Trop J Nat Prod Res, December 2025; 9(12): 5935 - 5944

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

Tropical Journal of Natural Product Research

Available online at https://www.tjnpr.org

Original Research Article

Development of 2D-QSAR Models and Molecular Docking of Protein Tyrosine Phosphatase 1B Inhibitors

Ismail Lamrani[1]*, Fathallaah Bazi[1], Mounir Ghamali[2], Fatiha Amegrissi[1], Bahija Mounir[1]

[1]*Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, Casablanca, Morocco*
[2]*Molecular Chemistry and Natural Substances Laboratory, Faculty of Sciences, University Moulay Ismail, Meknes, Morocco*

ARTICLE INFO

Article history:
Received 11 June 2025
Revised 02 November 2025
Accepted 13 November 2025
Published online 01 January 2026

ABSTRACT

Quantitative structure-activity relationships (QSARs) are widely used in drug discovery and design to quantitatively analyze the relationships between the structures and biological activities of compounds. The present study aimed to develop a two-dimensional (2D) QSAR and molecular docking studies to predict the biological activities of protein tyrosine phosphatase 1B (PTP1B) inhibitors. A QSAR was carried out to study a series of fifty-three (53) compounds based on protein tyrosine phosphatase 1B (PTP1B) inhibitors. The study was performed using principal components analysis (PCA), multiple linear regression (MLR) and multiple non-linear regression (MNLR) to predict unambiguous QSAR models of studied compounds toward PTP1B inhibitory activity. Molecular docking was used to elucidate the inhibitory mechanisms of the most active compound from the data set against PTP1B. The statistical results of the MLR and MNLR indicate that the determination coefficients (R^2) were similar (R^2 = 0.796). To validate the predictive power of the resulting models, the determination coefficients of external validation were 0.815 and 0.639 for the MLR and MNLR, respectively. These results showed that both models possess favorable estimation stability and good prediction power. The docking of the most active compound **46** showed many hydrogen bond formations with the active site residues ASP (A:48) and TYR (A:46) of PTP1B. The study successfully developed a simple, convenient quantitative structure-activity relationship (QSAR) model that can be used to screen chemical databases or design new PTP1B inhibitor-derived compounds.

Keywords: Diabetes, Protein tyrosine phosphatase 1B, Molecular docking, 2D-QSAR

Introduction

The World Health Organization recognizes type 2 diabetes mellitus as the only non-infectious disease that is considered an epidemic due to its global prevalence, particularly in countries with a Western lifestyle.[1] It is characterized by chronically elevated blood glucose levels. The increasing prevalence of type 2 diabetes mellitus and obesity in the general population has intensified the search for new therapeutic options.[2] Both type 2 diabetes and obesity are characterized by resistance to the hormone insulin in the muscles, liver, and central nervous system.[2,3] Therefore, drugs that can ameliorate this resistance should be effective in treating these diseases.[4,5] Tyrosine phosphorylation of specific intracellular proteins is a critical process that enables the biological effects of various polypeptide hormones and growth factors to be transduced and coordinated *in vivo*. This process is regulated by two types of enzymes: protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs).[6]

*Corresponding author. Email:
ismail.lamrani1-etu@etu.univh2c.ma
Tel: +212672798106

Citation: Lamrani I, Bazi F, Ghamali M, Amegrissi F, Mounir B. Development of 2D-QSAR Models and Molecular Docking of Protein Tyrosine Phosphatase 1B Inhibitors. Trop J Nat Prod Res. 2025; 9(12): 5935 – 5944 https://doi.org/10.26538/tjnpr/v9i12.5

Official Journal of Natural Product Research Group, Faculty of Pharmacy, University of Benin, Benin City, Nigeria

Recent insights into the mechanism of insulin action have demonstrated that reversible tyrosine phosphorylation of the insulin receptor and its cellular substrate proteins play a central role in the mechanism of insulin action.[7] Preliminary biochemical and cellular studies have indicated that protein tyrosine phosphatases (PTPs) play a significant role in regulating the process of insulin signal transduction.[8] Protein tyrosine phosphatase 1B (PTP1B), a cytosolic PTP, plays a pivotal role in the regulation of insulin sensitivity and the dephosphorylation of the insulin receptor. PTP1B has been implicated as a negative regulator of insulin receptor signaling.[9,10] Clinical studies have identified a correlation between states of insulin resistance and levels of PTP1B expression in muscle and adipose tissues. This finding suggests that PTP1B plays a significant role in the insulin resistance associated with obesity.[11] These results establish a direct role for PTP1B in down regulating the insulin functions. Consequently, potent, orally active, and selective PTP1B inhibitors have emerged as promising pharmacological agents for the treatment of obesity and diabetes. Quantitative structure-activity relationship (QSAR) has been widely used in drug discovery and drug design by medicinal chemists,[12,13] and in various practical applications[14,15] to provide quantitative analysis of structure and biological activity relationships of compounds. A multitude of QSAR studies have been documented, aiming to ascertain the pivotal structural characteristics that underpin biological activity and to formulate predictive models for a myriad of chemicals, as posited by various authors.[16,17] Consequently, the development of a QSAR model for the prediction of activity prior to the synthesis of new Protein Tyrosine Phosphatase 1B inhibitors becomes imperative. The development of a successful QSAR model facilitates a comprehensive understanding of the relationships between physicochemical properties and biological activity within a

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

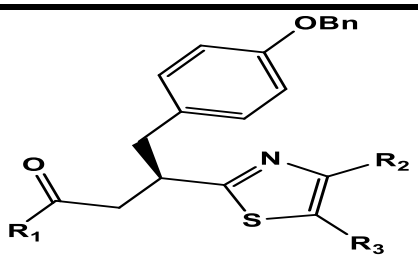
given class of molecules. In addition, it provides researchers with a thorough analysis of the lead molecules to be utilized in subsequent studies.[18] Among the various PTP1B inhibitors that have been documented in the extant literature, mono, bi, and tricyclic thiophene derivatives emerge as compelling small-molecule targets for drug design, a consequence of their synthetic accessibility and high potency.[19-21] In the present study, 2D-QSAR studies were performed on a series of 53 PTP1B derivatives, followed by molecular docking simulation, with the objective of identifying the key structural features required to design new potent lead candidates of this class. The findings from this study may contribute to the development of highly potent anticancer drugs.

