Journal of Innovations in Pharmaceutical and Biological Sciences (JIPBS)
ISSN: 2349-2759
Available online at www.jipbs.com

Research article

3D-descriptors in modeling the dipeptidyl peptidase IV (DPP-4) inhibition activity: Derivatives of β-Aminoamide bearing subsituted triazolopiperazines

Raghuraj Parihar, Afsar Jahan, Brij Kishore Sharma*

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

Received on: 08/07/2020, Revised on: 19/07/2020, Accepted on: 26/07/2020, Published on: 08/08/2020.

***Corresponding Author: Brij Kishore Sharma,** Department of Chemistry, Government College, Bundi-323 001 (Rajasthan), India.
Email id: bksharma_sikar@rediffmail.com

Keywords: QSAR, DPP-4 and DPP-8 inhibitors, Combinatorial protocol in multiple linear regression (CP-MLR) analysis, 3D-Dragon descriptors, Triazolopiperazines.

Vol. 7 (3): 39-53, Jul-Sep, 2020.

Abstract

The DPP-4 and DPP-8 inhibitory activity of triazolopiperazines have been quantitatively analyzed in terms of 3D-Dragon descriptors. The derived QSAR models have shown that atomic properties played pivotal role in terms of weighted radial distribution functions, 3D-MoRSE signals, component symmetry directional WHIM index and moment expansions. The CP-MLR indentified RDF, 3D-MoRSE, WHIM and GETAWAY descriptors, unweighted or weighted with atomic properties endow relevant molecular 3D informations about molecular size, shape, symmetry, atom distribution, effective position of substituents and fragments in the molecular space, hold promise for rationalizing the DPP-4 and DPP-8 inhibitory actions of triazolopiperazines. The values of statistical parameters, Q^2_{LOO} and r^2_{Test} ensure that the models have validated internally and externally, both and the predictions are reliable and acceptable. 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.

Introduction

Elevated plasma glucose in the presence of high endogenous insulin levels is characteristic of type 2 diabetes (T2D). T2D is a chronic disease and causes serious vascular complications, significant morbidity, and mortality. This metabolic disorder may be considered as a growing public health problem [1]. The T2D therapies, which increase the concentration of circulating insulin, are favorable therapeutically but show some undesirable side effects like weight gain and hypoglycemia [2]. A new potential approach for the treatment of T2D is based on inhibition of a serine protease the dipeptidyl peptidase IV (DPP-4[a]) [3-6]. DPP-4 inhibitors are indirect stimulators of insulin secretion and this stimulation is mediated by boosting the action of the incretin hormone glucagon-like peptide 1 (GLP-1). Ingestion of food releases this hormone in the gut. The stimulation of insulin biosynthesis and secretion and inhibition of release of glucagon is caused, in turn, by GLP-1 [7-13]. The GLP-1 therapy is beneficiary due to the regulation of insulin in a strictly glucose-dependent manner. Little or no risk of hypoglycemia, slowing down of gastric emptying [14, 15] and reduction of appetite [16] are the beneficial effects of GLP-1 therapy.

A potential role in restoration of β-cell function in rodents point out that this mechanism may actually slow or even reverse disease progression [17-22]. DPP-4 degrades GLP-1, which cleaves a dipeptide from the N-terminus to give the inactive GLP-1[9–36] amide [23, 24]. As a result of inhibition of DPP-4 the half-life of GLP-1 is increased and thus the beneficial effects of this incretin hormone are prolonged. Sitagliptin [25, 26], LAF-237 [27] and BMS-477118 [28] are examples of DPP-4 inhibitors. Detailed structure–activity relationships (SARs) of Sitagliptin scaffold as DPP-4 inhibitors are

reported in literature with a variety of substituents on the left phenyl and the right triazolopiperazine [29]. Alkyl substitution around the β-aminoamide backbone was found to be detrimental to potency. Other modifications such as lengthening, shortening, or tethering along with alkyl substitution of triazolopiperazine series were discarded due to the similar ineffective SAR trends of corresponding thiazolidine [30] and the piperazine series [31]. A series of β-aminoamides bearing triazolopiperazines having alkyl substitutions around the triazolopiperazine moiety has been reported by Kim *et al.* [32]. 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.

Materials and methods

Biological actions and theoretical molecular descriptors

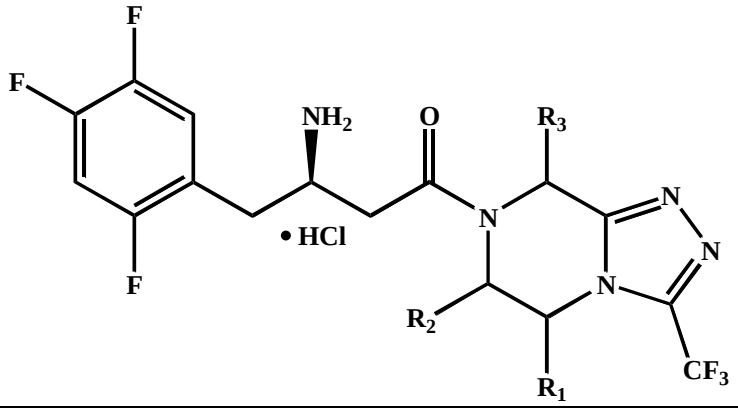
The reported thirtynine β-aminoamides bearing triazolopiperazine derivatives are considered as the data set for this study [32]. The structural variations of these analogues are mentioned in Table 1. These compounds were evaluated in vitro for their inhibition of DPP-4 and DPP-8. The reported inhibitory activity in terms of IC_{50}(nM) of these congeners is also presented in Table 1. For modeling purpose the data set has been sub-divided into training set (for model development) and test set (for external prediction or validation). The selection of test set compounds was made using an in-house written randomization program. The test and training set compounds are also mentioned in Table 1.

The structures of the all the compounds (listed in Table 1) were drawn in 2D ChemDraw [33] 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 3D-molecular descriptors of titled compounds were computed using DRAGON software [34]. This software offers a large number of descriptors corresponding to eight different classes of 3D-descriptor modules. The different 3D-descriptor classes include charge descriptors, aromaticity indices, Randic molecular profiles, geometrical descriptors, RDF descriptors, 3D-MoRSE descriptors, WHIM descriptors and GETAWAY descriptors. These descriptors are characteristic to the molecules under multi-descriptor environment. A total number of 673 descriptors, belonging to 3D-modules, have been computed to obtain most appropriate models describing the biological activity. Prior to model development procedure, all those descriptors that are intercorrelated beyond 0.90 and showing a correlation of <0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. This procedure has reduced the total descriptors from 673 to 158 as relevant ones to explain the biological actions of titled compounds.

Development and validation of model

The combinatorial protocol in multiple linear regression (CP-MLR) [35-39] and partial least squares (PLS) [40-42] 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 [43, 44]. 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 [45, 46] by repeated randomization (100 simulation runs) of the biological responses. The datasets with randomized response vector have been reassessed by multiple regression analysis. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. This has been used as a measure to express the percent chance correlation of the model under scrutiny. Validation of the derived model is necessary to test its prediction and generalization within the study domain. For each model, derived by involving n data points, a number of statistical parameters such as r (the multiple correlation coefficient), s (the standard deviation), F (the F ratio between the variances of calculated and observed activities), and Q^2_{LOO} (the cross-validated index from leave-one-out procedure) have been obtained to access its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5.

Table 1. Structural variations and reported DPP-4 and DPP-8 inhibitory activities of triazolopiperazines.



