Research Article

Pharma Research Library

Chemometric descriptor based QSAR rationales for the MMP-13 Inhibition Activity of non-zinc-chelating compounds

Brij Kishore Sharma[1]*, Prithvi Singh[2]*

[1]Department of Chemistry, Government Post-Graduate College, Bundi, India
[2]Department of Chemistry, Shri Kalyan Government Post-Graduate College, Sikar, India
*E-mail: bksharma_sikar@rediffmail.com, psingh_sikar@rediffmail.com

ABSTRACT

The MMP-13 inhibition activity of non-zinc-chelating compounds has been quantitatively analyzed in terms of chemometric descriptors. The statistically validated quantitative structure-activity relationship (QSAR) models provided rationales to explain the inhibition activity of these compounds. The descriptors, identified through combinatorial protocol in multiple linear regression (CP-MLR) analysis, have highlighted the role of 3-path Kier alpha-modified shape index (S3K), complementary information content index of 1-order neighborhood symmetry (CIC1), eigenvalue sum from mass weighted distance matrix (SEigm), lowest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes (BELv6) and by atomic polarizabilities (BELp6), 3-order topological charge index (GGI3) and the functionality, R--CR--R (C-025). From statistically validated models, it appeared that the descriptors S3K, BELv6, BELp6 and SEigm make positive contribution to activity and their higher values are conducive in improving the MMP-13 inhibition activity of a compound. On the other hand, the descriptors CIC1, GGI3 and C-025 render detrimental effects to activity. Therefore, the absence of functionality, R--CR--R and lower values of descriptors CIC1 and GGI3 would be advantageous. PLS analysis has further corroborated the dominance of the CP-MLR identified descriptors. Applicability domain analysis revealed that the suggested models have acceptable predictability. All the compounds are within the applicability domain of the proposed models and were evaluated correctly.

Key words: QSAR, MMP-13 inhibitors, chemometric descriptors, combinatorial protocol in multiple linear regression (CP-MLR) analysis, non-Zn-chelating compounds.

INTRODUCTION

The matrix metalloproteinases (MMPs), a family of more than 27 zinc- and calcium-containing enzymes, are involved in the degradation of extracellular matrix and tissue remodeling [1-3]. Of the collagenase family (MMP-1, MMP-8 and MMP-13) MMP-13, the most efficient type II collagen-degrading MMP [4,5], has now become an attractive therapeutic target because its inhibition reduces cartilage degradation associated with the progression of rheumatoid arthritis and osteoarthritis in animal models [6,7]. However, broad-spectrum MMP inhibitors exhibit a dose-limiting toxicity leading to side effects such as a painful joint stiffening (musculoskeletal syndrome, MSS) and inflammation [8-15]. It was suggested that MSS is caused by the inhibition of normal extracellular matrix turnover due to inhibition of other MMPs rather than MMP-13 [16-21]. At present, it is unclear which MMP isoforms may be involved [22] and to what extent they contribute to MSS. Thus, selective inhibition of MMP-13, devoid of MSS, may prove to be better therapeutic research area.

MMPs having a tris(histidine)-bound zinc(II), acts as the catalytic site for the hydrolysis of substrate. Most MMP inhibitors achieve affinity through interaction with the catalytic zinc via a chelating moiety such as hydroxamic acid and by locating hydrophobic functionality in the S1′ pocket [8,28]. The S1′ pocket varies in length and amino acid sequence for different MMP isoforms. Such variations between MMP family members were, therefore, used to design MMP inhibitors with different selectivity profiles [23]. MMP-13 has additional region, S1′*, for inhibitor binding that has not been identified in other MMP isoforms. Most potent and selective MMP-13 inhibitors occupy both the S1′ and S1′* pockets only [24-27] and reduces the need to have a Zn-binding functionality.

In view of this, two new classes of potent and selective MMP-13 inhibitors involving unique binding mode at the active site and not interacting with the catalytic zinc, have recently been reported [28,29]. The general structure of these classes is shown in Figure 1. In the first series, the structural variations appeared at position R_1 and in incision X while in the second series, positions R_2 and R_3 have been varied. These functional variations are given in Table 1. The first series of compounds (**1**-**23**) were obtained through optimization with the aid of co-crystal structural information [28]. For this, the hit compound (**1**) was extended out from the active site into the S1′ pocket by adding an aryl group through two different linking functionalities. The aryl ring occupies the entrance to the S1′ pocket thus providing the opportunity to grow into the S1′ pocket to improve the potency against MMP-13 and the selectivity profile against other MMP isoforms. Depending on the linkage different trends for potency and selectivity for the respective aryl groups were observed.

To further improve potency against MMP-13, the second series (compounds **23**-**55**) was explored to investigate alternative ways of interacting with the S1′ pocket [29]. The starting point of this optimization was the result of a hybridization of hit structure (compound **1**) with another series of MMP-13 inhibitors based on an overlay of their crystal structures. In these analogues, the aryl groups was appended at the C-3 position of phenyl ring through a methyl amide linkage occupying the MMP S1′ pockets somewhat differently than the analogous functionality in first class of compounds. Additionally, the cyclohexyl group, able to bind in the S2′ pocket, was also replaced by other smaller substituents to modify the lipophilicity of some of the congeners.

In both reported studies, the structure-activity relationships (SARs) were, however, targeted at the alteration of substituents at different positions and provided no rationale to reduce the trial-and-error factors. Hence, in the present communication a 2D-quantitative SAR (2D-QSAR) has been conducted to provide the rationale for drug-design and to explore the possible mechanism of the action. In the congeneric series, where a relative study is being carried out, the 2D-descriptors may play important role in deriving the significant correlations 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 inhibition actions of compounds at molecular level.

MATERIAL AND METHODS

2.1 Data-set

For present work the non-Zn-chelating compounds (Table 1), along with their in vitro inhibition activity of MMP-13, have been taken from the literature [28,29]. The inhibition activity, IC_{50}, represents the concentration of a compound to achieve 50% inhibition of MMP-13 against type II collagen. The same is expressed as pIC_{50} on a molar basis and stand as the dependent descriptor for present quantitative analysis.

For modeling purpose, the complete data-set was divided into training- and test-sets. The training-set was used to derive statistical significant models while the test-set, consisting nearly 25% of total compounds, was employed to validate such models. The selection of test-set compounds was made through SYSTAT [30] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the inhibition activity, pIC_{50} values. The test-set compounds 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 intervals. Due to this reason, a single linkage clustering procedure was applied.

2.2 Molecular descriptors

The structures of the compounds (Table 1), under study, have been drawn in 2D ChemDraw [31] using the standard procedure. These structures were converted into 3D objects using the default conversion procedure implemented in the CS Chem3D Ultra. The generated 3D-structures of the compounds were subjected to energy minimization in the MOPAC module, using the AM1 procedure for closed shell systems, implemented in the CS Chem3D Ultra. This will ensure a well defined conformer relationship across the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software [32] for computing the descriptors corresponding to 0D-, 1D-, and 2D-classes. The combinatorial protocol in multiple linear regression (CP-MLR) [33] analysis and partial least-squares (PLS) [34-36] procedures have been used in the present work for developing QSAR models. A brief description of the computational procedure is given below.