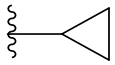
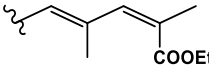
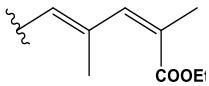
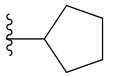
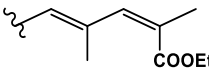
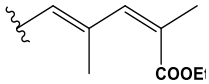
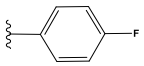
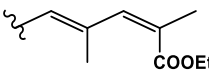
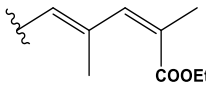
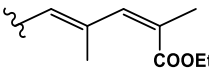
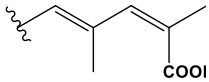
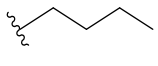
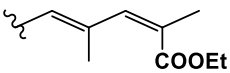
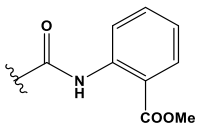
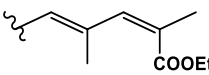
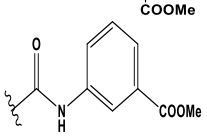
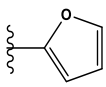
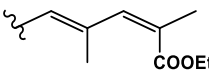
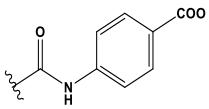
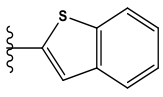
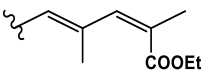
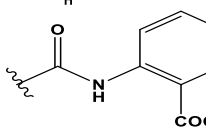
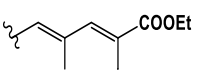
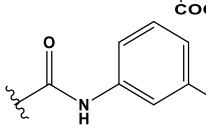
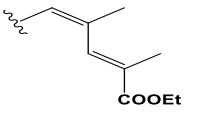
Materials and Methods

Data sources

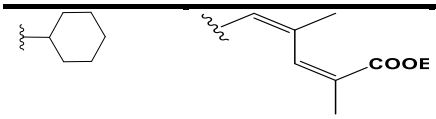
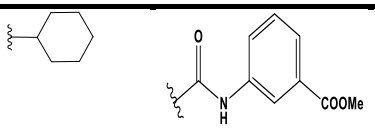
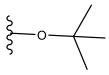
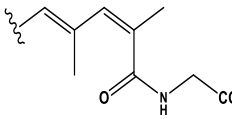
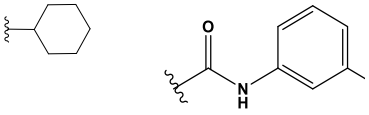
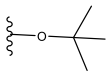
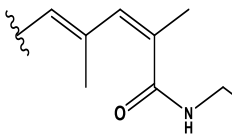
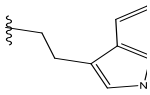
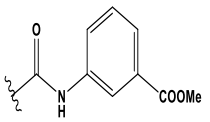
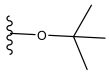
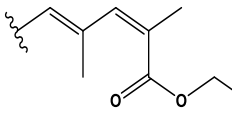
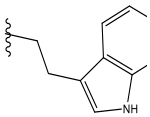
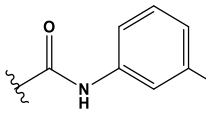
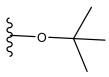
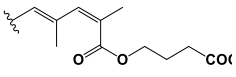
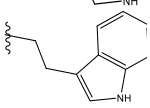
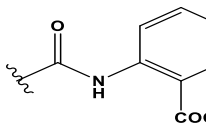
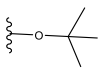
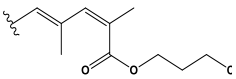
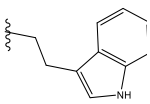
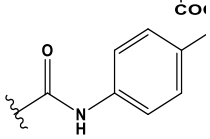
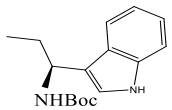
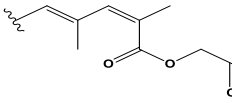
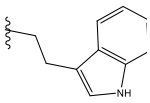
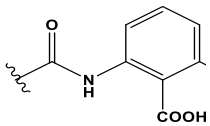
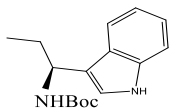
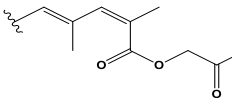
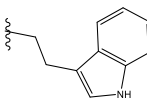
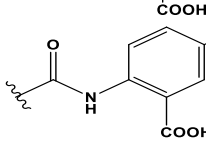
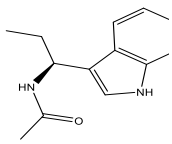
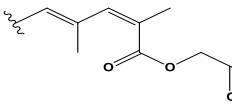
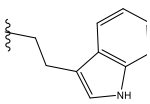
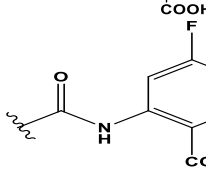
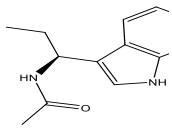
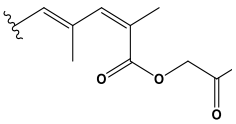
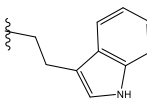
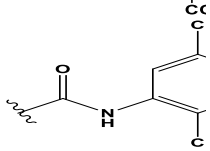
In the present study, fifty-three (53) substituted PTP1B molecules for which activities have been reported in the literature[22] were selected. The activities were collected as pIC_{50} values, where IC_{50} refers to the molar concentration of the compound required for 50% inhibition of PTP1B activity. Table 1 presents the substituted structures of the studied compounds and the corresponding experimental activities for pIC_{50}. To ensure the validity of the data set for the QSAR model, the 53 substituted PTP1B were divided into training and test sets. A total of 42 molecules were utilized as the training set for the construction of QSAR models, while a subset of 11 molecules were designated as the test set. The division was executed through the utilization of the K-means method.

Table 1: Structures and inhibitory activity of PTP B1 compounds

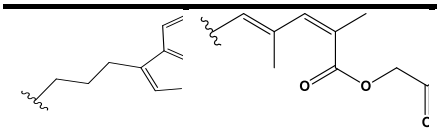
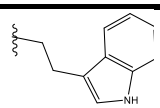
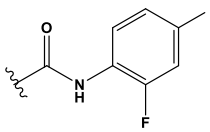
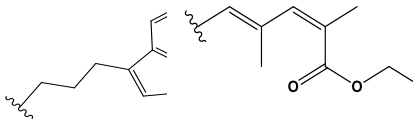
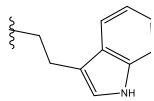
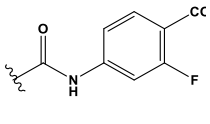
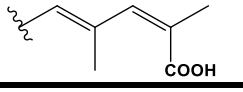


R_1	R_2	R_3	pIC50	No	R_1	R_2	R_3	pIC50
		H	4.70	28			CH_3	4.87
		H	4.74	29			$CH_2CH(CH_3)_2$	5.02
		H	5.04	30			Ph	5.07
		H	4.79	31			Ph	4.96
		H	4.94	32			H	5.05
		H	5.03	33			H	4.97
		H	4.74	34			H	5.35
		H	5.02	35			H	5.09
		H	4.72	36			H	5.22
		H	4.62	37			H	5.31

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

		H	4.88	38			Ph	5.42
		H	4.90	39			Ph	5.93
		H	4.79	40			Ph	5.57
		H	5.14	41			Ph	5.99
		H	4.90	42			Ph	5.25
		H	5.01	43			Ph	5.64
		H	5.25	44			Ph	5.37
		H	5.79	45			Ph	5.54
		H	5.68	46			Ph	6.26
		H	5.52	47			Ph	6.14
		H	5.22	48			Ph	5.92
		H	5.43	49			Ph	5.79
		H	5.26	50			Ph	5.84
		H	5.63	51			Ph	5.95

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

		H	5.51	5 2				Ph	6.11
		H	5.61	5 3				Ph	5.59
		H	4.83	-	6666	-666		-	-

Molecular descriptors

Presently, a multitude of molecular descriptors were employed in the QSAR studies. After validation, the findings were used to predict the activity of untested compounds. The ACD/ChemSketch and Chem 3D programs[23,24] were used in this study to calculate all molecular descriptors.

