

# Pharmacophore Modeling and Virtual Screening Studies to Design Potential Protein Tyrosine Phosphatase 1B Inhibitors as New Leads

Neelakantan Suresh<sup>1\*</sup> and N. S. Vasanthi<sup>2</sup>

<sup>1</sup>GVK Biosciences Pvt. Ltd., 37, Sterling Road, Nungambakkam, Chennai 600034, T.N., India

<sup>2</sup>Head, Department of Biotechnology, Bannari Amman Institute of technology, Erode 638401, T.N., India

## Abstract

Protein Tyrosine Phosphatase 1B (PTP-1B) is one of the important targets in the treatment of diabetes and obesity. They play a very important role in cellular signaling within and between cells. The best pharmacophore hypothesis (Hypo 1), consisting of four features, namely, one hydrogen-bond acceptor (HBA), one hydrophobic point (HY), and two ring aromatics (RA), has a correlation coefficient of 0.961, a root mean square deviation (RMSD) of 0.885, and a cost difference of 62.436, suggesting that a highly predictive pharmacophore model was successfully obtained. A chemical feature based pharmacophore model has been generated from known PTP-1B inhibitors (25 training set compounds) by HypoGen module implemented in CATALYST software. The top ranked hypothesis (Hypo1) contained four chemical feature types such as hydrogen-bond acceptor (HA), hydrophobic aromatic (HY), and two ring aromatic (RA) features. Hypo1 was further validated by 125 test set molecules giving a correlation coefficient of 0.905 between experimental and estimated activity. This was also validated using CatScramble method. Thus, the Hypo1 was exploited for searching new lead compounds over chemical compounds in Medichem database and then the selected compounds were screened based on restriction estimated activity. Finally, we obtained 30 new lead candidates and the one best highly active compound structure was selected as a lead compound. The results demonstrate that hypothesis derived in this study could be considered to be a useful and reliable tool in identifying structurally diverse compounds with desired biological activity.

**Keywords:** Protein tyrosine phosphatase 1B; Diabetes; Insulin; Obesity; Catalyst; Pharmacophore

## Introduction

Protein tyrosine phosphatases (PTPs) constitute a family of receptor-like and cytoplasmic enzymes that catalyze the dephosphorylation of phosphotyrosine residues in protein substrates. PTPs together with protein tyrosine kinases (PTKs) play critical roles in regulating intracellular signal transduction pathways responsible for controlling cell growth, differentiation, motility, and metabolism.

Protein tyrosine phosphatases (PTPs) have emerged as a new and promising class of signaling targets, since the discovery of PTP-1B as a major drug target for diabetes and obesity. Biochemical and cellular studies have provided evidences that PTPs have an important role in the regulation of insulin signal trans-

duction. Protein tyrosine phosphatases 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 Reference (Zhang and Zhang, 2007).

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. Blocking one or more phosphatases could enhance the phosphorylation state of the insulin receptor kinase/subunit and/or its downstream signaling partners and restore the insulin resistance, which is a characteristic of type II diabetes (van Huijsduijnen et al., 2004).

Since then, many drugs have been synthesized by various companies for targeting PTP-1B, which is very challenging due to the closed form of the catalytic site of PTPs containing a highly polar phosphotyrosine (pTyr) binding site. The quest for oral PTP-1B inhibitors, with a satisfactory balance between physicochemical properties and selectivity, is still in its early stages, but despite the recent progress, compounds with optimal oral activity remain to be discovered. A pharmacophore model represents the 3D arrangements of the structural or chemical features of a drug (small organic compounds, peptides, peptidomimetics, etc.) that may be essential for interacting with the protein for optimum binding.

These pharmacophore models can be used differently in drug design programs such as

- (i) 3D query tool for virtual screening to identify potential new compounds from 3D databases of "drug-like" molecules that have patentable structures different from those that currently exist, and
- (ii) Tool to predict the activities of a set of new compounds that remain to be synthesized Reference (Bharatham et al., 2007).