2.3 Model development

The CP-MLR is a 'filter'-based variable selection procedure for model development in QSAR studies [33]. Its procedural aspects and implementation are discussed in some of our recent publications [37-42]. 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 r^2 or 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. Furthermore, in order to discover any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been put to a randomization test [43,44] 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.

To support the findings, a partial least squares (PLS) analysis has been carried out on descriptors identified through CP-MLR. The study facilitates the development of a 'single window' structure-activity model and help to categorize the potentiality of identified descriptors in explaining the MMP-13 inhibition activity profiles of the compounds. It also gives an opportunity to make a comparison of the relative significance among the descriptors. The fraction contributions obtainable from the normalized regression coefficients of the descriptors allow this comparison within the modeled activity.

2.4 Applicability Domain

The utility of a QSAR model is based on its accurate prediction ability for new compounds. 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 is assessed by the leverage values for each compound [45,46]. The Williams plot (the plot of standardized residuals versus leverage values, h) can then be used for an immediate and simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in the model. In this plot, the applicability domain is established inside a squared area within $\pm$ x (s.d.) and a leverage threshold h^*. The threshold h^* 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. Prediction must be considered unreliable for compounds with a high leverage value (h > h*). 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.

RESULTS AND DISCUSSION

3.1 QSAR results

For the compounds in Table 1, a total number of 495 descriptors belonging to 0D- to 2D- classes of DRAGON have been computed and were subjected to CP-MLR analysis. The preliminary assessment of complete data-set suggested that the lone compound **25**, having a methyl group at R$_2$ remained as an 'outlier'. Similarly compound **15**, due to its uncertain activity value, could not fit into the trend of remaining compounds of the series. Both these compounds were, therefore, ignored in the subsequent analyses. The remaining 53 compounds were further divided into training-set and test-set.

Thirteen compounds (nearly 25% of total population) have been selected for test-set through SYSTAT. The identified test-set was then used for external validation of models derived from remaining forty compounds in the training-set. The squared correlation coefficient between the observed and predicted values of compounds from test-set, r^2$_{Test}$, was calculated to explain the fraction of explained variance in the test-set which is not part of regression/model derivation. It is a measure of goodness of the derived model equation. A high r^2$_{Test}$ value is always good. But considering the stringency of test-set procedures, often r^2$_{Test}$ values in the range of 0.500–0.600 are regarded as indicative predictive models. Following the strategy to explore only predictive models, CP-MLR resulted into 70 models in two descriptors, 99 models in three descriptors, 8 models in four descriptors and 13 models in five descriptors. However, the highest significant of them, in statistical sense, are given through Equations (1)-(10)

$pIC_{50} = 5.983 + 2.242(0.324)S3K + 0.874(0.228)nRORPh$
$n = 40, r = 0.764, s = 0.525, F = 25.915, q^2_{LOO} = 0.505, q^2_{L5O} = 0.511, r^2_{Test} = 0.634$ (1)

$pIC_{50} = 6.230 + 1.434(0.417)VAR + 1.141(0.241)N\text{-}075$
$n = 40, r = 0.736, s = 0.551, F = 21.870, q^2_{LOO} = 0.470, q^2_{L5O} = 0.474, r^2_{Test} = 0.670$ (2)

$pIC_{50} = 6.598 + 3.470(0.415)S3K - 1.183(0.432)GGI3 - 1.541(0.323)C\text{-}025$
$n = 40, r = 0.846, s = 0.440, F = 30.112, q^2_{LOO} = 0.629, q^2_{L5O} = 0.616, r^2_{Test} = 0.504$ (3)

$pIC_{50} = 6.549 + 2.513(0.317)S3K - 1.745(0.327)C\text{-}025 + 0.556(0.260)C\text{-}027$
$n = 40, r = 0.833, s = 0.456, F = 27.299, q^2_{LOO} = 0.613, q^2_{L5O} = 0.615, r^2_{Test} = 0.553$ (4)

$pIC_{50} = 5.617 + 2.491(0.533)SEigm + 2.703(0.433)BELv6 - 3.152(0.648)GGI3$
$+ 0.779(0.232)N\text{-}075$
$n = 40, r = 0.856, s = 0.432, F = 23.993, q^2_{LOO} = 0.646, q^2_{L5O} = 0.652, r^2_{Test} = 0.751$ (5)

$pIC_{50} = 7.052 + 1.354(0.518)S3K - 0.515(0.224)PJI2 + 1.521(0.510)C\text{-}006$
$- 1.316(0.343)C\text{-}025$
$n = 40, r = 0.854, s = 0.435, F = 23.665, q^2_{LOO} = 0.644, q^2_{L5O} = 0.634, r^2_{Test} = 0.658$ (6)

$pIC_{50} = 7.363 + 2.626(0.491)S3K - 1.629(0.553)CIC1 + 1.677(0.632)BELv6$
$- 2.601(0.627)GGI3 - 0.967(0.356)C\text{-}025$
$n = 40, r = 0.880, s = 0.404, F = 23.260, q^2_{LOO} = 0.661, q^2_{L5O} = 0.648, r^2_{Test} = 0.652$ (7)

$pIC_{50} = 7.358 + 2.591(0.504)S3K - 1.570(0.541)CIC1 + 1.602(0.614)BELp6$
$- 2.528(0.612)GGI3 - 0.957(0.359)C\text{-}025$
$n = 40, r = 0.879, s = 0.405, F = 23.101, q^2_{LOO} = 0.661, q^2_{L5O} = 0.675, r^2_{Test} = 0.675$ (8)

$pIC_{50} = 6.140 + 2.230(0.654)S3K + 1.551(0.650)SEigm + 1.304(0.607)BELp6$
$- 2.483(0.679)GGI3 - 1.102(0.359)C\text{-}025$

$n = 40$, $r = 0.870$, $s = 0.418$, $F = 21.176$, $q^2_{LOO} = 0.639$, $q^2_{L5O} = 0.654$, $r^2_{Test} = 0.620$ (9)

$pIC_{50} = 6.115 + 2.267(0.642)S3K + 1.576(0.660)SEigm + 1.332(0.622)BELv6$
$- 2.520(0.693)GGI3 - 1.120(0.356)C\text{-}025$
$n = 40$, $r = 0.870$, $s = 0.419$, $F = 21.161$, $q^2_{LOO} = 0.637$, $q^2_{L5O} = 0.618$, $r^2_{Test} = 0.619$ (10)

Where n and F represent respectively the number of data points and the F-ratio between the variances of calculated and observed activities. The data within the parentheses are the standard errors associated with regression coefficients. In all above equations, the F-values remained significant at 99% level. The indices q^2_{LOO} and q^2_{L5O} (> 0.5), except baseline Equation (2), have accounted for their internal robustness. For all above models the r^2_{Test} values, obtained greater than 0.5, specified that the selected test-set is fully accountable for their external validation. The descriptors, in all above models, have been scaled between the intervals 0 to 1 [47] to ensure that a descriptor will not dominate simply because it has larger or smaller pre-scaled value compared to the other descriptors. In this way, the scaled descriptors would have equal potential to influence the QSAR models.

