



(RESEARCH ARTICLE)



Quantitative structure-activity relationship study on the MMP-13 inhibitory activity of fused pyrimidine derivatives possessing a 1,2,4-Triazol-3-yl group as a ZBG

Afsar Jahan ¹, Brij Kishore Sharma ^{1,*} and Vishnu Dutt Sharma ²

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

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

GSC Biological and Pharmaceutical Sciences, 2021, 16(01), 251–265

Publication history: Received on 09 June 2021; revised on 13 July 2021; accepted on 15 July 2021

Article DOI: <https://doi.org/10.30574/gscbps.2021.16.1.0199>

Abstract

QSAR study has been carried out on the MMP-13 inhibitory activity of fused pyrimidine derivatives possessing a 1,2,4-triazol-3-yl group as a ZBG in 0D- to 2D-Dragon descriptors. The derived QSAR models have revealed that the number of Sulfur atoms (descriptor nS), Balaban mean square distance index (descriptor MSD), molecular electrotopological variation (descriptor DELS), structural information content index of neighborhood symmetry of 2nd and 3rd order (descriptors SIC2 and SIC3), average valence connectivity index chi-4 (descriptor X4Av) in addition to 1st order Galvez topological charge index (descriptor JG11) and global topological charge index (descriptor JGT) played a pivotal role in rationalization of MMP-13 inhibition activity of titled compounds. Atomic properties such as mass and volume in terms of atomic properties weighted descriptors MATS5m and MATS3v, and certain atom centred fragments such as CH2RX (descriptor C-006), X--CX--X (descriptor C-044), H attached to heteroatom (descriptor H-050) and H attached to C0(sp3) with 1X attached to next C (descriptor H-052) are also predominant to explain MMP-13 inhibition actions of fused pyrimidines.

PLS analysis has also corroborated the dominance of CP-MLR identified descriptors. Applicability domain analysis revealed that the suggested model matches the high-quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Keywords: QSAR; MMP-13 inhibitory activity; Combinatorial protocol in multiple linear regression (CP-MLR) analysis; PLS analysis; Dragon descriptors; Fused pyrimidines; Zinc binding group (ZBG)

1. Introduction

More than 30 million patients worldwide are affected from osteoarthritis (OA) which is the most common forms of arthritis [1]. The pain and reduced mobility in affected joints due to progressive cartilage damage is the principal morphological characteristic of OA. The lessening of pain and inhibition of inflammation is the only recommended pharmaceutical therapies [2-5] which include oral treatment with acetaminophen, nonsteroidal anti-inflammatory drugs (NSAIDs), or cyclooxygenase-2 (COX-2) inhibitors and intra-articular injections of hyaluronic acid or corticosteroids. The withdrawal of some COX-2 selective inhibitors (rofecoxib and valdecoxib) [6,7] rendered an unmet medical need for safe oral disease-modifying osteoarthritis drugs to prevent, slow down, or reverse any advanced cartilage destruction [8,9].

The matrix metalloproteinases (MMPs) are structurally related zinc-dependent endopeptidases which degrade varied extracellular matrix. Among MMPs the hydrolysis of type II collagen, the main structural component of the cartilage

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

matrix, is specifically catalyzed by MMP-13 (collagenase-3) [10,11] and the protein is resistant to most proteases. MMP-13 has high levels of expression at OA chondrocytes as compared to normal chondrocytes [12]. MMP-13 plays a crucial role in the destruction of articular cartilage in OA may be evinced from the findings that in genetically modified mice the regulated expression of human MMP-13 in joint cartilages induced OA [13] and the preferential inhibition of MMP-13 blocked the degradation of explanted human OA cartilage [14]. The inhibition of destruction of cartilage by MMP inhibitors had shown in some animal models of OA in preclinical testing [15]. The concerns of dose limiting toxicity such as skin rash and musculoskeletal side effects (MSS) characterized by joint stiffness and pain has discontinued the most clinical trials of broad-spectrum MMP inhibitors. The reason behind these side effects is unclear pharmacologically [16-18] however, it is hypothesized that a nonselective inhibition of other metalloproteinases or a combined inhibition of a series of critical MMPs is the cause of MSS. To avoid undesirable side effects interest has been raised toward potent MMP-13 inhibitors having a high degree of selectivity over other MMPs [19-29]. Most of the MMP inhibitors generally have a P1' fragment that may be lodged in the S1' subsite of the enzyme active site and a functional group capable of binding the catalytic zinc ion. The modulation of potency and selectivity against various MMPs is allowed in a number of P1' fragments but there are only a small identified set of zinc binding groups (ZBGs) employed in the design of selective MMP inhibitors [27,30-32].

The successful cocrystallization of quinazoline-2-carboxamide [23,25,33] and triazole in MMP-13 catalytic domain in high-throughput screening for MMP-13 inhibition allowed the designing of a hybrid molecule which combined the quinazoline with the triazole ZBG. Fused pyrimidine system has been applied successfully by several groups to obtain inhibitory potency and selectivity for MMP-13 inhibition [23,25,33-37]. A series of novel fused pyrimidine derivatives possessing a 1,2,4-triazol-3-yl group as a ZBG with potent MMP-13 inhibitory activities has been synthesized by Nara *et al.* [38]. The aim of present communication is to establish the quantitative relationships between the reported activities and molecular descriptors unfolding the substitutional changes in titled compounds.

2. Material and methods

2.1. Biological actions and theoretical molecular descriptors

The reported twenty-nine fused pyrimidine derivatives possessing a 1,2,4-triazol-3-yl group as a ZBG are considered as the data set for present study [38]. These derivatives were evaluated for their MMP-13 inhibitory activities and were reported as IC_{50} . The reported MMP-13 activity on molar basis (as pIC_{50}) along with the structures of these analogues is shown in Table 1. The data set was sub-divided into training set to develop models and test set to validate the models externally. The test set compounds which were selected using an in-house written randomization program, are also mentioned in Table 1.

