BJMHR
British Journal of Medical and Health Research
Journal home page: www.bjmhr.com

# QSAR study on the Dipeptide-Derived Inhibitor of Dipeptidyl Peptidase IV: The (2S)-Cyanopyrrolidine analogues

**Raghuraj Parihar[1], Brij Kishore Sharma[1]***
*Department of Chemistry, Government College, Bundi-323 001, India*

## ABSTRACT

The inhibition activity of (2*S*)-cyanopyrrolidine analogues for dipeptidyl peptidase IV has been quantitatively analyzed in terms of topological 0D-, 1D- and 2D-descriptors based on molecular graph theory. Statistically sound models have been obtained between the activity and various DRAGON descriptors through combinatorial protocol-multiple linear regression (CP-MLR) computational procedure. Amongst the large number of such derived models, the most significant ones have only been discussed to draw meaningful conclusions. Additionally the inhibition activity for DPP8 enzyme, reported for a limited number of such congeners, has also been correlated with such descriptors. From the final statistically significant models, it appeared that the mode of actions of titled compounds were different for DPP IV and DPP8 enzyme systems. Applicability domain analysis carried out for DPP IV inhibitors revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

**Keywords:** QSAR, DPP IV inhibitors, Combinatorial protocol in multiple linear regression (CP-MLR) analysis, Dragon descriptors, (2*S*)-cyanopyrrolidine analogues.

*Corresponding Author Email: bksharma_sikar@rediffmail.com
Received 05 June 2020, Accepted 19 June 2020

Please cite this article as: Sharma BK *et al*., QSAR study on the Dipeptide-Derived Inhibitor of Dipeptidyl Peptidase IV: The (2S)-Cyanopyrrolidine analogues . British Journal of Medical and Health Research 2020.

## INTRODUCTION

The most potent known insulinotropic hormone is glucagon-like peptide-1 (GLP-1) [1-3]. This hormone, containing 30 amino acids, is produced by L-cells of the intestinal mucosa that results from tissue specific processing of the proglucagon gene [4,5]. The stimulation of insulin secretion, inhibition of glucagons release [6-9] and slow down of gastric emptying [10-13] is due to the active form of GLP-1. These beneficial effects are helpful in controlling the glucose homeostasis in patients with type 2 diabetes [14-16]. However, GLP-1 is rapidly degraded by plasma DPP-IV and is lacking of oral activity; its use as a therapeutic agent is, therefore, restricted. In view of this a small molecule, as the inhibitor of DPP-IV, may extend the duration of action of GLP-I and result in the beneficial effects of this hormone for a long period of time. Through human clinical trials, it was shown that inhibition of DPP-IV may improve glucose tolerance in diabetic patients and healthy volunteers and leads to a new strategy for the treatment of type 2 diabetes [17-21]. DPP-IV is a serine protease, able to cleave the N-terminal dipeptide having preference for L-proline or L-alanine at the penultimate position [22-25]. A large number of DPP-IV inhibitors resemble the P2-P1 dipeptidyl substrate cleavage product. The simplest inhibitors are the compounds which are not having a carbonyl functionality of the proline residue, e.g., aminoacyl pyrrolidines and thiazolidines, possessing moderate inhibition activity for DPP-IV. Replacement of hydrogen with an electrophilic nitrile group at the 2-position of the pyrrolidine, in some compounds, elicited a 1000-fold increase in potency compared to the unsubstituted pyrrolidines [26].

One of the potent and stable representatives of the nitrile class is cyclohexylglycine-(2*S*)-cyanopyrrolidine, having a $K_i$ value of 1.4 nM and an excellent chemical stability $t_{1/2}$ ~ 48 h at pH 7.4[27]. Another class, similar to proline inhibitors, was synthesized with diverse N-substituted glycines in the P2 site [17]. In this class, the side chain was moved from the α-carbon to the terminal nitrogen, led to two potent derivatives which have showed greater efficacy in clinical trial [28]. From this study, it was concluded that (2*S*)-cyanopyrrolidine derivatives with *N*-substituted glycine in the P2 site are more selective for DPP-IV than α-carbon-substituted glycine. An interesting study has recently been reported to develop a new pharmacophore in the P2 site with N-substituted glycine [29]. Initially, the P2 site amine extension was designed using β-alanine as building block and it was coupled the C-terminal with various substituted amines to generate a novel pharmacophore in the P2 site. Then, the N-terminal of the β-alanine derivative was combined with the P1 site α-bromoacetyl (2*S*)-cyanopyrrolidide to design 2-[3-[[2-[(2*S*)-2-cyano-1-pyrrolidinyl]-2-oxoethyl]amino]-1-oxopropyl]-based DPP-IV inhibitors. The structure-activity relationships of several series (I–III) of these DPP-IV inhibitors were explored to discover the potent and selective DPP-IV inhibitors. Series I, II, and III, being N-

substituted glycine derivatives include, respectively, the bicyclic ring system, monocyclic piperazine ring, and phenylalkyl groups. These compounds were tested for inhibition of DPP-IV, DPP8, and DPP-II. The activity was evaluated in terms of the concentration of a compound required to bring out 50% inhibition of the enzyme concerned. 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.

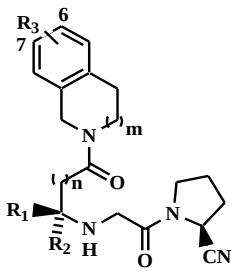
## MATERIALS AND METHOD

***Data set***

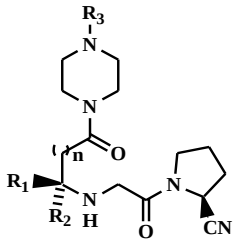
The reported twenty five (2*S*)-cyanopyrrolidine analogues, belonging to series I, II, and III are considered to formulate the data set for present study [29]. The structural variations and the activity values, expressed as $IC_{50}$(nM), of the reported analogues are given in Table 1. Since the activity variation for DPP-II is very small, therefore, inhibition profiles for DPP-IV and DPP8 have only been considered for quantitative analysis.

For the purpose of modeling study, the data set has been divided into training and test sets. One fifth of the compounds, from this data set, have been included in the test set for the validation of derived models while remaining compounds were used to derive the model correlating biological activity with descriptors unfolding molecular structures. The test-set, containing 5 compounds out of the 25 active ones, was generated in the SYSTAT [30] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the activity. The selection of test set from the cluster tree was done in such a way to keep the test compounds at a maximum possible distance from each other. In this way, the identified test set will further ensure the statistical significance and reasonable predictability of derived models. As the leave-one-out (LOO) procedure has been applied to each model, therefore, corresponding to test set the derived model would be validated both internally and externally. The training and test set compounds are listed in Table 1.

