

# 3D-QSAR of Protein Tyrosine Phosphatase 1B Inhibitors by Genetic Function Approximation

Sree M. Vadlamudi,<sup>1,\*</sup> Vithal M. Kulkarni,<sup>2</sup>

<sup>1</sup>Topotarget UK Ltd, 87 A, Milton Park, Abingdon, Oxon, OX144RY, United Kingdom

<sup>2</sup>Institute of Chemical Technology, Department of Pharmaceutical Sciences and Technology, University of Mumbai, Matunga (E), Mumbai 400019, India

Received xxx; Preprint published xxx; Accepted xxx ; Published xxx

*Internet Electron. J. Mol. Des.* 2003, 1, 000–000

## Abstract

Protein tyrosine phosphatase 1B (PTP 1B) has been implicated as negative regulator of insulin receptor signaling system. Design of small molecule PTP 1B inhibitors has received considerable attention as inhibition of PTP 1B enzyme is expected to improve insulin action, to treat non insulin dependent diabetes mellitus (NIDDM). In this work, we report three dimensional quantitative structure activity relationship (3D-QSAR) study performed by genetic function approximation (GFA) technique on a series of benzofuran/benzothiophene biphenyls as PTP 1B inhibitors. The QSAR models were generated using 92 compounds, and the predictive ability of the resulting each model was evaluated against a test set of 26 compounds. The internal and external consistency of the final QSAR model was 0.694 and 0.672. Analyses of results from the present QSAR study indicate that electronic, structural, and shape descriptors govern the PTP 1B inhibitory activity.

**Keywords.** Three-dimensional quantitative structure activity relationship; 3D-QSAR, Genetic function approximation; GFA; Protein Tyrosine Phosphates 1B; PTP 1B, non insulin dependent diabetes mellitus; NIDDM; descriptor.

## 1 INTRODUCTION

Resistance to the biological actions of insulin in its target tissues is a major feature of the path-physiology in human obesity and non-insulin dependent diabetes mellitus (NIDDM). Tyrosine phosphorylation of specific intracellular proteins controlled by the actions of protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs) is recognized as a key process by which a number of polypeptide hormones and growth factors transduce and coordinate their biological effects *in vivo* [1]. Recent insights into the mechanism of insulin actions 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 [2]. Biochemical and cellular studies have provided evidences that PTPs have an important role in the regulation of insulin signal transduction [3].

\* Correspondence author; phone: 00-44-1235-443740; fax: 00-44-1235-835-557; E-mail: sreemv2003@yahoo.co.uk

Protein tyrosine phosphates 1B (PTP 1B), a cytosolic PTP play a major role in the regulation of insulin sensitivity and dephosphorylation of the insulin receptor. PTP 1B has been implicated as negative regulator of insulin receptor signaling [4,5]. Clinical studies have found a correlation between insulin resistance states and levels of PTP 1B expression in muscle and adipose tissue, suggesting that PTP 1B has a major role in the insulin resistance associated with obesity and NIDDM [6, 7]. A recent pivotal PTP 1B knock out study revealed that mice lacking functional PTP 1B were viable, healthy, and lean; displayed enhanced insulin sensitivity and resistance to diet induced obesity [8]. All these results establish a direct role for PTP 1B in down regulating the insulin functions. Hence potent, selective and orally active PTP 1B inhibitors could be potential pharmacological agents for the treatment of obesity and NIDDM.

Malamas et al. reported a series of 118 molecules belonging to benzofuran/benzothiophene biphenyls as PTP 1B inhibitors with antihyperglycemic activity [9]. The logarithm of measured  $IC_{50}$  ( $\mu\text{M}$ ), against human recombinant PTP 1B enzyme (h-PTP 1B) as  $pIC_{50}$  was used as dependent variable for the present QSAR analysis. The activity data used in the present study is *in vitro* data and such a type of activity data could have contributions not only from steric and electrostatic interactions but also from other physicochemical parameters. The factors contributing to the biological activity can be understood through use of different physicochemical descriptors in the generation of QSAR models. Analysis of antibacterial activity of macrolide compounds has resulted in the identification of physicochemical descriptors such as  $\log P$ ,  $\log D$ , CMR,  $pK_a$  and HPLC capacity factor deriving correlation with *in vitro* MIC and *in vivo* activity [10]. Various quantum chemical and quantum mechanical descriptors are being applied for the quantitative structure activity relationship and quantitative structure pharmacokinetic relationship studies involving complex biological phenomenon [11]. Use of descriptors, which characterize the molecular shape related properties, might be especially useful to explain variance in the biological activity among a series of the compounds [12].

We have used genetic function approximation (GFA) technique to generate different 3D-QSAR models from various descriptors available within Cerius2 modeling software [13] in order to deduce correlation between the structure and biological activity of the present series of molecules. Our strategy follows the methodology used previously to generate successful 3D-QSAR models for antifungal [14], antibacterial [15], antitubercular [16], and anti-inflammatory agents [17]. GFA technique was used since it generates a population of equations rather than one single equation for correlation between biological activity and physicochemical properties. GFA developed by Rogers, involved the combination of Friedman's multi variate adaptive regression

splines (MARS) algorithm with Holland's genetic algorithm to evolve population of equations that best fit the training set data [18]. This is done as follows: (i) An initial population of equations is generated by random choice of descriptors. The fitness of each equation is scored by Lack-of-Fit (LOF) measure,  $LOF = LSE / \{1 - [c + d * p / m]\}^2$ , where LSE is least square error,  $c$  is the number of basis functions in the model,  $d$  is the smoothing parameter which controls the number of terms in the equations and  $p$  is the number of features contained in all terms of the models, and  $m$  is the number of compounds in the training set. (ii) Pairs from the population of equations are chosen at random and "crossovers" are performed and progeny equations are generated. (iii) The fitness of each progeny equation is assessed by LOF measure. (iv) If the fitness of new progeny equation is better, then it is preserved. The model with proper balance of all statistical terms will be used to explain variance in the biological activity.

A distinctive feature of GFA is that it produces a population of models (eg. 100), instead of generating a single model, as do most other statistical methods. The range of variations in this population gives added information on the quality fit and importance of the descriptors. By examining these models, additional information can be obtained. For example, the frequency of use of a particular descriptor in the population of equations may indicate how relevant the descriptor is to the prediction of activity. Combination of robust statistical technique GFA coupled with the use of different types of descriptors would result in better prediction of biological activity for PTP1B enzyme inhibitors.

## **2 MATERIALS AND METHODS**

### **2.1 Chemical Data**

#### **2.1.1 Molecules**

A series of 118 molecules belonging to benzofuran/benzothiophene biphenyls as PTP 1B inhibitors were taken from the literature and used for the present study [9]. The 3D-QSAR models were generated using a training set of 92 molecules. The structures observed and predicted biological activities of the training set molecules are presented in Table 1. Predictive power of the resulting models was evaluated by a test set of 26 molecules with uniformly distributed biological activity. The structures observed and predicted biological activities of the test set molecules are presented in Table 2. Selection of test set molecules was made by considering the fact that, test set molecules represent range of biological activity similar to training set. The mean of biological activity of training and test set was 0.65 and 0.69, respectively. Thus test set is the true representative of the training set.

### **2.1.2 Biological activity**

The logarithm of measured  $IC_{50}$  ( $\mu\text{M}$ ) against human recombinant PTP 1B (h-PTP 1B) enzyme as  $pIC_{50}$  ( $pIC_{50} = \log 1/IC_{50}$ ) was used as dependent variable, thus correlating the data linear to the free energy change. Since  $IC_{50}$  against h-PTP 1B for compounds exhibiting  $>70\%$  inhibition at 2.5  $\mu\text{M}$  concentration (average of quadruplet) was not determined, such compounds were excluded from the present study. Further details regarding the biological testing can be found in [9].

## **2.2 Molecular Modeling**

### **2.2.1 Software**

All molecular modeling studies were carried out using Cerius2 (version 3.5) running on Silicon Graphics O2 R5000 workstation [13]. Structures were constructed from the builder module and partial charges were assigned using charge equilibration method within Cerius2 [19]. Throughout the study, Universal forcefield 1.02 was used [20]. The molecules were subsequently minimized until a root mean square deviation 0.001 kcal/mol  $\text{\AA}$  was achieved and used in the study.

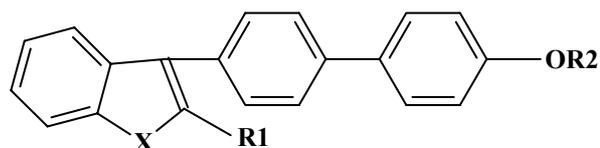
### **2.2.2 Calculation of descriptors**

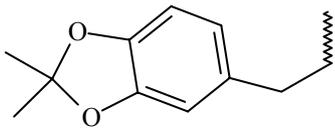
Different types of descriptors were calculated for each molecule in the study table using default settings within Cerius2. These descriptors include electronic, spatial, structural, thermodynamic and molecular shape analysis (MSA). A complete list of descriptors used in the study is given in given Table 3.

