QSAR Rationale of Matrix Metalloproteinase Inhibition Activity in a Class of Carboxylic Acid Based Compounds

B. K. Sharma[1], P. Singh[2*] and Y. S. Prabhakar[3]

[1]*Department of Chemistry, Government College, Bundi-323 001, India.*
[2]*Department of Chemistry, S.K. Government College, Sikar-332 001, India.*
[3]*Medicinal and Process Chemistry Division, CSIR, Central Drug Research Institute, CSIR, Lucknow-226 001, India.*

Authors' contributions

This work was carried out in collaboration between all authors. Author BKS designed the study, performed the statistical analysis, wrote the protocol. Author PS wrote the first draft of the manuscript. Authors PS and YSP managed the analyses of the study. Author YSP managed the literature searches. All authors read and approved the final manuscript.

Research Article

ABSTRACT

The matrix metalloproteinase-13 (MMP-13) inhibitory activities of carboxylic acid based compounds, in presence and absence of bovine serum albumin (BSA), have been analyzed quantitatively in terms of chemometric descriptors. The statistically validated quantitative structure-activity relationship (QSAR) models obtained through combinatorial protocol in multiple linear regression (CP-MLR) analysis and the participated descriptors in these models provided rationales to explain the inhibitory activities of these congeners. For MMP-13 inhibition activity, the identified descriptors (BEHm1, BELm1 and BEHm8) have highlighted the role of the atomic mass in terms of the highest and lowest eigenvalues derived from Burden matrix. The positive correlation with activity suggested that their higher values are desirable in improving the activity of a compound. Additionally, the descriptor C-027 representing R-CH-X type fragment in a molecular structure advocates the absence of such type of fragment for the improved activity. On the other hand presence of RCO-N< or >N-X=X type fragment (descriptor N-072) would be beneficiary to the MMP-13 inhibitory activity.
The structural features, rationalized by the descriptors MSD (Balaban's mean square

---

Corresponding author: Email: psingh_sikar@rediffmail.com;

distance index), nCrHR (number of ring tertiary C (sp^3), H-047 (H attached to C1(sp^3)/C0(sp^2)) and H-050 (H attached to heteroatom) have imparted positive impact on the MMP-13 w/BSA inhibition activity. The atomic properties such as atomic polarizability and atomic Sanderson's electronegativity have shown their positive impact on the activity via descriptors BELp4 and GATS3e in respective eigenvalues or lag. The other descriptors, MATS1m and MATS3e, have revealed the negative influence of atomic mass and electronegativity on the of MMP-13 w/BSA inhibition activity. The results obtained from CP-MLR analysis have been supported further through partial least-squares (PLS) study.

1. INTRODUCTION

Matrix metalloproteinases (MMPs) belong to a family of zinc-endopeptidases that use the electrophilic zinc ion to degrade the components of the extracellular matrix [1]. A number of physiological processes such as ovulation, embryogenesis, angiogenesis, cellular differentiation, and wound healing [2,3] have been connected to these enzymes. Under normal physiological conditions, endogenous tissue inhibitors of MMPs (TIMPs) control their activity [4]. However, over expression of MMP activity, or poor control by TIMPs, have been related with a variety of pathological conditions such as psoriasis [5], multiple sclerosis [6,7], osteoarthritis [8], rheumatoid arthritis [9,10] osteoporosis [11,12] Alzheimer's disease [13], tumor growth and metastasis [4,14,15], and undesirable degradation of extracellular proteins.

The development of low molecular weight synthetic inhibitors of MMPs is one approach to treat these diseases [16-18]. These inhibitors generally include a zinc binding group (ZBG), capable to chelate the catalytic zinc ion, bound to a substrate-like fragment designed to fit the S1′ primary subsite and adjacent subsites [19]. Hydroxamate is considered the most effective ZBG because it forms five-membered chelates and two additional H-bonds with the enzyme [20-22]. Hydroxamates, however are affected by lack of selectivity [23], not only toward the other members of the same family but even to other physiologically important metalloenzymes. Moreover, they show poor pharmacokinetic properties [24] and may cause toxicity resulting from metabolic activation of hydroxylamine [25].

The initial therapeutic approach was based on the use of broad-based hydroxamic acid inhibitors. However, the approach was failed at the clinical level because of musculoskeletal side effects [26-30], the exact nature of which is not completely clear.

Two hypotheses have been proposed to explain the musculoskeletal side effects. The first hypothesis is based on the inhibition of tumor necrosis factor (TNF)-α converting enzyme (TACE) [31-35], and/or other sheddases [36], by hydroxamates. The other hypothesis ascribed the musculoskeletal side-effects to nonselective inhibition within the MMP family. Specifically MMP-1, which is linked to normal tissue turnover and repair, was implicated as an antitarget.
Considering the high affinity of the hydroxamate moiety for zinc relative to other coordinating moieties [37], it was further hypothesized that selective inhibition of MMP-13 versus other

metalloproteinases [38] (specifically TACE and MMP-1) could be achieved with alternate [39,40] or no [41] zinc-binding moiety. Thus, the primary goal remains to replace the hydroxamic acid zinc-binding moiety so as to reduce the potential for pan-MMP and off-target sheddase inhibition. Based on optimization of a lead compound, Monovich et al. have recently reported [42] potent, orally active, nonhydroxamic acid MMP-13 inhibitors lacking sheddase and MMP-1 activities.

The present communication is aimed to perform a 2D-quantitative structure-activity relationship (2D-QSAR) for the reported compounds so as to provide the rationale for drug-design and to explore the possible mechanism of action. In the congeneric series, where a relative study is being carried out, the 2D-descriptors may play important role in deriving the significant relationships with biological activities of the compounds. The novelty and importance of a 2D-QSAR study is due to its simplicity for the calculations of different descriptors and their interpretation (in physical sense) to explain the biological activities of compounds at molecular level.

2. MATERIALS AND METHODS

The carboxylic acid based compounds along with their inhibition activity, IC_{50} values were taken from the literature [42]. The generalized structure of these compounds is shown in Fig. 1. The structural variations of these compounds are presented in Table 1.
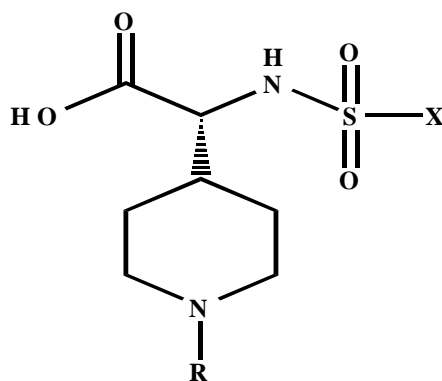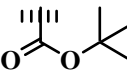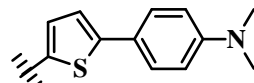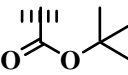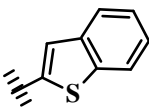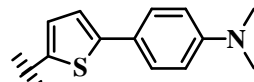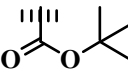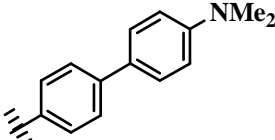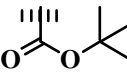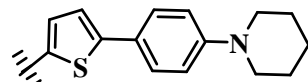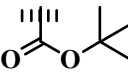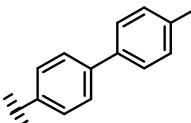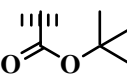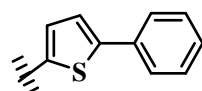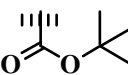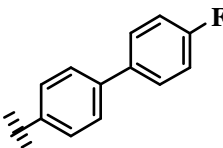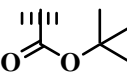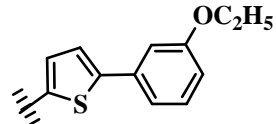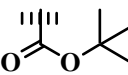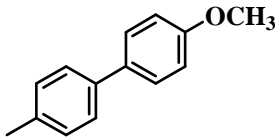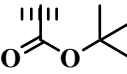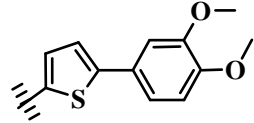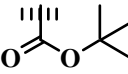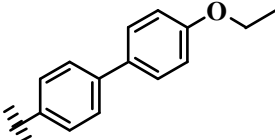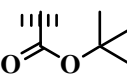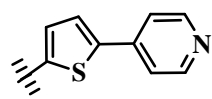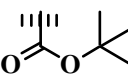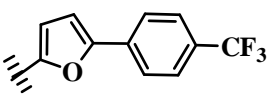

Fig. 1. General structure of carboxylic acid based compounds

Table 1. Structures[a] of the carboxylic acid based analogues

Cpd	R	X	Cpd	R	X
1			11		
2	H		12		
3			13		
4			14		
5			15		
6			16		
7			17		

#		#	
8		18	
9		19	
10		20	
21		34	
22		35	
23		36	
24		37	
25		38	
26		39	

27		—OCH(CH$_3$)$_2$	**40**		
28		O(CH$_2$)$_2$OMe	**41**		
29			**42**		
30		CF$_3$	**43**		
31			**44**		
32			**45**		
33			**46**		
47			**55**		
48			**56**		
49			**57**		