The signs of the regression coefficients have indicated the direction of influence of explanatory variables in above 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. Though Equations (1)-(10) emerged as significant predictive models but Equations (7)-(10) remained statistically more efficient. The later four models, involving five descriptors in each, could estimate up to 77.44 percent of variance in observed activity of the compounds. In fact, a total number of 13 such models, sharing 15 descriptors among them, have been obtained through CP-MLR and only four of them, being most significant have been documented through Equation (7)-(10). The shared 15 descriptors along with their brief description, average regression coefficients and total incidences are given in Table 2. Besides listed descriptors in Table 2, the other identified descriptors PJI2 and VAR are from topological class, nRORPh is from functional class and C-027 is from atom centred fragment class. The PJI2 represents the 2D Petitjean shape index (Equation 6), VAR explains the variation in a molecular structure (Equation 2), nRORPh accounts for the number of ethers (aromatic) (Equation 1) and C-027 encodes the functionality R--CH--X (Equation 4).

The further discussion is, however, based on the highest significant Equations (7)-(10). The derived statistical parameters of these four models have indicated that their level of significance is almost the same. These models were, therefore, used to calculate the activity profiles of all the compounds and are included in Table 3 for the sake of comparison with observed ones. A close agreement between them has been observed. Additionally, the graphical display, showing the variation of observed versus calculated activities is given in Figure 2 to insure the goodness of fit for each of these four models.

The participated descriptors in these models are S3K, CIC1, SEigm, BELv6, BELp6, GGI3 and C-025. These descriptors represent, respectively, 3-path Kier alpha-modified shape index, complementary information content index of 1-order neighborhood symmetry, eigenvalue sum from mass weighted distance matrix, lowest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes (v) and by atomic polarizabilities (p), 3-order topological charge index and the functionality, R--CR--R.

The S3K encodes information about the centrality of branching in the H-depleted molecular graph. The CIC1 measures the deviation of the information content pertaining to neighbourhood symmetry of 1-order (IC1) from its maximum value. The descriptor SEigm determines the sum of all the eigenvalues of atomic mass weighted distance matrix of the H-depleted molecular graph.

From Equations (7)-(10), it appeared that the descriptors S3K, BELv6, BELp6 and SEigm make positive contribution to activity while the descriptors CIC1, GGI3 and C-025 render the negative role to it. Thus to explore more potential analogues of the series, the values of prevalent descriptors of a given model may be decided according to delineated strategy. For example, Equation (7) has revealed that the higher values of descriptors S3K and BELv6, the lower (or more negative) value of descriptor CIC1 and absence of functionality R--CR--R, are all conducive in improving the MMP-13 inhibition activity of a compound. To corroborate the study further, a PLS analysis has also been carried out on 15 descriptors identified through CP-MLR and results are given in Table 4. For this purpose, the descriptors have been autoscaled (zero mean and unit s.d.) to give each one of them equal weight in the analysis.

In the PLS cross-validation, four components have been found to be the optimum for these 15 descriptors and they explained 78% variance in the activity ($r^2 = 0.780$). The MLR-like PLS coefficients of these 15 descriptors are given in Table 4. The calculated activity values of training- and test-set compounds are in close agreement to that of the observed ones and are listed in Table 2. For the sake of comparison, the plot between observed and calculated activities (through PLS analysis) for the training- and test-set compounds is given in Figure 2. Figure 3 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity (Table 4). Actually, the 15 identified descriptors have shared 55 PLS models and the analysis could reveal four components (Table 4) as optimum to explain the MMP-13 inhibition activity.

The top ten descriptors in decreasing order of significance are C-025, nNR2, MATS4m, GGI3, C-006, BELv6, BELp6, S3K, N-075, ATS5m (Table 4, figure 3). Among these descriptors, C-025, GGI3, C-006, BELv6, BELp6, S3K and N-075 are part of Equations discussed above and convey same inferences in PLS analysis. The negative contribution of functional group count descriptor nNR2 (number of tertiary aliphatic amine functionality in a molecule) advocates that a higher number of such functional groups are detrimental to activity. The positive contribution of atomic mass weighted 2D-autocorrelation descriptors (Moran autocorrelation, MATS4m and Broto-Moreau autocorrelation, ATS5m) suggest that higher value of these are helpful in improving the activity profile. It is also observed that PLS model from the dataset devoid of 15 descriptors (Table 4) remained inferior in explaining the activity of the analogues.

3.2 Applicability domain

On analyzing the applicability domain (AD) in the Williams plot (Figure 3) of the model based on the whole data set (Table 5), one compound (**25**; Table 1) has been 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 3×(standard deviation) units. None of the compounds was found to have leverage (h) values greater than the threshold leverage (h*). 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.

CONCLUSION

The MMP-13 inhibition activity of non-zinc-chelating compounds has been quantitatively analyzed in terms of chemometric descriptors. The statistically validated quantitative structure-activity relationship (QSAR) models provided rationales to explain the inhibition activity of these congeners. The descriptors identified through combinatorial protocol in multiple linear regression (CP-MLR) analysis have highlighted the role of 3-path Kier alpha-modified shape index (S3K), complementary information content index of 1-order neighborhood symmetry (CIC1), eigenvalue sum from mass weighted distance matrix (SEigm), lowest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes (BELv6) and by atomic polarizabilities (BELp6), 3-order topological charge index (GGI3 and the functionality, R--CR--R (C-025).

From statistically validated models, it appeared that the descriptors S3K, BELv6, BELp6 and SEigm make positive contribution to activity and their higher values are conducive in improving the MMP-13 inhibition activity of a compound. On the other hand, the descriptors CIC1, GGI3 and C-025 render detrimental effect to activity. Therefore, the absence of functionality, R--CR--R and lower values of descriptors CIC1 and GGI3 would be advantageous. Such guidelines may be helpful in exploring more potential analogues of the series. The statistics emerged from the test sets have validated the identified significant models. PLS analysis has further confirmed the dominance of the CP-MLR identified descriptors. Applicability domain analysis revealed that the suggested models have acceptable predictability. All the compounds are within the applicability domain of the proposed models and were evaluated correctly.
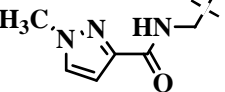
Acknowledgement
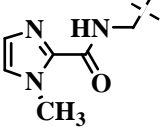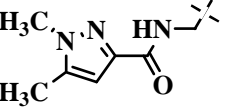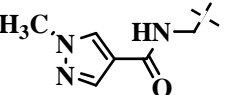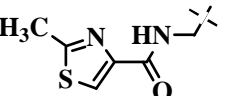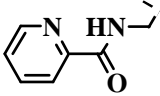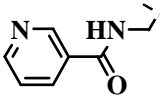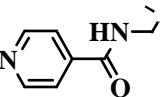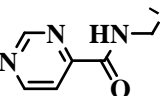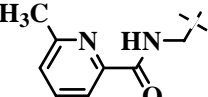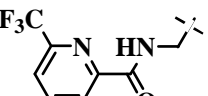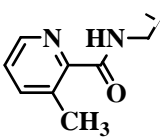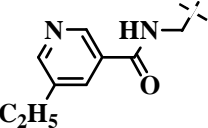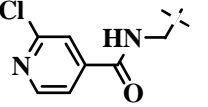
Table.1. Structural variations and MMP-13 inhibition activity of non-zinc-chelating compounds
(see Figure 1 for general structure)