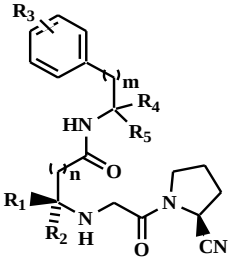
**Table 1**: **General structures and structural variations of (2*S*)-Cyanopyrrolidine analogues.**



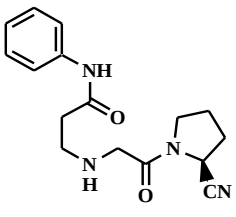

**Series I (1-12)**




**Series II (13-16)**




**Series III (17-24)**




**25**

| Cpd. | n | m | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $IC_{50}$(nM)[a] | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | DPP-IV | DPP8 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1. | 0 | 1 | H | H | H | | | 3236 | 4169 |
| 2. | 1 | 1 | H | H | H | | | 116 | 3583 |
| 3. | 1 | 1 | H | H | 6,7-diOMe | | | 651 | 3340 |
| 4.[b] | 1 | 1 | H | H | 6-F | | | 83 | 1700 |
| 5. | 1 | 0 | H | H | H | | | 132 | 2121 |
| 6. | 2 | 1 | H | H | H | | | 428 | 1407 |
| 7. | 1 | 1 | $CH_3$ | H | H | | | 54 | 5346 |
| 8.[b] | 1 | 1 | *i*-Pr | H | H | | | 811 | 41859 |
| 9. | 1 | 1 | $CH_3$ | $CH_3$ | H | | | 49 | >100000 |
| 10. | 1 | 1 | $CH_3$ | $CH_3$ | 6-F | | | 30 | >100000 |
| 11. | 1 | 1 | $CH_3$ | $CH_3$ | 6,8-$F_2$ | | | 22 | >100000 |
| 12. | 1 | 0 | $CH_3$ | $CH_3$ | H | | | 15 | >100000 |
| 13. | 1 | | H | H | CO(3,5-$F_2C_6H_3$) | | | 676 | 202 |
| 14. | 1 | | H | H | $SO_2C_6H_4$-4-$NHCOCH_3$ | | | 418 | 3416 |
| 15. | 1 | | H | H | nicotinonitrile | | | 629 | 2000 |
| 16. | 1 | | H | H | benzothiazole | | | 527 | 2117 |
| 17. | 1 | 0 | H | H | H | H | H | 452 | 10744 |
| 18. | 1 | 0 | H | H | 4-$NO_2$ | H | H | 317 | 2387 |
| 19.[b] | 1 | 0 | H | H | H | H | ethyl[c] | 447 | 21961 |
| 20. | 1 | 0 | H | H | 3,5-$F_2$ | H | H | 369 | 5532 |
| 21. | 1 | 0 | H | H | H | H | *i*-Pr[c] | 784 | 12847 |
| 22.[b] | 1 | 0 | H | H | H | $CH_3$ | $CH_3$ | 119 | 8338 |
| 23. | 1 | 0 | $CH_3$ | $CH_3$ | H | $CH_3$ | $CH_3$ | 1108 | >100000 |
| 24. | 1 | 1 | H | H | H | H | H | 564 | 2592 |
| 25.[b] | | | | | | | | 298 | 855 |

[a]$IC_{50}$ represents the concentration of a compound required to bring out 50% inhibition of DPP-IV and DPP8, taken from ref [29]; [b]compound of test set; [c]The stereochemistry at the benzylic carbon is not defined (mixture of diastereomers).

***Theoretical molecular descriptors***

The structures of the compounds were drawn in 2D ChemDraw [31], converted into 3D modules and subjected to energy minimization in the MOPAC using the AM1 procedure for closed shell system to ensure a well defined conformer relationship among the compounds under investigation. DRAGON software [32] was then used to compute the molecular descriptors of titled compounds. This software offers several hundreds of descriptors corresponding to 0D-, 1D-, and 2D-descriptor modules. The modules include ten different classes, namely, the constitutional (CONST), the topological (TOPO), the molecular walk counts (MWC), the BCUT descriptors (BCUT), the Galvez topological charge indices (GALVEZ), the 2D-autocorrelations (2D-AUTO), the functional groups (FUNC), the atom-centered fragments (ACF), the empirical descriptors (EMP), and the properties describing descriptors (PROP). For each of these classes the DRAGON software computes a large number of descriptors which are characteristic to the molecules under multi-descriptor environment. A total number of 484 descriptors, belonging to above classes, have been computed to generate most appropriate models describing the biological activity. The combinatorial protocol in multiple linear regression (CP-MLR) [33] method, discussed below, has been used further for developing

statistical significant models divulging quantitative structure-activity relationship (QSAR). Before the application of CP-MLR procedure, all those descriptors which are intercorrelated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor vs. activity, $r < 0.1$) were excluded. This has reduced the total dataset of the compounds from 484 to 90 and 84 descriptors as relevant ones for the DPP-IV and DPP8 inhibitory activity, respectively. The descriptors, in all above models, have been scaled between the intervals 0 to 1 [34] 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.

***Model development***

The CP-MLR software is developed for the selection of most appropriate descriptors from a pool, which are subsequently used to develop statistical significant models in a systematic manner. Its procedural aspects and implementation are discussed in some of our publications [35-40]. The thrust of this software is implicated mainly on its embedded "filters" which have been interfaced with multiple linear regression (MLR) to extract diverse structure-activity models, each having unique combination of descriptors from the dataset under investigation. In this procedure, the contents and number of variables to be evaluated are mixed according to the predefined confines and the 'filters' are significance evaluators of the variables in regression at different stages of model development. Of these, filter-1 is set in terms of inter-parameter correlation cutoff criteria for variables to stay as a subset (filter-1, default value 0.3 and upper limit $\leq 0.79$). In this, if two variables are correlated higher than a predefined cutoff value the respective variable combination is forbidden and will be rejected. The second filter is in terms of t-values of regression coefficients of variables associated with a subset (filter-2, default value 2.0). Here, if the ratio of regression coefficient and associated standard error of any variable is less than a predefined cutoff value then the variable combination will be rejected. Since successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient (r) values, square-root of adjusted multiple correlation coefficient of regression equation, r-bar, has been used to compare the internal explanatory power of models with different number of variables. Accordingly, a filter has been set in terms of predefined threshold level of r-bar (filter-3, default value 0.79) to decide the variables' 'merit' in the model formation. Finally, to exclude false or artificial correlations, the external consistency of the variables of the model have been addressed in terms of cross-validated $Q^2$ criteria from the leave-one-out (LOO) cross-validation procedure as default option (filter-4, default threshold value $0.3 \leq Q^2 \leq 1.0$). All these filters make the variable selection process efficient and lead to unique solution. In order to collect the descriptors with higher information content and explanatory power, the threshold of filter-3 may be incremented successively with

increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation.

***Model validation***

The subdivision of data set into training set and test set have been used, respectively, for model development and external prediction. Goodness of fit of the models was assessed by examining the multiple correlation coefficient (r), the standard deviation (s) and the F-ratio between the variances of calculated and observed activities (F). The internal validation of derived model was ascertained through the cross-validated index, $Q^2$, from leave-one-out ($Q^2_{LOO}$) and leave-five-out ($Q^2_{L5O}$) procedures. The LOO method creates a number of modified data sets by taking away one compound from the parent data set in such a way that each observation has been removed once only. Then one model is developed for each reduced data set, and the response values of the deleted observations are predicted from these models.

The external validation or predictive power of derived model is based on test set compounds. The index $r^2_{Test}$, representing the squared correlation coefficient between the observed and predicted data of the test-set, has been used to infer the same. A value greater than 0.5 of $r^2_{Test}$ suggests that the model obtained from training set has a reliable predictive power.

***Y-randomization***

Chance correlations, if any, associated with the CP-MLR models were explored through randomization test [41,42] by repeated scrambling of the biological response. The data sets with scrambled response vector have been reassessed by multiple regression analysis (MRA). The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to the unscrambled response data were counted. Every model has been subjected to 100 such simulation runs. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

***Applicability domain***

The utility of a QSAR model is based on its accurate predictive ability for new congeners. A model is valid only within its training domain, and new compounds must be assessed as belonging to this domain before the model is applied. The applicability domain is assessed by the leverage values for each compound [43,44]. A 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 ± β.(standard deviations) and a leverage threshold h*. The threshold h* is generally fixed at 3(k+1)/n (n is the number of compounds in the training-set and k is the number of independent descriptors of the model) whereas β = 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 agreement between predicted and observed values is as high as that for the training set compounds.