Statistical analysis

A quantitative structure-activity relationship (QSAR) study is a statistical approach used to create empirical models relating the biological activity of compounds to their chemical structures. This QSAR study use sixteen (16) quantitative descriptors to explain chemical structures and describe their relationship to biological activity. These 16 descriptors were calculated for the 53 molecules using the ACD/ChemSketch and Chem 3D programs to explain the structure-activity relationship. The substituted PTPB1's quantitative descriptors were studied using statistical methods based on principal component analysis (PCA)[25] with the software XLSTAT version 2014.[26] The PCA is a useful statistical technique to obtain the maximal amount of encoded information in the compound structures and understand the distribution of the compounds.[27] This method is essentially a descriptive statistical technique used to graphically present as much information as possible from the data. Multiple linear regression (MLR) analysis with stepwise variable selection is used to model structure-activity relationships. This mathematical technique minimizes the difference between the actual and predicted values. It is also used to select descriptors as input variables in multiple nonlinear regression (MNLR).
The MLR and MNLR models were generated using the XLSTAT software version 2014. To justify the equations for predicting pIC_{50}, we used the determination coefficient (R^2), mean squared error (MSE), Fisher's criterion (F), and significance level (P).

Validation

The main objective of a QSAR study is to develop a model with the greatest predictive and generalization capabilities. To evaluate the predictive ability of the developed QSAR models, two principles were applied: internal and external validation. For internal validation, leave-one-out cross-validation (R^2_{cv}) was used to evaluate the stability of the models. A high R^2_{cv} value indicates high internal predictive power and robustness of a QSAR model. However, Globarikh's study[28] revealed that there is no correlation between the R²cv value for the training set and the predictive ability for the test set. This finding suggests that R^2_{cv} is insufficient for reliably estimating the predictive power of models for new compounds. Therefore, external validation is the only way to determine the generalizability and true predictive ability of QSAR models for new chemicals. For this reason, statistical external validation was applied to the models, as described by Globarikh and Tropsha using a test set.[28-30]

Molecular docking

Molecular docking is one of the most important methods for discovering novel small-molecule drugs.[31-33] This study used Surflex-Dock, which is implemented in SYBYL-X.2.0, to perform this technique. The ligand and protein preparation steps for the docking protocol were performed in SYBYL-X.2.0 with Gasteiger-Hückel atomic partial charges,[34] and then results were analyzed using Discovery Studio 2016 software.[35]
The crystal structure of PTP1B was downloaded from the Protein Data Bank (PDB entry code 1YNN). After removing the original ligand, cofactors, and water molecules, the most active compound from the dataset and the best designed hit were docked into the active site.

Results and Discussion

Data set for analysis

A QSAR study was carried out for the first time on fifty-three PTP1B compounds, in order to establish quantitative relationships between their structures and their inhibitory activities. The values of the calculated descriptors are shown in Table 2.

Table 2: Values for the training and test of PTP1B compounds

N	pIC_{50}	MW	MR	VM	Pc	IR	ST	D	1.4VDW	E_T	log *P*	HBA	HBD	IB	IW	E_{LUMO}	E_{HOMO}
1	4.70	544.70	157.87	456.00	1 211.80	1.60	49.80	1.19	31.60	31.73	6.86	4	1	3931624	6078	-4.98	-8.19
2	4.74	558.73	162.31	471.50	1 262.80	1.60	51.40	1.18	31.73	48.62	7.52	4	1	4351900	6407	-4.96	-8.48
3	5.04	584.70	166.32	476.70	1 262.30	1.61	49.10	1.22	35.99	33.94	8.02	5	1	5412328	7253	-4.96	-8.36
4	4.79	532.69	155.25	460.10	1 196.30	1.58	45.60	1.15	29.95	36.00	7.18	4	1	4263919	5682	-4.96	-8.22
5	4.94	546.72	159.92	476.20	1 238.70	1.58	45.70	1.14	30.61	32.61	7.45	4	1	4825920	6115	-4.97	-8.24
6	5.03	572.75	166.92	489.30	1 302.90	1.59	50.20	1.17	34.78	36.00	7.93	4	1	4851876	6811	-4.97	-8.29
7	4.74	556.67	158.63	455.30	1 209.20	1.61	49.70	1.22	28.89	55.89	6.47	5	1	4351900	6407	-4.95	-8.46
8	5.02	622.79	182.55	496.10	1 342.20	1.65	53.50	1.25	36.41	44.79	8.89	4	1	5635576	8109	-4.97	-7.94
9	4.72	572.75	166.92	489.30	1 302.90	1.59	50.20	1.17	34.83	31.87	7.93	4	1	4851876	6811	-4.93	-8.33

