*Available online at* www.derpharmachemica.com

Scholars Research Library

**Der Pharma Chemica, 2011, 3 (4):158-170**
*(http://derpharmachemica.com/archive.html)*

**ISSN 0975-413X**
**CODEN (USA): PCHHAX**

# 2D QSAR Analysis on 5β-Methylprolyl-2-Cyanopyrrolidine Derivatives as DPP IV Inhibitors

**Sanmati K. Jain*, Sarika Vishwakarma and Pragya Nayak**

*SLT Institute of Pharmaceutical Sciences, Guru Ghasidas Vishwavidyalaya, Bilaspur (Chhattisgarh), India*

---

## ABSTRACT

*Two dimensional quantitative structure activity relationship (2D QSAR) study was performed on 5β-methylprolyl-2-cyanopyrrolidine derivatives as dipeptidyl peptidase IV (DPP IV) inhibitors using molecular design suite software (VLifeMDS). This study was performed with 30 compounds (data set) using sphere exclusion (SE) algorithm, random and manual selection methods for the division of the data set into training and test set. Partial least square (PLS) linear regression analysis coupled with stepwise variable selection method was applied to derive QSAR models which were further validated for statistical significance by internal and external validation. The most significant model has squared correlation coefficient (r2), cross validated correlation coefficient (q2) and predictive correlation coefficient (pred_r2) 0.6231, 0.5109 and 0.3862 respectively. The QSAR model indicates that the descriptors T_C_O_4 [This is the count of number of Carbon atoms (single, double or triple bonded) separated from Oxygen atom (single or double bonded) by 4 bond distance in a molecule], SdssCE-index [Electrotopological state indices for number of carbon atom connected with two single bonds] and XlogP [This descriptor signifies ratio of solute concentration in octanol & water and generally termed as octanol water partition coefficient] contributing 48.19%, 28.93% and 22.88 % respectively. Negative coefficient value of T_C_O_4 and XlogP indicated that lower value leads to better dipeptidyl peptidase inhibitory activity whereas higher value leads to decrease activity whereas positive coefficient value of SdssCE-index indicated that higher value leads to good dipeptidyl peptidase inhibitory activity while lower value leads to reduced activity.*

**Keywords:** 2D-QSAR, PLSR, DPP IV inhibitors, 5β-methylprolyl-2-cyanopyrrolidine derivatives.

---

## INTRODUCTION

Diabetes mellitus is a metabolic disorder, which is a major public health issue all over the world. In the year 2025, it is expected that about 333 million people will suffer from the disease, with type 2 diabetes mellitus (T2DM) [1]. Type 2 diabetes mellitus (T2DM) is a persistent disease, which is characterized by insulin resistance, excess hepatic glucose production and progressive pancreatic β cell dysfunction [2]. Among the various promising targets, the development of dipeptidyl peptidase IV inhibitors appears to be one of the most attractive, rational agents for the treatment of T2DM [3].

Dipeptidyl peptidase-IV inhibitors (DPP-IV inhibitors) inhibit the enzyme dipeptidyl peptidase-IV and having usefulness in the treatment of type 2 diabetes. Inhibition of the DPP- IV prolongs and enhances the activity of incretins that play an important role in insulin secretion and blood glucose control regulation. Dipeptidyl peptidase IV also known as T-cell antigen CD26 [4,5] is a member of serine protease family that selectively cleaves dipeptide from polypeptides, including proline or alanine at the N-terminal penultimate position [6,7]. DPP- IV acting as a peptidase that implicated in the degradation of two insulin-sensing hormones, glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP) [8,9]. GLP-1 is an incretin hormone, secreted by intestinal L-cells in response to meal ingestion. It stimulates the secretion of insulin, inhibits glucagon release, slows gastric emptying and induces satiety, helps in the control of glucose homeostasis in patients with type-2 diabetes [10]. Thus inhibition of DPP-IV extends the half-life of endogenously secreted GLP-1, which in turn enhances insulin secretion and improves the glucose tolerance.

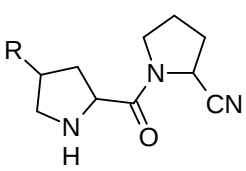
Quantitative structure activity relationship (QSAR) is an accepted means for establishing quantitative relationship between biological activity and descriptors representing physicochemical properties of the compounds in a series using statistical methods and it helps to predict the biological activities of newly designed analogues contributing to the drug discovery processes [11].

The main aim of the present QSAR study is to establish quantitative relationship between physiochemical properties and biological activities of the compounds in order to search for novel 2-cyanopyrrolidine derivatives that would show a promise to become useful DPP-IV inhibitors. A series of 5β-methylprolyl-2-cyanopyrrolidine derivatives [12] which were reported as DPP-IV inhibitors selected for QSAR study using VlifeMDS software [13].