## RESULTS AND DISCUSSION

### Qsar results

Initially, the DPP-IV inhibition activity of titled compounds was investigated with a variety of 0D-, 1D- and 2D-descriptors obtained from DRAGON software. For the development of QSAR models, 20 compounds were considered in training set while 05 (one-fifth of the total) compounds were included in test set for external validation of derived significant models. Though each individual descriptor class is enriched with information corresponding to the activity, different descriptors classes together have led to the models with optimum explained variance. These models 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 models, in three parameters of the descriptor pool of 90 descriptors, emerged in CP-MLR for the DPP-IV inhibition actions are given in Table 2 as Equations (i) to (vi). 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. The maximum number of descriptors, participated in these models ATS8p, GATS8p, GATS7e, MATS3e and MATS8e, belong to 2D-autocorrelations (2D-AUTO) class. The 2D-AUTO descriptors, ATSke, GATSke and MATSke have their origin in autocorrelation of topological structure of Broto-Moreau, of Moran and of Geary, respectively. The computation of these descriptors involves the summation of different autocorrelation functions corresponding to the different fragment lengths and lead to different autocorrelation vectors corresponding to the lengths of the structural fragments. Also a weighting component in terms of a physicochemical property has been embedded in these descriptors. As a result, these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. In these descriptors' nomenclature, the penultimate character, a number, indicates the number of consecutively connected edges considered in its computation and is called as the autocorrelation vector of lag k (corresponding to the number of edges in the unit fragment). The very last character of the descriptor's nomenclature indicates the physicochemical property considered in the weighting component – m for atomic mass, e for atomic Sanderson electronegativity and p for atomic polarizability - for its computation.

**Table 2**: **CP-MLR models[a] derived in three parameters for the DPP-IV inhibition activity.**

| Model | r | s | F | $q^2_{LOO}$ | $r^2_{Test}$ | Eq. |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| $pIC_{50}$ = 5.090 + 1.709(0.280)JGI4 – 1.150(0.331)ATS8p + 2.074(0.297)GATS8p | 0.934 | 0.242 | 36.897 | 0.806 | 0.279 | (i) |
| $pIC_{50}$ = 5.538 + 0.599(0.248)BELm1 – 0.902(0.364)GATS7e + 1.784(0.358)GATS8p | 0.854 | 0.353 | 14.465 | 0.615 | 0.331 | (ii) |
| $pIC_{50}$ = 7.174 – 1.493(0.444)Uindex + 1.200(0.347)JGI4 – 1.492(0.442)MATS3e | 0.850 | 0.358 | 13.927 | 0.533 | 0.250 | (iii) |
| $pIC_{50}$ = 5.939 – 0.976(0.400)GATS7e + 1.442(0.410)GATS8p + 0.668(0.325)C-002 | 0.842 | 0.367 | 13.021 | 0.562 | 0.331 | (iv) |
| $pIC_{50}$ = 6.094 – 0.663(0.327)RBN – 0.725(0.361)GATS7e + 1.848(0.374)GATS8p | 0.841 | 0.368 | 12.913 | 0.511 | 0.191 | (v) |
| $pIC_{50}$ = 5.651 + 1.109(0.418)JGI4 + 1.124(0.524)MATS8e – 0.806(0.400)GATS7e | 0.804 | 0.404 | 9.812 | 0.521 | 0.233 | (vi) |

[a]The models, in three parameters, emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.5 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 20 compounds.

Desciptors, GATS8p and MATS8e both added positively to the inhibitory activity whereas ATS8p, GATS7e and MATS3e contributed negatively to the activity advocating that higher values of descriptors GATS8p and MATS8e and lower values of descriptors ATS8p, GATS7e and MATS3e would be beneficiary to the activity. Constitutional class descriptors are dimensionless or 0D descriptors and are independent from molecular connectivity and conformations. The appeared constitutional class descriptor RBN (number of rotatable bonds) favors the least preference of rotatable bonds.

Descriptor Uindex, corresponds to Balaban U-index, is a topological class descriptor. Topological class descriptors are based on a graph representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. They can be sensitive to one or more structural features of the molecules such as size, shape, symmetry, branching and cyclicity and can also encode chemical information concering atom type and bond multiplicity. The negative contribution of descriptor Uindex suggested that a lower value of it would be supportive to the activity. The other participated descriptors are JGI4 (from the Galvez topological charge indices), BELm1 (from the modified Burden eigenvalues class, BCUT descriptors) and C-002 (from the atom-centered fragments). GALVEZ class descriptors are 2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix. BCUT descriptors are also 2D-descriptors representing positive and negative eigenvalues of the adjacency matrix weighting the diagonal elements and atoms. Atom centered fragments (ACF) are molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen. The $4^{th}$ order mean Galvez topological charge index (JGI4), the lowest eigenvalue n.1 of Burden matrix/ weighted by atomic masses (BELm1) and CH2R2 type atom centered fragment (C-002) correlated positively to the activity suggested that a higher value of these will augment the activity. However, for all the models mentioned in Table 2 the $r^2_{Test}$ values (<0.5) are inferior to a specified value.

Considering the number of observation in the data set, models with up to four descriptors were explored. A total number of seven such models, sharing 18 descriptors among them, have been obtained through CP-MLR. The shared 18 descriptors along with their brief description, average regression coefficients and total incidences are given in Table 3. Following are the selected four-descriptor models, obtained from CP-MLR, for the DPP-IV inhibitory activity.

**Table 3**. **Identified descriptors[a] along with their physical meaning, average regression coefficient and incidence[b], in modeling the DPP-IV inhibitory activity.**

| S. No. | Descriptor | Descriptor class | Physical meaning | Average regression coefficient (incidence) |
|---|---|---|---|---|
| 1 | RBN | Constitutional | Number of rotatable bonds | -1.302(1) |
| 2 | PJI2 | Topological | 2D Petitijean shape index | 0.360(1) |
| 3 | BIC3 | Topological | Bond information content (neighborhood symmetry of 3 order) | 3.300(2) |
| 4 | BIC5 | Topological | Bond information content (neighborhood symmetry of 5 order) | -2.310(1) |
| 5 | SRW09 | Molecular walk counts | Self- returning walk count of order 09 | 1.341(1) |
| 6 | BELm1 | BCUT | Lowest eigenvalue n.1 of Burden matrix/ weighted by atomic masses | 0.713(1) |
| 7 | BEHv1 | BCUT | Highest eigenvalue n.1 of Burden matrix/ weighted by atomic van der Waals volumes | 0.821(1) |
| 8 | BELe1 | BCUT | lowest eigenvalue n.1 of Burden matrix/ weighted by atomic Sanderson electro negativities | 0.655(1) |
| 9 | BELp2 | BCUT | lowest eigenvalue n.2 of Burden matrix/ weighted by atomic polarizabilities | -0.675(1) |
| 10 | JGI4 | Galvez topological charge indices | Mean topological charge index of order 4 | 1.993(1) |
| 11 | ATS8p | 2D autocorrelations | Broto-Moreau autocorrelation of a topological structure - lag8/ weighted by atomic polarizabilities | -0.992(2) |
| 12 | GATS7e | 2D autocorrelations | Geary autocorrelation of lag-7/ weighted by atomic Sanderson electro negativities | -0.901(4) |
| 13 | GATS8p | 2D autocorrelations | Moran autocorrelation of lag-8/ weighted by atomic polarizabilities | 2.020(5) |
| 14 | C-024 | Atom-centred fragments | R--CH--R | 0.405(1) |
| 15 | C-040 | Atom-centred fragments | R-C(=X)-X/ R-C#X/X-=C=X | 0.853(1) |

| | | | | |
|---|---|---|---|---|
| 16 | H-052 | Atom-centred fragments | H attached to C0(sp3) with 1X attached to next C | 2.214(2) |
| 17 | MR | Properties | Ghose-Crippen molecular refractivity | -1.058(1) |
| 18 | MLOGP | Properties | Moriguchi octanol-water partition coefficient (logP) | 1.331(1) |

[a]The descriptors are identified from the four parameter models, emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.809 (r-bar of the three parameter model having the highest $r^2_{Test}$ ), and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 20 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.

