training set compounds

Compd	Actual activity ($-\log IC_{50}$) [μM]	Predicted activity(μM)		
		MLR	PLS	NN
2	-1.40993	-1.138	-1.15568	-1.51037
3	-3.65648	-2.91956	-3.31094	-3.55523
4	-0.63347	-0.60791	-0.76042	-0.66211
5	-0.90309	-1.06864	-1.20427	-0.91252
7	-0.90309	-1.04072	-1.16659	-0.90149
8	-1.25042	-0.92441	-1.03443	-0.8407
9	-0.44716	-0.79156	-0.87999	-0.76058
11	-0.50515	-0.68787	-0.6889	-0.7619
13	-0.47712	-0.66455	-0.72118	-0.62832
14	-0.88649	-0.79601	-0.90583	-0.75814
15	-3.67897	-4.08786	-4.7129	-4.00357
16	-0.17609	-0.05127	-0.10306	-0.42711
17	-1.01703	-1.03832	-1.224	-0.86892
18	-1.86332	-1.01368	-1.18468	-0.82098
19	-0.34242	-0.75192	-0.86845	-0.7108
20	-0.82608	-0.45835	-0.52441	-0.58358
22	-1.14613	-0.79264	-0.72498	-0.97134
24	-0.63347	-0.76228	-0.77909	-0.78308
25	-0.27875	-0.80025	-0.87973	-0.82
26	-0.89763	-0.54336	-0.32902	-0.98671
27	-0.99564	-0.63038	-0.48694	-0.51308
29	-0.23045	-0.47212	-0.37982	-0.41573
30	-0.8451	-0.14661	-0.01681	-0.26786
33	-0.32222	-1.03661	-1.14074	-0.81847
34	-0.53148	-0.80246	-0.91161	-0.74035
35	-0.14613	-0.34227	-0.12874	-0.10147
36	-1.14613	-0.70593	-0.7727	-0.77921
37	-1.60206	-0.71892	-0.71655	-0.79949
39	-0.83885	-0.88542	-0.92073	-0.9315
41	0	0.099801	0.185572	0.138676
43	-0.83251	-1.06175	-1.20757	-0.906
44	0.154902	-0.24669	-0.1732	-0.52286
45	-0.14613	-0.78086	-0.88272	-0.73513
46	-0.20412	0.007172	0.142749	-0.30559
48	-0.5563	-0.20352	-0.20199	-0.509
49	-0.20412	0.007172	0.142749	-0.30559
50	-0.23045	-0.48156	-0.29524	-0.10128
51	0.09691	-0.40964	-0.24558	0.134131
52	0.30103	-0.19922	-0.0266	0.22004
53	0.522879	0.11284	0.351653	0.414856
55	-0.63347	-0.25978	-0.2247	-0.48805
56	-0.44716	-0.20588	0.118293	-0.59785
57	-0.07918	-0.33118	-0.16615	0.182638
58	-0.27875	-0.21951	-0.24624	-0.66243
59	0.39794	0.114501	0.259959	0.322204
61	-0.04139	0.036627	0.223125	0.074446
62	-0.53148	-0.60038	-0.42268	-0.48672

Table 5 — Actual and predictive activity data obtained from multivariate analysis for test set compounds

Compd	Actual activity ($-\log IC_{50}$) [μM]	Predicted activity(μM)		
		MLR	PLS	NN
6	-2.86864	-2.46344	-2.2005	-2.89778
10	-0.34242	0.023436	-0.04404	-0.4181
12	-0.04139	-0.33712	-0.3203	-0.60569
21	-2.05308	-1.75684	-1.72422	-2.03952
23	-0.97313	-1.01487	-1.10555	-0.75547
28	-0.57978	-1.3426	-1.34187	-0.70548
31	-2.27184	-2.11753	-2.3794	-2.2673
32	-0.27875	-0.47274	-0.3268	-0.68314
38	-0.83885	-0.79306	-0.83876	-0.11396
40	-1.65321	-1.00905	-0.88015	-1.05574
42	-0.14613	-0.84803	-0.12808	-0.1594
47	-0.36173	0.225759	-0.05288	-0.26762
54	-0.64345	-1.10088	-1.36273	-0.67067
60	-0.34242	-0.38789	-0.68955	-0.44658