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

10	4.62	572.75	166.92	489.30	1 302.90	1.59	50.20	1.17	34.19	30.48	7.93	4	1	4851876	6811	-5.08	-8.35
11	4.88	572.75	166.92	489.30	1 302.90	1.59	50.20	1.17	34.32	37.21	7.93	4	1	4851876	6811	-5.10	-8.24
12	4.90	619.77	174.52	521.60	1 366.00	1.58	47.00	1.18	30.04	30.73	6.27	5	2	8622394	8632	-4.10	-8.07
13	4.79	605.74	169.89	505.10	1 326.20	1.58	47.50	1.19	29.15	30.83	5.93	5	2	7670110	8031	-4.10	-8.08
14	5.14	620.75	172.67	519.10	1 357.70	1.57	46.70	1.19	35.62	34.95	6.95	5	1	8622394	8632	-4.95	-8.32
15	4.90	648.80	181.94	552.20	1 437.30	1.57	45.90	1.17	37.25	35.97	7.52	5	1	10933789	10039	-4.92	-8.27
16	5.01	662.83	186.57	568.70	1 477.10	1.56	45.50	1.16	38.42	36.92	7.94	5	1	12278406	10811	-4.89	-8.37
17	5.25	806.96	226.96	648.50	1 744.40	1.61	52.30	1.24	39.91	41.51	0.00	7	3	19920969	16831	-4.96	-8.26
18	5.79	778.91	217.49	606.60	1 661.90	1.63	56.30	1.28	36.54	37.59	0.00	7	4	16679443	15078	-4.96	-8.37
19	5.68	748.88	211.35	592.60	1 610.20	1.63	54.50	1.26	35.64	43.28	5.89	7	3	14591298	14156	-4.97	-8.24
20	5.52	720.83	201.88	550.70	1 527.70	1.65	59.20	1.30	32.43	38.78	5.28	7	4	12059943	12581	-4.94	-8.38
21	5.22	677.80	193.83	536.40	1 459.20	1.64	54.70	1.26	35.26	49.10	6.87	6	2	9614159	11252	-4.98	-8.23
22	5.43	649.75	184.35	494.50	1 376.70	1.66	60.00	1.31	31.98	44.59	6.27	6	3	7799399	9888	-4.94	-8.38
23	5.26	691.83	198.46	552.90	1 499.00	1.63	54.00	1.25	36.69	50.71	7.29	6	2	10686235	12029	-4.96	-8.38
24	5.63	663.78	188.98	511.00	1 416.50	1.66	59.00	1.29	33.45	43.41	6.69	6	3	8716657	10612	-4.94	-8.38
25	5.51	705.86	203.09	569.40	1 538.80	1.63	53.30	1.23	37.45	55.52	7.71	6	2	11865022	12856	-4.96	-8.38
26	5.61	677.80	193.62	527.50	1 456.30	1.65	58.00	1.28	33.96	48.30	7.10	6	3	9729147	11384	-4.94	-8.37
27	4.83	544.70	157.45	447.40	1 219.40	1.62	55.10	1.21	29.68	38.64	7.33	4	2	3797992	5871	-4.95	-8.30
28	4.87	586.78	171.75	505.50	1 341.20	1.59	49.50	1.16	35.46	39.42	8.27	4	1	5312513	7119	-4.89	-8.01
29	5.02	628.86	185.70	555.50	1 459.40	1.58	47.60	1.13	37.66	41.55	9.50	4	1	7042641	8257	-4.88	-7.97
30	5.07	648.85	191.51	554.50	1 475.00	1.60	50.00	1.16	40.97	59.61	9.66	4	1	7142248	9055	-5.10	-7.08
31	4.96	620.80	182.04	512.70	1 391.50	1.62	54.20	1.21	36.19	57.19	9.06	4	2	5794005	7983	-5.09	-7.08
32	5.05	597.72	168.07	467.30	1 300.20	1.63	58.90	1.27	30.50	36.19	7.02	5	2	4977453	7477	-3.36	-8.28
33	4.97	597.72	168.07	467.30	1 300.20	1.63	58.90	1.27	30.79	28.35	7.02	5	2	5063183	7609	-3.71	-8.29
34	5.35	597.72	168.07	467.30	1 300.20	1.63	58.90	1.27	30.84	30.70	7.02	5	2	5147818	7741	-3.19	-8.21
35	5.09	583.69	163.22	443.90	1 256.80	1.65	64.20	1.31	26.53	34.71	6.75	5	3	4460872	7010	-3.35	-8.28
36	5.22	583.69	163.22	443.90	1 256.80	1.65	64.20	1.31	26.90	25.93	6.75	5	3	4522185	7109	-3.69	-8.25
37	5.31	583.69	163.22	443.90	1 256.80	1.65	64.20	1.31	26.77	25.24	6.75	5	3	4582630	7208	-3.18	-8.21
38	5.42	673.81	192.66	534.60	1 472.30	1.64	57.50	1.26	38.74	54.25	8.75	5	2	7431755	9955	-3.75	-7.16
39	5.93	659.79	187.82	509.20	1 428.90	1.65	61.90	1.29	34.73	52.46	8.49	5	3	6737376	9386	-3.73	-7.16
40	5.57	734.86	213.16	561.70	1 566.50	1.68	60.40	1.30	40.86	42.03	8.57	6	3	10460242	13075	-3.81	-7.51
41	5.99	720.83	208.31	536.30	1 523.80	1.70	65.10	1.34	36.79	40.23	8.31	6	4	9581092	12408	-3.79	-7.50
42	5.25	720.83	208.31	536.30	1 523.80	1.70	65.10	1.34	37.86	51.44	8.31	6	4	9481653	12276	-3.48	-7.17
43	5.64	720.83	208.31	536.30	1 523.80	1.70	65.10	1.34	36.82	39.09	8.31	6	4	9679070	12540	-3.41	-7.50
44	5.37	738.82	208.31	540.50	1 530.90	1.69	64.30	1.36	37.87	53.11	8.47	7	4	10283328	12848	-3.11	-7.17
45	5.54	738.82	208.31	540.50	1 530.90	1.69	64.30	1.36	37.75	52.81	8.47	7	4	10319734	12895	-3.46	-7.16
46	6.26	738.82	208.31	540.50	1 530.90	1.69	64.30	1.36	37.95	53.38	8.47	7	4	10287768	12854	-2.85	-7.15
47	6.14	755.27	213.21	548.30	1 559.70	1.70	65.40	1.37	38.38	53.45	8.87	6	4	10287768	12854	-2.90	-7.17
48	5.92	799.73	216.00	552.50	1 574.30	1.71	65.80	1.44	38.55	53.51	9.14	6	4	10287768	12854	-3.01	-7.16
49	5.79	738.82	208.31	540.50	1 530.90	1.69	64.30	1.36	36.84	41.72	8.47	7	4	10421187	13024	-3.43	-7.50
50	5.84	755.27	213.21	548.30	1 559.70	1.70	65.40	1.37	37.80	47.39	8.87	6	4	10421187	13024	-3.48	-7.50
51	5.95	738.82	208.31	540.50	1 530.90	1.69	64.30	1.36	36.73	47.20	8.47	7	4	10356370	12942	-3.01	-7.50
52	6.11	738.82	208.31	540.50	1 530.90	1.69	64.30	1.36	36.79	47.11	8.47	7	4	10456004	13071	-3.40	-7.50
53	5.59	738.82	208.31	540.50	1 530.90	1.69	64.30	1.36	36.53	47.90	8.47	7	4	10488856	13112	-3.13	-7.49

Molecular Weight MW, Molecular Refractivity MR, Molar Volume VM, Parachor Pc, Density D, Partition coefficient Log P, Van Der Waals VDW, Total Energy E_T, HOMO energy E_{HOMO}, LUMO energy E_{LUMO}, Hydrogen Bond Donor HBD*, Hydrogen Bond Acceptor* HBA*, Surface Tension ST, Index of Refraction IR*

Principal component analysis

A total of 16 descriptors encoding the 53 molecules were submitted to a principal component analysis (PCA).[36] The first three principal axes sufficiently described the information provided by the data matrix. The percentages of variance were 60.86%, 18.74%, and 11.38% for axes F1, F2, and F3, respectively. The total variance was estimated to be 90.98%. PCA was conducted to identify the link among the different variables.[37] Table 3 presents the correlation matrix representing the correlations among the sixteen descriptors. The correlation matrix confirmed the absence of serious collinearity between the descriptors present in the model; therefore, the highly correlated descriptors ($R \geq 0.9$) were excluded.

