

Available online at [GSC Online Press Directory](https://www.gsconlinepress.com/)

GSC Biological and Pharmaceutical Sciences

e-ISSN: 2581-3250, CODEN (USA): GBPSC2

Journal homepage: <https://www.gsconlinepress.com/journals/gscbps>

(RESEARCH ARTICLE)



## QSAR rationales for the dipeptidyl peptidase-4 (DPP-4) inhibitors: The imidazolopyrimidine amides

Parihar Raghuraj and Sharma Brij Kishore \*

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

Publication history: Received on 09 June 2020; revised on 18 June 2020; accepted on 19 June 2020

Article DOI: <https://doi.org/10.30574/gscbps.2020.11.3.0169>

### Abstract

The DPP4 inhibition activity of imidazolopyrimidine amides has been quantitatively analyzed in terms of chemometric descriptors. The statistically validated QSAR models provided rationales to explain the inhibition activity of these congeners. The descriptors identified through CP-MLR analysis have highlighted the role of mean electrotopological state (Ms), number of double bonds in molecular structure (nDB), 2D Petitjean shape index (PJI2), Moran autocorrelation of lag-2/weighted by atomic polarizabilities (MATS2p), Moran autocorrelation of lag-6 and lowest eigenvalue n.5 of Burden matrix /weighted by atomic Sanderson electronegativities (MATS6e and BELv5), lowest eigenvalue n.3 and highest eigenvalue n.1 of Burden matrix/weighted by atomic van der Waals volumes (BELv3 and BEHv1). In addition to these 2<sup>nd</sup> order mean Galvez topological charge index (JGI2), number of ring tertiary C(sp<sup>3</sup>) (nCrHR) and R--CR--X type structural fragments (C-028) have also shown prevalence to model the inhibitory activity.

From statistically validated models, positive contribution of descriptors Ms, PJI2, JGI2, MATS2p, BELv5, BELv3 and BEHv1 suggested that higher values of these are conducive in improving the DPP4 inhibition activity. On the other hand, negative contribution of descriptors nDB, C-028, nCrHR and MATS6e advocated that absence of number of double bonds (nDB), R--CR--X type structural fragment (C-028), number of ring tertiary C(sp<sup>3</sup>) (nCrHR) and lower value of descriptor MATS6e would be advantageous. PLS analysis has confirmed the dominance of the CP-MLR identified descriptors and applicability domain analysis revealed the acceptable predictability of suggested models. All the compounds are within the applicability domain of the proposed models and were evaluated correctly.

**Keywords:** Quantitative structure-activity relationship (QSAR); DPP-4 inhibitors; Combinatorial protocol in multiple linear regression (CP-MLR) analysis; Chemometric descriptors; Imidazolopyrimidine amides.

### 1. Introduction

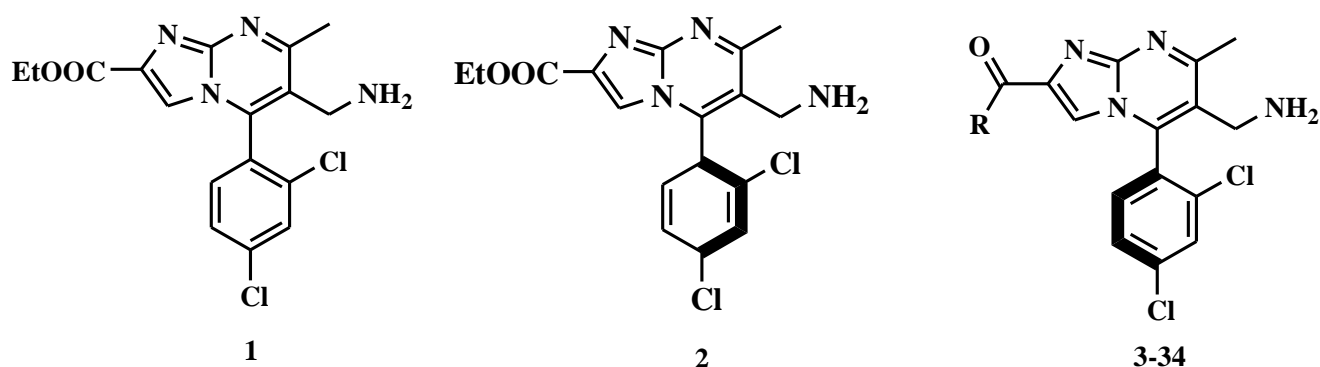
Therapeutics based on Glucagon-like peptide-1 (GLP-1<sup>a</sup>) is among the novel and promising targets to cure type 2 diabetes [1-3]. The active and natural form of GLP-1, the incretin hormone GLP-1 (7-36), is secreted from intestinal L-cells after the intake of meals. The stimulation of insulin secretion, inhibition of glucagon release, delay in gastric emptying and promotion of  $\beta$ -cell trophism in intestinal L-cells are advantageous to glucose homeostasis in both the animal models and human [4,5]. Studies revealed that GLP-1 levels are noticeably reduced in type 2 diabetics and exogenous infusion of it may lead to normal insulin response to glucose [6-8] and this fact is the basis for GLP-1 and its analogues as novel treatments of type 2 diabetes. One such example of a GLP-1 analogue is exenatide [9,10]. Half-maximal effective concentration of 10 pM of the most potent incretin hormone, GLP-1 (7-36), is required to show its effects on pancreatic  $\beta$ -cells [11]. The biological functions of GLP-1 (7-36) are exerted through circulation and binding to the GLP-1 receptor that is highly expressed in pancreatic  $\beta$ -cells. After secretion GLP-1 (7-36) is rapidly degraded by DPP4 (EC 3.4.14.5) to afford inactive GLP-1 (9-36) under normal physiological conditions. The apparent half-life for GLP-1 (7-36) in this quick inactivation process is 60-90s. It is evinced that due to this natural degradation mechanism

\* Corresponding author: Sharma Brij Kishore

less than 50% of released active GLP-1 (7-36) can reach circulation [12]. Thus it is apparent that a DPP4 inhibitor can prevent degradation of and lead to potentiation of GLP-1 and further improve glucose and insulin homeostasis [13,14].