Table 6 — Values of the descriptors for each compound

Compd	-Log value	Log P	Total lipole	VAMP LUMO
1	-1.6532	4.6319	9.64459	-0.660007
2	-1.40993	3.3513	12.5088	-0.785081
3	-3.65648	1.323	4.64461	-0.740982
4	-0.63347	5.7031	4.33799	-0.59306
5	-0.90309	4.4225	7.16895	-0.712197
6	-2.86864	2.3942	9.16203	-0.670869
7	-0.90309	4.4225	7.45753	-0.713238
8	-1.25042	4.607	7.6035	-0.702887
9	-0.44716	4.8463	7.93394	-0.731649
10	-0.34242	5.1469	8.51616	-0.735201
11	-0.50515	4.4035	11.1693	-0.722198
12	-0.04139	4.7468	7.52143	-0.717483
13	-0.47712	5.2395	8.79843	-0.941192
14	-0.88649	4.9744	7.03701	-0.705871
15	-3.67897	0	0	-0.783947
16	-0.17609	6.4949	6.12278	-0.624708
17	-1.01703	4.8677	4.89537	-0.672823
18	-1.86332	4.9626	5.44958	-0.776572
19	-0.34242	5.2143	6.572	-0.74769
20	-0.82608	5.6106	7.38865	-0.73378
21	-2.05308	3.186	12.5144	-0.769614
22	-1.14613	3.5285	14.5482	-0.716331
23	-0.97313	3.9971	14.0667	-0.704119
24	-0.63347	4.3934	10.8552	-0.772242
25	-0.27875	4.5838	8.30538	-0.615843
26	-0.89763	2.8072	19.5841	-0.564447
27	-0.99564	3.1497	16.9704	-0.566171
28	-0.57978	3.6183	14.0813	-0.562211
29	-0.23045	4.0146	14.0905	-0.560211
30	-0.8451	4.8303	14.2787	-0.702993
31	-2.27184	2.7987	22.6234	-0.859636
32	-0.27875	4.5838	5.14056	-0.606792
33	-0.32222	4.5824	8.45581	-0.945733
34	-0.53148	5.0303	7.1175	-0.762363

(*Contd.*)

Table 6 — Values of the descriptors for each compound (*Contd.*)

Compd	-Log value	Log P	Total lipole	VAMP LUMO
35	-0.14613	3.4464	18.6658	-0.607805
36	-1.14613	4.7108	8.41018	-0.59051
37	-1.60206	4.2457	11.4409	-0.692904
38	-0.83885	4.1573	19.4862	0
39	-0.83885	3.9953	10.565	-0.629555
40	-1.65321	3.9953	7.67171	-0.635686
41	0	3.4549	10.5987	1.02506
42	-0.14613	3.9729	6.30737	1.02952
43	-0.83251	4.5034	6.69939	-0.695838
44	0.154902	4.9706	12.3284	-0.673293
45	-0.14613	5.0214	7.32674	-0.755376
46	-0.20412	5.2952	13.8554	-0.756658
47	-0.36173	5.3669	14.8139	-0.707579
48	-0.5563	5.6974	9.08796	-0.679085
49	-0.20412	5.2952	13.8554	-0.756658
50	-0.23045	3.5186	18.2075	-0.779508
51	0.09691	3.8611	16.9283	-0.746808
52	0.30103	4.3297	16.3302	-0.705501
53	0.522879	4.726	17.7526	-0.746595
54	-0.64345	3.9148	18.8248	-0.809249
55	-0.63347	5.5417	10.8441	-0.878826
56	-0.44716	3.5101	22.9152	-1.02755
57	-0.07918	4.0559	16.6114	-0.732551
58	-0.27875	2.4894	6.97575	1.72628
59	0.39794	3.0074	13.064	1.02138
60	-0.34242	4.7085	16.44	-0.865118
61	-0.04139	5.1048	15.9243	-0.861279
62	-0.53148	3.2721	18.3815	-0.793529