50			**58**		
51			**59**		
52			**60**		
53			**61**		
54					
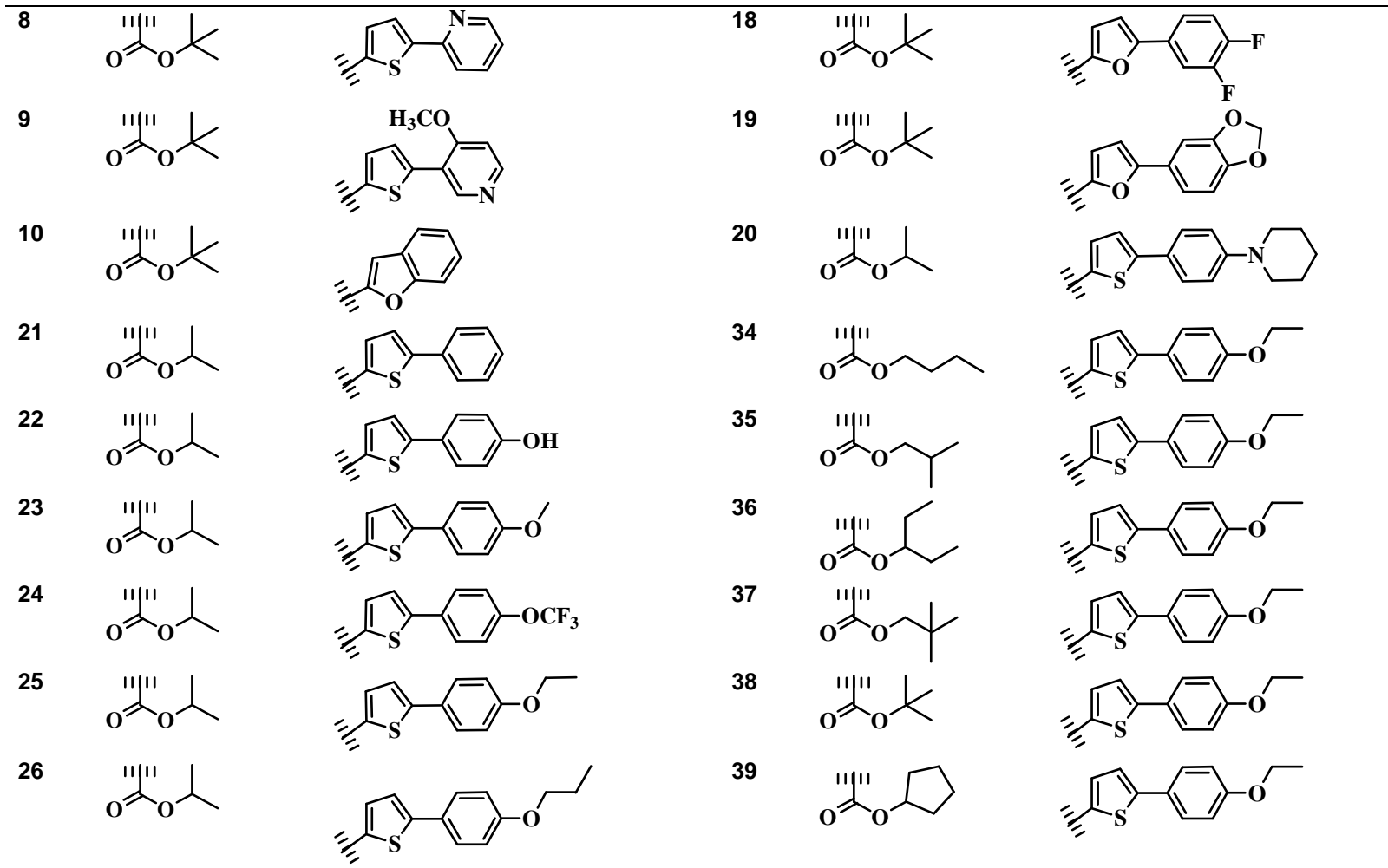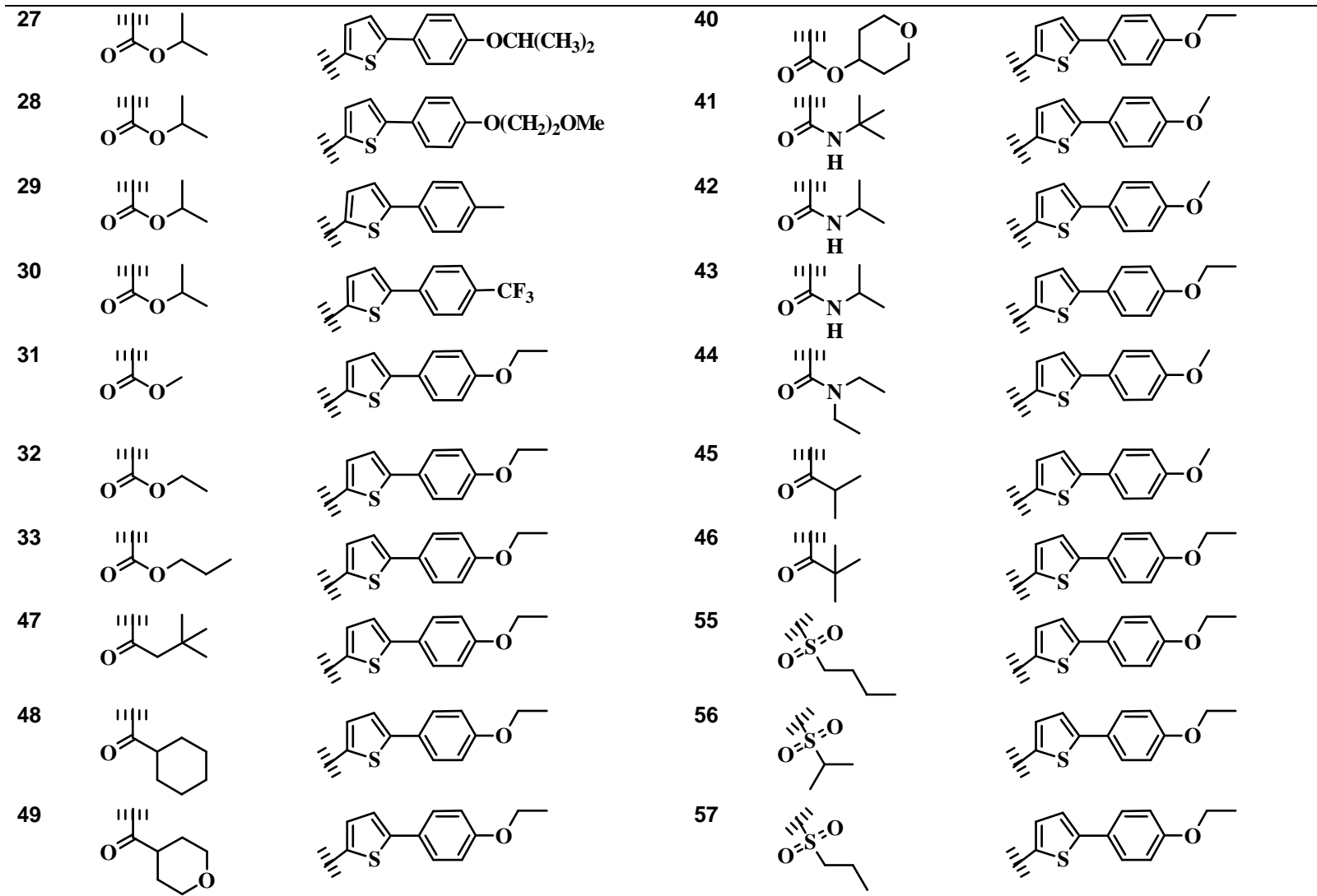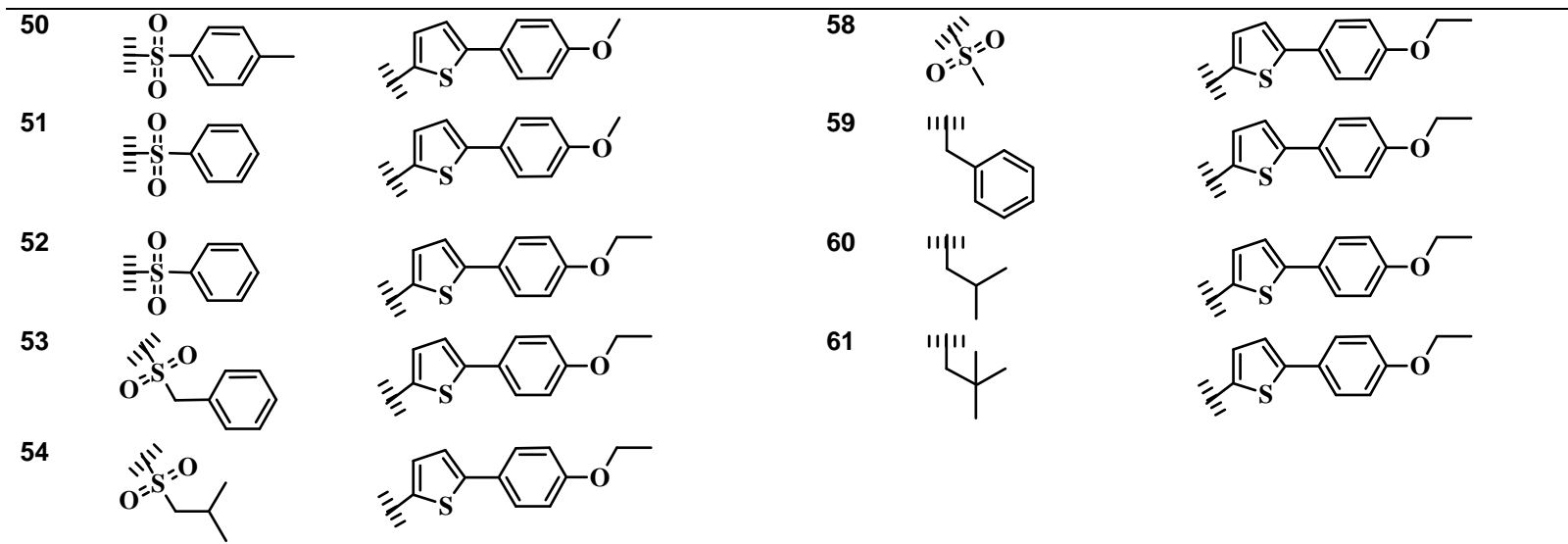
^a*Reference [42].*