Cpd.	R1	R2	R3	IC_{50}(nM)[a]	
				DPP-4	DPP-8
1[b]	H	H	H	18	48000
2	(*S*)-CH_3	H	H	23	23000
3	(*R*)-CH_3	H	H	14	33000
4	H	(*S*)-CH_3	H	91	>100000[c]
5	H	(*R*)-CH_3	H	42	75000
6[b]	H	H	(*S*)-CH_3	88	>100000[c]
7	H	H	(*R*)-CH_3	4.3	17000
8	*di*-CH_3	H	H	92	66000
9	H	H	*di*-CH_3	175	6000
10[b]	CH_3	H	CH_3	100	>100000[c]
11[b]	CH_3	H	CH_3	209	>100000[c]
12	CH_3	H	CH_3	12	70000
13	CH_3	H	CH_3	11	44000
14	H	H	Et	113	>100000[c]
15	H	H	Et	5	8000
16	H	H	CH_2CF_3	123	>100000[c]
17	H	H	CH_2CF_3	5.7	1600
18	H	H	$CH_2CH=CH_2$	1.5	3000
19	H	H	$CH_2CH=CH_2$	32	72000
20	H	H	$CH_2CON(CH_3)_2$	377	>100000[c]
21[b]	H	H	$CH_2CON(CH_3)_2$	2.8	30000
22	H	H	CH_2Ph	140	>100000[c]
23	H	H	CH_2Ph	0.66	622
24	H	H	CH_2(4-methoxyphenyl)	320	>100000[c]
25	H	H	CH_2(4-methoxyphenyl)	0.43	367
26[b]	H	H	CH_2(2-trifluoromethylphenyl)	438	>100000[c]
27[b]	H	H	CH_2 (2-trifluoromethylphenyl)	0.31	8000
28[b]	H	H	CH_2(2-fluorophenyl)	131	>100000[c]
29[b]	H	H	CH_2(2-fluorophenyl)	0.46	1103
30	H	H	CH_2(4-fluorophenyl)	116	>100000[c]
31	H	H	CH_2(4-fluorophenyl)	0.18	332
32	H	H	CH(OH)(4-fluorophenyl)	430	>100000[c]
33	H	H	CH(OH)(4-fluorophenyl)	0.32	326
34	H	H	CH(OH)(4-fluorophenyl)	90	40000
35	H	H	CH(OH)(4-fluorophenyl)	0.5	628
36	H	H	CH_2(3,5-*bis*-trifluoromethylphenyl)	587	>100000[c]
37[b]	H	H	CH_2 (3,5-*bis*-trifluoromethylphenyl)	6.3	>100000[c]
38	H	H	CH_2(2-pyridyl)	132	>100000[c]
39	H	H	CH_2(2-pyridyl)	0.4	5000

[a]Concentration of a compound to bring out 50% inhibition (IC_{50}), taken from reference [32]; [b]Compound included in test set; [c]Compound with uncertain activity, not part of data set.

Additional statistical parameters such as, the Akaike's information criterion, AIC [47, 48], the Kubinyi function, FIT [49, 50] and the Friedman's lack of fit, LOF [51], 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.

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 [52]. 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.

Results and discussion

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, 10 compounds have been included in the test set for the validation of the models derived from 29 training set compounds. A total number of 156 significant descriptors from 3D-classe have been subjected to CP-MLR analysis with default "filters" set in it. Statistical models in two, three and four descriptor(s) have been derived successively to achieve the best relationship correlating DPP-4 inhibitory activity. A total number of 10, 20 and 22 models in two, three and four descriptors, respectively, were obtained through CP-MLR. These models (with 158 descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this, the optimum r-bar value of the preceding level model has been used as the new threshold of filter-3 for the next generation. The selected models in two, three and four descriptors are given below.

$pIC_{50} = 7.563 – 3.848(0.768)RDF075m + 6.324(1.145)RDF085m$
$n = 29, r = 0.765, s = 0.722, F = 18.405, Q^2_{LOO} = 0.499, Q^2_{L5O} = 0.503$
$r^2_{Test} = 0.372$, FIT = 1.115, LOF = 0.629, AIC = 0.642 (1)

$pIC_{50} = 7.642 – 3.011(0.737)RDF075m + 2.883(0.568)RDF105p$
$n = 29, r = 0.740, s = 0.754, F = 15.752, Q^2_{LOO} = 0.469, Q^2_{L5O} = 0.465$
$r^2_{Test} = 0.165$, FIT = 0.954, LOF = 0.687, AIC = 0.701 (2)

$pIC_{50} = 8.702 – 5.173(0.780)RDF075m + 6.241(0.985)RDF085m – 1.785(0.560)G3e$
$n = 29, r = 0.840, s = 0.621, F = 19.979, Q^2_{LOO} = 0.615, Q^2_{L5O} = 0.603$
$r^2_{Test} = 0.646$, FIT = 1.577, LOF = 0.528, AIC = 0.509 (3)

$pIC_{50} = 10.050 – 2.942(0.734)DISPv + 3.853(0.976)RDF085m – 3.848(0.602)RDF110e$
$n = 29, r = 0.833, s = 0.631, F = 19.029, Q^2_{LOO} = 0.588, Q^2_{L5O} = 0.538$
$r^2_{Test} = 0.505$, FIT = 1.502, LOF = 0.547, AIC = 0.526 (4)

$pIC_{50} = 5.359 –2.502(0.665)RDF075m + 4.134(0.557)RDF085p + 1.904(0.474)Mor10m$
$+ 2.268(0.534)RTu+$
$n = 29, r = 0.889, s = 0.534, F = 22.626, Q^2_{LOO} = 0.696, Q^2_{L5O} = 0.615$
$r^2_{Test} = 0.684$, FIT = 2.011, LOF = 0.451, AIC = 0.405 (5)

pIC_{50} = 6.016 –1.956(0.525)RDF110e + 2.590(0.414)RDF105p + 2.139(0.497)Mor10m
+ 1.442(0.516)RTp+
n = 29, r = 0.877, s = 0.560, F = 20.083, Q^2_{LOO} = 0.621, Q^2_{L5O} = 0.633
r^2_{Test} = 0.525, FIT = 1.785, LOF = 0.495, AIC = 0.444 (6)

In above and all follow-up regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation.

Most of the participated descriptors RDF075m, RDF085m, RDF110e, RDF085p and RDF105p in above models belong to RDF descriptors class. RDF (Radial Distribution Function) descriptors [53] are based on a radial distribution function that may be considered as the probability distribution of finding an atom in a spherical volume of radius (R). The RDF descriptors are represented as RDFkw where k is step size and w is weighting scheme such as the unweighted case (u), atomic mass (m), the van der Waals volume (v), the Sanderson atomic electronegativity (e) and the atomic polarizability (p). The RDF descriptors not only provide information about interatomic distances in the whole molecule but bond distances, atom types, ring types and planar and non-planar systems also. Atomic masses weighted radial distribution function 7.5 (descriptor RDF075m) and atomic Sanderson electronegativities weighted radial distribution functions 11.0 (descriptor RDF110e) contributed negatively to the activity suggesting that a higher values of these radial distribution functions would be detrimental to DPP-4 inhibition actions. On the other hand positive contribution of radial distribution function-8.5/ weighted by atomic masses (descriptor RDF085m) and atomic polarizabilities weighted radial distribution functions 8.5 and 10.5 (RDF085p and RDF105p, respectively) advocated a higher value of these to augment the inhibitory activity.

Descriptor DISPv is representative of geometrical class of descriptors. Geometrical descriptors are derived from the three-dimensional structure of the molecule and calculation of these is based on some optimized molecular geometry obtainable by the methods of the computational chemistry or on crystallographic coordinates. Geometrical descriptors offer more information and discrimination power for similar molecular structures and molecule conformations because a geometrical representation of a molecule involves the knowledge of the relative positions of the atoms in 3D space. Descriptor DISPv is among the COMMA2 descriptors [54]. COMMA2 descriptors are given by moment expansions for which the zero-order moment of a considered property (such as mass (m), van der Waals volume (v), Sanderson electronegativity (e) and polarizability (p)) field is non-vanishing. The negative contribution of descriptor DISPv (the displacement between the geometric centre and the centre of the van der Waals volume field, calculated with respect to the molecular principal axes) hints that a lower value of it would be beneficiary to the activity.

Descriptor Mor10m is a 3D-MoRSE (3D-Molecule Representation of Structures based on Electron diffraction) descriptor [55]. These descriptors (Mor*sw*) represent the scattered electron intensity (signals). The term *s* represents the scattering in various directions by a collection of atoms and *w* is the atomic property or may be unweighted case. The positive contribution of 3D-MoRSE - signal 10/ weighted by atomic masses (Mor10m) suggests that a higher value of it would be incremental to the activity. Descriptors RTu+ and RTp+ are from the GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) class of descriptors. GETAWAYs [56] are geometrical descriptors which encode information on the effective position of substituents and fragments in the molecular space. These descriptors are independent of molecule alignment and account for information on molecular size and shape and for specific atomic properties. Both the descriptors RTu+ (unweighted R maximal index) and RTp+ (atomic polarizabilities weighted R maximal index) shown positive correlation to the activity advocating higher values of these for augmented activity. The remaining descriptor G3e is a Weighted Holistic Invariant Molecular (WHIM) descriptor. These descriptors are geometrical descriptors and are based on statistical indices calculated on the projections of the atoms along principal axes [57]. WHIM descriptors are free from prior alignment of molecules because these are invariant to translation and rotation. WHIM descriptors (categorized as directional and global) furnish relevant molecular 3D information about molecular size, shape, symmetry, and atom distribution with respect to invariant reference frames. The appeared WHIM descriptor, G3e (3rd component symmetry directional WHIM index/weighted by atomic Sanderson electronegativities) correlated negatively to the activity suggesting lower value of it for elevated DPP-4 activity.