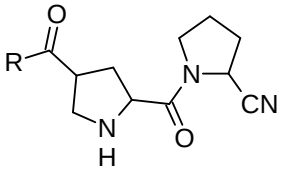
## MATERIALS AND METHODS

***Data Set:*** In the present study a data set of 5β-methylprolyl-2-cyanopyrrolidine derivatives (30 molecules) [12] has been taken from the literature for QSAR studies reported in Table-1. The reported $IC_{50}$ values (nM), have been changed to the logarithmic scale [$pIC_{50}$ (moles)], for QSAR study.

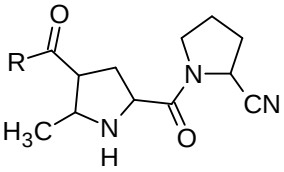
**Table-1: General structure of the compounds of 5β-methylprolyl 2-cyanopyrrolidine derivatives and their biological activities (data set of 30 molecules)**



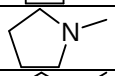
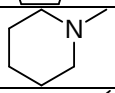
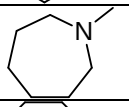
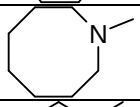
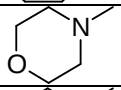
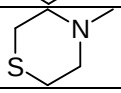
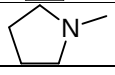

1-9

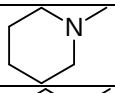
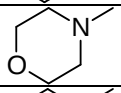
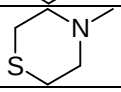
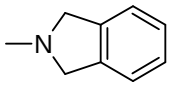



10-22




23-30

| S. No. | Compound | R | $IC_{50}$ (nM) | log(1/$IC_{50}$) |
|---|---|---|---|---|
| 1 | 4 | -H | 20 | 8.69 |
| 2 | 6 | -Allyl | 3.5 | 8.45 |
| 3 | 7 | - *n*-propyl | 3.4 | 8.46 |
| 4 | 8 | -$CH_2OH$ | 6.3 | 8.2 |
| 5 | 9 | -$CH_2CH_2OH$ | 3.8 | 8.42 |
| 6 | 10 | -$CH_2COOH$ | 16 | 7.79 |
| 7 | 11 | -$CONMe_2$ | 1.8 | 8.74 |
| 8 | 12 | - $CH_2CONMe_2$ | 4.5 | 8.34 |
| 9 | 13 | -$CH_2CH_2CONMe_2$ | 2.4 | 8.61 |
| 10 | 14 | -$NH_2$ | 7.7 | 8.11 |
| 11 | 15 | -NHMe | 3.5 | 8.45 |
| 12 | 16 | -NHEt | 4.4 | 8.35 |
| 13 | 17 | -NHPr | 5.7 | 8.24 |
| 14 | 18 | -N(Me)Et | 4.9 | 8.30 |
| 15 | 19 | -$NEt_2$ | 6.0 | 8.22 |
| 16 | 20 |  | 4.0 | 8.39 |
| 17 | 21 |  | 5.6 | 8.25 |
| 18 | 22 |  | 4.9 | 8.30 |
| 19 | 23 |  | 8.3 | 8.08 |
| 20 | 24 |  | 11 | 7.95 |
| 21 | 25 |  | 4.3 | 8.36 |
| 22 | 26 |  | 4.8 | 8.31 |
| 23 | 27 | -$NMe_2$ | 10 | 9 |
| 24 | 28 | -N(Me)Et | 15 | 7.82 |
| 25 | 29 |  | 13 | 7.88 |
| 26 | 30 |  | 6.5 | 8.18 |

| | | | | |
|---|---|---|---|---|
| 27 | 31 |  | 13 | 7.88 |
| 28 | 32 |  | 20 | 7.69 |
| 29 | 33 |  | 6.2 | 8.20 |
| 30 | 34 |  | 5.6 | 8.25 |

∗ *Compounds having $IC_{50}$ >316 (nM) in the selected series were excluded and the resulting total data set of 30 molecules were used for the present study.*

***Molecular modeling:*** Molecular modeling and PLS studies were performed on HCL computer having genuine Intel Pentium Dual Core Processor and Windows XP operating system using the software Molecular Design Suite (VLifeMDS).

Structures were drawn using the 2D draw application and converted to 3D structures. Structures were optimized by energy minimization and geometry optimization was done using Universal Force Field method (**UFF)** with 10000 as maximum number of cycles, 0.01 as convergence criteria (root mean square gradient) and 1.0 as constant (medium's dielectric constant which is 1 for in vacuo) in dielectric properties. The default values of 20.0 and 10.0 Kcal/mol were used for electrostatic and steric energy cutoff.