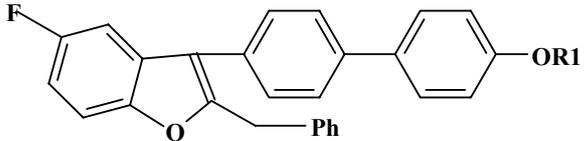
### **2.2.3 MSA descriptors**

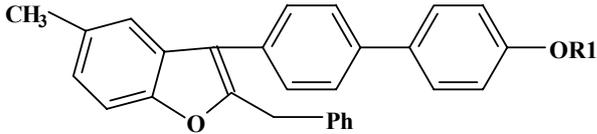
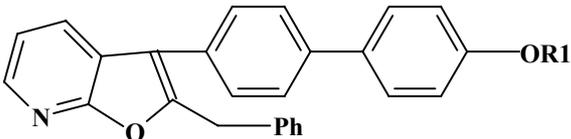
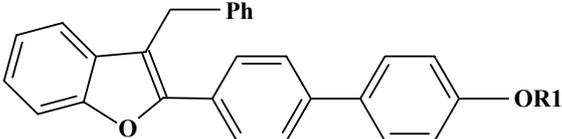
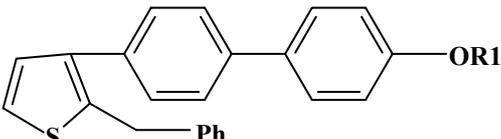
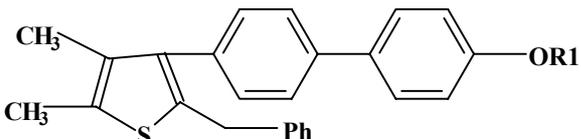
MSA descriptors [21] were calculated using MSA module within Cerius2. As MSA descriptors calculate three-dimensional properties of the ligands, knowledge of active conformer of the molecules under study is essential. The crystallographic conformation of the present series of molecules was not available/deposited at protein data bank. Hence conformational analysis on all molecules was performed using random sampling search [22] and Universal force field [20], with maximum number of conformers set equal to 10. The lowest energy conformer of the molecule with the highest biological activity (compound 54, Table 1) was used as reference for calculation of MSA descriptors. Crystallographically two different orientations were shown for phenyllactic acid and sulfosalicylic acid type of inhibitors [9]. In the present QSAR study no descriptor related to ligand-enzyme interactions (such as binding energy) has been used. Hence we think that differences in binding orientations would have no effect on the conclusions from present QSAR analysis.

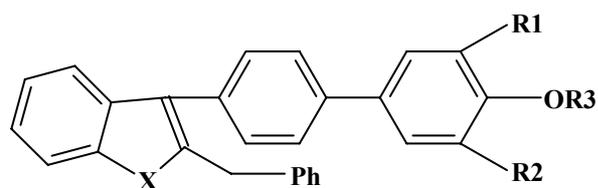
**Table 1.** Structures and biological activities of the training set molecules.



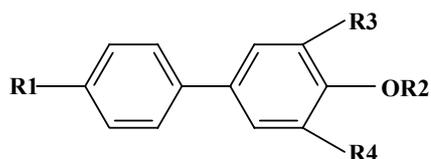
No	R1	R2	X	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
1	butyl	H	O	0.130	-0.048
2	benzyl	H	O	0.0362	0.147
3	benzoyl	H	O	0.130	0.065
4	butyl	H	S	0.154	0.164
5	4-OH benzyl	H	S	-0.033	-0.045
6	benzyl	CH(CH <sub>2</sub> Ph)COOH (R)	O	0.455	0.676
7	benzyl	CH(CH <sub>2</sub> CH <sub>2</sub> Ph)COOCH <sub>3</sub> (S)	O	0.657	0.740
8	benzyl	CH(CH <sub>2</sub> CH <sub>2</sub> -N-phtalinimide) – COOH(S)	O	0.468	1.153
9	benzyl	CCH <sub>3</sub> (CH <sub>2</sub> Ph)COOH(R)	O	0.537	0.845
10	benzyl	CH(CH <sub>3</sub> )COOH(R)	O	0.120	0.368
11	benzoyl	CH(CH <sub>2</sub> Ph)COOH(R)	O	0.167	0.580
12	CH(OH)phenyl	CH(CH <sub>2</sub> Ph)COOH(R)	O	0.958	0.587
13	benzyl	CH <sub>2</sub> Ph-4-COOH	O	0.443	0.778
14	butyl	CH(CH <sub>2</sub> Ph)COOH(R)	S	0.769	0.518
15	benzyl	CH(CH <sub>2</sub> Ph)COOH(R)	S	1.022	0.743
16	butyl	CH(Ph)COOH (R)	S	0.958	0.713
17	4-F-benzyl	CH(CH <sub>2</sub> Ph)COOH (R)	S	0.920	0.681
18	4-OCH <sub>3</sub> -benzyl	CH(CH <sub>2</sub> Ph)COOH(R)	S	1.113	0.769
19	2,4-di-OH-benzyl	CH(CH <sub>2</sub> Ph)COOH (R)	S	0.920	0.681
20		CH(CH <sub>2</sub> Ph)COOH(R)	S	1.113	0.769
21	2-methyl thiazolo	CH(CH <sub>2</sub> Ph)COOH(R)	S	-0.064	0.616
22	2-methyl pyridyl	CH(CH <sub>2</sub> Ph)COOH(R)	S	-0.190	0.655

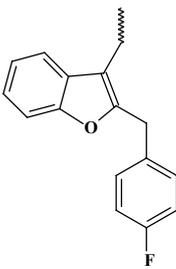
No	R	R1	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
23		CH(CH <sub>2</sub> Ph)COOH(S)	0.886	0.609

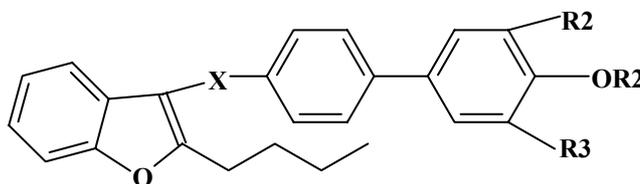
24		CH(CH <sub>2</sub> Ph)COOH(R)	0.387	0.862
25		CH(CH <sub>2</sub> Ph)COOH(S)	0.229	0.578
26		CH(CH <sub>2</sub> Ph)COOH(R)	0.455	0.678
27		CH(CH <sub>2</sub> Ph)COOH(R)	0.0132	0.361
28		CH(CH <sub>2</sub> Ph)COOH(R)	0.292	0.672



No	R1	R2	R3	X	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
29	Br	H	H	S	-0.029	0.456
30	Br	Br	H	S	0.346	0.806
31	I	I	H	S	0.283	0.813
32	Br	Br	CH(CH <sub>2</sub> Ph)COOH(R)	S	1.602	0.894
33	4-OCH <sub>3</sub> -Ph	H	CH(CH <sub>2</sub> Ph)COOH(R)	S	1.275	1.278
34	Br	Br	CH(CH <sub>2</sub> CH <sub>2</sub> Ph)COOH(S)	S	0.537	1.166
35	Br	Br	CH(CH <sub>2</sub> CH <sub>2</sub> -N-phtalimide)-COOH	S	1.356	1.563
36	Br	Br	CH(CH <sub>2</sub> CH <sub>2</sub> NHCOPh-2-COOH)COOH	S	0.744	1.466
37	Br	Br	CH(CH <sub>2</sub> CH <sub>2</sub> NHCOPh-2-COOH)COOCH <sub>3</sub>	S	1.267	1.540
38	Br	H	CH <sub>2</sub> COOH	S	0.443	0.574
39	Br	Br	CH <sub>2</sub> COOH	S	1.000	0.820
40	4-OCH <sub>3</sub> -Ph	H	CH <sub>2</sub> COOH	S	1.096	0.955
41	4-OC <sub>2</sub> H <sub>5</sub> -Ph	H	CH <sub>2</sub> COOH	S	1.283	0.968
42	2,3-di-OCH <sub>3</sub> -Ph	H	CH <sub>2</sub> COOH	S	1.148	1.094
43	3,4,5-tri-OCH <sub>3</sub> -Ph	H	CH <sub>2</sub> COOH	S	1.000	1.210
44	4-OCH <sub>3</sub> -Ph	Br	CH <sub>2</sub> COOH	S	1.537	1.226
45	2,4-di-OCH <sub>3</sub> -Ph	Br	CH <sub>2</sub> COOH	S	1.327	1.321
46	3-OCH <sub>3</sub> -Ph	3-OMe-Ph	CH <sub>2</sub> COOH	S	1.602	1.487
47	4-OCH <sub>3</sub> -Ph	4-OMe-Ph	CH <sub>2</sub> COOH	S	1.602	1.406
48	Br	H	CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> COOH	S	0.769	0.685
49	Br	H	CH(CH <sub>2</sub> PhCOOH(S))	O	1.251	0.835
50	4-OCH <sub>3</sub> -Ph	H	CH(CH <sub>2</sub> Ph)COOH(S)	O	1.366	1.303
51	NO <sub>2</sub>	H	CH(CH <sub>2</sub> Ph)COOH(R)	O	0.638	0.713
52	Br	Br	CH(C <sub>2</sub> H <sub>5</sub> )COOH(S)	O	0.886	0.902
53	Br	Br	CH[CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub> ]COOH(R)	O	1.267	0.572
54	Br	Br	CH[CH <sub>2</sub> ] <sub>5</sub> CH <sub>3</sub> ]COOH	O	1.638	0.932
55	CH <sub>3</sub>	CH <sub>3</sub>	CH(CH <sub>2</sub> Ph)COOH(R)	O	1.130	0.965
56	Cyclopentyl	H	CH(CH <sub>2</sub> Ph)COOH(S)	O	1.259	1.208
57	Cyclopentyl	H	CH <sub>2</sub> COOH	O	0.769	0.715
58	NHCH <sub>2</sub> CH <sub>2</sub> COOH	H	CH <sub>2</sub> CH <sub>2</sub> Ph	O	0.853	0.715
59	NHCOCH <sub>2</sub> CH <sub>2</sub> COOH	H	H	O	0.036	0.743
60	NHCOCH=CHCOOH	H	H	O	0.337	0.356
61	NHCO-C <sub>6</sub> H <sub>4</sub> -2-COOH	H	H	O	0.795	0.568