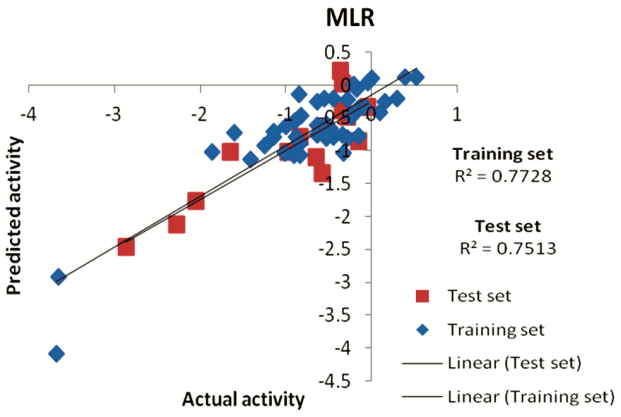


Fig. 2 — Graph between Actual *vs*. Predicted value of training and test set obtained from MLR analysis

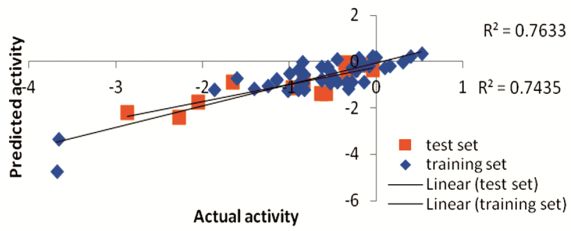


Fig. 3 — Graph between Actual *vs*. Predicted value of training and test set obtained from PLS analysis

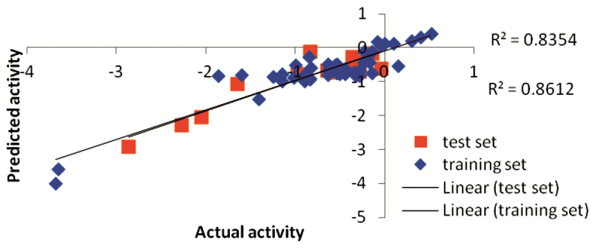


Fig. 4 — Graph between Actual *vs*. Predicted value of training and test set obtained from FFNN analysis

log P value. According to the present model the equation obtained in linear regression, total lipole is also positively correlated with compounds biological activity[27].

VAMP LUMO

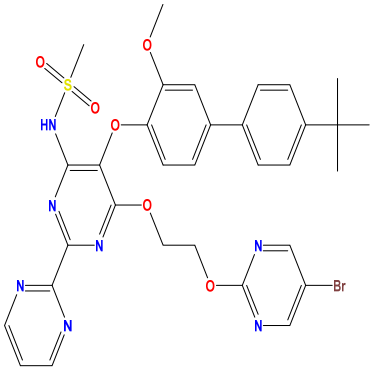
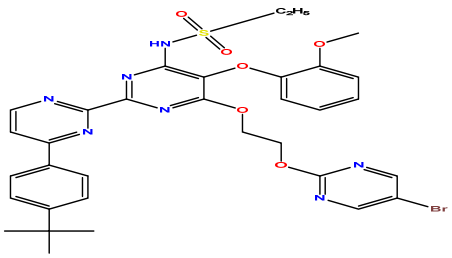
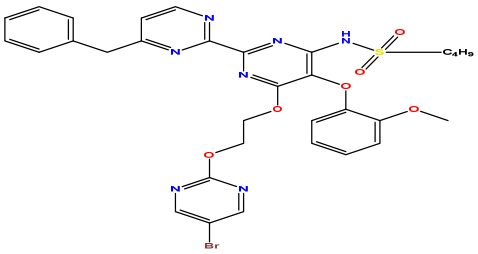
The LUMO description for each model adds the LUMO's energy (in electron volts). LUMO stands for the lowest energy level of an electron-free molecule. It's crucial for regulating molecular reactivity and characteristics. Incoming electron pairs are accepted

