Journal of Reports in Pharmaceutical Sciences 2018, 7 (2), 138-155

Applications of Chemometric Methods to Elucidate Physicochemical Requirements for Binding of PTP1B Inhibitors to Its Target

Monika Chauhan*, Sarvesh K. Paliwal, SeemaKesar, Neetika, Ashima Nagpal

Department of Pharmacy, Banasthali University, 304022, Rajasthan, India

ARTICLE INFO

Article Type:
Research Article

Article History:

Received: 2018-01-09
Revised: 2018-02-15
Accepted: 2018-02-17
ePublished: 2018-03-05

***Keywords*:**
FFNN
MLR
PTP1B
PLS
QSAR
Similarity Indices

A B S T R A C T

The quantitative structure activity relationship (QSAR) models were developed using multiple linear regression (MLR), partial least square (PLS) and feed forward neural network (FFNN) for a set of 49 PTP1B inhibitors of diabetes. The MLR,PLS and FFNN generated analogous models with good prognostic ability and all the other statistical values, such as r, r^2, r^2cv and F and S values, remained satisfactory. The results obtained from this study indicate the importance of dipole moment Y component, Number of H- bond and VAMP polarization (whole molecule) in determining the inhibitory activity of PTP1B inhibitor. The best artificial neural network model is a fully-connected, feed forward back propagation network with a 2-5-1 architecture. This statistics is appropriate to the further design of novel PTP1B receptor. The similarity (CARBO and HODGKIN) analysis was also done on the same series which positively support the previous results. The QSAR study reported in the present study provide important structural situation, related to anti-diabetic activity. Present study enlightens the path of determining the potent lead compounds of PTP1B antagonist.

*Corresponding Author: **Monika Chauhan**, E-mail: monachauhan99@gmail.com

Introduction

Diabetes "a polygenic phenomenon has both genetic and environmental etiological factors".[1-5]This is a chronic disease related with ten-year-shorter life expectancy or a group of numerous diseases producing the phenotype of hyperglycemia. "Diabetes currently affects more than 371 million people worldwide and is expected to affect 552 million by 2030. In U.S., a new case of diabetes is diagnosed every 30 seconds; more than 1.9 million people are diagnosed each year''. In Type I diabetes, there is an absolute lack of insulin due to breakdown of islet cells in the pancreas. Type II diabetes mellitus and its precursor state, was also known as non-insulin–dependent diabetes mellitus (NIDDM) or adult- onset diabetes and is considered by lack in insulin signaling in spite of normal or greatly elevated levels of insulin, which makes up about 90% of cases of diabetes.

The most undoubted evidence for its use as a therapeutic target is that PTP1B knock-out mice injected with antisense oligonucleotide shows increased insulin sensitivity in case of normal and diabetic mice through improved IR signaling in peripheral tissues. There have been increasing interests in discovering and developing novel and potent molecules that can inhibit PTP1B. The choice of anti-diabetic study is massive from which the emergent small molecular objectives have been pursued for the treatment of diabetes.[6] Pharmaceutical industries have invested heavily in arena of kinase research with over huge clinical trials concerning kinase inhibitors ongoing. It has been expected that half of today's drug discovery is focused on the identification of novel kinase inhibitor as drug candidate, but therapeutic target class phosphatase surprisingly in comparison with kinase results are still in its infancy.

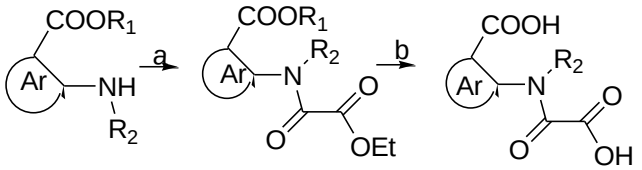
In the emerging era of technology, computational resources are becoming an absolute requirement for the development of scientific knowledge based product.[7] It has got a large number of applications, which include structure analysis, structure comparisons, lead compound design, identification of active conformations and pharmacophores, combinatorial library design, protein and binding structure, ligand binding and QSAR studies yield very good druggable compounds. The QSAR is a balanced approach for lead optimization when the structure of the target is not known. The underlying evidence of QSAR exists a relationship between the physicochemical properties of a compound and its observed biological activity. This relationship can be exploited for the design and optimization of novel ligands with desired selectivity and potency. The QSAR studies can help to accelerate the drug development during lead optimization stage and can reduce the huge time and money involved. QSAR analysis is become an economic necessity to reduce the empiricism in drug design to ensure that every molecule designed and tested should be as meaningful as possible.

Material and methods

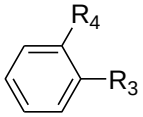
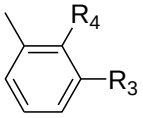
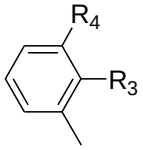
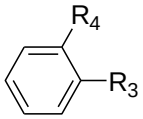
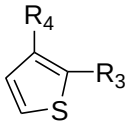
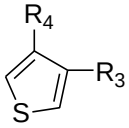
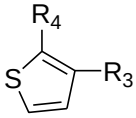
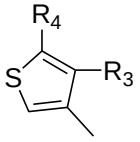
A dataset of 49 compounds of 2-(oxalylamino) benzoic acid analogues (Table 1) was selected to perform QSAR studies.[8-9]Experimentally procured IC50 values (μM) of the derivatives were converted into the negative logarithm (Log IC50).Chemical structures of respective compounds present in the series were drafted using standalone module of [32]discovery studio (version 2.0) and were finally placed in the TSAR work sheet ([32]version 3.3; Accelrys Inc., Oxford, England) five substituents[10] for each compound were identified using "define substituent" option. Two-dimensional structures were converted into three- dimension in order to understand the ligand-receptor interactions well. For each input structure the CORINA generates one low energy conformation and predicts 3D-coordinates. The information regarding stereo chemical property and the presence of rings along with chain has been considered in order to quantify the model to be generated. Nextly, "Charge 2-derive charges" option was employed to enumerate the partial charge which is a prerequisite for various structural manipulations. The 3D optimization of structures has been carried out by means of "Cosmic- optimize 3D" which comprises valence terms and non-bonded terms. It evaluates the total molecular energies by considering the summation of bond angle, bond length, torsional

angle, Van der Waals and columbic terms for suitable sets of atoms.