In the present study, we have generated pharmacophore model using Catalyst software Reference (Catalyst 4.11, Accelrys, Inc.,

**\*Corresponding author:** Neelakantan Suresh, GVK Biosciences Pvt. Ltd., 37, Sterling Road, Nungambakkam, Chennai 600034, T.N., India

**Received** September 27, 2009; **Accepted** January 12, 2010; **Published** January 12, 2010

**Citation:** Suresh N, Vasanthi NS (2010) Pharmacophore Modeling and Virtual Screening Studies to Design Potential Protein Tyrosine Phosphatase 1B Inhibitors as New Leads. J Proteomics Bioinform 3: 020-028. doi:10.4172/jpb.1000117

**Copyright:** © 2010 Suresh N, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

San Diego, CA, 2005) for diverse set of molecules of PTP 1B with an aim to obtain Pharmacophore model which could provide a rational hypothetical picture of the primary chemical features responsible for activity.

This is expected to provide useful knowledge for developing new potentially active candidates targeting the PTP-1B which can be useful for treatment of obesity and diabetes. Pharmacophore modeling correlates activities with the spatial arrangement of various chemical features.

## Materials and Methods

### Selection of molecules

We selected a set of 150 compounds which are reported to be inhibitors of PTP-1B. The inhibitory activity of these compounds was expressed as IC<sub>50</sub> (i.e., concentration of compound required to inhibit 50% of PTP-1B was taken). The activity reported for these compounds was measured according to assay procedures. The IC<sub>50</sub> values spanned across a wide range from 0.039  $\mu$ M to 1800  $\mu$ M. Of these 150 compounds, 25 compounds were taken as training set (Table 2) and the rest of the 125 compounds as test set (Chart 2 Table S1 Additional information). The dataset was divided into training set and test set. The training set was selected by considering both structural diversity and wide coverage of the activity range. They were distributed into most active, moderately active and least inactive compounds based on their IC<sub>50</sub> values in order to obtain critical information on pharmacophore requirements. The important aspect of this selection was to ensure that each active compound would teach

something new to the HypoGen module thus it can be able to uncover as much as critical information possible for predicting biological activity.

### Molecular modeling

The geometry of a compound is built with the Catalyst builder and optimized by the CHARMM like force field. All molecules were built using the builder module of Cerius2. All the structures were minimized using steepest descent algorithm with a convergence gradient value of 0.001 kcal/mol. Partial atomic charges were calculated using Gasteiger method Reference (Gasteiger and Marsilli, 1980). Further geometry optimization was carried out for each compound with the MOPAC 6 package using the semi-empirical AM1 Hamiltonian.

### Pharmacophore modeling

Multiple acceptable conformations were generated for all of 20.0 kcal/mol above the global energy minimum. Instead of using lowest energy conformation of each compound, multiple acceptable conformations were generated for all ligands within the Catalyst ConForm module using the "Poling" algorithm. A maximum of 250 conformations were generated for each molecule within an energy threshold all conformational models for each molecule in training set were used in Catalyst for pharmacophore hypothesis generation.

The training set molecules (25) associated with their conformations were submitted to the Catalyst hypothesis generation (HypoGen) (Table 1). Features of hydrogen-bond acceptor (HBA), hydrophobic features (HY), hydrogen-bond donor