$pIC_{50}$ =4.722+1.993(0.276) JGI4–1.295(0.300) ATS8p+2.163(0.265) GATS8p+ 0.405(0.173)C-024

n = 20, r = 0.952, s = 0.214, F = 36.718, $q^2_{LOO}$ = 0.832, $q^2_{L5O}$ = 0.805, $r^2_{Test}$= 0.576 (1)

$pIC_{50}$=5.165–0.688(0.336) ATS8p–0.690(0.267) GATS7e+2.112(0.344) GATS8p+ 1.331(0.288) MLOGP

n = 20, r = 0.921, s = 0.273, F = 21.086, $q^2_{LOO}$ = 0.715, $q^2_{L5O}$ = 0.704, $r^2_{Test}$= 0.730 (2)

$pIC_{50}$ = 5.910 – 1.302(0.282) RBN + 3.530(0.486) BIC3 – 2.310(0.364) BIC5 + 1.897(0.247) H-052

n = 20, r = 0.920, s = 0.274, F = 20.846, $q^2_{LOO}$ = 0.561, $q^2_{L5O}$ = 0.713, $r^2_{Test}$= 0.801 (3)

$pIC_{50}$ = 3.727 +3.070(0.467) BIC3+1.341(0.226) SRW09+0.853(0.243) C-040+ 2.530(0.335) H-052

n = 20, r = 0.907, s = 0.295, F = 17.516, $q^2_{LOO}$ = 0.558, $q^2_{L5O}$ = 0.582, $r^2_{Test}$= 0.543 (4)

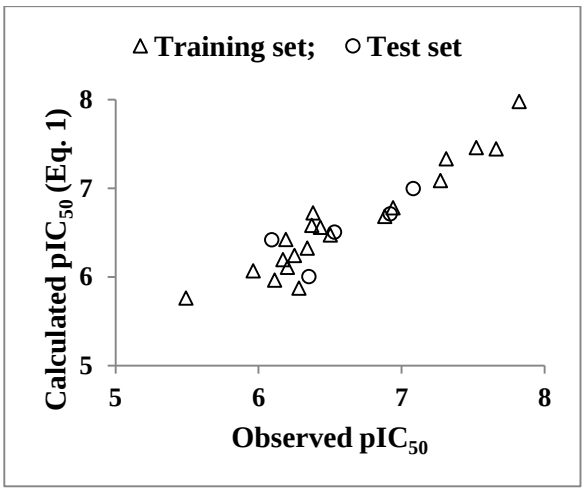
Where n and F represent respectively the number of data points and the F-ratio between the variances of calculated and observed activities. The data within the parentheses are the standard errors associated with regression coefficients. In all above equations, the F-values remained significant at 99% level. The indices $q^2_{LOO}$ and $q^2_{L5O}$ (> 0.5) have accounted for their internal robustness. For all above models the $r^2_{Test}$ values, obtained greater than 0.5, specified that the selected test-set is fully accountable for their external validation.

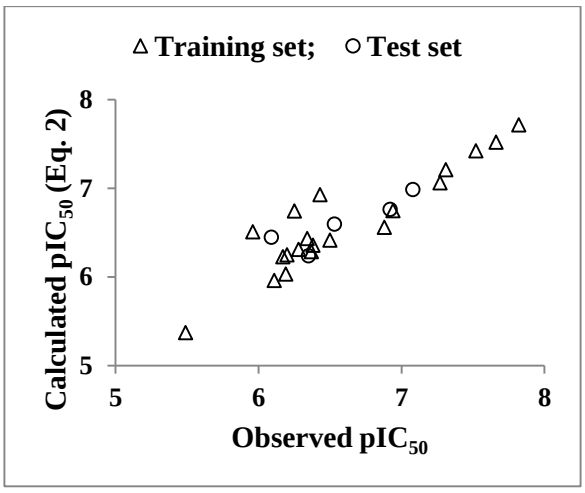
These models are able to estimate up to 90.73 percent of variance in observed activity of the compounds. The derived statistical parameters of these four models have shown that these models are significant. These models were, therefore, used to calculate the activity profiles of all the compounds and are included in Table 4 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 1 to ensure the goodness of fit for each of these four models.

**Table 4: Observed, calculated and predicted DPP-IV inhibition activities of (2*S*)-Cyanopyrrolidine analogues.**

| Cpd. | $pIC_{50}$[a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Obsd[b]. | Eq. (1) Calc. | Pred[c]. | Eq. (2) Calc. 2 | Pred[c]. | Eq. (3) Calc.3 | Pred[c]. | Eq. (4) Calc. 4 | Pred[c]. |
| 1 | 5.49 | 5.76 | 5.90 | 5.37 | 5.31 | 5.75 | 5.98 | 5.73 | 5.79 |
| 2 | 6.94 | 6.78 | 6.76 | 6.75 | 6.72 | 6.59 | 6.52 | 6.64 | 6.61 |
| 3 | 6.19 | 6.42 | 6.53 | 6.04 | 6.00 | 5.84 | 5.76 | 5.86 | 5.80 |
| 4[d] | 7.08 | 7.00 | -[d] | 6.99 | -[d] | 7.26 | -[d] | 7.22 | -[d] |
| 5 | 6.88 | 6.68 | 6.65 | 6.56 | 6.50 | 7.03 | 7.11 | 6.84 | 6.82 |
| 6 | 6.37 | 6.58 | 6.60 | 6.29 | 6.18 | 6.40 | 6.41 | 6.78 | 6.82 |
| 7 | 7.27 | 7.09 | 7.07 | 7.06 | 7.04 | 6.93 | 6.85 | 7.27 | 7.27 |
| 8[d] | 6.09 | 6.42 | -[d] | 6.45 | -[d] | 6.26 | -[d] | 6.32 | -[d] |
| 9 | 7.31 | 7.33 | 7.34 | 7.21 | 7.19 | 7.13 | 7.09 | 7.10 | 7.06 |
| 10 | 7.52 | 7.46 | 7.44 | 7.42 | 7.40 | 7.70 | 7.77 | 7.60 | 7.63 |
| 11 | 7.66 | 7.44 | 7.34 | 7.52 | 7.47 | 7.70 | 7.72 | 7.60 | 7.58 |
| 12 | 7.82 | 7.98 | 8.07 | 7.72 | 7.68 | 7.45 | 7.18 | 7.33 | 6.99 |
| 13 | 6.17 | 6.20 | 6.20 | 6.23 | 6.24 | 6.15 | 6.14 | 5.88 | 5.71 |
| 14 | 6.38 | 6.72 | 6.80 | 6.36 | 6.34 | 6.15 | 6.06 | 6.57 | 6.68 |
| 15 | 6.20 | 6.11 | 5.99 | 6.25 | 6.28 | 6.21 | 6.21 | 6.30 | 6.34 |
| 16 | 6.28 | 5.88 | 5.78 | 6.31 | 6.32 | 6.27 | 6.27 | 6.81 | 7.10 |
| 17 | 6.34 | 6.33 | 6.32 | 6.44 | 6.45 | 6.56 | 6.59 | 6.27 | 6.26 |
| 18 | 6.50 | 6.47 | 6.47 | 6.42 | 6.41 | 6.67 | 6.71 | 6.44 | 6.43 |
| 19[d] | 6.35 | 6.00 | -[d] | 6.24 | -[d] | 6.14 | -[d] | 6.42 | -[d] |
| 20 | 6.43 | 6.56 | 6.59 | 6.93 | 7.16 | 6.88 | 6.99 | 6.55 | 6.57 |
| 21 | 6.11 | 5.97 | 5.88 | 5.96 | 5.88 | 6.06 | 6.05 | 5.87 | 5.83 |
| 22[d] | 6.92 | 6.71 | -[d] | 6.76 | -[d] | 6.97 | -[d] | 6.62 | -[d] |
| 23 | 5.96 | 6.07 | 6.20 | 6.51 | 6.80 | 6.32 | 7.14 | 6.26 | 6.87 |
| 24 | 6.25 | 6.24 | 6.24 | 6.75 | 6.84 | 6.24 | 6.23 | 6.36 | 6.37 |
| 25[d] | 6.53 | 6.51 | -[d] | 6.60 | -[d] | 6.68 | -[d] | 6.16 | -[d] |

[a]$IC_{50}$ represents the concentration of a compound required to bring out 50% inhibition of DPP-IV and the same is expressed as $pIC_{50}$ on molar basis; [b]Taken from ref. [29]; [c]Leave-one-out (LOO) procedure; [d]Compound included in test set.