Table 1. Inhibitory Effect of 3a-n on PTP1B.



1a-n 2a-n 3a-n

COMPOUND	AR[C]	R_2	KI(μM)[B] PTP1B
3a[4]		H	23
3b		H	1700
3c		H	1100
3d		Me	1100
3e[14]		H	108
3f		H	75
3g[14]		H	37
3h		H	2000

Continue of Table 1. Inhibitory Effect of 3a-n on PTP1B.

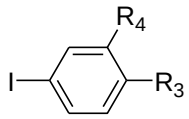
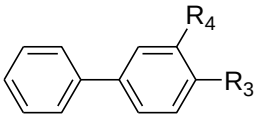
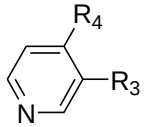
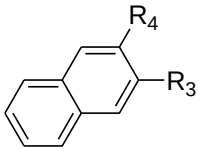
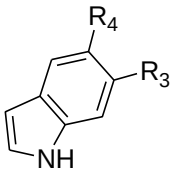
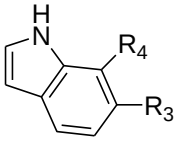
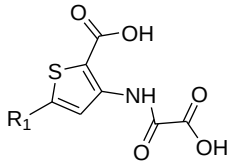
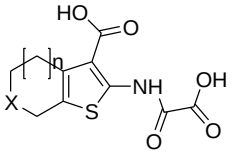
Compound	Ar[c]	R_2	Ki(μM)[b]
3i[4]		H	14
3j		H	14
3k		H	160
3i[4]		H	9.9
3m[4]		H	8
3n[4]		H	18

Table 2. Inhibitory Effect of 3g and 8a-o on PTP 1B.



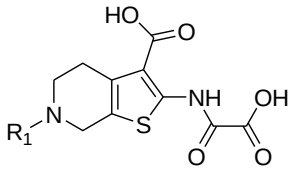
Compound	R_1	Ki(μM)[a]
		PTP1B
3g	H	37
8a	Ph	13
8b	3-thienyl	11
8c	4-(i-Bu)-Ph	13
8d	4-F-Ph	8
8e	4-Cl-Ph	10
8f	4-HO-Ph	4.5
8g	4-MeO-Ph	9
8h	4-PhO-Ph	26
8i	4-BnO-Ph	16
8j	4-($HOOCCH_2$-O)-Ph	2.5
8k	3-NO_2-Ph	4.7
8l	3-H_2N-Ph	6.1
8m	3-MeO-Ph	12
8n	3,4-MeO-Ph	25
8o	3,5-MeO-Ph	8

Table 3. Inhibitory Effect of Fused 2-Aminothiophenes 12a-k on PTP 1B.



Compound	X	n	Ki(μM)[a]
			PTP1B
12a	CH_2	0	12
12b	CH_2	1	8.1
12c	CH2	2	12
12d	O	1	14
12e	S	1	8.3
12f	SO	1	5.1
12g	SO2	1	6.1
12i	C=O	1	1.3
12j	CH-OH	1	2.0
12k		1	2.2

Table 4. Inhibitory Effect of Analogues of 12h on PTP1B.



Compound	R_1	Ki(μM) PTP1B
12h	H	0.29
22a	$-CH_3$	1.0
22b	$-CH_2$-Ph	1.0
22c	$-CH_2$-(3-MeO-Ph)	0.7
22e	$-CH_2$-(2-pyridyl)	1.0
22f	$-CH_2$-(3-pyridyl)	0.9
22g	$-CH_2$-(4-pyridyl)	1.1
22h	$-CH_2$-(2-quinolyl)	1.8
22i	$-(CH_2)_2$-Ph	0.27

Data set preparation and data reduction

The statistical robustness and accuracy of the developed QSAR model depends upon the predictivity of test set compounds by the training set driven model. So, the selection of the training set is significantly important in QSAR analysis. In the present study, data set of each series under investigation was divided randomly into training set and test set. During the selection of training set molecule it was kept in mind that the most active and the least active compound of the series remain in training set. For QSAR analysis the training set molecules were used to build linear and non-linear models so that an accurate relationship could be found between structures and biological activity. The test set molecules were not used to develop the regression model but served to check the predictive power of the developed model.49 compounds from the series remained distinguished into training set comprising of 36 compounds and test set with 13 compounds. Further training set has been used for multiple regression analysis (MLR), partial least square (PLS) and feed forward neural network (FFNN) for model improvement whereas, test set was kept to cross check the predictive power of the developed model[31]. There is a significant requirement of data reduction to eliminate the chance of correlation. It has been observed that, out of large number of descriptors only few are significantly correlated with the activity and many of the descriptors are inter-correlated which causes negative effect on interpretability of the final model. To overcome this problem, only significant descriptors are used for model development, which can be obtained by data reduction technique that reduces the number of non-relevant descriptors to obtain the most informative ones.

Techniques like pair wise correlation analysis, backward elimination and forward addition were used to reduce the large number of variables to smaller number without losing information. These techniques are particularly useful for reducing the dimension of a dataset, enabling a greater quantity of information to be visualized, studying the relationships between different descriptors and preparing data for further analysis.

TSAR methodology involves around 150 descriptors in order to generate QSAR model. Being an integrated analysis, it aims at cooperative investigation[32] of quantitative structure-activity relationships. Assumption of TSAR methodology depicts the importance of appropriate sampling of these structural descriptors in understanding their biological properties. The calculation of numerical

descriptors and employment of statistics to have a correlation has been automatically enumerates by TSAR software.