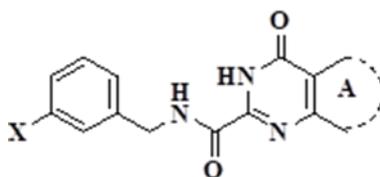
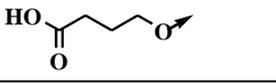
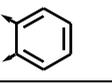
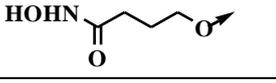
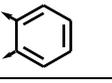
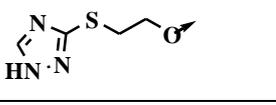
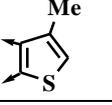
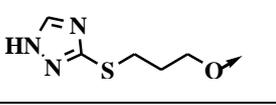
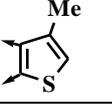
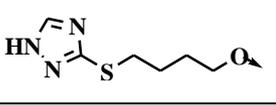
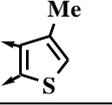
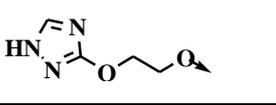
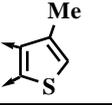
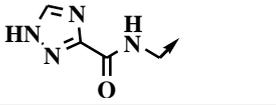
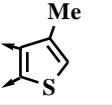
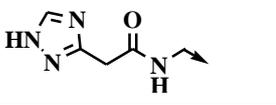
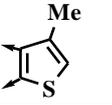
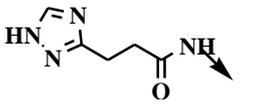
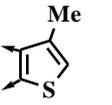
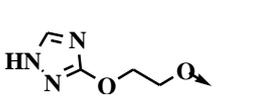
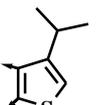
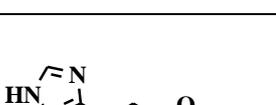
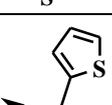
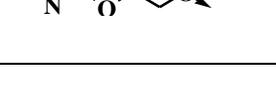
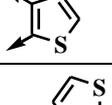
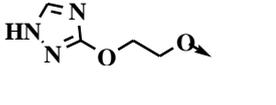
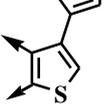
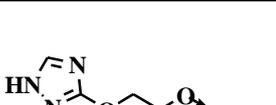
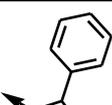
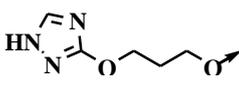
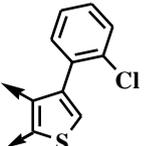
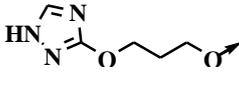
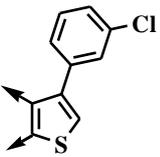


Table 1 Structures, observed and calculated MMP-13 inhibitory activities of fused pyrimidine derivatives

| Cpd. | X | | pIC_{50}^a | | | | | |
|----------------|---|--|-------------------|------------|---------|---------|---------|------|
| | | | Obs. ^b | Calculated | | | | |
| | | | | Eq. (6) | Eq. (7) | Eq. (8) | Eq. (9) | PLS |
| 1 | | | 7.92 | 8.40 | 8.37 | 8.32 | 8.57 | 7.89 |
| 2 ^c | | | 9.70 | 9.63 | 9.52 | 9.55 | 10.20 | 9.54 |
| 3 | | | 7.54 | 7.38 | 7.37 | 7.48 | 7.15 | 7.42 |

| | | | | | | | | |
|-----------------|---|---|-------|-------|-------|-------|-------|-------|
| 4 |  |  | 8.03 | 8.22 | 8.32 | 8.53 | 8.14 | 8.21 |
| 5 ^c |  |  | 8.80 | 8.17 | 8.26 | 8.60 | 8.47 | 8.54 |
| 6 ^c |  |  | 10.13 | 10.17 | 10.11 | 10.16 | 10.47 | 9.93 |
| 7 ^c |  |  | 10.48 | 10.42 | 10.59 | 10.40 | 10.66 | 9.89 |
| 8 |  |  | 8.59 | 8.70 | 8.66 | 8.42 | 9.02 | 8.99 |
| 9 |  |  | 10.19 | 9.64 | 9.72 | 10.13 | 10.19 | 9.63 |
| 10 |  |  | 8.82 | 8.72 | 8.60 | 8.37 | 9.01 | 9.16 |
| 11 ^c |  |  | 8.31 | 8.92 | 8.93 | 9.08 | 9.17 | 9.19 |
| 12 |  |  | 9.54 | 9.36 | 9.47 | 9.86 | 9.44 | 9.38 |
| 13 |  |  | 10.04 | 9.94 | 10.26 | 9.88 | 9.58 | 9.65 |
| 14 |  |  | 10.57 | 10.96 | 10.79 | 10.57 | 10.76 | 10.48 |
| 15 ^c |  |  | 10.37 | 10.77 | 10.38 | 10.33 | 10.72 | 10.44 |
| 16 |  |  | 10.47 | 10.17 | 10.18 | 9.74 | 10.15 | 9.96 |
| 17 |  |  | 10.62 | 10.21 | 10.04 | 10.22 | 10.29 | 10.47 |

| | | | | | | | | |
|-----------------|--|--|-------|-------|-------|-------|-------|-------|
| 18 | | | 10.44 | 10.17 | 10.08 | 10.30 | 9.97 | 10.33 |
| 19 ^c | | | 10.03 | 10.17 | 10.08 | 10.21 | 9.90 | 10.17 |
| 20 | | | 10.27 | 10.40 | 10.55 | 10.42 | 10.29 | 10.33 |
| 21 | | | 10.24 | 10.07 | 10.14 | 10.20 | 9.97 | 10.17 |
| 22 | | | 9.36 | 10.07 | 10.05 | 10.11 | 9.90 | 9.99 |
| 23 | | | 9.54 | 10.20 | 10.12 | 10.11 | 10.31 | 10.35 |
| 24 | | | 10.59 | 10.11 | 10.01 | 10.04 | 9.93 | 10.14 |
| 25 | | | 10.00 | 10.11 | 10.01 | 9.92 | 9.83 | 9.97 |
| 26 | | | 10.21 | 10.19 | 10.11 | 10.24 | 10.44 | 10.42 |
| 27 ^c | | | 9.85 | 10.14 | 10.14 | 10.31 | 10.14 | 10.31 |

| | | | | | | | | |
|-----------------|---|---|-------|-------|-------|-------|-------|-------|
| 28 ^c |  |  | 10.23 | 10.38 | 10.52 | 10.44 | 10.44 | 10.28 |
| 29 |  |  | 10.11 | 10.07 | 10.24 | 10.24 | 10.14 | 10.14 |

^aIC₅₀ on molar basis; ^bTaken from reference [38]; ^cCompound included in test set.

The structures of the all the compounds (listed in Table 1) were drawn in 2D ChemDraw [39] and subjected to energy minimization in the MOPAC using the AM1 procedure for closed shell system after converting these into 3D modules. The energy minimization was carried out to attain a well-defined conformer relationship among the congeners under study. The 0D- to 2D-molecular descriptors of titled compounds was computed using DRAGON software [40]. This software offers a large number of descriptors corresponding to ten different classes of 0D- to 2D-descriptor modules. The different descriptor classes include the constitutional, topological, molecular walk counts, BCUT descriptors, Galvez topological charge indices, 2D-autocorrelations, functional groups, atom-centered fragments, empirical descriptors and the properties describing descriptors. These descriptors offer characteristic structural information specific to the descriptor class. The definition and scope of these descriptor's classes is given in Table 2.

Table 2 Descriptor classes used for the modeling of MMP-13 inhibitory activity of fused pyrimidine derivatives [40]

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

A total number of 492 descriptors, belonging to 0D- to 2D- modules, have been computed to obtain most appropriate models describing the biological activity. Prior to model development procedure, all those descriptors that are inter-

correlated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. This procedure has reduced the total descriptors from 492 to 93 as relevant ones to explain the biological actions of titled compounds.

