WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.084
Volume 10, Issue 7, 967-985. Research Article ISSN 2277– 7105

CP-MLR/PLS DERIVED QSAR RATIONALES FOR THE MMP-13 INHIBITORY ACTIVITY OF THE INDOLE DERIVATIVES

Jahan Afsar[1], Sharma Brij Kishore[1]* and Sharma Vishnu Dutt[2]

[1]Department of Chemistry, Government College, Bundi-323 001 (Rajasthan), India.

[2]Department of Shalya Tantra, Dr. SR Rajasthan Ayurved University Jodhpur-342 037 (Rajasthan), India.

Article Received on
01 May 2021,

Revised on 21 May 2021,
Accepted on 11 June 2021

DOI: 10.20959/wjpr20217-20799

***Corresponding Author**
Sharma Brij Kishore
Department of Chemistry, Government College, Bundi-323 001 (Rajasthan), India.

ABSTRACT

QSAR study has been carried out on the MMP-13 inhibitory activity of indole derivatives in 0D- to 2D-Dragon descriptors. The derived QSAR models have revealed that the information content index of 2nd order neighborhood symmetry (descriptor IC2) and average valence connectivity index chi-2 (descriptor X2A) in addition to molecular weight (descriptor MW), number of rotatable bonds (descriptor RBN), number of Oxygen atoms (descriptor nO) and aromatic ester functionality (descriptor nCOORPh) played a pivotal role in rationalization of MMP-13 inhibition activity of titled compounds. Atomic mass and polarizability weighted descriptors (MATS5m, BELm4 and GATS1p) and certain structural fragments such as H attached to C0(sp3) with 1X attached to next C (descriptor H-052) and CH3R/CH4 (descriptor C-001) are also predominant to explain MMP-13 inhibition actions of indole derivatives. PLS analysis has also corroborated the dominance of CP-MLR identified descriptors. Applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

KEYWORDS: QSAR; MMP-13 inhibitory activity; Combinatorial protocol in multiple linear regression (CP-MLR) analysis; PLS analysis; Dragon descriptors; Indole derivatives.

1. INTRODUCTION

Matrix metalloproteinases (MMPs) which are involved in the cleavage of collagen, gelatin, and other proteins in the extracellular matrix are zinc- and calcium- dependent peptidases.

The implications of MMPs in a wide range of disorders ranging from oncology to infectious diseases led to a number of pan-MMP inhibitors.[1-5] MMPs are grouped into subtypes based upon their substrates. The broad spectrum MMP inhibitors have primarily shown side effects such as muscular skeletal syndrome (MSS) and that is characterized by painful stiffening of the joints. It is thought that MSS occurs by inhibition of intrinsic cellular matrix degradation due to inhibition of MMPs other than MMP-13.[6,7] Being the most efficient enzyme MMP-13 in degrading collagen it has implications in the development of osteoarthritis and rheumatoid arthritis.[8] Thus, selective inhibition of MMP-13 has become an attractive target for small molecule intervention. Although there is difference in subunits of each class of MMPs but the catalytic domains share the canonical zinc-binding amino acid sequence HExxHxxGxxH.[9]

The potency gained by direct interaction with the zinc residue of classical broad spectrum inhibitors inevitably made designing selectivity into these inhibitors challenging.[10,11] The MMPs differ specifically in the S1′ pocket in a profound manner outside the active site[12] which provides an opportunity to gain selectivity. A large S1′ pocket possessed by MMP-13 upon appropriately occupied by a ligand opens up a side pocket (S1′*) that is not observed in many other MMPs.[13] The structural significance of this pocket resulted in non zinc interacting potent selective MMP-13 inhibitors. The common binding motifs shared by these inhibitors are an aromatic group π-stacks against histidine 222, direct or water-mediated hydrogen bonds are made with residues (threonine 245, threonine 247, and methionine 253) in the selectivity loop, and an aromatic group is located in the hydrophobic portion of the S1′* pocket formed by residues phenylalanine 217 and leucine 218. Few examples of selective inhibitors have been cited in literature.[14-19] As an attempt to discover high ligand efficiency leads, based on a virtual screening and fragment-based approach, with the aim of evolving selective inhibitors using structure-guided methods to harmonize chemistry strategy a new series of indole-based derivatives have been reported by Taylor et. al.[20] The aim of present communication is to establish the quantitative relationships between the reported activities and molecular descriptors unfolding the substitutional changes in titled compounds.

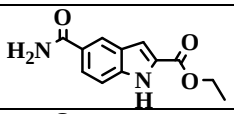
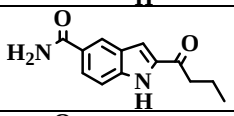
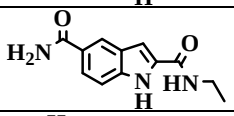
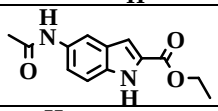
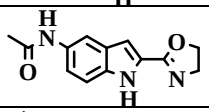
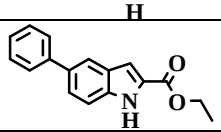
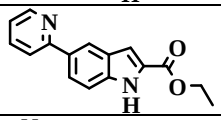
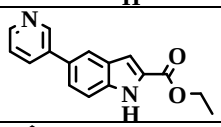
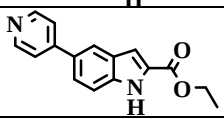
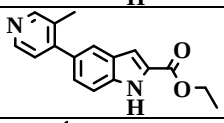
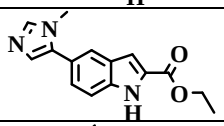
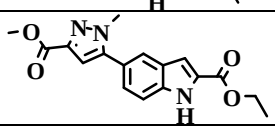
2. MATERIALS AND METHODS

