3D-QSAR CoMFA and CoMSIA on Protein Tyrosine Phosphatase 1B Inhibitors

V. Sreenivasa Murthy and Vithal M. Kulkarni*

Pharmaceutical Technology and Pharmacy Division, Institute of Chemical Technology, University of Mumbai, Matunga, Mumbai 400 019, India

Abstract—3D-QSAR and molecular modeling was performed on a series of benzofuran/benzothiophene biphenyls as protein tyrosine phosphatase 1B (PTP 1B) inhibitors with anti-hyperglycemic activity. Evaluation of 92 compounds served to establish the model, which was validated by evaluation of an external set of 26 compounds. The lowest energy conformer of most active compound (compound **54**) obtained from simulated annealing was used as a template structure for the alignment. The best predictions were obtained with the CoMFA model from RMS fit and A log P as additional descriptor ($r_{cv}^2 = 0.615$, $r^2 = 0.842$), and with the CoMSIA combined steric, electrostatic, and lipophilic fields ($r_{cv}^2 = 0.597$, $r^2 = 0.910$). The 3D-QSAR model was then superimposed to the PTP 1B active site, giving direct contour maps of the different fields. Further comparison of the contour maps from the 3D-QSAR showed high level of compatibility with the active site of PTP 1B enzyme. © 2002 Elsevier Science Ltd. All rights reserved.

Introduction

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

Protein tyrosine phosphatase 1B (PTP 1B), a cytosolic PTP play a major role in 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 tissues, 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 exhibit increased sensitivity toward insulin resistance and are resistant to obesity.[8] All these results establish a direct role for PTP 1B in down regulating the insulin functions. Hence potent, orally active, and selective PTP 1B inhibitors could be potential pharmacological agents for the treatment of obesity and NIDDM.

The dephosphorylation mechanism of tyrosine residue by PTPs involves the formation of a covalent thiophosphate intermediate by the invariant cysteine reside in PTP signature motif. The invariant arginine residue plays a role in substrate binding and transition state stabilization. The invariant aspertic acid residue on the surface loop may act as general acid in the first step and general base in the second step of dephosphorylation by PTPs.[9] The majority of PTP 1B inhibitors are compounds with non-hydrolyzable phosphotyrosine mimetic incorporated into a peptide scaffold or aromatic functionality.[10–12] Structure–activity data generated so far suggest that the potency of PTP 1B inhibitors is related to the extent of non-covalent interactions rather than its vulnerability to undergo hydrolysis by PTP 1B enzyme.[13–15] Therefore, the reactivity of the inhibitors with PTP 1B to mimic the phosphotyrosine alone can not explain the efficiency of the inhibitors. The intermolecular non-covalent interactions of the non-reacting parts of the inhibitor and enzyme may contribute toward the activity of the inhibitors.

*Corresponding author. Tel.: +91-22-414 5616; fax: +91-22-414 5614; e-mail: vithal@biogate.com or vmk@pharma.udct.ernet.in

In the present paper, a quantitative SAR (QSAR) study by molecular field analysis of benzofuran/benzothiophene biphenyls as protein tyrosine phosphatase 1B (PTP 1B) inhibitors with anti-hyperglycemic activity[16] has been carried out. Both CoMFA (comparative molecular field analysis),[17–20] and CoMSIA (comparative molecular similarity analysis)[21,22] techniques were employed. The CoMSIA technique is of particular interest since it includes a lipophilic field and two hydrogen bond fields and is less alignment-sensitive than CoMFA.[23] The predictiveness of each of our optimized models was evaluated using a test set of 26 compounds that were not included in the model.

The contour maps derived from both the CoMFA and CoMSIA 3D-QSAR models permitted an understanding of the steric, electrostatic, and lipophilic requirements for ligand binding. As a consequence, the structural variations in the training set that give rise to variations in the molecular fields at particular regions of the space are correlated to the biological properties. The comparison of 3D-QSAR models with PTP 1B enzyme active site provided information, which can be used in the development of new PTP 1B inhibitors with anti-hyperglycemic activity.

Results

The CoMFA and CoMSIA techniques were used to derive a 3D-QSAR model for benzofuran/benzothiophene biphenyls as PTP 1B inhibitors with anti-hyperglycemic activity. The in vitro inhibitory activity pIC_{50} (µM) was used as dependent variable. Conformation of the molecules used in the study was obtained by simulated annealing. Relative alignment of the molecules was then carried out by using three techniques namely RMS fitting (atom-based), multi fit (flexible fitting), and SYBYL QSAR rigid body field fit. CoMFA models were generated with column filtering value (σ min) 2.0 kcal/mol (Table 1). The 3D-QSAR models were validated using a test set of 26 molecules (Table 2). External predictions were used to select best model. Results form the PLS analysis are reported in Tables 3–5. To ascertain the true predictivity of the model a harder test using leave-half-out method of cross-validation was performed for all the analyses. The analyses were performed 100 times and the mean and standard deviation of the r^2 values are reported in the Table 6. Negative values of cross-validated r^2 in the randomized biological activity test (Table 6) revealed that the results were not based on chance correlation.

Analysis A (Table 3) shows results obtained form the three different alignments. The atom-based alignment shows a cross-validated r^2 of 0.498 with first three components. A non-cross-validated r^2 of 0.724, F vale of 76.99 and boot-strapped r^2 of 0.753 was observed with this model. The steric and electrostatic contributions were 47.4 and 52.6%, respectively. The atom-based alignment model exhibited good external predictions with r^2_{pred} of 0.513.

CoMFA model generated from multi fit (Table 3) showed cross-validated r^2 of 0.470 with three components, non-cross-validated r^2 of 0.698, F value of 67.933, boot-strapped r^2 of 0.713, and predictive r^2 of 0.395. The steric and electrostatic contributions of this model were 47 and 53%, respectively. Realignment of the molecules by field fit with respect to the fields of template molecule (compound **54**) yielded a cross-validated r^2 of 0.488 with two components, non-cross-validated r^2 of 0.637, F value of 74.726, and boot-strapped r^2 of 0.513. Steric and electrostatic contributions for this model were 46.9 and 53.1% respectively. However, this model exhibited poor predictive r^2 of 0.163 (Figs 1–5).
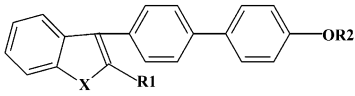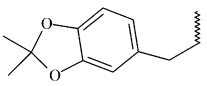
Based on the predictive ability of the three CoMFA models from analysis A (Table 3), the model generated with atom-based alignment with good predictive ability r^2_{pred} of 0.513 was selected to carry out further analyses.
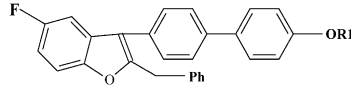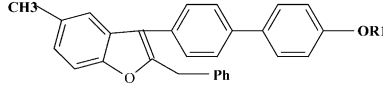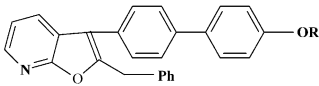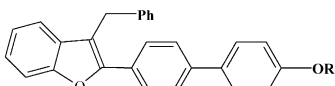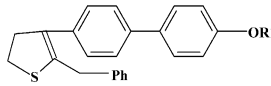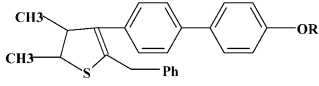
The activity data used in the present study is in vitro data and such a type of activity data could have contributions not only from steric and electrostatic fields but also other physico-chemical properties. These properties account for the transport phenomenon and pharmacokinetic profile of the molecules. To account for the lipophilic properties, which may influences this type of activity, A log P calculated by using Cerius2 software was included in the analysis. As the CoMFA model generated with atom-based alignment showed good statistical results we have decided to add additional descriptors to this model. Inclusion of A log P contributed significantly to the CoMFA model. Table 4 depicts the results obtained from the CoMFA model after inclusion of additional descriptors.

Analysis B (Table 4) showed that inclusion of A log P has improved the r^2_{cv} from 0.498 to 0.615, non-cross-validated r^2 from 0.724 to 0.842, F value from 76.99 to 79.449, r^2_{pred} from 0.513 to 0.752. A high boot-strapped (100 runs) r^2 of 0.878 ± 0.026 adds a high confidence limit to this analysis. In this analysis the contribution of steric, electrostatic, A log P are 34.8, 43.1, 22.1%, respectively. The analysis predicted the activity of the molecules of the test set with predictive r^2 of 0.752. This model showed cross-validated r^2 of 0.580 by leave-half-out method (100 runs) and randomized cross-validated r^2 of −0.218 (Table 6) indicating the statistical significance of the model. A graph depicting actual versus fitted activities of the training set is shown in shown in Figure 6. Figure 7 displays a plot of actual verses predicted activities of test set molecules.