2.2. Development and validation of model

The combinatorial protocol in multiple linear regression (CP-MLR) [41-45] and partial least squares (PLS) [46-48] procedures were used in the present work for developing QSAR models. The CP-MLR is a “filter”-based variable selection procedure, which employs a combinatorial strategy with MLR to result in selected subset regressions for the extraction of diverse structure–activity models, each having unique combination of descriptors from the generated dataset of the compounds under study. The embedded filters make the variable selection process efficient and lead to unique solution. Fear of “chance correlations” exists where large descriptor pools are used in multilinear QSAR/QSPR studies [49,50]. In view of this, to find out any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been subjected to randomization test [51,52] by repeated randomization (100 simulation runs) of the biological responses. The datasets with randomized response vector have been reassessed by multiple regression analysis. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

Validation of the derived model is necessary to test its prediction and generalization within the study domain. For each model, derived by involving n data points, a number of statistical parameters such as r (the multiple correlation coefficient), s (the standard deviation), F (the F ratio between the variances of calculated and observed activities), and Q^2_{LOO} (the cross-validated index from leave-one-out procedure) have been obtained to assess its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5. Additional statistical parameters such as, the Akaike’s information criterion, AIC [53,54], the Kubinyi function, FIT [55,56] and the Friedman’s lack of fit, LOF [57], have also been calculated to further validate the derived models. The AIC takes into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. The FIT, closely related to the F -value, proved to be a useful parameter for assessing the quality of the models. A model which is derived in k independent descriptors, its F -value will be more sensitive if k is small while it becomes less sensitive if k is large. The FIT, on the other hand, will be less sensitive if k is small whereas it becomes more sensitive if k is large. The model that produces the lowest AIC value and highest FIT value is considered potentially the most useful and the best. The LOF factor takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large number of parameters.

2.3. Applicability domain

The usefulness of a model is based on its accurate prediction ability for new congeners. A model is valid only within its training domain and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain (AD) is evaluated by the leverage values for each compound [58]. A Williams plot (the plot of standardized residuals versus leverage values (h)) is constructed, which can be used for a simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in the model. In this plot, the AD is established inside a squared area within $\pm x$ standard deviations and a leverage threshold h^* , which is generally fixed at $3(k + 1)/n$ (n is the number of training set compounds and k is the number of model parameters), whereas $x = 2$ or 3 . If the compounds have a high leverage value ($h > h^*$), then the prediction is not trustworthy. On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training set compounds.

3. Results and discussion

3.1. QSAR results

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modeling study, 9 compounds have been included in the test set for the validation of the models derived from 20 training set compounds. A total number of 93 significant descriptors from 0D- to 2D- classes have been subjected to CP-MLR analysis with default “filters” set in it. Statistical models in one, two and three descriptors have been derived to achieve the best relationship correlating MMP-

13 inhibitory activity. Only one model in one descriptor, nine models in two descriptors and eight models in three descriptors, having $r^2_{\text{Test}} > 0.5$, were obtained through CP-MLR. The three parameter models have shared 14 descriptors among them. All these 14 descriptors along with their brief meaning, average regression coefficients, and total incidence are listed in Table 3, which will serve as a measure of their estimate across these models.

Table 3 Identified descriptors along with their class, physical meaning, average regression coefficient and incidence

| Descriptor class, average regression coefficient and (incidence) | |
|--|--|
| Constitutional descriptors (CONST) | nS (number of Sulfur atoms), 1.670(2) |
| Topological descriptors (TOPO) | MSD (Balaban mean square distance index), -1.153(1); DELS (molecular electrotopological variation), 1.190(1); X4Av (average valence connectivity index chi-4), 1898(1); SIC2 (structural information content index, neighborhood symmetry of 2-order), 1.568(2); SIC3 (structural information content index, neighborhood symmetry of 3-order), 0.706(1) |
| Galvez topological charge indices (GALVEZ) | JGI1 (mean topological charge index of order 1), -1.566(1); JGT (global topological charge index), -1.913(1) |
| 2D autocorrelations (2D-AUTO) | MATS5m (Moran autocorrelation of lag-5/ weighted by atomic masses), -1.476(1); MATS3v (Moran autocorrelation of lag-3/ weighted by atomic van der Waals volumes), 1.961(3) |
| Atom-centred fragments (ACF) | C-006 (CH2RX), 1.669(1); C-044 (X--CX..X), 1.546(1); H-050 (H attached to heteroatom), 1.238(1); H-052 (H attached to C0(sp3) with 1X attached to next C), -2.431(7) |

^aThe descriptors are identified from the three parameter models for activity emerged from CP-MLR protocol with filter-1 as 0.30, filter-2 as 2.0, filter-3 as 0.872 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 20 compounds. ^bThe 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.

The alone one parameter and selected two parameter models obtained through CP-MLR are mentioned below:

$$\begin{aligned} \text{pIC}_{50} &= 8.370 + 1.712(0.316) \text{ C-044} \\ n &= 20, r = 0.786, s = 0.613, F = 29.196, Q^2_{\text{L00}} = 0.502, Q^2_{\text{L50}} = 0.605 \\ r^2_{\text{Test}} &= 0.853, \text{FIT} = 1.390, \text{LOF} = 0.418, \text{AIC} = 0.460 \end{aligned} \quad (1)$$

$$\begin{aligned} \text{pIC}_{50} &= 8.516 + 1.657(0.330) \text{ C-006} - 2.459(0.379) \text{ H-052} \\ n &= 20, r = 0.887, s = 0.471, F = 31.421, Q^2_{\text{L00}} = 0.601, Q^2_{\text{L50}} = 0.543 \\ r^2_{\text{Test}} &= 0.608, \text{FIT} = 2.618, \text{LOF} = 0.295, \text{AIC} = 0.301 \end{aligned} \quad (2)$$

$$\begin{aligned} \text{pIC}_{50} &= 8.469 + 2.058(0.426) \text{ IC3} - 2.228(0.387) \text{ H-052} \\ n &= 20, r = 0.881, s = 0.482, F = 29.690.421, Q^2_{\text{L00}} = 0.694, Q^2_{\text{L50}} = 0.558 \\ r^2_{\text{Test}} &= 0.534, \text{FIT} = 2.474, \text{LOF} = 0.309, \text{AIC} = 0.314 \end{aligned} \quad (3)$$

$$\begin{aligned} \text{pIC}_{50} &= 7.429 + 1.724(0.585) \text{ PW2} + 1.686(0.265) \text{ C-044} \\ n &= 20, r = 0.864, s = 0.513, F = 25.170, Q^2_{\text{L00}} = 0.653, Q^2_{\text{L50}} = 0.638 \\ r^2_{\text{Test}} &= 0.607, \text{FIT} = 2.097, \text{LOF} = 0.350, \text{AIC} = 0.357 \end{aligned} \quad (4)$$

$$\begin{aligned} \text{pIC}_{50} &= 7.645 + 1.712(0.270) \text{ C-044} + 1.450(0.524) \text{ H-050} \\ n &= 20, r = 0.858, s = 0.524, F = 23.801, Q^2_{\text{L00}} = 0.651, Q^2_{\text{L50}} = 0.663 \\ r^2_{\text{Test}} &= 0.738, \text{FIT} = 1.983, \text{LOF} = 0.365, \text{AIC} = 0.372 \end{aligned} \quad (5)$$

where n , r , s and F represent respectively the number of data points, the multiple correlation coefficient, the standard deviation and the F -ratio between the variances of calculated and observed activities. In above regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models. The positive regression coefficient associated to a descriptor will augment the activity profile of a compound while the negative coefficient will cause

detrimental effect to it. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation.