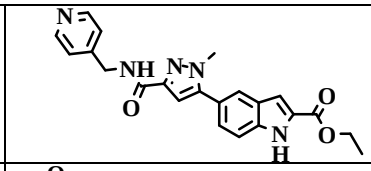
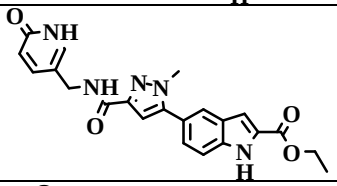
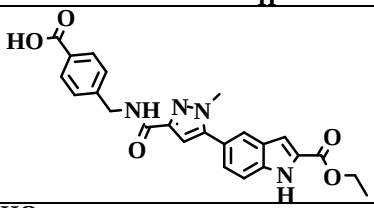
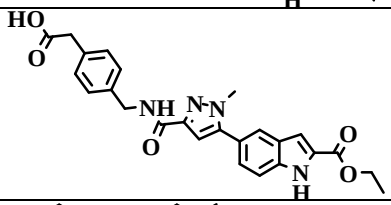
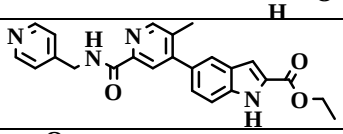
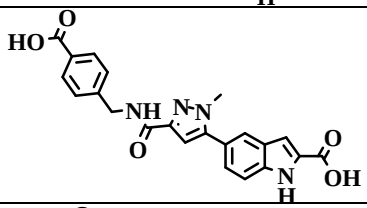
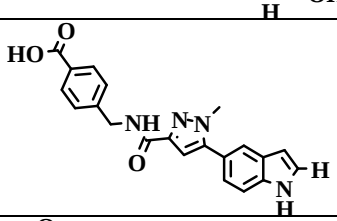
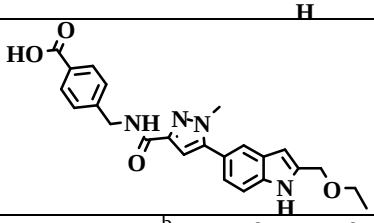
2.1 Biological actions and theoretical molecular descriptors

The reported twenty indole derivatives are considered as the data set for present study.[20] These derivatives were evaluated for their inhibitory activities and were reported as IC_{50}. The reported activity on molar basis (as pIC_{50}) along with the structures of these analogues is

shown in Table 1. The data set was sub-divided into training set to develop models and test set to validate the models externally. The test set compounds which were selected using an in-house written randomization program, are also mentioned in Table 1.

Table 1: Structures, observed and calculated MMP-13 inhibitory activities of indole derivatives.

Cpd.	Structure	pIC_{50}[a]					
		Obsd.[b]	Calculated				
			Eq. (3)	Eq. (4)	Eq. (5)	Eq. (6)	PLS
1		4.41	3.74	3.91	4.46	3.25	4.29
2		3.30	3.74	3.35	4.29	3.06	3.16
3		3.30	3.74	3.89	4.27	3.16	3.41
4		4.51	4.52	4.16	4.74	4.82	4.92
5		3.66	3.36	2.53	3.77	4.23	3.27
6		3.30	3.91	4.51	2.56	4.40	3.06
7[a]		3.30	3.91	4.53	4.03	4.72	4.53
8		4.82	3.91	4.53	4.04	4.34	4.57
9		4.09	3.91	4.53	3.70	4.04	4.23
10		5.02	4.77	4.78	4.70	4.71	5.11
11[c]		5.60	5.04	4.58	4.56	4.56	4.84
12		5.62	6.47	5.63	5.48	5.93	6.23

13[c]		7.05	6.95	7.01	6.80	7.01	6.66
14		6.92	6.86	7.30	6.94	7.12	7.03
15		9.00	7.52	7.79	7.35	7.54	7.77
16		8.52	8.67	8.05	8.72	8.14	8.60
17		6.82	6.86	7.21	7.29	7.32	7.19
18		4.60	5.91	5.71	5.58	5.82	5.06
19[c]		4.59	5.34	4.91	4.94	5.05	3.69
20[c]		6.60	8.19	7.54	7.90	7.10	7.02

[a]IC_{50} on molar basis; [b]taken from reference,[20] [c]Compound included in test set.

The structures of the all the compounds (listed in Table 1) were drawn in 2D ChemDraw[21] and subjected to energy minimization in the MOPAC using the AM1 procedure for closed shell system after converting these into 3D modules. The energy minimization was carried out to attain a well defined conformer relationship among the congeners under study. The 0D- to 2D-molecular descriptors of titled compounds was computed using DRAGON software.[19] This software offers a large number of descriptors corresponding to ten different

classes of 0D- to 2D-descriptor modules. The different descriptor classes include the constitutional, topological, molecular walk counts, BCUT descriptors, Galvez topological charge indices, 2D-autocorrelations, functional groups, atom-centered fragments, empirical descriptors and the properties describing descriptors. These descriptors offer characteristic structural information specific to the descriptor class. The definition and scope of these descriptor's classes is given in Table 2.

Table 2: Descriptor classes used for the modeling of MMP-13 inhibitory activity of indole derivatives.

S. no.	Descriptor Class (Acronyms)[a]	Definition and Scope
1	Constitutional (CONST)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations
2	Topological (TOPO)	2D-descriptor from molecular graphs and independent conformations
3	Molecular walk counts (MWC)	2D-descriptors representing self-returning walk counts of different lengths
4	Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights of the diagonal elements and atoms
5	Galvez topological charge indices (GALVEZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix
6	2D-autocorrelatons (2D-AUTO)	Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag)
7	Functional groups (FUN)	Molecular descriptors based on the counting of the chemical functional groups
8	Atom centered fragments (ACF)	Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen
9	Empirical (EMP)	1D-descriptors represent the counts of nonsingle bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule
10	Properties (PROP)	1D-descriptors representing molecular properties of a molecule

[a]Reference[22]

A total number of 473 descriptors, belonging to 0D- to 2D- modules, have been computed to obtain most appropriate models describing the biological activity. Prior to model development procedure, all those descriptors that are inter-correlated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. This procedure has reduced the total descriptors from 473 to 74 as relevant ones to explain the biological actions of titled compounds.