***Descriptors used in the QSAR analysis:*** Number of physicochemical, alignment and atom type independent descriptors was calculated using Molecular Design Suite software after optimization or minimization of the energy of the data set molecules.

**Physicochemical descriptors:**

1. Individual –( H-Acceptor count, H-Donor count , X log p, smr, polarisablity)
2. Chi – (Chi0-05)
3. ChiV – (ChiV 0-05)
4. Path count - (0pathcount-05pathcount)
5. Chi chain – (Chi3chain – Chi6chain)
6. Chi v chain – (Chi3chain – Chi6chain)
7. Chain path count – 3chain path count, 6chain path count)
8. Cluster – (Chi3 cluster, ChiV6 cluster, 3cluster count)
9. Path cluster – (Chi4path cluster ChiV4 path cluster, 4path cluster count)
10. Kapa – (Kapa1,2,3), ( k1 alpha – k3 alpha)
11. Element count – (H, N, C, S, O, Cl)
12. Estate numbers – (SsCH3 Count, SdCH2 Count, SssCH2 Count, StCH count etc.)
13. Estate contribution – (SsCH3-index., SdCH2- index, SssCH2 – index , StCH – index_
14. Polar surface area – Polar surface area excluding, polar surface including.

**Alignment independent descriptors:** More than 200 descriptors are calculated using the following attributes. A few examples are T_2_O_7, T_2_N_5, T_2_2_6, T_C_O_1, T_O_Cl_5 etc.

| Structural descriptors | Selected Attributes |
|---|---|
| *Topological | 2 |
| **Range** | 3 |
| Min – 0 | T (any) |
| Max. - 7 | C |
| | N |
| | O |
| | F |
| | S |
| | Cl |

**Atom Type Count Descriptors:** The atom type count descriptors are based on MMFF atom types and their count in each molecule. In MMFF, there are 99 atom types and hence 99 descriptors indicating number of times that atom has occurred in a given molecule are generate.

***Generation of training and test set of compounds:*** In order to evaluate the QSAR model, data set was divided into training and test set using Sphere Exclusion, random and manual data selection methods. Training set is used to develop the QSAR model for which biological activity data are known. Test set is used to challenge the QSAR model developed based on the training set to assess the predictive power of the model which is not included in model generation.

**Data selection:** Biological activity taken as dependent variable and descriptors as independent variable. Following methods were for creation of training and test set.

- Sphere Exclusion method
- Random selection method
- Manual data selection

**Sphere Exclusion method:** In this method initially data set were divided into training and test set using sphere exclusion method. In this method dissimilarity value provides an idea to handle training and test set size. It needs to be adjusted by trial and error until a desired division of training and test set is achieved. Increase in dissimilarity value results in increase in number of molecules in the test set.

**Random selection:** In order to construct and validate the QSAR models, both internally and externally, the data sets were divided into training [85% of total data set] and test sets (15%) in a random manner. Ten trials were run.

**Manual data selection:** Whole range of activities was sorted on the basis of results obtained in sphere exclusion and random methods.

After the creation of training and test set, Min and Max value of the test and training set is checked, using the QSAR tool, if the values are not following the Min – Max, then the training / test set is again set and procedure is repeated. If the Min – Max is following, then Partial Least Squares Regression (PLSR) used for model building.

These methods were run using the following criteria.

| | |
|---|---|
| **Step wise variable selection method**<br>* Forward -back ward<br>Cross correlation Limit – < 0.5<br>No. of variables – $1/5^{th}$ of total training set<br>Term selection – $r^2$ | F test<br>In – 4.00<br>Out – 3.99 |

Model building criteria – Cross validation

***Partial least square regression (PLSR):*** PLSR was used for model generation. PLSR is an expansion of the multiple linear regression (MLR). In its simplest form, a linear model specifies the (linear) relationship between a dependent variable and a set of predictor variables. In PLSR, prediction functions are represented by factors extracted from the Y'XX'Y matrix. The number of such prediction functions that can be extracted typically will exceed the maximum of the number of Y and X variables. PLSR is probably the least restrictive of the various multivariate extensions of the multiple linear regression models. PLSR can be used as an exploratory analysis tool to select suitable predictor variables and to identify outliers before classical linear regression. All the calculated descriptors were considered as independent variable and biological activity as dependent variable.

## RESULTS AND DISCUSSION

Selected data set, 5β-methylprolyl-2-cyanopyrrolidine derivatives was subjected to partial least square regression analysis method for model building. Result of PLSR analysis using sphere exclusion, random and manual data selection methods is shown in Table-3, 4 and 5 respectively. The statistically significant model obtained is shown in Table-6.