The descriptors IC3 and PW2 participated in above models are the topological descriptors representing information content index of 3rd order neighborhood symmetry and path/walk 2 Randic shape index, respectively. The positive influence of descriptors IC3 and PW2 on the activity suggested that higher values of descriptor IC3 and PW2 would be beneficiary to the activity. The other participated descriptor C-006, C-044, H-050 and H-052 belong to atom centered fragment class. The descriptors C-006, C-044 and H-050 shown positive and descriptor H-052 showed negative contribution to the activity. Thus, presence of CH2RX (descriptor C-006), X--CX.X (descriptor C-044) and H attached to heteroatom (descriptor H-050) in addition to absence of H attached to C0(sp3) with 1X attached to next C (descriptor H-052) type structural fragments in a molecular structure would be favorable to the activity.

This model in two descriptors could account for nearly 79% variance in the observed activities. Considering the number of observations models up to three descriptors have been explored and a total number of 8 models having test set r^2 greater than 0.5 were obtained. The representative models in three descriptors are presented below:

$$\begin{aligned} \text{pIC}_{50} &= 7.914 + 1.605(0.352) \text{ nS} + 1.808(0.378) \text{ MATS3v} - 2.345(0.319) \text{ H-052} \\ n &= 20, r = 0.931, s = 0.384, F = 34.779, Q^2_{\text{L00}} = 0.791, Q^2_{\text{L50}} = 0.757 \\ r^2_{\text{Test}} &= 0.763, \text{FIT} = 3.597, \text{LOF} = 0.241, \text{AIC} = 0.221 \quad (6) \end{aligned}$$

$$\begin{aligned} \text{pIC}_{50} &= 7.787 + 1.898(0.438) \text{ X4Av} + 2.183(0.391) \text{ MATS3v} - 2.874(0.327) \text{ H-052} \\ n &= 20, r = 0.926, s = 0.395, F = 32.555, Q^2_{\text{L00}} = 0.782, Q^2_{\text{L50}} = 0.709 \\ r^2_{\text{Test}} &= 0.800, \text{FIT} = 3.367, \text{LOF} = 0.255, \text{AIC} = 0.234 \quad (7) \end{aligned}$$

$$\begin{aligned} \text{pIC}_{50} &= 7.814 + 1.664(0.417) \text{ SIC2} + 1.894(0.405) \text{ MATS3v} - 2.452(0.339) \text{ H-052} \\ n &= 20, r = 0.920, s = 0.412, F = 29.455, Q^2_{\text{L00}} = 0.744, Q^2_{\text{L50}} = 0.674 \\ r^2_{\text{Test}} &= 0.785, \text{FIT} = 3.047, \text{LOF} = 0.278, \text{AIC} = 0.255 \quad (8) \end{aligned}$$

$$\begin{aligned} \text{pIC}_{50} &= 9.798 + 1.472(0.439) \text{ SIC2} - 1.566(0.361) \text{ JGI1} - 2.046(0.348) \text{ H-052} \\ n &= 20, r = 0.912, s = 0.430, F = 26.640, Q^2_{\text{L00}} = 0.659, Q^2_{\text{L50}} = 0.692 \\ r^2_{\text{Test}} &= 0.658, \text{FIT} = 2.755, \text{LOF} = 0.302, \text{AIC} = 0.278 \quad (9) \end{aligned}$$

Table 4 PLS and MLR-like PLS models from the 14 descriptors of four parameter CP-MLR models for MMP-13 inhibitory activities

| A: PLS equation | | | | |
|--------------------------|------------|-------------------------------------|--------------------------------------|-------|
| PLS components | | PLS coefficient (s.e.) ^a | | |
| Component-1 | | -0.394(0.040) | | |
| Component-2 | | 0.176(0.050) | | |
| Constant | | 9.654 | | |
| B: MLR-like PLS equation | | | | |
| S. No. | Descriptor | MLR-like coefficient ^b | (Fraction contribution) ^c | Order |
| 1 | nS | 0.130 | 0.083 | 6 |
| 2 | MSD | -0.210 | -0.134 | 2 |
| 3 | DELS | 0.135 | 0.086 | 5 |
| 4 | X4Av | 0.017 | 0.011 | 14 |
| 5 | SIC2 | 0.124 | 0.079 | 7 |
| 6 | SIC3 | 0.031 | 0.020 | 12 |
| 7 | JGI1 | -0.098 | -0.062 | 8 |
| 8 | JGT | -0.024 | -0.015 | 13 |

| | | | | |
|--------------------------------|--------|--------|--------|----|
| 9 | MATS5m | -0.033 | -0.021 | 11 |
| 10 | MATS3v | 0.038 | 0.025 | 10 |
| 11 | C-006 | 0.087 | 0.055 | 9 |
| 12 | C-044 | 0.204 | 0.131 | 3 |
| 13 | H-050 | 0.138 | 0.088 | 4 |
| 14 | H-052 | -0.296 | -0.190 | 1 |
| Constant | | | 8.406 | |
| C: PLS regression statistics | | | Values | |
| n | | | 20 | |
| r | | | 0.929 | |
| s | | | 0.377 | |
| F | | | 53.878 | |
| FIT | | | 4.489 | |
| LOF | | | 0.189 | |
| AIC | | | 0.192 | |
| Q ² _{L00} | | | 0.823 | |
| Q ² _{L50} | | | 0.657 | |
| r ² _{Test} | | | 0.668 | |

^aRegression coefficient of PLS factor and its standard error. ^bCoefficients of MLR-like PLS equation in terms of descriptors for their original values; ^cf.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit s.d.) data.

These models have accounted for nearly 87% variance in the observed activities. The values greater than 0.5 of Q² index is in accordance to a reasonable robust QSAR model. The pIC₅₀ values of training set compounds calculated using Eqs. (6) to (9) have been included in Table 1. The models (6) to (9) are validated with an external test set of 9 compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r² (r²_{Test}) values and the same is reported in Table 1. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

It is evident from the signs of the regression coefficients that the newly appeared constitutional class descriptor nS, topological class descriptors X4Av and SIC2, 2D autocorrelation descriptor MATS3v contributed positively to the activity. Thus, higher number of Sulfur atoms (descriptor nS) in a molecule and higher values of average valence connectivity index chi-4 (descriptor X4Av), structural information content index of neighborhood symmetry of 2-order (descriptor SIC2) and Moran autocorrelation of lag-3/ weighted by atomic van der Waals volumes (descriptor MATS3v) would be beneficiary to the activity. On the other hand, Galvez charge index JGI1 contributed negatively to the activity suggesting that a lower value of mean topological charge index of order 1 would be favorable to the activity.