The four descriptor models could estimate nearly 79% in observed activity of the compounds. Considering the number of observations in the dataset, models with up to five descriptors were explored. It has resulted in 4 five-parameter models with test set $r^2 > 0.50$. These models have shared 10 descriptors among them. All these 10 descriptors along with their brief meaning, average

regression coefficients, and total incidence are listed in Table 2, which will serve as a measure of their estimate across these models. Following are the emerged five-descriptor models for the DPP-4 inhibitory activity of titled compounds.

pIC_{50} = 7.266 – 1.712(0.685)DISPv – 2.653(0.634)RDF110e + 3.126(0.567)RDF085p
+ 1.564(0.507)Mor10m + 1.795(0.570)RTu+
n = 29, r = 0.900, s = 0.518, F = 19.774, Q^2_{LOO} = 0.668, Q^2_{L5O} = 0.680
r^2_{Test} = 0.606, FIT = 1.830, LOF = 0.496, AIC = 0.409 (7)

pIC_{50} = 5.240 – 2.398(0.595)RDF110e + 3.094(0.581)RDF085p + 2.392(0.481)Mor10m
+ 1.436(0.603)Mor12m + 1.740(0.586)RTu+
n = 29, r = 0.898, s = 0.523, F = 19.296, Q^2_{LOO} = 0.715, Q^2_{L5O} = 0.726
r^2_{Test} = 0.512, FIT = 1.786, LOF = 0.506, AIC = 0.417 (8)

pIC_{50} = 6.452 – 2.969(0.510)RDF110e – 2.016(0.719)RDF155e + 2.402(0.505)RDF085p
+ 2.802(0.501)Mor10m + 1.809(0.580)Mor12m
n = 29, r = 0.895, s = 0.531, F = 18.573, Q^2_{LOO} = 0.711, Q^2_{L5O} = 0.690
r^2_{Test} = 0.735, FIT = 1.719, LOF = 0.522, AIC = 0.430 (9)

pIC_{50} = 7.507 – 4.112(0.760)RDF075m + 5.814(0.901)RDF085m + 1.151(0.497)Mor10m
– 1.879(0.485)G3e + 1.303(0.500)RTu+
n = 29, r = 0.894, s = 0.533, F = 18.437, Q^2_{LOO} = 0.611, Q^2_{L5O} = 0.638
r^2_{Test} = 0.670, FIT = 1.707, LOF = 0.525, AIC = 0.432 (10)

These models have accounted for nearly 81% variance in the observed activities. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The values greater than 0.5 of Q^2 index is in accordance to a reasonable robust QSAR model. The pIC_{50} values of training set compounds calculated using Eqs. (7) to (10) and predicted from LOO procedure have been included in Table 3. The models (7) to (10) are validated with an external test set of 10 compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the same is reported in Table 3. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

The newly appeared descriptors in above models are RDF155e (a RDF class descriptor) and Mor12m (from 3D-MoRSE class). The signs of regression coefficients of these descriptors suggest that a lower value of descriptor RDF155e (radial distribution function – 15.5/weighted by atomic Sanderson electronegativities) and a higher value of 3D-MoRSE - signal 12/ weighted by atomic masses (descriptor Mor12m) would be incremental to the activity. In this way the descriptors identified for rationalizing the activity give avenues to modulate the structure to a desirable biological endpoint.

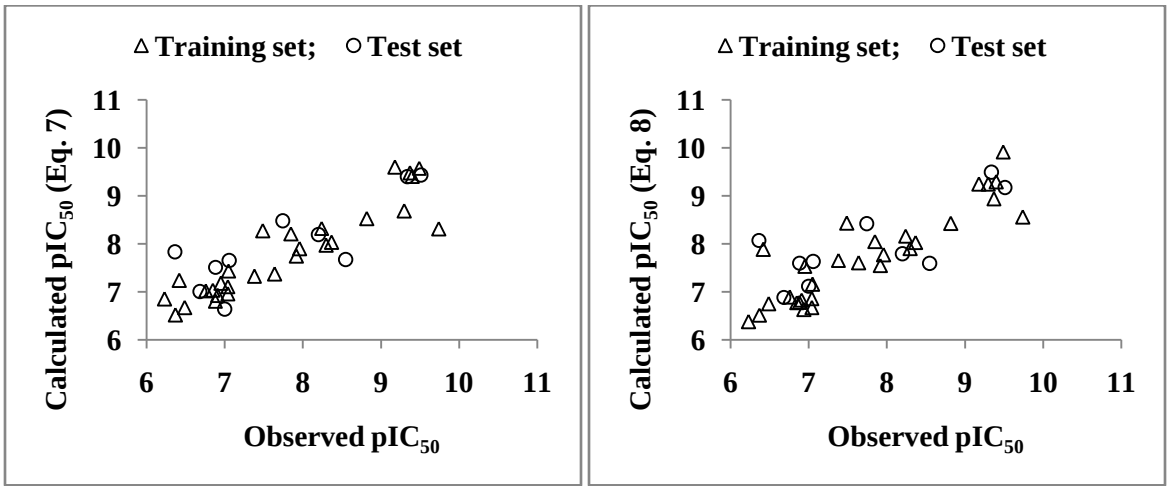
A partial least square (PLS) analysis has been carried out on these 10 CP-MLR identified descriptors (Table 2) to facilitate the development of a "single window" structure–activity model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, three components are found to be the optimum for these 10 descriptors and they explained 93.9% variance in the activity (r^2 = 0.939, Q^2_{LOO} = 0.853, s = 0.391, F = 62.879, r^2_{Test} = 0.763). The MLR-like PLS coefficients of these 10 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.

The PLS analysis has suggested RDF110e as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 4 in Table 4; Figure 2). The other nine significant descriptors in decreasing order of significance are Mor10m, RDF075m, RDF085p, RDF085m, RTU+, G3e, Mor12m, DISPv and RDF155e. All descriptors are part of Eqs. (1) to (10) and convey same inference in the PLS model as well. It is also observed that PLS model from the dataset devoid of 10 descriptors (Table 2) is inferior in explaining the activity of the analogues.

Table 2. Identified descriptors[a] along with their physical meaning, average regression coefficient and incidence[b], in modeling the DPP-4 inhibitory activity of triazolopiperazines.

S. No.	Descriptor class	Descriptor	Physical meaning, average regression coefficient (**incidence**)
1	Geometrical descriptors	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes -1.712(**1**)
2	RDF descriptors	RDF075m	Radial distribution function at 7.5 Å / weighted by atomic masses -4.112(**1**)
3		RDF085m	Radial Distribution Function at 8.5 Å / weighted by atomic polarizabilities 5.814(**1**)
4		RDF110e	Radial Distribution Function at 11.0 Å / weighted by atomic Sanderson electronegativities -2.674(**3**)
5		RDF155e	Radial Distribution Function at 15.5 Å / weighted by atomic Sanderson electronegativities -2.016(**1**)
6		RDF085p	Radial Distribution Function at 8.5 Å / weighted by atomic polarizabilities 2.874(**3**)
7	3D-MoRSE descriptors	Mor10m	3D-MoRSE - signal 10 / weighted by atomic masses 1.977(**4**)
8		Mor12m	3D-MoRSE - signal 12 / weighted by atomic masses 1.622(**2**)
9	WHIM descriptors	G3e	3rd component symmetry directional WHIM index /weighted by atomic Sanderson electronegativities -1.879(**1**)
10	GETAWAY descriptor	RTu+	R maximal index / unweighted 1.613(**3**)

[a]The descriptors are identified from the five parameter models, emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.869, and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 29 compounds. [b]The average regression coefficient of the descriptor corresponding to all models and the total number of its incidence. The arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.