We have also calculated various descriptors using Cerius2 software and used in analysis B. These descriptors include electronic descriptors (superdelocalizability, dipole moment, atomic polarizabilities, HOMO and LUMO), spatial descriptors (molecular surface area, molecular volume, molecular density), structural descriptors (molecular weight, number of rotatable bonds) and thermodynamic descriptors (desolvation free energy, molecular refractivity). These descriptors are calculated and directly used as additional descriptors in the PLS table.

Table 1. Structures and biological activities of the training set of 92 molecule used in the present study

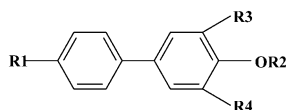| Compound	R1	R2	X	Biological activity[a]
1	Butyl	H	O	0.130
2	Benzyl	H	O	0.0362
3	Benzoyl	H	OS	0.130
4	Butyl	H	S	0.154
5	4-OH benzyl	H	O	−0.033
6	Benzyl	$CH(CH_2Ph)COOH$ (R)	O	0.455
7	Benzyl	$CH(CH_2CH_2Ph)COOCH_3$(S)	O	0.657
8	Benzyl	$CH(CH_2CH_2$-*N*-pthalinimide)–COOH(S)	O	0.468
9	Benzyl	$CCH_3(CH_2Ph)COOH$(R)	O	0.537
10	Benzyl	$CH(CH_3)COOH$(R)	O	0.120
11	Benzoyl	$CH(CH_2Ph)COOH$(R)	O	0.167
12	CH(OH)phenyl	$CH(CH_2Ph)COOH$(R)	O	0.958
13	Benzyl	CH_2Ph-4-COOH	O	0.443
14	Butyl	$CH(CH_2Ph)COOH$(R)	S	0.769
15	Benzyl	$CH(CH_2Ph)COOH$(R)	S	1.022
16	Butyl	$CH(Ph)COOH$ (R)		0.958
17	4-F-benzyl	$CH(CH_2Ph)COOH$ (R)	S	0.920
18	4-OCH_3-benzyl	$CH(CH_2Ph)COOH$(R)	S	1.113
19	2,4-di-OH-benzyl	$CH(CH_2Ph)COOH$ (R)	S	0.920
20		$CH(CH_2Ph)COOH$(R)	S	1.113
21	2-Methyl thiazolo	$CH(CH_2Ph)COOH$(R)	S	−0.064
22	2-Methyl pyridyl	$CH(CH_2Ph)COOH$(R)	S	−0.190

Compound	R	R1	Biological activity[a]
23		$CH(CH_2Ph)COOH$(S)	0.886
24		$CH(CH_2Ph)COOH$(R)	0.387
25		$CH(CH_2Ph)COOH$(S)	0.229
26		$CH(CH_2Ph)COOH$(R)	0.455
27		$CH(CH_2Ph)COOH$(R)	0.0132
28		$CH(CH_2Ph)COOH$(R)	0.292
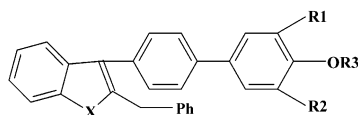
Compound	R1	R2	R3	X	Biological activity[a]
29	Br	H	H	S	−0.029
30	Br	Br	H	S	0.346
31	I	I	H	S	0.283
32	Br	Br	CH(CH$_2$Ph)COOH(R)	S	1.602
33	4-OCH$_3$–Ph	H	CH(CH$_2$Ph)COOH(R)	S	1.275
34	Br	Br	CH(CH$_2$CH$_2$Ph)COOH(S)	S	0.537
35	Br	Br	CH(CH$_2$CH$_2$–*N*-pthalinimide)–COOH	S	1.356
36	Br	Br	CH(CH$_2$CH$_2$NHCOPh-2-COOH)COOH	S	0.744
37	Br	Br	CH(CH$_2$CH$_2$NHCOPh–2-COOH)COOCH$_3$	S	1.267
38	Br	H	CH$_2$COOH	S	0.443
39	Br	Br	CH$_2$COOH	S	1
40	4-OCH$_3$–Ph	H	CH$_2$COOH	S	1.096
41	4-OC$_2$H$_5$–Ph	H	CH$_2$COOH	S	1.283
42	2,3-di-OCH$_3$–Ph	H	CH$_2$COOH	S	1.148
43	3,4,5-tri-OCH$_3$–Ph	H	CH$_2$COOH	S	1
44	4-OCH$_3$–Ph	Br	CH$_2$COOH	S	1.537
45	2,4-di-OCH$_3$–Ph	Br	CH$_2$COOH	S	1.327
46	3-OCH$_3$–Ph	3-OCH$_3$–Ph	CH$_2$COOH	S	1.602
47	4-OCH$_3$–Ph	4-OCH$_3$–Ph	CH$_2$COOH	S	1.602
48	Br	H	CH$_2$CH$_2$CH$_2$COOH	S	0.769
49	Br	H	CH(CH$_2$PhCOOH(S))	O	1.251
50	4-OCH$_3$–Ph	H	CH(CH$_2$Ph)COOH(S)	O	1.366
51	NO$_2$	H	CH(CH$_2$Ph)COOH(R)	O	0.638
52	Br	Br	CH(C$_2$H$_5$)COOH(S)	O	0.886
53	Br	Br	CH[CH$_2$CH(CH$_3$)$_2$]COOH(R)	O	1.267
54	Br	Br	CH[CH$_2$)$_5$CH$_3$]COOH	O	1.638
55	CH$_3$	CH$_3$	CH(CH$_2$Ph)COOH(R)	O	1.130
56	Cyclopentyl	H	CH(CH$_2$Ph)COOH(S)	O	1.259
57	Cyclopentyl	H	CH$_2$COOH	O	0.769
58	NHCH$_2$CH$_2$COOH	H	CH$_2$CH$_2$Ph	O	0.853
59	NHCOCH$_2$CH$_2$COOH	H	H	O	0.036
60	NHCOCH=CHCOOH	H	H	O	0.337
61	NHCO–C$_6$H$_4$–2-COOH	H	H	O	0.795

Compound	R1	R2	R3	R4	Biological activity[a]
62		CH$_2$COOH	4-OCH$_3$–Ph	4-OCH$_3$–Ph	1.508

[2-butyl benzofuran biphenyls]
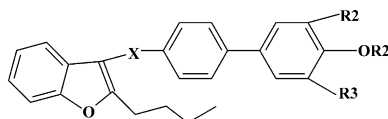
Compound	R1	R2	R3	X	Biological activity[a]
63	H	H	H	CH$_2$	−0.0755
64	H	Br	Br	CH(OH)	−0.146
65	CH$_2$COOH	H	H	CH$_2$	−0.060
66	CH$_2$–tetrazole	H	H	CH$_2$	0.292

[substituted oxazole biphenyls]
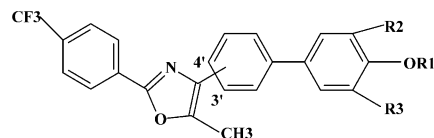
Compound	R1	R2	R3	P.O.A.[b]	Biological activity[a]
67	CH_2COOH	H	H	4′	0.096
68	$CH(CH_2Ph)COOH$	H	H	4′	−0.113
69	CH_2-tetrazole	H	H	4′	0.045
70	$CH(CH_2Ph)COOH$	H	H	3′	−0.204
71	H	Br	Br	4′	0.187
72	CH_2COOH	Br	Br	4′	0.327

[2-butyl benzofuran napthnlenes]
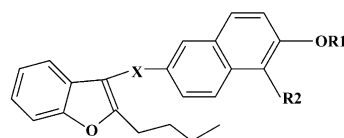
Compound	R1	R2	X	Biological activity[a]
73	H	H	CH(OH)	−0.041
74	H	Br	CH(OH)	0.318
75	H	Br	CH_2	0.481
76	H	I	CH_2	0.420
77	CH_2COOH	Br	CH_2	−0.146
78	$CH(CH_2Ph)COOH$	Br	CH_2	0.431
79	$CH(CH_2Ph)COOH$	Br	CO	−0.079
80	$CH(CH_2Ph)COOH$	I	CH_2	0.494
81	CH_2–tetrazole	Br	CH_2	0.154
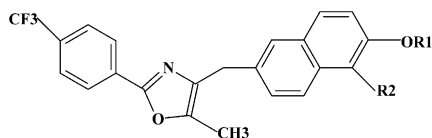
Compound	R1	R2	Biological activity[a]
82	CH_2COOH	Br	−0.113

[sulpho biphenyls]
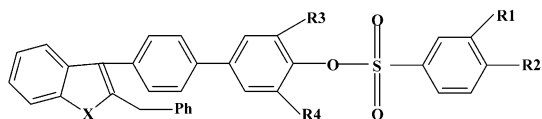
Compound	R1	R2	R3	R4	X	Biological activity[a]
83	H	COOH	H	H	O	1.124
84	COOH	H	H	H	O	0.974
85	OH	COOH	H	H	O	1.408
86	OH	COOH	CH_3	CH_3	O	1.468
87	OH	COOH	H	H	S	1.552
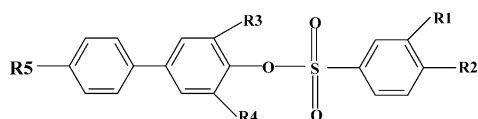
Compd	R1	R2	R3	R4	R5	Biological activity[a]
88	OH	COOH	H	H		1.494
89	OH	COOH	Cyclopentyl	H		1.397
90	OH	COOH	H	H		0.450
91	OAc	COOH	H	H		−0.064
92	OH	COOH	NO$_2$	H		0.749

[a]Biological activity is expressed as log 1/ IC $_{50}$ against human recombinant PTP 1B enzyme (h-PTP 1B) in µM.
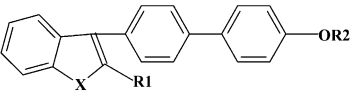[b]P.O.A., point of attachment.