Cpd.	R_1	R_2	X	R_3	pIC_{50} (M)[a]
1	Cl–	H	CH	cyclohexyl	6.37
2	benzofuran-2-carboxamide	H	CH	cyclohexyl	6.21
3	4,5-dimethylfuran-2-carboxamide	H	CH	cyclohexyl	7.20
4	5-methylfuran-2-carboxamide	H	CH	cyclohexyl	7.24
5	1-methylpyrazole-3-carboxamide	H	CH	cyclohexyl	7.51
6	1-methylpyrazole-4-carboxamide	H	CH	cyclohexyl	6.74
7	5-methyl-1H-pyrazole-3-carboxamide	H	CH	cyclohexyl	7.06
8	6-methylpyridine-2-carboxamide	H	CH	cyclohexyl	7.77
9	Cl–	H	N	cyclohexyl	6.61
10	H_3CC(O)NH	H	CH	cyclohexyl	5.81
11	H_3CNHC(O)	H	CH	cyclohexyl	5.22
12	H_3CS(O)$_2$NH	H	CH	cyclohexyl	6.34
13	H_3CNHS(O)$_2$	H	CH	cyclohexyl	6.32
14	H_3C–	H	CH	cyclohexyl	6.54
15	phenoxy	H	CH	cyclohexyl	<4.30

16	H₃C–O–⁄	H	CH	(cyclohexyl)	6.66
17	H₃C–O–⁄	H	N	(cyclohexyl)	7.49
18	(benzofuran)–O–	H	N	(cyclohexyl)	8.22
19	H₃C–(pyridine)–O–	H	N	(cyclohexyl)	7.40
20	(pyrazole)–O–	H	N	(cyclohexyl)	8.22
21	H₃C–(pyrazole)–O–	H	N	(cyclohexyl)	7.60
22	H₃C–(pyrazole)–O–	H	N	(cyclohexyl)	8.70
23	H₃C,H₃C–(pyrazole)–O–	H	N	(cyclohexyl)	8.40
24	H	(imidazopyridine)–C(O)–HN–		CH	(cyclohexyl)	7.00
25	H	H₃C–		CH	(cyclohexyl)	5.48
26	H	CH₃–(imidazopyridine)–C(O)–HN–		CH	(cyclohexyl)	6.96
27	H	H₃C–(imidazopyridine)–C(O)–HN–		CH	(cyclohexyl)	6.85
28	H	(furan)–C(O)–HN–		CH	(cyclohexyl)	7.12
29	H	H₃C–(furan)–C(O)–HN–		CH	(cyclohexyl)	7.96
30	H	(furan)–C(O)–HN–, CH₃		CH	(cyclohexyl)	6.74
31	H	CH₃–(furan)–C(O)–HN–		CH	(cyclohexyl)	6.48

32	H		CH		6.51
33	H		CH		7.62
34	H		CH		7.14
35	H		CH		6.85
36	H		CH		8.10
37	H		CH		8.10
38	H		CH		7.38
39	H		CH		7.51
40	H		CH		8.00
41	H		CH		8.40
42	H		CH		8.05
43	H		CH		7.00
44	H		CH		7.54
45	H		CH		7.57

46	H	Cl-pyrimidine-C(O)NH–C(CH₃)₃	CH	cyclohexyl	8.52
47	H	F₃C-pyrimidine-C(O)NH–C(CH₃)₃	CH	cyclohexyl	8.15
48	H	F₃C-pyridine-C(O)NH–C(CH₃)₃	CH	isobutyl	8.22
49	H	F₃C-pyridine-C(O)NH–C(CH₃)₃	CH	cyclopropyl	7.55
50	H	F₃C-pyridine-C(O)NH–C(CH₃)₃	CH	tetrahydropyranyl	8.70
51	H	F₃C-pyridine-C(O)NH–C(CH₃)₃	CH	C(CH₃)₂OH	7.60
52	H	F₃C-pyrimidine-C(O)NH–C(CH₃)₃	CH	isobutyl	8.05
53	H	F₃C-pyrimidine-C(O)NH–C(CH₃)₃	CH	cyclopropyl	7.44
54	H	F₃C-pyrimidine-C(O)NH–C(CH₃)₃	CH	tetrahydropyranyl	8.70
55	H	F₃C-pyrimidine-C(O)NH–C(CH₃)₃	CH	C(CH₃)₂OH	8.30

[a]IC_{50} represents the concentration of a compound to bring out 50% inhibition of MMP-13; expressed as pIC_{50} on molar basis; taken from reference [28,29].

Table 2. Observed and modeled MMP-13 inhibition activity of non-zinc-chelating compounds.

Cpd.	pIC_{50} (M)					Cpd.	pIC_{50} (M)				
	Obsd.	Calculated Eq.					Obsd.	Calculated Eq.			
		(7)	(8)	(9)	(10)	PLS			(7)	(8)	(9)	(10)	PLS
1	6.37	6.43	6.44	6.42	6.41	6.51	29	7.96	7.63	7.55	7.58	7.65	7.57
2	6.21	6.75	6.76	6.81	6.81	6.95	30	6.74	6.72	6.71	6.73	6.74	7.03
3	7.20	6.90	6.91	6.90	6.90	6.99	31[b]	6.48	6.50	6.50	6.48	6.48	6.75
4[b]	7.24	7.51	7.45	7.50	7.54	7.29	32	6.51	6.95	6.96	7.02	7.02	6.53
5	7.51	7.55	7.54	7.49	7.49	7.30	33[b]	7.62	7.62	7.62	7.56	7.56	7.52
6[b]	6.74	7.16	7.20	7.16	7.12	7.02	34	7.14	7.29	7.29	7.27	7.28	7.47
7	7.06	7.65	7.64	7.41	7.40	7.36	35	6.85	7.24	7.28	7.24	7.20	7.26
8	7.77	7.36	7.41	7.44	7.41	7.28	36	8.10	7.89	7.80	7.83	7.90	7.55
9	6.61	6.90	6.91	6.78	6.77	7.17	37	8.10	7.79	7.80	7.76	7.76	7.74
10[b]	5.81	6.91	6.91	6.83	6.83	6.65	38[b]	7.38	7.47	7.50	7.47	7.44	7.50