The inhibitory activity, IC_{50}, represents the concentration of a compound to bring out 50% inhibition of the MMP-13 enzyme. The activity data have been expressed on the negative logarithm as pIC_{50} ($-logIC_{50}$) on the molar basis and regarded as the dependent variable for quantitative analysis. The empirical observations have indicated that activity in the presence of serum proteins was important for activity in the rat intra-articular (IA) model, the *in vitro* MMP-13 assay was, therefore, carried out both in the presence and absence of 1% bovine serum albumin (BSA). Thus inhibition activity data, determined under two conditions, have been considered in present study. A total number of 61 compounds including the lead compound (cpd 1; Table 2) have been considered as the data-set for present investigation.

Table 2. Experimental and modeled matrix metalloproteinase-13 inhibition activity of carboxylic acid based analogues

Cpd.	$-logIC_{50}$ (M)			$-logIC_{50}$ (M)		
	MMP-13			MMP-13 w/BSA		
	Obsd.[a]	Calc.		Obsd.[a]	Calc.	
		Eq. (12)	PLS		Eq. (14)	PLS
1	9.52	9.05	9.19	7.92	7.28	7.36
2	7.96	7.80	8.06	7.33	7.09	7.20
3[b]	9.12	9.40	9.23	7.20	7.15	7.21
4[b]	8.17	8.34	7.97	5.14	6.37	6.35
5	5.98[c]	8.80	8.61	<5.00[d]	6.79	6.78
6	6.24[c]	8.56	8.51	<5.00[d]	6.63	6.62
7	7.05	6.92	6.99	5.70	6.00	5.94
8[b]	7.74	7.36	7.10	5.73	6.17	6.08
9	7.30	7.43	7.28	5.60	5.98	6.01
10	5.88	5.86	5.89	<5.00[d]	5.85	5.72
11[b]	6.84	8.16	8.18	<5.00[d]	5.90	5.91
12	9.52	9.29	9.86	8.05	8.12	8.24
13	9.15	9.18	9.19	7.14	7.09	7.16
14	8.62	8.77	8.91	7.32	7.01	6.99
15[b]	9.30	9.17	9.30	7.96	7.70	7.75
16	9.30	9.21	9.45	8.01	7.95	7.95
17	7.21	6.91	6.60	<5.00[d]	5.57	5.41
18	5.05	5.83	5.82	<5.00[d]	5.64	5.30
19	6.71	6.36	6.48	<5.00[d]	5.74	5.61
20	8.89	9.34	9.25	7.00	6.96	7.02
21[b]	7.92	8.34	8.06	<5.00[d]	6.24	6.17
22	8.48	8.26	8.15	6.52	6.98	6.90
23[b]	9.22	8.72	8.59	7.17	6.79	6.76
24	8.66	8.68	8.61	5.74	5.61	5.50
25	9.30	8.74	8.71	7.42	7.04	6.97
26	9.10	8.75	8.82	7.24	7.17	7.09
27	8.40	8.77	8.68	6.41	7.19	7.11
28	9.05	8.74	8.93	7.60	7.99	7.84
29	8.89	8.96	8.58	5.94	6.11	6.11
30	8.41	8.86	8.61	5.35	5.48	5.36
31	8.46	8.17	8.29	6.62	6.48	6.49
32	8.70	8.58	8.59	6.85	6.74	6.72
33	9.00	8.83	8.87	7.09	6.95	6.88
34	9.10	8.98	9.08	7.05	7.12	7.04

35	9.22	8.94	9.05	6.98	6.98	6.93
36[b]	9.40	8.87	8.94	7.22	6.96	6.90
37	9.05	8.98	9.11	6.76	6.90	6.92
38	9.22	8.80	8.69	7.84	7.13	7.07
39	9.30	8.96	9.12	7.22	7.29	7.26
40	9.40	8.99	9.25	8.15	8.34	8.24
41	8.85	9.26	8.92	6.96	7.25	7.39
42	8.80	9.26	9.01	7.71	7.39	7.42
43	9.10	9.28	9.13	7.87	7.68	7.68
44[b]	8.55	9.11	8.74	6.67	6.66	6.78
45[b]	7.80	8.30	8.18	6.61	6.11	6.16
46[b]	8.10	8.39	8.28	6.60	6.28	6.34
47	8.44	8.83	8.72	6.79	6.54	6.66
48	9.05	9.30	9.10	7.10	6.86	6.90
49	8.77	8.85	8.89	7.70	7.68	7.69
50[b]	8.85	8.67	8.70	7.19	6.30	6.32
51	8.92	8.67	8.56	6.24	6.42	6.40
52	9.15	9.12	9.00	6.74	6.69	6.61
53	9.15	8.91	9.03	7.60	7.49	7.42
54	8.26	8.84	8.58	6.64	6.71	6.85
55[b]	8.40	8.84	8.78	6.97	6.83	6.93
56	8.33	8.54	8.42	6.55	6.44	6.56
57	8.16	8.70	8.53	6.84	6.74	6.84
58	7.90	7.77	7.86	6.91	6.97	7.10
59	8.12	8.51	8.70	7.19	7.74	7.61
60	7.55	7.67	7.89	6.73	6.83	6.97
61[b]	7.31	7.82	7.90	6.49	6.74	6.95

[a]IC_{50} *represents the concentration of compound to bring out 50% inhibition of MMP-13 enzyme; taken from ref. [42],* [b]*compound in test-set,* [c] *'outlier' compound,* [d]*compound ignored due to uncertain activity.*

For modeling purpose, the data-set was divided into training- and test-sets to insure external validation of models derived from the appropriate descriptors. Additionally, leave-one-out (LOO) and leave-five-out (L5O) procedures were employed for internal validation of derived models. The selection of compounds for test-set has been made through SYSTAT [43] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the activity, pIC_{50} values. Nearly 25% of the compounds, from total population, were selected from the generated cluster tree in such a way to keep them at a maximum possible distance from each other. In SYSTAT, by default, the normalized Euclidean distances are computed to join the objects of cluster. The normalized distances are root mean-squared distances. The single linkage uses distance between two closest members in clustering. It generates long clusters and provides scope to choose objects at different intervals. Due to this reason, a single linkage clustering procedure was applied.

2.1 Theoretical molecular descriptors

The structures of the compounds under study have been drawn in 2D ChemDraw [44] using the standard procedure. All these structures have been ported to DRAGON software [45] for computing the descriptors corresponding to 0D-, 1D-, and 2D-classes. Table 3 provides the definition and scope of these descriptor-classes in addressing the structural features which were employed in present QSAR work.

Table 3. Descriptor classes used for the analysis of inhibition activities of the compounds

Descriptor class (acronyms)	Definition and scope
Constitutional (CONST)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations.
Topological (TOPO)	2D-descriptor from molecular graphs and independent conformations.
Molecular walk counts (MWC)	2D-descriptors representing self-returning walk counts of different lengths.
Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms.
Galvez topological charge indices (GLVZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix.
2D-autocorrelations (2DAUTO)	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).
Functional groups (FUNC)	Molecular descriptors based on the counting of the chemical functional groups.
Atom-centred fragments (ACF)	Molecular descriptors based on the counting of 120 atom-centred fragments, as defined by Ghose-Crippen.
Empirical (EMP)	1D-descriptors represent the counts of non-single bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule.
Properties (PROP)	1D-descriptors representing molecular properties of a molecule.

The combinatorial protocol in multiple linear regression (CP-MLR) computational procedure [46] has been used for present work in developing QSAR models. Prior to application of the CP-MLR procedure, all those descriptors which 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. The remaining descriptors, able to address the biological activity of these compounds, serve as the database (pool) at the end of this initial stage. The descriptors have been scaled [47] using the formula:

$$X_{ij}^n = \frac{X_{ij} - X_{ij,min}}{X_{ij,max} - X_{ij,min}} \tag{1}$$

where X_{ij}^n and X_{ij} are the scaled and non-scaled values of jth descriptor for compound i, respectively, and $X_{ij,max}$ and $X_{ij,min}$, in that order, are the maximum and minimum values for j^{th} descriptor. In this way, the scaled values of each descriptor would be between 0 and 1.