A partial least square (PLS) analysis has been carried out on these 14 CP-MLR identified descriptors (Table 3) to facilitate the development of a “single window” structure–activity model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, two components are found to be the optimum for these 14 descriptors and they explained 86.30% variance in the activity. The MLR-like PLS coefficients of these 14 descriptors are given in Table 4.

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

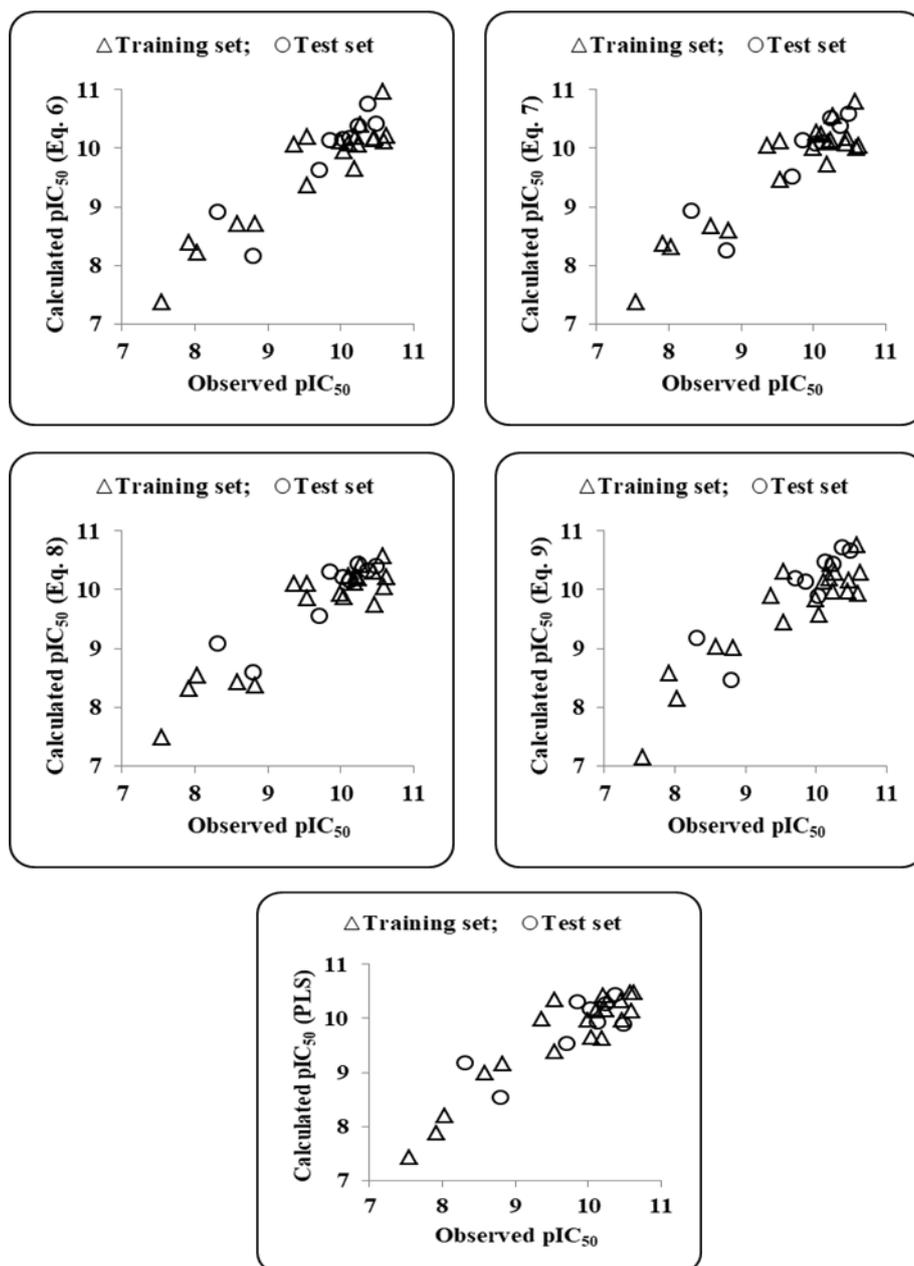


Figure 1 Plot of observed and calculated pIC_{50} values of training- and test-set compounds for MMP-13 inhibition

The PLS analysis has suggested H-052 as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 14 in Table 4; Figure 2). The other descriptors in decreasing order of significance are MSD, C-044, H-050, DELS, nS, SIC2, JG11, C-006, MATS3v, MATS5m, SIC3, JGT and X4Av. Except descriptors MSD, DELS, SIC3, MATS5m and JGT all these descriptors are part of Eqs. (1) to (9) and convey same inference in the PLS model as well. It is inferred from the PLS analysis that a higher values of descriptors DELS (molecular electrotopological variation) and SIC3 (structural information content index of neighborhood symmetry of 3-order) and lower values of descriptors MSD (Balaban mean square distance index), MATS5m (Moran autocorrelation of lag-5/ weighted by atomic masses) and JGT ((global topological charge index) would be advantageous to the activity. It is also observed that PLS model from the dataset devoid of CP-MLR identified 14 descriptors (Table 3) is inferior in explaining the activity of the analogues.

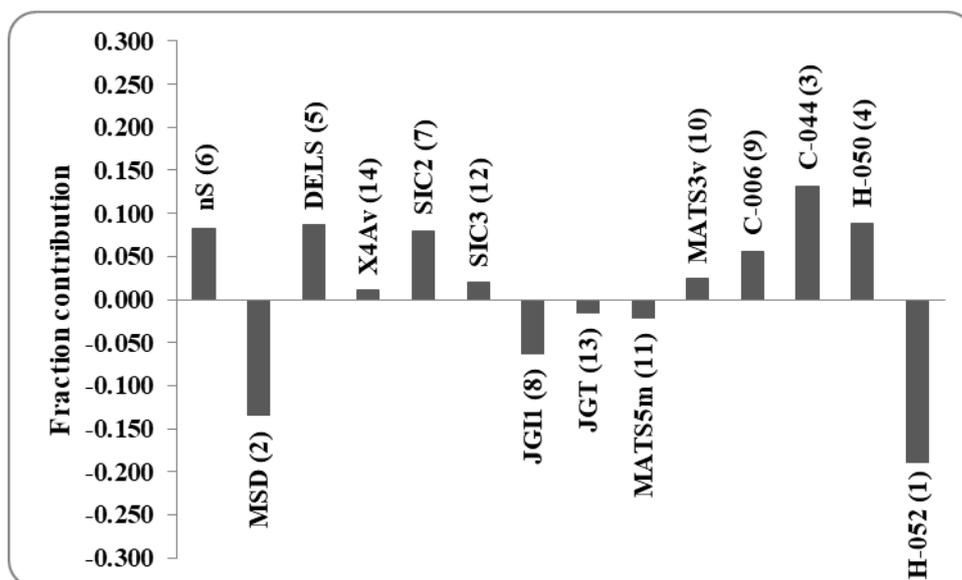


Figure 2 Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 14 CP-MLR identified descriptors (Table 3) associated with MMP-13 inhibitory activity of fused pyrimidine derivatives

3.2. Applicability domain (AD)

On analyzing the model AD in the Williams plot, shown in Figure 3, of the model based on the whole dataset (Table 5), it has appeared that none of the compound was identified as an obvious outlier for the MMP-13 inhibitory activity if the limit of normal values for the Y outliers (response outliers) was set as 2.5 times of standard deviation units.