11	5.22	6.13	6.13	6.20	6.21	5.92	39	7.51	7.45	7.48	7.46	7.43	7.50
12 [b]	6.34	6.78	6.75	7.09	7.12	6.74	40	8.00	7.98	8.00	7.77	7.74	7.92
13	6.32	5.88	5.89	6.41	6.42	6.25	41 [b]	8.40	7.78	7.79	7.74	7.74	7.73
14	6.54	5.92	5.91	5.65	5.65	6.03	42 [b]	8.05	8.07	8.11	8.48	8.46	8.17
15 [c]	-	-	-	-	-	-	43	7.00	6.81	6.84	6.81	6.78	7.27
16	6.66	6.23	6.22	6.34	6.36	6.19	44	7.54	7.49	7.47	7.51	7.52	7.55
17	7.49	7.29	7.28	7.26	7.28	7.20	45	7.57	7.73	7.75	7.93	7.91	7.75
18 [b]	8.22	7.36	7.36	7.53	7.54	8.08	46	8.52	8.15	8.17	8.24	8.23	8.17
19	7.40	7.96	7.97	8.18	8.18	8.11	47 [b]	8.15	8.20	8.26	8.50	8.46	8.39
20	8.22	8.17	8.18	8.05	8.04	7.92	48	8.22	8.00	7.93	8.00	8.06	7.83
21	7.60	8.27	8.26	8.28	8.28	8.23	49 [b]	7.55	7.42	7.44	7.51	7.50	7.04
22	8.70	8.27	8.27	8.12	8.11	8.27	50	8.70	8.26	8.29	8.65	8.64	8.57
23	8.40	7.84	7.86	7.92	7.91	8.14	51	7.60	7.87	7.83	7.78	7.81	8.05
24	7.00	7.33	7.32	7.27	7.28	6.96	52	8.05	8.33	8.31	8.09	8.10	8.06
25 [d]	5.48	7.43	7.43	7.30	7.31	7.06	53	7.44	7.72	7.73	7.58	7.57	7.31
26	6.96	6.73	6.74	6.69	6.68	6.58	54	8.70	8.47	8.52	8.67	8.64	8.80
27 [b]	6.85	6.94	6.94	6.87	6.86	6.78	55	8.30	8.11	8.14	7.86	7.83	8.31
28	7.12	7.46	7.43	7.51	7.53	7.32							

[a] See footnote under Table 1, [b]compounds in test set, [c]compound with uncertain activity and not included in the study, [d]outlier compound in present study.

Table3. Identified descriptors[a] along with their physical meaning, average regression coefficient and incidence[b], in modeling the MMP-13 inhibition activity

S. No.	Descriptor	Descriptor class	Physical meaning	Average regression coefficient (incidence)
1	S3K	Topological	3-Path Kier alpha-modified shape index	2.135 (10)
2	CIC1	Topological	Complementary information content index of 1-order neighborhood symmetry	-1.949 (4)
3	SEigZ	Topological	Eigenvalue sum from Z weighted distance matrix (Barysz matrix)	1.733 (4)
4	SEigm	Topological	Eigenvalue sum from mass weighted distance matrix	1.811 (4)
5	BELv6	BCUT	Lowest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes	1.861 (7)
6	BELe1	BCUT	Lowest eigenvalue n. 1 of Burden matrix / weighted by atomic Sanderson electronegativities	-0.854 (4)
7	BELp6	BCUT	Lowest eigenvalue n. 6 of Burden matrix / weighted by atomic polarizabilities	1.382 (5)

8	GGI3	Galvez topological charge indices	3-Order topological charge index	-2.524 (12)
9	ATS5m	2D-autocorrelation	Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic masses	1.765 (1)
10	MATS4m	2D-autocorrelation	Moran autocorrelation - lag 4 / weighted by atomic masses	0.747 (1)
11	nCs	Functional	number of total secondary C(sp3)	1.379 (1)
12	nNR2	Functional	number of tertiary amines (aliphatic)	-0.741 (2)
13	C-006	Atom centered fragment	CH2RX	1.198 (2)
14	C-025	Atom centered fragment	R--CR--R	-1.161 (7)
15	N-075	Atom centered fragment	R--N--R / R--N--X	0.668 (1)

[a]The descriptors are identified from the five parameter models, emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.837, and filter-4 as $0.3 \le q^2 \le 1.0$ with a training set of 40 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.

Table 4. PLS and MLR-like PLS models from the descriptors of five parameter CP-MLR models for MMP-13 inhibition activity.

A: PLS equation		
PLS components	PLS coefficient (s.e.)[a]	
Component-1	-0.234 (0.025)	
Component-2	0.207 (0.039)	
Component-3	-0.079 (0.039)	
Component-4	0.149(0.072)	
Constant	7.423	
B: MLR-like PLS equation		
S. No.	Descriptor	MLR-like coefficient (f.c.)[b] Order
1	S3K	0.500(0.077) 8
2	CIC1	-0.060(-0.008) 13
3	SEigZ	0.038(0.005) 14
4	SEigm	0.023(0.003) 15
5	BELv6	0.523(0.086) 6
6	BELe1	-0.268(-0.042) 12
7	BELp6	0.496(0.083) 7
8	GGI3	-0.623(-0.093) 4
9	ATS5m	0.414(0.060) 10
10	MATS4m	0.909(0.110) 3
11	nCs	0.313(0.044) 11
12	nNR2	-0.745(-0.114) 2
13	C-006	0.689(0.089) 5
14	C-025	-0.837(-0.122) 1
15	N-075	0.294(0.063) 9
	Constant	6.276
C: PLS regression statistics	Values	

n	40
r	0.883
s	0.392
q^2_{LOO}	0.709
q^2_{L5O}	0.710
r^2_{Test}	0.771

[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 s.d.) data.

Table 5: Models derived for the whole data set (n = 54) for the MMP-13 inhibition activity in descriptors identified through CP-MLR