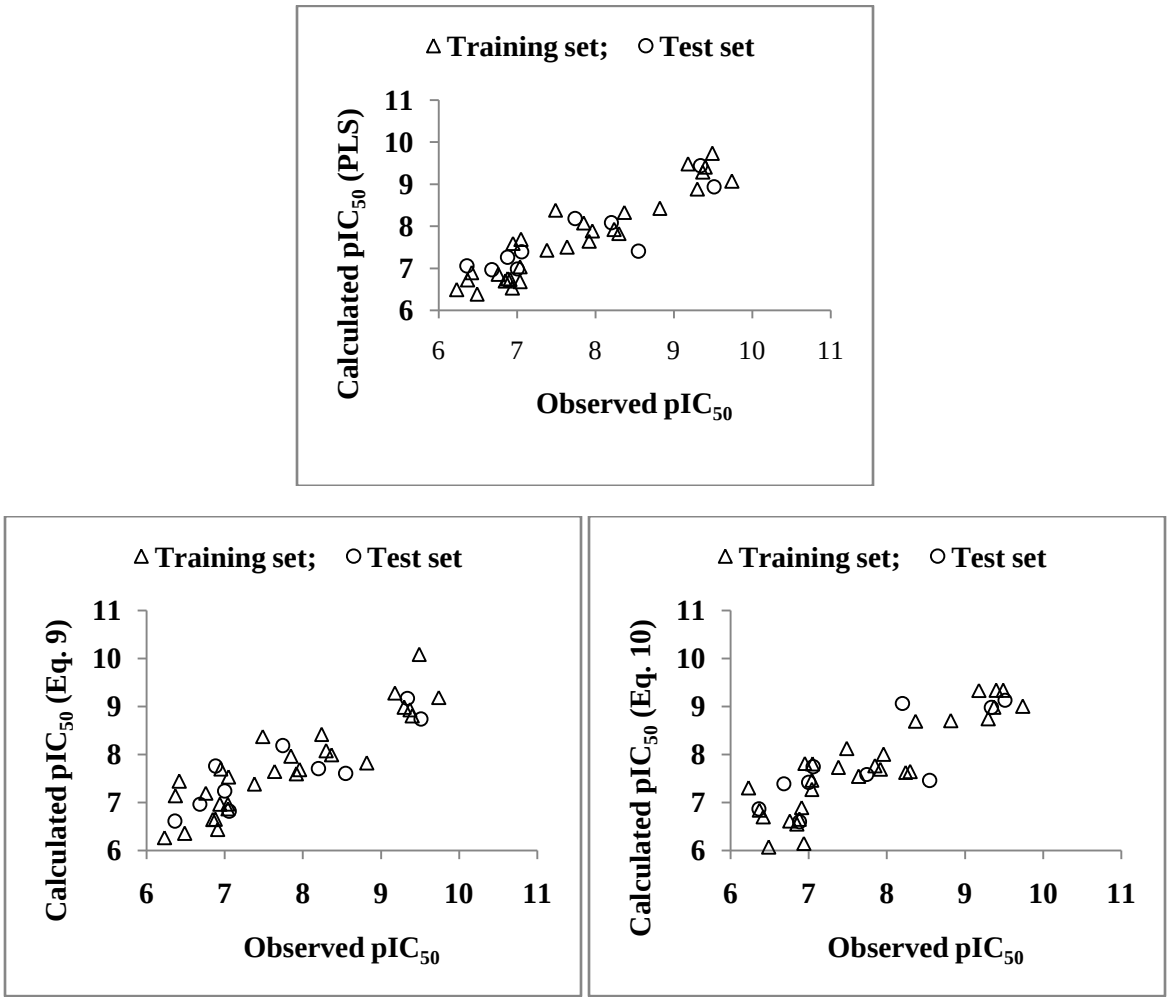


Figure 1. Plot of observed and calculated pIC_{50} values of training- and test-set compounds for DPP-4.

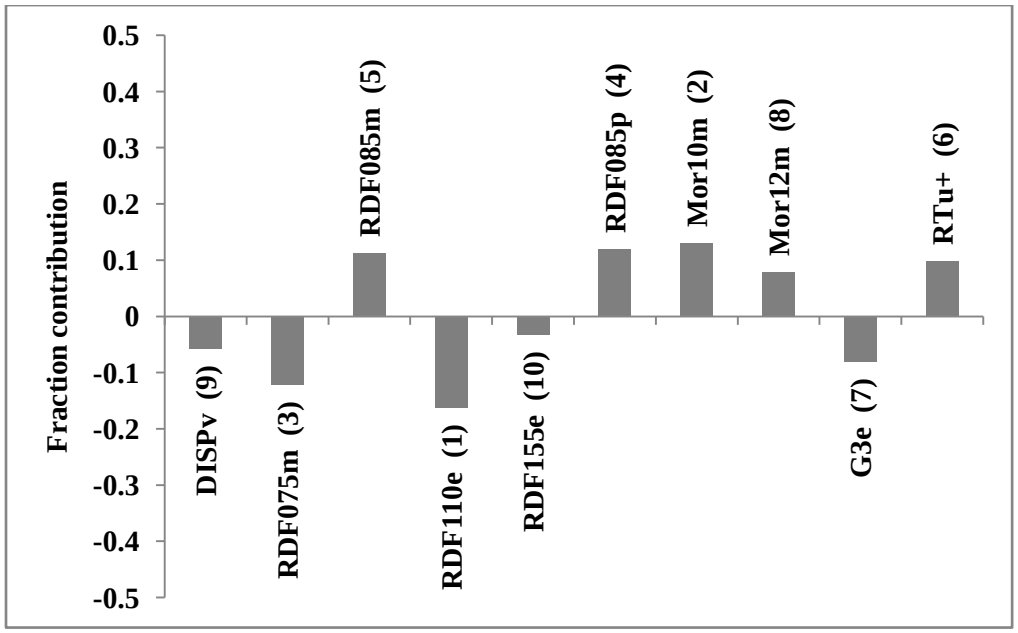


Figure 2. Plot of fraction contribution of MLR-like PLS coefficients (normalized) against ten CP-MLR identified descriptors (Table 4) associated with DPP-4 inhibitory activity of triazolopiperazines.

Table 3. Observed and modeled DPP-4 inhibitory activities of triazolopiperazines.

S. No.	pIC_{50}(M)[a]										
		Eq. (7)		Eq. (8)		Eq. (9)		Eq. (10)		PLS	
	Obsd[b]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]
1[d]	7.74	8.48	-[d]	8.42	-[d]	8.19	-[d]	7.59	-[d]	8.18	-[d]
2	7.64	7.37	7.32	7.60	7.60	7.65	7.65	7.54	7.53	7.50	7.49
3	7.85	8.21	8.29	8.05	8.08	7.96	7.98	7.76	7.75	8.08	8.11
4	7.04	6.96	6.94	6.67	6.52	6.87	6.79	7.27	7.28	6.68	6.64
5	7.38	7.32	7.31	7.65	7.68	7.38	7.38	7.73	7.76	7.43	7.44
6[d]	7.06	7.65	-[d]	7.63	-[d]	6.82	-[d]	7.74	-[d]	7.39	-[d]
7	8.37	8.03	7.98	8.03	7.97	7.99	7.94	8.69	8.75	8.33	8.33
8	7.04	7.11	7.15	6.86	6.73	6.97	6.93	7.47	7.70	7.04	7.04
9	6.76	7.02	7.17	6.89	6.95	7.19	7.31	6.61	6.54	6.86	6.87
10[d]	7.00	6.64	-[d]	7.12	-[d]	7.24	-[d]	7.42	-[d]	6.99	-[d]
11[d]	6.68	7.00	-[d]	6.89	-[d]	6.97	-[d]	7.39	-[d]	6.97	-[d]
12	7.92	7.74	7.72	7.55	7.50	7.60	7.56	7.69	7.67	7.65	7.62
13	7.96	7.90	7.89	7.77	7.75	7.68	7.65	8.00	8.01	7.89	7.87
14	6.95	7.18	7.20	7.53	7.57	7.70	7.75	7.81	7.87	7.59	7.63
15	8.30	7.97	7.94	7.90	7.87	8.08	8.06	7.64	7.60	7.83	7.80
16	6.91	6.92	6.93	6.83	6.81	6.44	6.36	6.89	6.89	6.74	6.72
17	8.24	8.32	8.33	8.16	8.15	8.42	8.44	7.63	7.36	7.92	7.86
18	8.82	8.52	8.42	8.42	8.27	7.82	7.74	8.70	8.66	8.42	8.39
19	7.49	8.27	8.36	8.43	8.56	8.37	8.49	8.12	8.21	8.38	8.45
20	6.42	7.24	7.71	7.88	8.03	7.44	7.68	6.70	6.77	6.89	6.94
21[d]	8.55	7.67	-[d]	7.59	-[d]	7.60	-[d]	7.45	-[d]	7.40	-[d]
22	6.85	7.03	7.06	6.77	6.76	6.64	6.61	6.55	6.50	6.70	6.68
23	9.18	9.60	9.72	9.24	9.26	9.28	9.32	9.33	9.37	9.48	9.54
24	6.49	6.67	6.74	6.75	6.85	6.36	6.27	6.07	5.91	6.39	6.35
25	9.37	9.48	9.54	8.94	8.81	8.93	8.80	8.98	8.87	9.29	9.24
26[d]	6.36	7.83	-[d]	8.07	-[d]	6.61	-[d]	6.86	-[d]	7.06	-[d]
27[d]	9.51	9.43	-[d]	9.17	-[d]	8.74	-[d]	9.13	-[d]	8.94	-[d]
28[d]	6.88	7.51	-[d]	7.60	-[d]	7.76	-[d]	6.60	-[d]	7.26	-[d]
29[d]	9.34	9.40	-[d]	9.49	-[d]	9.17	-[d]	8.98	-[d]	9.44	-[d]
30	6.94	5.96	5.68	6.63	6.44	6.96	6.98	6.15	5.96	6.53	6.48
31	9.74	8.31	8.02	8.56	8.19	9.19	9.04	9.00	8.70	9.07	8.97
32	6.37	6.52	6.54	6.51	6.53	7.14	7.28	6.84	6.96	6.73	6.82
33	9.49	9.57	9.62	9.91	10.11	10.09	10.35	9.34	9.30	9.73	9.81
34	7.05	7.43	7.50	7.16	7.17	7.53	7.58	7.80	7.95	7.69	7.72
35	9.30	8.68	8.24	9.24	9.20	8.98	8.71	8.75	8.36	8.89	8.78
36	6.23	6.85	7.24	6.38	6.66	6.27	6.33	7.30	8.24	6.49	6.60
37[d]	8.20	8.19	-[d]	7.79	-[d]	7.71	-[d]	9.07	-[d]	8.08	-[d]
38	6.88	6.80	6.79	6.78	6.76	6.65	6.61	6.65	6.61	6.75	6.74
39	9.40	9.40	9.40	9.29	9.27	8.80	8.64	9.34	9.32	9.41	9.41