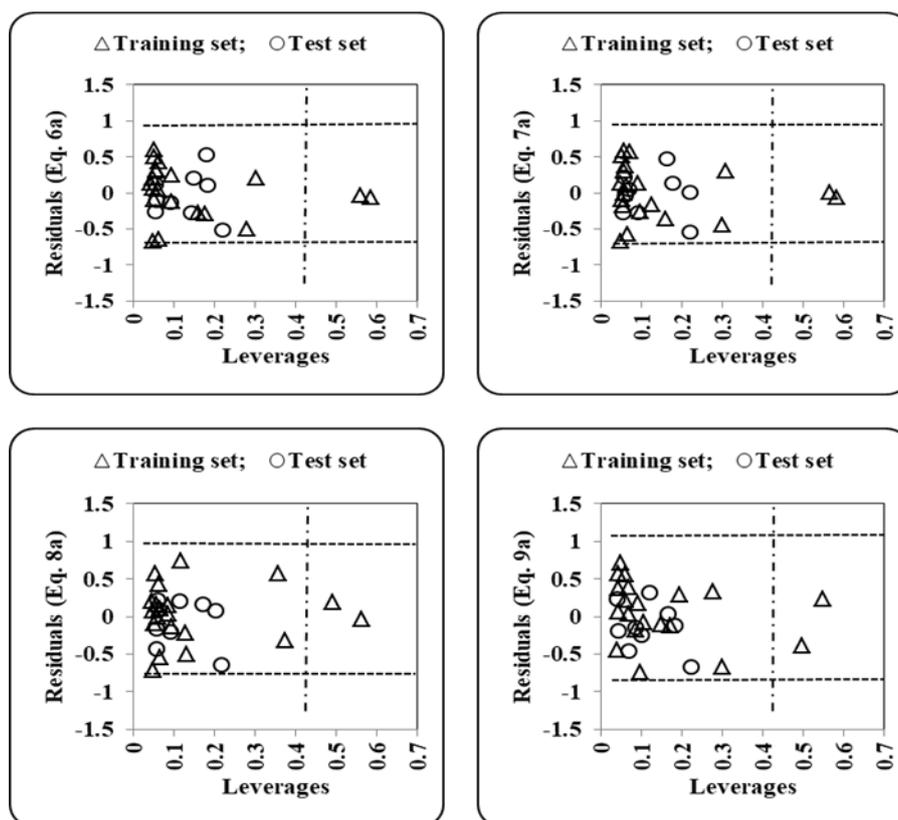


Figure 3 Williams plot for the training-set and test-set compounds for MMP-13 inhibitory activity. The horizontal dotted line refers to the residual limit ($\pm 2.5 \times$ standard deviation) and the vertical dotted line represents threshold leverage h^* ($= 0.413$)

An outlier to a QSAR is identified normally by having a large standard residual activity and can indicate the limits of applicability of QSAR models. Two compounds listed in Table 1 at S. Nos. 3 and 8 were found to have leverage (h) values greater than the threshold leverage ($h^*=0.413$) suggesting them as chemically influential compounds. For both the training-set and test-set, the suggested model matches the high-quality parameters with good fitting power and the capability of assessing external data. Furthermore, all of the compounds were within the applicability domain of the proposed model and were evaluated correctly.

Table 5 Models derived for the whole data set (n = 29) in descriptors identified through CP-MLR

| Model | r | s | F | Q ² _{loo} | Eq. |
|---|-------|-------|--------|-------------------------------|------|
| $pIC_{50} = 7.909 + 1.423(0.250) nS + 1.897(0.287) MATS3v - 2.228(0.289) H-052$ | 0.923 | 0.360 | 48.608 | 0.809 | (6a) |
| $pIC_{50} = 7.771 + 1.747(0.311) X4Av + 2.246(0.298) MATS3v - 2.729(0.280) H-052$ | 0.922 | 0.362 | 47.924 | 0.807 | (7a) |
| $pIC_{50} = 7.702 + 1.619(0.312) SIC2 + 2.029(0.303) MATS3v - 2.376(0.298) H-052$ | 0.915 | 0.378 | 43.384 | 0.771 | (8a) |
| $pIC_{50} = 9.895 + 1.205(0.339) SIC2 - 1.678(0.281) JG11 - 1.890(0.317) H-052$ | 0.902 | 0.406 | 36.553 | 0.725 | (9a) |

4. Conclusion

QSAR study has been carried out on the MMP-13 inhibitory activity of fused pyrimidine derivatives possessing a 1,2,4-triazol-3-yl group as a ZBG in 0D- to 2D-Dragon descriptors. The derived QSAR models have revealed that the number of Sulfur atoms (descriptor nS), Balaban mean square distance index (descriptor MSD), molecular electrotopological variation (descriptor DELS), structural information content index of neighborhood symmetry of 2nd and 3rd order (descriptors SIC2 and SIC3), average valence connectivity index chi-4 (descriptor X4Av) in addition to 1st order Galvez topological charge index (descriptor JG11) and global topological charge index (descriptor JGT) played a pivotal role in rationalization of MMP-13 inhibition activity of titled compounds. Atomic properties such as mass and volume in terms of atomic properties weighted descriptors MATS5m and MATS3v, and certain atom centred fragments such as CH2RX (descriptor C-006), X--CX--X (descriptor C-044), H attached to heteroatom (descriptor H-050) and H attached to C0(sp3) with 1X attached to next C (descriptor H-052) are also predominant to explain MMP-13 inhibition actions of fused pyrimidines.

PLS analysis has also corroborated the dominance of CP-MLR identified descriptors. Applicability domain analysis revealed that the suggested model matches the high-quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Compliance with ethical standards

Acknowledgments

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

Disclosure of conflict of interest

The authors declare no conflict of interest.

References

- [1] Wieland HA, Michaelis M, Kirschbaum BJ, Rudolphi KA. Osteoarthritis – an untreatable disease? *Nature Reviews Drug Discovery*. 2005; 4: 331-344.
- [2] Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, Moskowitz RW, Schnitzer TJ. Guidelines for the medical management of osteoarthritis. Part I. Osteoarthritis of the hip. *American College of Rheumatology. Arthritis and Rheumatology*. 1995; 38: 1535-1540.