2.2 Development and validation of model

The combinatorial protocol in multiple linear regression (CP-MLR)[23-27] and partial least squares (PLS)[28-30] procedures were used in the present work for developing QSAR models. The CP-MLR is a "filter"-based variable selection procedure, which employs a combinatorial strategy with MLR to result in selected subset regressions for the extraction of diverse structure–activity models, each having unique combination of descriptors from the generated dataset of the compounds under study. The embedded filters make the variable selection process efficient and lead to unique solution. Fear of "chance correlations" exists where large descriptor pools are used in multilinear QSAR/QSPR studies.[31,32] In view of this, to find out any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been subjected to randomization test[33,34] by repeated randomization (100 simulation runs) of the biological responses. The datasets with randomized response vector have been reassessed by multiple regression analysis. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

Validation of the derived model is necessary to test its prediction and generalization within the study domain. For each model, derived by involving n data points, a number of statistical parameters such as r (the multiple correlation coefficient), s (the standard deviation), F (the F ratio between the variances of calculated and observed activities), and Q^2_{LOO} (the cross-validated index from leave-one-out procedure) have been obtained to access its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5. Additional statistical parameters such as, the Akaike's information criterion, AIC,[35,36] the Kubinyi function, FIT[37,38] and the Friedman's lack of fit, LOF,[39] have also been calculated to further validate the derived models. The AIC takes into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. The FIT, closely related to the F-value, proved to be a useful parameter for assessing the quality of the models. A model which is derived in k independent descriptors, its F-value will be more sensitive if k is small while it becomes less

sensitive if k is large. The FIT, on the other hand, will be less sensitive if k is small whereas it becomes more sensitive if k is large. The model that produces the lowest AIC value and highest FIT value is considered potentially the most useful and the best. The LOF factor takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large number of parameters.

2.3 Applicability domain

The usefulness of a model is based on its accurate prediction ability for new congeners. A model is valid only within its training domain and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain (AD) is evaluated by the leverage values for each compound.[40] A Williams plot (the plot of standardized residuals versus leverage values (h)) is constructed, which can be used for a simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in the model. In this plot, the AD is established inside a squared area within $\pm x$ standard deviations and a leverage threshold h^*, which is generally fixed at $3(k + 1)/n$ (n is the number of training set compounds and k is the number of model parameters), whereas $x = 2$ or 3. If the compounds have a high leverage value ($h > h^*$), then the prediction is not trustworthy. On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training set compounds.

3. RESULTS AND DISCUSSION

3.1 QSAR results

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modelling study, 5 compounds have been included in the test set for the validation of the models derived from 15 training set compounds. A total number of 74 significant descriptors from 0D- to 2D- classes have been subjected to CP-MLR analysis with default "filters" set in it. Statistical models in one and two descriptors have been derived to achieve the best relationship correlating MMP-13 inhibitory activity. A total number of two models in one descriptor, having $r^2_{Test} > 0.5$, were obtained through CP-MLR. The models in one descriptor are given below.

pIC_{50} = 3.537 + 5.267(0.795) RBN

n = 15, r = 0.878, s = 0.912, F = 43.803, Q^2_{LOO} = 0.697, Q^2_{L5O} = 0.745

r^2_{Test} = 0.545, FIT = 2.737, LOF = 0.960, AIC = 1.088 (1)

pIC_{50} = 2.026 + 5.637(1.109) IC2

n = 15, r = 0.815, s = 1.103, F = 25.810, Q^2_{LOO} = 0.506, Q^2_{L5O} = 0.596

r^2_{Test} = 0.598, FIT = 1.613, LOF = 1.406, AIC = 1.593 (2)

Where n, r, s and F represent respectively the number of data points, the multiple correlation coefficient, the standard deviation and the F-ratio between the variances of calculated and observed activities. In above and all follow-up regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models. The positive regression coefficient associated to a descriptor will augment the activity profile of a compound while the negative coefficient will cause detrimental effect to it. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation.

The descriptor RBN participated in above models is from the CONST class and the other one IC2 is the topological class descriptor. Both the descriptors have shown positive influence on the activity as evident from the signs of regression coefficients. Thus a higher value of descriptor RBN (number of rotatable bonds) and IC2 (information content index of 2^{nd} order neighborhood symmetry) would be beneficiary to the activity.

The one descriptor models could estimate nearly 77% in observed activity of the compounds. Considering the number of observation in the dataset, models with up to two descriptors were explored. It has resulted in 16 two-parameter models with test set $r^2 > 0.50$. These models (with 74 descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this, the optimum *r*-bar value of the preceding level model (=0.868) has been used as the new threshold of filter-3 for the next generation. These models have shared 17 descriptors among them. All these 17 descriptors along with their brief meaning, average regression coefficients, and total incidence are listed in Table 3, which will serve as a measure of their estimate across these models.

Table 3: Identified descriptors[a] along with their class, physical meaning, average regression coefficient and incidence[b].

Descriptor class, average regression coefficient and (incidence)	
Constitutional descriptors (**CONST**):	MW (molecular weight), 4.423(6); RBN (number of rotatable bonds), 4.497(4); nO (number of Oxygen atoms), 2.038(1)
Topological descriptors (**TOPO**):	SPI (superpendentic index), 2.374(1); X2A (average valence connectivity, index chi-2), -2.046(1); Lop (Lopping centric index) 2.166(1); IC2 (information content index, neighborhood symmetry of 2-order), 4.556(7); SIC2 (structural information content index, neighborhood symmetry of 2-order), -2.764(1); CIC3 CIC3 (complementary information content index, neighborhood symmetry of 3-order), 3.287(1)
Modified Burden Eigen values (**BCUT**):	BELm4 (lowest eigenvalue n.4 of Burden matrix/weighted by atomic masses), 2.116(2)
2D autocorrelations (**2D-AUTO**):	MATS5m (Moran autocorrelation of lag-5/ weighted by atomic masses), -2.622(1); MATS1e (Moran autocorrelation of lag-1/ weighted by atomic Sanderson electronegativities), 1.863(1); MATS6e (Moran autocorrelation of lag-6/ weighted by atomic Sanderson electronegativities), -2.723(1); GATS1p (Geary autocorrelation of lag-1/ weighted by atomic polarizabilities), 1.946(1)
Functional group counts (**FUNC**):	nCOORPh (number of aromatic esters), 1.697(1)
Atom-centered fragments (**ACF**):	C-001 (CH3R/CH4), 1.859(1); H-052 (H attached to C0(sp3) with 1X attached to next C), 1.574(1)

[a]The descriptors are identified from the two parameter models for activity emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.868 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 15 compounds. [b]The average regression coefficient of the descriptor corresponding to all models and the total number of its incidence. The arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

Following are the selected two-descriptor models for the MMP-13 inhibitory activities of indole derivatives emerged through CP-MLR.