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

Table 3: Correlation matrix between different descriptors obtained

	PIC_{50}	MW	MR	VM	Pc	IR	ST	D	1,4VDW	ET	log *P*	HBA	HBD	IB	IW	E_{LUMO}	E_{HOMO}
PIC_{50}	1																
MW	0.810	1															
MR	0.788	0.988	1														
VM	0.509	0.852	0.874	1													
Pc	0.715	0.969	0.979	0.944	1												
IR	0.804	0.684	0.672	0.229	0.526	1											
ST	0.760	0.582	0.551	0.105	0.425	0.951	1										
D	0.817	0.708	0.657	0.236	0.530	0.961	0.951	1									
1.4VDW	0.441	0.697	0.758	0.800	0.773	0.292	0.124	0.212	1								
ET	0.433	0.514	0.560	0.445	0.509	0.442	0.294	0.349	0.590	1							
log *P*	0.074	-0.113	0.064	-0.236	-0.157	0.232	0.172	0.104	0.237	0.324	1						
HBA	0.780	0.870	0.827	0.631	0.785	0.696	0.636	0.766	0.373	0.343	-0.259	1					
HBD	0.841	0.791	0.754	0.410	0.666	0.889	0.889	0.919	0.235	0.304	-0.079	0.848	1				
IB	0.572	0.879	0.865	0.933	0.921	0.313	0.221	0.377	0.609	0.314	-0.446	0.794	0.555	1			
IW	0.769	0.982	0.975	0.877	0.969	0.619	0.516	0.642	0.669	0.465	-0.217	0.899	0.758	0.932	1		
E_{LUMO}	0.595	0.422	0.366	0.002	0.265	0.730	0.803	0.784	0.049	0.108	0.286	0.468	0.697	0.086	0.333	1	
E_{HOMO}	0.547	0.531	0.560	0.320	0.478	0.626	0.559	0.548	0.564	0.602	0.524	0.304	0.531	0.179	0.414	0.615	1

Base data division

In this study, the k-means method was used to divide the observations into homogeneous clusters based on their descriptions. To divide the dataset into a training set (80% of the compounds) and a test set (20% of the compounds), a combination of k-means clustering results was used (Table 4). The test set consists of eleven molecules (2, 14, 15, 18, 19, 41, 23, 34, 47, 53, and 48) one from each cluster. The remaining 42 molecules were used for the training set.

Table 4: The k-means clustering results

1	*2,34,14,15,18,19,25,26,41,53,47*
2	2,14,15,17,19,41,23,24,34,53,48
3	2,34,14,15,18,19,41,23,39,53,47
4	2,8,38,18,19,41,25,34,53,47,48

Multiple linear regression

A mathematical linear model was proposed based on the selected descriptors to quantitatively predict the physicochemical effects of substituents on the inhibitory activity of the 53 molecules using multiple linear regression A total of forty-two molecules were placed in the training set to build the QSAR models. The remaining eleven molecules comprised the test set. (Equation 1).

$$Y = a_0 + \sum_{i=1}^{n} a_i x_i \quad \text{(Eq. 1)}$$

The linear model using this method includes two molecular descriptors: the molecular weight and the molecular volume. The following Equation 2 represents the best obtained linear QSAR model using the regression linear multiple regression (MLR) method:

$$\mathbf{PIC_{50}} = 3.106 + 0.007\times\mathbf{MW} - 0.005\times\mathbf{VM} \quad \text{(Eq. 2)}$$

N = 42; **R** = 0.892; $\mathbf{R^2}$ = 0.796; $\mathbf{R^2_{cv}}$= 0.769; **MSE** = 0.042; **F** = 76.20; **P** < 0.0001

The established models were judged by the statistical keys, such as, $\mathbf{R^2}$ which is the coefficient of determination, **F** is the Fisher statistic and **MSE** is the mean squared error. The model is more reliable, as demonstrated by its higher coefficient of determination and lower mean squared error. A **P** less than 0.05 means that the obtained equation is statistically significant at the 95% level. The leave one out cross-validated correlation coefficient LOO ($\mathbf{R^2_{cv}}$ = 0.769) illustrates the reliability of the model by focusing on the sensitivity of the model towards the elimination of any single data point. A value of $\mathbf{R^2_{cv}}$ greater than 0.5 is the basic criteria to qualify a model as valid.[28]

The multi-collinearity between the two chosen descriptors was evaluated by calculating their variation inflation factors **VIF** as shown in Table 5.

The **VIF** was defined as $1/(1 - R^2)$

Where R is the coefficient of correlation between one descriptor and all the other descriptors in the proposed model.[38]

Table 5: Multi-collinearity test

Descriptors	**MW**	**MV**
VIF	3.693	3.693

A **VIF** value greater than 5.0 indicates that the model is unstable; a value between 1.0 and 4.0 indicates that the model is acceptable. Accordingly, it was found that the descriptors used in the proposed model have very low-inter-correlation and the obtained model is stable.

In comparing the importance of each descriptor on pIC_{50} of substituted PTP1B, it is important to know their standardized coefficient and the **t-test** values in the MLR equations. The

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

bigger the absolute value of the **t-test** value, the greater the influence of the descriptor.
In Equation 1, the **t-test** values were 9.362 and -3.700 for **MW** and **MV** respectively.
By interpreting the molecular descriptors in the regression model, it is possible to gain insight into structural features that are likely to govern the activity of the studied compounds, which can be used to study the activity of structurally related compounds.
There were two descriptors in the regression model that proved to be important and made statistically significant contribution to the model.
As indicated by the **t-test** values, **MW** appeared as the most significant descriptor for the derived QSAR model. Thus, it can be inferred that the molecular weight is the most important descriptor for the activity. The developed model suggests that higher weight results in higher activity.
Molar volume (**MV**) followed molecular weight (**WM**) in importance for their contribution to the output of this QSAR model. This descriptor is one of the important polarizability parameters and can represent the volume occupied by an atom or a molecule. According to the model, **MV** has a negative contribution towards the activity value as evidenced by the negative regression coefficient. The smaller the **MV** value, the higher the activity of the compound.
The predicted values computed using this MLR model with the experimental values for the training and test sets are plotted in Figure 1. The selected descriptors (Eq. 2) in the MLR model were then used as the input variables to perform the multiple nonlinear regression (MNLR).