Model	r	s	F	q^2_{LOO}	Eq.
$pIC_{50} = 7.264 + 2.303(0.489)S3K$ $- 1.822(0.590)CIC1 + 2.187(0.601)BELv6$ $- 2.456(0.662)GGI3 - 1.069(0.387)C\text{-}025$	0.827	0.487	20.822	0.596	(7a)
$pIC_{50} = 7.280 + 2.207(0.502)S3K$ $- 1.777(0.578)CIC1 + 2.140(0.585)BELp6$ $- 2.384(0.645)GGI3 - 1.056(0.388)C\text{-}025$	0.828	0.486	20.109	0.599	(8a)
$pIC_{50} = 5.838 + 1.606(0.627)S3K$ $+ 1.937(0.642)SEigm + 1.990(0.553)BELp6$ $- 2.596(0.704)GGI3 - 1.063(0.390)C\text{-}025$	0.826	0.487	20.717	0.587	(9a)
$pIC_{50} = 5.794 + 1.692(0.614)S3K$ $+ 1.969(0.653)SEigm + 2.018(0.567)BELv6$ $- 2.655(0.720)GGI3 - 1.080(0.389)C\text{-}025$	0.825	0.489	20.592	0.584	(10a)
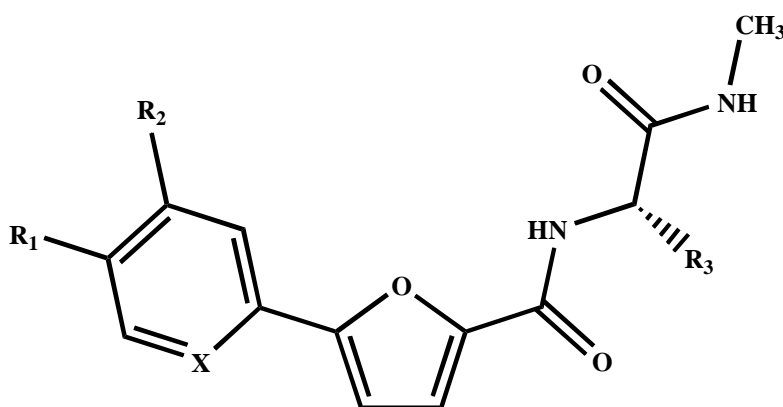
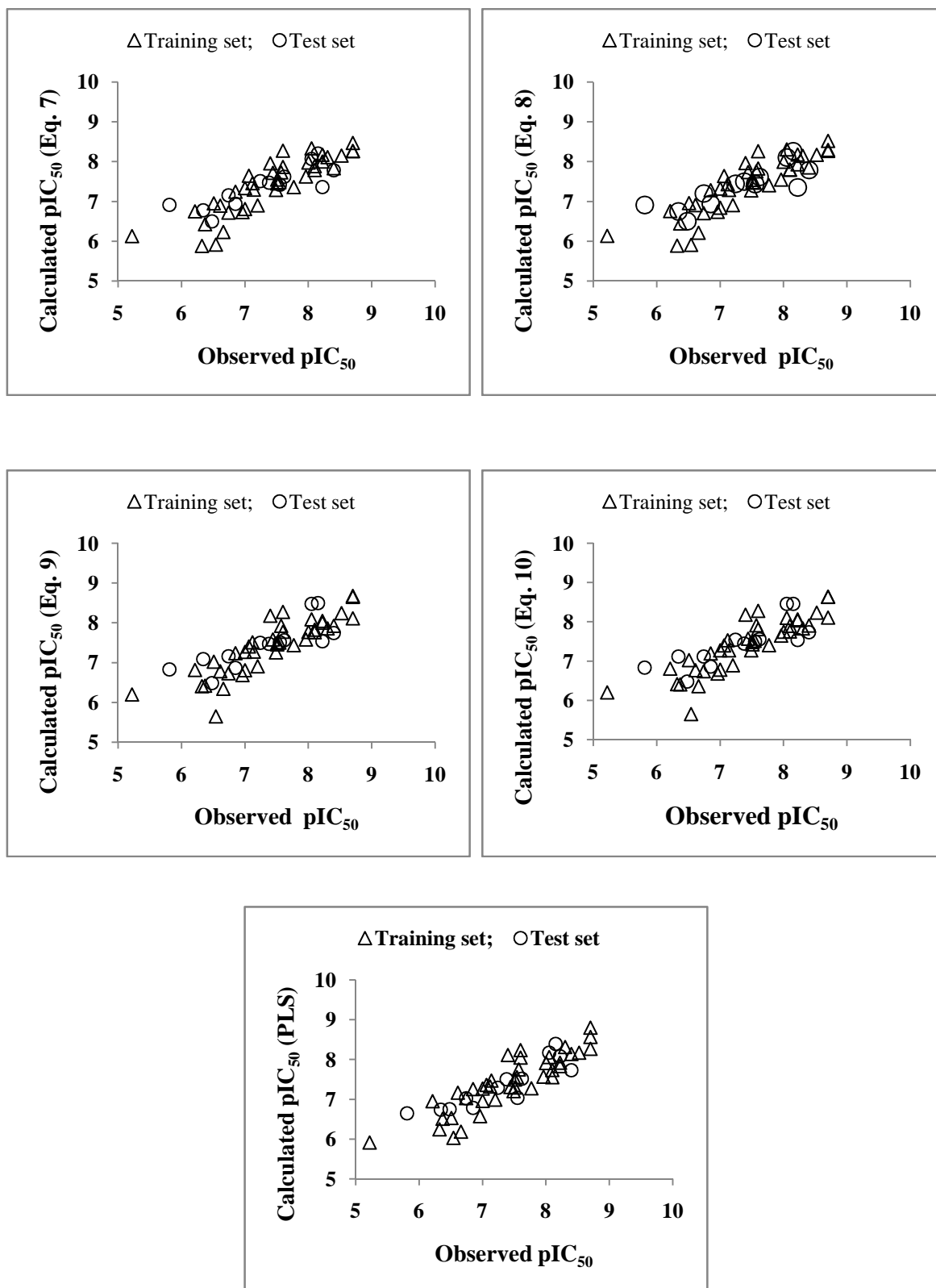**Figure 1. General structure of non-zinc-chelating compounds.**