DPP4, ubiquitously expressed throughout the body, is a nonclassical and sequence-specific serine protease. Membrane-bound DPP4 is highly expressed in the endothelium of the capillary bed in close proximity to intestinal L-cells where secretion of GLP-1 takes place. The other form which circulates in plasma is soluble form of DPP4 plays a little role in the cleavage of GLP-1 [15,16]. Vildagliptin [17], sitagliptin [18], saxagliptin [19] and alogliptin [20] are examples of small molecule DPP4 inhibitors which have demonstrated ability to lower blood glucose and HbA1c levels and to improve glucose tolerance in type 2 diabetic patients [21]. Several novel series of azolopyrimidine amines, containing an aromatic or heteroaromatic group on the azolo ring, as potent and selective DPP4 inhibitors were reported in view of medicinal chemistry efforts to discover novel scaffolds [22]. The substitution of aromatic or heteroaromatic group on the azolo ring in these compounds showed enhancement in the binding affinity to DPP4 but displayed high levels of the human ether-à-go-go related gene (hERG) and sodium channel inhibition.

As an attempt to minimize undesired hERG and sodium channel activities a novel series of imidazolopyrimidine amides as a highly potent and selective class of DPP4 inhibitors has been reported by Meng *et al.* [23]. The general structure of these analogues is shown in Figure 1 and structural variations are given in Table 1.



**Figure 1** General structures of imidazolopyrimidine derivatives




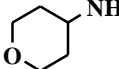
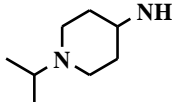
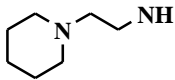
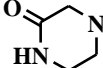
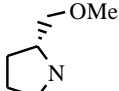
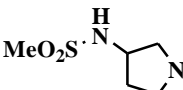
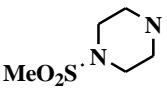
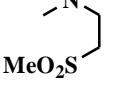
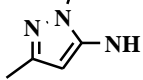
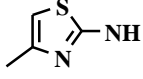
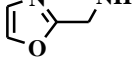
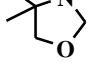
In the present communication a 2D-quantitative SAR (2D-QSAR) has been conducted to provide the rationale for drug-design and to explore the possible mechanism of the action. In the congeneric series, where a relative study is being carried out, the 2D-descriptors may play important role in deriving the 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.

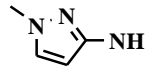
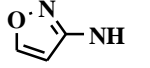
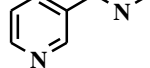
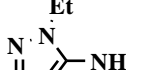
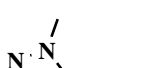
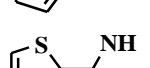
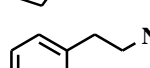
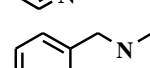
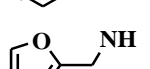
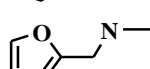
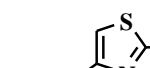
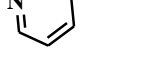
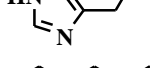
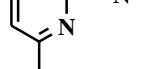
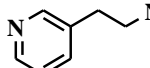
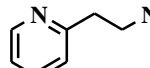
## 2. Material and methods

### 2.1. Data-set

For present work the imidazolopyrimidine amides (Table 1), along with their in vitro human DPP4 inhibition activity, have been taken from the literature [23]. The inhibition activity reported in terms of  $K_i$  is expressed as  $pK_i$  on a molar basis and considered as the dependent variable for the present quantitative analysis.

**Table 1** Structural variations and DPP-4 binding affinities of imidazolopyrimidine amides (see fig. 1 for general structure)

Cpd.	R	pK <sub>i</sub> (M) <sup>a</sup>					
		Obsd <sup>b</sup> .	Calculated				
			Eq. (7)	Eq. (8)	Eq. (9)	Eq. (10)	PLS
1	racemic	9.15	8.98	8.94	9.01	9.06	8.99
2	chiral	7.93	8.08	8.03	7.93	8.10	7.83
3 <sup>c</sup>	OEt (chiral)	9.40	8.98	8.94	9.01	9.06	8.99
4		8.62	8.81	8.80	8.77	8.64	8.94
5		8.54	8.45	8.46	8.43	8.68	8.75
6		8.66	8.47	8.68	8.50	8.32	8.50
7 <sup>c</sup>		8.51	8.40	8.42	8.52	8.29	8.44
8		8.30	8.26	8.30	8.18	8.26	8.22
9		8.06	8.30	8.33	8.33	8.28	8.14
10		9.00	8.82	8.95	8.95	8.72	8.81
11		8.51	8.52	8.53	8.63	8.32	8.59
12		9.70	9.52	9.55	9.49	9.59	9.56
13		9.30	9.37	9.41	9.33	9.42	9.39
14		9.52	9.70	9.54	9.76	9.56	9.56
15		9.05	8.75	8.73	8.81	8.74	8.87
16		8.18	8.43	8.25	8.43	8.56	8.21
17 <sup>c</sup>		8.70	8.54	8.49	8.91	8.65	8.71
18		9.22	9.22	9.19	9.01	8.86	9.10

19		8.74	8.70	8.68	8.88	8.72	8.82
20		8.42	8.73	8.91	8.59	8.58	8.59
21 <sup>c</sup>		8.59	8.49	8.67	8.63	8.40	8.58
22		8.96	8.75	8.73	8.62	9.03	8.94
23		8.70	8.74	8.93	8.53	8.97	8.83
24		8.49	8.44	8.27	8.48	8.56	8.51
25 <sup>c</sup>		9.15	9.18	8.98	9.04	8.78	8.85
26		8.59	8.46	8.65	8.61	8.47	8.60
27 <sup>c</sup>		8.4	8.75	8.56	8.69	8.67	8.50
28		8.52	8.70	8.51	8.86	8.64	8.56
29		8.30	8.56	8.55	8.32	8.67	8.27
30		8.33	8.06	8.22	8.38	8.30	8.48
31		8.27	8.57	8.40	8.61	8.30	8.30
32		8.82	8.68	8.67	8.68	8.83	8.81
33		8.85	8.68	8.50	8.61	8.55	8.56
34 <sup>c</sup>		8.89	8.68	8.67	8.68	8.76	8.77

<sup>a</sup>On molar basis; <sup>b</sup>Taken from reference [23]; <sup>c</sup>Compounds in test set.