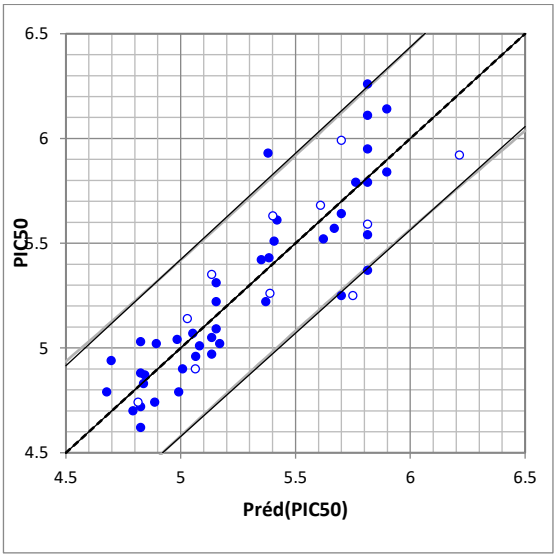


Figure 1: Predicted and observed activity values calculated by MLR

Multiple nonlinear regression
The nonlinear regression model was used to evaluate how the substituents in the studied PTP1B compounds affect inhibitory activity. This improved the quantitative structure-activity relationship.
The 42 molecules in the training set were used to build a non-linear model with the descriptors proposed by multiple linear regression. The best regression performance was selected according to the coefficient of determination (R^2) and the mean squared error MSE, a pre-programmed function in the XLSTAT was used to evaluate the nonlinear regression model as follows (Equation 3):

$$Y = a + (bX_1 + cX_2 + dX_3 + eX_4 \ldots) + (fX_1^2 + gX_2^2 + hX_3^2 + iX_4^2 \ldots) \quad \text{(Eq. 3)}$$

Where X_1, X_2, X_3, X_4…: represent the variables, and a, b, c, d…: represent the parameters.
The resulting Equation 4 is as follows:

$$\mathbf{PIC_{50}} = 4.058 + 0.005\times\mathbf{MW} - 0.006\times\mathbf{VM} + 2.00410^{-6}\times\mathbf{MW^2} + 4.93710^{-7}\times\mathbf{VM^2} \quad \text{(Eq. 4)}$$

N = 42; **R** = 0.892; $\mathbf{R^2}$ = 0.796; $\mathbf{R^2_{cv}}$= 0.755; **MSE** = 0.044
The leave one out cross-validated correlation coefficient LOO ($\mathbf{R^2_{cv}}$ = 0.755) illustrates the reliability of the model by focusing on the sensitivity of the model towards the elimination of any single data point. A value of $\mathbf{R^2_{cv}}$ greater than 0.5 is the basic criteria to qualify a model as valid.[28] It can be seen clearly from the key statistical indicators, coefficient of determination ($\mathbf{R^2}$), mean squared error (**MSE**) and, value of $\mathbf{R^2_{cv}}$, that the predicting ability of this model is like that of the linear model (MLR). The correlations of predicted and observed activities and the residual graph of absolute numbers are illustrated in Figure 2.

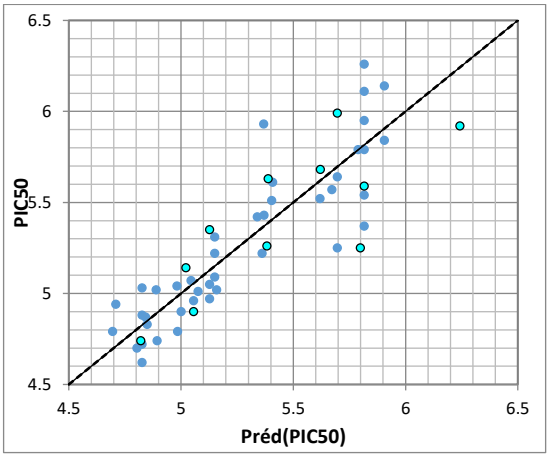


Figure 2: Predicted and observed activity values calculated by MNLR

Applicability domain
The utility of a QSAR model is its accurate prediction ability for new chemical compounds. So, once the QSAR model is built, its domain of applicability (AD) must be defined. A model is regarded valid only within its training domain and only the prediction for new compounds falling within its applicability domain can be considered reliable and not model extrapolations. The most common method to define the AD is based on the determination of the leverage value of each compound.[30] The Williams plot [the plot of standardized residuals versus leverage values (h)] was used in the present study to visualize the AD of the QSAR model (Equation 5).

$$h_i = x_i^T (X^T X)^{-1} x_i \quad \text{(Eq. 5)}$$

Where x_i is the descriptor vector of the considered compound, X is the descriptor matrix derived from the training set descriptor values, the threshold is defined as Equation 6:

$$h^* = \frac{3(k+1)}{n} \quad \text{(Eq. 6)}$$

Where n is the number of compounds in the training set, k is the number of descriptors in the proposed model, a leverage (h) greater than the threshold (h^*) indicates that the predicted response is an extrapolation of the model and, consequently, it can be unreliable.
The Williams plot of the presented MLR model is shown in Figure 3, the applicability domain is established inside a squared area within ±3 standard deviation and a leverage threshold h^* of 0.214 for MLR model.

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

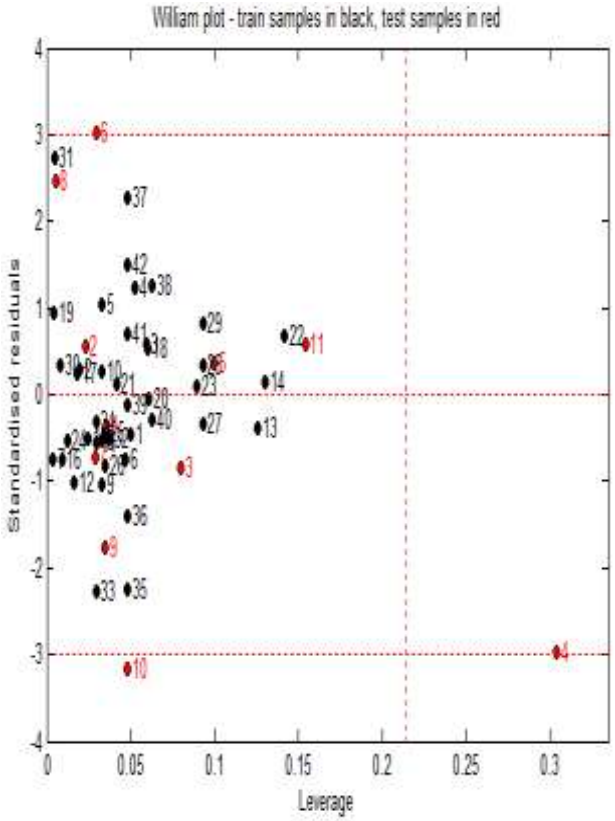


Figure 3: Williams plot of standardized residual versus leverage for the presented MLR model (with h^* = 0.214 and residual limits ±3)

As shown in the Williams plot, the majority of the compounds in the data set were in this area, except compound 17 (numbered as compound 4 in the Figure according to the test set) in the test set which exceeded the threshold and was considered as an outlier compound. In addition, the standardized residuals of all compounds were less than ±3 standard deviation, except compound 53 (numbered as compound 10 in the Figure according to the test set) which was wrongly predicted (>3σ) with a lower leverage value ($h < h^*$). Therefore, the predicted activity by the developed MLR model is reliable.