Different training and test set of 5β-methylprolyl-2-cyanopyrrolidine derivatives were constructed using sphere exclusion (dissimilarity level 1.0 to 2.3), random and manual data selection methods. Training and test set were selected if they follow the Unicolumn statistics (Table-2). This result shows that the test is interpolative i.e., derived from the min-max range of training set. The mean and standard deviation of the training and test set provides insight to the relative difference of mean and point density distribution of the two sets.

Partial least squares regression analysis (PLSR) in conjunction with stepwise (SW) forward-backward was applied for building QSAR models.

**Table-2: Uni-Column Statistics for Model 1 for training and test set activity.**

| Column Name | Average | Max | Min | StdDev | Sum |
|---|---|---|---|---|---|
| **Training set** | 8.2378 | 8.7400 | 7.7900 | 0.2412 | 189.4700 |
| **Test set** | 8.3486 | 9.0000 | 7.6900 | 0.4319 | 58.4400 |

**Table-3: Result of PLSR study using sphere exclusion selection method**

| Trial | Dissimilarity value | Test Set | PLSR | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $r^2$ | $q^2$ | pred_$r^2$ | $r^2$se | $q^2$se | pred_$r^2$ se |
| 1 | 1 | 28,11,19,16 | 0.4805 | 0.3398 | 0.4597 | 0.2097 | 0.2364 | 0.2794 |
| 2 | 1.1 | 28,11,19,24,16 | 0.4706 | 0.3140 | 0.5139 | 0.2108 | 0.2400 | 0.2553 |
| 3 | 1.2 | 18,19,28,15,17,24 | 0.5212 | 0.3102 | -0.8543 | 0.2146 | 0.2576 | 0.3693 |
| 4 | 1.4 | 11,18,19,28,20,24,16 | 0.4807 | 0.1667 | -0.1158 | 0.2177 | 0.2757 | 0.3196 |
| 5 | 1.42 | 11,18,19,28,20,21, 24,16 | 0.4901 | 0.1661 | -0.1646 | 0.2210 | 0.2826 | 0.3024 |
| 6 | 1.45 | 11,18,19,28,20,21,22, 24,16,25 | 0.6004 | 0.2817 | -0.7500 | 0.2057 | 0.2757 | 0.3299 |
| 7 | 1.56 | 11,18,19,28,20,21,22, 24,16,25,26 | 0.6443 | 0.3248 | -1.1115 | 0.1995 | 0.2749 | 0.3447 |
| 8 | 1.69 | 11,18,19,28,20,21,22, 24,15,16,25,26 | 0.6396 | 0.2887 | -0.9557 | 0.2057 | 0.2879 | 0.3287 |
| 9 | 1.93 | 11,18,19,28,20,21,22, 23,15,16,25,6,26 | 0.6389 | 0.3171 | -1.0876 | 0.2126 | 0.2924 | 0.3287 |
| 10 | 2.3 | 11,18,19,28,20,21,22, 23,15,16,25,9,6,26 | 0.7770 | 0.2608 | -1.8896 | 0.1770 | 0.2608 | 0.3921 |

**Table-4: Result of PLSR study using Random data selection method (85% training set)**

| Trial | Test Set | PLSR | | | | | |
|---|---|---|---|---|---|---|---|
| | | $r^2$ | $q^2$ | pred_$r^2$ | $r^2$se | $q^2$se | pred_$r^2$ se |
| 1 | 19,20,26,4,9 | 0.3536 | 0.2469 | 0.7701 | 0.2497 | 0.2695 | 0.1248 |
| 2 | 10,14,33,8,9 | 0.4536 | 0.3585 | -0.0643 | 0.2270 | 0.2460 | 0.2881 |
| 3 | 19,23,29,32,9 | 0.3230 | 0.2134 | 0.6028 | 0.2356 | 0.2540 | 0.2500 |
| 4 | 14,20,26,32,9 | 0.3801 | 0.2819 | 0.4714 | 0.2358 | 0.2358 | 0.2325 |
| 5 | 25,30,32,7,9 | 0.3649 | 0.2529 | 0.5455 | 0.2382 | 0.2584 | 0.2164 |
| 6 | 13,28,29,6,9 | 0.3905 | 0.2769 | 0.3853 | 0.2268 | 0.2470 | 0.2840 |
| 7 | 14,17,29,32,9 | 0.5976 | 0.5249 | -0.3932 | 0.1865 | 0.2020 | 0.4606 |
| 8 | 14,28,29,33,9 | 0.6435 | 0.5254 | -0.8956 | 0.1817 | 0.2098 | 0.4697 |
| 9 | 24,28,30,32,9 | 0.4362 | 0.2259 | 0.1736 | 0.2069 | 0.2424 | 0.4042 |
| 10 | 19,28,30,33,9 | 0.5267 | 0.3352 | -1.1545 | 0.2138 | 0.2534 | 0.3702 |