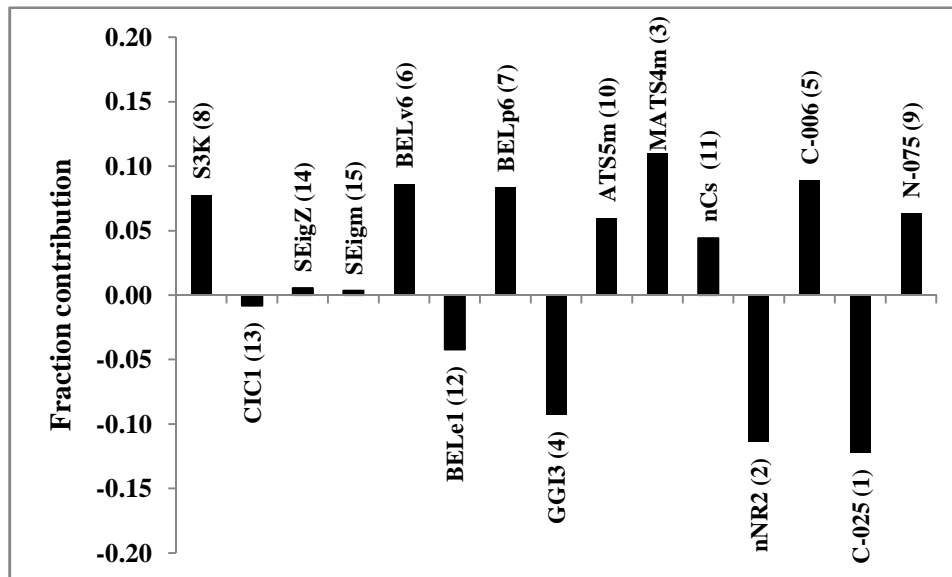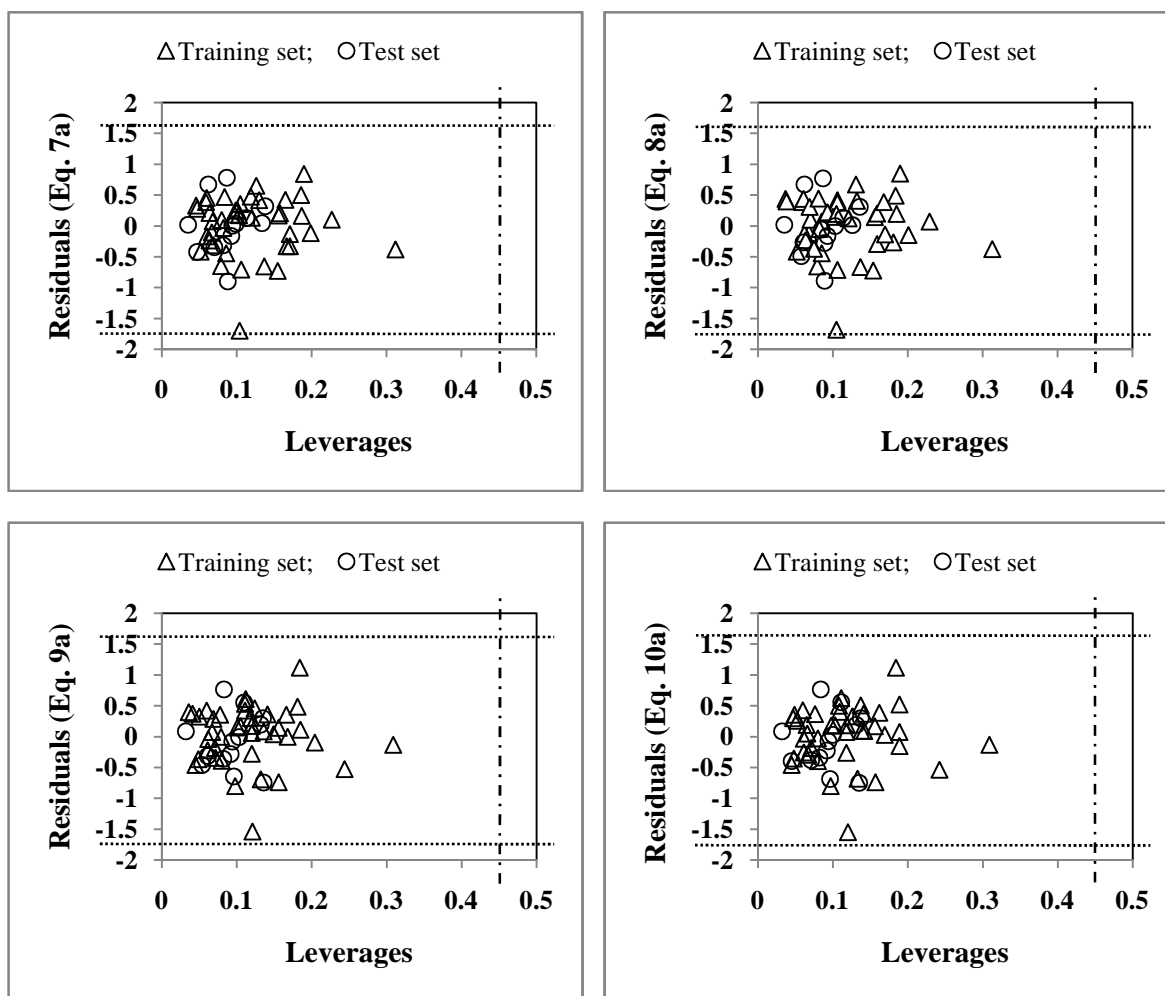**Figure 2. Plot of observed versus caculated pIC$_{50}$ values for training- and test-set compounds.**

Figure 3. Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 15 identified descriptors (Table 4) associated with MMP-13 inhibition activity of the compounds.

Figure 4. Williams plot for the training-set and test- set for inhibition activity of MMP-13 for the compounds in Table 1. The horizontal dotted line refers to the residual limit (±3×standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.439).

REFERENCES

[1].   Woessner, J. F.; Nagase, H., Editors, Matrix Metalloproteinases and TIMPs, Oxford University Press: New York, **2000**.

[2].   Cawston, T. E.; Wilson, A. J. *Best. Pract. Res. Clin. Rheumatol.* **2006**, *20*, 983-1002.

[3].   Huxley-Jones, J.; Foord, S. M.; Barnes, M. R. *Drug Discovery Today* **2008**, *13*, 685-694.

[4].   Wernicke, D.; Seyfert, C.; Hinzmann, B.; Gromnica, I. E. *J. Rheum.* **1996**, *23*, 590-595.

[5].   Knäuper, V.; Will, H.; López-Otín, C.; Smith, B.; Atkinson, S. J.; Stanton, H.; Hembry, R. M.; Murphy, G. *J. Biol. Chem.* **1996**, *271*, 17124-17131.

[6].   Takaishi, H.; Kimura, T.; Dalal, S.; Okada, Y.; D'Armiento, J. *Curr. Pharm. Biotechnol.* **2008**, *9*, 47-54.

[7].   Rowan, A. D.; Litherland, G. J.; Hui, W.; Milner, J. M. *Exp. Opin. Ther. Targets* **2008**, *12*, 1-18.

[8].   Drummond, A. H.; Beckett, P.; Brown, P. D.; Bone, E. A.; Davidson, A. H.; Galloway, W. A.; Gearing, A. J.; Huxley, P.; Laber, D.; McCourt, M.; Whittaker, M.; Wood, L. M.; Wright, A. *Ann. N. Y. Acad. Sci.* **1999**, *878*, 228-235.

[9].   Hutchinson, J. W.; Tierney, G. M.; Parsons, S. L.; Davies, T. R. C. *J. Bone Joint Surg. Br.* **1998**, *80*, 907-908.

[10].   Wojtowicz-Praga, S.; Torri, J.; Johnson, M.; Steen, V.; Marshall, J.; Ness, E.; Dickson, R.; Sale, M.; Rasmussen, H. S.; Chiodo, T. A.; Hawkins, M. J. *J. Clin. Oncol.* **1998**, *16*, 2150-2156.

[11].   Griffioen, A. W. AG-3340 (Agouron Pharmaceuticals Inc). *IDrugs* **2000**, *3*, 336-345.

[12].   Levitt, N. C.; Eskens, F. A. L. M.; O'Byrne, K. J.; Propper, D. J.; Denis, L. J.; Owen, S. J.; Choi, L.; Foekens, J. A.; Wilner, S.; Wood, J. M.; Nakajima, M.; Talbot, D. C.; Steward, W. P.; Harris, A. L.; Verweij, J. *Clin. Cancer. Res.* **2001**, *7*, 1912-1922.

[13].   Skiles, J. W.; Gonnella, N. C.; Jeng, A. Y. *Curr. Med. Chem.* **2001**, *8*, 425-474.

[14].   Clark, I. M.; Parker, A. E. *Exp. Opin Ther Targets* **2003**, *7*, 19-34.

[15].   Peterson, J. T. *Heart Fail Rev.* **2004**, *9*, 63-79.

[16].   Stickens, D.; Behonick, D. J.; Ortega, N.; Heyer, B.; Hartenstein, B.; Yu, Y.; Fosang, A. J.; Schorpp-Kistner, M.; Angel, P.; Werb, Z. *Development* **2004**, *131*, 5883-5895.

[17].   Inada, M.; Wang, Y.; Byrne, M. H.; Rahman, M. U.; Miyaura, C.; López-Otín, C.; Krane, S. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17192-17197.