$pIC_{50} = 4.285 + 5.312(0.660)RBN - 2.046(0.779)X2A$

$n = 15, r = 0.924, s = 0.756, F = 35.266, Q^2_{LOO} = 0.791, Q^2_{L5O} = 0.744$

$r^2_{Test} = 0.604$, FIT = 3.712, LOF = 0.852, AIC = 0.859 (3)

$pIC_{50} = 2.298 + 4.173(0.531)$ MW $+1.574(0.577)$ H-052

$n = 15, r = 0.923, s = 0.763, F = 34.549, Q^2_{LOO} = 0.602, Q^2_{L5O} = 0.644$

$r^2_{Test} = 0.630$, FIT = 3.636, LOF = 0.867, AIC = 0.874 (4)

pIC_{50} = 2.558 +3.619(0.973) RBN +2.664(1.122) IC2

n = 15, r = 0.918, s = 0.783, F = 32.539, Q^2_{LOO} = 0.760, Q^2_{L5O} = 0.723

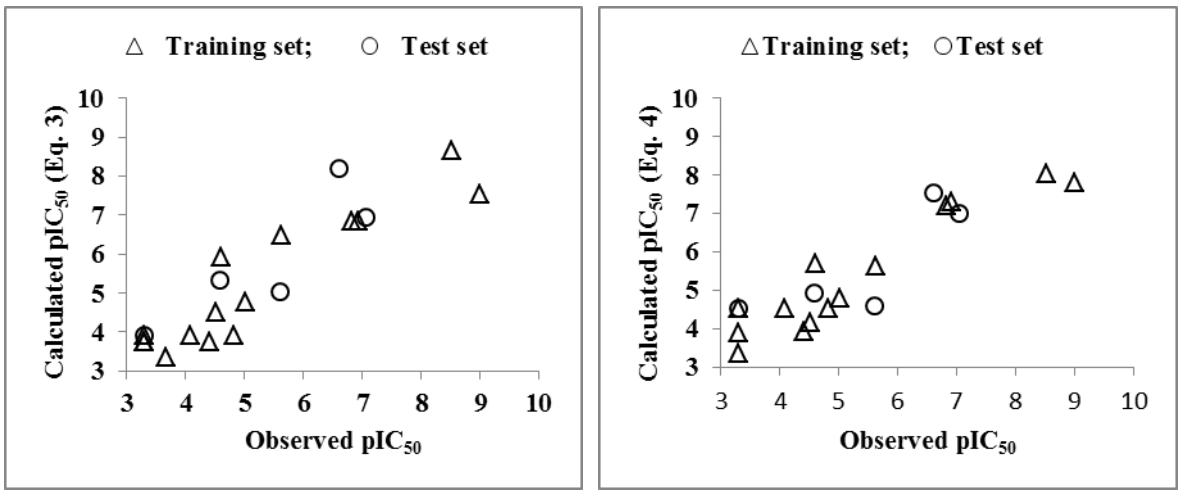
r^2_{Test} = 0.635, FIT = 3.425, LOF = 0.912, AIC = 0.920 (5)

pIC_{50} = 4.484 +4.850(0.616) MW -2.622(1.066) MATS5m

n = 15, r = 0.916, s = 0.792, F = 31.658, Q^2_{LOO} = 0.653, Q^2_{L5O} = 0.582

r^2_{Test} = 0.627, FIT = 3.332, LOF = 0.934, AIC = 0.941 (6)

These models have accounted for nearly 85% variance in the observed activities. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The values greater than 0.5 of Q^2 index is in accordance to a reasonable robust QSAR model. The pIC_{50} values of training set compounds calculated using Eqs. (3) to (6) have been included in Table 1. The models (3) to (6) are validated with an external test set of 5 compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the same is reported in Table 1. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.