in a molecule's LUMO when it functions as a Lewis acid in bond formation. Because molecules with low-lying LUMOs are better at accepting electrons than those with high LUMOs, the LUMO descriptor should be used to determine the molecule's nucleophilicity[28]. According to the present model the equation obtained in linear regression, VAMP LUMO is positively correlated with compounds biological activity which clearly explains that the groups with optimum LUMO are required for good biological activity.

Designed New Compounds with high predicted Anti- hypertensive Activity

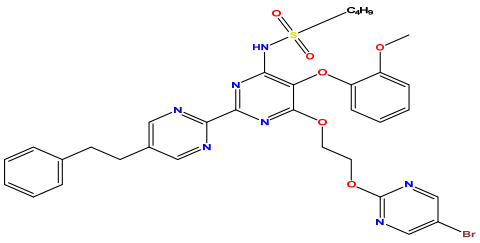
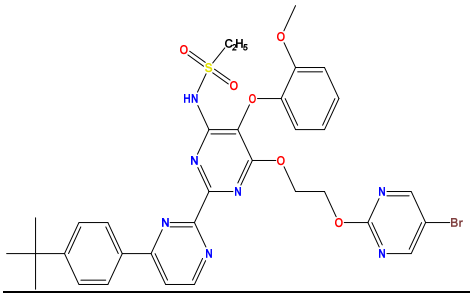
Based on findings, three factors that contribute to anti-hypertensive action in eq.1, it was discovered that these three factors are extremely essential and have a higher association with $-\log IC_{50}$. We created 5 novel compounds with stronger anti-hypertensive activity than the existing series by changing the substituents and improving the values of the three variables. These newly developed compounds' structure and values of the three parameters, as well as their projected $-\log IC_{50}$ values, were determined using the same procedure as before. According to the findings of the current QSAR investigation, the values of the three factors indicated that all 5 compounds were correctly created. As a result of these findings, we may conclude that our model based on QSAR investigations is substantial and predictive, and that including molecular design is also realistic[29]. The structure of the predicted new compounds as shown in Table 7 along with their three descriptors values and its IC_{50} (biological value).

Table 7 — Structure, descriptors values and biological values of new designed compounds

Structure	Log P	Total Lipole	VAMP LUMO	IC_{50} [μM]
A 	6.497	23.795	-0.7079	0.0226
B 	7.2392	18.086	-0.6936	0.0342
C 	6.7286	19.99	-0.6682	0.03849

(*Contd.*)

Table 7 — Structure, descriptors values and biological values of new designed compounds (*Contd.*)

Structure	Log P	Total Lipole	VAMP LUMO	IC_{50} [μM]
D 	6.8704	19.404	-0.7044	0.03973
E 	6.7743	19.497	-0.6927	0.04264

Study Limitations

Discovery of current QSAR analysis will be excellent just for modeling of potent Endothelin-A receptor inhibitor as active hypertensive agents.

Conclusion

A proficient QSAR analysis completed on 62 compounds of alkyl sulfamide substituted pyrimidines derivatives of endothelin receptor antagonist using both linear (MLR, PLS) and non-linear (FFNN) analysis. Both linear statistical analysis had comparable results which proved that model generated has good predictability and hence can be used to design endothelin receptor antagonist's. According to the developed model presented in the current work, the descriptors log P, total lipole and VAMP LUMO of endothelin receptor inhibitors are viewed as significant contributor to their biological properties'. Thus, an effort was made to develop an authentic model for ET_A receptor that possesses anti-hypertensive activity and the outcome of present model would help for future generation and the predicted new compounds with high potency, bioavailability and less toxicity.