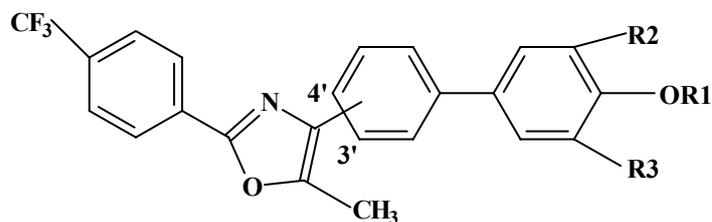


No	R1	R2	R3	R4	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
62		CH2COO H	4-OCH3- Ph	4-OCH3-Ph	1.508	1.318



No	R1	R2	R3	X	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
63	H	H	H	CH2	-0.0755	-0.077
64	H	Br	Br	CH(OH)	-0.146	0.458
65	CH2COOH	H	H	CH2	-0.060	0.138
66	CH2-tetrazole	H	H	CH2	0.292	0.240

[substituted oxazole biphenyls]

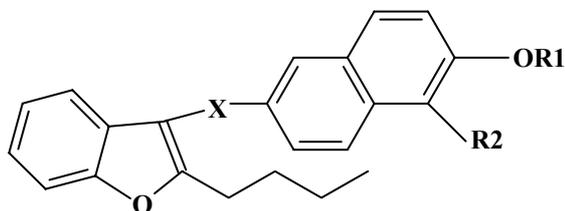


No	R1	R2	R3	P.O.A <sup>1</sup>	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
67	CH2COOH	H	H	4'	0.096	-0.278
68	CH(CH2Ph)COOH	H	H	4'	-0.113	0.114
69	CH2-tetrazole	H	H	4'	0.045	0.047
70	CH(CH2Ph)COOH	H	H	3'	-0.204	0.114
71	H	Br	Br	4'	0.187	0.308

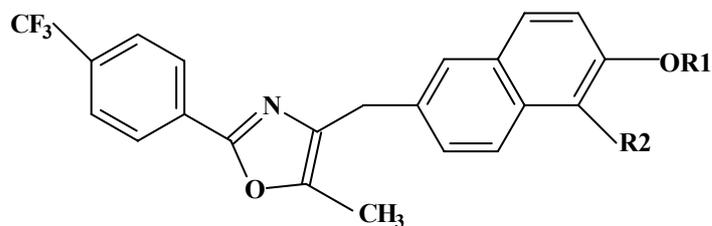
72	CH <sub>2</sub> COOH	Br	Br	4'	0.327	0.319
----	----------------------	----	----	----	-------	-------

<sup>1</sup>Point of attachment

[2-butyl benzofuran naphthalenes]

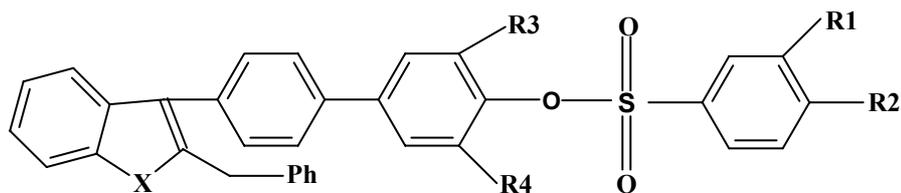


No	R1	R2	X	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
73	H	H	CH(OH)	-0.041	-0.227
74	H	Br	CH(OH)	0.318	-0.014
75	H	Br	CH <sub>2</sub>	0.481	0.068
76	H	I	CH <sub>2</sub>	0.420	0.082
77	CH <sub>2</sub> COOH	Br	CH <sub>2</sub>	-0.146	0.181
78	CH(CH <sub>2</sub> Ph)COOH	Br	CH <sub>2</sub>	0.431	0.539
79	CH(CH <sub>2</sub> Ph)COOH	Br	CO	-0.079	0.455
80	CH(CH <sub>2</sub> Ph)COOH	I	CH <sub>2</sub>	0.494	0.624
81	CH <sub>2</sub> -tetrazole	Br	CH <sub>2</sub>	0.154	0.344

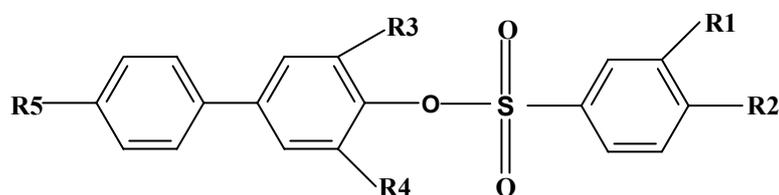


No	R1	R2	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
82	CH <sub>2</sub> COOH	Br	0.113	0.120

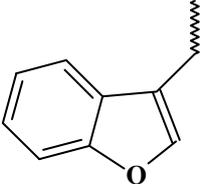
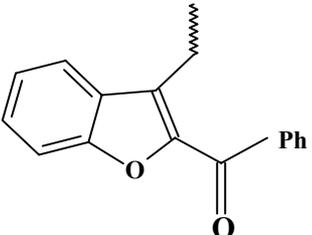
[sulphono biphenyls]



No	R1	R2	R3	R4	X	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
83	H	COOH	H	H	O	1.124	1.00
84	COOH	H	H	H	O	0.974	0.926
85	OH	COOH	H	H	O	1.408	1.110
86	OH	COOH	CH3	CH3	O	1.468	1.266
87	OH	COOH	H	H	S	1.552	1.562



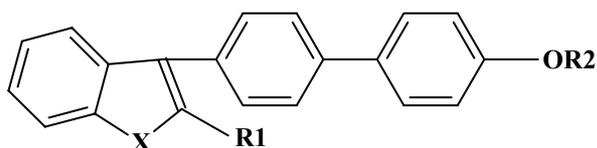
No	R1	R2	R3	R4	R5	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>
88	OH	COOH	H	H		1.494	1.032
89	OH	COOH	cyclopentyl	H		1.397	1.460
90	OH	COOH	H	H		0.450	0.347

91	OAc	COOH	H	H		-0.064	0.731
92	OH	COOH	NO <sub>2</sub>	H		0.749	1.010

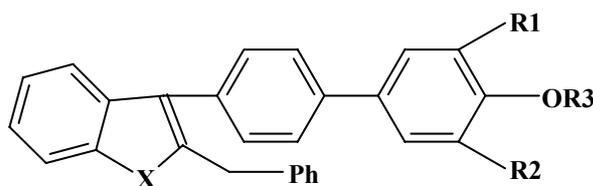
<sup>a</sup> Obsd. act = observed biological activity is defined as  $\log 1/IC_{50}$  against human recombinant PTP 1B enzyme (h-PTP 1B) in  $\mu\text{M}$ .

<sup>b</sup> Pred. act = Predicted biological activity calculated using Equation 6 in Table 4.

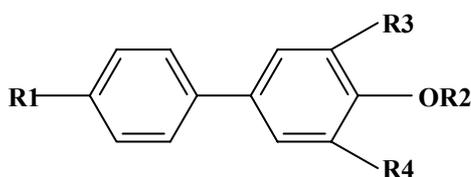
**Table 2.** Structures and observed, predicted activities along with residuals for the test set molecules.