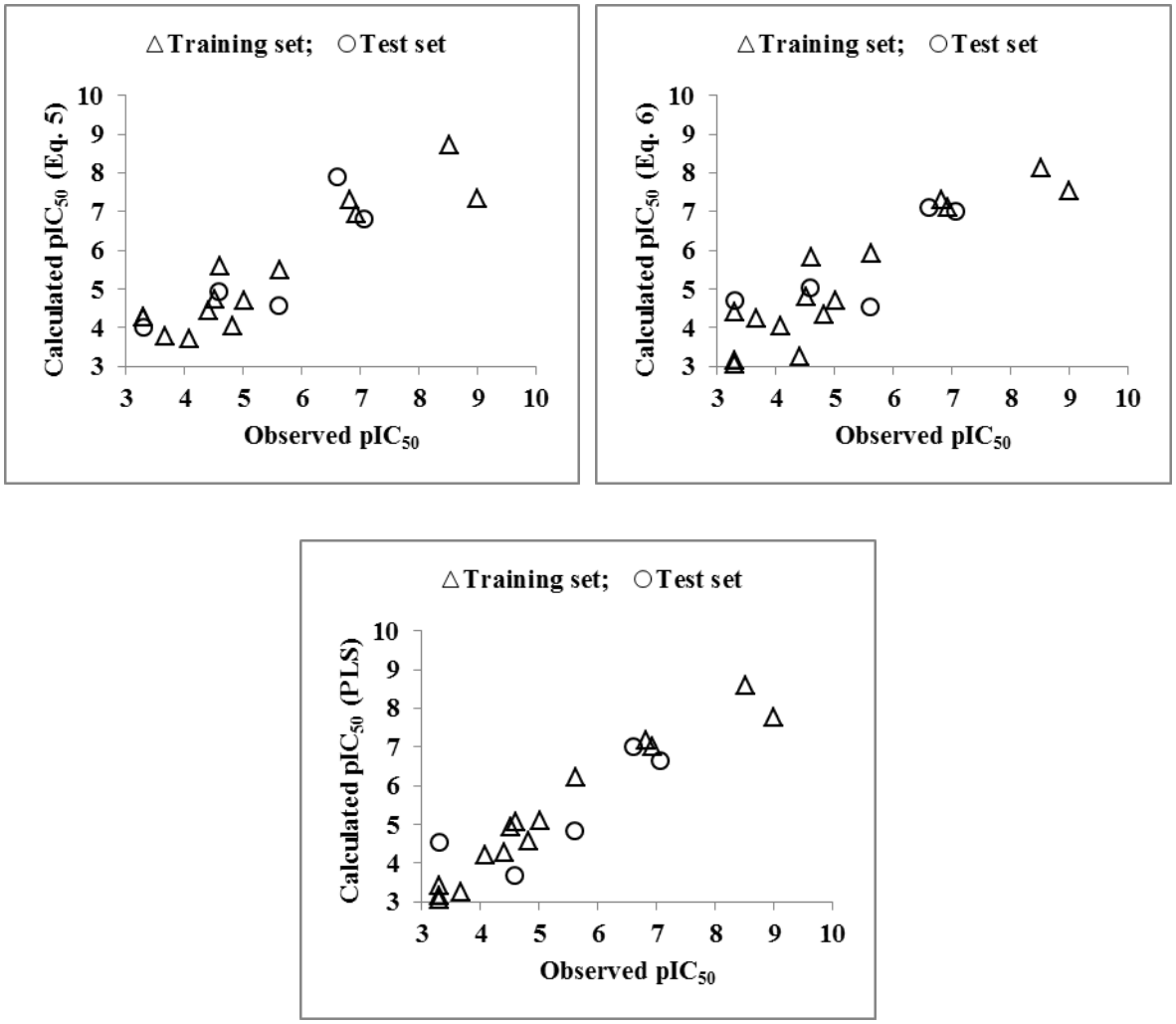


Figure 1: Plot of observed and calculated pIC_{50} values of training- and test-set compounds for MMP-13 inhibition.

The newly appeared descriptors in above models are X2A (topological class), MW (constitutional class), H-052 (atom-centered fragments class) and MATS5m (2D autocorrelations). Descriptors MW and H-052 have shown positive and descriptors MATS5m and X2A negative correlation to the activity. The signs of regression coefficients advocated that higher values of molecular weight (descriptor MW) and presence of H attached to C0(sp3) with 1X attached to next C type structural fragments in a molecular structure would be incremental to the activity. On the other hand a higher value of average valence connectivity index chi-2 (descriptor X2A) and Moran autocorrelation of lag-5/ weighted by atomic masses (descriptor MATS5m) would be deleterious to the activity.

A partial least square (PLS) analysis has been carried out on these 17 CP-MLR identified descriptors (Table 3) to facilitate the development of a "single window" structure–activity

model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, two components are found to be the optimum for these 17 descriptors and they explained 94.09% variance in the activity (r = 0.970, Q^2_{LOO} = 0.904, s = 0.479, F = 96.842, r^2_{Test} = 0.661). The MLR-like PLS coefficients of these 17 descriptors are given in Table 5.

Table 5: PLS and MLR-like PLS models from the 17 descriptors of two parameter CP-MLR models for MMP-13 inhibitory activities.

A: PLS equation									
PLS components					**PLS coefficient (s.e.)[a]**				
Component-1					-0.675(0.051)				
Component-2					0.376(0.083)				
Constant					5.192				
B: MLR-like PLS equation									
S. no	**Descriptor**	**MLR-like coefficient[b]**	**(f.c.)[c]**	**Order**	**S. no.**	**Descriptor**	**MLR-like coefficient[b]**	**(f.c.)[c]**	**Order**
1	MW	0.775	0.105	3	10	BELm4	0.666	0.077	7
2	RBN	1.191	0.128	2	11	MATS5m	-0.023	-0.002	16
3	nO	0.619	0.089	5	12	MATS1e	0.247	0.022	13
4	SPI	0.837	0.102	4	13	MATS6e	-0.284	-0.026	12
5	X2A	-0.004	0.000	17	14	GATS1p	0.556	0.049	10
6	Lop	-0.040	-0.004	15	15	nCOORPh	0.536	0.053	9
7	IC2	1.596	0.149	1	16	C-001	0.597	0.054	8
8	SIC2	0.412	0.038	11	17	H-052	0.660	0.082	6
9	CIC3	-0.245	-0.019	14	Constant = 1.408				
C: PLS regression statistics					Values				
n					15				
r					0.970				
s					0.479				
F					96.842				
FIT					10.193				
LOF					0.342				
AIC					0.344				
Q^2_{LOO}					0.904				
Q^2_{L5O}					0.896				
r^2_{Test}					0.661				

[a]Regression coefficient of PLS factor and its standard error. [b]Coefficients of MLR-like PLS equation in terms of descriptors for their original values; [c]f.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit s.d.) data.

For the sake of comparison, the plot showing goodness of fit between observed and calculated activities (through PLS analysis) for the training and test set compounds is also given in Figure 1. Figure 2 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity.