Inclusion of these descriptors did not improve the predictivity of the model. Inclusion of HOMO descriptor improved r_{cv}^2 from 0.498 to 0.503, with five components and r_{pred}^2 from 0.513 to 0.537. CoMFA model was generated by inclusion of A log P and HOMO as additional descriptors to check the individual contribution of these descriptors to the CoMFA model. This model exhibited cross-validated r^2 of 0.613 with four components, non-cross-validated r^2 of 0.732, F value of 60.371 with predictive r^2 of 0.663. The contributions of steric, electrostatic, A log P, HOMO are 34.7, 42.6, 22.1, 0.6%, respectively.

As the HOMO descriptor contribution was very less compared to A log P, we have selected CoMFA model with atom based alignment and A log P as additional descriptor (Analysis B, Table 4) with good internal as well as external predictivity. Hence all the CoMFA contours were analyzed using this model. The field values were calculated at each grid point as the scalar product of the associated QSAR coefficient and the standar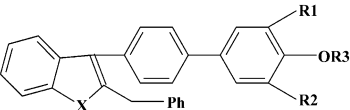d deviation of all values in the corresponding column of the data table (STDDEV*COEFF) and are plotted as percentage contribution to QSAR equation. The steric and electrostatic maps of CoMFA model with atom-based alignment and A log P as additional descriptor (Analysis B) are shown in Figures 2 and 3.

Figure 2 represents the steric contour plot. The green contours represents regions of high steric tolerance (80% contribution), while yellow contours represents regions of unfavorable steric effect (20% contribution). Sterically favorable green contours were observed around the hexyl side chain, in the vicinity of 2- substituted benzyl group, and around the phenyl ring bearing the oxo-acetic acid functional group of compound **54**. The green region is accompanied by the sterically forbidden yellow region.
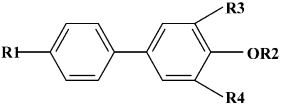
Figure 3 displays the electrostatic contour plot. The blue contours represent the regions where positively charged groups enhance the activity (80% contribution) and red region where the negatively charged groups enhance the activity (20% contribution). Electrostatic fields are observed in the vicinity of the substituted
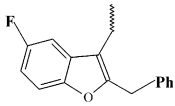
Table 2. Structures and activities for the test set of 26 molecules used in the present study

Compound	R1	R2	X	Biological activity[a]
1	2,4-di-OH-benzyl	H	S	0.236
2	Butyl	CH_2COOH	O	−0.340
3	Butyl	$CH(CH_2Ph)COOH$	O	0.356
4	Benzyl	$CH(CH_2Ph)COOH$	O	0.568
5	Benzyl	$CH(CH_2Ph)COOH(S)$	O	0.494
6	Benzyl	$CH(Ph)COOH(R)$	S	0.397
7	3,4-OCH_3-benzyl	$CH(CH_2Ph)COOH(R)$	S	0.920
8	2,4-di-OCH_3-benzyl	$CH(CH_2Ph)COOH$	S	1.070
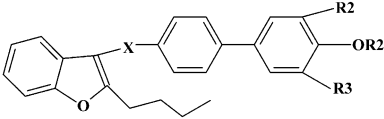
Compound	R1	R2	R3	X	Biological activity[a]
9	Br	H	$CH(CH_2Ph)COOH(R)$	S	1.236
10	4-Cl–Ph	H	$CH(CH_2Ph)COOH(R)$	S	1.283
11	Ph	H	CH_2COOH	S	1.000
12	3-OCH_3–Ph	Br	CH_2COOH	S	1.552
13	Br	Br	$CH(CH_2Ph)COOH(S)$	o	1.420
14	Br	Br	$CH[(CH_2)_3CH_3]COOH$	O	1.283
15	$NHCH_2COOH$	H	CH_2CH_2Ph	O	1.086

Compound	R1	R2	R3	R4	Biological activity[a]
16		CH_2COOH	4-OCH_3–Ph	4-OCH_3–Ph	1.318

[2-butyl benzofuran biphenyls]

Compound	R1	R2	R3	X	Biological activity[a]
17	CH_2COOH	H	H	CH(OH)	0.267

[oxaxolo biphenyls]
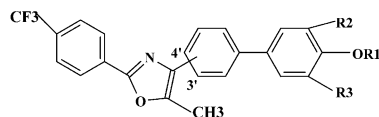
Compd	R1	R2	R3	P.O.A.[b]	Biological activity[a]
18	CH(CH$_2$Ph)COOH	H	H	3′	−0.204
19	CH(CH$_2$Ph)COOH	Br	Br	4′	0.886

[2-butyl benzofuran naphthalenes]
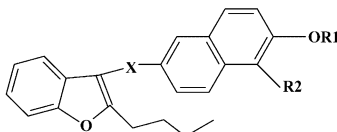
Compd	R1	R2	X	Biological activity[a]
20	H	H	CH$_2$	−0.113
21	CH$_2$-tetrazole	Br	CO	−0.415

[sulfonobiphenyls]
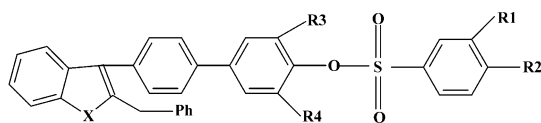
Compd	R1	R2	R3	R4	X	Biological activity[a]
22	COOH	OH	H	H	O	1.585
23	OH	COOH	NO$_2$	H	O	1.537
24	OH	COOH	Cyclopentyl	H	O	1.552
25	OH	COOH	Br	H	S	1.619
26	OH	COOH	Br	Br	S	1.522

[a]Biological activity is expressed as log $1/IC_{50}$ against human recombinant PTP 1B enzyme (h-PTP 1B) in μM.
[b]P.O.A., point of attachment.

Table 3. Summary of CoMFA results (Analysis A)

Alignment	1[a]	2[b]	3[c]
r_{cv}^2 [d]	0.498	0.470	0.488
Components	3	3	2
SEP[e]	0.394	0.404	0.318
r_{ncv}^2 [f]	0.724	0.698	0.637
SEE[g]	0.292	0.305	0.124
F value	76.99	67.93	74.726
$Pr_{=0}^2$	0	0	0
Contrib. steric	47.4	47.0	46.9
Elect.	52.6	53.0	53.1
r_{pred}^2 [h]	0.513	0.395	0.163
r_{bs}^2 [i]	0.753	0.713	0.513
Standard deviation[i]	0.023	0.044	0.147

[a]Alignment by RMS fit.
[b]Alignment by Multi fit.
[c]Alignment by Field fit.
[d]Cross-validated r^2.
[e]Standard error of prediction.
[f]Non cross-validated r^2.
[g]Standard error of estimate.
[h]Predictive r^2.
[i]From 100 bootstrapping runs.

Table 4. Summary of CoMFA results (Analysis B)

	CoMFA[a]	CoMFA[a] A log P	CoMFA[a] HOMO	CoMFA[a] A log P HOMO
r_{cv}^2 [b]	0.498	0.615	0.503	0.613
Components	3	3	5	4
SEP[c]	0.394	0.345	0.398	0.348
r_{ncv}^2 [d]	0.724	0.842	0.745	0.735
SEE[e]	0.292	0.236	0.285	0.288
F value	76.99	79.449	62.125	60.371
$P r_{=0}^2$	0	0	0	0
Contrib. Steric	47.4	34.8	47.5	34.7
Elect.	52.6	43.1	51.0	42.6
A log P	—	22.1	—	22.1
HOMO	—	—	1.5	0.6
r_{pred}^2 [f]	0.513	0.752	0.537	0.663
r_{bs}^2 [g]	0.753	0.878	0.513	0.728
Standard deviation[e]	0.023	0.026	0.044	0.052

[a]Alignment by RMS fit (Analysis A).
[b]Cross-validated r^2.
[c]Standard error of prediction.
[d]Non cross-validated r^2.
[e]Standard error of estimate.
[f]predictive r^2.
[g]From 100 bootstrapping runs.