No	R1	R2	X	Obsd.act <sup>a</sup>	Perd.act <sup>b</sup>	Residual
1	2,4-di-OH-benzyl	H	S	0.236	0.294	-0.058
2	butyl	CH <sub>2</sub> COOH	O	-0.340	0.527	0.187
3	butyl	CH(CH <sub>2</sub> Ph)COOH	O	0.356	0.504	-0.148
4	benzyl	CH(CH <sub>2</sub> Ph)COOH	O	0.568	0.775	-0.207
5	benzyl	CH(CH <sub>2</sub> Ph)COOH(S)	O	0.494	0.775	-0.281
6	benzyl	CH(Ph)COOH(R)	O	0.397	0.678	-0.281
7	3,4-OCH <sub>3</sub> -benzyl	CH(CH <sub>2</sub> Ph)COOH(R)	S	0.920	0.657	0.263
8	2,4-di-OCH <sub>3</sub> -benzyl	CH(CH <sub>2</sub> Ph)COOH	S	1.070	0.646	0.424

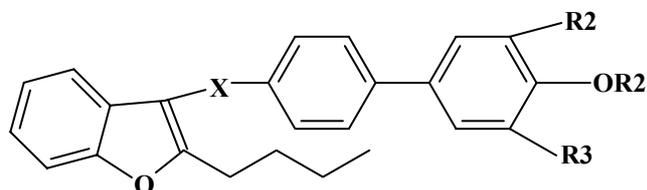


No	R1	R2	R3	X	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>	Residual
9	Br	H	CH(CH <sub>2</sub> Ph)COOH(R)	S	1.236	0.723	0.513
10	4-Cl-Ph	H	CH(CH <sub>2</sub> Ph)COOH(R)	S	1.283	0.955	0.328
11	Ph	H	CH <sub>2</sub> COOH	S	1.00	1.251	-0.257
12	3-OCH <sub>3</sub> -Ph	Br	CH <sub>2</sub> COOH	S	1.552	1.105	0.447
13	Br	Br	CH(CH <sub>2</sub> Ph)COOH(S)	O	1.420	0.850	0.570
14	Br	Br	CH[(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub> ]COOH	O	1.283	1.075	0.208
15	NHCH <sub>2</sub> COOH	H	CH <sub>2</sub> CH <sub>2</sub> Ph	O	1.086	0.699	0.387



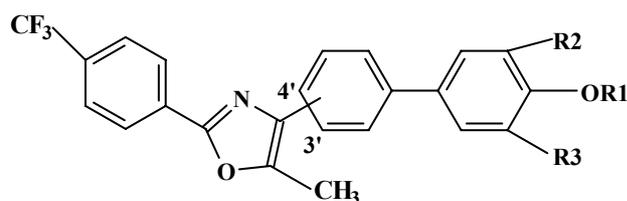
No	R1	R2	R3	R4	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>	Residual
16		CH <sub>2</sub> COOH	4-OCH <sub>3</sub> -Ph	4-OCH <sub>3</sub> -Ph	1.318	1.431	0.113

[2-butyl benzofuran biphenyls]



No	R1	R2	R3	X	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>	Residual
17	CH <sub>2</sub> COOH	H	H	CH(OH)	0.267	0.064	0.203

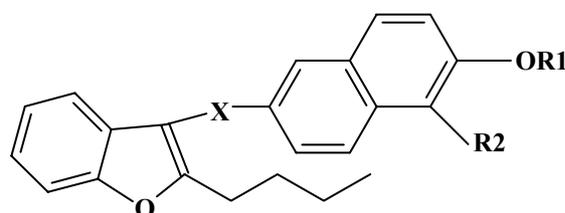
[oxazolo biphenyls]



No	R1	R2	R3	P.O.A <sup>1</sup>	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>	Residual
18	CH(CH <sub>2</sub> Ph)COOH	H	H	3'	-0.204	0.284	0.080
19	CH(CH <sub>2</sub> Ph)COOH	Br	Br	4'	0.886	0.5245	0.361

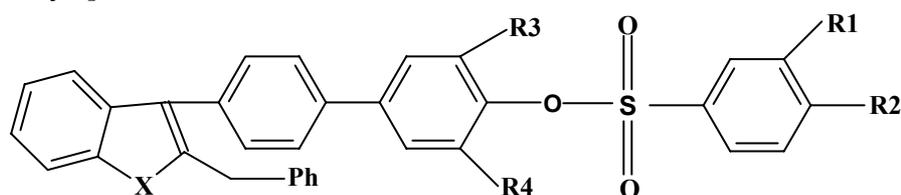
<sup>1</sup>Point of attachment

[2-butyl benzofuran naphthalenes]



No	R1	R2	X	Obsd.act <sup>a</sup>	Pred.act <sup>b</sup>	Residual
20	H	H	CH <sub>2</sub>	-0.113	-0.115	0.002
21	CH <sub>2</sub> -tetrazole	Br	CO	-0.04	0.173	0.133

[sulphono biphenyls]



No	R1	R2	R3	R4	X	Obsd.act <sup>a</sup>	Perd.act <sup>b</sup>	Residual
22	COOH	OH	H	H	O	1.585	0.981	0.604
23	OH	COOH	NO <sub>2</sub>	H	O	1.537	1.073	0.464
24	OH	COOH	Cyclopentyl	H	O	1.552	1.4612	0.090
25	OH	COOH	Br	H	S	1.619	1.238	0.381
26	OH	COOH	Br	Br	S	1.522	1.311	0.211

<sup>a</sup> Obsd. act = observed biological activity is defined as  $\log 1/IC_{50}$  against human recombinant PTP 1B enzyme (h-PTP 1B) in  $\mu\text{M}$

<sup>b</sup> Pred. act = Predicted biological activity calculated using Equation 6 in Table 4.

**Table 3.** Descriptors used in the present study

No	Descriptor	Type	Descriptors
1	Vm	Spatial	Molecular volume
2	Area	Spatial	Molecular surface area
3	Density	Spatial	Molecular density
4	RadOfGyr	Spatial	Radius of Gyration
5	PMI-mag	Spatial	Principle moment of inertia
6	PMI_X	Spatial	Principle moment of inertia X- component
7	PMI_Y	Spatial	Principle moment of inertia Y-component
8	PMI_Z	Spatial	Principle moment of inertia Z-component <sup>9</sup>
9	MW	Structural	Molecular weight
10	RotlBonds	Structural	Number of rotatable bonds
11	HbondAcc	Structural	Number of hydrogen bond acceptors
12	HbondDon	Structural	Number of hydrogen bond donors
13	AlogP	Thermodynamic	Logarithm of partition coefficient
14	MolRef	Thermodynamic	Molar refractivity
15	Dipole-mag	Electronic	Dipole moment
16	Dipole-X	Electronic	Dipole moment- X-component
17	Dipole-Y	Electronic	Dipole moment-Y-component
18	Dipole-Z	Electronic	Dipole moment-Z-component
19	Charge	Electronic	Sum of partial charges
20	Apol	Electronic	Sum of atomic polarizabilities
21	HOMO	Electronic	Highest occupied molecular orbital energy
22	LUMO	Electronic	Lowest unoccupied molecular orbital energy
23	Sr	Electronic	Superdelocalizability
24	Foct	Thermodynamic	Desolvation free energy for octanol
25	Fh2o	Thermodynamic	Desolvation free energy for water
26	Hf	Thermodynamic	Heat of formation
27	DIFFV	MSA	Difference volume
28	COSV	MSA	Common overlap steric volume
29	Fo	MSA	Common overlap volume ratio
30	NCOSV	MSA	Non-common overlap steric volume
31	Shape RMS	MSA	RMS to shape reference
32	SR Vol	MSA	Volume of shape reference compound

### 2.2.4 Generation of QSAR models

QSAR analysis in computational research is responsible for the generation of models to correlate biological activity and physicochemical properties of a series of compounds. The underlying assumption is that the variations of biological activity within a series can be

correlated with changes in measured or computed molecular features of the molecules. In the present study, QSAR model generation was performed by GFA technique. The application of the GFA algorithm allows the construction of high-quality predictive models and makes available additional information not provided by standard regression techniques, even for data sets with many features. GFA was performed using 100,000 crossovers, smoothness value of 2.0 and other default settings for each combination. The number of terms in the equation was fixed to five including constant in the training set. The set of equations generated were evaluated on the following basis: (a) LOF measure; (b) Variable terms in the equations; (c) Cross validated and non cross validated  $r^2$ ; (d) Randomized cross validated  $r^2$ ; (e) Predictive ability of equation. Cross validated  $r^2$  ( $r^2_{cv}$ ), Randomized cross validated  $r^2$ , were calculated using cross validated test option in the statistical tools supported in Cerius2.