2.2 Model Development

The CP-MLR is a 'filter'-based variable selection procedure for model development in QSAR studies [46]. Its procedural aspects and implementation are discussed in some of our recent publications [48-53]. The thrust of this procedure is in its embedded 'filters'. They are briefly as follows: filter-1 seeds the variables by way of limiting inter-parameter correlations to predefined level (upper limit ≤ 0.79); filter-2 controls the variables entry to a regression equation through t-values of coefficients (threshold value ≥ 2.0); filter-3 provides comparability of equations with different number of variables in terms of square root of adjusted multiple correlation coefficient of regression equation, r-bar; filter-4 estimates the consistency of the equation in terms of cross-validated Q^2 with leave-one-out (LOO) cross-validation as default option (threshold value $0.3 \leq Q^2 \leq 1.0$). All these filters make the variable selection process efficient and lead to a unique solution. In order to collect the descriptors with higher information content and explanatory power, the threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation.

In order to discover any chance correlations associated with the models obtained through CP-MLR, each cross-validated model has been put to a randomization test [54,55] by repeated randomization of the activity to ascertain the chance correlations, if any, associated with them. For this, every model has been subjected to 100 simulation runs with scrambled activity. The scrambled activity models with regression statistics better than or equal to that of the original activity model have been counted, to express the percent chance correlation of the model under scrutiny.

Validation of the derived model is necessary to test the prediction and generalization of the method. In the present study, the data set has been divided into training-set for model development and test-set for external prediction. Goodness of fit of the models was assessed by examining the multiple correlation coefficient (r), the standard deviation (s), the F-ratio between the variances of calculated and observed activities (F). A number of additional statistical parameters such as the Akaike's information criterion, AIC [56,57], the Kubinyi function, FIT [58,59], and the Friedman's lack of fit, LOF [60], Equations (2)-(4), have also been derived to evaluate the best model.

$$\text{AIC} = \frac{\text{RSS} \cdot (n + p')}{(n - p')^2} \tag{2}$$

$$\text{FIT} = \frac{r^2 \cdot (n - k - 1)}{(n+k)^2 \cdot (1 - r^2)} \tag{3}$$

$$\text{LOF} = \frac{\text{RSS}/n}{[1 - \frac{k(1+d)}{n}]^2} \tag{4}$$

In above equations, RSS is the sum of the squared differences between the observed and estimated activity values, k is the number of variables in the model, p' is the number of adjustable parameters in the model and d is the smoothing parameter. 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 (Fisher-ratio), was proved to be a useful parameter for assessing the quality of the models. The model that

produces the minimum value of AIC and the highest value of FIT is considered potentially the most useful and the best. The LOF takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large numbers of parameters. A minimum LOF value infers that the derived model is statistically sound.

The internal validation of derived model was ascertained through the cross-validated index, Q^2, from leave-one-out and leave-five-out procedures. The LOO method creates a number of modified data sets by taking away one compound from the parent data set in such a way that each observation has been removed once only. Then one model is developed for each reduced data set and the response values of the deleted observations are predicted from these models. The squared differences between predicted and actual values are added to give the predictive residual sum of squares, PRESS. In this way, PRESS will contain one contribution from each observation. The cross-validated Q^2_{LOO} value may further be calculated as

$$Q^2_{LOO} = 1 - \frac{PRESS}{SSY} \tag{5}$$

where, SSY represents the variance of the observed activities of molecules around the mean value. In leave-five-out procedure a group of five compounds is randomly kept outside the analysis each time in such a way that all compounds, for once, become the part of the predictive groups. A value greater than 0.5 of Q^2-index hints towards a reasonable robust model.

The external validation or predictive power of derived model is based on test-set compounds. The squared correlation coefficient between the observed and predicted values of compounds from test-set, r^2_{Test}, has been calculated as

$$r^2_{Test} = 1 - \frac{\sum[Y_{Pred(Test)} - \overline{Y}_{(Test)}]^2}{\sum[Y_{(Test)} - \overline{Y}_{(Training)}]^2} \tag{6}$$

where, $Y_{Pred(Test)}$ and $Y_{(Test)}$ are the predicted and the observed activity values, respectively, of the test-set compounds and $\overline{Y}_{(Training)}$ and $\overline{Y}_{(Test)}$ are, correspondingly, the mean observed activity values of the training-set and test-set compounds. r^2_{Test} is the squared correlation coefficient between the observed and predicted data of the test-set. A value greater than 0.5 of r^2_{Test} suggests that the model obtained from training-set has a reliable predictive power.

2.3 Partial Least-Squares Analysis

Partial Least-Squares (PLS) [61-63] linear regression is a method suitable for overcoming the problems in MLR related to multicollinear or over-abundant descriptors. This is a modeling technique where information in the descriptor matrix X is projected onto a small number of latent variables (LV) called PLS components, which are linear combination of the original variables. The matrix Y is simultaneously used in estimating the "latent" variables in X that will be most relevant to predict the Y variables. All descriptor variables are preprocessed by autoscaling, using weights based on the variables' standard deviation and the data are mean-centered prior to PLS processing. Scaling of descriptors is necessary because the values have different orders of magnitude.

Cross-validation was employed to select the used optimum number of LVs. With cross-validation, some samples were kept out of the calibration and used for prediction. The process was repeated so that each of the samples was kept out once. The predicted values of left-out samples were then compared to the observed values using predicted residual sum of squares (PRESS). The PRESS obtained in the cross-validation was calculated each time that a new LV was added to the model.

Tables should be explanatory enough to be understandable without any text reference. Double spacing should be maintained throughout the table, including table headings and footnotes. Table headings should be placed above the table. Footnotes should be placed below the table with superscript lowercase letters.

3. RESULTS AND DISCUSSION

A total number of 484 descriptors, belonging to 0D-2D classes of DRAGON, have been computed for 61 compounds of Table 1. Next, the descriptors which were inter-correlated above 0.90 and exhibited correlation less than 0.1 with biological activities have been eliminated in the initial stage. The remaining 141 and 131 descriptors able to address, respectively, the MMP-13 and the MMP-13 w/BSA inhibition activities of the compounds have been scaled and collated in the separate pools for CP-MLR analyses. A test-set has been selected through SYSTAT and the same was used for external validation of the models, derived from the training-set compounds. Fourteen compounds (S. Nos. **3**, **4**, **8**, **11**, **15**, **21**, **23**, **36**, **44**, **45**, **46**, **50**, **55** and **61**; Table 2) were identified for the test-set while remaining compounds constitute the training-set for MMP-13 and MMP-13 w/BSA activities. A number of models in two-, three-, four-, five- and six-descriptors have been derived in succession. In doing so, filter-3 was in turn incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation.

In order to quantify MMP-13 inhibition activity in terms of molecular descriptors, compounds **5** and **6** (Table 2) appeared to behave indifferently from other compounds of the series. The *ortho*- or *para*-substitution of the phenyl ring, in these two congeners appeared to be poorly tolerated. Both these compounds have been treated at the "outliers". The training-set was then employed to explore predictive models through CP-MLR. 4 Models in four-descriptors and 2 models in five-descriptors only remained statistically significant and the same, in increasing level of significance, are given through Equations (7)-(12).

pIC_{50} (MMP-13) = 5.256 + 0.926(0.345)TIE + 2.090(0.291)S2K + 2.968(0.342)BEHp1 + 0.591(0.267)nNR2Ph
$n = 45$, $r = 0.907$, $s = 0.417$, $F(4, 40) = 46.489$, AIC = 0.217, LOF = 229, FIT = 3.048,
$Q^2_{LOO} = 0.750$, $Q^2_{L5O} = 0.752$, $r^2_{Test} = 0.523$ (7)

pIC_{50} (MMP-13) = 6.012 + 0.878(0.337)TIE + 1.884(0.277)S2K + 2.938(0.338)BEHp1 − 0.867(0.351) MATS7m
$n = 45$, $r = 0.910$, $s = 0.411$, $F(4, 40) = 48.027$, AIC = 0.212, LOF = 222, FIT = 3.149,
$Q^2_{LOO} = 0.785$, $Q^2_{L5O} = 0.798$, $r^2_{Test} = 0.503$ (8)

pIC_{50} (MMP-13) = 5.527 + 1.699(0.325)MW + 3.423(0.317)BEHp1 − 1.070(0.244)JGI3 + 0.893(0.238)N-072
$n = 45$, $r = 0.913$, $s = 0.404$, $F(4, 40) = 50.172$, AIC = 0.204, LOF = 215, FIT = 3.290,