Table 5. Summary of CoMSIA results[a]

	S	E	D	A	H	S + E	S + E + H	ALL
r_{cv}^2 [b]	0.521	0.481	0.317	0.423	0.496	0.541	0.597	0.576
Components	8	3	3	10	4	2	4	5
SEP[c]	0.411	0.416	0.477	0.453	0.412	0.390	0.375	0.380
r_{ncv}^2 [d]	0.828	0.612	0.519	0.672	0.757	0.821	0.910	0.866
SEE[e]	0.247	0.360	0.401	0.294	0.322	0.248	0.177	0.214
F value	51.471	59.833	40.986	33.033	57.833	84.707	119.395	86.949
$Pr_{=0}^2$	0	0	0	0	0	0	0	0
Contrib. S	100	—	—	—	—	42.8	37.30	14.1
E	—	100	—	—	—	57.2	35.70	17.3
D	—	—	100	—	—	—	—	29.00
A	—	—	—	100	—	—	—	21.50
H	—	—	—	—	100	—	37.30	18.1
r_{pred}^2 [f]	0.349	−4.082	−1.095	−0.565	0.269	0.412	0.648	0.562
r_{bs}^2 [g]	0.864	0.692	0.569	0.712	0.812	0.869	0.987	0.941
Standard deviation	0.005	0.004	0.023	0.061	0.074	0.069	0.008	0.019

[a]S, steric; E, electrostatic; D, hydrogen bond donor; A, hydrogen bond acceptor; H, hydrophobic; S + E, steric + electrostatic; S + E + H, steric + electrostatic + hydrophobic; ALL, steric + electrostatic + hydrogen bond donor + hydrogen bond acceptor + hydrophobic.
[b]Cross-validated r^2.
[c]Standard error of prediction.
[d]Non cross-validated r^2.
[e]Standard error of estimate.
[f]Predictive r^2.
[g]From 100 bootstrapping runs.

Table 6. Results of analyses with leave-half-out crossvalidation[a] and randomized biological activities

	r_{cv}^2 [a]					r_{cv}^2 [b]				
	CoMFA[c]	CoMFA[c] A log P	CoMFA[c] HOMO	CoMFA[c] A log P HOMO	CoMSIA[d] SEH	CoMFA[c]	CoMFA[c] A log P	CoMFA[c] HOMO	CoMFA[c] A log P HOMO	CoMSIA[d] SEH
Mean	0.46	0.580	0.496	0.472	0.486	−0.185	−0.218	−0.232	−0.170	−0.412
[e]SD	0.06	0.04	0.11	0.19	0.26	0.152	0.146	0.189	0.156	0.292
High	0.65	0.61	0.58	0.67	0.63	0.190	0.242	0.138	0.147	0.167
Low	0.25	0.43	0.21	0.29	0.37	−0.734	−0.821	−0.901	−0.663	−0.543

[a]Cross-validated r^2 by leave-half-out with optimum number of components, average of 100 runs.
[b]Cross-validated r^2 with randomized biological activity, average of 100 runs.
[c]Alignment by RMS fit (Analysis A).
[d]CoMSIA analysis by combined steric, electrostatic, and hydrophobic fields.
[e]Standard deviation.

oxo-acetic acid group and around the hexyl side chain of compound **54**. Figures 4 and 5 display CoMFA steric and electrostatic fields mapped on to the active site of PTP 1B enzyme.

By use of steric, electrostatic, hydrogen bond donor, acceptor, and lipophilicity as descriptors CoMSIA analysis was performed. The atom-based alignment used for CoMFA study served as alignment for CoMSIA. The results of CoMSIA are summarized in Table 5. In CoMSIA combination of steric, electrostatic, and lipophilic fields gave good r_{cv}^2 and r_{pred}^2 (Table 5). It was observed in CoMFA that combination of steric, electrostatic fields with A log P as additional descriptor gave good r_{cv}^2 and r_{pred}^2 (Table 4). This demonstrates that these fields are necessary to fully describe the field properties of PTP 1B inhibitors. The CoMSIA steric, electrostatic, and lipophilic fields explain variance of 37.3, 35.7, and 27.0%, respectively (Table 5). The

contribution of steric, electrostatic and A log P in CoMFA model are 34.8, 43.1, and 22.1%, respectively (Table 4). Thus CoMFA model with A log P as additional descriptor revealed the importance of lipophilicity for the present series of molecules.

Analysis of CoMSIA lipophilic contour maps (Fig. 8) indicates that lipophilic favorable yellow region was found around the hexyl side chain, benzyl group attached to benzofuran heterocycle of most active compound, compound **54**. This indicates that bulky substituents in this region are necessary to exhibit good inhibitor potency. Figure 9 displays the hydrophobic fields of compound **54** mapped on to the PTP 1B active site. A graph depicting actual versus fitted activities of the training set obtained from CoMSIA steric, electrostatic, and hydrophobic fields is shown in shown in Figure 10. Figure 11 displays a plot of actual verses predicted activities of test set molecules for the

CoMSIA analysis. The analysis predicted the activity of the molecules of the test set with predictive r^2 of 0.648. This model showed cross-validated r^2 of 0.486 by leave-half-out method (100 runs) and randomized cross-validated r^2 of -0.412 (Table 6) indicating the statistical significance.
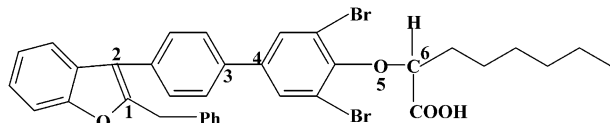

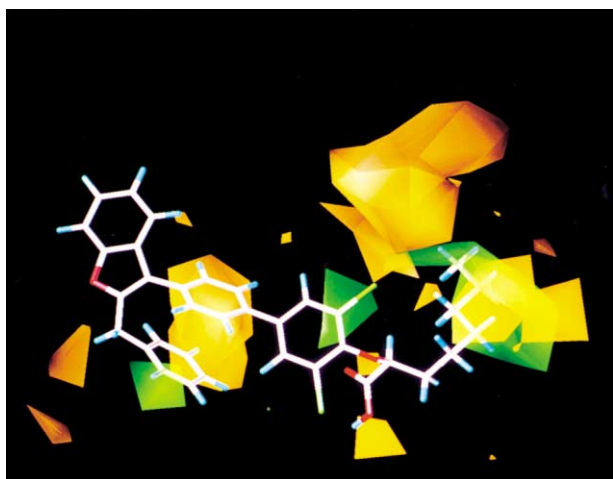**Figure 1.** Compound **54** with atoms used for superimposition are marked.

Figure 2. CoMFA steric STDE V*COEFF contour plots from atom-based alignment with Alog P as additional descriptor (Analysis B). Sterically favored areas (contribution level 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level 20%) are represented by yellow polyhedra. The active compound **54** show in capped sticks.
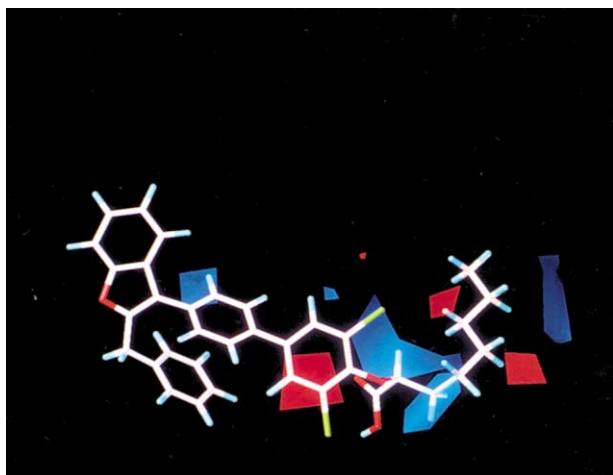

Figure 3. CoMFA electrostatic STDEV*COEFF contour plots from atom-based alignment with Alog P as additional descriptor (Analysis B). Positive charge favored areas (contribution level 80%) are represented by blue polyhedra. Negative charge favored areas (contribution level 20%) are represented by red polyhedra. The active compound **54** shown in capped sticks.