[18].   Kennedy, A. M.; Inada, M.; Krane, S. M.; Christie, P. T.; Harding, B.; López-Otín, C.; Sánchez, L. M.; Pannett, A. J.; Dearlovem, A.; Hartley, C.; Byrne, M. H.; Reed, A.; Nesbit, M. A.; Whyte, M. P.; Thakker, R. V. *J. Clin. Invest.* **2005**, *115*, 2832-2842.

[19].   Johnson, A. R.; Pavlovsky, A. G.; Ortwine, D. F.; Prior, F.; Man, C. F.; Bornemeier, D. A.; Banotai, C. A.; Mueller, W. T.; McConnell, P.; Yan, C.; Baragi, V.; Lesch, C.; Roark, W. H.; Wilson, M.; Datta, K.; Guzman, R.; Han, H. K.; Dyer, R. D. *J. Biol. Chem.* **2007**, *282*, 27781-27791.

[20].   Baragi, V. M.; Becher, G.; Bendele, A. M.; Biesinger, R.; Bluhm, H.; Boer, J.; Deng, H.; Dodd, R.; Essers, M.; Feuerstein, T.; Gallagher, B. M.; Gege, C.; Hochgürtel, M.; Hofmann, M.; Jaworski, A.; Jin, L.; Kiely, A.; Korniski, B.; Kroth, H.; Nix, D.; Nolte, B.; Piecha, D.; Powers, T. S.; Richter, F.; Schneider, M.; Steeneck, C.; Sucholeiki, I.; Taveras, A.; Timmermann, A.; Veldhuizen, J. V.; Weik, J.; Wu, X.; Xia, B. *Arthr. Rheum.* **2009**, *60*, 2008-2018.

[21].   Reiter, L. A.; Freeman-Cook, K. D.; Jones, C. S.; Martinelli, G. J.; Antipas, A. S.; Berliner, M. A.; Datta, K.; Downs, J. T.; Eskra, J. D.; Forman, M. D.; Greer, E. M.; Guzman, R.; Hardink, J. R.; Janat, F.; Keene, N. F.; Laird, E. R.; Liras, J. L.; Lopresti-Morrow, L. L.; Mitchell, P. G.; Pandit, J.; Robertson, D.; Sperger, D.; Vaughn-Bowser, M. L.; Waller, D. M.; Yocum, S. A. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5822-5826.

[22].   Reiter, L. A.; Robinson, R. P.; McClure, K. F.; Jones, C. S.; Reese, M. R.; Mitchell, P. G.; Otterness, I. G.; Bliven, M. L.; Liras, J.; Cortina, S. R.; Donahue, K. M.; Eskra, J. D.; Griffiths, R. J.; Lame, M. E.; Lopez-Anaya, A.; Martinelli, G. J.; McGahee, S. M.; Yocum, S. A.; Lopresti-Morrow, L. L.; Tobiassen, L. M.; Vaughn-Bowser, M. L. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3389-3395.

[23].   Matter, H.; Schudok, M. *Curr. Opin. Drug. Disc. Dev.* **2004**, *7*, 513-535.

[24].   Skiles, J. W.; Gonnella, N. C.; Jeng, A. Y. *Curr. Med. Chem.* **2004**, *11*, 2911-2977.

[25].   Rao, B. G. *Curr. Pharm. Des.* **2005**, *11*, 295-322.

[26]. Engel, C. K.; Pirard, B.; Schimanski, S.; Kirsch, R.; Habermann, J.; Klingler, O.; Schlotte, V.; Weithmann, K. U.; Wendt, K. U. *Chem. Biol.* **2005**, *12*, 181-189.

[27]. Pirard, B. *Drug Disc. Today* **2007**, *12*, 640-646.

[28]. Heim-Riether, A.; Taylor, S. J.; Liang, S.; Gao, D. A.; Xiong, Z.; August, E. M.; Collins, B. K.; Farmer, B. T., II; Haverty, K.; Hill-Drzewi, M.; Junker, H. -D.; Margarit, S. M.; Moss, N.; Neumann, T.; Proudfoot, J. R.; Smith-Keenan, L.; Sekul, R.; Zhang, Q.; Li, J.; Farrow, N. A. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5321-5324.

[29]. Gao, D. A.; Xiong, Z.; Heim-Riether, A.; Amodeo, L.; August, E. M.; Cao, X.; Ciccarelli, L.; Collins, B. K.; Harrington, K.; Haverty, K.; Hill-Drzewi, M.; Li, X.; Liang, S.; Margarit, S. M.; Moss, N.; Nagaraja, N.; Proudfoot, J.; Roman, R.; Schlyer, S.; Keenan, L. S.; Taylor, S.; Wellenzohn, B.; Wiedenmayer, D.; Li, J.; Farraw, N. A. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 5039-5043.

[30]. SYSTAT, Version 7.0; SPSS Inc 444 North Michigan Avenue, Chicago IL, 60611.

[31]. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. http://www.cambridgesoft.com

[32]. Dragon software (version 1.11-2001) by Todeschini R.; Consonni V. Milano, Italy. http//www.talete.mi.it/dragon.htm

[33]. Prabhakar, Y. S. *QSAR Comb. Sci.* **2003**, *22*, 583-595.

[34]. Wold, S. *Technometrics* **1978**, 20, 397-405.

[35]. Kettaneh, N.; Berglund, A.; Wold, S. *Comput. Stat. Data Anal.* **2005**, 48, 69-85.

[36]. Stahle, L.; Wold, S. Biomedical Research, Progress in Medicinal Chemistry, vol. 25, Elsevier Science Publishers, BV, 1998.

[37]. Sharma, S.; Prabhakar, Y. S.; Singh, P.; Sharma, B. K. Eur J Med Chem **2008**, *43*, 2354-2360.

[38]. Sharma, S.; Sharma, B. K.; Sharma, S. K.; Singh, P.; Prabhakar, Y. S. *Eur. J. Med. Chem.* **2009**, *44*, 1377-1382.

[39]. Sharma, B. K.; Pilania, P.; Singh, P.; Prabhakar, Y. S. *SAR QSAR Environ. Res.* **2010**, *21*, 169-185.

[40]. Sharma, B. K.; Singh, P.; Sarbhai, K.; Prabhakar, Y. S. *SAR QSAR Environ. Res.* **2010**, *21*, 369-388.

[41]. Sharma, B. K.; Pilania, P.; Sarbhai, K.; Singh, P.; Prabhakar, Y. S. *Mol. Divers.* **2010**, *14*, 371-384.

[42]. Sharma, B. K.; Singh, P.; Shekhawat, M.; Sarbhai, K.; Prabhakar, Y. S. *SAR QSAR Environ. Res.* **2011**, *22*, 365-383

[43]. So, S. –S.; Karplus, M. *J. Med. Chem.* **1997**, *40*, 4347-4359.

[44]. Prabhakar, Y. S.; Solomon, V. R.; Rawal, R. K.; Gupta, M. K.; Katti, S. B. *QSAR Comb. Sci.* **2004**, *23*, 234-244.

[45]. Gramatica, P. *QSAR Comb. Sci.* **2007**, *26*, 694-701.

[46]. Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. *Env. Health Persp.* **2003**, *111*, 1361-1375.

[47]. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Model.* **2002**, *20*, 269-76.