$Q^2_{LOO} = 0.760$, $Q^2_{L5O} = 0.755$, $r^2_{Test} = 0.502$ \hfill (9)

pIC_{50} (MMP-13) = 5.010 + 2.505(0.245)BEHm1 + 1.240(0.230)BEHm8 + 2.225(0.254)BELm1 + 1.080(0.230)N-072
n = 45, r = 0.921, s = 0.385, F(4, 40) = 56.175, AIC = 0.186, LOF = 195, FIT = 3.684,
$Q^2_{LOO} = 0.808$, $Q^2_{L5O} = 0.802$, $r^2_{Test} = 0.638$ \hfill (10)

pIC_{50} (MMP-13) = 5.501 + 1.994(0.303)MW + 3.175(0.293)BEHp1 − 1.226(0.223)JGI3 + 0.791(0.237)nNR2Ph + 0.962(0.214)N-072
n = 45, r = 0.933, s = 0.361, F(5, 39) = 52.589, AIC = 0.170, LOF = 186, FIT = 3.756,
$Q^2_{LOO} = 0.799$, $Q^2_{L5O} = 0.788$, $r^2_{Test} = 0.608$ \hfill (11)

pIC_{50} (MMP-13) = 5.098 + 2.582(0.220)BEHm1 + 1.076(0.211)BEHm8 + 2.066(0.232)BELm1 − 0.883(0.264)C-027 + 1.076(0.207)N-072
n = 45, r = 0.939, s = 0.344, F(5, 39) = 58.598, AIC = 0.155, LOF = 170, FIT = 4.186,
$Q^2_{LOO} = 0.840$, $Q^2_{L5O} = 0.828$, $r^2_{Test} = 0.538$ \hfill (12)

In all above equations, the F-values remained significant at 99% level [$F_{4,40}(0.01) = 3.828$, $F_{5,39}(0.01) = 3.528$]. The descriptor, MW (from CONST class) accounts for the molecular weight of a compound. The descriptors, BEHkw and BELkw (from BCUT class) represent, respectively the highest and the lowest eigenvalues of Burden matrices, in which k is eigenvalue rank and w is atomic property, such as, mass (m) and polarizability (p). The descriptors, TIE and S2K (from TOPO class) are representative of E-state topological parameter and 2-path Kier alpha-modified shape index respectively. The descriptor, MATS7m (from 2DAUTO class) stands for the Moran autocorrelation – lag 7/ weighted by atomic masses. The nNR2Ph (from FUNC class) corresponds to the number of tertiary amines (aromatic). The C-027 and N-072 (from ACF class) signify the functionality, such as, R-CH-X and RCO-N< / >N-X=X respectively. The sign of the regression coefficient of a specified descriptor indicated the direction of its influence in above models. The positive regression coefficient will augment the activity profile of a compound while the negative coefficient will cause detrimental effect to it.

Equations (7)-(10) are the four-descriptor models while Equation (11)-(12) are the five-descriptor models. However, Equations (10) and (12) only remained highest significant amongst these models which have accounted for 85 and 88 percent of variances in observed activity values respectively. In fact Equation (10) represents the subset of Equation (12), therefore, the latter model is retained for further discussion. The participated descriptors in Equation (12) are mainly obtained from Burden matrix. The descriptors, from this matrix, are then calculated as an ordered sequence of the highest and lowest eigenvalues, which have been demonstrated to reflect relevant aspects of molecular structure, and are therefore useful for similarity searching. The eigenvalue ranks 1 and 8, each of them has been weighted by atomic masses (m), have emerged as important measures to influence the inhibition of MMP-13. The higher values of the descriptors, BEHm1, BELm1 and BEHm8 would, therefore, augment the activity of a compound. Additionally, the functionality RCO-N< or >N-X=X is desirable but R-CH-X is unwanted in a molecule for improvement of its inhibition activity.

This model has been used to calculate the MMP-13 inhibition activity profiles of all the compounds. The same are included in Table 2 for the sake of comparison with observed ones. A close agreement between them has been observed. Moreover, the graphical display

showing the variation of observed versus calculated activities is given in Fig. 2 to depict the goodness of fit.
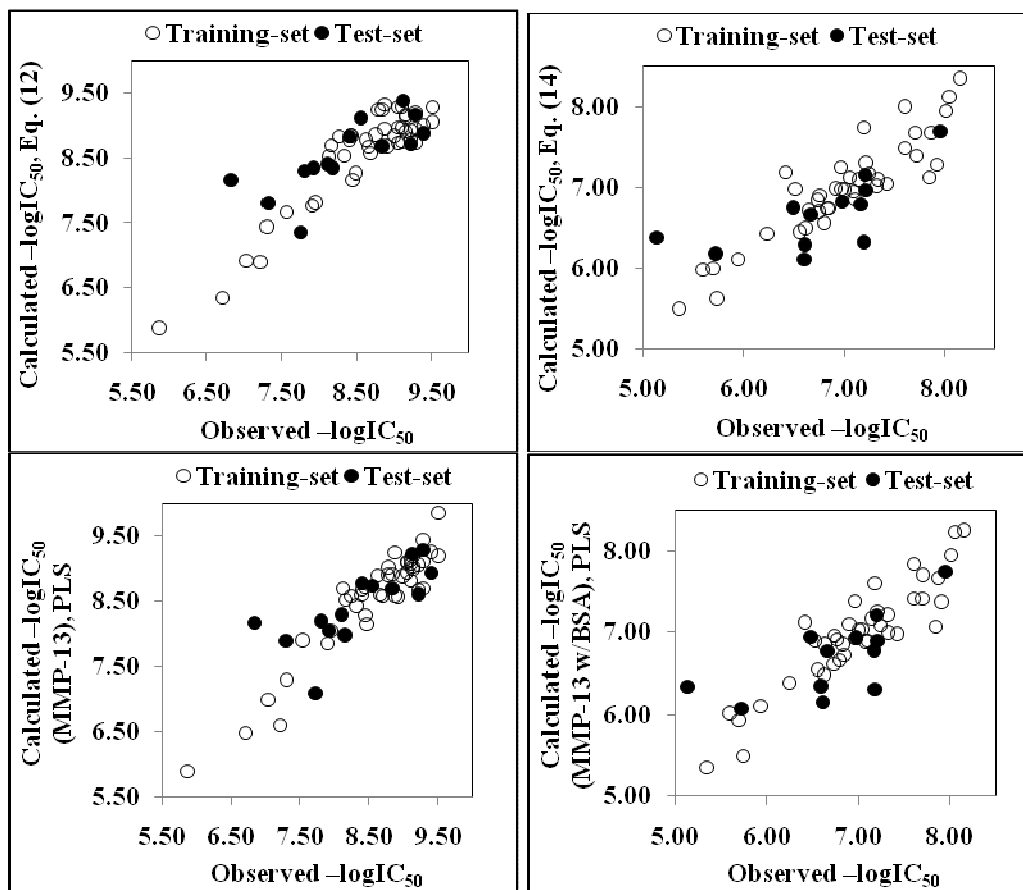
Fig. 2. Plot of observed versus caculated –logIC$_{50}$ values relating to inhibition of MMP-13 and MMP-13 w/BSA for training-set and test-set compounds

Actually, the five-descriptor models, Equations (11)-(12), could explain the highest variance in observed inhibition activity. All statistical parameters have remained superior to those of four-descriptor models (Equations 7-10). These two models have shared 9 descriptors among them and the class, brief description, average regression coefficient and total incidences, for individual descriptor, are given in Table 4.

Table 4. Identified descriptors[a] along with their physical meaning, average regression coefficient and incidence[b], in modeling the inhibition activities

S. No.	Descriptor	Descriptor class	Physical meaning	Average regression coefficient (incidence)	
				MMP-13	**MMP-13 w/BSA**
1	MW	CONST	Molecular weight.	1.994 (1)	
2	MSD	TOPO	Mean square distance index (Balaban).		1.202 (2)
3	BEHm1	BCUT	Highest eigenvalue n. 1 of Burden matrix/ weighted by atomic masses.	2.582 (1)	
4	BEHm8	BCUT	Highest eigenvalue n. 8 of Burden matrix/ weighted by atomic masses.	1.076 (1)	
5	BEHp1	BCUT	Highest eigenvalue n. 1 of Burden matrix/ weighted by atomic polarizabilities.	3.175 (1)	
6	BELm1	BCUT	Lowest eigenvalue n. 1 of Burden matrix/ weighted by atomic masses.	2.066 (1)	
7	BELp4	BCUT	Lowest eigenvalue n. 4 of Burden matrix/ weighted by atomic polarizabilities.		0.522 (1)
8	JGI3	GLVZ	Mean topological charge index of order 3.	-1.226 (1)	
9	MATS1m	2DAUTO	Moran autocorrelation – lag 1/ weighted by atomic masses.		-1.034 (1)
10	MATS3e	2DAUTO	Moran autocorrelation – lag 3/ weighted by atomic Sanderson electronegativities.		-1.180 (1)
11	GATS3e	2DAUTO	Geary autocorrelation – lag 3/ weighted by atomic Sanderson electronegativities.		1.427 (1)
12	nNR2Ph	FUNC	Number of tertiary amines (aromatic).	0.791 (1)	
13	nCrHR	FUNC	Number of ring tertiary C (sp^3).		0.725 (2)
14	C-027	ACF	Corresponds to R--CH--X.	-0.883 (1)	
15	N-072	ACF	Corresponds to RCO-N< / >N-X=X.	1.019 (2)	
16	H-047	ACF	H attached to C1(sp^3)/C0(sp^2).		1.632 (2)
17	H-050	ACF	H attached to heteroatom.		0.752 (2)

[a]*The descriptors have been identified from the models, emerged from CP-MLR protocol with a training-set of 45 and 41 compounds for MMP-13 and MMP-13 w/BSA inhibition activities respectively.* [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.*