The correlation between two of the descriptors may lead to false prediction of the QSAR model due to statistical instability. Also, it results in over prediction with difficult mechanistic interpretation.[11]The nature of descriptors used and extent to which they encode the structural features related to the biological activity is a crucial part of QSAR study. Molecular descriptors are terms that characterize a specific aspect of a molecule. It also contributes towards the better understanding of structural, steric, electronic and multidimensional properties responsible for activity. Mathematical methods such as, pair wise correlation analysis and backward elimination were used as descriptor segregation methods to reduce the collinearity and correlation between descriptors, as discussed earlier. The descriptor with correlation coefficient[31] of greater than 0.6 and poor correlation with biological activity was eliminated from the work sheet for reducing data and inter-correlation among two successive descriptors. Therefore, three independent descriptors, first dipole moment Y, H bond donor and VAMP polarization with good correlation with biological activity were retrieved.

Model development

The multiple linear regression method used the biological data as dependent and the calculated descriptors as independent variable to derive finest QSAR model. The generated model is used to compute the relationship between X and Y variable. The positive and negative sign of regression coefficient in the attained regression equation directs the relation of descriptor and biological activity. [31]Conventional regression coefficient (R^2), Fischer's statistic (F) and the standard deviation (s) facilitate the geometric consequence of the regression equations.[12]

Two methods internal and external validation techniques have been employed to assure the authenticity of the generated model. The model is internally validated by [31]Leave-one-out (LOO) method in which one molecule stays eradicated from the training set followed by recalculation of the model. This has been repeated with each compound until each of them is once omitted. The actual activity of each compound is then compared with its predicted activity. The actual activities of the test set compound were predicted to validate the generated model externally.

Partial least square (PLS) study works as a substitute approach to avoid risk of over fitting and enlarges the information regarding each model. Partial least square (PLS) study has been carried away over the same training set which was used to generate the model by multiple linear analysis (MLR) in order to check the stoutness and predictive capability of the model. Leave out one row (LOO) was used to validate the generated model using partial least square (PLS).

Table 5. Actual and predicted activity data obtained from multivariate analysis of the training set compounds.

Compound name	Actual activity -log MIC (µM)	Predicted activity (µM)			
		MLR	PLS	FFNN	
3A	-1.3617	-1.3434	-1.3568	-1.3467	
3D	-3.0414	-2.5439	-2.5584	-3.0535	-
3E	-2.0334	-2.0601	-2.1006	-1.8911	
3G	-1.5682	-1.3453	-1.3671	-1.614	
3H	-3.301	-1.9751	-1.9747	-2.3687	
3J	-1.1461	-1.025	-1.0095	-0.9410	
3L	-0.99564	-0.9504	-0.9398	-0.9408	
8A	-1.1139	-1.0105	-0.9941	-0.9372	
8B	-1.0414	-0.9480	-0.9302	-0.9584	
8D	-0.90309	-0.98439	0.97997	-0.9037	
8E	-1	-0.83883	-0.8324	-0.9247	
8G	-0.95424	-0.6707	-0.6525	-1.041	
8M	-1.0792	-0.6822	-0.6640	-1.0384	
12D	-1.1461	-1.084	-1.0969	-0.9059	
12E	-0.91908	-0.9589	-0.9674	-0.8257	
12F	-0.70757	-0.9612	-0.9625	-0.8817	
12K	-0.34242	-0.1477	-0.1684	-0.1971	
12H	0.5376	0.3244	0.2995	0.2878	
22B	0	0.2526	0.2615	0.0438	
22C	0.1549	0.3393	0.3455	0.0311	
22F	0.045757	0.0802	0.0823	0.1047	
22I	056864	-0.1870	-0.1869	-0.1487	
3N	-1.2553	-1.3663	-1.3495	-1.2788	
8F	-0.65321	-0.5177	-0.5050	-0.5512	
8O	-0.90309	-1.0932	-1.0713	-0.9363	
3K	-2.2041	-2.3219	-2.3207	-2.4956	
3M	-0.90309	-0.8075	-0.7959	-0.5895	
22E	0	0.3089	0.2887	0.1211	
22G	-0.041393	-0.3019	-0.2915	-0.0932	
12J	-0.30103	0.2466	0.2313	-0.0997	
12G	-0.78533	-1.2026	-1.2052	-0.9516	
8C	-1.1139	-1.668	-1.6651	-1.6443	
3G2	-1.5682	-1.9324	-1.9367	-2.3467	
3I	-1.1461	-1.7326	-1.7227	-1.6048	
12I	-0.11394	-0.5826	-0.5988	0.1886	
22A	0	-0.6317	-0.6399	-0.0687	

Table 6. Actual and predicted activity data obtained from multivariate analysis for test set compounds.

Compound name	Actual activity -log MIC (µM)	Predicted activity (µM)		
		MLR	PLS	FFNN
3F	-1.8751	-2.1742	-2.8365	-1.9581
8L	-0.7853	-0.2735	-0.9479	-0.8384
8I	-1.2041	-0.5324	-1.2644	-1.0738
8K	-0.6721	-0.1943	-1.0675	-0.7636
12B	-0.9084	-0.3352	-1.2593	-1.0112
12A	-1.0792	-0.3574	-1.2792	-1.0513
3B	-3.2304	-1.9475	-2.4251	-3.2272
8H	-1.415	-0.5116	-1.2485	-1.0416
8N	-1.3979	-0.7130	-1.4039	-1.393
3C	-3.0414	-1.3357	-2.0169	-2.847
8J	-0.3979	0.6898	-0.2908	-0.3681
12C	-1.0792	-0.3027	-1.2301	-0.9530
22H	-0.2552	0.9937	-0.0711	-0.2521

The R^2 value for both the analysis was found to be 0.793 which signifies the strong relation. The actual and predicted biological activity values of MLR, PLS and FFNN analysis for training and test set are shown in Table 5& 6 and the respective plots are shown in Fig. 1-3 respectively.

Result

Multiple linear regression (MLR) model developed with around 150 descriptors initially resulted in poor statistical values and high inter-descriptors correlation. The model had R^2 = 0.73, s = 0.54, f = 32.63, r^2_{cv} = 0.57 and r^2 = 0.79, [32]Residual sum of square=5.316, Predictive sum of square=6.342.