Attempts to combine CoMSIA steric, electrostatic with hydrogen bond fields (ST + EL + D + A) results in reduced r^2_{cv} where as combination of all five fields (ALL) did not lead to any significant improvement (Table 5).
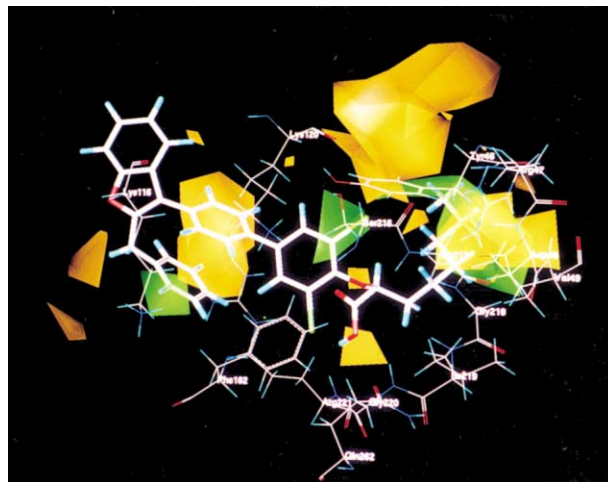

Figure 4. Superimposition of CoMFA steric fields on the active site of PTP 1B enzyme. Sterically favored (contribution level 80%), disfavored areas (contribution level 20%) are represented in green and yellow contours, respectively.
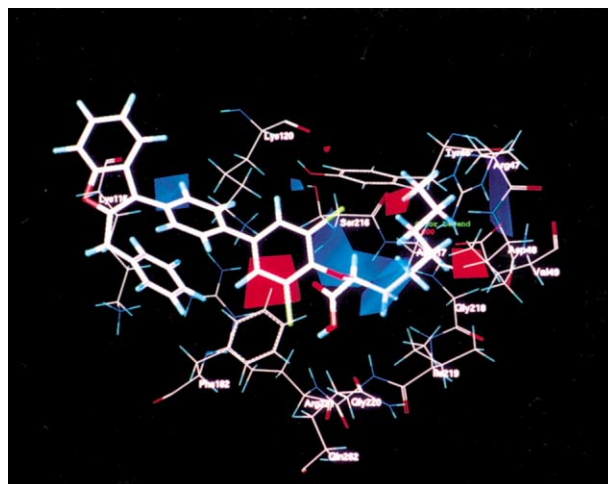

Figure 5. Superimposition of CoMFA electrostatic fields on the active site of PTP 1B enzyme. Positive charge favored (contribution level 80%), negative charge favored areas (contribution level 20%) are represented in blue and red contours, respectively.
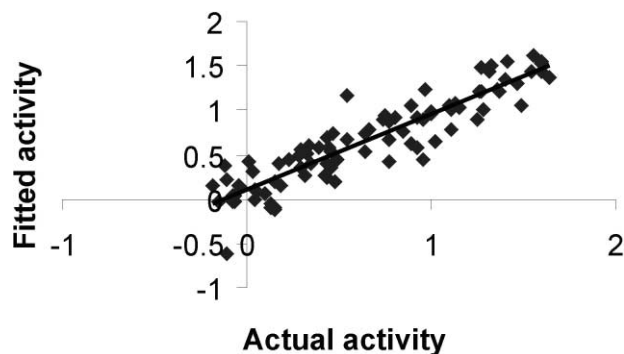

Figure 6. A graph of actual versus fitted activities of the training set molecules from atom-based alignment with A log P as additional descriptor (Analysis B).
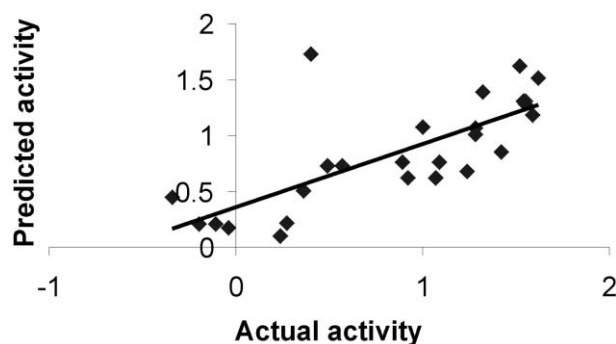
Figure 7. A graph of actual versus predicted activities of the test set molecules from atom-based alignment with A log P as additional descriptor (Analysis B).
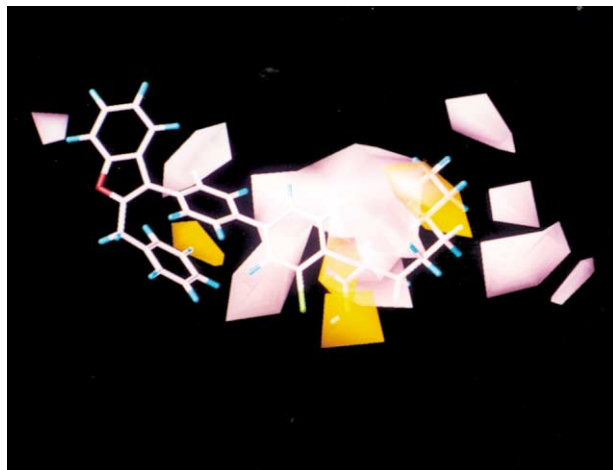

Figure 8. CoMSIA hydrophobic fields. Yellow indicates regions where hydrophobic substituents enhance activity; white indicates hydrophobic substituents decrease activity.
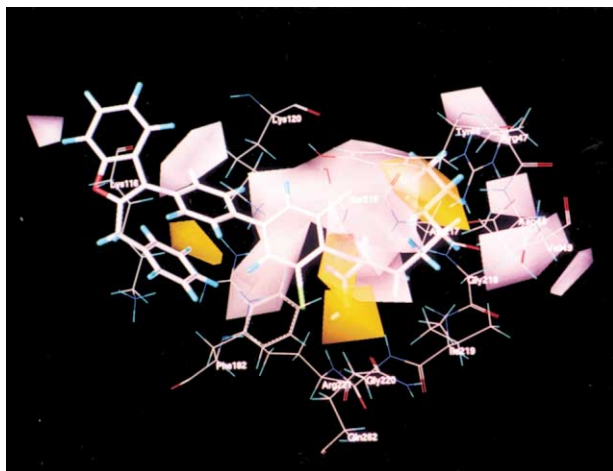

Figure 9. CoMSIA hydrophobic favorable (yellow) and unfavorable (white) superimposed on the active site of PTP 1B enzyme.

Hence, results from CoMFA and CoMSIA, suggests that variation of biological activity of inhibitors is dominated by steric and lipophilic interactions with PTP 1B active site in addition to the electrostatic interactions.
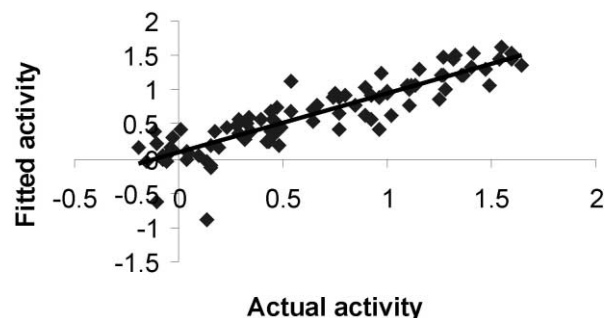

Figure 10. A graph of actual versus fitted activities of the training set molecules from CoMSIA analysis with steric, electrostatic, and hydrophobic fields.
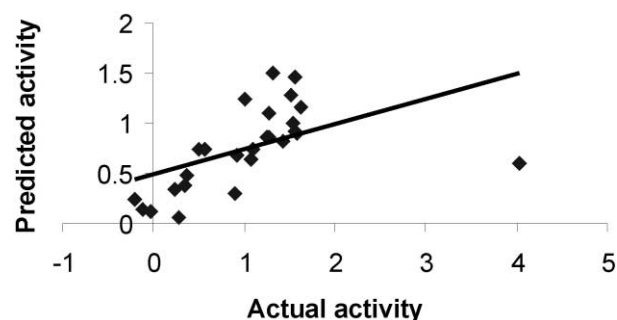

Figure 11. A graph of actual versus predicted activities of the test set molecules from CoMSIA analysis with steric, electrostatic, and hydrophobic fields.

Discussion

Three-dimensional quantitative structure-activity relationship by comparative molecular field analysis and comparative molecular similarity indices analysis was performed on a series of 118 molecules belonging to benzofuran/benzothiophene biphenyls as PTP 1B inhibitors with anti-hyperglycemic activity. The lowest energy conformer of most active compound, compound **54** obtained from simulated annealing was used as template for 3D-QSAR studies.

The alignment of the compound is one of the critical inputs for the CoMFA studies. The alignments define the putative pharmacophore for the series of ligands. In the present study we have aligned the ligands onto a template structure compound **54** using three alignment rules (Table 3). These alignments were validated using the CoMFA studies. The CoMFA models were validated by predicting the activity of the external test set.

The PLS analyses on all three alignments (atom-based, multi fit, field fit) are reported in Table 3. Alignment 1, rms fitting of atoms shows better r^2 values than other two alignments. This indicates that all ligands have to be superimposed by the atoms used for rms fitting. This superimposition produced a good external prediction. Flexible fitting of ligands using multi-fit algorithm did not improve the predictiveness of the model. Alignment by field-fit produced CoMFA models with poor external predictive ability. This shows that exact superimposition

of ligands is essential for good predictions. Cross-validation analyses was also preformed by the leave-half out technique. Final r_{cv}^2 value was calculated by taking the average of 100 runs. In some cases the average r_{cv}^2 was less than the leave-one-out r_{cv}^2 but in no case were these values negative. The boot-strapping analyses shows that models are stable and statistically robust. The mean and standard deviation values from 100 runs with randomized biological activity are shown in Table 6. These results indicate good internal consistency of the underlying data set.