Carboxylic acids and other lipophilic compounds are known to bind to plasma proteins [64]. These compounds, therefore, lose potency in the presence of serum obtained from human, rat, and rabbit. Similar results have been observed when such compounds were tested in presence of bovine serum albumin (BSA). Further, the activity determined in the presence of BSA was reported to be important for *in vivo* activity in the rat intra-articular (IA) model. Thus, the *in vitro* MMP-13 inhibition activity, in presence of 1% BSA, i.e., pIC_{50} (MMP-13 w/BSA) has been quantitatively analyzed further in terms of chemometric descriptors, able to encode molecular structures of the compounds in Table 1. For this, a total number of 131 descriptors, identified at initial stage, have been subjected to CP-MLR to discover significant models. As the mode of action, in the presence and in the absence of BSA, was different, therefore, it is interesting to investigate the models which could explain it in terms of molecular features of the compounds. To impress upon similar structural features of the compounds, which are able to address MMP-13 inhibition action in presence of BSA, the same test-set has been chosen to identify significant models through C-MLR. Thus, models in two-, three-, four- and five-descriptors have been derived successively but none of them could divulge statistical significant results. Finally, 2 models, each derived in six-descriptors, have satisfactorily explained the variance in observed activity.

pIC_{50} (MMP-13 w/BSA) = 6.134 + 1.181(0.247)MSD − 1.034(0.265)MATS1m
− 1.180(0.229)MATS3e + 0.757(0.172)nCrHR + 1.687(0.230)H-047 + 0.682(0.157)H-050
n = 41, r = 0.907, s = 0.315, F(6, 34) = 26.345, AIC = 0.140, LOF = 0.164, FIT = 2.053,
Q^2_{LOO} = 0.757, Q^2_{L5O} = 0.767, r^2_{Test} = 0.527 (13)

pIC_{50} (MMP-13 w/BSA) = 4.132 + 1.224(0.232)MSD + 0.522(0.223)BELp4
+ 1.427(0.244)GATS3e + 0.694(0.168)nCrHR + 1.577(0.223)H-047 + 0.822(0.158)H-050
n = 41, r = 0.907, s = 0.315, F(6, 34) = 26.426, AIC = 0.140, LOF = 0.164, FIT = 2.059,
Q^2_{LOO} = 0.748, Q^2_{L5O} = 0.685, r^2_{Test} = 0.523 (14)

Eight compounds (S. Nos. **5**, **6**, **10**, **11**, **17**, **18**, **19** and **21**, Table 2), with uncertain activities (IC_{50} > 10000 or pIC_{50} < 5.00), have been removed in the derivation of Equation (13) and (14). A total number of 8 descriptors, participating in above models, have been identified through CP-MLR and are listed in Table 4 along with their class, brief description, average regression coefficient and total incidences. Four descriptors, MSD, nCrHR, H-047 and H-050, representing, respectively, mean square distance index (Balaban), the number of ring tertiary C (sp^3), H attached to C1(sp^3)/C0(sp^2) and H attached to heteroatom are the part of both above models. Since these descriptors make positive impact on activity, structural features incorporating them are desirable to enhance its inhibition action. The Moran autocorrelation – lag 1 and – lag 3, weighted respectively by atomic masses (MATS1m) and atomic Sanderson electronegativities (MATS3e) have both negative impact on activity (Equation 13). Therefore, lower or more negative values of these descriptors would be beneficial in improving the inhibition activity. Additionally, the higher values of the lowest eigenvalue n. 4 of Burden matrix/ weighted by atomic polarizabilities (BELp4) and the Geary autocorrelation – lag 3/ weighted by atomic Sanderson electronegativities (GATS3e) will enhance the activity of a compound (Equation 14).

Further, the PLS analyses have also been performed on 9 and 8 identified descriptors related, respectively, to MMP-13 and MMP-13 w/BSA inhibition activities of the compounds and the results are summarized in Table 5.

Table 5. PLS and MLR-like PLS models from the descriptors of five and six parameter CP-MLR models for MMP-13 and MMP-13 w/BSA inhibition activities

A: PLS equation

PLS components	PLS coefficient (s.e.)[a]		
	MMP-13	**MMP-13 w/BSA**	
Component-1	0.580 (0.030)	0.440 (0.035)	
Component-2	0.192 (0.038)	-0.163 (0.036)	
Component-3	---	0.119 (0.047)	
Constant	8.499	6.986	

B: MLR-like PLS equation

S. No.	MMP-13			S. No.	MMP-13 w/BSA		
	Descriptor	**MLR-like coefficient (f. c.)[b]**	**Order**		**Descriptor**	**MLR-like coefficient (f. c.)[b]**	**Order**
1	MW	0.644 (0.066)	8	1	MSD	1.060(0.146)	3
2	BEHm1	1.050 (0.148)	2	2	BELp4	0.355(0.054)	8
3	BEHm8	0.756 (0.107)	5	3	MATS1m	-0.674(-0.087)	6
4	BELm1	0.888 (0.113)	4	4	MATS3e	-0.498(-0.075)	7
5	BEHp1	1.947 (0.205)	1	5	GATS3e	0.769(0.107)	5
6	JGI3	-0.780 (-0.106)	6	6	nCrHR	0.694(0.133)	4
7	nNR2Ph	0.534 (0.072)	7	7	H-047	1.594(0.231)	1
8	C-027	-0.574(-0.064)	9	8	H-050	0.794(0.168)	2
9	N-072	0.839(0.117)	3		Constant	5.159	
	Constant	5.532					

C: PLS regression statistics

Symbol	Value	
	MMP-13	**MMP-13 w/BSA**
n	45	41
r	0.952	0.913
s	0.297	0.292
F	201.061	62.041
Q^2_{LOO}	0.884	0.797
Q^2_{L5O}	0.816	0.794
r^2_{Test}	0.877	0.533

[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; f.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit standard deviation) data.*

In the study, the descriptors were autoscaled (zero mean and unit standard deviation) to provide each one of them equal weightage. In the PLS cross-validation, two- and three-components remained optimum for each of these 9 and 8 descriptors and they have explained, respectively, 90.6% and 83.4% of variances in the said activities. The PLS equations of optimum two- and three-components and MLR-like PLS coefficients of identified descriptors for MMP-13 and MMP-13 w/BSA activities are given in Table 5. The calculated activity values of training- and test-set compounds remained in close agreement to that of the observed ones and are listed in Table 2. For comparison, the plot between observed and calculated activities (through PLS analyses) for the training- and test-set compounds is given in Fig. 2. Fig. 3 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity (Table 5).
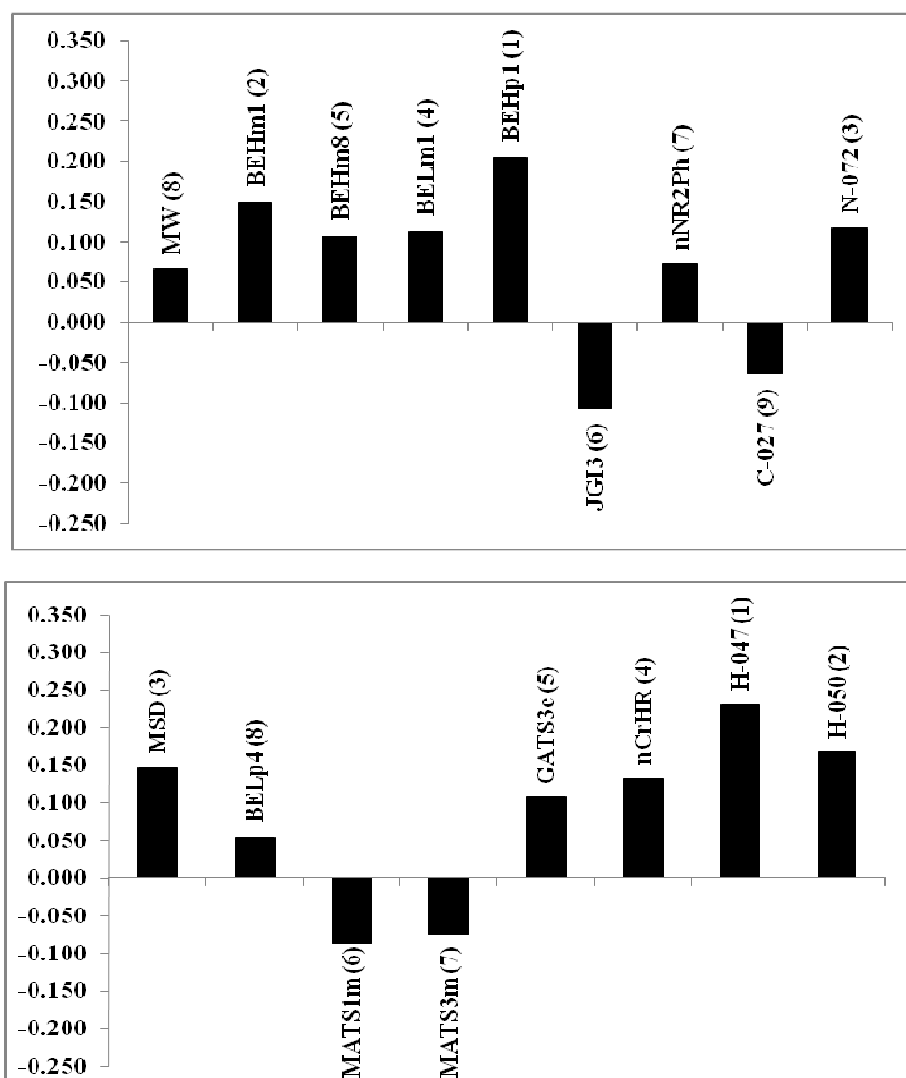