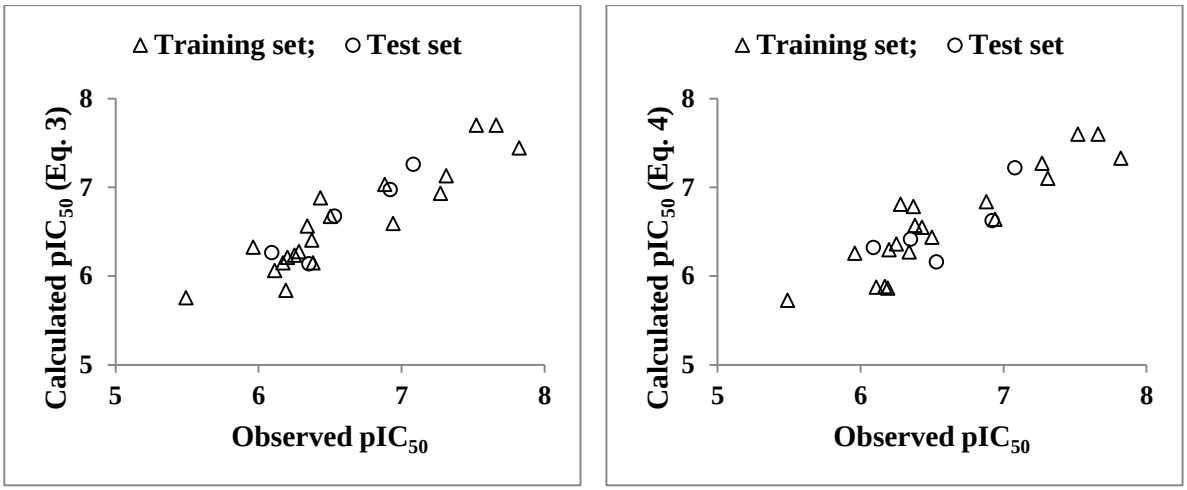



**Figure 1**: **Plots showing variations of observed versus calculated $pIC_{50}$ values of training and test set compounds.**

The newly emerged descriptors C-024, H-052 and C-040 in these models are atom centered fragments and shown positive correlation to the activity. Therefore, presence of R--CH--R (descriptor C-024), H attached to C0(sp3) with 1X attached to next C (descriptor H-052) and R-C(=X)-X/R-C#X/X-=C=X (descriptor C-040) type atom centered fragments in a molecular structure would enhance the activity. Topological class descriptor BIC3 (bond information content of 3rd order neighborhood symmetry) contributed positively whereas BIC5 (bond information content of 5th order neighborhood symmetry) contributed negatively to the activity revealed that a higher value of descriptor BIC3 and a lower value of descriptor BIC5 would be beneficiary to the activity. Descriptor SRW09 represents self- returning walk count of order 09 is from molecular walk counts class. Molecular walk counts are 2D-descriptors representing self-returning walk counts of different lengths. The descriptor MLOGP is from properties class representing Moriguchi octanol-water partition coefficient (logP). It is evinced from the models that higher values of both of these descriptors (SRW09 and MLOGP) would augment the activity.

CP-MLR analysis has also been performed for the DPP-8 inhibitory activity with the descriptor pool of 84 descriptors with the same test which was used for the DPP-IV inhibitory activity. All the emerged four models in three descriptors are given below through Equations (5) to (8).

$pIC_{50} = 5.283 + 0.863(0.234)MW – 0.745(0.207)X1Av + 0.544(0.123)PJI2$

$n = 15, r = 0.885, s = 0.218, F = 13.264, q^2_{LOO} = 0.552, q^2_{L5O} = 0.578, r^2_{Test} = 0.683$ (5)

$pIC_{50} = 5.419 – 0.536(0.210)X1Av + 0.533(0.129)PJI2 + 0.497(0.148)C\text{-}040$

$n = 15, r = 0.871, s = 0.229, F = 11.623, q^2_{LOO} = 0.507, q^2_{L5O} = 0.612, r^2_{Test} = 0.761$ (6)

$pIC_{50} = 5.300 – 0.778(0.223)X1Av + 0.575(0.130)PJI2 + 0.771(0.231)BEHm8$

$n = 15, r = 0.871, s = 0.230, F = 11.525, q^2_{LOO} = 0.506, q^2_{L5O} = 0.726, r^2_{Test} = 0.646$ (7)

$pIC_{50} = 6.037 + 0.609(0.135)PJI2 – 0.601(0.254)GATS1e – 0.691(0.205)C\text{-}024$

$n = 15, r = 0.860, s = 0.238, F = 10.493, q^2_{LOO} = 0.533, q^2_{L5O} = 0.528, r^2_{Test} = 0.511$ (8)

The derived statistical parameters of these four models have shown that these models are significant and are able to explain up to 78.34 percent of variance in observed DPP8 activity of the compounds. A close agreement between observed and calculated activity has been observed and is given in Table 5 for the sake of comparison. The participated descriptors in above models suggested that higher values of molecular weight (MW, constitutional class descriptor), 2D- Petitijean shape index (PJI2, topological class), highest eigenvalue n.8 of Burden matrix weighted by atomic masses (BEHm8, BCUT descriptor) and presence of R-C(=X)-X/R-C#X/X-=C=X type atom centered fragments (descriptor C-040, atom centered fragment descriptor) would be beneficiary to DPP8 inhibitory activity. Another emerged topological class descriptor X1Av (average valence connectivity index, chi-1), 2D-AUTO class descriptor GATS1e (Geary autocorrelation of lag-1/weighted by atomic Sanderson electro negativities) advocated that a lower value of these descriptors and absence of R--CH--R type fragment (descriptor C-024) would augment the activity.