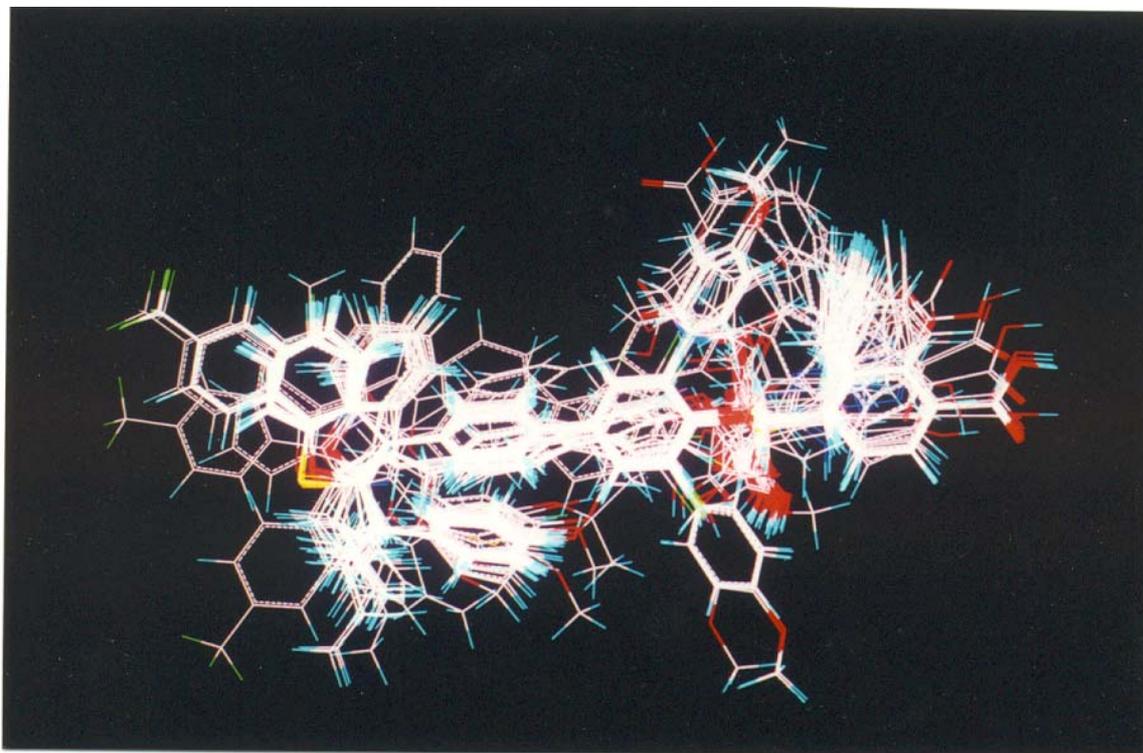
The predictive  $r^2$  was based only molecules not included in the training set and is defined as:  $r^2_{pred} = (SD - PRESS) / SD$ , where SD is the sum of the squared deviations between the biological activity of molecules in the test set and the mean biological activity of the training set molecules and PRESS is the sum of squared deviations between predicted and actual activity values for every molecule in the test set. Like  $r^2_{cv}$  the predictive  $r^2$  can assume a negative value reflecting a complete lack of predictive ability of the training set for the molecules included in the test set [23, 24].

### 3. RESULTS AND DISCUSSIONS

#### 3.1 Results

In the present study, QSAR models were generated using a training set of 92 molecules (Table 1). Test set of 26 molecules (Table 2) with regularly distributed biological activities was used to assess the predictive ability of the generated QSAR models. Biological activity was expressed in terms of pIC<sub>50</sub>, the logarithm of measured IC<sub>50</sub> (μM) against human recombinant PTP 1B (h-PTP 1B) enzyme. The conformational space of the rotatable bonds in the molecules was explored using random sampling technique in order to obtain sterically accessible conformations within optimum computational time. Conformational search was performed during molecular shape analysis (MSA) technique and the lowest energy conformer of each molecule was used for alignment using MSA technique. All the molecules were superimposed on the lowest energy conformer of the molecule with highest biological activity (compound 54). Pharmacophoric superposition of the molecules used in the present study is shown in Figure 1. The alignment resulted in the orientation of the molecule in such a way that, oxo-acetic acid functional group of compound 54 was oriented to z-axis (perpendicular to the plane of biphenyl ring), and substituents on the chiral carbon atom of oxo-

acetic acid group oriented to Y-axis. Conformations obtained from the random sampling technique were found to be similar to the conformations derived in our 3D-QSAR study [25]. GFA technique was used to generate QSAR models. It was observed that in each case 100,000 crossovers and smoothing factor  $d = 2.0$  resulted in optimum internal and external predictivity. Hence the number of crossovers has been set to 100,000 for all other models.



**Figure 1.** Pharmacophoric superposition of PTP 1B inhibitors used in the present QSAR study.

### 3.1.1 Significance of molecular descriptors

The Cerius2 QSAR generates different descriptors belonging to different categories like conformational, electronic, shape, spatial, thermodynamic, *etc.* Interpretation of QSAR models with more terms becomes difficult for the drug design. Moreover all the terms may not be relevant. Equations generated without restricting the number of descriptors (infinite chain length) showed good internal predictivity but have poor external predictivity.

$$\text{BA} = -2.90589 - 0.00229 \text{ Area} + 0.00105 \text{ A Pol} - 0.00237 \text{ PMI-Y} + 0.150797 \text{ H bond acceptor} - 0.02825 \text{ Dopole}_X - 0.00289 \text{ MW} - 0.0001867 \text{ Vm} + 0.395774 \text{ A log P}$$

$$\text{LOF} = 0.124, F_{\text{test}} = 33.096, r^2 = 0.708, r^2_{\text{pred}} = 0.132$$

Hence to determine more relevant descriptors, GFA equations were generated with an option that they have no more than five terms including a constant. In order to obtain stable and consistent

**Table 4.** Summary of the best equations selected from different GFA models. Equation 6 from Model C was selected to explain the observed PTP 1B inhibitory activity of benzofuran/benzothiophen biphenyls derivatives.

No.	Equation	LOF	r <sup>2</sup>	r <sup>2</sup> <sub>cv</sub>	F- value	r <sup>2</sup> <sub>pred</sub>
Model A						
1	BA = -2.37109 – 0.01869 (Dipole_X) + 0.1349 (Alog P) – 0.00064 (A pol)	0.150	0.549	0.507	35.722	0.328
2	BA = -2.1029 + 0.006549 (Vm) – 0.07565 (Rotlbonds) + 0.0471 (A log P)	0.155	0.534	0.490	33.563	0.364
Model B						
3	BA = -0.8588 - 0.015386 (Dipole_X) + 0.0151(Molref) + 0.0966 (HOMO)	0.150	0.547	0.508	35.538	0.484
4	BA = 0.31685 + 0.2600 (HOMO) + 0.007381(Vm) - 0.090578 (Dipole_X)	0.152	0.542	0.504	35.773	0.513
5	BA = -1.783611+ 0.19583(HOMO) + 0.00491(Vm) - 0.0168 (Dipole_X)	0.146	0.551	0.513	36.050	0.444
Model C						
6	BA = 3.7344 - 0.0193 (Dipole_X) + 0.2465(HOMO) - 0.007399 (DIFFV) -0.07644 (Rotlbonds)	0.118	0.694	0.658	45.960	0.672
7	BA = 1.48729 - 0.007342 (DIFFV) + 0.02055 (Dipole_X) –0.0722 (Rotlbonds) + 0.0001883 (NCOSV)	0.122	0.685	0.646	44.00	0.562
8	BA = 0.12294 - 0.018696 (Dipole_X) + 0.23116 (HOMO) + 0.01655 (MolRef) –0.00116 (Hf)	0.131	0.681	0.644	43.322	0.617
9	BA = -2.172 - 0.02066 (Dipole_X) - 0.00109 (Hf) + 0.01452 (MolRef) + 0.06182 (Alog P)	0.118	0.678	0.635	42.675	0.537

results from GFA and also to determine relevant descriptors, we have used a procedure to select a subset of descriptors, from a much large pool of descriptor. In order to explore more relevant descriptors contributing to the biological activity of PTP 1B inhibitors, GFA was run for several times by using molecular descriptors. Three models were generated using combination of different descriptors: Model A: Using 20 default descriptors (Table 3, from 1-20 descriptors); Model B: Default+ Descriptors from 21-26 in Table 3 (Three electronic and Three thermodynamic); Model C: Descriptors of Model A+ Model B+ six MSA descriptors (Table 3). We have checked the “intervariable correlation matrix” (option available within Cerius2) for the equations in all the models (Models A, B, C). This parameter was used to filter off the equations that were showing intercorrelation among the descriptors, even though those equations showed good statistical data.

All the statistically significant equations for each QSAR model have been represented in Table 4. The term BA in the equations represent biological activity expressed as  $\text{pIC}_{50}$  ( $\mu\text{M}$ ).