Fig. 3. Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 9 and 8 identified descriptors (Table 5) associated, resepectively, with MMP-13 and MMP-13 w/BSA inhibition activities of the compounds

In decreasing level of significance, 9 descriptors, being the part of Equations (11) and (12) have been arranged as BEHp1, BEHm1, N-072, BELm1, BEHm8, JGI3, nNR2Ph, MW and C-027 for MMP-13 inhibition activity while 8 descriptors, shared Equations (13) and (4), have been arranged as H-047, H-050, MSD, nCrHR, GATS3e, MATS1m, MATS3e and BELp4 for MMP-13 w/BSA activity. Similar conclusions have been observed from the PLS models for these two activities. Further the descriptors, BEHp1, BEHm1, N-072, BELm1, BEHm8, nNR2Ph and MW have positive contribution to MMP-13 activity and the descriptors JGI3 and C-027 have negative contribution to it. The order of the descriptors in PLS analysis suggested that MMP-13 inhibition activity of titled compounds is positively related to the molecular bulk or polarizabilty (descriptors MW, BEHm1, BEHm8, BELm1 and BEHp1). The presence of aromatic tertiary amine functionality and RCO-N< / >N-X=X type fragment and absence of R--CH--X type fragment in molecular structure are also important for the elevated MMP-13 inhibition activity.

Likewise, the descriptors H-047, H-050, MSD, nCrHR, GATS3e, and BELp4 have positively contributed to MMP-13 w/BSA activity while the descriptors MATS1m and MATS3e have negative contribution to it. The descriptors, in a given significant model, having positive contribution will augment the activity and their higher values are desirable to further improve it. On the other hand, the descriptors having negative contribution will diminish the activity. The lower or more negative values of such descriptors may, therefore, enhance the activity of a compound. Thus, more number of hydrogen atoms attached to primary and secondary carbon atoms and heteroatom (descriptors H-047 and H-050) in addition to the higher number of ring tertiary carbons in a molecular structure would be beneficiary to MMP-13 w/BSA activity. The atomic properties like mass, electronegativity and polarizability have also shown their relevance to MMP-13 w/BSA inhibitory activity.

4. CONCLUSION

The MMP-13 inhibition activity (in presence and in absence of BSA), of carboxylic acid based compounds have been quantitatively analyzed in terms of chemometric descriptors. The statistically validated QSAR models provided rationales to explain the inhibition activities of these congeners.

For MMP-13 inhibition activity, the identified descriptors have highlighted the role of the atomic properties such as mass and/or molecular bulk and polarizability. The MMP-13 inhibition activity has shown positive correlation to molecular bulk or polarizability accounting descriptors MW, BEHm1, BEHm8, BELm1 and BEHp1 and their higher values are desirable in improving the activity of a compound. The presence of aromatic tertiary amine functionality and RCO-N< / >N-X=X type fragment and absence of R--CH--X type fragment in molecular structure are also important for the elevated MMP-13 inhibition activity.

For MMP-13 w/BSA inhibition activity, the structural features (or properties), rationalized by the descriptors MSD, nCrHR, H-047 and H-050, have imparted positive impact on it. Similarly, the descriptors, BELp4 and GATS3e have shown their affirmative role whereas the descriptors, MATS1m and MATS3e have shown their negative influence on the activity. More number of hydrogen atoms attached to primary and secondary carbon atoms and heteroatom (descriptors H-047 and H-050) in addition to the higher number of ring tertiary carbons in a molecular structure would be beneficiary to MMP-13 w/BSA activity. The atomic properties like mass, electronegativity and polarizability have also shown their relevance to MMP-13 w/BSA inhibitory activity.

These guidelines may be used for further synthesis of potential analogues of the series. The PLS analysis has been carried out on identified descriptors related, respectively, to MMP-13 and MMP-13 w/BSA inhibition activities of the compounds. The results obtained from the CP-MLR approach and the PLS analysis have corroborated each other.

CONSENT

Not applicable.

ETHICAL APPROVAL

Not applicable.

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper

REFERENCES

1. Sternlicht MD, Werb Z. How matrix metalloproteinases regulate cell behavior. Annu Rev Cell Dev Biol. 2001;17:463–516.
2. Shapiro SD. Matrix metalloproteinase degradation of extracellular matrix: biological consequences. Curr Opin Cell Biol. 1998;10:602–608.
3. Whittaker M, Floyd CD, Brown P, Gearing AJ. Design and therapeutic application of matrix metalloproteinase inhibitors. Chem Rev. 1999;99:2735–2776.
4. Overall CM, Lo´pez-Otin C. Strategies for MMP inhibition in cancer: innovations for the post-trial era. Nat Rev Cancer. 2002;2:657–672.
5. Holleran WM, Galardy RE, Gao WN, Levy D, Tang PC, Elias PM. Matrix metalloproteinase inhibitors reduce phorbol ester induced cutaneous inflammation and hyperplasia. Arch Dermatol Res. 1997;289:138–144.
6. Hewson AK, Smith T, Leonard JP, Cuzner ML. Suppression of experimental allergicencephalomyelitis in the Lewis rat by the matrix metalloproteinase inhibitor Ro31-9790. Inflamm Res. 1995;44:345–349.
7. Chandler S, Coates R, Gearing A, Lury J, Wells G, Bone E. Matrix metalloproteinases degrade myelin basic protein. Neurosci Lett. 1995;201:223–226.
8. Lohmander LS, Hoerrner LA, Lark MW. Metalloproteinases, tissue inhibitor, and proteoglycan fragments in knee synovial fluid in human osteoarthritis. Arthritis Rheum. 1993;36:181–189.
9. Ahrens D, Koch AE, Pope RM, Stein-Picarella M, Niedbala MJ. Expression of matrix metalloproteinase 9 (96-kd gelatinase B) in human rheumatoid arthritis. Arthritis Rheum. 1996;39:1576–1587.

10. Blaser J, Triebel S, Maasjosthusmann U, Romisch J, Krahl-Mateblowski U, Freudenberg W, Fricke R, Tschesche H. Determination of metalloproteinases, plasminogen-activators and their inhibitors in the synovial fluids of patients with rheumatoid arthritis during chemical synoviorthesis. Clin Chim Acta. 1996;244:17–33.

11. Ohishi K, Fujita N, Morinaga Y, Tsuruo T. H-31 human breast cancer cells stimulate type I collagenase production in osteoblast-like cells and induce bone resorption. Clin Exp Metastasis. 1995;13:287–295.

12. Witty JP, Foster SA, Stricklin GP, Matrisian LM, Stern PH. Parathyroid hormone-induced resorption in fetal rat limb bones is associated with production of the metalloproteinases collagenase and gelatinase B. J Bone Miner Res. 1996;11:72–78.

13. Peress N, Perillo E, Zucker S. Localization of tissue inhibitor of matrix metalloproteinases in Alzheimer's disease and normal brain. J Neuropathol Exp Neurol. 1995;54:16–22.

14. Chang C, Werb Z. The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis. Trends Cell Biol. 2001;11:S37-S43.

15. Coussens LM, Fingleton B, Matrisian LM. Matrix metalloproteinase inhibitors and cancer: trials and tribulations. Science. 2002;295:2387–2392.

16. Puerta DT, Cohen SM. A bioinorganic perspective on matrix metalloproteinase inhibition. Curr Top Med Chem. 2004;4:1551–1573.

17. Matter H, Schudok M. Recent advances in the design of matrix metalloprotease inhibitors. Curr Opin Drug Discovery Dev. 2004;7:513–535.

18. Fingleton B. Matrix metalloproteinases as valid clinical targets. Curr Pharm Des. 2007;13:333–346.

19. Babine RE, Bender SL. Molecular recognition of protein-ligand complexes: applications to drug design. Chem Rev. 1997;97:1359–1472.

20. Grams F, Crimmin M, Hinnes L, Huxley P, Pieper M, Tschesche H, Bode W. Structure determination and analysis of human neutrophil collagenase complexed with a hydroxamate inhibitor. Biochemistry. 1995;34:14012–14020.

21. Muri EM, Nieto MJ, Sindelar RD, Williamson JS. Hydroxamic acids as pharmacological agents. Curr Med Chem. 2002;9:1631–1653.

22. Levin JI. The design and synthesis of aryl hydroxamic acid inhibitors of MMPs and TACE. Curr Top Med Chem. 2004;4:1289–1310.

23. Freskos JN, McDonald JJ, Mischke BV, Mullins PB, Shieh HS, Stegeman RA, Stevens AM. Synthesis and identification of conformationally constrained selective MMP inhibitors. Bioorg Med Chem Lett. 1999;9:1757–1760.

24. Tamura Y, Watanabe F, Nakatani T, Yasui K, Fuji M, Komurasaki T, Tsuzuki H, Maekawa R, Yoshioka T, Kawada K, Sugita K, Ohtani M. Highly selective and orally active inhibitors of type IV collagenase (MMP-9 and MMP-2): *N*-sulfonylamino acid derivatives. J Med Chem. 1998;41:640–649.