**Table-5: Result of PLSR study using Manual data selection method**

| Trial | Trial | Test Set | PLSR | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $r^2$ | $q^2$ | pred_$r^2$ | $r^2$se | $q^2$se | pred_$r^2$ se |
| 1 | 1 | 4,9,32 | 0.2835 | 0.1770 | 0.7535 | 0.2353 | 0.2519 | 0.2569 |
| 7 | 2 | 4,9,32,15 | 0.2703 | 0.1578 | 0.7639 | 0.2399 | 0.2578 | 0.2124 |
| 28. | 3 | 4,9,32,8 | 0.6789 | 0.2016 | 0.6789 | 0.2354 | 0.2531 | 0.2403 |
| 29. | 4 | 4,9,32,14,27 | 0.6291 | 0.5378 | 0.3735 | 0.1509 | 0.1684 | 0.4211 |
| 36. | 5 | 4,9,32,23,14,15,27 | **0.6231** | **0.5109** | **0.3862** | **0.1553** | **0.1769** | **0.3511** |
| 57. | 6 | 4,9,27 | 0.7267 | 0.5758 | 0.1623 | 0.1401 | 0.1746 | 0.6086 |
| 82. | 7 | 4,32,27 | 0.5880 | 0.4584 | 0.4164 | 0.1550 | 0.1777 | 0.5614 |
| 91. | 8 | 4,32,27,8,15 | **0.6211** | **0.4745** | **0.4106** | 0.1526 | 0.1798 | 0.4092 |
| 101. | 9 | 4,32,27,8,24 | **0.6079** | **0.4512** | **0.4216** | 0.1527 | 0.1806 | 0.4095 |
| 107. | 10 | 4,32,27,8,31 | **0.6090** | **0.4724** | **0.4231** | 0.1496 | 0.1738 | 0.4171 |

**Table-6: Statistical significant models generated**

| Model | Trial no. (Manual) | Test set molecules | Equation |
|---|---|---|---|
| 1 | 5 | 4,9,32,23, 14,15,27 | $pIC_{50}$ = -0.1277 T_C_O_4 + 0.3360 SdssCE-index - -0.1194 XlogP + 9.14873<br>Optimum Components = 2; n = 23; Degree of freedom = 20;<br>r2 = 0.6231; q2 = 0.5109; r2 se = 0.1553;<br>q2 se = 0.1769; pred_r2 = 0.3862; pred_r2se = 0.3511;<br>F test = 16.5340 Alpha Rand R^2 = 0.00004;<br>Alpha Rand Q^2 = 0.001; Alpha Rand Pred R^2 = 0.05 |
| 2 | 8 | 4,32,27,8, 15 | $pIC_{50}$ = -0.1286 T_C_O_4 + 0.38427 SdssCE-index - 0.08778 XlogP<br>Optimum Components = 2; n = 23; Degree of freedom = 20;<br>r2 = 0.6079; q2 = 0.4512; r2 se = 0.1527;<br>q2 se = 0.1806; pred_r2 = 0.4216; pred_r2se = 0.4095;<br>F test = 17.0542 Alpha Rand R^2 = 0.00000;<br>Alpha Rand Q^2 = 0.001; Alpha Rand Pred R^2 = 0.05 |
| 3 | 10 | 4,32,27,8, 31 | $pIC_{50}$ = -0.1277 T_C_O_4 + 0.3360 SdssCE-index - -0.1194 XlogP + 9.14873<br>Optimum Components = 2; n = 23; Degree of freedom = 20;<br>r2 = 0.6090; q2 = 0.4724; r2 se = 0.1496;<br>q2 se = 0.1738; pred_r2 = 0.4231; pred_r2se = 0.4171;<br>F test = 17.1356 Alpha Rand R^2 = 0.00001 ;<br>Alpha Rand Q^2 = 0.001; Alpha Rand Pred R^2 = 0.05 |

Data fitness plot for model 1 is shown in Figure 2. Result of the observed and predicted biological activity for the training and test compounds for the Model 1 is shown in Table 7. Descriptors used in the Model 1 are shown in Table 8. The plot of observed vs. predicted activity of training and test sets for model 1 is shown in Figure 3. From the plot it can be seen that model is able to predict the activity of training set quite well (all points are close to regression line) as well as external. Sphere exclusion (SE) algorithm, random and manual selection methods were used for constructing training and test sets. PLSR was used for building the QSAR models. In the present data set, sphere exclusion and random selection method does not result in any predictive model.

**Interpretation of the Model 01 (Most significant)**

Among the three significant models generated (Table-06), model 1 is the most significant one as it is having the highest cross validated correlation coefficient value. The equation explains 62% (r2 = 0.6231) of the total variance in the training set and has an internal (q2) and external (pred_r2) predictive ability of ~51% and ~39% respectively. The F test shows the statistical significance of 99.99 % of the model which means that probability of failure of the model is 1 in 10000. In addition, the randomization test shows confidence of 95 (Alpha Rand Pred R^2 = 0.05) that the generated model is not random and hence may be chosen as the QSAR model.