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 20% of total compounds, was employed to validate such models. The selection of test-set compounds was made through SYSTAT [24] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the binding activity,  $pK_i$  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 [25] and 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 [26] for computing the descriptors corresponding to 0D-, 1D-, and 2D-classes. The combinatorial protocol in multiple linear regression (CP-MLR) [27] analysis and partial least-squares (PLS) [28-30] 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. Its procedural aspects and implementation are discussed in some of our publications [31-36]. The thrust of this procedure is in its embedded four 'filters'. They are briefly as follows: filter-1 seeds the variables by way of limiting inter-parameter correlations to predefined level (upper limit  $\leq 0.79$ ); filter-2 controls the variables entry to a regression equation through t-values of coefficients (threshold value  $\geq 2.0$ ); filter-3 provides comparability of equations with different number of variables in terms of square root of adjusted multiple correlation coefficient of regression equation,  $\bar{r}$ ; 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  $\bar{r}$  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 [37,38] 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 DPP4 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 [39,40]. 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  $\pm x(\text{s.d.})$  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.

### 3. Results and discussion

#### 3.1. QSAR results

For the compounds in Table 1, a total number of 479 descriptors belonging to 0D- to 2D- classes of DRAGON have been computed and were subjected to CP-MLR analysis. All the 34 compounds of data set were further divided into training-set and test-set. Seven compounds (nearly 20% 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 twenty seven compounds in the training-set. The squared correlation coefficient between the observed and predicted values of compounds from test-set,  $r^2_{\text{Test}}$ , 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  $r^2_{\text{Test}}$  value is always good. But considering the stringency of test-set procedures, often  $r^2_{\text{Test}}$  values in the range of 0.5 to 0.6 are regarded as logical models. Following the strategy to explore only predictive models, CP-MLR resulted one model in three descriptors, five models in four descriptors and sixteen models in five descriptors at upper limit of filter-1. The highest significant of them, in statistical sense, are given through Equations (1)-(10):

$$\text{pKi} = 7.583 + 1.556(0.241)\text{Ms} + 0.716(0.215)\text{BELe5} + 0.729(0.241)\text{MATS2p}$$

$$n = 27, r = 0.812, s = 0.271, F = 14.904, q^2_{\text{L00}} = 0.515, q^2_{\text{L50}} = 0.504, r^2_{\text{Test}} = 0.125 \quad (1)$$

$$\text{pKi} = 7.267 + 1.503(0.229)\text{Ms} + 0.300(0.121)\text{PJI2} + 0.985(0.256)\text{BELv3}$$

$$+ 0.738(0.223)\text{MATS2p}$$

$$n = 27, r = 0.844, s = 0.254, F = 13.731, q^2_{\text{L00}} = 0.565, q^2_{\text{L50}} = 0.516, r^2_{\text{Test}} = 0.632 \quad (2)$$

$$\text{pKi} = 8.043 - 0.677(0.253)\text{Mv} + 2.308(0.381)\text{Ms} - 0.880(0.255)\text{nDB}$$

$$+ 0.635(0.248)\text{BELv3}$$

$$n = 27, r = 0.836, s = 0.260, F = 12.856, q^2_{\text{L00}} = 0.524, q^2_{\text{L50}} = 0.537, r^2_{\text{Test}} = 0.527 \quad (3)$$

$$\text{pKi} = 7.935 + 2.678(0.448)\text{Ms} - 0.806(0.249)\text{nDB} - 0.619(0.246)\text{IC1}$$

$$+ 0.575(0.199)\text{BELe5}$$

$$n = 27, r = 0.832, s = 0.263, F = 12.380, q^2_{\text{L00}} = 0.507, q^2_{\text{L50}} = 0.594, r^2_{\text{Test}} = 0.525 \quad (4)$$

$$\text{pKi} = 8.031 + 1.239(0.235)\text{Ms} + 0.315(0.127)\text{PJI2} + 0.991(0.272)\text{BELv3}$$

$$- 0.679(0.232)\text{GATS2v}$$

$$n = 27, r = 0.831, s = 0.264, F = 12.295, q^2_{\text{L00}} = 0.501, q^2_{\text{L50}} = 0.513, r^2_{\text{Test}} = 0.501 \quad (5)$$

$$\text{pKi} = 8.080 + 1.325(0.227)\text{MAXDN} - 0.556(0.219)\text{BELm8} + 0.706(0.345)\text{BELv3}$$

$$+ 0.957(0.256)\text{MATS2p}$$

$$n = 27, r = 0.821, s = 0.270, F = 11.449, q^2_{\text{L00}} = 0.503, q^2_{\text{L50}} = 0.525, r^2_{\text{Test}} = 0.513 \quad (6)$$

$$\text{pKi} = 7.360 + 1.914(0.288)\text{Ms} - 0.744(0.185)\text{nDB} + 0.214(0.101)\text{PJI2}$$

$$+ 1.110(0.215)\text{BELv3} + 0.765(0.192)\text{JGI2}$$

$$n = 27, r = 0.901, s = 0.210, F = 18.170, q^2_{\text{L00}} = 0.651, q^2_{\text{L50}} = 0.600, r^2_{\text{Test}} = 0.548 \quad (7)$$

$$\begin{aligned} \text{pKi} &= 7.644 + 1.753(0.277)\text{Ms} - 0.774(0.188)\text{nDB} + 1.059(0.211)\text{BELv3} \\ &+ 0.729(0.196)\text{JGI2} - 0.357(0.171)\text{C-028} \\ n &= 27, r = 0.900, s = 0.211, F = 18.042, q^2_{\text{L00}} = 0.667, q^2_{\text{L50}} = 0.690, r^2_{\text{Test}} = 0.579 \end{aligned} \quad (8)$$

$$\begin{aligned} \text{pKi} &= 6.629 + 1.573(0.194)\text{Ms} + 0.331(0.102)\text{PJI2} + 0.798(0.249)\text{BEHv1} \\ &+ 0.705(0.232)\text{BELv3} + 0.603(0.191)\text{MATS2p} \\ n &= 27, r = 0.898, s = 0.213, F = 17.656, q^2_{\text{L00}} = 0.644, q^2_{\text{L50}} = 0.696, r^2_{\text{Test}} = 0.616 \end{aligned} \quad (9)$$

$$\begin{aligned} \text{pKi} &= 8.310 + 1.327(0.208)\text{Ms} + 0.265(0.107)\text{PJI2} + 0.669(0.172)\text{BELe5} \\ &- 0.723(0.208)\text{MATS6e} - 0.771(0.199)\text{nCrHR} \\ n &= 27, r = 0.883, s = 0.227, F = 14.972, q^2_{\text{L00}} = 0.569, q^2_{\text{L50}} = 0.550, r^2_{\text{Test}} = 0.511 \end{aligned} \quad (10)$$