The statistical equation was represented as:

Original Data:

Y = -0.16149086*X1+0.5170663*X2+0.065410152*X3-5.3257103

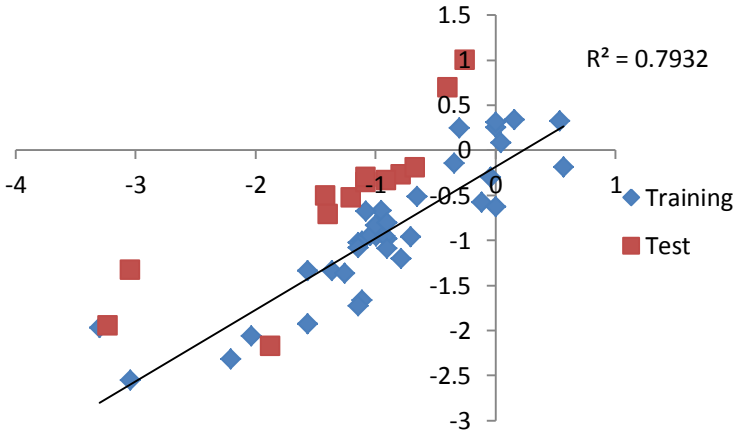


Fig. 1. Plot between actual and estimated value of training and test set of MLR.

The improvement of the statistical values is attained by data reduction. Data reduction leads to refinement of descriptors which gives an outcome of better model generation with good statistical values and low inter-correlation. The 1st condition for the model validity deals with the ratio of the number of the molecules over the number of selected descriptors. This is called the Topliss ratio. As a rule-of-thumb, it is recommended that the Topliss ratio should have a value of at least 5 (ratio of 5:1 for molecule/variable), MLR models with a maximum no. of 3 or 4 variables were selected and given model completely stands on the parameter of this ratio. The three independent descriptors, first atom dipole moment Y, H-bond and VAMP polarization were selected with high t value and good correlation with biological activity. [13-15]These has been chosen after marking stepping as zero with elimination of less significant variables. The statistical significance was polished by identifying potential 1 outliers and was deleted. These outliers were having high residual values.

Table 7. Equation Statistical values and the descriptors used for the development of MLR & PLS model.

Equation **Y= Biological Activity** **X= Descriptors**	**r**	**r^2**	**r^2_{cv}**	**s-value**	**F-value**	**Descriptors**		
						X_1	**X_2**	**X3**
MLR Y=0.16149086*X1+0.5170663*X2+0.065410152*X3-5.3257103	0.79	0.79	0.75	0.54	32.63	Dipole moment Y	H bond donor	VAMP polarization
PLS Y= -0.17298132*X1+ 0.71130329*X2+ 0.0721552*X3- 5.987565	0.79	0.79	0.75	0.50	--			

Finally, the model developed after descriptor refinement and deletion of outliers has been recognized as robust model and fulfilled statistical criteria. The statistical values obtained through this model with three descriptors are R^2 = 0.79, s = 0.40, f = 40.92, r^2_{cv} = 0.75 and r = 0.89. The remaining compounds of the dataset were used as test set for the external validation of the generated model. A correlation coefficient graph between actual activity and estimated activity shows R^2 value of 0.793. Value of R^2 clearly validates the generated model and illustrates its good predictive ability along with high statistical significance. All the t-test values, Jacknife SE and Covariance SE values (Table 8) were significant for best model that confirms the importance of each selected descriptor.

Table 8. t-test values, Jacknife SE and Covariance SE for the selected descriptors.

Descriptor	t-value	Jacknife SE	Covariance SE
Dipole Moment Y	-5.3653	0.030304	0.030099
VAMP	5.7913	0.012135	0.011294
H_2 Bond Donor	2.9249	0.16471	0.17678

Partial least square (PLS) was performed over the same model to confirm the predictive ability and improvement of the generated model.[16-19]The r^2 value of 0.793 proves the high predictive ability of the model generated by MLR and PLS.

The statistical equation is

Y = -0.17298132*X1+ 0.71130329*X2+ 0.0721552*X3- 5.987565

Where X1 = [31]Dipole Moment Y Component (whole molecule), X2 = No of H Bond (whole molecule), X3=VAMP Polarization YY (whole molecule)

Statistical significance = 1.09, Residual sum of square = 7.320, Predictive sum of squares = 8.649, r^2_{cv} = 0.75, r^2 = 0.79, s value=.50

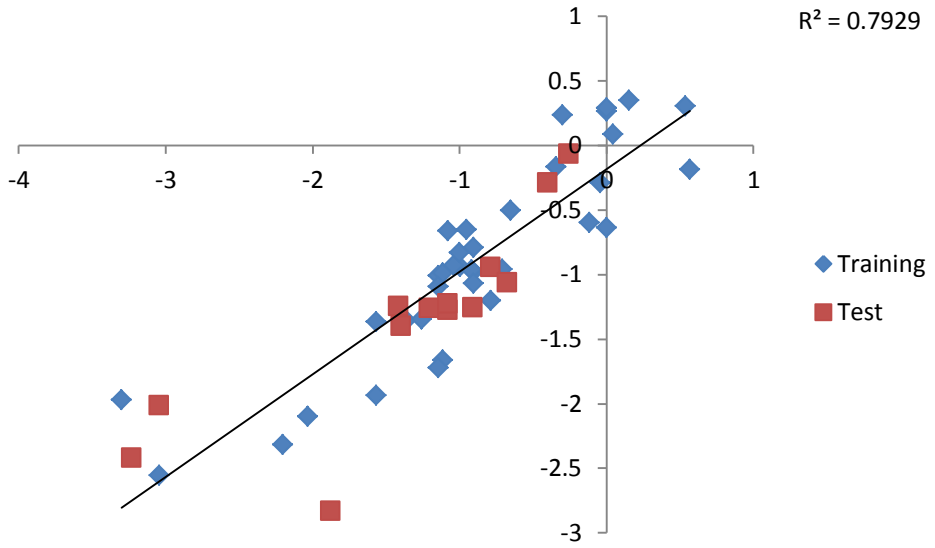


Fig. 2. Plot between actual and estimated value of training and test set of PLS.