[a]On molar basis; [b]Taken from ref. [32]; [c]Leave-one-out (LOO) procedure; [d]Compound included in test set.

Table 4. PLS and MLR-like PLS models from the descriptors of five parameter CP-MLR models for DPP-4 inhibitory activity.

A: PLS equation			
PLS components		PLS coefficient (s.e.)[a]	
Component-1		-0.859(0.064)	
Component-2		0.091(0.036)	
Component-3		0.182(0.069)	
Constant		7.737	
B: MLR-like PLS equation			
S. No.	Descriptor	MLR-like coefficient (f.c.)[b]	Order
1	DISPv	-0.159(-0.057)	9
2	RDF075m	-0.339(-0.122)	3
3	RDF085m	0.314(0.113)	5
4	RDF110e	-0.449(-0.162)	1
5	RDF155e	-0.092(-0.033)	10
6	RDF085p	0.333(0.120)	4
7	Mor10m	0.361(0.130)	2
8	Mor12m	0.220(0.079)	8
9	G3e	-0.226(-0.081)	7
10	RTu+	0.272(0.098)	6
	Constant	7.196	
C: PLS regression statistics		Values	
n		29	
r		0.939	
s		0.391	
F		62.879	
FIT		4.964	
LOF		0.210	
AIC		0.202	
Q^2_{LOO}		0.853	
Q^2_{L5O}		0.858	
r^2_{Test}		0.763	

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

The other inhibitory activity reported for DPP-8 enzyme system has also analyzed quantitatively. A total number of 10 models in two parameters and 46 models in three parameters, having $r^2_{Test} > 0.5$, were obtained on applying CP-MLR. For the sake of brevity, highly significant four models in three parameters emerged through CP-MLR are shown below.

$pIC_{50} = 3.193 + 2.361(0.305)RDF085m – 2.142(0.718)RDF150p + 1.732(0.582)Mor10p$
$n = 19, r = 0.915, s = 0.390, F = 25.793, Q^2_{LOO} = 0.741, Q^2_{L5O} = 0.738$
$r^2_{Test} = 0.523, FIT = 2.763, LOF = 0.257, AIC = 0.234$ (11)

$pIC_{50} = 3.847 + 2.302(0.325)RDF085m – 1.778(0.726)RDF150p + 1.237(0.479)Mor10m$
$n = 19, r = 0.906, s = 0.409, F = 23.001, Q^2_{LOO} = 0.747, Q^2_{L5O} = 0.724$
$r^2_{Test} = 0.697, FIT = 2.464, LOF = 0.283, AIC = 0.257$ (12)

$pIC_{50} = 7.192 + 2.588(0.594)RDF115m – 5.919(1.176)Mor23m – 1.685(0.728)E3e$
$n = 19, r = 0.905, s = 0.411, F = 22.788, Q^2_{LOO} = 0.709, Q^2_{L5O} = 0.756$
$r^2_{Test} = 0.577, FIT = 2.441, LOF = 0.285, AIC = 0.259$ (13)

$pIC_{50} = 9.741 – 1.025(0.450)RDF145p – 8.388(1.092)Mor23m – 1.702(0.442)H5m$
$n = 19, r = 0.900, s = 0.422, F = 21.355, Q^2_{LOO} = 0.703, Q^2_{L5O} = 0.719$
$r^2_{Test} = 0.574, FIT = 2.288, LOF = 0.300, AIC = 0.273$ (14)

These models are able to explain nearly 84% variance in the observed DPP-8 inhibitory activities. None of the identified models has shown any chance correlation in the randomization study (100 simulations per model). The values greater than 0.5 of Q^2 index is in accordance to a reasonable internal validation and r^2_{Test} values reflect upon the good predictive power of above mentioned QSAR models. The pIC_{50} values of training and test set compounds calculated using Eqs. (11) to (14) and predicted from LOO procedure have been included in Table 5. The goodness of fit or agreement between observed and calculated activities for the training and test set compounds is shown in Figure 3.

It is evident from the models that higher values of atomic mass weighted radial distribution functions 8.5 and 11.5 (descriptors RDF085m and RDF115m, respectively), and lower values of atomic polarizability weighted radial distribution functions 14.5 and 15.0 (descriptors RDF145p and RDF150p, respectively) would supplement the activity. Atomic mass weighted 3D-MoRSE signals 10 and 23 (descriptors Mor10m and Mor23m, respectively) in addition to atomic polarizability weighted 10 (descriptor Mor10p) have shown prevalence to explain the DPP-8 inhibitory activity. A higher value of descriptors Mor10p and Mor10m is conducive to activity whereas a higher value of descriptor Mor23m is unfavorable to the activity. The negative correlation of WHIM descriptor (E3e, 3rd component accessibility WHIM index/ weighted by atomic Sanderson electronegativities) and physicochemical properties weighted spatial autocorreation GETAWAY descriptor (H5m, H autocorrelation of lag 5/weighted by atomic masses) recommended a lower value of these descriptors for elevated DPP-8 inhibitory activity.