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  $q^2_{\text{L00}}$  and  $q^2_{\text{L50}}$  ( $> 0.5$ ) have accounted for their internal robustness. For all above models except equation (1) the  $r^2_{\text{Test}}$  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 [41] 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 81.22 percent of variance in observed activity of the compounds. In fact, a total number of sixteen such models, sharing 19 descriptors among them, have been obtained through CP-MLR and the most significant four of them have been documented through Equation (7)-(10). The shared 19 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  $M_v$  is from constitutional and  $\text{MAXDN}$  is from topological class. The  $M_v$  represents mean atomic Van der Waals volume (scaled on carbon atom) (Equation 3) and  $\text{MAXDN}$  is maximal electrotopological negative variation (Equation 6). The further discussion is, however, based on the highest significant Equations (7)-(10). The derived statistical parameters of these four models have shown that these models are significant. These models were, therefore, used to calculate the activity profiles of all the compounds and are included in Table 1 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 ensure the goodness of fit for each of these four models.

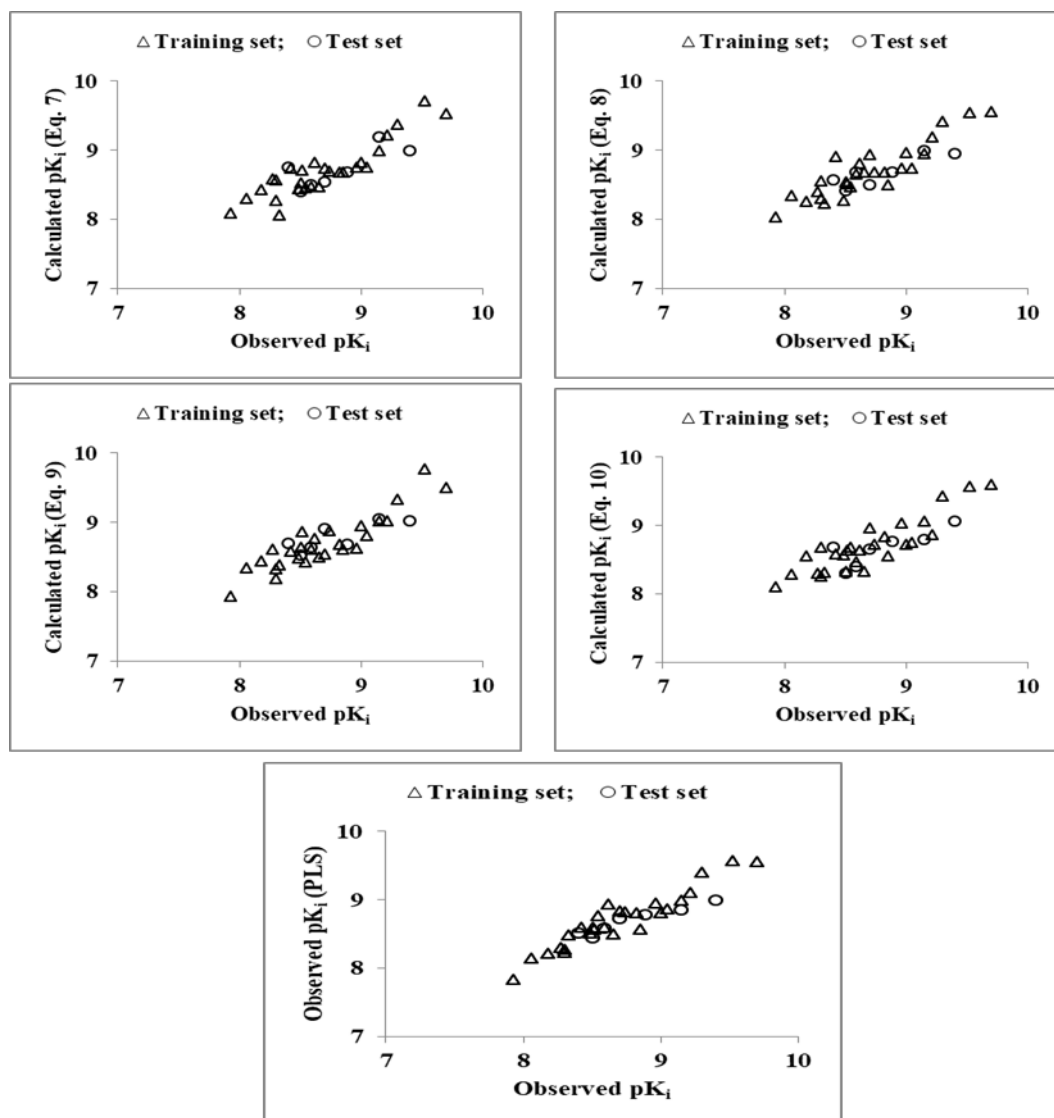
Descriptors  $M_s$  (mean electrotopological state) and  $n\text{DB}$  (number of double bonds in molecular structure) belong to constitutional class. From the sign of regression coefficients it is evident that higher value of mean electrotopological state (descriptor  $M_s$ ) and lower number of double bonds (descriptor  $n\text{DB}$ ) are helpful to augment the activity. The descriptor  $\text{PJI2}$  participated in these models is topological class descriptor and represents 2D Petitjean shape index. The positive sign of regression coefficient of descriptor  $\text{PJI2}$  suggest that a higher value of this descriptor is beneficiary to the DPP4 inhibition activity. The descriptors  $\text{MATS2p}$  (Moran autocorrelation of lag-2/weighted by atomic polarizabilities) and  $\text{MATS6e}$  (Moran autocorrelation of lag-6/weighted by atomic Sanderson electronegativities) are 2D autocorrelation descriptors. It is evinced from the models mentioned above the descriptor  $\text{MATS2p}$  contributed positively and descriptor  $\text{MATS6e}$  negatively to the activity. Thus a higher value of descriptor  $\text{MATS2p}$  and a lower value of descriptor  $\text{MATS6e}$  will be supportive to enhance the inhibition activity.

**Table 2** Identified descriptors<sup>a</sup> along with their physical meaning, average regression coefficient and incidence<sup>b</sup>, in modeling the DPP-4 binding affinity.