**Model A:** QSAR equations using GFA were generated using 20 default descriptors (Table 3). Observation of variable usage graph indicates that, dipole\_X, AlogP, Vm, and rotl.bonds contribute more significantly than all other descriptors for this model (Table 5). The resultant equations were evaluated for their predictive power. The best equation from the set of equations was selected on the basis of predictivity, LOF and other statistical parameters such as F value. Equations 1 and 2 (Table 4) showed better internal predictivity and also resulted in better predictions for test set of molecules. The variable terms in the equations show low correlation among themselves indicating low probability of chance correlation. Equation 2 with better predictive  $r^2$  value is proposed as the QSAR model with 20 default descriptors for the present series of molecules.

**Model B:** This model was built by combination of default, three thermodynamic, and three electronic descriptors. The generated set of QSAR equations were evaluated on the basis of cross validated  $r^2$ , non-cross validated  $r^2$ , LOF and frequency of variables used (Table 5) for model generation. As indicated by variable usage graph, HOMO (highest occupied molecular orbital) and dipole\_X descriptors were repeatedly used for the generation of set of equations. This resulted in the identification of three best equations 3-5 (Table 4). These three equations were analyzed for their predictive power. Equation 4 with highest external predictivity was selected as the best QSAR equation for Model B (Table 4). Addition of six descriptors to QSAR table increased the internal predictivity of the model moderately.

**Model C:** Deviation of biological activity for a series of molecules can be explained on the basis of differences in the physico-chemical descriptors. Hence, we considered using shape related descriptors in the generation of QSAR models. Addition of six MSA descriptors to QSAR table resulted in the generation of equations with all thirty-two descriptors (Table 4) for model C. These equations were analyzed on the basis of cross validate and non-cross validated  $r^2$ , LOF, F value and variable terms in the equation. Analyses of frequency of variables used in the model generation (Table 5) indicate that dipole\_X, HOMO, diff.vol. contribute more significantly than all other descriptors. The equations and variable terms in the equations clearly indicate the importance of electronic, shape and structure based factors in governing the biological activity of these compounds. Detailed statistical analysis of equations resulted in the identifications of five equations 6-10 (Table 4) for Model C. Equation 6 (Table 4) was selected as a single best equation with proper balance of statistical terms for Model C. Inclusion of MSA based parameters clearly shows the improvement in the internal and external predictivity of the model C. The internal and

**Table 5.** The frequency of use of variables for each QSAR model generation.

Variable	Usage
<b>Model A</b>	
a) Diploe_X	65
b) Rotlbonds	45
c) Vm	42
d) A log P	32
e) Area	30
f) A pol	22
g) MolWt	18
h) PMI_Z	12
i) PMI_Y	10
<b>Model B</b>	
a) Diploe_X	75
b) Vm	42
c) HOMO	35
d) Area	30
e) A pol	34
f) Rotlbonds	27
h) MolRef	29
) RadOfGyr	27
j) Alog P	24
k) Hf	15
l) PMI_X	09
<b>Model C</b>	
a) Dipole_X	95
b) HOMO	44
c) DIFFV	35
d) Rotlbonds	32
e) NCOSV	32
f) A log P	29
g) Mol.Ref	17
h) RadOfGyr	15
i) LUMO	12
j) Hf	12

external predictivity of equation 6 from Model C is more than Model A and B equations. This equation also has low LOF and higher F value than Model A and B. Therefore, this equation clearly shows the importance of shape related descriptors. Figure 2 shows the graph of actual and fitted activities of the training set molecules using Equation 6 from Model C. Figure 3 shows graph of actual and predicted activities of test set molecules using Equation 6 from Model C.

Equation 6 from Model C with proper balance of all statistical terms was selected as the best equation to explain the variance in the biological activity of benzofuran/benzothiophene biphenyls as PTP 1B enzyme inhibitors from the present QSAR analysis.

### 3.2 Randomization tests

To determine model's reliability and significance, the randomization procedure was performed at 95 % (19 trials) and 98 % (49 trials) confidence levels. The randomization was carried out by repeatedly permuting the dependent variable set. If the score of the original QSAR model proved better than those from the permuted data sets, the model would be considered statistically significant.

**Table 6.** Results of randomized  $r^2$  for Equation 6 (Model C).

Confidence Level	Trials	$r^2_{nonrandom}$	$r^2_{random}$	SD <sup>a</sup>	SD <sup>b</sup>	$r^2 <^c$	$r^2 <^d$
95 %	19	0.694	0.173	2.976	0.173	19	0
98 %	49	0.694	0.145	3.237	0.145	49	0

<sup>a</sup> Number of standard deviations of the mean value of  $r^2$  of all random trials to the non-random  $r^2$  value.

<sup>b</sup> Standard deviation of the  $r^2$  values of all random trials from the mean value of  $r^2$

<sup>c</sup> Number of  $r^2$  values from random trials that are less than the  $r^2$  value for the non-random trial

<sup>d</sup> Number of  $r^2$  values from random trials that are greater than the  $r^2$  value for the non-random trial

The results of randomization tests are shown in Table 6. The correlation coefficient  $r^2$  for the non-randomized QSAR model was 0.698, better than those obtained from randomized data. None of the permuted data sets produced an  $r^2$  comparable to 0.698; hence, the value obtained from original GFA model is significant.

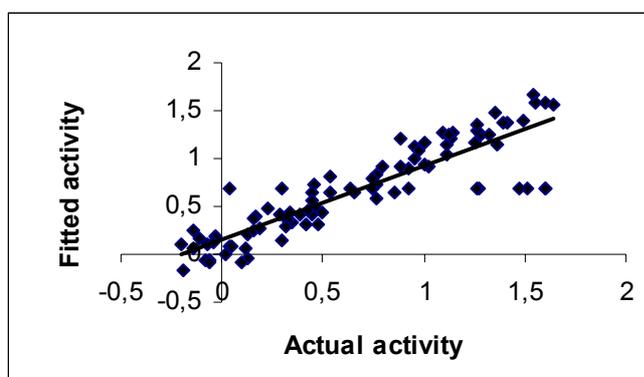


Figure 2. A graph of actual versus fitted activities of training set molecules using Equation 6 of from Model C.

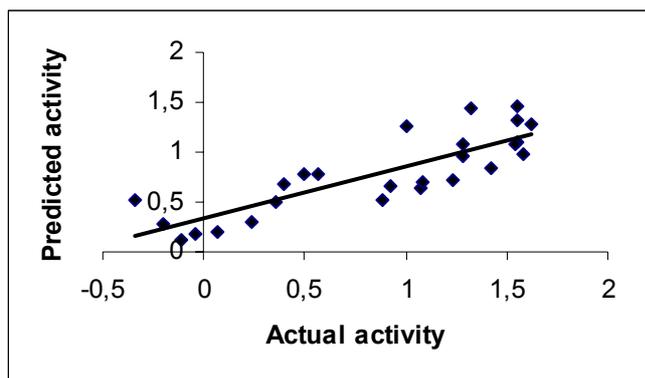


Figure 3. A graph of actual versus predicted activities of test set molecules using Equation 6 of from Model C.

### 3.3 Discussion

The Cerius2 QSAR module provides different descriptors divided into categories like spatial, structural, electronic, conformational, thermodynamic and receptor. Among those some descriptors constitute a default set. Using this default set we have obtained reasonably well-predicted model (Model A) with cross validated  $r^2$  ( $r^{2cv}$ ) of 0.507. Therefore in order to optimize the internal and external predictivity, the default descriptor set was extended in two different ways by including (a) three electronic and three thermodynamic (Model B), (b) Descriptors of Model B + six MSA descriptors (Model C); available in the Cerius2 QSAR module to generate different models using GFA. With these additions the models were greatly improved in terms of internal and external consistency.

#### 3.3.1 Interpretation of models

##### Model A

The equation describing biological activity for this model is equation 2 (Table 4) containing  $V_m$  – spatial descriptor,  $Rotl.bonds$  – structural descriptor and  $AlogP$  – thermodynamic descriptor. The spatial descriptors,  $V_m$  and structural descriptor,  $rotl.bonds$  describes the molecular volume and rigidity of the molecules respectively. These two descriptors reflect the importance of size and conformation of the molecule to exhibit PTP 1B inhibitory activity.  $AlogP$  is the partition coefficient calculated using atom based approach and represents the hydrophobicity of the molecules [26].  $AlogP$  is positively correlated with the biological activity. This property assumes significance in the present case because of the fact that the molecules under study contain lipophilic groups. This equation showed low internal as well as external predictivity. This indicates that other physicochemical parameters may be responsible for the variance in the biological activity of present set of compounds.

## **Model B**

An important observation in Model B QSAR equations was the occurrence of HOMO and dipole\_X as common descriptors in statistically significant equations 3, 4, and 5 (Table 4). Equation 4 with better predictive  $r^2$  of 0.513 was selected as the representative for Model B. The variable terms contribute for the biological activity in equation 4 from Model B include two electronic descriptors – dipole\_X, HOMO and one spatial descriptor – Vm.