**Table 7: Actual and predicted biological activity for model 1**

| Compound | Actual | Predicted |
|---|---|---|
| 4* | 8.69 | 8.52 |
| 6 | 8.45 | 8.45 |
| 7 | 8.46 | 8.39 |
| 8 | 8.20 | 8.47 |
| 9* | 8.42 | 8.31 |
| 10 | 7.79 | 7.75 |
| 11 | 8.74 | 8.48 |
| 12 | 8.34 | 8.35 |
| 13 | 8.61 | 8.45 |
| 14* | 8.11 | 8.40 |
| 15* | 8.45 | 8.47 |
| 16 | 8.35 | 8.30 |
| 17 | 8.24 | 8.26 |
| 18 | 8.30 | 8.32 |
| 19 | 8.22 | 8.32 |
| 20 | 8.39 | 8.40 |
| 21 | 8.25 | 8.22 |
| 22 | 8.30 | 8.17 |
| 23* | 8.08 | 7.93 |
| 24 | 7.95 | 7.88 |
| 25 | 8.36 | 8.18 |
| 26 | 8.31 | 8.32 |
| 27* | 9.00 | 8.29 |
| 28 | 7.82 | 8.13 |
| 29 | 7.88 | 8.21 |
| 30 | 8.18 | 8.03 |
| 31 | 7.88 | 7.98 |
| 32* | 7.69 | 7.99 |
| 33 | 8.20 | 8.13 |
| 34 | 8.25 | 8.26 |

**Indicates that compounds are in the test set*

**Table 8: List of descriptors used in the training set for model 1**

| Compound | T_C_O_4 | SdssCE-index | XlogP |
|---|---|---|---|
| 6 | 5 | 0.11 | 0.78 |
| 7 | 5 | 0.15 | 1.44 |
| 8 | 6 | 0.03 | -0.69 |
| 10 | 9 | -0.88 | -0.39 |
| 11 | 6 | 0.03 | -0.77 |
| 12 | 7 | 0.12 | -0.48 |
| 13 | 6 | 0.19 | -0.03 |
| 16 | 7 | -0.03 | -0.49 |
| 17 | 7 | -0.01 | -0.04 |
| 18 | 7 | 0.07 | -0.35 |
| 19 | 7 | 0.07 | -0.35 |
| 20 | 7 | 0.17 | -0.73 |
| 21 | 8 | 0.18 | -0.28 |
| 22 | 8 | 0.19 | 0.17 |
| 24 | 9 | 0.2 | 1.57 |
| 25 | 9 | 0.09 | -1.29 |
| 26 | 8 | 0.24 | -0.94 |
| 28 | 8 | 0.06 | 0.15 |
| 29 | 8 | 0.17 | -0.23 |
| 30 | 9 | 0.18 | 0.22 |
| 31 | 9 | 0.19 | 0.67 |
| 33 | 9 | 0.23 | -0.43 |
| 34 | 10 | 1.14 | 0 |

In the QSAR model 1, the negative coefficient value of T_C_O_4 [This is the count of number of Carbon atoms (single, double or triple bonded) separated from Oxygen atom (single or double bonded) by 4 bond distance in a molecule] on the biological activity indicated that lower value leads to better dipeptidyl peptidase inhibitory activity (compound 7, 11, 13 etc.) whereas higher value leads to decrease activity (compound 8, 24, 30, 31, 33 etc.). Positive coefficient value of SdssCE-index [Electrotopological state indices for number of carbon atom connected with two single bonds] on the biological activity indicated that higher values leads to good dipeptidyl peptidase inhibitory activity (compound 7, 13, 20, 22 etc.) while lower value leads to reduced activity (compound 8, 10, 17, 28 etc.). Negative coefficient value of XlogP [This descriptor signifies ratio of solute concentration in octanol & water and generally termed as octanol water partition coefficient] indicated that lower values leads to better DPP IV inhibitory activity (compound 11, 20, 25, 26 etc.) while higher value leads to reduced activity (compound 24, 28, 30, 31 etc.). Contribution

chart for model 1 reveals that the descriptors T_C_O_4, SdssCE-index and XlogP contributing 48.19%, 28.93% and 22.88 % respectively.