**Table 5**: **Observed, calculated and predicted DPP-8 inhibition activities of (2*S*)-Cyanopyrrolidine analogues.**

| Cpd. | $pIC_{50}$[a] Obsd[b]. | Eq. (5) | | Eq. (6) | | Eq. (7) | | Eq. (8) | |
|---|---|---|---|---|---|---|---|---|---|
| | | Calc. | Pred[c]. | Calc. 2 | Pred[c]. | Calc. 3 | Pred[c]. | Calc. 4 | Pred[c]. |
| 1 | 5.38 | 5.72 | 5.82 | 5.79 | 5.89 | 5.64 | 5.74 | 5.81 | 5.89 |
| 2 | 5.45 | 5.11 | 5.04 | 5.16 | 5.11 | 5.20 | 5.16 | 5.20 | 5.16 |
| 3 | 5.48 | 5.60 | 5.63 | 5.34 | 5.32 | 5.69 | 5.75 | 5.50 | 5.53 |
| 4[d] | 5.77 | 5.88 | -[d] | 5.80 | -[d] | 6.01 | -[d] | 6.20 | -[d] |
| 5 | 5.67 | 5.72 | 5.73 | 5.79 | 5.82 | 5.90 | 5.94 | 5.81 | 5.84 |
| 6 | 5.85 | 5.61 | 5.55 | 5.62 | 5.56 | 5.79 | 5.77 | 5.81 | 5.81 |
| 7 | 5.27 | 5.12 | 5.09 | 5.13 | 5.10 | 5.23 | 5.23 | 5.20 | 5.19 |
| 8[d] | 4.38 | 5.11 | -[d] | 5.02 | -[d] | 5.17 | -[d] | 5.20 | -[d] |
| 9 | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] |
| 10 | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] |
| 11 | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] |
| 12 | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] |
| 13 | 6.69 | 6.40 | 6.11 | 6.43 | 6.11 | 6.29 | 6.02 | 6.21 | 5.97 |
| 14 | 5.47 | 5.55 | 5.70 | 5.53 | 5.62 | 5.45 | 5.43 | 5.55 | 5.56 |
| 15 | 5.7 | 5.66 | 5.66 | 5.91 | 6.04 | 5.64 | 5.63 | 5.67 | 5.65 |
| 16 | 5.67 | 5.87 | 5.95 | 5.57 | 5.54 | 5.81 | 5.85 | 5.69 | 5.69 |
| 17 | 4.97 | 5.01 | 5.02 | 5.18 | 5.21 | 4.99 | 5.00 | 4.88 | 4.84 |
| 18 | 5.62 | 5.61 | 5.61 | 5.48 | 5.43 | 5.41 | 5.33 | 5.65 | 5.67 |
| 19[d] | 4.66 | 4.92 | -[d] | 5.02 | -[d] | 4.84 | -[d] | 4.89 | -[d] |
| 20 | 5.26 | 5.49 | 5.57 | 5.41 | 5.46 | 5.60 | 5.72 | 5.58 | 5.68 |
| 21 | 4.89 | 4.97 | 5.00 | 5.01 | 5.05 | 4.88 | 4.87 | 4.89 | 4.90 |
| 22[d] | 5.08 | 5.05 | -[d] | 5.11 | -[d] | 4.90 | -[d] | 4.89 | -[d] |
| 23 | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] | -[e] |
| 24 | 5.59 | 5.51 | 5.48 | 5.63 | 5.65 | 5.46 | 5.39 | 5.50 | 5.46 |
| 25[d] | 6.07 | 5.60 | -[d] | 5.79 | -[d] | 5.65 | -[d] | 5.49 | -[d] |

[a]$IC_{50}$ represents the concentration of a compound required to bring out 50% inhibition of DPP-IV and the same is expressed as $pIC_{50}$ on molar basis; [b]Taken from ref. [29]; [c]Leave-one-out

(LOO) procedure; [d]Compound included in test set. [e]Compound with uncertain activity, not part of data set.

CP-MLR has also been carried out on DPP8 inhibitory activity from the pool of 90 descriptors which was used to find rationales for DPP-IV inhibitory activity. The analysis resulted into 10 models having $r^2_{Test} > 0.5$ and the highest significant four models are listed in Table 6. These are able to estimate up 84.82 percent of variance in observed DPP8 activity of the compounds. The newly appeared descriptors in above models are MR (property class descriptor), N-072 and H-047 (ACF class descriptor), X0Av and X2Av (TOPO class descriptor) and BELp2 (BCUT descriptor). Tabled Equations reveal that lower values of average valence connectivity indices (X0Av and X2Av, chi-0 and chi-2) would be advantageous to enhance the activity. On the other hand, a higher lower value of Ghose-Crippen molecular refractivity (MR) and lowest eigenvalue n.2 of Burden matrix weighted by atomic polarizabilities are incremental to the activity. Counts for certain structural fragments, H attached to C1(sp3) /C0(sp2) (descriptor H-047) and R-CO-N</>N-X=X (descriptor N-072) strongly recommend the presence of such structural features favorable to activity. Thus the descriptors identified for rationalizing the DPP-IV activity give avenues to rationalize the DPP8 inhibitory activity. From the different nature of emerged descriptors in final statistically significant models for DPP IV and DPP8 inhibition actions, it appeared that the mode of actions of titled compounds were different for DPP IV and DPP8 enzyme systems.

**Table 6**: **Three parameter CP-MLR models for the DPP-8 inhibition activity from the descriptor pool of DPP-IV.**

| Model | r | s | F | $q^2_{LOO}$ | $r^2_{Test}$ | Eq. |
|---|---|---|---|---|---|---|
| $pIC_{50}$ = 5.218 – 0.999(0.193)X2Av + 0.523(0.103)PJI2 + 1.075(0.230)MR | 0.920 | 0.182 | 20.490 | 0.690 | 0.545 | (vii) |
| $pIC_{50}$ = 5.510 – 0.737(0.203)X2Av + 0.464(0.118)PJI2 + 0.613(0.161)N-072 | 0.896 | 0.208 | 14.982 | 0.590 | 0.755 | (viii) |
| $pIC_{50}$ = 5.405 – 0.865(0.217)X2Av + 0.411(0.125)PJI2 + 0.702(0.195)BELp2 | 0.889 | 0.214 | 13.934 | 0.539 | 0.516 | (ix) |
| $pIC_{50}$ = 5.221 – 1.066(0.276)X0Av + 0.410(0.125)PJI2 + 0.871(0.231)H-047 | 0.888 | 0.215 | 13.730 | 0.562 | 0.762 | (x) |

[a]The models, in three parameters, emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.5 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 15 compounds.

***Applicability domain***

On analyzing the applicability domain (AD) in the Williams plot (Figure 2) of the model based on the whole data set (Table 7), none of the compound has been identified as an obvious 'outlier' for the DPP-IV inhibitory activity if the limit of normal values for the Y outliers (response outliers) was set as 3×(standard deviation) units. None of the compounds was found to have leverage (h) values greater than the threshold leverage (h*). For both the training-set and test-set, the suggested model matches the high quality parameters with good fitting power

and the capability of assessing external data. Furthermore, all of the compounds were within the applicability domain of the proposed model and were evaluated correctly.

**Table 7**: **Models derived for the whole data set (n = 25) for the DPP-IV inhibitory activity in descriptors identified through CP-MLR.**

| Model | r | s | F | $q^2_{LOO}$ | Eq. |
|---|---|---|---|---|---|
| $pIC_{50}$=4.707+2.088(0.260)JGI4 1.513(0.230)ATS8p+2.197(0.246)GATS8p+ 0.478(0.158)C-024 | 0.942 | 0.212 | 39.987 | 0.815 | (1a) |
| $pIC_{50}$=5.216–0.847(0.249)ATS8p–0.805(0.207)GATS7e+ 2.143(0.298)GATS8p + 1.365(0.254)MLOGP | 0.920 | 0.249 | 27.603 | 0.760 | (2a) |
| $pIC_{50}$=5.902–1.229(0.233)RBN+3.431(0.415)BIC3– 2.257(0.315)BIC5 + 1.866(0.217)H-052 | 0.919 | 0.250 | 27.370 | 0.658 | (3a) |
| $pIC_{50}$=3.827+2.966(0.412)BIC3+1.301(0.208)SRW09+0.809(0.223)C-040+ 2.471(0.305)H-052 | 0.895 | 0.283 | 20.219 | 0.582 | (4a) |