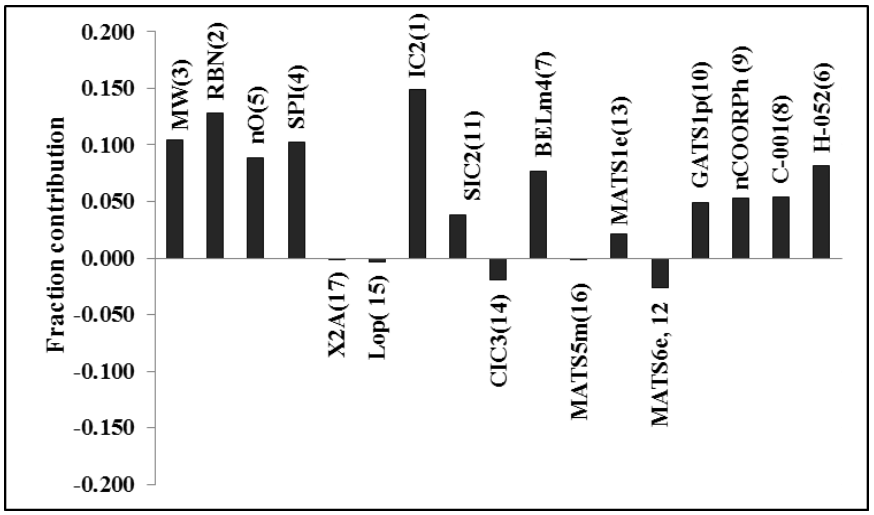


Figure 2: Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 17 CP-MLR identified descriptors (Table 3) associated with MMP-13 inhibitory activity of indole derivatives.

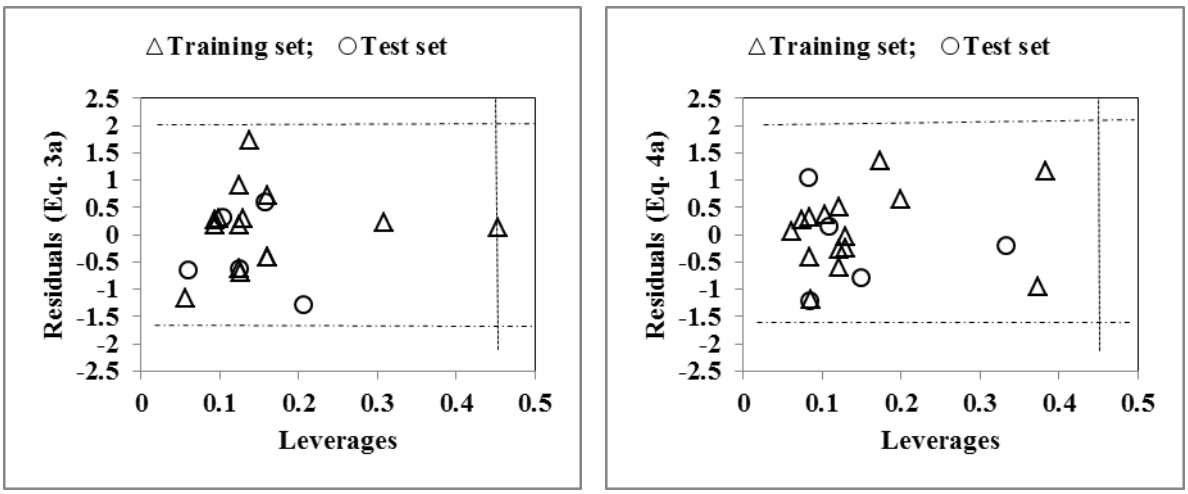
The PLS analysis has suggested IC2 as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 7 in Table 5; Figure 2). The other nine significant descriptors in decreasing order of significance are RBN, MW, SPI, nO, H-052, BELm4, C-001, nCOORPh and GATS1p. Descriptors IC2, RBN, MW and H-052 are part of Eqs. (1) to (6) and convey same inference in the PLS model as well. It is inferred from the PLS analysis that a higher values of topological descriptor SPI (superpendentic index), constitutional descriptor nO (number of Oxygen atoms), modified Burden eigenvalue class descriptor BELm4 (lowest eigenvalue n.4 of Burden matrix/weighted by atomic masses) and 2D autocorrelation descriptors GATS1p (Geary autocorrelation of lag-1/ weighted by atomic polarizabilities) in addition to presence of structural fragment CH3R/CH4 (descriptor C-001) and aromatic ester functionality (descriptor nCOORPh) would be advantageous to the activity. It is also observed that PLS model from the dataset devoid of CP-MLR identified 17 descriptors (Table 3) is inferior in explaining the activity of the analogues.

3.2 Applicability domain (AD)

On analyzing the model AD in the Williams plot, shown in Figure 3, of the model based on the whole dataset (Table 6), it has appeared that none of the compound was identified as an obvious outlier for the MMP-13 inhibitory activity if the limit of normal values for the *Y* outliers (response outliers) was set as 2.5 (standard deviation) units. An outlier to a QSAR is identified normally by having a large standard residual activity and can indicate the limits of applicability of QSAR models. None of the compounds found to have leverage (h) values greater than the threshold leverage (h*=0.45). For both the training-set and test-set, the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data. Furthermore, all of the compounds were within the applicability domain of the proposed model and were evaluated correctly.

Table 6: Models derived for the whole data set (n = 20) in descriptors identified through CP-MLR.