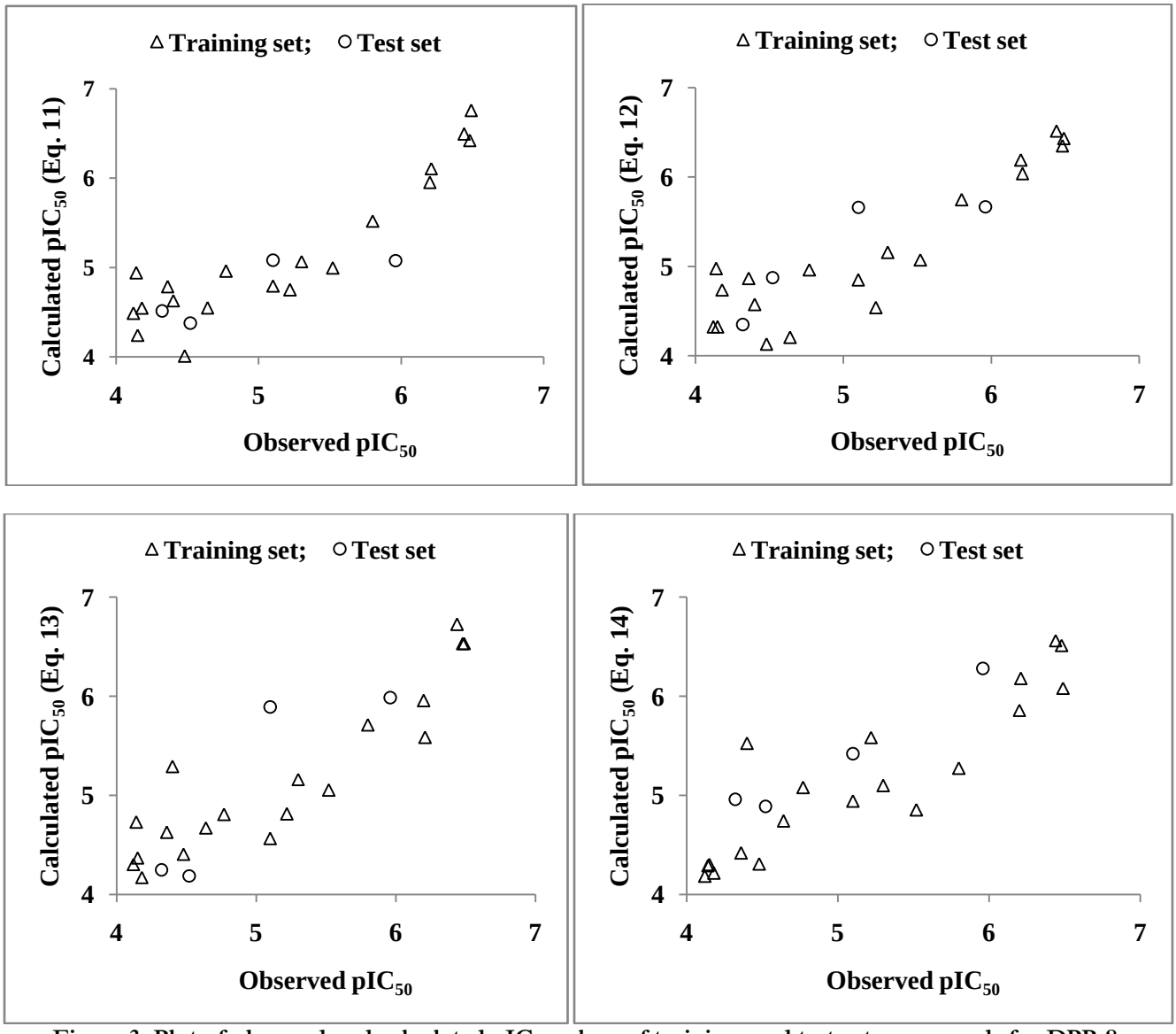


Figure 3. Plot of observed and calculated pIC_{50} values of training- and test-set compounds for DPP-8.

Table 5. Observed and modeled DPP-8 inhibitory activities of triazolopiperazines.

S. No.	$pIC_{50}(M)$[a]								
		Eq. (11)		Eq. (12)		Eq. (13)		Eq. (14)	
	Obsd[b]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]
1[d]	4.32	4.51	-[d]	4.35	-[d]	4.25	-[d]	4.96	-[d]
2	4.64	4.55	4.53	4.21	4.10	4.67	4.68	4.74	4.75
3	4.48	4.01	3.91	4.13	4.07	4.41	4.39	4.31	4.27
4[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
5	4.12	4.49	4.60	4.32	4.38	4.30	4.34	4.18	4.21
6[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
7	4.77	4.96	4.99	4.96	4.99	4.81	4.81	5.08	5.18
8	4.18	4.55	4.62	4.74	4.80	4.17	4.17	4.22	4.23
9	5.22	4.75	4.50	4.54	4.33	4.81	4.62	5.58	5.61
10[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
11[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
12	4.15	4.24	4.26	4.33	4.35	4.37	4.40	4.30	4.33
13	4.36	4.78	4.81	4.87	4.90	4.63	4.65	4.42	4.43
14[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
15	5.1	4.79	4.76	4.85	4.82	4.56	4.49	4.94	4.91
16[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
17	5.8	5.52	5.48	5.75	5.74	5.71	5.69	5.27	5.05
18	5.52	4.99	4.89	5.07	5.00	5.05	4.94	4.85	4.76
19	4.14	4.94	5.02	4.98	5.07	4.73	4.97	4.29	4.33
20[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
21[d]	4.52	4.38	-[d]	4.88	-[d]	4.19	-[d]	4.89	-[d]
22	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
23	6.21	6.11	6.09	6.04	6.00	5.58	5.51	6.18	6.17
24	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
25	6.44	6.49	6.51	6.52	6.54	6.72	6.85	6.56	6.60
26[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
27[d]	5.1	5.08	-[d]	5.66	-[d]	5.89	-[d]	5.42	-[d]
28[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
29[d]	5.96	5.08	-[d]	5.67	-[d]	5.98	-[d]	6.28	-[d]
30[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
31	6.48	6.42	6.40	6.35	6.30	6.53	6.55	6.51	6.52
32[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
33	6.49	6.76	6.92	6.43	6.42	6.53	6.56	6.08	5.80
34	4.4	4.63	4.72	4.57	4.66	5.29	5.55	5.52	5.74
35	6.2	5.95	5.88	6.19	6.18	5.96	5.89	5.86	5.80
36[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
37[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
38[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]	-[e]
39	5.3	5.07	4.76	5.16	4.96	5.16	5.13	5.10	5.00

Applicability domain (AD)

On analyzing the model AD in the Williams plot (Figure 4) of the model based on the whole dataset (Table 6), it has appeared that none of the compounds were identified as an obvious outlier for the DPP-4 inhibitory activity if the limit of normal values for the *Y* outliers (response outliers) was set as 3 (standard deviation) units. None of the compounds was found to have leverage (*h*) values greater than the threshold leverages (*h**). For both the training set and test set, the suggested model matches the high-quality parameters with good fitting power and the capability of assessing external data. Furthermore, almost all of the compounds was within the AD of the proposed model and were evaluated correctly.

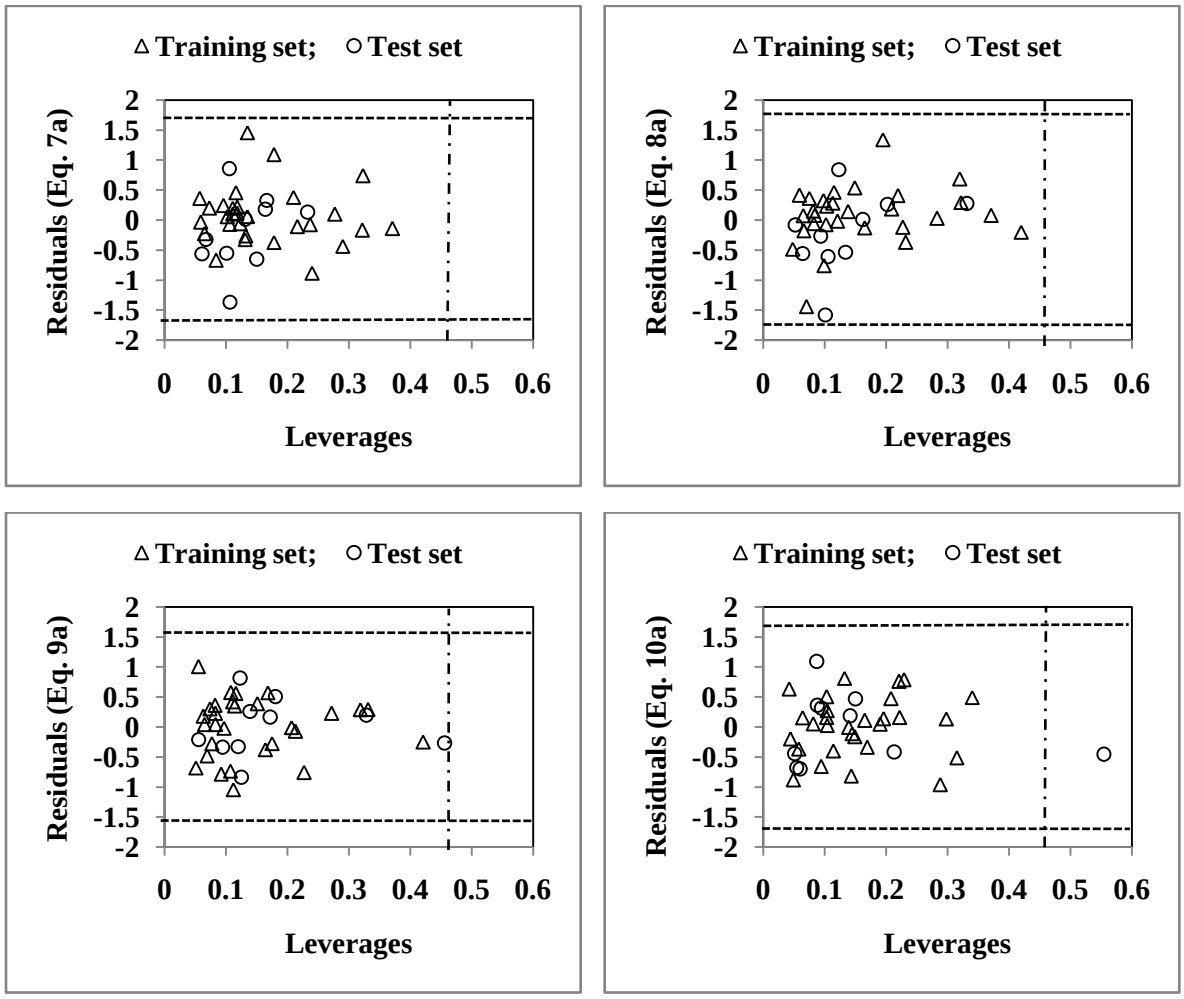


Figure 4. Williams plot for the training-set and test- set compounds for DPP-4 inhibitory activity. The horizontal dotted line refers to the residual limit (±3×standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.46).