Non Linear Regression analysis: The feed forward neural network (FFNN) study was carried out using the same data set, by Net Configuration 2-5-1, and amount of data excluded for testing was set to 25%, which was used for cross-validation to assess the performance of the trained net. The best model having closer values of test RMS fit = 0.0660 and best RMS fit = 0.0668 and r^2 = 0.881 of the training and r^2 = 0.859 of the test set was obtained. Dependency plots were drawn to analyze the influence of each independent parameter versus biological activity.

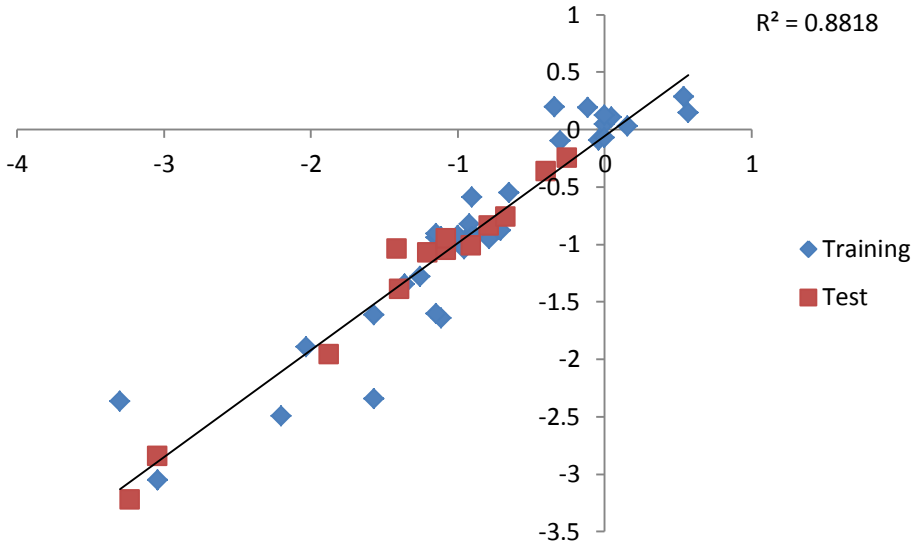


Fig. 3. Plot between actual and estimated value of training and test set of FFNN.

Table 9. Correlation matrix of three most important descriptors used to develop the model.

	Dipole Moment X_1	**No. of H_2 Bond X_2**	**VAMP Polarization X_3**
Dipole Moment X_1	1		
No. of H_2 Bond X_2	0.048271	1	
VAMP Polarization X_3	0.067538	0.28305	1
Log IC_{50} value Y	-0.43272	0.40392	0.58733

Interpretation of Descriptors

Total dipole moment is electrostatic descriptor which explains the charge distribution in a molecule. Dipole moment is a 3D electronic descriptor that specifies the strength and orientation conduct of a molecule in an electrostatic field. It is projected by utilizing partial atomic charges and atomic coordinates. The dipole moment is often considered as the direct characteristic of the polarity of a molecule.[20-22] So as according to all the three models (MLR, PLS and FFNN) total dipole moment descriptor is negatively interrelated with the biological activity, this indicates that decreasing the polarity of the molecule or lead compound by substituting such groups that decrease the polarity of the molecule as a whole will account for an increase in the biological activity which is well described T value of -6.8075

Hydrogen bond donor (whole molecule) is negatively interrelated with biological activity by decreasing the number of hydrogen bond donor, inhibitory activity can be increased. The correlation is further explained by the descriptor T value 4.4693.

[31]Vamp is a semi empirical molecular orbital package in TSAR Version 3.3 which illustrates the nuclear repulsion energy and reckons the electrostatics properties. The vamp nuclear energy explains the nuclear repulsion-driven developments in the molecule and might acquire conformational changes or atomic reactivity in the molecule. Vamp depicts negative correlation with the biological activity thereby concluding that with the decrease in electrostatic nature of the substituents there would be an impetus in biological activity of prime molecule. If there are more electron withdrawing groups there will be additional atoms to contribute to the energy terms resulting in increased nuclear energy.[23-25]The increase in nuclear energy goes with the reduction of biological activity. The correlation is further explained by the descriptor T value 6.8075.

Similarity Based analysis on 2-(oxalylamino) benzoic acid derivatives

[33]Similarity indices signify a quantitative measure of the similarity among two molecules on the basis of their size, shape, electronic distribution, lipid solubility, water solubility, or chemical reactivity .Molecular similarity was introduced as a concept by Carbo.

The most broadly used form of similarity index applied to calculation of 3D molecular similarity was proposed by Carbo:

$$R_{AB} = \frac{\int P_A P_B d\nu}{(\int P_A^2 d\nu)^{1/2} (\int P_B^2 d\nu)^{1/2}}$$

The numerator in the equation measures property overlap while denominator indicates similarity result. The difference between equations for Carbo and Petke(Hodgkin) index is only in the denominator part With respect to Carbo index, it is less sensitive to shape of the property but more sensitive to its magnitude [Hodgkin and Richards,1987;Richards and Hodgkin 1988]. It is defined as:

$$H_{st} = \frac{2 \cdot \sum_{k=1}^{N} P_{sk} \cdot P_{tk}}{\sum_{k=1}^{N} P_{sk}^2 + \sum_{k=1}^{N} P_{tk}^2} = 1 - \frac{d_{st}^2}{\sum_{k=1}^{N} P_{sk}^2 + \sum_{k=1}^{N} P_{tk}^2}$$

The molecular[33] similarity indices were calculated with the ASP similarity program in the TSAR software. Between the two approaches (a grid-based method and Gaussian approximation), the Gaussian approximation was used for calculating similarity indices because it thoroughly reflects that of the grid-based calculations and is much faster.[26] A Gaussian approximation based NxN similarity matrix was built and subjected to data reduction techniques.

Descriptor Calculation

Primarily, more than 300 descriptor were calculated for series for (whole molecule and substituents) by Carbo index (NxN) Similarity matrix. The descriptors with the same values for all the compounds were discarded. The correlation matrix was generated to study the data patterns and to reduce data redundancy by Pair wise correlation.