S. No.	Descriptor	Descriptor class	Physical meaning	Average regression coefficient (incidence)
1	Ms	Constitutional	Mean electrotopological state	1.835(15)
2	nDB	Constitutional	Number of double bonds	-0.685(11)
3	HNar	Topological	Narumi harmonic topological index	-1.304(2)
4	PJI2	Topological	2D Petitjean shape index	0.280(7)
5	IC1	Topological	information content index of neighborhood symmetry of 1-order	-0.585(1)
6	BELm8	BCUT	Lowest eigenvalue n.8 of Burden matrix/ weighted by atomic masses	-0.479(1)
7	BEHv1	BCUT	Highest eigenvalue n.1 of Burden matrix/ weighted by atomic van der Waals volumes	0.989(2)
8	BELv3	BCUT	lowest eigenvalue n.3 of Burden matrix/ weighted by atomic van der Waals volumes	0.935(12)
9	BELe5	BCUT	lowest eigenvalue n.5 of Burden matrix/ weighted by atomic Sanderson electronegativities	0.653(6)
10	JGI2	Galvez topological charge indices	Mean topological charge index of order 2	0.747(2)
11	MATS5e	2D autocorrelations	Moran autocorrelation of lag-5/ weighted by atomic Sanderson electronegativities	-0.597(1)
12	MATS6e	2D autocorrelations	Moran autocorrelation of lag-6/ weighted by atomic Sanderson electronegativities	-0.594(2)
13	MATS2p	2D autocorrelations	Moran autocorrelation of lag-2/ weighted by atomic polarizabilities	0.694(5)
14	GATS2v	2D autocorrelations	Geary autocorrelation of lag-2/ weighted by atomic van der Waals volumes	-0.794(1)
15	nCp	Functional	Number of total primary C (sp <sup>3</sup> )	0.274(1), 0.612(1)
16	nCrHR	Functional	Number of ring tertiary C(sp <sup>3</sup> )	-0.771(1)
17	nNR2	Functional	Number of tertiary aliphatic amines	-0.702(1)
18	C-028	Atom-centered fragments	R--CR--X	-0.464(6)
19	C-032	Atom-centered fragments	X--CX--X	-0.458(2)

<sup>a</sup>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 0.813, and filter-4 as  $0.3 \leq q_2 \leq 1.0$  with a training set of 27 compounds. <sup>b</sup>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





**Figure 2** Plot of observed versus calculated  $pK_i$  values for training- and test-set compounds.

The participated descriptors BELe5 (lowest eigenvalue n.5 of Burden matrix/weighted by atomic Sanderson electronegativities), BELv3 (lowest eigenvalue n.3 of Burden matrix/weighted by atomic van der Waals volumes) and BEHv1 (highest eigenvalue n.1 of Burden matrix/weighted by atomic van der Waals volumes) belong to BCUT class. All these descriptors contributed positively to the activity suggesting that higher value of these will augment the activity.

From Equations (7)-(10), it appeared that the descriptors nCrHR, a functional group accounting descriptor representing number of ring tertiary C(sp<sup>3</sup>) functionality in a structure and atom centered fragment accounting descriptor C-028 showing R—CR—X type fragment in a molecular structure make negative contribution to activity and JGI2, mean Galvez topological charge index of order 2 shown positive correlation to the activity. In this way absence of number of ring tertiary C(sp<sup>3</sup>) functionality along with R—CR—X type fragment in a molecular structure and higher value of mean Galvez topological charge index of order 2 would be advantageous in improving the DPP4 inhibition activity of a compound.

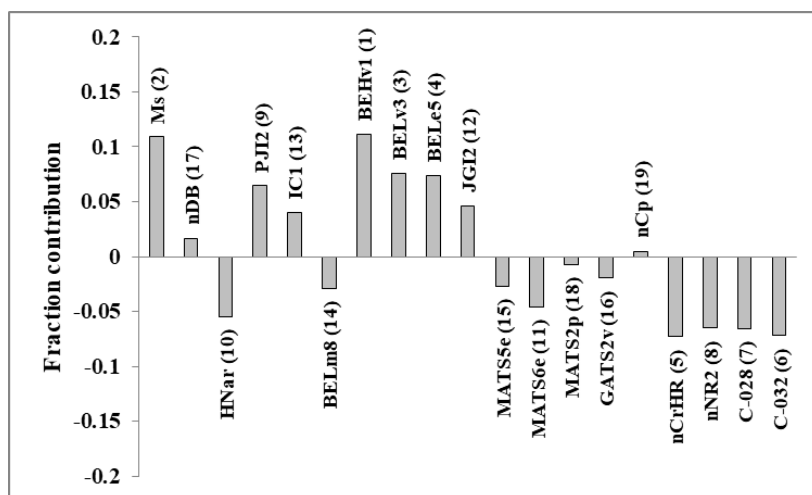
To corroborate the study further, a PLS analysis has also been carried out on 19 descriptors identified through CP-MLR and results are given in Table 3. For this purpose, the descriptors have been autoscaled (zero mean and unit s.d.) to give each one of them equal weight in the analysis. In the PLS cross-validation, three components have been found to be the optimum for these 19 descriptors and they explained 89.7% variance in the activity ( $r^2 = 0.897$ ). The MLR-like PLS coefficients of these 19 descriptors are given in Table 3.

**Table 3** PLS and MLR-like PLS models from the descriptors of five parameter CP-MLR models for DPP-4 binding affinity.

A: PLS equation			
PLS components		PLS coefficient (s.e.) <sup>a</sup>	
Component-1		0.196(0.015)	
Component-2		-0.113(0.019)	
Component-3		0.078(0.023)	
Constant		8.693	
B: MLR-like PLS equation			
S. No.	Descriptor	MLR-like coefficient (f.c.) <sup>b</sup>	Order
1	Ms	0.311(0.109)	2
2	nDB	0.045(0.016)	17
3	HNar	-0.157(-0.55)	10
4	PJl2	0.184(0.064)	9
5	IC1	0.115(0.040)	13
6	BELm8	-0.083(-0.029)	14
7	BEHv1	0.318(0.111)	1
8	BELv3	0.215(0.075)	3
9	BELe5	0.210(0.073)	4
10	JGI2	0.130(0.045)	12
11	MATS5e	-0.077(-0.027)	15
12	MATS6e	-0.131(-0.045)	11
13	MATS6p	-0.020(-0.007)	18
14	GATS2v	-0.055(-0.019)	16
15	nCp	0.011(0.004)	19
16	nCrHR	-0.208(-0.072)	5
17	nNR2	-0.184(-0.064)	8
18	C-028	-0.187(-0.065)	7
19	C-032	-0.205(-0.071)	6
Constant		7.909	
C: PLS regression statistics		Values	
n		27	
r		0.947	
s		0.148	
F		67.415	
q <sup>2</sup> <sub>L00</sub>		0.851	
q <sup>2</sup> <sub>L50</sub>		0.860	
r <sup>2</sup> <sub>Test</sub>		0.662	

<sup>a</sup>Regression coefficient of PLS factor and its standard error. <sup>b</sup>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.