External validation

To estimate the predictive ability of the MLR and MNLR models, we have to use a set of compounds that were not used as the training set to establish the QSAR model. The models established in the computation process using the 42 substituted PTP1B were used to predict the inhibitory activity of the rest 11 compounds. The main performance parameters of the two models are shown in Table 6. As seen from this table, for the studied series of compounds, the statistical parameters of the two models were nearly the same. Among the obtained models for this series, the MLR model had the highest prediction ability for the test set ($\mathbf{R^2_{test}}$ = 0.815), as well as the highest cross-validation coefficient ($\mathbf{R^2_{cv}}$ = 0.769), all of which supported the applicability of the proposed MLR prediction model. However, both the results obtained by the MLR and MNLR should be regarded as satisfactory for predicting the inhibitory activity using the proposed descriptors.

Table 6: The statistical results of MLR and MNLR models with validation techniques

Method/ parameter						
MLR	**MLR**	**0.892**	**0.796**	**0.769**	**0.815**	**0.042**
MNLR	**MNLR**	**0.892**	**0.796**	**0.755**	**0.639**	**0.044**

Docking results

To investigate the probable binding mode between this series of studied compounds and the PTP1B receptor as well as to better understand and support the *in vitro* activity of the studied compounds for the rational design of drugs, a molecular docking study of the most active compound **46** was carried out into the active site of the PTP1B protein (PDB ID: **1YNN**) and the binding interactions of most active molecule, compound **46** with PTP1B binding pocket are shown in Figure 4. The binding pocket of PTP1B is mainly contributed by residues ASP (A:48), TYR (A:46), ASP (A:265), ARG (A:221), CYS (A:215), THR (A:263), GLY (A:183). The docking of most active compound **46** showed many hydrogen bond formations with the active site residues ASP (A:48), TYR (A:46).

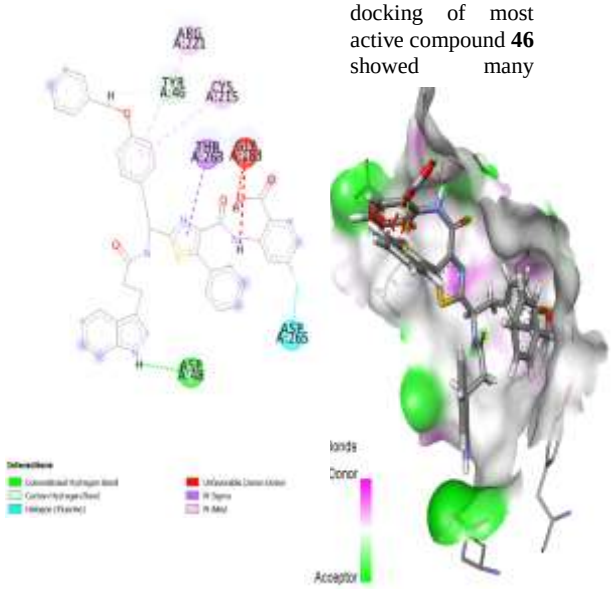


Figure 4: The Binding conformations (2D and 3D Binding pose views) and ligand interaction of the most active molecule (compound 46) at active site of the PTPB1 receptor (PDB ID: **1YNN**), visualized with Discovery studio program.

The present *in silico* experiments has demonstrated that compound **46** may exhibit different type of interactions in the binding site of PTP1B, which may lead to a high inhibitory activity against PTP1B, and thus may act as a drug.

Conclusion

This study underscores the potential of PTP1B as a promising therapeutic target in the treatment of diabetes mellitus and obesity. A comprehensive review of the existing literature yielded a series of PTP1B inhibitors, which were then evaluated for their efficacy through the application of 2D QSAR modeling and molecular docking techniques.
Furthermore, molecular docking provided valuable insights

into the inhibitory mechanisms of these compounds against PTP1B. The quality studies of the MLR and MNLR models have shown that both models have significantly better predictive capacity. Future studies will focus on the investigation of this series of heterocyclic compounds for other types of biological activity and to use other methods, as well as to synthesize the proposed molecules.
As computational power and algorithmic sophistication continue to evolve, 2D-QSAR is poised to remain a cornerstone of rational drug design, especially when integrated with emerging AI-driven methodologies.

Conflict of Interest

The authors declare no conflict of interest.

Authors' Declaration

The authors hereby declare that the work presented in this article is original and that any liability for claims relating to the content of this article will be borne by them.