Additional descriptors were added to the CoMFA table to study the influence of the other physico-chemical factors on the CoMFA results. The inclusion of A log P contributed significantly to CoMFA models with atom-based alignment (Table 4).

Log p the octanol/water coefficient related to the hydrophobicity or lipophilicity of the molecule. A log P is the atom based approach, in which each atom of the molecule is assigned to a particular class, with additive contribution to the total value of log P. Lipophilicity encodes two major contributions, namely a bulk term accounting for lipophilic and dispersive forces, and a polar term reflecting electrostatic interactions and hydrogen bonds.[37] A certain account of correlation between lipophilic and steric (or electrostatic) fields must be expected and should be properly taken into account in any CoMFA model. Thus A log P as additional descriptor yielded improved CoMFA model by atom-based alignment in statistical terms reveal the importance of lipophilicity in ligand–receptor interactions.

To gain further insights regarding ligand–receptor interactions 3D-QSAR contours were mapped onto the active site of PTP 1B enzyme and fields were analyzed with respect to the amino acid residues of PTP 1B enzyme.

Analysis B with CoMFA and A log P was used to analyze the CoMFA contour maps as this model exhibits good internal as well as external predictivity. The CoMFA steric and electrostatic contour maps of this analysis are shown in Figures 2 and 3.

The contour plots are to be considered as a representation of the lattice points, where differences in the field values are strongly associated with differences in the receptor binding affinity. The absence of lattice points does not indicate that a given substructure element has no influence on the biological activity. It is likely that all the compounds studied exert same steric and/or electrostatic influence in a certain area. Though CoMFA contour maps cannot be used as receptor maps, some useful interpretation can be derived from these contour maps.

Figure 2 represents the steric contour plot of compound **54**. The lipophilic fragment of the molecule is surrounded by the steric contours. The most active compound **54** has a hexyl side chain embedded in the

sterically favorable green region and compounds **2**, **5** (Table 1) have no functional groups to extend into this region showed less activity. Sterically favorable green regions were also observed in the vicinity of benzyl side chain, terminal phenyl ring of the biphenyl ring system bearing α-substituted oxo-acetic acid functional group. The sterically favorable green contours are embedded in the surrounding yellow contours, suggesting that there is a definite requirement of substructure with appropriate shape to exhibit high activity. The substructure fragments with less or very high steric bulk in this area reduces the activity. The steric contours in this region of the molecule underlay the importance of the steric interactions of the ligands with the receptor.

Figure 3 depicts the electrostatic contour plot of compound **54**. A positive charge favorable blue contour was found at the junction of heterocycle and biphenyl ring, and a large blue contour was found in the vicinity of α-substituted oxo-acetic acid functional group suggests that substructure with high electron density in this region reduce the activity. Negative charge favorable red contour surrounding the α-substituted oxo-acetic acid and two red contours were observed in the vicinity of hexyl side chain of compound **54**. The negative charge favorable red contour in the sterically favorable green region of hexyl side chain suggesting that substituents in this region should have a bulky substitution with electron rich group for better activity. Electrostatic contour plots indicate the importance and orientation of oxo-acetic acid functional group to exhibit PTP 1B inhibitory activity. CoMFA steric and electrostatic contour maps are compared with the active site of PTP 1B enzyme. Figures 4 and 5 shows mapped CoMFA steric and electrostatic fields on to the PTP 1B enzyme.

As combination of CoMSIA steric, electrostatic, and hydrophobic fields gave good statistical results (Table 5), this model was used to analyze CoMSIA 3D-plots. The hydrophobic field (Fig. 8) in CoMSIA brings to light the importance of hydrophobic substituents. Hydrophobic favorable yellow regions were around the hexyl side chain, in the vicinity of phenyl ring bearing α-substituted oxo-acetic acid functional group, and around the benzyl group of benzofuran heterocycle of compound **54**. Hydrophobically disfavored white regions were embedded in the favorable yellow regions. This indicates that to exhibit good PTP 1B inhibitor activity molecules should possess substructures with appropriate size and shape to occupy the hydrophobically favorable regions. In the present series of compounds, compound **54** with hexyl side chain on oxo-acetic acid functional group occupied the favorable yellow region showed highest activity, where as compounds **2** and **5** with phenolic hydroxyl group showed least activity. To gain further insight about the importance of hydrophobic interactions, CoMSIA hydrophobic 3D plots are compared with the active site of PTP 1B enzyme. Figure 9 shows the CoMSIA hydrophobic favorable (yellow), disfavored (white) fields superimposed on to the active site of PTP 1B enzyme.

Analysis of contour maps with PTP 1B active site

Because of our alignment rules are based on the conformation of the molecules obtained from simulated annealing, we compared the 3D-QSAR model of PTP 1B inhibitors with X-ray crystallographic structure of the PTP 1B enzyme (PDB code 1ecv). Mapping of CoMFA steric and electrostatic contours, CoMSIA hydrophobic contours onto the active site of PTP 1B enzyme revealed the interaction of amino acid residues with the substructures of the molecules with steric, electrostatic, and hydrophobic fields around them.

Steric interactions (Fig. 4)

Sterically favorable green polyhedra was found around the benzyl substituent at the second position benzofuran/benzothiophene oxo-acetic acid. Phe182 with π-stacking interaction and Lys116 with van der Waals contact were observed around the benzyl substituent. α-substituent of oxo-acetic acid functional group and orthoaromatic position of oxo-acetic acid moiety are surrounded by sterically favorable large green polyhedra. This green polyhedra was effectively occupied by the lipophilic fragment of compound **54** (hexyl side chain). Tyr46, Arg47, Asp48, Val49 amino acid residues form a lipophilic pocket in this region. In the case of sulfonobiphenyls, replacement of α-substituted oxo-acetic acid functional group with sulfonosalicylic acid retained approximately similar activity as that of compound **54**. The functional groups ortho to the sulfonosalicylic acid moiety on the terminal ring of biphenyl ring system in sulfonobiphenyls occupied the hydrophobic pocket formed by Tyr46, Arg47, Asp48, Val49 amino acid residues.

Compounds **2** and **5** with phenolic –OH functional group showed less activity. Compounds from oxazolobiphenyl class where the benzofuran/benzothiophene moiety replaced by oxazole and point of attachment between oxazole and biphenyl is third or fourth showed less activity. Because of the above modifications change in the orientation of the α-substituted oxo-acetic acid functional group did not take the advantage of the steric interactions in this region. Due to spacer groups (C=O, –CH$_2$) between 2-substituted benzofuran and biphenyl or naphthalene ring system (Table 1) compounds from this class lost favorable steric interactions of α-substituted oxo-acetic acid functional group showed less activity. Sterically favorable region was also observed around the phenyl ring bearing α-substituted oxo-acetic acid functional group. This phenyl ring is stacked between Tyr46 and Phe182 residues by aromatic π–π stacking interactions. Hydrophobic contacts by side chain of Lys116 and Ser216 with this phenyl ring were observed.

Sterically forbidden yellow region was embedded into the favorable green region around the hexyl side chain of compound **54** (lipophilic pocket formed by Tyr46, Arg47, Asp48, Val49). This indicates that functional groups protruding beyond this region show less activity. Compounds **2** and **5** lacking functional group to extend into sterically favorable region showed less activity. Due to change in the orientation of terminal lipophilic substituent on the

α-carbon atom of oxo-acetic acid functional group compounds from oxazolo biphenyls and 2-substitued benzofuran naphthalenes extended into sterically forbidden yellow region showed less activity. Sterically unfavorable region was also found in the vicinity of Gly220 and Arg221 amino acid residues. Inclusion of any steric bulk in this region will prevent the interaction of oxo-acetic acid functional group of these compounds with amide protons of Gly220 and Arg221 of PTP 1B.

Electrostatic interactions (Fig. 5)

The electrostatic field map shows negative charge favorable red polyhedra around the α-substituted oxo-acetic acid functional group. Amide protons of Gly218, Gly220, Arg221 and guanidium side chain of Arg221 was observed in this region. It is evident from literature that to exhibit PTP 1B inhibitory activity compounds should possess ionizable functional groups to interact with signature motif residues (Cys215-Arg221) of PTP 1B, and this interaction serves as key recognition element in ligand–receptor interactions.[38] Compounds **2** and **5** with out oxo-acetic functional group and compounds from oxazolobiphenyls, 2-substituted benzofuran biphenyl/naphthalenes due to unfavorable orientation of oxo-acetic acid functional group showed less activity than compound **54**.

Negative charge favorable red polyhedra was also observed in the vicinity of hexyl side chain of compound **54**. Amide protons of Arg47, Asp48, Val49 and guanidium group of Arg47 were observed in this region. This indicates that lipophilic moiety with electron-rich functional group in this region may show enhanced activity due to favorable electrostatic interactions. In the case of sulfonobiphenyls, terminal salicylic acid moiety seems to utilize this favorable interaction and showed good inhibitor activity. Red contour was also found in the vicinity of Ala217, Gly218, and Val49 residues of PTP 1B. Electron rich functional groups in this region may show enhanced activity by interacting with amide protons of Ala217, Gly218, and Val49 residues. Another red contour was found around the electronegative atom of heterocyclic moiety of these compounds.