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 1. 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 3).



**Figure 3** Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 19 identified descriptors (Table 3) associated with DPP-4 binding affinity of the compounds.

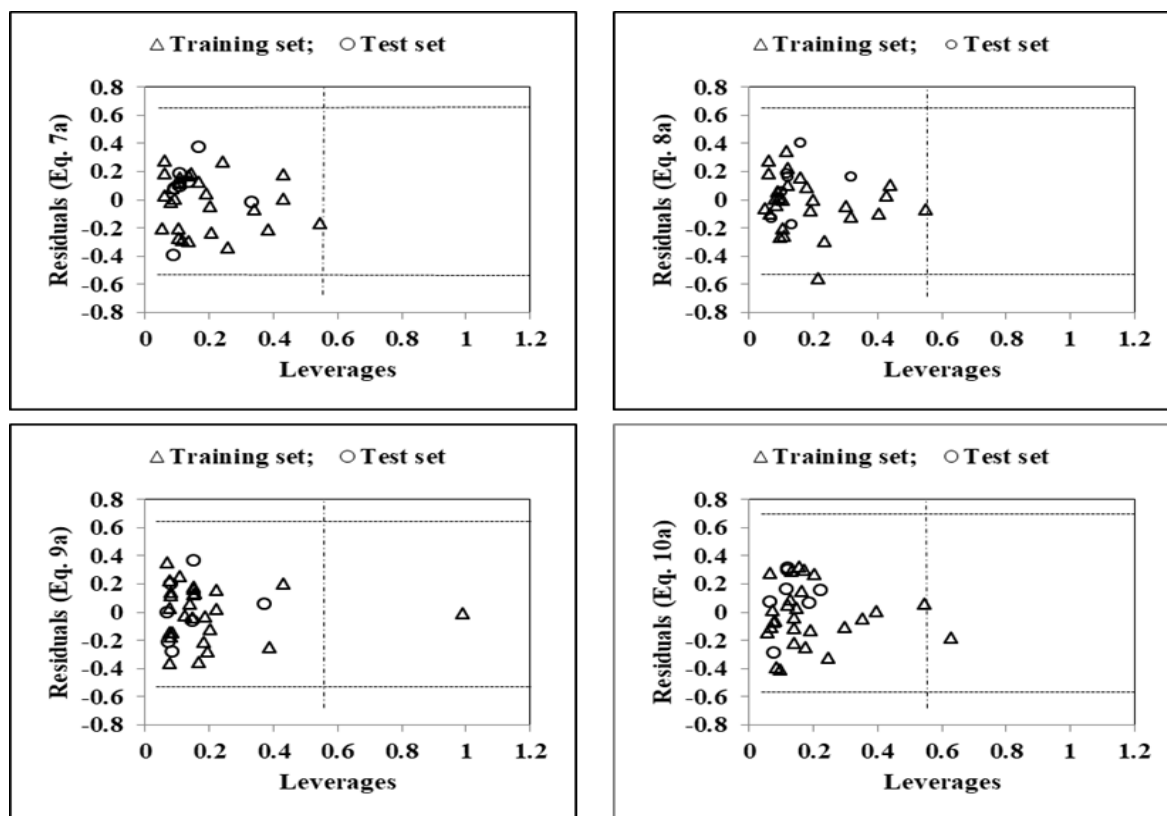
The PLS analysis in 19 identified descriptors revealed three components (Table 3) as optimum to explain the DPP4 inhibition activity. The top ten descriptors in decreasing order of significance are BEHv1, Ms, BELv3, BELe5, nCrHR, C-032, C-028, nNR2, PJI2 and HNar (Table 3, figure 3). Among these descriptors, BEHv1, Ms, BELv3, BELe5, nCrHR, C-028 and PJI2 are part of Equations discussed above and convey same inferences in PLS analysis. The negative contributions of atom centered fragment descriptor C-032 (X-CX-X type fragment), functional group count descriptor nNR2 (number of tertiary aliphatic amine functionality in a molecule) and topological descriptor HNar (Narumi harmonic topological index) advocated lower value of these are helpful in improving the activity profile. It is also observed that PLS model from the dataset devoid of 19 descriptors (Table 3) 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 dataset (Table 4), no any compound has been identified as an obvious 'outlier' for the DPP4 inhibitory activity if the limit of normal values for the Y outliers (response outliers) was set as  $3 \times$  (standard deviation) units. One of the compound (**2**; Table 1) was found to have leverage (h) values greater than the threshold leverage ( $h^*$ ); suggesting it as chemically influential compound.

**Table 4** Models derived for the whole data set ( $n = 34$ ) for the DPP-4 binding affinity in descriptors identified through CP-MLR.

Model	r	s	F	q2LOO	Eq.
$pKi = 7.366 + 1.980(0.276)Ms - 0.762(0.176)nDB$ + $0.231(0.092)PJI2 + 1.073(0.178)BELv3$ + $0.731(0.181)JGI2$	0.880	0.216	19.331	0.653	(7a)
$pKi = 7.653 + 1.858(0.265)Ms - 0.844(0.176)nDB$ + $1.081(0.175)BELv3 + 0.694(0.180)JGI2$ - $0.426(0.156)C-028$	0.884	0.212	20.176	0.680	(8a)
$pKi = 6.571 + 1.600(0.184)Ms + 0.356(0.092)PJI2$ + $0.764(0.241)BEHv1 + 0.777(0.180)BELv3$ + $0.698(0.171)MATS2p$	0.884	0.213	20.026	0.156	(9a)
$pKi = 8.402 + 1.269(0.194)Ms + 0.295(0.098)PJI2$ + $0.685(0.159)BELe5 - 0.834(0.187)MATS6e$ - $0.792(0.185)nCrHR$	0.865	0.228	16.727	0.623	(10a)



**Figure 4** Williams plot for the training-set and test- set for inhibition activity of DPP4 for the compounds in Table 1. The horizontal dotted line refers to the residual limit ( $\pm 3 \times$  standard deviation) and the vertical dotted line represents threshold leverage  $h^*$  ( $= 0.529$ ).