- [3] Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, Moskowitz RW, Schnitzer TJ. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. American College of Rheumatology. *Arthritis and Rheumatology*. 1995; 38: 1541-1546.
- [4] Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. (2000). American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. *Arthritis and Rheumatology*. 2000; 43: 1905-1915.
- [5] Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, Towheed T, Welch V, Wells G, Tugwell P. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care and Research*. 2012; 64: 465-474.
- [6] FitzGerald GA. Coxibs and cardiovascular disease. *New England Journal of Medicine*. 2004; 351: 1709-1711.
- [7] Steinmeyer J, Konttinen YT. Oral treatment options for degenerative joint disease—present and future. *Advanced Drug Delivery Reviews*. 2006; 58: 168-211.
- [8] Cooper C, Adachi JD, Bardin T, Berenbaum F, Flamion B, Jonsson H, Kanis JA, Pelousse F, Lems WF, Pelletier JP, Martel-Pelletier J, Reiter S, Reginster JY, Rizzoli R, Bruyere O. How to define responders in osteoarthritis. *Current Medical Research and Opinion*. 2013; 29: 719-729.
- [9] Davies PS, Graham SM, MacFarlane RJ, Leonidou A, Mantalaris A, Tsiridis E. Disease-modifying osteoarthritis drugs: in vitro and in vivo data on the development of DMOADs under investigation. *Expert Opinion on Investigational Drugs*. 2013; 22: 423-441.
- [10] Knäuper V, López-Otin C, Smith B, Knight G, Murphy G. Biochemical characterization of human collagenase-3. *Journal of Biological Chemistry*. 1996; 271: 1544-1550.
- [11] Mitchell PG, Magna HA, Reeves LM, Lopresti-Morrow LL, Yocum SA, Rosner PJ, Geoghegan KF, Hambor JE. Cloning, expression, and type II collagenolytic activity of matrix metalloproteinase-13 from human osteoarthritic cartilage. *Journal of Clinical Investigation*. 1996; 97: 761-768.
- [12] Yocum SA, Lopresti-Morrow LL, Reeves LM, Mitchell PG. MMP-13 and MMP-1 expression in tissues of normal articular joints. *Annals of the New York Academy of Sciences*. 1999; 878: 583-586.
- [13] Neuhold LA, Killar L, Zhao W, Sung ML, Warner L, Kulik J, Turner J, Wu W, Billingham C, Meijers T, Poole AR, Babij P, DeGennaro LJ. Postnatal expression in hyaline cartilage of constitutively active human collagenase-3 (MMP-13) induces osteoarthritis in mice. *Journal of Clinical Investigations*. 2001; 107: 35-44.
- [14] Billingham RC, Dahlberg L, Ionescu M, Reiner A, Bourne R, Rorabeck C, Mitchell P, Hambor J, Diekmann O, Tschesche H, Chen J, Van Wart H, Poole AR. Enhanced cleavage of type II collagen by collagenases in osteoarthritic articular cartilage. *Journal of Clinical Investigations*. 1997; 99: 1534-1545.
- [15] Sabatini M, Lesur C, Thomas M, Chomel A, Anract P, de Nanteuil G, Pastoureau P. Effect of inhibition of matrix metalloproteinases on cartilage loss *in vitro* and in a Guinea pig model of osteoarthritis. *Arthritis and Rheumatology*. 2005; 52: 171-180.
- [16] Hoekstra R, Eskens FALM, Verweij J. Matrix metalloproteinase inhibitors: current developments and future perspectives. *Oncologist*. 2001; 6: 415-427.
- [17] Reiter LA, Robinson RP, McClure KF, Jones CS, Reese MR, Mitchell PG, Otterness IG, Bliven ML, Liras J, Cortina SR, Donahue KM, Eskra JD, Griffiths RJ, Lame ME, Lopez-Anaya A, Martinelli GJ, McGahee SM, Yocum SA, Lopresti-Morrow LL, Tobiassen LM, Vaughn-Bowser ML. Pyran-containing sulfonamide hydroxamic acids: potent MMP inhibitors that spare MMP-1. *Bioorganic and Medicinal Chemistry Letters*. 2004; 14: 3389-3395.
- [18] Peterson JT. The importance of estimating the therapeutic index in the development of matrix metalloproteinase inhibitors. *Cardiovascular Research*. 2006; 69: 677-687.
- [19] Nara H, Sato K, Kaieda A, Oki H, Kuno H, Santou T, Kanzaki N, Terauchi J, Uchikawa O, Kori M. Design, synthesis, and biological activity of novel, potent, and highly selective fused pyrimidine-2-carboxamide-4-one-based matrix metalloproteinase (MMP)-13 zinc-binding inhibitors. *Bioorganic and Medicinal Chemistry*. 2016; 24: 6149-6165.
- [20] Puerta DT, Cohen SM. A bioinorganic perspective on matrix metalloproteinase inhibition. *Current Topics in Medicinal Chemistry*. 2004; 4: 1551-1573.
- [21] Breuer E, Frant J, Reich R. Recent non-hydroxamate matrix metalloproteinase inhibitors. *Expert Opinion on Therapeutic Patents*. 2005; 15: 253-269.