Stastical value obtained for MLR analysis

r^2 = 0.72, s = 0.51, f = 39.92, r^2_{cv} = 0.64 and r = 0.84, Residual sum of square=8.08, Predictive Sum of square=10.13

Finally two parameters, highly correlated with biological activity were used to generate regression equation and evaluated for their relative impact on the activity of the compound in Table 10.

Where X1= Combined Similarity vs. Molecule and X2= Charge Similarity vs Molecule

Original Data: Y = -5.0566635*X1 - 1.1768534*X2 + 0.5929991

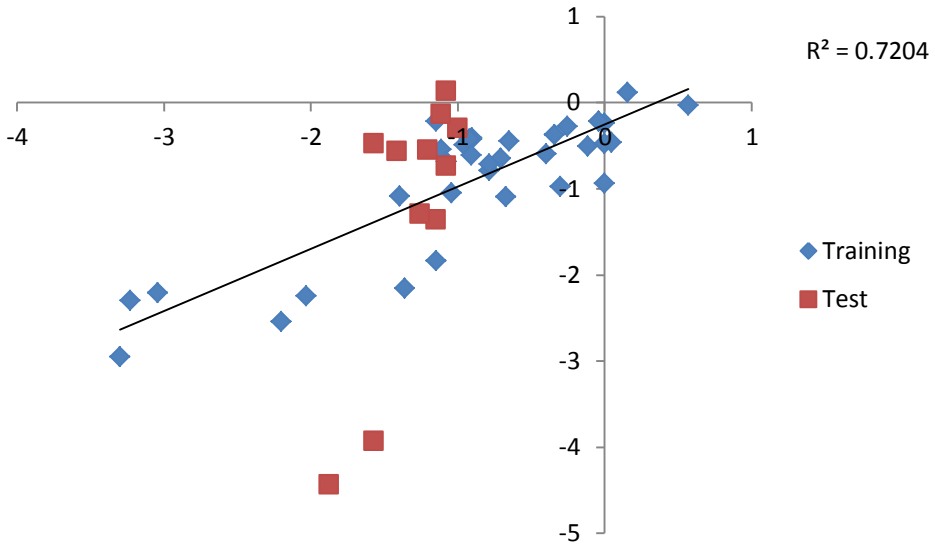


Fig. 4. Plot between actual and estimated value of training and test set of MLR.

The PLS study was also performed using the same data set and the resulted r^2cv value and statistical signifies the high predictive ability of the developed PLS model (Equation 2), mentioned in Fig 5.

Y=-4.5275674*X1-1.4307652*X2+0.45086622

r^2 = 0.78, r^2_{cv} = 0.65 and r = 0.84, Residual sum of square=7.222, Predictive Sum of square=11.496

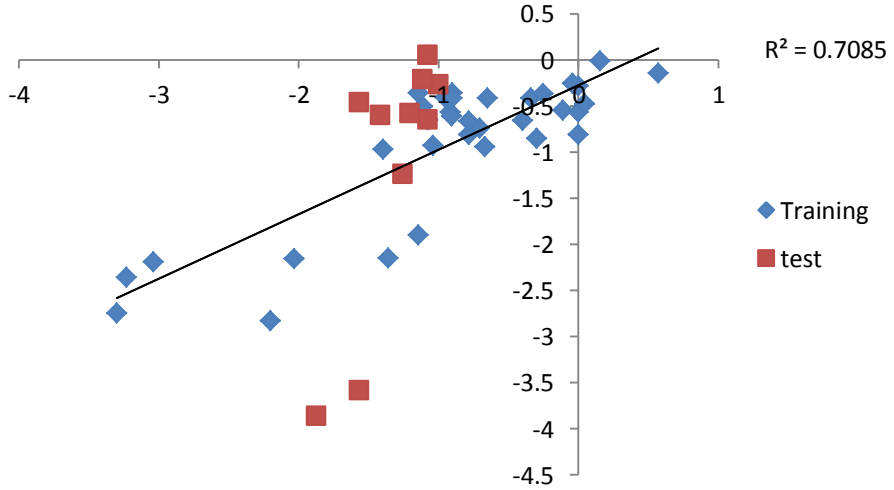


Fig. 5. Plot between actual and estimated value of training and test set of PLS

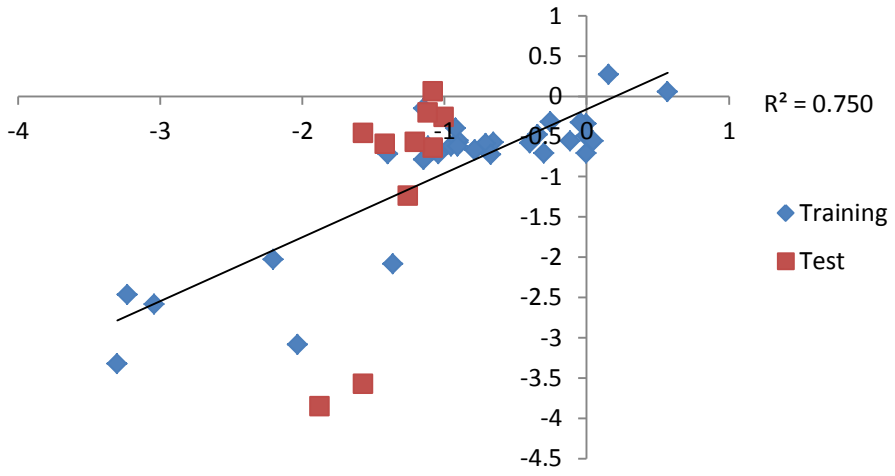


Fig. 6. Plot between actual and estimated value of training and test set of FFNN.

The MLR, PLS and FFNN graphs plotted between the actual and predicted activities of training set as well as test set of compound,[27-28] as shown Fig. 4-6. The correlation of biological activity of compounds is given in Table 11.

Non Linear Regression analysis: The feed forward neural network (FFNN) study was performed using the similar data set, by net configuration 2-5-1, and amount of data excluded for testing was set to 30%, which was used for cross-validation to assess the performance of the trained net. The best model having closer values of test RMS fit = 0.1456 and best RMS fit = 0.033 and r^2 = 0.864 of the training and r^2 = 0.734 of the test set was obtained.