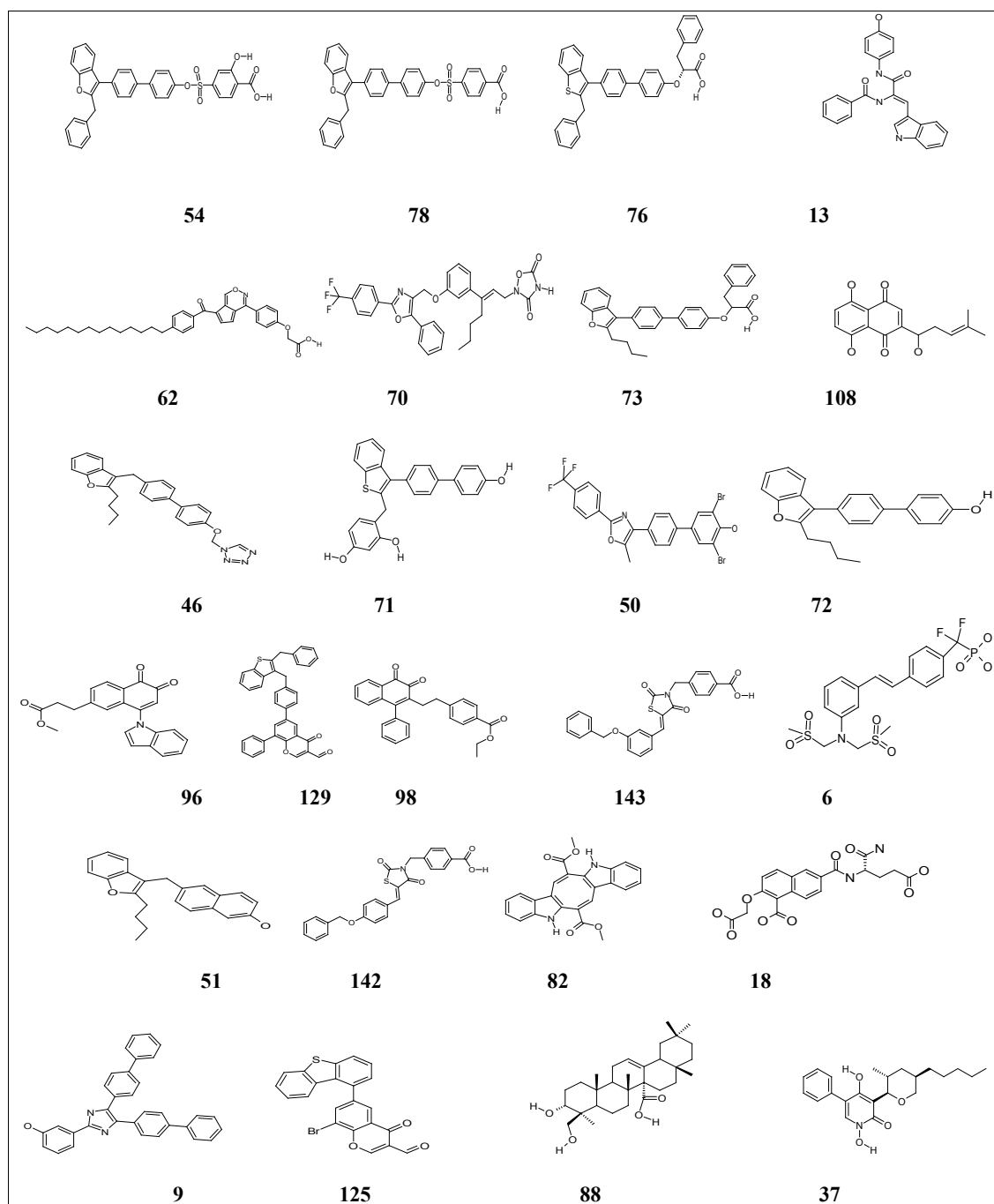
Compound	True IC <sub>50</sub> ( $\mu$ M)	Estimated IC <sub>50</sub> ( $\mu$ M)	Error factor (a)	Fit Value (b)	Activity Scale (c)	Estimated activity scale
54	0.039	0.064	1.6	12.45	+++	+++
78	0.075	0.054	-1.4	11.32	+++	+++
76	0.095	0.05	-1.9	11.31	+++	+++
62	0.15	0.74	4.9	11.58	+++	+++
70	0.37	0.65	1.8	11.66	+++	+++
73	0.44	0.37	-1.2	11.93	+++	+++
46	0.51	0.51	-1	11.55	+++	+++
71	0.58	0.35	-1.6	11.42	+++	+++
50	0.65	1.4	2.2	11.45	+++	+++
72	0.74	0.36	-2	11.34	+++	+++
96	0.92	0.88	-1.1	10.69	++	++
129	1	0.46	-2.2	11.66	++	++
98	1.01	0.51	-2	10.76	++	++
143	1.1	3.5	3.2	10.80	++	++
51	1.3	1.2	-1.1	10.65	++	++
142	1.6	1.2	-1.4	9.92	++	++
82	3.77	2.5	-1.5	9.67	++	++
9	8.6	1.4	-1	8.92	+	+
125	10	25	2.5	8.57	+	+
88	11.7	25	2.1	8.95	+	+
37	12	6	-2	9.28	+	+
13	21.5	25	1.1	9.58	+	+
108	25	25	-1	8.58	+	+
6	57.9	25	-2.4	8.93	+	+
18	3800	1800	-2.1	8.46	+	+