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.

#### 4. Conclusion

The DPP4 inhibition activity of imidazolopyrimidine amides 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 mean electrotopological state (Ms), number of double bonds in molecular structure (nDB), 2D Petitjean shape index (PJI2), Moran autocorrelation of lag-2/weighted by atomic polarizabilities (MATS2p), Moran autocorrelation of lag-6/weighted by atomic Sanderson electronegativities (MATS6e), lowest eigenvalue n.5 of Burden matrix/ weighted by atomic Sanderson electronegativities (BELe5), lowest eigenvalue n.3 of Burden matrix/ weighted by atomic van der Waals volumes (BELv3), highest eigenvalue n.1 of Burden matrix/ weighted by atomic van der Waals volumes (BEHv1). In addition to these 2<sup>nd</sup> order mean Galvez topological charge index (JGI2), number of ring tertiary C(sp<sup>3</sup>) (nCrHR) and R--CR--X type structural fragments (C-028) have also shown prevalence to model the inhibitory activity.

From statistically validated models, it appeared that the descriptors Ms, PJI2, JGI2, MATS2p, BELe5, BELv3 and BEHv1 make positive contribution to activity and their higher values are conducive in improving the DPP4 inhibition activity of a compound. On the other hand, the descriptors nDB, C-028, nCrHR and MATS6e render detrimental effect to activity. Therefore, absence or lower number of double bonds (nDB), R--CR--X type structural fragment (C-028), number of ring tertiary C(sp<sup>3</sup>) (nCrHR) and lower value of descriptor MATS6e 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.

---

## 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] Wideman RD and Kieffer TJ. (2009). Mining incretin hormone pathways for novel therapies. *Trends in Endocrinology and Metabolism*, 20, 280-286.
- [2] Triplitt C, Wright A and Chiquette E. (2006). Incretin mimetics and dipeptidyl peptidase-IV inhibitors: potential new therapies for type 2 diabetes mellitus. *Pharmacotherapy*, 26, 360-374.
- [3] Gallwitz B. (2005). Glucagon-like peptide-1-based therapies for the treatment of type 2 diabetes mellitus. *Treatments in Endocrinology*, 4, 361-370.
- [4] Holst JJ. (1994). Glucagon-like peptide-1 (GLP-1) a newly discovered GI hormone. *Gastroenterology*, 107, 1048-1055.
- [5] Drucker DJ. (1998). Glucagon-like peptides. *Diabetes*, 47, 159-169.
- [6] Vilsboll T and Holst JJ. (2004). Incretins, insulin secretion and type 2 diabetes mellitus. *Diabetologia*, 47, 357-366.
- [7] Elahi D, McAloon-Dyke M, Fukagawa NK, Meneilly GS, Sclater AL, Minaker KL, Habener JF and Andersen DK. (1994). The insulinotropic actions of glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (7-37) in normal and diabetic subjects. *Regulatory Peptides*, 51, 63-75.
- [8] Nauck MA, Bartels E, Oerskov C, Ebert R and Creutzfeldt WJ. (1993). Additive insulinotropic effects of exogenous synthetic human gastric inhibitory polypeptide and glucagon-like peptide-1-(7-36) amide infused at near-physiological insulinotropic hormone and glucose concentrations. *Clinical Endocrinology and Metabolism*, 76, 912-917.
- [9] Gentilella R, Bianchi C, Rossi A and Rotella CM. (2009). Exenatide: a review from pharmacology to clinical practice. *Diabetes, Obesity and Metabolism*, 11,544-556.
- [10] Gonzalez C, Beruto V, Keller G, Santoro S and Di Girolamo G. (2006). Investigational treatments for type 2 diabetes mellitus: Exenatide and Liraglutide. *Expert Opinion on Investigational Drugs*, 15, 887-895.
- [11] Holst JJ. (2002). Therapy of type 2 diabetes mellitus based on the actions of glucagon-like peptide-1. *Diabetes/ Metabolism Research and Reviews*, 18, 430-441.
- [12] Hansen L, Deacon CF, Orskov C and Holst JJ. (1999). Glucagon-like peptide-1-(7-36) amide is transformed to glucagon-like peptide-1-(9-36)amide by dipeptidyl peptidase IV in the capillaries supplying the L cells of the porcine intestine. *Endocrinology*, 140, 5356-5363.
- [13] Nagakura T, Yasuda N, Yamazaki K, Ikuta H, Yoshikawa S, Asano O and Tanaka I. (2001). Improved glucose tolerance via enhanced glucose-dependent insulin secretion in dipeptidyl peptidase IV-deficient Fischer rats. *Biochemical and Biophysical Research Communications*, 284, 501-506.
- [14] Marguet D, Baggio L, Kobayashi T, Bernard A, Pierres M, Nielsen PF, Ribet U, Watanabe T, Drucker DJ and Wagtmann N. (2000). Enhanced insulin secretion and improved glucose tolerance in mice lacking CD26. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 6874-6879.
- [15] Durinx C, Lambeir A, Bosmans E, Falmagne J, Berghmans R, Haemers A, Scharpe S and De Meester I. (2000). Molecular characterization of dipeptidyl peptidase activity in serum. Soluble CD26/dipeptidyl peptidase IV is responsible for the release of X-Pro dipeptides. *European Journal of Biochemistry*, 267, 5608-5613.
- [16] Iwaki-Egawa S, Watanabe Y, Kikuya Y and Fujimoto Y. (1998). Dipeptidyl peptidase IV from human serum: purification, characterization, and N-terminal amino acid sequence. *Journal of Biochemistry*, 124, 428-433.