Negative charge disfavored blue region was found around oxo-acetic functional group. Compounds **2** and **5** with phenolic hydroxyl functional group extended into this region showed less activity. This indicates that the orientation of α-substituted oxo-acetic acid functional group is important to have favorable electrostatic interactions with signature motif residues (Cys215-Arg221) of PTP 1B in this region. These results indicate that compounds should possess electron rich functional groups that mimic the phosphate group of phosphotyrosine (pTyr), a natural substrate for PTP 1B to exhibit PTP 1B inhibitory activity.

Hydrophobic interactions (Fig. 9)

CoMSIA hydrophobic 3D-contour plot showed hydrophobically favorable yellow regions around hexyl side chain. Comparison of hydrophobic fields with PTP 1B

active site revealed the lipophilic pocket formed by Tyr46, Arg47, Asp48, Val49 residues. Hence, bulky substituents to occupy this region are necessary to show good PTP 1B inhibitor activity. Presence of Phe182 amino acid residue around the hydrophobically favorable yellow region occupied by the phenyl ring bearing the oxo-acetic acid functional group, indicates the importance of aromatic ring system in this region to have favorable π–π stacking interactions. Amino acid residues Phe182 and Lys116 were found around the favorable yellow region near to benzyl substituent of benzofuran heterocycle of compound **54**. van der Waals and hydrophobic interactions seems to be important in this region.

Thus, CoMFA steric fields can be seen as comprehensive contribution of the pure steric and hydrophobic effects as the combination of steric and electrostatic fields with A log P gave good statistical model. CoMSIA hydrophobic 3D-contour maps revealed the favored and disfavored regions of PTP 1B inhibitors.

Conclusions

The 3D-QSAR model using CoMFA and CoMSIA methods were used to rationalize the PTP 1B inhibitory activity of 118 compounds belonging to benzofuran/benzothiophene biphenyls with anti-hyperglycemic activity. The CoMFA model obtained from atom-based alignment and A log P as additional descriptor showed good correlation with biological activity and predictive ability. A high boot-strapped r^2 value and small standard deviation indicate a similar relationship exists in all compounds. Negative value of cross-validated r^2 in the randomized test revealed that the results were not based on chance correlation. Inclusion of A log P a lipophilic parameter improved the significance of CoMFA model indicating the role of lipophilicity in ligand–receptor interactions in addition to steric and electrostatic interaction. The comparison of these 3D-QSAR models with the active site of PTP 1B validated the conformation and alignment rules that we employed. 3D-QSAR contour maps showed good compatibility with the receptor properties even though the conformations and ligands is not based on the receptor structure. In the present study, we have identified the importance of non-covalent interactions of these inhibitors. The lipophilic substituents on the α- carbon of oxo-acetic acid functional group interact with the hydrophobic pocket (Tyr46, Arg47, Asp48, Val49) produced from the surface of the Cys215, and such interaction seems to play significant role in governing the inhibitor potency of these compounds. This suggests that inherent structure of the inhibitors as well as the α-substituted oxo-acetic acid, sulfo-salicylic acid functional group that mimics the phosphate group of phosphotyrosine (pTyr) is essential for the activity. The structural requirements of the present series of compounds identified through 3D-QSAR contour plots will be helpful in the design of new PTP 1B inhibitors with better activity.

Methods

Biological data

118 molecules belonging to benzofuran/benzothiophene biphenyls have been reported as inhibitors of PTP 1B with anti-hyperglycemic activity.[16] The logarithm of measured IC$_{50}$ (μM) against human recombinant PTP 1B enzyme (h-PTP 1B) as pIC$_{50}$ was used in 3D-QSAR, thus correlating the data linear to the free energy change. As IC$_{50}$ against h-PTP 1B for compounds exhibiting >70% inhibition at 2.5 μM concentration (average of quadruplet) was not determined, such compounds are excluded from the present study. The training set of 92 molecules with structures and pIC$_{50}$ values are shown in Table 1. It is essential to access the predictive power of the models by using a test set of compounds. The test set of 26 molecules with structures and pIC$_{50}$ values are shown 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 0.86 and 0.66, respectively. Thus, the test set is the true representative of the training set.

Molecular modeling

All molecular modeling and 3D-QSAR studies were performed on Silicon Graphics Indy R5000 computer. Structural manipulations were performed with molecular modeling package SYBYL 6.6,[24] using standard Tripos force field.[25] Partial atomic charges of the molecules were calculated using AM$_1$ model Hamiltonian[26] within the MOPAC (Key words: 1SCF, MMOK, PARASOCK).

An essential requirement for 3D-QSAR techniques using molecular field analysis is knowledge of the active conformation of the inhibitor understudy. As the crystallographic conformation of present series of molecules was not available/deposited at protein data bank, the initial conformation of most active molecule (compound **54**) was obtained from simulated annealing as it enables the rapid identification of good solutions, ideally the global minimum.[27] The system was heated to 1000 K for 1.0 ps and then annealed to 250 K for 1.5 ps. The annealing function was exponential, 50 such cycles were run and those resulting 50 conformers were optimized using Powell method with a convergence of 0.001 kcal/mol Å. The default settings for all other minimization options and a distance dependent dielectric of 1.0 were employed through out the calculation. The active conformation for the most active compounds from 2-butyl benzofuran biphenyl (compound **66**), substituted oxazolobiphenyls (compound **72**), 2-butyl benzofuran naphthalenes (compound **80**), sulfonobiphenyls (compound **87**) was obtained from simulated annealing technique as mentioned above. All the other molecules were constructed using standard geometries and standard bond lengths from the nearest structures obtained from simulated annealing. A constrained minimization was performed initially in which the unchanged part of the new inhibitor molecule was defined as aggregate. The constraints were then removed and minimization was repeated by Powell

method till the root mean square deviation (RMSD) 0.001 kcal/mol Å was achieved.

The docking and molecular dynamics simulations were carried out on PTP 1B enzyme (PDB code 1ecv). The enzyme was refined using stepwise refinement protocol.[28] The active conformer of the template molecule (compound **54**) was docked into the active site of PTP 1B using interactive docking option available with in SYBYL. The binding orientation obtained from docking was subjected to molecular dynamics simulations at 300 K with 10 ps equilibration and 40 ps simulation using NTV ensemble to obtain favorable binding orientation of the inhibitor with enzyme. The trajectories obtained from molecular dynamics simulations were analyzed based on the interaction energy, hydrogen bonding interactions, hydrophobic interactions and van der Waals contacts. The selected trajectory was finally minimized to rms gradient of 0.01 kcal/mol Å. The active conformer of compound **54** used for 3D-QSAR study obtained from simulated annealing was superimposed on the conformer obtained from the docking and molecular dynamics simulations using match option in Sybyl. A low root mean square deviation (RMSD) between the heavy atoms of the ligands supports our choice of the conformations obtained from simulated annealing technique for the molecular field analysis. The 3D-QSAR model was then superimposed to the PTP 1B active site, giving direct contour maps of the different fields.

Alignment rules

The alignment rule, that is molecular conformation and orientation, is one of the most sensitive input areas for 3D-QSAR study. In the present study, three different alignment rules were adopted.

Alignment 1. This alignment involved RMS fitting (atom-based fitting) of the heavy atoms of the ligands. The compounds were fitted to the template molecule, compound **54** (Fig. 1), by:

a. C_2 atom of the heterocycle ring (1)

b. C_3 atom of the heterocycle ring (2)

c. Two connecting atoms of the biphenyl ring (3,4)

d. Oxygen of substituted oxo-acetic acid moiety (5)

e. Carbon atom of the substituted oxo-acetic acid moiety (6)

This alignment maximizes the overlap of the heavy atoms of all the ligands.

Alignment 2. In this case, alignment of the molecules was carried out by flexible fitting (multi fit) of the atoms of the ligands to the template molecule, compound **54**. This involved energy calculation and fitting on the template molecule by applying force (force constant 20 kcal/ mol Å) and subsequent energy minimization.

Alignment 3. This was carried out by using the SYBYL QSAR rigid body field fit command within SYBYL and using compound **54** as template molecule. Field fit adjusts the geometry of the ligands such that the fields match with the template molecule.

CoMFA and CoMSIA studies

The steric and electrostatic potential fields for CoMFA were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å. The CoMFA region was defined automatically, and is extended past the van der Waals volume of all the molecules in *X*, *Y*, and *Z* directions. The van der Waals potential (Leonard–Jones, 6–12) and coloumbic term, which represent, respectively, steric and electrostatic fields were calculated using the TRI-POS force field. A distance dependent dielectric expression $\in = \in_0$ Rij with $\in = 1.0$ was used. An sp^3 carbon atom with van der Waals radius of 1.52 Å and a + 1.0 charge was served as the probe atom to calculate steric and electrostatic fields. The steric and electrostatic contributions were truncated to ± 30 kcal/mol, and the electrostatic contributions were ignored at lattice intersections with maximum steric interactions.