Table10. Correlattion matrix showing descriptor entered model

	Combined Similarity vs. Molecule (X_1)	**Charge Similarity vs Molecule (X_2)**	**EC_{50}**
Combined Similarity vs. Molecule (X_1)	1	0.3541	-0.55365
Charge Similarity vs Molecule (X_2)	0.3541	1	-0.50657
EC_{50}	-0.55365	-0.50657	1

Table 11. Correlation of biological activity of active and inactive molecules of training set with all three descriptors.

	Name of Compound	Biological activity – log MIC (µM)	Dipole moment Y component (whole molecule)	VAMP Polarisation YY (whole molecule)	Hydrogen Bond donor (whole molecule)
Active Compound	22 i	0.56864	1.17949	41.786	3
Inactive Compound	3H	-3.301	0.17402	26.418	3

(Fig. 7) shows the alignment of most active compound (22 i) within the active site of PTP1B receptor.[29-30]The compound showed hydrogen bond interaction with LYS 1116, PHE 1182 and Van der Waal interaction with ARG 1112.

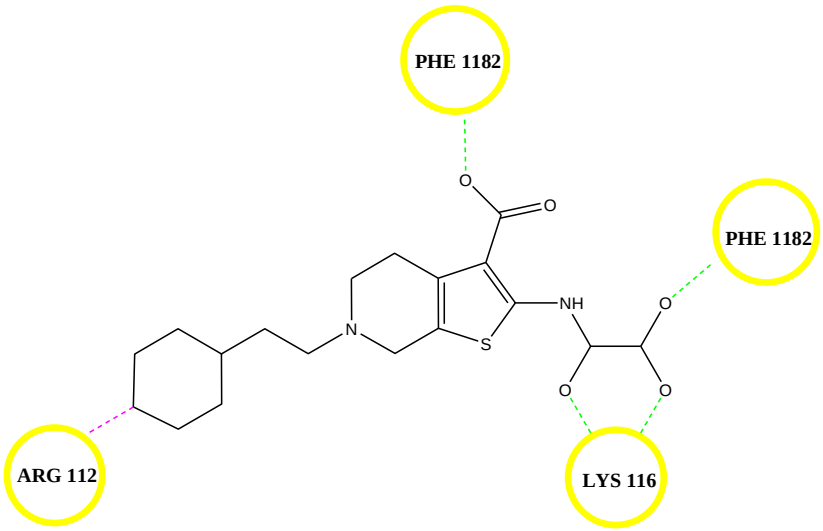


Fig. 7. Binding mode of most active compound into the crystal structure of PTP1B. Green dotted line represents hydrogen bond interaction and pink dotted line represents Van der Waal interaction.

Conclusion

All the results were discussed above, indicate that the using MLR, PLS and FFNN analysis with molecular descriptors[34-35] belongs to 2-(oxalylamino) benzoic acid derivatives and generated highly robust QSAR models. Molecular parameters such as lipophilic, shape, and electronic design of 2-(oxalylamino) benzoic acid are considered to be important contributors to their biological properties and it can also be used for the designing of further new anti-diabetic compounds with more potency and reduced mechanism based side effect of traditional anti-diabetic agents. The results obtained from same series in similarity analysis (charge, combined,) also support the MLR, PLS and FFNN results with correlation coffecient greater than 7. This information about the 2D-requirement of the compound showed the robustness of the model i.e. of

great value for the novel design of new PTP1B derivatives of pharmaceutical importance.

Acknowledgement

Computational resources were provided by Banasthali University, and the authors thank Vice Chancellor for providing all the necessary facilities.

Conflict of Interests

Authors certify that there is no actual or potential conflict of interest in relation to this article.

References

[1] Korner A, Kiess W, Stumvoll M, Kovacs P. Polygenic contribution to obesity Genome-wide strategies reveal new targets: Front Horm Res. 2008:36:12–36.

[2] Guillausseau P J, Tielmans D, Virally-Monod M, Assayag M. Diabetes: From phenotypes to genotypes: Diabetes Metab.1997:23:2:14–21.

[3] Romao I, Roth J. Genetic and environmental interactions in obesity and type 2 diabetes: J Am Diet Assoc. 2008:108:S24–S28.

[4] Florez J C. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes Where are the insulin resistance genes?: Diabetologia. 2008:51:1100–1110.

[5] Alonso A, Sasin J, Bottini N, FriedbergI,Osterman A, Godzik A, Hunter T, Dixon J, T Mustelin. Protein tyrosine phosphatases in the human genome: Cell. 2004:117:699–71.

[6] Kenner K A, Anyanwu E, Olefsky J M, Kusari J. Protein- tyrosine phosphatase 1B is a negative regulator of insulin- and insulin-like growth factor-I-stimulated signaling: J. Biol. Chem. 1996; 19810–19816.

[7] Ahmad F, Li P M, Meyerovitch J, Goldstein BJ. Osmotic loading of neutralizing antibodies demonstrates a role for protein-tyrosine phosphatase 1 Bin negative regulation of the insulin action pathway: J. Biol. Chem. 1995:270: 20503–20508.

[8] Elchebly M, Payette P, Michaliszyn E, Cromlish W, Collins S, Loy A L, Normandin D, Cheng A, Himms-Hagen J, Chan C, Ramachandran C, Gresser M J, Tremblay M L, Kennedy B P. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene: Science.1999:283:1544–1548.

[9] You-Ten KE, Muise ES, Itie A, Michaliszyn E, Wagner J, Jothy S, Lapp W S, Tremblay M L. Impaired bone marrow microenvironment and immune function in T cell protein tyrosine phosphatase-deficient mice: J. Exp. Med. 1997:186:683–693.

[10] Traurig M, Hanson RL, Kobes S, Bogardus C, Baier L J. Protein tyrosine phosphatase 1B is not a major susceptibility gene for type 2 diabetes mellitus or obesity among Pima Indians: Diabetologia. 2007:50:985–989.