- [22] Puerta DT, Griffin MO, Lewis JA, Romero-Perez D, Garcia R, Villarreal FJ, Cohen SM. Heterocyclic zinc-binding groups for use in next-generation matrix metalloproteinase inhibitors: potency, toxicity, and reactivity. *Journal of Biological Inorganic Chemistry*. 2006;11:131-138.
- [23] Nara H, Sato K, Naito T, Mototani H, Oki H, Yamamoto Y, Kuno H, Santou T, Kanzaki N, Terauchi J, Uchikawa O, Kori M. Thieno[2,3-d]pyrimidine-2-carboxamides bearing a carboxybenzene group at 5-position: Highly potent, selective, and orally available MMP-13 inhibitors interacting with the S1' binding site. *Bioorganic and Medicinal Chemistry*. 2014;22: 5487-5505.
- [24] Ruminski PG, Massa M, Strohbach J, Hanau CE, Schmidt M, Scholten JA, Fletcher TR, Hamper BC, Carroll JN, Shieh HS, Caspers N, Collins B, Grapperhaus M, Palmquist KE, Collins J, Baldus JE, Hitchcock J, Kleine HP, Rogers MD, McDonald J, Munie GE, Messing DM, Portolan S, Whiteley LO, Sunyer T, Schnute ME. Discovery of N-(4-fluoro-3-methoxybenzyl)-6-(2-(((2S,5R)-5-(hydroxymethyl)-1,4-dioxan-2-yl) methyl))-2H-tetrazol-5-yl)-2-methylpyrimidine-4-carboxamide. A highly selective and orally bioavailable matrix metalloproteinase-13 inhibitor for the potential treatment of osteoarthritis. *Journal of Medicinal Chemistry*. 2016;59: 313-327.
- [25] Nara H, Sato K, Naito T, Mototani H, Oki H, Yamamoto Y, Kuno H, Santou T, Kanzaki N, Terauchi J, Uchikawa O, Kori M. Discovery of novel, highly potent, and selective quinazoline-2-carboxamide-based matrix metalloproteinase (MMP)-13 inhibitors without a zinc binding group using a structure-based design approach. *Journal of Medicinal Chemistry*. 2014;57: 8886-8902.
- [26] Fields GB. New strategies for targeting matrix metalloproteinases. *Matrix Biology*. 2015;44-46: 239-246.
- [27] Vandenbroucke RE, Libert C. Is there new hope for therapeutic matrix metalloproteinase inhibition? *Nature Reviews Drug Discovery*. 2014;13: 904-927.
- [28] Cathcart JM, Cao J. MMP Inhibitors: Past, present and future. *Frontiers in Biosciences Landmark Ed*. 2015;20: 1164-1178.
- [29] Wang K, Xu J, Hunter DJ, Ding C. Investigational drugs for the treatment of osteoarthritis. *Expert Opinion on Investigational Drugs*. 2015;24:1539-1556.
- [30] Li NG, Shi ZH, Tang YP, Wang ZJ, Song SL, Qian LH, Qian DW, Duan JA. New hope for the treatment of osteoarthritis through selective inhibition of MMP-13. *Current Medicinal Chemistry*. 2011; 18: 977-1001.
- [31] Jacobsen JA, Major Jourden JL, Miller MT, Cohen SM. To bind zinc or not to bind zinc: an examination of innovative approaches to improved metalloproteinase inhibition. *Biochimica et Biophysica Acta Molecular Cell Research*. 2005; 1803: 72-94.
- [32] Devel L, Czarny B, Beau F, Georgiadis D, Stura E, Dive V. Third generation of matrix metalloprotease inhibitors: Gain in selectivity by targeting the depth of the S1' cavity. *Biochimie*. 2010;92:1501-1508.
- [33] Terauchi J, Kuno H, Nara H, Oki H and Sato K. Preparation of heterocyclic amides as MMP-13 inhibitors for treating osteoarthritis and rheumatoid arthritis. 2005;WO2005105760A1.
- [34] Nara H, Kaieda A, Sato K, Terauchi J. Preparation of heterocyclic amide compounds as matrix metalloproteinase inhibitors. WO2007049820A1. 2007.
- [35] Gege C, Chevrier C, Schneider M, Bluhm H, Hochguertel M, Deng H, Gallagher BM, Sucholeiki I, Taveras AG. Preparation of heterocyclic metalloprotease inhibitors. WO2008063667A1. 2008.
- [36] Pochetti G, Montanari R, Gege C, Chevrier C, Taveras AG, Mazza F. Extra binding region induced by non-zinc chelating inhibitors into the S1' subsite of matrix metalloproteinase 8 (MMP-8). *Journal of Medicinal Chemistry*. 2009; 52: 1040-1049.
- [37] Schnute ME, Ruminski PG, Hanau CE, Strohbach JW, Carroll JN and Sample K. Heterobicyclic carboxamide derivatives, processes for preparing them, pharmaceutical compositions containing them, and their uses as inhibitors of matrix metalloproteinase enzymes. WO2008149191A1. 2008.
- [38] Nara H, Kaieda A, Sato K, Naito T, Mototani H, Oki H, Yamamoto Y, Kuno H, Santou T, Kanzaki N, Terauchi J, Uchikawa O, Kori M. Discovery of novel, highly potent, and selective matrix metalloproteinase (MMP)-13 inhibitors with a 1,2,4-Triazol-3-yl moiety as a zinc binding group using a structure-based design approach. *Journal of Medicinal Chemistry*. 2017; 60: 608-626.
- [39] Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. <http://www.cambridgesoft.com>

- [40] Dragon software (version 1.11-2001) by Todeschini R.; Consonni V. Milano, Italy. <http://www.taletе.mi.it/dragon.htm>
- [41] Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: Analysis of Selwood et al. Data Set-a case study. *QSAR and Combinatorial Science*. 2003; 22: 583-595.
- [42] Sharma S, Prabhakar YS, Singh P, Sharma BK. QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach. *European Journal of Medicinal Chemistry*. 2008; 43:2354-2360.
- [43] Sharma S, Sharma BK, Sharma SK, Singh P, Prabhakar YS. Topological descriptors in modeling the agonistic activity of human A3 adenosine receptor ligands: The derivatives of 2-Chloro-N6-substituted-4'-thioadenosine-5'-uronamide. *European Journal of Medicinal Chemistry*. 2009; 44: 1377-1382.
- [44] Sharma BK, Singh P, Sarbhai K, Prabhakar YS. A quantitative structure-activity relationship study on serotonin 5-HT6 receptor ligands: Indolyl and piperidinyl sulphonamides. *SAR and QSAR in Environmental Research*. 2010; 21: 369-388.
- [45] Sharma BK, Pilania P, Sarbhai K, Singh P, Prabhakar YS. Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. *Molecular Diversity*. 2010; 14: 371-384.
- [46] Wold S. Cross-validated estimation of the number of components in factor and principal components models. *Technometrics*. 1978; 20: 397-405.
- [47] Kettaneh N, Berglund A, Wold S. PCA and PLS with very large data sets. *Computational Statistics and Data Analysis*. 2005; 48: 69-85.
- [48] Stahle L, Wold S. Multivariate data analysis and experimental design, in *Biomedical Research, Progress in Medicinal Chemistry*, Ellis GP and West WB, eds., Elsevier Science Publishers, BV, The Netherlands. 1998;25: 291-338.
- [49] Topliss JG, Edwards RP. Chance factors in studies of quantitative structure-activity relationships. *Journal of Medicinal Chemistry*. 1979; 22: 1238-1244.
- [50] Katritzky AR, Dobchev DA, Slavov S, Karelson M. Legitimate utilization of large descriptor pools for QSAR/QSPR models. *Journal of Chemical Information and Modeling*. 2008; 48:2207-2213.
- [51] So S-S, Karplus M. Three-dimensional quantitative structure activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations. *Journal of Medicinal Chemistry*. 1997; 40:4347-4359.
- [52] Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR/PLS directed structure-activity modeling of the HIV-1RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. *QSAR and Combinatorial Science*. 2004; 23: 234-244.
- [53] Akaike H. Information theory and an extension of the minimum likelihood principle. In: Petrov BN and Csaki F. (Eds.). *Second international symposium on information theory*. Budapest: Akademiai Kiado. 1973; 267-281.
- [54] Akaike H. A new look at the statistical identification model. *IEEE Transactions on Automatic Control*, AC-19. 1974; 716-723.
- [55] Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. *Quantitative Structure-Activity Relationship*. 1994; 13: 285-294.
- [56] Kubinyi H. Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. *Quantitative Structure-Activity Relationship*. 1994; 13: 393-401.
- [57] Friedman J. In *Technical Report No. 102*. Laboratory for Computational Statistics, Stanford University: Stanford. 1990.
- [58] Gramatica P. Principles of QSAR models validation: internal and external. *QSAR and Combinatorial Science*. 26: 694-701.