<sup>a</sup>The error factor is computed as the ratio of the measured activity to the activity estimated by the hypothesis or the inverse if estimated is greater than measured

<sup>b</sup>Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule.

<sup>c</sup>Activity scale: +++, <0.5  $\mu$ M (highly active); ++, >0.5-10  $\mu$ M (moderately active); +, >10  $\mu$ M (inactive).

**Table 1:** Output of the score hypothesis process on the training set.



**Table 2:** Molecular structures of the 25 training set compounds.

(HBD), ring aromatic (RA) features were included for the pharmacophore generation on the basis of common features present in the study molecules. The statistical parameters like cost values determine the significance of the model. Ten pharmacophore models with significant statistical parameters were generated.

The best model was selected on the basis of a high correlation coefficient ( $r$ ), lowest total cost, and RMSD values. The final model was further validated by a test set of 125 molecules.

#### Generation of pharmacophore model

Based on the structures of the training set compounds and their experimentally determined inhibitory activities against PTP-1B, 10 best pharmacophore (or hypotheses) were generated using HypoGen module implemented in Catalyst 4.11 software. On

further analysis, it was observed that four chemical feature types such as hydrogen-bond acceptor (HA), hydrophobic aromatic (HY), and two ring aromatic (RA) features could effectively map all critical chemical features of all molecules in the training and test sets. These features were further selected and used to build a series of hypotheses using the HypoGen module in Catalyst using default uncertainty value 3 (defined by Catalyst software as the measured value being within three times higher or three times lower of the true value). Catalyst thereby generates a chemical feature based model on the basis of the most active compounds.

In hypothesis generation, the structure and activity correlations in the training set were examined. HypoGen identifies those features that were common to the active compounds but excluded from the inactive compounds within conformationally allowable regions of space. It also further estimates the activity of each

training set compound using regression parameters. These parameters are computed by the regression analysis using the relationship of geometric fit value versus the negative logarithm of activity. The greater is the geometric fit, greater would be the activity prediction of the compound.

The fit function not only checks the feature is mapped or not but whether it contains a distance term, which measures the distance that separates the feature on the molecule from the centroid of the hypothesis feature. Both these terms are used to calculate the geometric fit value.

### Pharmacophore validation

The generated pharmacophore model should be able to also predict the activity of the molecules accurately and also identify the active compound from the database. Therefore, the derived pharmacophore map was validated using (i) cost analysis, (ii) test set prediction and (iii) Fisher's test.

### Cost analysis

The HypoGen module in Catalyst performs two important theoretical cost calculations determining the success of any pharmacophore hypothesis. One is the 'fixed cost' (termed as ideal cost), representing the simplest model that fits all data perfectly, and the second is the 'null cost' (termed as no correlation cost), representing the highest cost of a pharmacophore with no features and estimates activity to be the average of the activity data of the training set molecules.

A meaningful pharmacophore hypothesis may also result when the difference between null and fixed cost value is large; with values of 40-60 bits for a pharmacophore hypothesis may indicate that it has 75-90% probability of correlating the data (Catalyst 4.11 documentation).