Dipole\_X, an electronic parameter indicates the dipole moment in X-axis. This term was negatively correlated and indicates that the compounds having dipole moment in X-axis may show less activity. Compounds 2 and 5 (Table 1) with functional groups orienting towards X-axis showed less activity.

HOMO is an electronic parameter. When a molecule acts as an electron pair donor, electrons from its HOMO are supplied. This term indicates the importance of hydrogen bonding interactions and was positively correlated. Compound 54 and 87 (Table 1) with high HOMO energy were more active than compounds 2 and 5 with low HOMO energy (Table 1).

Vm, a spatial descriptor defines the molecular volume of ligand inside the contact surface with receptor during ligand-receptor interactions. Molecular volume is related to binding and transport. This descriptor represents the importance of size and shape of the molecule to bind tightly with enzyme during ligand-receptor interactions. Positively correlated Vm underlines the importance of essential volume of the molecules under study required to possess as that of the shape reference molecule (compound 54) to bind effectively with the receptor. Compounds 2 and 5 (Table 1) with less molecular volume compared to the shape reference compound (compound 54) showed less biological activity because of the absences of terminal hydrophobic functional groups.

## **Model C**

QSAR equations for model C were generated using thirty two descriptors (Table 3). Equation 6 (Table 4) with better internal and external predictivity was selected as the representative equation for model C. The variable terms in this equation are dipole\_X, HOMO, diff.vol and rotl.bonds.

### **3.3.2 Structure activity relationship of PTP 1B Inhibitors with representative QSAR equation**

Equation 6 (Model C, Table 4) with good internal and external predictivity was selected as representative equation to explain the variance in the biological activity of PTP 1B inhibitors from the QSAR models A, B, and C. This equation includes two electronic parameters – Dipole\_X,

HOMO, a shape parameter – diff.vol and a structural parameter – rotl.bonds contributing for the biological activity.

**Dipole moment:** Dipole moment, an electronic parameter and is important in case when dipole-dipole interactions are involved in ligand-receptor interactions. Dipole\_X describes the dipole moment along the X-axis i.e. parallel to the plane of biphenyl ring system. Thus the interaction between the electron rich functional groups of PTP 1B inhibitors and corresponding amino acid residues in the enzyme active site play an important role during enzyme inhibition. It is evident from our docking and molecular dynamics simulations on a series of PTP 1B inhibitors that, the interaction between ionizable functional group of ligands and Arg221 amino acid residue of PTP 1B active site residue is critical for enzyme inhibition [27]. It was observed during conformational analysis of compounds that oxo-acetic acid, substituted oxo-phenyllactic acid, sulfosalicylic acid functional groups oriented perpendicular to the biphenyl ring system (along Z-axis). Compounds having the aforementioned functional groups showed better inhibitor potency than compounds having phenolic hydroxyl group oriented towards X-axis. The term dipole\_X in QSAR equation was correlated negatively. This indicates the importance of ionizable functional group and its orientation in determining the activity of PTP 1B inhibitors.

The conformation obtained from random sampling method [22] used in the present QSAR study was found to be similar with the conformation used in our 3D-QSAR CoMFA (Comparative molecular field analysis) and docking studies on PTP 1B inhibitors [25]. The CoMFA fields were mapped on to the active site of PTP 1B enzyme. High level of correlation between CoMFA fields and amino acid residues of PTP 1B enzyme active site validates our choice of conformation used for CoMFA study. Hence we justify the use of dipole\_X descriptor in equation 6 (Table 4) to explain variance in the biological activity of PTP 1B inhibitors in the present study.

**HOMO:** HOMO is an electronic parameter and is the highest energy level in the molecule that contains electrons. It is crucially important in governing the molecular reactivity and properties. When a molecule acts as an electron pair donor in bond formation, the electrons are supplied from the molecule's HOMO. In the present series of molecules compounds having oxo-acetic acid, substituted phenyllactic acid, and sulfosalicylic acid functional groups with high HOMO energy (compounds 7, 54, 87 - Table 1) are more active than compounds bearing phenolic hydroxyl group (compounds 2, 5 - Table 1) with low HOMO energy. Hence HOMO descriptor denotes nucleophilicity of the molecule and this term was correlated positively. This indicates that hydrogen bonding interactions between the terminal ionizable functional group on biphenyl ring system in the present series of molecules and amino acid residues of PTP 1B enzyme active site

are crucial in the ligand-receptor interactions. It is evident from the literature that high-density basic residues surround the active site cleft of PTP 1B enzyme [28]. Hydrogen bonding interactions between ionizable functional groups of inhibitor molecules and signature motif amino acid residues (Cys215-Arg221) of PTP 1B enzyme serve as key recognition elements in the ligand-receptor interactions [29]. These evidences support the importance of positively correlated HOMO descriptor in the QSAR equation in governing the potency of PTP 1B inhibitors.

**Diff. Vol:** A shape descriptor, Differential volume (diff.vol) represents the difference between the volume of the individual molecule and volume of the shape reference compound. This term was correlated negatively with the biological activity in the present QSAR analysis. The sulphonyl biphenyl compounds (83-87, Table 1) having approximately same volume compared to the shape reference compound showed good PTP 1B enzyme inhibition. The 2-substituted-benzofuran naphthalene compounds (73-82, Table 1) with less volume compared to the shape reference compound showed less PTP 1B inhibitory activity. This descriptor indicates the importance of 2-substituted benzofuran/benzothiophene biphenyl ring system to effectively occupy the available space in the active site of PTP 1B enzyme [30].

**Rotl. Bonds:** Rotl.bonds, a structural descriptor correlated negatively with biological activity. This term indicates that conformational rigidity of the molecules is important for the activity. This is evident from better inhibitory activity of compounds with rigid 2-substituted benzofuran/benzothiophene biphenyls ring system compared to the compounds with spacer group between heterocycle and the biphenyls ring system. Compounds 63-66 (Table 1) with spacer groups  $-\text{CH}_2$ ,  $-\text{CH}(\text{OH})$  in training set and compound 17 (Table 2) with  $-\text{C}=\text{O}$  spacer group in test set between benzofuran heterocycle and biphenyls ring system showed less activity than compounds without spacer group (Compounds 7, 54, 87, Table 1). Similarly compounds 73-81 in training set (Table 1) with spacer groups  $-\text{CH}_2$ ,  $-\text{CH}(\text{OH})$ ,  $-\text{C}=\text{O}$  and compounds 20-21 with  $-\text{CH}_2$ ,  $-\text{CH}(\text{OH})$  spacer groups in test set (Table 2) between 2-butyl benzofuran heterocycle and naphthalene ring showed less activity compared to compounds without spacer groups (Compounds 7, 54, 87, Table 1). Hence conformational rigidity is important for the present series of molecules to exhibit better PTP 1B enzyme inhibition.

#### 4. CONCLUSIONS

In conclusion, 3D-QSAR analysis on a series of benzofuran/benzothiophene biphenyls with PTP 1B inhibitory activity expressed as  $\text{pIC}_{50}$  ( $\mu\text{M}$ ) against human recombinant PTP 1B enzyme was performed using robust statistical technique GFA, coupled with the use of combination of different classes of descriptors. The generated equations in each QSAR model were analyzed for

their statistical significance and predictive ability by using test set of 26 molecules that were not used in model generation. Randomization test and intervariable correlations matrix were used to check the possibility of “chance correlation” for the generated equations. GFA handled the physico-chemical descriptors effectively in the generation of QSAR models with significant statistical terms including external predictivity. Equation 6 from model C was selected as representative equation to explain the variance in the biological activity for present series of PTP 1B inhibitors. This equation explains about 70% ( $r^2 = 0.694$ ) variance in the biological activity. The variables in the equation reveal that electronic, spatial and structural descriptors contribute significantly for the biological activity of PTP 1B inhibitors. Two electronic descriptors dipole\_X, HOMO underlines the importance of electron rich functional groups (oxo-acetic acid, substituted oxo-phenyllactic acid, sulfosalicylic acid) and their orientation on biphenyls ring system for PTP 1B enzyme inhibition. The spatial descriptor diff.vol indicates essential volume of the inhibitors required to show better PTP 1B inhibitory activity. This descriptor underlines the importance of the benzofuran/benzothiophene biphenyl ring system for the effective binding of inhibitors with PTP 1B enzyme. The structural descriptor rotl.bonds indicates the importance of conformational rigidity of the compounds required for enzyme inhibition. This QSAR equation agrees with the structure activity relationship of present series of PTP 1B inhibitors. The results from the present QSAR analysis are presently being used for the design of newer compounds with better PTP 1B inhibitory activity.

## Acknowledgements

The authors thank University Grants Commission (UGC), New Delhi for financial support under its Department of Special Assistance and COSIST programs. The authors also thank Department of Science and Technology, Government of India for Financial support.