Model	r	s	F	Q^2_{LOO}	Eq.
pIC_{50} = 4.292 +4.931(0.576)RBN – 2.036(0.726)X2A	0.906	0.771	39.130	0.759	(3a)
pIC_{50} = 2.275 + 4.008(0.485)MW +1.598(0.488)H-052	0.904	0.780	38.109	0.711	(4a)
pIC_{50} = 2.552 + 3.227(0.851)MW +2.800(1.077)H-052	0.901	0.789	37.053	0.740	(5a)
pIC_{50} = 4.482 +4.800(0.577)MW - 2.718(0.900)MATS5m	0.905	0.803	35.424	0.723	(6a)



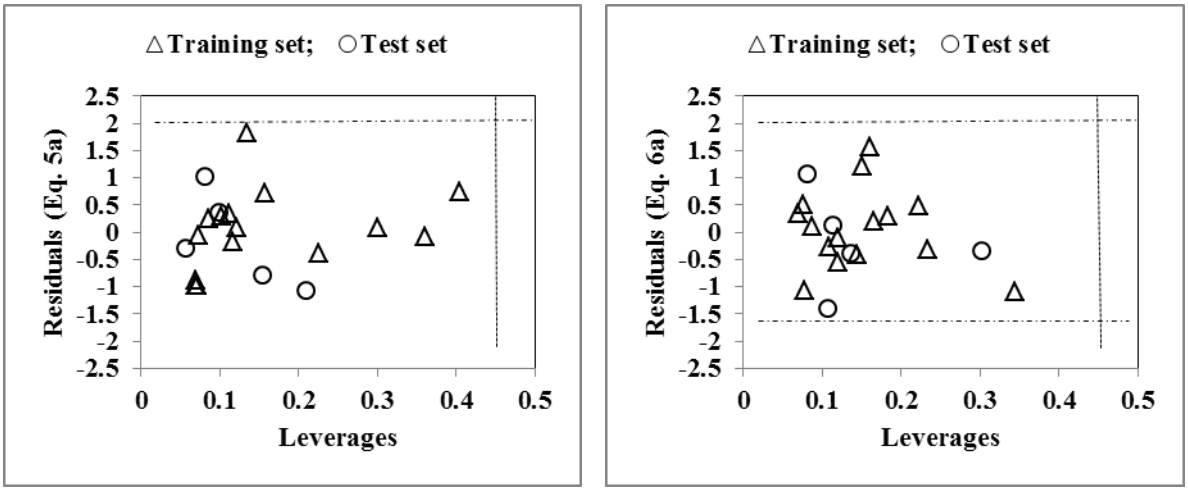


Figure 3: Williams plot for the training-set and test- set compounds for MMP-13 inhibitory activity. The horizontal dotted line refers to the residual limit (±2.5×standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.45).

CONCLUSIONS

QSAR study has been carried out on the MMP-13 inhibitory activity of indole derivatives in 0D- to 2D-Dragon descriptors. The derived QSAR models have revealed that the information content index of 2nd order neighborhood symmetry (descriptor IC2) and average valence connectivity index chi-2 (descriptor X2A) in addition to molecular weight (descriptor MW), number of rotatable bonds (descriptor RBN), number of Oxygen atoms (descriptor nO) and aromatic ester functionality (descriptor nCOORPh) played a pivotal role in rationalization of MMP-13 inhibition activity of titled compounds. Atomic mass and polarizability weighted descriptors (MATS5m, BELm4 and GATS1p) and certain structural fragments such as H attached to C0(sp3) with 1X attached to next C (descriptor H-052) and CH3R/CH4 (descriptor C-001) are also predominant to explain MMP-13 inhibition actions of indole derivatives. PLS analysis has also corroborated the dominance of CP-MLR identified descriptors. Applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Compliance with ethical standards

Acknowledgements

Authors are thankful to their institution for providing necessary facilities to complete this study.

Disclosure of conflict of interest

The authors declare no conflict of interest.

REFERENCES

1. Morrison CJ, Butler GS, Rodriǵuez D and Overall CM. Matrix metalloproteinase proteomics: substrates, targets, and therapy. Current Opinion in Cell Biology, 2009; 5: 645-653.
2. Cudic M and Fields GB. Extracellular proteases as targets for drug development. Current Protein and Peptide Science, 2009; 4: 297-307.
3. Tu G, Xu W, Huang H and Li S. Progress in the development of matrix metalloproteinase inhibitors. Current Medicinal Chemistry, 2008; 15: 1388-1395.
4. Fingleton B. Matrix metalloproteinases as valid clinical targets. Current Pharmaceutical Design, 2007; 13: 333-346.
5. Li X and Li J. Recent advances in the development of MMPIs and APNIs based on the pyrrolidine platforms. Mini Reviews in Medicinal Chemistry, 2010; 9: 794-805.
6. Reiter LA, Freeman-Cook KD, Jones CS, Martinelli GJ, Antipas AS, Berliner MA, Datta K, Downs J, Eskra JD, Forman MD, Greer EM, Guzman R, Hardink JR, Janat F, Keene NF, Laird E R, Liras J, Lopresti-Morrow LL, Mitchell PG, Pandit J, Robertson D, Sperger D, Vaughn-Bowser ML, Waller DM and Yocum SA. Potent, selective pyrimidinetrione-based inhibitors of MMP-13. Bioorganic Medicinal Chemistry Letters, 2006; 5822-5826.
7. Rowan AD, Litherland GJ, Hui W and Milner JM. Metalloproteases as potential therapeutic targets in arthritis treatment. Expert Opinion on Therapeutic Targets, 2008; 12: 1-18.
8. Vincenti MP and Brinckerhoff CE. Transcriptional regulation of collagenase MMP-1, MMP-13) genes in arthritis: Integration of complex signaling pathways for therecruitment of gene-specific transcription factors. Arthritis Research, 2002; 4: 157-164.
9. Tallant C, Marrero A and Gomis-Rüth FX. Matrix metalloproteinases: Fold and function of their catalytic domains. Biochimica et Biophysica Acta, 2010; 1: 20-28.
10. Jacobsen JA, Major Jourden JL, Miller MT and Cohen SM. To bind zinc or not to bind zinc: an examination of innovative approaches to improved metalloproteinase inhibition. Biochimica et Biophysica Acta, 2010; 1: 72-94.