Two other parameters determine the quality of any pharmacophore configuration cost or entropy cost depending on the complexity of the pharmacophore hypothesis space and should have a value <17, and the error cost, which is dependent on the root mean square differences between the estimated and the actual activities of the training set molecules. The RMSD represents the quality of the correlation between the estimated and the actual activity data. The best pharmacophore model has highest cost difference, lowest RMSD and best correlation coefficient.

### Test set activity prediction

In addition to the estimation of activity of the 25 training set molecules, the pharmacophore model should also be able to estimate the activity of new compounds. For external validation of the pharmacophore model, we have considered 125 compounds as test set (Table S1 Supporting information), having wide range of activities ( $IC_{50}$ , spanning from 0.5 to 10.00  $\mu$ M) and structural diversity. The best pharmacophore (Hypo1) having high correlation coefficient ( $r$ ), lowest total cost, and lower RMSD value was chosen to estimate the activity of test set. Test set compounds were classified on the basis of their activity as highly active < 0.5  $\mu$ M (highly active); ++, > 0.5-10  $\mu$ M (moderately active); +, > 10  $\mu$ M (inactive) (Doman et al., 2002; Lazo et al., 2001; Malamas et al., 2000a; Malamas et al., 2000b; Jia et al., 2001; Gao et al., 2001; Ripka, 2000; Lyon et al., 2002; Furstner et al., 2004; Taha and AIDamen, 2005; Cho et al., 2006; Cui et

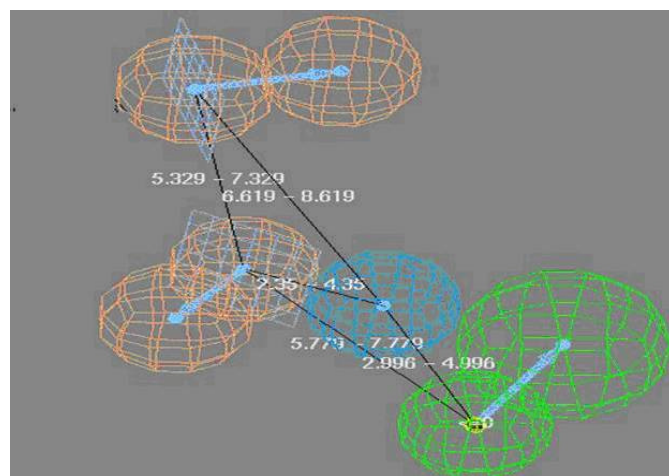
al., 2006; Dewang et al., 2005; Lazo et al., 2006; Mao et al., 2006; Na et al., 2006a; Na et al., 2006b; Wang et al., 1998; Liu et al., 2003; Wipf et al., 2001; Ahn et al., 2002; Chen et al., 2002; Shrestha et al., 2004; Black et al., 2005; Leung et al., 2002; Shim et al., 2005; Cao et al., 2005; Huang et al., 2003; Maccari et al., 2007; Wrobel et al., 1999).

### Fisher's test

Using the module CatScramble, the molecular spreadsheets of our training set were modified by arbitrary scrambling of the affinity data for all compounds. These randomized spreadsheets yield hypotheses without statistical significance; otherwise, the original model is also random. To achieve a statistical significance level of 98%, 41 random spreadsheets were generated for each of our three hypotheses. For all three targets, randomization tests gave hypotheses with total cost values lying well above those reported for the sets of original hypotheses, yielding lower values for the differences null hypothesis cost - total cost, further supporting the statistical significance of our models Sarma et al., 2008.