Acknowledgement

The late Prof. Aditya Shastri and Mrs. Ina Shastri, Vice Chancellor of Banasthali University in Rajasthan, India, are sincerely thanked by the authors for providing the computational resources they needed to finish the project quickly.

Conflict of Interest

Not applicable.

References

1 https://www.indiaspend.com/high-blood-pressure-killed-1-6-mn-indians-in-2016-but-most-are-unaware-of-its-dangers-94201/.
2 Hoeper M M, Humbert M & Souza R, *Lancet Respir Med,* 4 (2016) 306.
3 Humbert M, Sitbon O & Simonneau G, *N Engl J Med*, 351 (2004) 1425.
4 Sokolovsky M & Shraga-Levine Z, *Handbook of Experimental Pharmacology*, (Springer, Berlin, Heidelberg) 2001, 152.
5 Arai H, Hori S, Aramori I, Ohkubo H & Nakanishi S, *Nature,* 348 (1990) 730.
6 Haynes W G & Webb D J, *Lancet*, 344 (1994) 852.
7 Russell F D, Coppell A L & Davenport A P, *Biochem Pharmacol*, 55 (1998) 697.
8 Russell F D & Davenport A P, *Br J Pharmacol*, 126 (1999) 391.
9 Maguire J J & Davenport A P, *Semin Nephrol*, 35 (2015) 125.
10 Morley S D, Abraham R J, Haworth I S, Jackson D E, Saunders M R & Vinter J G, *J Comp Aided Mol Des*, 5 (1991) 475.
11 Kesar S, Paliwal S K, Mishra P & Chauhan M, *Turk J Pharm Sci*, 16 (2019) 141.

12 Wu W, Walczak B, Massart D L, Heuerding S, Erni F, Last I R & Prebble K A, *Chemomet Intell Lab Sys*, 33 (1996) 35.
13 Yasri A & Hartsough D, *J Chem Inf Comput*, 41 (2001) 1218.
14 Mehta R S, Prajapati H R, Thakkar D V & Brahmkshatriya P S, *Int J Biol Sci*, 4 (2008) 266.
15 Bansal R, Karthikeyan C, Hari NS, Moorthy NS Hari & Trivedi P, *ARKIVOC,* xv (2007) 66. (https://www.arkat-usa.org/get-file/23189/)
16 Bottcher C J F, *Theory of Electric Polarization,* 2nd ed, (Elsevier Press, Amsterdam), 1973.
17 Paliwal S, Singh S & Pal M, *Drug Disc Ther*, 6 (2012) 69.
18 Myint K Z & Xie X Q, *Int J Mol Sci*, 11 (2010) 3846.
19 Wu B, Rega M F & Wei J, *Chem Biol Drug Des*, 73 (2009) 369.
20 Datar P A & Aher S B, *Int J Curr Pharm*, 10 (2012) 07.
21 Tetko I V, Luik A I & Poda G I, *J Med Chem*, 36 (1993) 811.
22 Pratim R P, Paul S, Mitra I & Roy K, *Molecules*, 14 (2009) 1660.
23 Patel D K & Patel N M, *J Sci Res*, 1 (2009) 594.
24 Prathipati P & Saxena A K, *J Comp Aided Mol Des*, 19 (2005) 93.
25 Yu Y, Su R, Wang L, Qi W & He Z, *Med Chem Res*, 19 (2010) 1233.
26 Ojha P, Mishra P, Kesar S & Singh S, *Int J Adv Res,* 4 (2016) 779.
27 Paliwal S, Das S, Yadav D, Saxena M & Paliwal S, *Med Chem Res*, 20 (2010) 1643.
28 Agrawal A & Mishra R, *Asian J. Exp. Sci*, 28 (2014) 11.
29 Paliwal S K, Verma A N & Paliwal S, *Monatsh Chem*, 142 (2011) 1069.