- [17] Villhauer EB, Brinkman JA, Naderi GB, Burkey BF, Dunning BE, Prasad K, Mangold BL, Russell ME and Hughes TE. (2003). 1-[[[3-Hydroxy-1-adamantyl]amino]acetyl]-2-cyano-(S)-pyrrolidine: A potent, selective, and orally bio available dipeptidyl peptidase IV inhibitor with antihyperglycemic properties. *Journal of Medicinal Chemistry*, 46, 2774-2789.
- [18] Kim D, Wang L, Beconi M, Eiermann GJ, Fisher MH, He H, Hickey GJ, Kowalchick JE, Leiting B, Lyons K, Marsilio F, McCann ME, Patel RA, Petrov A, Scapin G, Patel SB, Roy RS, Wu JK, Wyvratt MJ, Zhang BB, Zhu L, Thornberry NA and Weber AE. (2005). (2R)-4-Oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3- $\alpha$ ]pyrazin-7(8H)-yl]-1-(2,4,5-trifluorophenyl)butan-2-amine: a potent, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *Journal of Medicinal Chemistry*, 48, 141-151.
- [19] Augeri DJ, Robl JA, Betebenner DA, Magnin DR, Khanna A, Robertson JG, Wang A, Simpkins LM, Taunk P, Huang Q, Han SP, Abboa-Offei B, Cap M, Xin L, Tao L, Tozzo E, Welzel GE, Egan DM, Marcinkkeviciene J, Chang SY, Biller SA, Kirby MS, Parker RA and Hamann LG. (2005). Discovery and preclinical profile of saxagliptin (BMS-477118): a highly potent, long-acting, orally active dipeptidyl peptidase IV inhibitor for the treatment of Type 2 diabetes. *Journal of Medicinal Chemistry*, 48, 5025-5037.
- [20] Feng J, Zhang Z, Wallace MB, Stafford JA, Kaldor SW, Kassel DB, Navre M, Shi L, Skene RJ, Asakawa T, Takeuchi K, Xu R, Webb DR and Gwaltney SL. (2007). Discovery of alogliptin: a potent, selective, bio available, and efficacious inhibitor of dipeptidyl peptidase IV. *Journal of Medicinal Chemistry*, 50, 2297-2300.
- [21] Havale SH and Pal M. (2009). Medicinal chemistry approaches to the inhibition of dipeptidyl peptidase-4 for the treatment of type 2 diabetes. *Bioorganic and Medicinal Chemistry*, 17, 1783-1802.
- [22] Brigance RP, Meng W, Fura A, Harrity T, Wang A, Zahler R, Kirby MS and Hamann LG. (2010). Synthesis and SAR of azolopyrimidines as potent and selective dipeptidyl peptidase-4 (DPP4) inhibitors for type 2 diabetes. *Bioorganic and Medicinal Chemistry Letters*, 20, 4395-4398.
- [23] Meng W, Brigance RP, Chao HJ, Fura A, Harrity T, Marcinkkeviciene J, O'Connor SP, Tamura JK, Xie D, Zhang Y, Klei HE, Kish K, Weigett CA, Turdi H, Wang A, Zahler R, Kirby MS and Hamann LG. (2010). Discovery of 6-(Aminomethyl)-5-(2,4-dichlorophenyl)-7-methylimidazo[1,2- $\alpha$ ]pyrimidine-2-carboxamides as potent, selective dipeptidyl peptidase-4 (DPP4) inhibitors. *Journal of Medicinal Chemistry*, 53, 5620-5628.
- [24] SYSTAT, Version 7.0; SPSS Inc 444 North Michigan Avenue, Chicago IL, 60611.
- [25] ChemDraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation, Cambridge, USA.
- [26] Todeschini R and Consonni V. Dragon software (version 1.11-2001), Milano, Italy.
- [27] Prabhakar YS. (2003). 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*, 22, 583-595.
- [28] Wold S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics*, 20, 397-405.
- [29] Kettaneh N, Berglund A and Wold S. (2005). PCA and PLS with very large data sets. *Computational Statistics and Data Analysis*, 48, 69-85.
- [30] Stahle L and Wold S. (1998). 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, 25, 291-338.
- [31] Sharma S, Prabhakar YS, Singh P and Sharma BK. (2008). QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach. *European Journal of Medicinal Chemistry*, 43, 2354-2360.
- [32] Sharma S, Sharma BK, Sharma SK, Singh P and Prabhakar YS. (2009). Topological descriptors in modeling the agonistic activity of human A<sub>3</sub> adenosine receptor ligands: The derivatives of 2-Chloro-N<sup>6</sup>-substituted-4'-thioadenosine-5'-uronamide. *European Journal of Medicinal Chemistry*, 44, 1377-1382.
- [33] Sharma BK, Pilania P, Singh P and Prabhakar YS. (2010). Combinatorial protocol in multiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. *SAR and QSAR in Environmental Research*, 21, 169-185.
- [34] Sharma BK, Singh P, Sarbhai K and Prabhakar YS. (2010). A quantitative structure-activity relationship study on serotonin 5-HT<sub>6</sub> receptor ligands: Indolyl and piperidinyl sulphonamides. *SAR and QSAR in Environmental Research*, 21, 369-388.

- [35] Sharma BK, Pilia P, Sarbhai K, Singh P and Prabhakar YS. (2010). Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. *Molecular Diversity*, 14, 371-384.
- [36] Sharma BK, Singh P, Shekhawat M, Sarbhai K and Prabhakar YS. (2011). Modelling of serotonin reuptake inhibitory and histamine H<sub>3</sub> antagonistic activity of piperazine and diazepam amides: QSAR rationales for co-optimization of the activity profiles. *SAR and QSAR in Environmental Research*, 22, 365-383.
- [37] So S-S and Karplus M. (1997). Three-dimensional quantitative structure activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations. *Journal of Medicinal Chemistry*, 40, 4347-4359.
- [38] Prabhakar YS, Solomon VR, Rawal RK, Gupta MK and Katti SB. (2004). 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*, 23, 234-244.
- [39] Gramatica P. (2007). Principles of QSAR models validation: internal and external. *QSAR and Combinatorial Science*, 26, 694-701.
- [40] Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM and Gramatica P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. *Environmental and Health Perspectives*, 111, 1361-1375.
- [41] Golbraikh A and Tropsha A. (2002). Beware of q<sup>2</sup>! *Journal of Molecular Graphics and Modeling*, 20, 269-276.

---

### How to cite this article

Parihar R and Sharma BK. (2020). QSAR rationales for the dipeptidyl peptidase-4 (DPP-4) inhibitors: The imidazolopyrimidine amides. *GSC Biological and Pharmaceutical Sciences*, 11(3), 130-144.

---