## 5. REFERENCES

- [1] P. G. Darke, and B. I. Posner, Insulin Receptor-Associated Protein Tyrosine Phosphatase(s): Role in Insulin Action. *Mol. Cell. Biochem.* **1998**, *182*, 79-89.
- [2] M. F. White, and C. R. Khan, The Insulin Signaling System. *J. Biol. Chem.* **1994**, *269*, 1-4.
- [3] J. C. H Byon, and J. Kusari, Protein-Tyrosine Phosphatase-1B Acts as a Negative Regulator of Insulin Signal Transduction. *Mol. Cell. Biochem.* **1998**, *182*, 101-108.
- [4] F. Ahamd, P.-M. Li, J. Meyerovitch, B. J. Goldstein, J. Chrenoff, T. A. Gustafson, and J. Kusari, Osmotic Loading of Neutralizing Antibodies Demonstrates a Role for Protein-Tyrosine Phosphatase 1B in Negative Regulation of the Insulin Action Pathway. *J. Biol. Chem.* **1995**, *270*, 20503-20508.
- [5] D. Bandyopadhyay, A. Kusari, K. A. Kerner, F. Liu, J. Chernoff, T. Gustafson, and J. Kusari, 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*, 1639-1645.
- [6] J. Kusari, K. A. Kenner, K.-I. Shu, D. E. Hill, and R. R. Henry, Skeletal Muscle Protein Tyrosine Phosphatase Activity and Tyrosine Phosphatase 1B Protein Content are Associated with Insulin Action and Resistance. *J. Clin. Invest.* **1994**, *93*, 1156-1162.

- [7] F. Ahmad R. V. Lonsidane, T. L. Bauer, J. P. Ohannesian C. C. Marco, and B. J. Goldstein, Improved Sensitivity to Insulin in Obese Subjects Following Weight Loss is Accompanied by Reduced Protein-Tyrosine Phosphatases in Adipose Tissue. *Metabolism* **1997**, *46*, 1140-1145.
- [8] M. Elchelby, P. Payette, E. Michalisyn, W. Cromlish, C. C. Collinschan, C. Ramachandran, M. J. Gresser, M. L. Tremblay, B. P. Kennedy, Structure of protein tyrosine phosphatase 1B in complex with inhibitors bearing two phosphotyrosine mimetics. *Science* **1999**, *283*, 1544-1548.
- [9] M. S. Malamas, J. Sredy, C. Moxham, A. Katz, W. Xu, R. M. Devitt, F. O. Adibayo, D. R. Sawicki, L. Seestaller, D. Sullivan, and J. R. Taylor, Novel Benzofuran and Benzothiophene Biphenyls as Inhibitors of Protein Tyrosine Phosphatase 1B with Antihyperglycemic properties. *J. Med. Chem.* **2000**, *43*, 1293-1310.
- [10] J. W. Mcfarland, C. M. Berger, S. A. Froshauer, S. J. Hayashi, B. H. Hecker, M. R. Jaynes, B. J. Jefson Kamicker, C. A. Lipinski, K. M. Lundy, C. P. Reese, and C. B. Vu, Quantitative Structure-Activity Relationships Among Macrolide Antibacterial Agents: in vitro and in vivo Potency against *Pasteurella Multocida*. *J. Med. Chem.* **1997**, *40*, 1340-1346.
- [11] M. Karelson, V. S. Lobanov, and A. R. Katarizky, Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *90*, 1027-1040.
- [12] A. J. Hopfinger, B. J. Burke, (1999) *In concepts and applications of molecular similarity*, (John Wiley, New York).
- [13] Cerius2 Version 3.5 is available from Molecular Simulations Inc., 9685, Onton Road, San Diego, CA, 92121, USA.
- [14] V. M. Gokhlae, and V. M. Kulkarni, Understanding the Antifungal Activity of Terbinafine Analogues Using Quantitative Structure-Activity Relationship (QSAR) Models. *Bioorg. Med. Chem.* **2000**, *8*, 2487-2499.
- [15] R. G. Karki, V. M. Kulkarni, Three-dimensional Quantitative Structure-Activity Relationship (3D-QSAR) of 3-Aryloxazolidin-2-one Antibacterials. *Bioorg. Med. Chem.* **2001**, *9*, 3153- 3160.
- [16] P. S. Kharkar, B. Desai, B. Varu, R. Loriya, Y. Naliyapara, H. Gaveria, A. Shah, and V. M. Kulkarni. Three Dimensional Quantitative Structure Activity relationship of 1,4-Dihydropyridines as Antitubercular Agents, *J. Med. Chem.* **2002**, *45*, 4858-4867.
- [17] R. V. Anand, and V. M. Kulkarni, 3D-QSAR of Cyclooxygenase-2 Inhibitors by Genetic Functions Approximation. *Internet Electron. J. Mol. Des.* **2003**, *2*, 242-261.
- [18] D. Rogers, and A. J. Hopfinger, Application of Genetic Functional Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *4*, 854-866.
- [19] A. K. Rappe, and W. A. Goddard, III Charge equilibration for Molecular Dynamics Simulations, *J. Phys. Chem.* **1991**, *95*, 3358-3363.
- [20] A. K. Rappe, C. J. Casewit, K. S. Colwell, G. A. Goddard, and Skiff W. M. Application of a Universal Force Field to Organic Molecules. *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.
- [21] A. J. Hopfinger, A QSAR Investigation of DHFR Inhibition by Baker Triazines Based Upon Molecular Shape Analysis. *J. Am. Chem. Soc.* **1980**, *102*, 7196-7206.
- [22] D. Mayer, C. Naylor, I. Motoc, G. Marshall. A Unique Geometry of the Active Site of Angiotensin Converting Enzyme Consistent with Structure-Activity Studies. *J. Comput.- Aided Mol. Design.* 1987, *1*, 3-16.
- [23] C. L. Waller, T. I. Opera, A. Giollitti, and G. R. Marshal, Three-Dimensional QSAR of Human Immuno Deficiency Virus (1) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally Determined Alignment Rules, *J. Med. Chem.* **1993**, *36*, 4152-4160.
- [24] R. D. III. Crammer, J. D. Bunce, and D. E. Patterson, Cross-validation, Boot strapping and partial Least squares compared with multiple regressions in conventional QSAR studies, *Quant. Struct. Act. Relat.* **1998**, *7*, 18-25.
- [25] V. S. Murthy, and V. M. Kulkarni, 3D-QSAR CoMFA and CoMSIA on protein tyrosine phosphatase 1B inhibitors. *Bioorg. Med. Chem.* **2002**, *10*, 2267-2282.
- [26] V. M. Viswanathan, A. K. Goshe, G. R. Revanker, and R. K. Robbins, Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163-172.
- [27] V. S. Murthy, and V. M. Kulkarni, Molecular Modeling of Protein Tyrosine Phosphatase 1B (PTP 1B) Inhibitors. *Bioorg. Med. Chem.* **2002**, *10*, 897-906.
- [28] M. Taing, F.-Y. Keng, K. Shen, L. Wu, D. S. Lawrence, and Z.-Y. Zhang. Potent and Highly Selective Inhibitors of the Protein Tyrosine Phosphatase 1B. *Biochemistry*, **1999**, *38*, 3793-3803.
- [29] R. M. Groves, Z.-Y. Yao, P. R. Roller, T. R. Jr. Burke, and D. Bradford. Structural Basis for Inhibition of the Protein Tyrosine Phosphatase 1B by Phosphotyrosine Peptide Mimetics. *Biochemistry* **1998**, *37*, 17773-17783.
- [30] M. Sarmiento, Y. A. Puius, S. W. Vetter, F.-Y. Keng, L. Wu, Y. Zhao, D. S. Lawrence, S. C. Almo, and Z.-Y. Zhang. Structural Basis of Plasticity in Protein Tyrosine Phosphatase 1B Substrate Recognition. *Biochemistry* **2000**, *39*, 8171-8179.

## Biographies

**Vithal M. Kulkarni** is Professor of Medicinal Chemistry and Head of Pharmaceutical Technology and Pharmacy Division, Mumbai University, Institute of Chemical Technology. He was postdoctoral visiting fellow in T. I. F. R. (1974-1976) Mumbai and at Brostel Research Institute, Germany (1979-1981). His areas of research specialization include (1) Chemotherapy of tropical, viral and fungal diseases (2) Computer Aided Drug Design (3) Synthesis of novel drugs and intermediates. He has presented / published numerous papers in reputed national and international conferences and journals. He is research investigator and consultant for several pharmaceutical industries. He is recipient of Prof. (late) M. L. Khorana Memorial Lecture Award, Eminent Pharmacist Award of Indian Pharmaceutical Association and Best Teacher Award of University of Mumbai.

**Sree M. Vadlamudi** is a Computational Chemist at Topotarget UK Ltd.. He pursued doctoral studies under the guidance of Prof. V. M. Kulkarni at Institute of Chemical Technology, University of Mumbai. His research interests are (1) Computer Aided Drug Design (2) *In Silico* ADMET studies. He has presented / published research papers in national and international conferences and journals.