Table 6. Models derived for the whole data set (n = 39) for the DPP-4 inhibitory activity in descriptors identified through CP-MLR.

Model	r	s	F	q^2_{LOO}	Eq.
pIC_{50} = 7.197 – 1.490(0.627)DISPv – 2.823(0.602)RDF110e + 3.198(0.527)RDF085p + 1.505(0.497)Mor10m + 1.662(0.587)RTu+	0.875	0.559	21.562	0.669	(7a)
pIC_{50} = 5.414 – 2.495(0.586)RDF110e + 3.383(0.537)RDF085p + 2.209(0.474)Mor10m + 0.864(0.530)Mor12m + 1.788(0.608)RTu+	0.863	0.582	19.383	0.657	(8a)
pIC_{50} = 6.582 – 3.073(0.427)RDF110e – 2.047(0.468)RDF155e + 2.772(0.419)RDF085p + 2.889(0.448)Mor10m + 1.321(0.459)Mor12m	0.892	0.520	25.914	0.728	(9a)
pIC_{50} = 7.778 – 4.456(0.682)RDF075m + 5.559(0.715)RDF085m + 1.223(0.463)Mor10m – 2.183(0.455)G3e + 1.251(0.485)RTu+	0.880	0.547	22.780	0.665	(10a)

Conclusions

The DPP-4 and DPP-8 inhibitory activity of triazolopiperazines have been quantitatively analyzed in terms of 3D-Dragon descriptors. The derived QSAR models have shown that atomic properties played pivotal role in terms of weighted radial distribution functions, 3D-MoRSE signals, component symmetry directional WHIM index and moment expansions. The CP-MLR indentified RDF, 3D-MoRSE, WHIM and GETAWAY descriptors weighted or unweighted with atomic properties endow relevant molecular 3D information about molecular size, shape, symmetry, atom distribution, effective position of substituents and fragments in the molecular space hold promise for rationalizing the DPP-4 and DPP-8 inhibitory actions of triazolopiperazines. The values of statistical parameters, Q^2_{LOO} and r^2_{Test} ensure that the models have validated internally and externally, both and the predictions are reliable and acceptable. 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.

Conflicts of interest

The authors declare no conflicts of interest and its original work neither communicate for publication nor under consideration for publication. Further, all authors have contributed to the work immensely.

Acknowledgement

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

References

1. King, H.; Aubert, R. E. and Herman, W. H. Global burden of diabetes, 1995–2025: Prevalence, numerical estimates and projections. Diabetes Care 1998; 21: 1414–1431.
2. Wagman, A. S. and Nuss, J. M. Current therapies and emerging targets for the treatment of diabetes. Current Pharmaceutical Design 2001; 7: 417–450.
3. Weber, A. E. Dipeptidyl peptidase IV inhibitors for the treatment of diabetes. Journal of Medicinal Chemistry 2004; 47: 4135–4141.
4. Augustyns, K.; der Veken, P. V.; Senten, K. and Haemers, A. The therapeutic potential of inhibitors of dipeptidyl peptidase IV (DPP-4) and related proline-specific dipeptidyl aminopeptidases. Current Medicinal Chemistry 2005; 12: 971–998.
5. Augustyns, K.; der Veken, P. V. and Haemers, A. Inhibitors of proline-specific dipeptidyl peptidases: DPP IV inhibitors as novel approach for the treatment of type 2 diabetes. Expert Opinion on Therapeutic Patents 2005; 15: 1387–1407.
6. von Geldern, T. W. and Trevillyan, J. M. "The next big thing" in diabetes: Clinical progress on DPP-IV inhibitors. Drug Development Research 2006; 67: 627–642.
7. Orsakov, C. Glucagon-like peptide 1, a new hormone of the enteroinsular axis. Diabetologia 1992; 35: 701–711.
8. Zander, M.; Madsbad, S.; Madsen, J. L. and Holst, J. J. Effect of 6-week course of glucagon-like peptide 1 on glycaemic control, insulin sensitivity, and β-cell function in type 2 diabetes; a parallel-group study. Lancet 2002; 359: 824–830.
9. Knudsen, L. B. Glucagon-like peptide 1: A basis of a new class of treatment for type 2 diabetes. Journal of Medicinal Chemistry 2004; 47: 4128–4134.
10. Vahl, T. P. and D'Alessio, D. A. Gut peptides in the treatment of diabetes mellitus. Expert Opinion on Investigational Drugs 2004; 13: 177–188.
11. Meier, J. J and Nauck, M. A. Glucagon-like peptide 1 (GLP-1) in biology and pathology. Diabetes Metabolism Research and Reviews 2005; 21: 91–117.
12. Drucker, D. J. and Nauck, M. A. The inceretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase 4 inhibitors in type 2 diabetes. Lancet 2006; 368: 1696–1705.
13. Combettes, M. M. J. GLP-1 and type 2 diabetes: physiology and new clinical advances. Current Opinion in Pharmacology 2006; 6: 598–605.
14. Wettergren, A.; Schjoldager, B.; Martensen, P. E.; Myhre, J.; Christiansen, J. and Holst, J. J. Truncated GLP-1 (proglucagon 72–107 amide) inhibits gastric and pancreatic functions in man. Digestive Diseases and Sciences 1993; 38: 665–673.
15. Nauck, M. A,; Niedereichholz, U.; Ettler, R.; Holst, J. J.; Orskov, C.; Ritzel, R. and Schmigel, W. H. Glucagon-like peptide-1 inhibition of gastric-emptying outweighs its insulinotropic effects in healthy humans. American Journal of Physiology 1997; 273: E981–E988.
16. Flint, A.; Raben, A.; Ersboll, A. K.; Holst, J. J. and Astrup, A. The effect of physiological levels of glucagons-like peptide-1 on appetite, gastric emptying, energy and substrate metabolism in obesity. International Journal of Obesity 2001; 25: 781–792.
17. Ahren, B. E. Enhancement or prolongation of GLP-1 activity as a strategy for treatment of type 2 diabetes. Drug Discovery Today 2004; 1: 207–212.
18. Deacon, C. F. Therapeutic strategies based on glucagon-like peptide 1. Diabetes 2004; 53: 2181–2189.
19. Deacon, C. F. and Holst, J. J. Glucagon-like peptide 1 and inhibitors of dipeptidyl peptidase IV in the treatment of type 2 diabetes mellitus. Current Opinion in Pharmacolology 2004; 4: 589–596.
20. Mentlein, R. Therapeutic assessment of glucagon-like peptide-1 agonists compared with dipeptidyl peptidase IV inhibitors as potential antidiabetic drugs. Expert Opinion on Investigational Drugs 2005; 14: 57–64.
21. Nielsen, L. L. Incretin mimetics and DPP-4 Inhibitors for the treatment of type 2 diabetes. Drug Discovery Today 2005; 10: 703–710.
22. Gautier, J. F.; Fetita, S.; Sobngwi, E. and Salaun-Martin, C. Biological actions of theincretins GIP and GLP-1 and therapeutic perspectives in patients with type 2 diabetes. Diabetes and Metabolism 2005; 31: 233–242.
23. Kieffer, T. J.; McIntosh, C. H. S. and Pederson, T. A. Degradation of glucose-dependent insulinotropic polypeptide and truncated glucagon-like peptide 1 in vitro and in vivo by dipeptidyl peptidase IV. Endocrinology 1995; 136: 3585–3596.
24. Deacon, C. F.; Nauck, M. A.; Toft-Nielson, M.; Pridal, L.; Willms, B. and Holst, J. J. Both subcutaneously and intravenously administered glucagons-like peptide 1 are rapidly degraded from the NH2-terminus in type II diabetic patients and in healthy subjects. Diabetes 1995; 44: 1126–1131.
25. Kim, D.; Wang, L.; Beconi, M.; Eiermann, G. J.; Fisher, M. H.; He, H.; Hickey, G. J.; Kowalchick, J. E.; Leiting, B.; Lyons, K.; Marsilio, F.; McCann, M. E.; Patel, R. A.; Petrov, A.; Scapin, G.; Patel, S. B.; Sinha Roy, R.; Wu, J. K.; Wyvratt, M. J.; Zhang, B. B.; Zhu, L.; Thornberry, N. A. and Weber, A. E. (2R)-4-Oxo-4-[3-(trifluoromethyl)-5,6- dihydro[1,2,4]triazolo[4,3-a]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 2005; 48: 141–151.
26. Edmondson, S. E.; Fisher, M. H.; Kim, D.; Maccoss, M.; Parmee, E. R.; Weber, A. E. and Xu, J. U. S. Patent 6,699,871 B2, Mar. 2, 2004.
27. Villhauer, E. B.; Brinkman, J. A.; Naderi, G. B.; Burkey, B. F.; Dunning, B. E.; Prasad, K.; Mangold, B. L.; Russell, M. E. and Hughes, T. E. 1-[[(3-Hydroxy-1-adamantyl) amino]acetyl]-2-cyano-(S)-pyrrolidine: A potent selective, and orally bioavailable dipeptidyl peptidase IV inhibitor with antihyperglycemic properties. Journal of Medicinal Chemistry 2003; 46: 2774–2789.
28. Magnin, D. R.; Robl, J. A.; Sulsky, R. B.; Augeri, D. J.; Huang, Y.; Simpkins, L. M.; Taunk, P. C.; Betebenner, D. A.; Robertson, J. G.; Abboa-Offei, B. E.; Wang, A.; Cap,M.; Xin, L.; Tao, L.; Sitkoff, D. F.; Malley, M. F.; Gougoutas, J. Z.; Khanna, A.; Huang, Q.; Han, S.; Parker, R. A. and Hamann, L. G. Synthesis of novel potent dipeptidyl peptidase IV inhibitors with enhanced chemical stability: Interplay