25. Hodgson J. Remodeling MMPIs. Biotechnology (N.Y.). 1995;13:554–557.

26. Slawomir W-P, Torri J, Johnson M, Steen V, Marshall J, Ness E, Dickson R, Sole M, Rosmussen H, Avanzato C, Hawkins M. Phase I trials marimastat, a metalloproteinase inhibitor, administered orally to patients with advanced lung cancer. J Clin Oncol. 1998;16:2150–2156.

27. Steward WP. Marimastat (BB2516): current status of development, cancer. Chemother Pharmacol. 1999;43(Suppl):S56–S60.

28. (a)Levitt NC, Eskens F, Propper DJ. A phase one pharmacokinetic study of CGS27023A, a matrix metalloproteinase inhibitor. Proc Amer Soc Clin Oncol (Abstract). 1998;17:213.

(b) Levitt NC, Eskens FALM, O'Byrne KJ, Propper DJ, Denis LJ, Owen SJ, Choi L, Foekens JA, Wilner S, Wood JM, Nakajima M, Talbot DC, Steward WP, Harris AL, Verweij J. Phase I and pharmacological study of the oral matrix metalloproteinase inhibitor, CGS 27023 (CGS27023A), in patients with advanced solid cancer. Clin Cancer Res. 2001;7(7):1912–1922.

29. Hande K, Wilding G, Ripple G. A phase I study of AG3340, a matrix metalloprotease inhibitor, in patients having advanced cancer. Ann Oncol (Abstract). 1998;9(Suppl. 2):74.

(b) Clendeninn N, Yeun G, Alberti D, Binger K, Wilding G. Phase I and pharmacokinetic study of prinomastat, a matrix metalloprotease inhibitor. Clin Cancer Res. 2004;10(3):909–915.

30. Collier MA, Yuen GJ, Bansal SK. A phase I study of the matrix metalloproteinase inhibitor AG3340 given in single doses to healthy volunteers; 88th Annual Meeting of the American Association for Cancer Research, 1997;221, abstract.

31. McGeehan GM, Becherer JD, Bast RC, Boyer CM, Champlon B, Connolly KM, Conway JG, Furdon P, Karp S, Kidao S, McElroy AB, Nichols J, Pryzwansky KM, Schoenen F, Sekut L, Truesdale A, Verghese M, Warner J, Ways JP. Regulation of tumour necrosis factors – a processing by a metalloproteinase Inhibitor. Nature. 1994;370:558.

32. Peschon JJ, Slack JL, Reddy P, Stocking KL, Sunnarborg SW, Lee DC, Russell WE, Caster BJ, Johnson RE, Fitzner JN, Boyce RW, Nelson N, Kozlosky CJ, Wolfon MF, Rauch CT, Ceretti DP, Paxton RJ, March CJ, Black RA. An essential role for ectodomain shedding in mammalian development. Science. 1998;282:1281–1284.

33. Gallea-Robache S, Morand V, Millet S, Btuneau J-M, Bhatnagar N, Chouaib S, Roman-Roman S. A metalloproteinase inhibitor blocks the shedding of soluble cytokine receptors and processing of transmembrane cytokine precursors in human monocytic cells. Cytokine. 1997;9:340–346.

34. Hooper NM, Kariran EH, Turner AJ. Membrane protein secretases. Biochem J. 1997;321:265–279.

35. Montana J, Baxter A. The design of selective non-substrate-based matrix metalloproteinase inhibitors. Curr Opin Drug Discovery Dev. 2000;3:353–361.

36. (a) Farrah Kheradmand F, Zena Werb Z. Shedding light on sheddases: role in growth and development. BioEssays. 2002;24:8–12.

(b) Huovila A-PJ, Turner AJ, Pelto-Huikko M, Kaerkkaeinen I, Ortiz RM. Shedding light on ADAM metalloproteinases. Trends Biochem Sci. 2005;30(7):413–422.

37. (a) Parker MH, Lunney E, Ortwine D, Pavlovsky G, Humblet C, Brouillette CG. Analysis of the binding of hydroxamic acid and carboxylic acid inhibitors to the stromelysin-1 (matrix metalloproteinase-3) catalytic domain by isothermal titration calorimetry. Biochemistry. 1999;38:13592-13601.

(b) Garmer DR, Gresh N, Roques B-P. Modeling of inhibitor-metalloenzyme interactions and selectivity using molecular mechanics grounded in quantum chemistry. Proteins: Struct, Funct, Genet. 1998;31:42-60.

(c) Castelhano AL, Billedeau R, Dewdney N, Donnelly S, Horne S, Kurz LJ, Liak TJ, Martin R, Uppington R, Yuan Z, Krantz A. Novel Indolactam-based inhibitors of matrix metalloproteinases. Bioorg Med Chem Lett. 1995;5:1415–1420.

(d) Donini OAT, Kollman P. Calculation and prediction of binding free energies for the matrix metalloproteinases. J Med Chem. 2000;43:4180–4188.

38. Aureli L, Gioia M, Cerbara I, Monaco S, Francesco G, Marini S, Ascenzi P, Topai A, Coletta M. Structural bases for substrate and inhibitor recognition by matrix metalloproteinases. Curr Med Chem. 2008;15:2192–2222.

39. Whittaker M, Floyd CD, Brown P, Gearing AJH. Design and Therapeutic application of matrix metalloproteinase inhibitors. Chem Rev. 1999;99:2735–2776.

40. Rao BG. Recent developments in the design of specific matrix metalloproteinase inhibitors aided by structural and computational studies. Curr Pharm Des. 2005;11(3):295–322.

41. Andrianjara C, Ortwine DF, Pavlovsky AG, Roark WH. Matrix metalloproteinase inhibitors. Patent WO 02/064080 A2; 2002.

42. 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, Quadros E, Reel N, Roland DM, Sakane Y, Singh H, Skiles J, Somers J, Toscano K, Wigg A, Zhou S, Zhu L, Shieh W-C, Xue S, McQuire LW. Discovery of potent, selective, and orally active carboxylic acid based inhibitors of matrix metalloproteinase-13. J Med Chem. 2009;52:3523–3538.

43. SYSTAT, Version 7.0; SPSS Inc 444 North Michigan Avenue, Chicago IL, 60611.

44. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. Available: http://www.cambridgesoft.com.

45. Dragon software (version 1.11-2001) by Todeschini R, Consonni V. Milano, Italy. Available: http//www.talete.mi.it/dragon.htm.

46. Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: Analysis of Selwood et al Data Set – A Case Study. QSAR Comb Sci. 2003;22:583-595.

47. Golbraikh A, Tropsha A. Beware of q^2 ! J Mol Graph Model. 2002;20:269-276.

48. Sharma S, Prabhakar YS, Singh P, Sharma BK. QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach. Eur J Med Chem. 2008;43:2354-2360.

49. Sharma S, Sharma BK, Sharma SK, Singh P, Prabhakar YS. Topological descriptors in modeling the agonistic activity of human A_3 adenosine receptor ligands: The derivatives of 2-chloro-N^6-substituted-4′-thioadenosine-5′-uronamide. Eur J Med Chem. 2009;44:1377-1382.

50. Sharma BK, Pilania P, Singh P, 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 Environ Res. 2010;21:169-185.

51. Sharma BK, Singh P, Sarbhai K, Prabhakar YS. A quantitative structure-activity relationship study on serotonin 5-HT_6 receptor ligands: Indolyl and piperidinyl sulphonamides. SAR QSAR Environ Res. 2010;21:369-388.

52. Sharma BK, Pilania P, Sarbhai K, Singh P, Prabhakar YS. Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. Mol Divers. 2010;14:371-384.

53. Sharma BK, Singh P, Shekhawat M, Sarbhai K, Prabhakar YS. Modeling of serotonin reuptake inhibitory and histamine H_3 antagonistic activity of piperazine and diazepane amides: QSAR rationales for cooptimization of the activity profiles. SAR QSAR Environ Res. 2011;22:365-383

54. So S-S, Karplus M. Three-dimensional quantitative structure-activity relationship from molecular similarity matrices and genetic neural networks. 1. Method and validation. J Med Chem. 1997;40:4347-4359.

55. 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 Comb Sci. 2004;23:234-244.

56. Akaike H. Information theory and an extension of the minimum likelihood principle. In: Petrov BN, Csaki F. Editors. Second international symposium on information theory. Akademiai Kiado, Budapest. 1973;267-281.

57. Akaike H. A new look at the statistical identification model. IEEE Trans Automat Control. 1974;AC-19:716-23.

58. Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. Quant Struct-Act Relat. 1994;13:285-94.

59. Kubinyi H. Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. Quant Struct-Act Relat. 1994;13:393-401.

60. Friedman J. In: Technical report no 102. Laboratory for computational statistics Stanford. Stanford University; 1990.

61. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics. 1978;20:397-405.

62. Kettaneh N, Berglund A, Wold S. PCA and PLS with very large data sets. Comput Stat Data Anal. 2005;48:69-85.

63. Stahle L, Wold S. Multivariate data analysis and experimental design in biomedical research. In: Ellis GP, West GB. (Eds.) Progress in Medicinal Chemistry, vol 25, Elsevier Science Publishers, BV, 1988;25:291-338.

64. Kragh-Hansen U, Chuang VTG, Otagiri M. Practical aspects of the ligand-binding and enzymatic properties of human serum albumin. Biol Pharm Bull. 2002;25(6):695–704.

_____

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=234&id=14&aid=1532