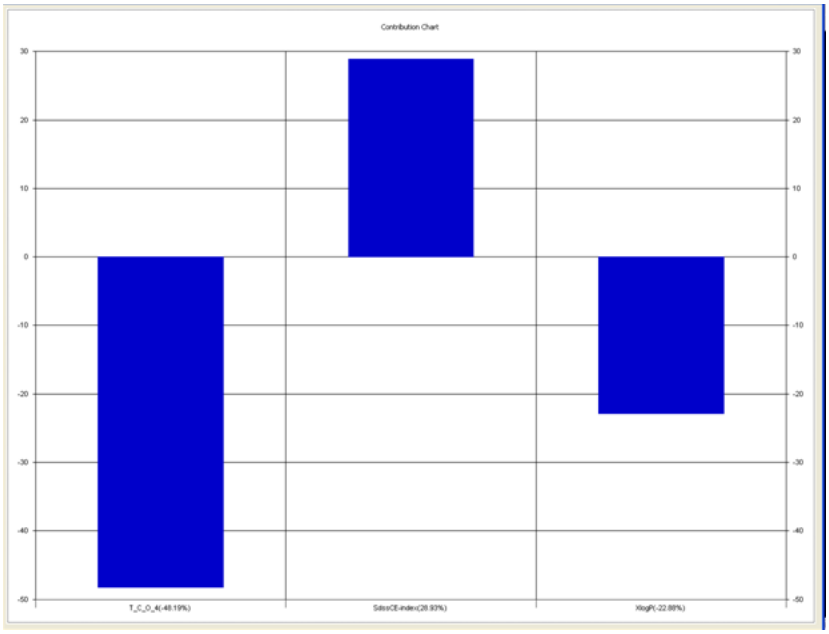



**Figure 1: Contribution chart for model 1 showing contribution of different descriptors.**

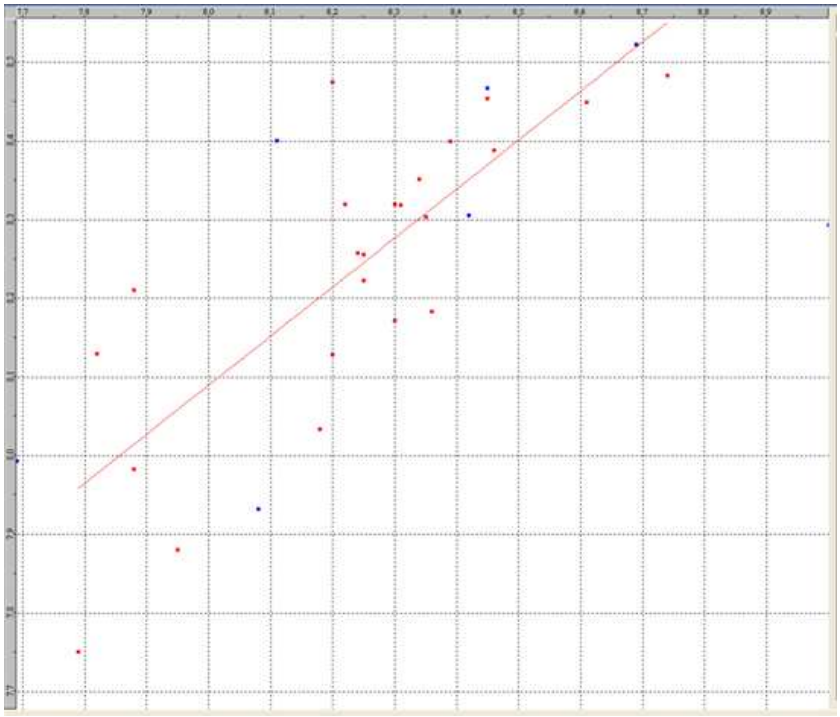


**Figure 2: Data fitness plot for models 1**

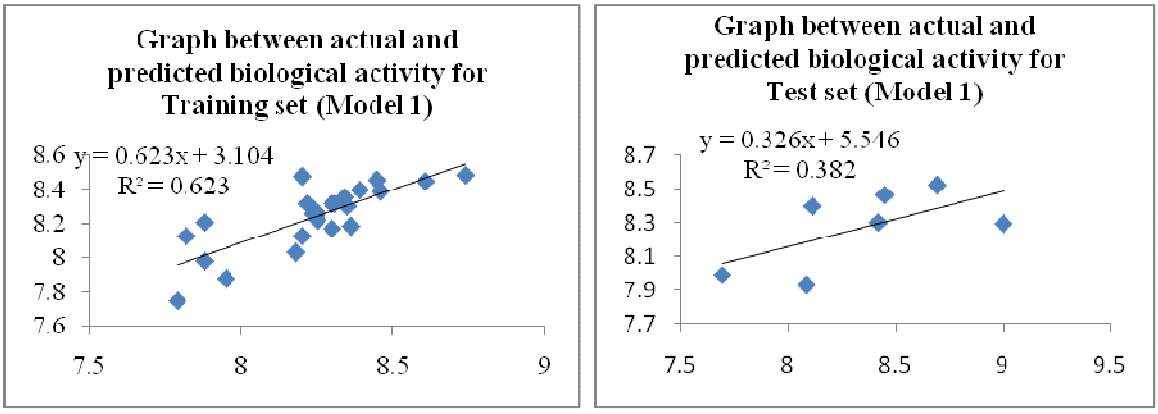



**Figure 3: Graph between actual and predicted biological activity for training and test set (Model-1).**

The observed vs. predicted activity provides an idea about how well the model was trained and how well it predicts the activity of the external test set. From the plot it can be seen that model is able to predict the activity of training set quite well (all points are close to the regression line) as well as external test set providing confidence in the predictive ability of the model.

## CONCLUSION

Two dimensional quantitative structure activity relationship (2D QSAR) study by means of partial least square regression (PLSR) method was performed on a series of 5β-methylprolyl-2-cyanopyrrolidine derivatives as dipeptidyl peptidase IV (DPP IV) inhibitors using molecular design suite (VLifeMDS). This study was performed with 30 compounds (data set) using sphere exclusion (SE) algorithm, random and manual selection methods for the division of the data set into training and test set. PLSR methodology with stepwise (SW) forward-backward variable selection method was used for building the QSAR models. Statistically significant QSAR models were generated. Among them most significant model has squared correlation coefficient (r2), cross validated correlation coefficient (q2) and predictive correlation coefficient (pred_r2) 0.6231, 0.5109 and 0.3862 respectively. The QSAR model indicates that the descriptors T_C_O_4, SdssCE-index and XlogP contributing 48.19%, 28.93% and 22.88 % respectively to biological activity. The negative coefficient value of T_C_O_4 and XlogP on the biological activity indicated that lower value leads to better dipeptidyl peptidase IV (DPP IV) inhibitory activity whereas higher value leads to decrease activity. Positive coefficient value of SdssCE-index indicates that higher value leads to better dipeptidyl peptidase IV (DPP IV) inhibitory activity whereas lower value leads to decrease activity.