[11] Fischer EH, Charbonneau H, Tonks NK. Protein tyrosine phosphatases: A diverse family of intracellular and transmem- brane enzymes. Science. 1991:13:1497-1506.

[12] Hunter T. Signaling - 2000 and beyond: Cell. 2000:100:113-127.

[13] Andersen HS, Iversen LF, Jeppesen CB, Branner S, Norris K, Rasmussen HB, Møller KB, Møller NP H. 2-(Oxalyl-amino)-benzoic acid is a general, competitive inhibitor of protein-tyrosine phosphatases: J. Biol. Chem. 2000:275:7101-7108.

[14] McCormack JC, Iversen LF, Andersen HS, Møller NPH. Protein-tyrosine phosphatases (PTPs) as drug targets: Inhibitors of PTP-1B for the treatment of diabetes. Curr. Opin. Drug Discovery Dev. 2000:3:527-540.

[15] Burke TR, Jr, Yao ZJ, Smyth MS, Ye B. Phosphotyrosylbased motifs in the structure-based design of protein-tyrosine kinase-dependent signal transduction inhibitors. Curr. Pharm. Des. 1997:3:291-304.

[16] Burke TR, Zhang YZ. Protein-tyrosine phosphatases: structure, mechanism, and inhibitor discovery. Biopolymers. 1998:47:225.

[17] Kole KH, Ye B, Akamatsu B, Yan X, Barford D, Roller P P, Burke TR. Tripeptide inhibitors of Yersinia protein-tyrosine phosphatase. J. Biochem. Biophys. Res. Com- mun. 1995:209:817.

[18] Ye B, Akamatsu M, Shoelson SE, Wolf G, Giorgetti- Peraldi S, Yan X J, Roller P P, Burke T R. Potential role of JAK kinases. J. Med. Chem. 1995; 38:4270.

[19] Burke TR, Ye B, Akamatsu M, Ford H, Yan XJ, Kole H K, Wolf G, Shoelson S E, Roller P P. Potent inhibition of protein-tyrosine phosphatese-1B using the phosphotyrosyl mimetic fluoro-O-malonyl tyrosine. J. Med. Chem. 1996: 39: 1021.

[20] Yao ZJ, King C R, Cao T, Kelley J, Milne G W A, Voigt J H, Burke T R .Amino acid Peptide and Proteins. J. Med. Chem. 1999:42:25.

[21] Burke TR, Yao ZJ, Zhao H, Milne GWA, Wu L, Zhang Z Y, Voigt J H. Enantio selective synthesis of non-phosphorus-containing phosphotyrosyl mimetics and their use in the preparation of tyrosine phosphatase inhibitory peptides. Tetrahedron. 1998:54:9981.

[22] Burke TR, Luo J, Yao ZJ, Gao Y, Milne GW A, Guo R, Voigt J H, King C R, Yang D. Development of Grb2 SH2 Domain Signaling Antagonists: A Potential New Class of Antiproliferative Agents. Bioorg. Med.Chem. Lett. 1999:9:347.

[23] Williams R M. Asymmetric synthesis of mono substituted and .alpha.,.alpha.-disubstituted .alpha.-amino acids via diastereoselective glycine enolate alkylations. J. Am. Chem. Soc. 1991:113: 9276.

[24] Sarmiento M, Wu L, Y Keng Y, Song L, Luo Z, Huang Z, Wu G Z,Yuan A K, Zhang Z Y. J. Med. Chem. 2000:43:146.

[25] Wrobel J, Li Z, Sredy J, Sawicki D R, L Seestaller, D Sullivan. Synthesis and PTP1B inhibition of novel 4-aryl-1-oxa-9-thiacyclopenta fluorenes. Bioorg. Med. Chem. Lett. 2000:10:1535.

[26] Chen YT, M B Onaran, CJ Doss, CT Seto. alpha-Ketocarboxylic acid-based inhibitors of protein tyrosine phosphatases. Bioorg. Med. Chem. Lett. 2001:11:1935.

[27] Malamas MS, Sredy J, Moxham C, Katz A, Xu WX, McDevitt R, Adebayo FO , Sawicki DR, Seestaller L, Sullivan D ,Taylor JR. Novel benzofuran and benzothiophene biphenyls as inhibitors of protein tyrosine phosphatase 1B with antihyperglycemic properties. J. Med. Chem. 2000: 43:1293.

[28] Cao X, Moran EJ, Siev D, Lio A, Ohashi C, Mjalli A M. One recent example is the Symposium-in-Print entitled "Molecular Diversity, Creation and Application. Bioorg. Med. Chem. Lett. 1995:5:2953.

[29] Huyer G, Kelly J, Moffat J, Zamboni R, Jia Z, Gresser MJ, Ramachandran C. Affinity selection from peptide libraries to determine substrate specificity of protein tyrosine phosphatases. Anal. Biochem. 1998:258:19.

[30] Wang P, Fu H, Snavley DF, Freitas MA, Pei D. Identification of optimal substrates of protein tyrosine phosphatase SHP-1.Biochemistry. 2002:41:6202.

[31] Paliwal SK. Quantitative structure activity relationship (QSAR) of N 6-substituted adenosine receptor agonists as potential antihypertensive agents: Medicinal Chemistry Research: 2010: 11:1643-1649.

[32] Paliwal SK. Quantitative structure activity relationship analysis of angiotensin II AT1 receptor antagonists, Medicinal Chemistry Research. 2009:5:475-489.

[33] Paliwal SK. Structure–activity relationship analysis of cationic 2-phenylbenzofurans as potent antitrypanosomal agents: a multivariate statistical approach, MonatsheftefürChemie–Chemical Monthly. 2011:1:85-104.

[34] Shahlaei M. Computational evaluation of some indenopyrazole derivatives as anticancer compounds: application of QSAR and docking methodologies, J.Enzyme Inhib.Med.Chem. 2011 oct; 281:16-32.

[35] Shahlaei M. QSAR study of some CCR5 antagonists as anti-HIV agents using radial basis function neural network and general regression neural network on the basis of principal components, Med Chem Res. 2012:21:10:3246–3262.