References

1. DeFronzo RA, Bonadonna RC, Ferrannini E. Pathogenesis of NIDDM. A Balanced Overview. Diabetes Care. 1992; 15(3):318-368. doi: 10.2337/diacare.15.3.318.
2. Zimmet P, Alberti KG, Shaw J. Global and Societal Implications of the Diabetes Epidemic. Nature. 2001; 414(6865):782-787. doi: 10.1038/414782a.
3. Saltiel AR and Kahn CR. Insulin Signalling and the Regulation of Glucose and Lipid Metabolism. Nature. 2001; 414(6865):799-806. doi: 10.1038/414799a.
4. Obici S, Feng Z, Karkanias G, Baskin DG, Rossetti L. Decreasing Hypothalamic Insulin Receptors Causes Hyperphagia and Insulin Resistance in Rats. Nat Neurosci. 2002; 5(6):566-572. doi: 10.1038/nn0602-861.
5. Johnson TO, Ermolieff J, Jirousek MR. Protein Tyrosine Phosphatase 1B Inhibitors for Diabetes. Nat Rev Drug Discov. 2002; 1(9):696-709. doi: 10.1038/nrd895.
6. Drake PG and Posner BI. Insulin Receptor-Associated Protein Tyrosine Phosphatase(s): Role in Insulin Action. Mol Cell Biochem. 1998; 182(1-2):79-89.
7. White MF and Kahn CR. The Insulin Signaling System. J Biol Chem. 1994; 269(1):1-4.
8. Byon JC, Kusari AB, Kusari J. Protein-Tyrosine Phosphatase-1B Acts as a Negative Regulator of Insulin Signal Transduction. Mol Cell Biochem. 1998; 182(1-2):101-108.
9. Dadke S and Chernoff J. Interaction of Protein Tyrosine Phosphatase (PTP) 1B With Its Substrates is Influenced by Two Distinct Binding Domains. Biochem J. 2002; 364(Pt 2):377-383. doi: 10.1042/BJ20011372.
10. Bandyopadhyay D, Kusari A, Kenner KA, Liu F, Chernoff J, Gustafson TA, Kusari J. Protein-Tyrosine Phosphatase 1B Complexes with the Insulin Receptor in Vivo and is Tyrosine-Phosphorylated in the Presence of Insulin. J Biol Chem. 1997; 272(3):1639-1645. doi: 10.1074/jbc.272.3.1639.
11. Kusari J, Kenner KA, Suh KI, Hill DE, Henry RR. Skeletal Muscle Protein Tyrosine Phosphatase Activity and Tyrosine Phosphatase 1B Protein Content are Associated with Insulin Action and Resistance. J Clin Invest. 1994; 93(3):1156-1162. doi: 10.1172/JCI117068.
12. González-Díaz H. Computational Prediction of Drug-Target Interactions in Medicinal Chemistry. Curr Top Med Chem. 2013; 13(14):1619-1621. doi: 10.2174/15680266113139990112.
13. González-Díaz H, Arrasate S, Sotomayor N, Lete E, Munteanu CR, Pazos A, Besada-Porto J, Ruso JM. MIANN Models in Medicinal, Physical and Organic Chemistry. Curr Top Med Chem. 2013; 13(5):619-641. doi: 10.2174/1568026611313050006.
14. Abeijon P, Garcia-Mera X, Caamano O, Yanez M, Lopez-Castro E, Romero-Duran FJ, Gonzalez-Diaz H. Multi-Target Mining of Alzheimer Disease Proteome with Hansch's QSBR-Perturbation Theory and Experimental-Theoretic Study of New Thiophene Isosters of Rasagiline. Curr Drug Targets. 2017; 18(5):511-521. doi: 10.2174/1389450116666151102095243.
15. Todeschini R, Pazos A, Arrasate S, González-Díaz H. Data Analysis in Chemistry and Bio-Medical Sciences. Int J Mol Sci. 2016; 17(12):2105. doi: 10.3390/ijms17122105.
16. González-Díaz H, Herrera-Ibatá DM, Duardo-Sánchez A, Munteanu CR, Orbegozo-Medina RA, Pazos A. ANN Multiscale Model of Anti-HIV Drugs Activity vs AIDS Prevalence in the US at County Level Based on Information Indices of Molecular Graphs and Social Networks. J Chem Inf Model. 2014; 54(3):744-755. doi: 10.1021/ci400716y.
17. Duardo-Sánchez A, Munteanu CR, Riera-Fernández P, López-Díaz A, Pazos A, González-Díaz H. Modeling Complex Metabolic Reactions, Ecological Systems, and Financial and Legal Networks with MIANN Models Based on Markov-Wiener Node Descriptors. J Chem Inf Model. 2014; 54(1):16-29. doi: 10.1021/ci400280n.
18. Agarwal N, Bajpai A, Gupta SP. A Quantitative Structure-Activity Relationship and Molecular Modeling Study on a Series of Heteroaryl- and Heterocyclyl-Substituted Imidazo[1,2-a] Pyridine Derivatives Acting as Acid Pump Antagonists. Biochem Res Int. 2013; 2013:141469. doi: 10.1155/2013/141469.
19. Moretto AF, Kirincich SJ, Xu WX, Smith MJ, Wan ZK, Wilson DP, Follows BC, Binnun E, Joseph-McCarthy D, Foreman K, Erbe DV, Zhang YL, Tam SK, Tam SY, Lee J. Bicyclic and Tricyclic Thiophenes as Protein Tyrosine Phosphatase 1B Inhibitors. Bioorg Med Chem. 2006; 14(7):2162-2177. doi: 10.1016/j.bmc.2005.11.005.
20. Erbe DV, Klaman LD, Wilson DP, Wan ZK, Kirincich SJ, Will S, Xu X, Kung L, Wang S, Tam S, Lee J, Tobin JF. Prodrug Delivery of Novel PTP1B Inhibitors to Enhance Insulin Signalling. Diabetes Obes Metab. 2009; 11(6):579-588. doi: 10.1111/j.1463-1326.2008.01022.x.
21. Wan ZK, Lee J, Xu W, Erbe DV, Joseph-McCarthy D, Follows BC, Zhang YL. Monocyclic Thiophenes as Protein Tyrosine Phosphatase 1B Inhibitors: Capturing Interactions with Asp48. Bioorg Med Chem Lett. 2006; 16(18):4941-4945. doi: 10.1016/j.bmcl.2006.06.051.
22. Wang F and Zhou B. Toward the Identification of a Reliable 3D-QSAR Model for the Protein Tyrosine Phosphatase 1B Inhibitors. J Mol Struct. 2018; 1158:75-87. doi.org/10.1016/j.molstruc.2018.01.011.
23. Advanced Chemistry Development Inc. Toronto, Canada; 2009. www.acdlabs.com/resources/freeware/chemsketch/.
24. ChemBioOffice, PerkinElmer Informatics, 2010. http://www.cambridgesoft.com.
25. Larif M, Adad, Hmammouchi R, Taghki AI, Soulaymani, Elmidaoui A, Bouachrine M, Lakhlifi T. Biological Activities of Triazine Derivatives Combining DFT and QSAR Results. Arab J Chem. 2017; 10:s946-s955.
26. XLSTAT Software, XLSTAT Company, 2014. http://www.xlstat.com.
27. Ghamali M, Chtita S, Aouidate A, Ghaleb A, Bouachrine M, Lakhlifi T. Combining DFT and QSAR Computation to Predict the Interaction of Flavonoids with the GABA (A) Receptor Using Electronic and Topological Descriptors. J Taibah Univ Sci. 2017; 11:422-433.
28. Golbraikh A and Tropsha A. Beware of q2!. J Mol Graph Model. 2002; 20(4):269-276. doi: 10.1016/s1093-3263(01)00123-1.
29. Roy PP and Roy K. On Some Aspects of Variable Selection for Partial Least Squares Regression Models. QSAR Comb Sci. 2008; 27:302-313. doi.org/10.1002/qsar.200710043.

30. Gramatica P. Principles of QSAR Models Validation: Internal & External. QSAR Comb Sci. 2007; 26:694-701. doi.org/10.1002/qsar.200610151.
31. Jain AN. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. J Med Chem. 2003; 46:499-511.
32. M Asmae, T Hamid, A Toufik, L Fatima. In Silico Screening of New Derivatives as Inhibitors of Enoyl-[Acyl-Carrier-Protein] Reductase From Staphylococcus Aureus Via 2D-QSAR Analysis, Molecular Docking and ADME/Tox Prediction. Trop. J. Nat. Prod. Res. 2025; 9(4):1616 – 1624.
33. Wang J, Kollman PA, Kuntz ID. Flexible Ligand Docking: A Multistep Strategy Approach. Proteins. 1999; 36(1):1-19.
34. Purcell WP and Singer JA. A Brief Review and Table of Semiempirical Parameters Used in the Hückel Molecular Orbital Method. J Chem Eng Data. 1967; 12(2):235–246. doi.org/10.1021/je60033a020.
35. Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016. http://accelrys.com/products/collaborative-science/biovia-discovery-studio/.
36. STATITCF Software, Technical Institute of Cereals and Fodder, Paris, France, 1987.
37. Erazua EA, Akintelu SA, Adelowo JM, Odoemene SN, Josiah OM, Raheem SF, Latona DF, Adeoye MD, Esan AO, Oyebamiji AK. QSAR and Molecular Docking Studies on Nitro (Triazole/Imidazole)-Based Compounds as Anti-Tubercular Agents. Trop. J. Nat. Prod. Res. 2021; 5(11):2022-2029. Doi.org/10.26538/tjnpr/v5i11.22.
38. O'Brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Qual Quant. 2007; 41:673-690. doi.org/10.1007/s11135-006-901