**Acknowledgement**
The authors are indebted to the Head, SLT Institute of Pharmaceutical Sciences, Guru Ghasidas University, Bilaspur (CG) for providing necessary facilities. PN is thankful to AICTE for GPAT scholarship.

## REFERENCES

[1] R.E. Pratley, A. Salsali, *Curr. Med. Res. Opin.*, **2007**, 23, 919-31.
[2] P.V. Bharatam, D.S. Patel, L. Adane, A. Mittal, S. Sundriyal, *Curr. Pharm. Des.*, **2007**, 13, 3518-30.
[3] A. Barnett, *Int. J. Clin. Pract.*, **2006**, 60, 1454-70.
[4] C.D. Haffner, D.L. McDougald, S.M. Reister, B.D. Thompson, D. Lenhard, P.R. Johnson, *Bioorg. Med. Chem. Lett.*, **2005**, 15, 5257-5261.
[5] V.K. Hopsu-Havu, G.G. Glenner, *Histochem.*, **1966**, 7, 197.
[6] M. Abe, T. Akiyama, Y. Umezawa, K. Yamamoto, H. Yamajaki, M. Nagai, Y. Muraoka, *Bioorg. Med. Chem.*, **2005**, 13,785-797.
[7] J. Heins, P. Welker, C. Schonlein, I. Born, B. Hartrodt, K. Neubert, A. Tsuru, *Biochem. Biophys. Acta*, **1988**, 954 (2), 161-9.
[8] H. Sakashita, F. Akahoshi, H. Kitajima, R. Tsutsumiuchi, Y. Hayashi, *Bioorg. Med. Chem.*, **2006**; 14, 3662-3671.
[9] C.F. Deacon, A.H. Johnson, J.J. Holst, *Clin. Endocrinol. Metab.*, **1995**, 80, 952-957.
[10] W.T. Jiaang, T.Y. Tsai, M.S. Coumar, T. Hsu, H.P. Hsieh, C.H. Chien, C.T. Chen, C.N. Chang, Y.W. Huang, X. Chen, *Bioorg. Med. Chem. Lett.*, **2006**, 16, 3268-3272.
[11] M.M.C. Ferreira, *J Braz. Chem. Soc.*, **2002**, 13, 742.
[12] T. Kondo, T. Nekado, I. Susimoto, K. Ochi, S. Takai, A. Kinoshita, Y. Tajima, S. Yamamoto, K. Kawabata, H. Nakai, M. Toda, *Bioorg. Med. Chem.*, **2007**, 15, 2631-2650.
[13] VLifeMDS 3.5, Molecular Design Suite, Vlife Sciences Technologies Pvt. Ltd., Pune, India (**2004**), www.vlifesciences.com.