The five CoMSIA similarity index fields available within Sybyl (steric, electrostatic, lipophilic, hydrogen bond donor, and hydrogen bond acceptor) were calculated at grid lattice point using a common probe atom of 1 Å radius, as well as the charge, hydrophobicity, and hydrogen bond properties of + 1 and an attenuation factor of 0.3.

PLS analysis

The partial lest-squares method (PLS)[29–31] was used to derive a linear relationship, and cross-validation was preformed using the leave-one-out method,[32,33] with a 2 kcal/mol column filter to check for consistency and predictiveness. The optimum number of components used to derive the nonvalidated model was defined as the number of components leading to the highest cross-validation r^2 (subsequently called r^2_{cv}) and lowest standard error of prediction (SEP). The leave-one-out method of cross-validation is rather obsolete and it generally gives high r^2 value, to ascertain the true predictivity of the model a harder test using leave-half-out method of cross-validation was performed for all the analyses.[34] The cross-validation analysis was performed by setting the number of cross-validation groups to 2 [leave-half-out (LHO)]. In this case, cross-validation groups were randomly selected and a model is derived. This was then used to predict the activity of the compounds from the other group. This analysis was repeated 100 times and the mean and standard deviation of r^2_{cv} values are reported (Table 6). To obtain the statistical confidence limit on the analysis, PLS analysis using 100 bootstrap groups with optimum number of components was performed. The PLS analysis with randomly interchanged biological activity was performed to estimate the extent of chance correlation (Table 6). All the cross-validated results were analyzed by considering the fact that a value of r^2_{cv} above 0.3

indicates that probability of chance correlation is less then 5%.[35]

Predictive r^2 values

The predictive ability of each analysis was determined from a set of 26 compounds that were not included in the training set. These molecules were aligned, and their activities were predicted by each PLS analysis. The predictive r^2 (r^2_{pred}) value will be based on molecules of the test set only and it is defined as:

$$r^2_{pred} = (SD - PRESS)/SD$$

where SD is the sum of the squared deviations between the biological activities of the test set and mean activity of the training set molecules and PRESS is the sum of squared deviation between predicted and actual activity values for every molecule in the test set.

Calculation of additional descriptors

Different types of physicochemical descriptors were calculated using Cerius2 version 3.5[36] running on Silicon Graphics O$_2$ R5000 workstation. These include electronic, spatial, structural, and thermodynamic descriptors. These descriptors were directly used as additional regressors in the PLS analysis.

Acknowledgements

References and Notes

1. Drake, P. G.; Posner, B. I. *Mol. Cell. Biochem.* **1998**, *182*, 79.

2. White, M. F.; Khan, C. R. *J. Biol. Chem.* **1994**, *269*, 1.

3. Byon, J. C. H.; Kusari, A. B.; Kusari, J. *Mol. Cell. Biochem.* **1998**, *182*, 101.

4. Ahmad, F.; Li, P.-M.; Meyerovitch, J.; Goldstein, B. J.; Chernoff, J.; Gustafson, T. A.; Kusari, J. *J. Biol. Chem.* **1995**, *270*, 20508.

5. Bandyopadhyay, D.; Kusari, A.; Kenner, K. A.; Liu, F.; Chernoff, J.; Gustafson, T. A.; Kusari, J. *J. Biol. Chem.* **1997**, *272*, 1639.

6. Kusari, J.; Kenner, K. A.; Shu, K. I.; Hill, D. E.; Henry, R. R. *J. Clin. Invest.* **1994**, *93*, 1156.

7. Ahmad, F.; Lonsidane, R. V.; Bauer, T. L.; Ohannesian, J. P.; Marco, C. C.; Goldstein, B. J. *Metabolism* **1997**, *46*, 1140.

8. Elchelby, M.; Payette, P.; Michaliszyn, E.; Cromlish, W.; Collinschan, C. C.; Ramachandran, C.; Gresser, M. J.; Tremblay, M. L.; Kennedy, B. P. *Science* **1999**, *283*, 1544.

9. Zhang, Z. Y. *CRC. Criti. Rev. Biochem. Mol. Biol.* **1998**, *33*, 1.

10. Burke, T. R., Jr.; Kole, H. K.; Roller, P. P. *Biochem. Biophys. Res. Commun.* **1994**, *204*, 129.

11. Kole, H. K.; Akamatsu, M.; Ye, B.; Yan, X.; Barford, D.; Roller, P. P.; Burke, T. R., Jr. *Biochem. Biophys. Res. Commun.* **1995**, *209*, 817.

12. Kole, H. K.; Smyth, M. S.; Russ, P. L.; Burke, T. R., Jr. *Biochem. J.* **1995**, *311*, 1025.

13. Sarmeinto, M.; Wu, L.; Keng, Y. F.; Song, L.; Luoz-Haung, Z.; Wu, G. Z.; Yaun, A. K.; Zhang, Z. Y. *J. Med. Chem.* **2000**, *43*, 146.

14. Worbel, J.; Li, Z.; Dietrich, A.; McCaleb, M.; Michan, B.; Sredy, J.; Sullivan, D. *J. Med. Chem.* **1998**, *41*, 1084.

15. Taing, M.; Keng, Y. F.; Shen, K.; Wu, L.; Lawrence, D. S.; Zhang, Z. Y. *Biochemistry* **1999**, *38*, 3797.

16. Malamas, M. S.; Sredy, J.; Moxham, C.; Katz, A.; Xu, W.; Devitt, R. M.; Adibayao, F. O.; Sawicki, D. R.; Seestaller, L.; Sullivan, D.; Taylor, X., Jr. *J. Med. Chem.* **2000**, *43*, 1293.

17. Crammer, R. D.III.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.

18. Gokhale, V. M.; Kulkarni, V. M. *J. Med. Chem.* **1999**, *42*, 5348.

19. Talele, T. T.; Kulkarni, S. S.; Kulkarni, V. M. *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 958.

20. Kulkarni, S. S.; Kulkarni, V. M. *Boiorg. Med. Chem.* **1999**, *7*, 1475.

21. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.

22. Makhija, M.T., Kulkarni, V.M. *J. Comput.-Aided. Mol. Des.* **2002** (In press).

23. Klebe, G.; Abraham, U. *J. Comput.-Aided. Mol. Des.* **1999**, *13*, 1.

24. *SYBYL 6.6*, Tripos Associates: 1669, S. Hanely Road, Suite 303, St. Louis, Missouri, MO 63144–2913, USA:

25. Clark, M.; Crammer, R. D.III.; Van Opdenbosh, N. *J. Comput. Chem.* **1989**, *10*, 982.

26. *MOPAC 6.00*; Quantum Chemical Program Exchange: Indiana University, no. 455.

27. Barakat, M. T.; Dean, P. M. *Comput.-Aided. Mol. Des.* **1990**, *4*, 295.

28. Levit, M.; Lifson, S. J. *J. Mol. Biol.* **1969**, *46*, 269.

29. Wold, S., Albano, C., Dunn, W. J., Edlund, U., Esbenson, K., Gelad, P., Hellberg, S., Lindberg, W., Sjostrom, M. In *Chemometrics*; Int. Ed. Kowalski, B., Reidel: Dortrecht, The Netherlands, 1984; p 17.

30. Dunn, W. J.; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. *Qaunt. Struct. Act. Relat. Chem. Bio.* **1984**, *3*, 31.

31. Geladi, P. *J. Chemom.* **1998**, *2*, 231.

32. Wold, S. *Technometrics,* **1978**, *4*, 397.

33. Crammer, R. D.III.; Bunce, J. D.; Patterson, D. E. *Quant. Struct. Act. Relat.* **1998**, *7*, 18.

34. Kulkarni, S. S.; Kulkarni, V. M. *J. Med. Chem.* **1999**, *42*, 373.

35. Kaminski, J. J.; Doweyko, A. M. *J. Med. Chem.* **1997**, *40*, 427.

36. *Cerius2, 3.5*; Molecular Simulations Inc.: Scranton Road, San Diego, CA, USA.

37. Testa B, Carrupt PA, Gaillard P, Tasai RS. Intra Molecular Interactions Encoded in Lipophilicity. In *Lipophilicity in Drug Action and Toxicology*; Pliska V, Testa B, Van der Waterbeemd H., Eds., VCH: Weinheim, 1997; p 49.

38. Chen, L.; Wu, L.; Otaka, A.; Smyth, M. S.; Roller, P. P.; Burke, T. R., Jr.; den Hertig, J.; Zhang, Z. Y. *Biochem. Biophys. Res. Commun.* **1995**, *21*, 697.