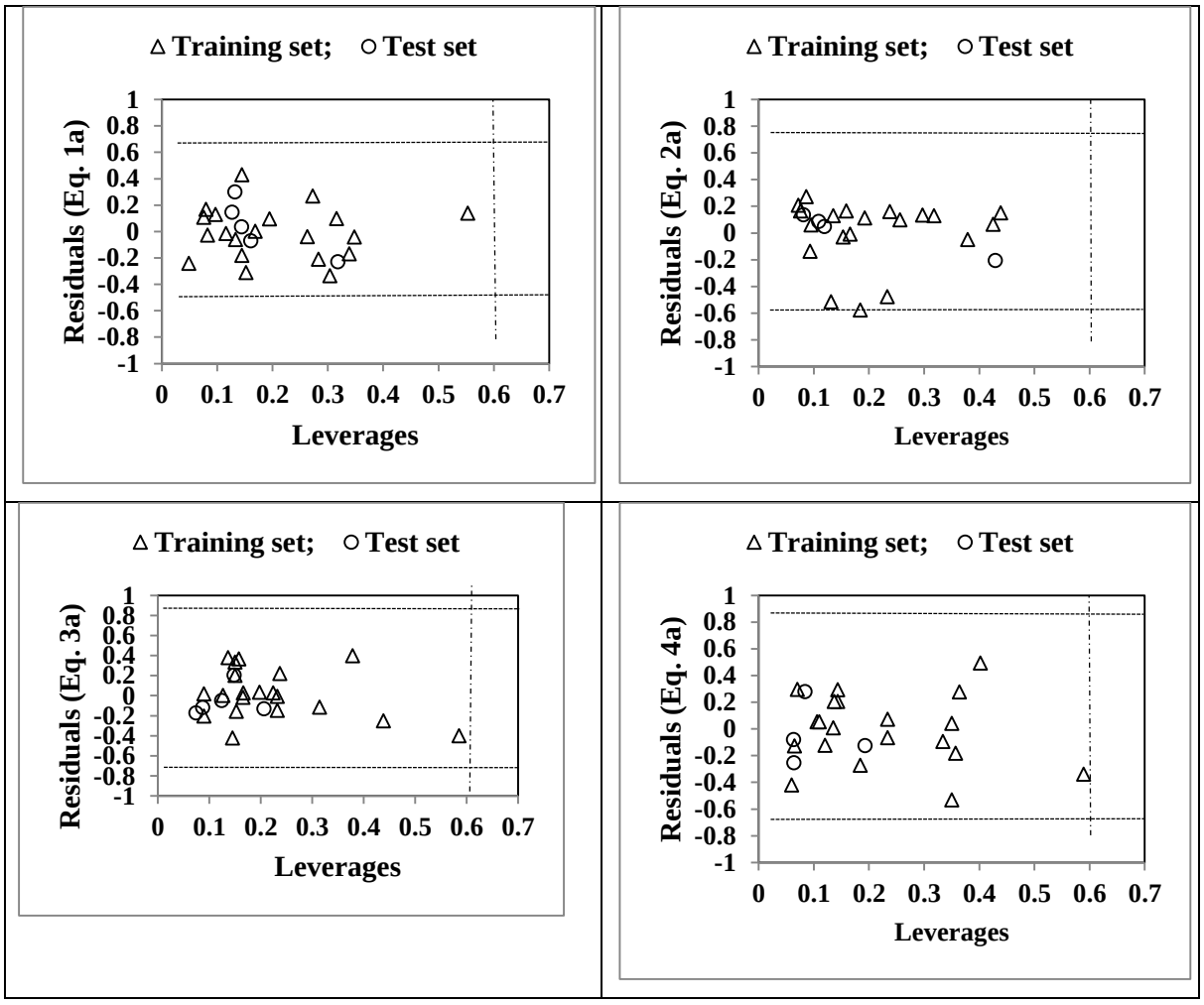



**Figure 2**. **Williams plot for the training-set and test- set for inhibition activity of DPP4 for the compounds in Table 1. The horizontal dotted line refers to the residual limit (±3×standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.6).**

## CONCLUSION

This study has provided a rational approach for the development of (2*S*)-cyanopyrrolidine analogues as DPP-IV inhibitors. The descriptors identified in CP-MLR analysis have highlighted the role of atomic properties in respective lags of 2D-autocorrelations (ATS8p, GATS8p and GATS7e), 4$^{th}$ order mean Galvez topological charge index (JGI4), 3$^{rd}$ and 5$^{th}$ order bond information content of neighborhood symmetry (BIC3 and BIC5) and 9$^{th}$ order self returning walk-count (SRW09) to explain the biological actions of (2*S*)-cyanopyrrolidine analogues as DPP IV inhibitors. Certain structural features or fragments (RBN, C-024, C-040 and H-052) in molecular structures in addition to hydrophobicity (MLOGP) of a molecule have also shown prevalence to optimize the DPP IV inhibitory activity of titled compounds. Applicability domain analysis revealed that the suggested model for DPP IV inhibitory activity matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

## REFERENCES

1. Holst, J. J. Glucagon-like peptide-1 (GLP-1): A newly discovered gastrointestinal hormone. Gastroenterology 1994, 107, 1848-1855.
2. Orskov C. Glucagon-like peptide-1, a new hormone of the enteroinsular axis. Diabetologia 1992, 35, 701-711.
3. Drucker, D. J. Glucagon-Like Peptides. Diabetes 1998, 47, 159-169.
4. Drucker, D. J. Minireview: The glucagon-like peptides. Endocrinology 2001, 142, 521-527.
5. Ross, S. A.; Gulve, E. A.; Wang, M. Chemistry and biochemistry of type 2 diabetes. Chem. Rev. 2004, 104, 1255-1282.
6. Ahren, B.; Larsson, H.; Holst, J. J. Effects of glucagon-like peptide-1 on islet function and insulin sensitivity in noninsulin dependent diabetes mellitus. J. Clin. Endocrinol. Metab. 1997, 82, 473-478.
7. Rachman, J.; Barrow, B. A.; Levy, J. C.; Turner, R. C. Near normalization of diurnal glucose concentrations by continuous administration of glucagon-like peptide 1 (GLP-1) in subjects with NIDDM. Diabetologia 1997, 40, 205-211.
8. Nauck, M. A.; Wollschlager, D.; Werner, J.; Holst, J. J.; Orskov, C.; Creutzfeldt, W.; Willms, B. Effects of subcutaneous glucagon-like peptide 1 (GLP-1 [7-36 amide]) in patients with NIDDM. Diabetologia 1996, 39, 1546-1553.
9. Drucker, D. J. Biological actions and therapeutic potential of the glucagon-like peptides. Gastroenterology 2002, 122, 531-544.

10. Wettergren, A.; Schjoldager, B.; Mortensen, P. E.; Myhre, J.; Christiansen, J.; Holst, J. J. Truncated GLP-1 (proglucagon 78-107-amide) inhibits gastric and pancreatic functions in man. Dig. Dis. Sci. 1993, 38, 665-673.
11. Nauck, M. A.; Niedereichholz, U.; Ettler, R.; Holst, J. J.; Orskov, C.; Ritzel, R.; Schmiegel, W. H. Glucagonlike peptide-1 inhibition of gastric emptying outweighs its Insulinotropic effects in healthy humans. Am. J. Physiol. 1997, 273, E981-E988.
12. Nauck, M. A. Is GLP-1 an incretin hormone? Diabetologia 1999, 42, 373-379.
13. Egan, J.; Clocquet, A. R.; Elahi, D. The insulinotropic effect of acute exendin-4-administered to humans: Comparison of nondiabetic state to type 2 diabetes. J. Clin. Endocrinol. Metab. 2002, 87, 1282-1290.
14. Gutniak, M.; Orskov, C.; Holst, J. J.; Ahren, B.; Efendic, S. Antidiabetogenic effect of glucagon-like peptide-1 (7-36)amide in normal subjects and patients with diabetes mellitus. N. Engl. J. Med. 1992, 326, 1316-1322.
15. Nauck, M. A.; Kleine, N.; Orskov, C.; Holst, J. J.; Willms, B.; Creutzfeldt, W. Normalization of fasting hyperglycaemia by exogenous glucagon-like peptide 1 (7-36 amide) in type 2 (non-insulin-dependent) diabetic patients. Diabetologia 1993, 36, 741-744.
16. Gutniak, M. K.; Linde, B.; Holst, J. J.; Efendic, S. Subcutaneous injection of the incretin hormone glucagon like peptide 1 abolishes postprandial glycemia in NIDDM. Diabetes Care 1994, 17, 1039-1044.
17. Villhauer, E. B.; Brinkman, J. A.; Naderi, G. B.; Burkey, B. F.; Dunning, B. E.; Prasad, K.; Mangold, B. L.; Russell. M. E.; 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. J. Med. Chem. 2003, 46, 2774-2789.
18. Deacon, C. F.; Ahren, B.; Holst, J. J. Inhibitors of dipeptidyl peptidase IV: A novel approach for the prevention and treatment of Type 2 diabetes? Expert. Opin. Investig. Drugs 2004, 13, 1091-1102.
19. 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.; Roy, R. S.; Wu, J. K.; Wyvratt, M. J.; Zhang, B. B.; Zhu, L.; Thornberry, N. A.; 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. J. Med. Chem. 2005, 48, 141-151.