11. Nuti E, Tuccinardi T and Rossello A. Matrix metalloproteinase inhibitors: new challenges in the era of post broadspectrum inhibitors. Current Pharmaceutical Design, 2007; 20: 2087-2100.
12. Engel CK, Pirard, B, Schimanski S, Kirsch R, Habermann J, Klingler O, Schlotte V, Weithmann KU and Wendt KU. Structural basis for the highly selective inhibition of MMP-13. Chemcal Biology, 2005; 2: 181-189.
13. Devel L, Czarny B, Beau F, Georgiadis D, Stura E and Dive V. Third generation of matrix metalloprotease inhibitors: Gain in selectivity by targeting the depth of the S1′ cavity. Biochimie, 2010; 11: 1501-1508.
14. Wu J, Rush TS 3rd, Hotchandani R, Du X, Geck M, Collins E, Xu ZB, Skotnicki J, Levin JI and Lovering FE. Identification of potent and selective MMP-13 inhibitors. Bioorganic and Medicinal Chemistry Letters, 2005; 18: 4105-4109.
15. Monovich LG, Tommasi RA, Fujimoto RA, Blancuzzi V, Clark K, Cornell WD, Doti R, Doughty J, Fang J, Farley D, Fitt J, Ganu V, Goldberg R, Goldstein R, Lavoie S, Kulathila R, Macchia W, Parker DT, Melton R, O'Byrne E, Pastor G, Pellas T, QuadrosE, Reel N, Roland DM, Sakane Y, Singh H, Skiles J, Somers J, Toscano K, Wigg A, Zhou S, Zhu L, Shieh WC, Xue S and McQuire LW. Discovery of potent, selective, and orally active carboxylic acid based inhibitors of matrix metalloproteinase- Journal of Medicinal Chemistry, 2009; 11: 3523-3538.
16. Levin JI, Chen JM, Du MT, Nelson FC, Wehr T, DiJoseph JF, Killar LM, Skala S, Sung A, Sharr MA, Roth CE, Jin G, Cowling R, Di L, Sherman M, Xu ZB, March CJ, Mohler KM, Black RA and Skotnicki JS. The discovery of anthranilic acid-based MMPinhibitors. Bioorganic and Medicinal Chemistry Letters, 2001; 22: 2975-2978.
17. Skiles JW, Gonnella NC and Jeng AY. The design, structure, and clinical update of small molecular weight matrix metalloproteinase inhibitors. Current Medicinal Chemistry, 2004; 11: 2911-2977.
18. Rao BG. Recent developments in the design of specific Matrix Metalloproteinase inhibitors aided by structural and computational studies. Current Pharmaceutical Design, 2005; 11: 295-322.
19. Wasserman ZR. Making a new turn in matrix metalloprotease inhibition. Chemical Biology, 2005; 2: 143-144.
20. Taylor SJ, Abeywardane A, Liang S, Muegge I, Padyana AK, Xiong Z, Hill-Drzewi M, Farmer B, Li X, Collins B, Li JX, Heim-Riether A, Proudfoot J, Zhang Q, Goldberg D, Zuvela-Jelaska L, Zaher H, Li J and Farrow NA. Fragment-based discovery of indole

inhibitors of matrix metalloproteinase-13. Journal of Medicinal Chemistry, 2011; 54: 8174-8187.

21. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. Dragon software (version 1.11-2001) by Todeschini R and Consonni V. Milano, Italy. Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: analysis of Selwood et al. Data Set-a case study. QSAR and Combinatorial Science, 2003; 22: 583-595.
22. Sharma S, Prabhakar YS, Singh P and Sharma BK. QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach. European Journal of Medicinal Chemistry, 2008; 43: 2354-2360.
23. Sharma S, Sharma BK, Sharma SK, Singh P and Prabhakar YS. Topological descriptors in modeling the agonistic activity of human A3 adenosine receptor ligands: The derivatives of 2-Chloro-N^6-substituted-4'-thioadenosine-5'-uronamide. European Journal of Medicinal Chemistry, 2009; 44: 1377-1382.
24. Sharma BK, Pilania P, Singh P and Prabhakar YS. Combinatorial protocol in multiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. SAR QSAR in Environmental Research, 2010; 21: 169-185.
25. Sharma BK, Singh P, Sarbhai K and Prabhakar YS. A quantitative structure- activity relationship study on serotonin 5-HT6 receptor ligands: Indolyl and piperidinyl sulphonamides. SAR QSAR in Environmental Research, 2010; 21: 369-388.
26. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics, 1978; 20: 397-405.
27. Kettaneh N, Berglund A and Wold S. PCA and PLS with very large data sets. Computational Statistics and Data Analysis, 2005; 48: 69-85.
28. Stahle L and Wold S. Multivariate data analysis and experimental design. In:Ellis GP, West WB. Eds., Biomedical research. Progress in medicinal chemistry. Elsevier Science Publishers, BV, Amsterdam, 1998; 25:291-338.
29. Topliss JG and Edwards RP. Chance factors in studies of quantitative structure- activity relationships. Journal of Medicinal Chemistry, 1979; 22: 1238-1244.
30. Katritzky AR, Dobchev DA, Slavov S and Karelson M. Legitimate utilization of large descriptor pools for QSAR/QSPR models. Journal of Chemical Information and Modelling, 2008; 48: 2207-2213.

31. So S-S and Karplus M. Three-dimensional quantitative structure–activityrelationship from molecular similarity matrices and genetic neural networks. 1. Method and validation. Journal of Medicinal Chemistry, 1997; 40: 4347-4359.
32. Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR/PLS directed structure–activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR and Combinatorial Science, 2004; 23: 234-244.
33. Akaike H. Information theory and an extension of the minimum likelihoodprinciple. In: Petrov BN, Csaki F, editors. Second international symposium on information theory. Budapest: Akademiai Kiado, 1973; 1973: 267-281.
34. Akaike H. A new look at the statistical identification model. IEEE Transactions on Automatic Control AC, 1974; 19: 716-723.
35. Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. Quantitative Structure-Activity Relationships, 1994; 13: 285-294.
36. Kubinyi H. Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. Quantitative Structure-Activity Relationships, 1994; 13: 393-401.
37. Friedman J. In Technical Report No. Laboratory for Computational Statistics, Stanford University: Stanford, 1990; 102.
38. Gramatica P. Principles of QSAR models validation: internal and external. QSAR and Combinatorial Science, 2007; 26: 694-701.