between the N- terminal amino acid alkyl side chain and the cyclopropyl group of β-aminoacyl-l-cis-4,5- methanoprolinenitrile based inhibitors. Journal of Medicinal Chemistry 2004; 47: 2587-2598.
29. Kim, D.; Kowalchick, J. E.; Edmondson, S. D.; Mastracchio, A.; Xu, J. M.; Eiermann, G. J.; Leiting, B.; Wu, J. K.; Pryor, K.; Patel, R. A.; He, H.; Lyons, K. A.; Thornberry N. A. and Weber, A. E. Triazolopiperazine-amides as dipeptidyl peptidase IV inhibitors: Close analogues of JANUVIATM (sitagliptin phosphate). Bioorganic and Medicinal Chemistry Letters 2007; 17: 3373-3377.
30. Xu, J.; Ok, H. O.; Gonzalez, E. J.; Colwell, L. F. Jr.; Habulihaz, B.; He, H.; Leiting, B.; Lyona, K. A.; Marsilio, F.; Patel, R. A.; Wu, J. K.; Thornberry, N. A.; Weber, A. E. and Parmee, E. R. Discovery of potent and selective β-Homophenylalanine based dipeptidyl peptidase IV inhibitors. Bioorganic and Medicinal Chemistry Letters 2004; 14: 4759–4762.
31. Brockunier, L.; He, J.; Colwell, L. F. Jr.; Habulihaz, B.; He, H.; Leiting, B.; Lyons, K. A.; Marsilio, F.; Patel, R.; Teffera, Y.; Wu, J. K.; Thornberry, N. A.; Weber, A. E. and Parmee, E. R. Substituted piperazines as novel dipeptidyl peptidase IV inhibitors. Bioorganic and Medicinal Chemistry Letters 2004; 14: 4763–4766.
32. Kim, D.; Kowalchick, J. E.; Brockunier, L. L.; Parmee, E. R.; Eiermann, G. J.; Fisher, M. H.; He, H.; Leiting, B.; Lyons, K.; Scapin, G.; Patel, S. B.; Petrov, A.; Pryor, K. D.; Roy, R. S.; Wu, J. K.; Zhang, X.; Wyvratt, M. J.; Zhang, B. B.; Zhu, L.; Thornberry, N. and Weber, A. E. Discovery of potent and selective dipeptidyl peptidase IV inhibitors derived from β-Aminoamides bearing subsituted triazolopiperazines. Journal of Medicinal Chemistry 2008; 51: 589–602.
33. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA.
34. Dragon software (version 1.11-2001) by Todeschini R.; Consonni V. Milano, Italy.
35. Prabhakar, Y. S. 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.
36. Sharma, S.; Prabhakar, Y. S.; Singh, P. and Sharma, B. K. QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach. European Journal of Medicinal Chemistry 2008; 43: 2354-2360.
37. Sharma, S.; Sharma, B. K.; Sharma, S. K.; Singh, P. and Prabhakar, Y. S. 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.
38. Sharma, B. K.; Pilania, P.; Singh, P. and Prabhakar, Y. S. Combinatorial protocol in multiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. SAR QSAR in Environmental Research 2010; 21: 169-185.
39. Sharma, B. K.; Singh, P.; Sarbhai, K. and Prabhakar, Y. S. A quantitative structure- activity relationship study on serotonin 5-HT6 receptor ligands: Indolyl and piperidinyl sulphonamides, SAR QSAR in Environmental Research 2010; 21: 369-388.
40. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 1978; 20: 397–405.
41. Kettaneh, N.; Berglund, A. and Wold, S. PCA and PLS with very large data sets. Computational Statistics and Data Analysis 2005; 48: 69–85.
42. Stahle, L. and Wold, S. Multivariate data analysis and experimental design. In: Ellis, G. P. and West, W. B. (Eds.), Biomedical research. Progress in medicinal chemistry. Elsevier Science Publishers, BV, Amsterdam.1988; 25: 291–338.
43. Topliss, J. G. and Edwards, R. P. Chance factors in studies of quantitative structure– activity relationships. Journal of Medicinal Chemistry 1979; 22: 1238–1244.
44. Katritzky, A. R.; Dobchev, D. A.; Slavov, S. and Karelson M. Legitimate utilization of large descriptor pools for QSAR/QSPR models. Journal of Chemical Information and Modeling 2008; 48: 2207–2213.
45. So, S. – S. and Karplus, M. Three-dimensional quantitative structure–activity relationship from molecular similarity matrices and genetic neural networks. 1. Method and validation. Journal of Medicinal Chemistry 1997; 40: 4347–4359.
46. Prabhakar, Y. S.; Solomon, V. R.; Rawal, R. K.; Gupta, M. K. and Katti, S. B. CP-MLR/ PLS directed structure–activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR and Combinatorial Science 2004; 23: 234–244.
47. Akaike, H. Information theory and an extension of the minimum likelihood principle. In: Petrov, B. N. and Csaki, F. (Eds.). Second international symposium on information theory. Budapest: Akademiai Kiado; 1973. p 267–281.
48. Akaike, H. A new look at the statistical identification model. IEEE Transactions on Automatic Control 1974; AC-19: 716–723.
49. Kubinyi, H. Variable selection in QSAR studies. I. An evolutionary algorithm. Quantitative Structure-Activity Relationship 1994; 13: 285–294.
50. 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.
51. Friedman J. In Technical Report No. 102. Laboratory for Computational Statistics, Stanford University: Stanford; 1990.
52. Gramatica, P. Principles of QSAR models validation: internal and external. QSAR and Combinatorial Science 2007; 26: 694-701.
53. Hemmer, M. C.; Steinhauer, V. and Gasteiger, J. Deriving the 3D structure of organic molecules from their infrared spectra. Vibrational Spectroscopy 1999; 19: 151-164.
54. Silverman, B. D. Three-dimensional moments of molecular property fields. The Journal for Chemical Information and Computer Scientists 2000; 40: 1470-1476.
55. Schuur, J. H.; Selzer, P. and Gasteiger, J. The Coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. Journal of the American Chemical Society 1996; 36: 334-344.
56. (a) Consonni, V.; Todeschini, R. and Pavan, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. The Journal for Chemical Information and Computer Scientists 2002; 42: 682-692. (b) Consonni, V.; Todeschini, R.; Pavan, M. and Gramatica, P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. The Journal for Chemical Information and Computer Scientists 2002; 42: 693- 705.
57. (a) Todeschini, R.; Lasagni, M. and Marengo, E. New molecular descriptors for 2D and 3D structures. Theory. Journal of Chemometrics 1994; 8: 263-273. (b) Todeschini, R. and Gramatica, P. 3D QSAR in drug design - vol. 2, Kubinyi, H.; Folkers, G. and Martin, Y. C. (Eds.), Kluwer/ESCOM, Dordrecht (The Netherlands), 1998; 355-380.