20. Ahren, B.; Landin-Olsson, M.; Jansson, P. A.; Svensson, M.; Holmes, D.; Schweizer, A. Inhibition of dipeptidyl peptidase-4 reduces glycemia, sustains insulin levels, and reduces glucagon levels in type 2 diabetes. J. Clin. Endocrinol. Metab. 2004, 89, 2078-2084.
21. Weber, A. E., Kim, D.; Beconi, M.; Eiermann, G.; Fisher, M.; He, H.; Hickey, G.; Jaspal, S.; Leiting, B.; Lyons, K.; Marsilio, F.; McCann, P.; Moller, D. E.; Patel, R.; Petrov, A.; Pryor, K.; Sinha Roy, R.; Wu, J. K.; Wyvratt, M.; Zhang, B. B.; Thornberry, N. A. MK- 0431 is a potent, selective dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. Presented at the American Diabetes Association 64th Scientific Session, Orlando, FL, June 4-8, 2004, Abstract 633-P.
22. Yaron, A.; Naider, F. Proline-Dependent Structural and Biological Properties of Peptides and Proteins. Crit. Rev. Biochem. Mol. Biol.1993, 28, 31-81.
23. Mentlein, R.; Gallwitz, B.; Schmidt, W. E. Dipeptidyl-peptidase IV hydrolyses gastric inhibitory polypeptide, glucagon-like peptide-1(7-36)amide, peptide histidine methionine and is responsible for their degradation in human serum. Eur. J. Biochem. 1993, 214, 829-835.
24. Kieffer, T. J.; McIntosh, C. H. S.; Pederson, R. 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.
25. Deacon, C. F.; Nauck, M. A.; Toft-Nielsen, M.; Pridal, L.; Willms, B.; Holst, J. J. Both subcutaneously and intravenously administered glucagon-like peptide I are rapidly degraded from the NH2-terminus in type II diabetic patients and in healthy subjects. Diabetes 1995, 44, 1126-1131.
26. Ashworth, D. M.; Atrash, B.; Baker, G. R.; Baxter, A. J.; Jenkins, P. D.; Jones, D. M.; Szelke, M. 4-Cyanothiazolidides as very potent, stable inhibitors of dipeptidyl peptidase IV. Bioorg. Med. Chem. Lett. 1996, 6, 2745-2748.
27. Ashworth, D. M.; Atrash, B.; Baker, G. R.; Baxter, A. J.; Jenkins, P. D.; Jones, D. M.; Szelke, M. Bioorg. 2-Cyanopyrrolidides as potent, stable inhibitors of dipeptidyl peptidase IV. Med. Chem. Lett. 1996, 6, 1163-1166.
28. Ahren, B.; Gomis, R.; Standl, E.; Mills, D.; Schweizer, A. A. prolonged efficacy of LAF237 in patients with Type 2 diabetes (T2DM) inadequately controlled with metformin. 64th Scientific Sessions of the American Diabetes Association. Orlando, FL, 2004, Late-breaking 7-LB (Abstract).
29. Tsu, H.; Chen, X.; Chen, C.-T.; Lee, S.-J.; Chang, C.-N.; Kao, K.-H.; Coumar, M. S.;Yeh, Y.-T.; Chien,C.-H.; Wang, H.-S.; Lin, K.-T.; Chang, Y.-Y.; Wu, S.-H.; Chen, Y.-S.; Lu, I-L.; Wu, S.-Y.; Tsai, T.-Y.; Chen, W.-C.; Hsieh, H.-P.; Chao, Y.-S.; Jiaang

W.-T. 2- [3-[2-[(2S)-2-Cyano-1-pyrrolidinyl]-2-oxoethylamino]-3-methyl-1-oxobutyl]-1,2,3,4-tetrahydroisoquinoline: A Potent, Selective, and Orally Bioavailable Dipeptide-Derived Inhibitor of Dipeptidyl Peptidase IV. J.Med. Chem. 2006, 49, 373-380.

30. SYSTAT, Version 7.0; SPSS Inc 444 North Michigan Avenue, Chicago IL, 60611.
31. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. http://www.cambridgesoft.com
32. Dragon software (version 1.11-2001) by Todeschini R.; Consonni V. Milano, Italy. http//www.talete.mi.it/dragon.htm
33. Prabhakar, Y. S. A combinatorial approach to the variable selection in multiple linear regression: analysis of Selwood et al. Data Set-a case study. QSAR Comb. Sci. 2003, 22,583-595.
34. Golbraikh, A.; Tropsha, A. Beware of q2! J. Mol. Graph. Model. 2002, 20, 269-276.
35. Sharma, S.; Prabhakar, Y. S.; Singh, P.; Sharma, B. K. QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach, Eur J Med Chem 2008, 43, 2354-2360.
36. Sharma, S.; Sharma, B. K.; Sharma, S. K.; Singh, P.; Prabhakar, Y. S. Topological Descriptors in modeling the agonistic activity of human A3 adenosine receptor ligands: The derivatives of 2-Chloro-$N^6$-substituted-4'-thioadenosine-5'-uronamide, Eur. J. Med. Chem. 2009, 44, 1377-1382.
37. Sharma, B. K.; Pilania, P.; Singh, P.; 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 Environ. Res. 2010, 21, 169-185.
38. Sharma, B. K.; Singh, P.; Sarbhai, K.; Prabhakar, Y. S. A quantitative structure-activity relationship study on serotonin 5-HT6 receptor ligands: Indolyl and piperidinyl sulphonamides, SAR QSAR Environ. Res. 2010, 21, 369-388.
39. Sharma, B. K.; Pilania, P.; Sarbhai, K.; Singh, P.; Prabhakar, Y. S. Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. Mol. Divers. 2010, 14, 371-384.
40. Sharma, B. K.; Singh, P.; Shekhawat, M.; Sarbhai, K.; Prabhakar, Y. S. Modelling of serotonin reuptake inhibitory and histamine $H_3$ antagonistic activity of piperazine and diazepane amides: QSAR rationales for co-optimization of the activity profiles. SARQSAR Environ. Res. 2011, 22, 365-383

41. So, S. –S.; Karplus, M. Three-dimensional quantitative structure activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations. J. Med. Chem. 1997, 40, 4347-4359.
42. Prabhakar, Y. S.; Solomon, V. R.; Rawal, R. K.; Gupta, M. K.; 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 Comb. Sci. 2004, 23, 234-244.
43. Gramatica, P. Principles of QSAR models validation: internal and external. QSAR Comb.Sci. 2007, 26, 694-701.
44. Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. Env. Health Persp. 2003, 111, 1361-1375.