## Results and Discussion

Pharmacophore models were generated by HypoGen present in (Catalyst 4.11, Accelrys, Inc., San Diego, CA, 2005) and top 10 hypotheses (Table 1) were exported. Most hypotheses showed high correlation (>0.90). Interestingly, in the training set, all highly active compounds map all the features that is hydrophobic (HY), hydrogen-bond acceptor (HBA), and two ring aromatics (RA1 and RA2). With a few exceptions, in moderately active and inactive compounds one feature is missing. All the compounds in the training set map HY and RA1 feature revealing that these two features should be mainly responsible for the high molecular bioactivity, thus, should be taken into account in discovering or designing novel PTP-1B. The most active compound, 54, has a fitness score of 12.45 when mapped to Hypo 1 (Figure 1) whereas the least active, 18, maps to a value of 8.46 as seen in Figure 2B (1). On the basis of similar composition of the 10 hypotheses, hypothesis 1 (Hypo1), characterized by the best statistical parameters (Table 1) in terms of its predictive



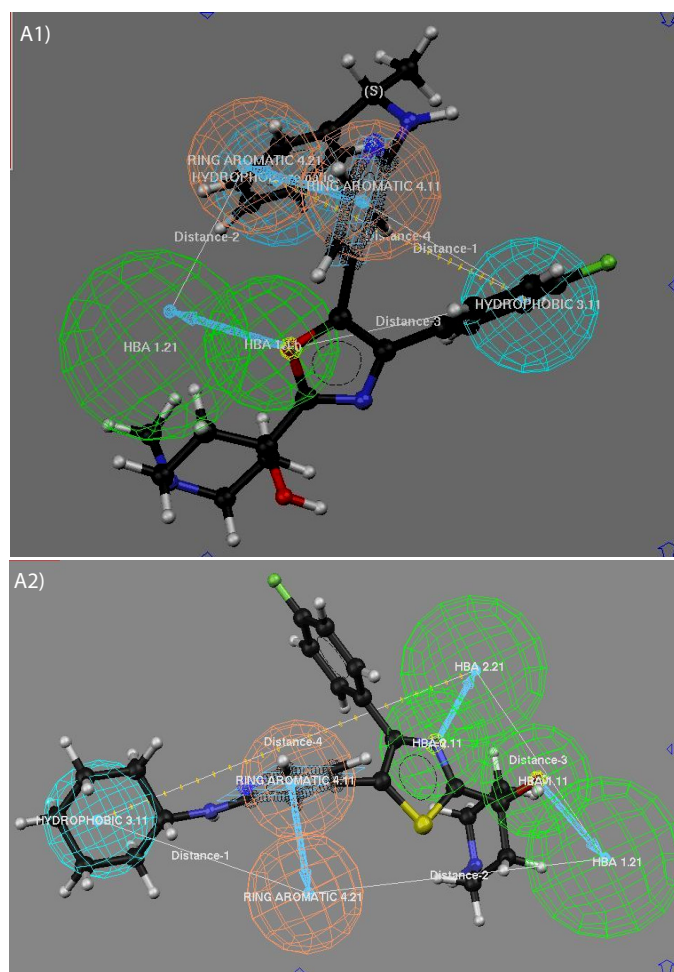
**Figure 1:** The best hypothesis model Hypo1 produced by the HypoGen module in Catalyst 4.11 software. The best hypothesis model Hypo 1 produced by the HypoGen module in Catalyst 4.11 software. Pharmacophore features are color-coded with orange, blue and green contours representing the ring aromatic features (RA), hydrophobic feature (HY) and hydrogenbond acceptor feature (HA) respectively. Distance between pharmacophore features is reported in angstroms.

ability, as indicated by the highest correlation coefficient and lowest RMS deviations, has been chosen to represent 'the pharmacophore model'. Remarkably, the highest active compound (compound 54) can be nicely mapped onto the Hypo1 model by the best fit values, which are shown in Figure 2A (1) indicating that the Hypo1 model provides reasonable pharmacophoric characteristics of the PTP-1B inhibitors for component of their activities.

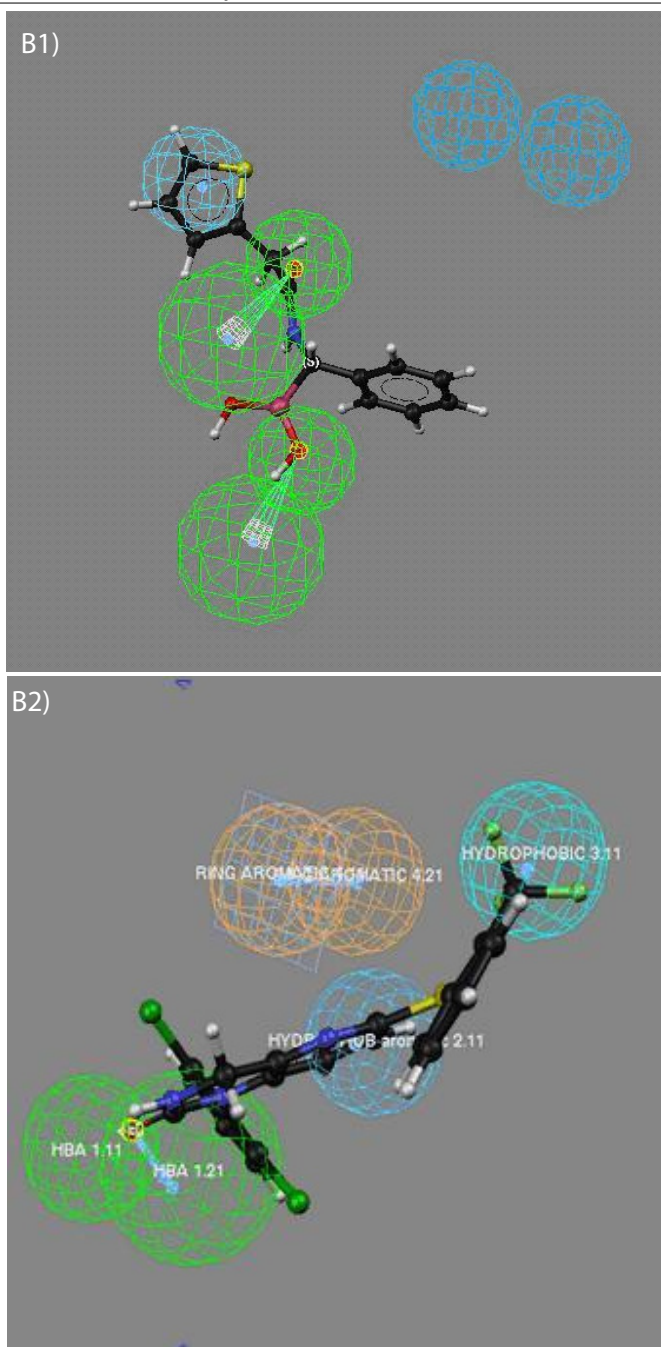
### Cost analysis

In addition to generating a hypothesis, Catalyst also provides two theoretical costs (represented in bit units) to help assess the validity of the hypothesis. The first is the cost of an ideal hypothesis (fixed cost), which represents the simplest model that fits all data perfectly. The second is the cost of the null hypothesis (null cost), which represents the highest cost of a pharmacophore with no features and which estimates activity to be the average of the activity data of the training set molecules. They represent the upper and lower bounds for the hypotheses that are generated. A generated hypothesis with a score that is substantially below that of the null hypothesis is likely to be statistically significant and bears visual inspection.

The greater the difference between the cost of the generated hypothesis and the cost of the null hypothesis, the less likely it is that the hypothesis reflects a chance correlation. A value of 40-60 bits between them for a pharmacophore hypothesis may indi-



**Figure 2A: Pharmacophore mapping of the most active compound on the best hypothesis model Hypo1.** (1) Compound 54 from the training set. (2) Compound 77 from the test set.



**Figure 2B: Pharmacophore mapping of the least active compound on the best hypothesis model Hypo1.** (1) Compound 18 from the training set. (2) Compound 110 from the test set.

cate that it has 75-90% probability of correlating the data. The total fixed cost of the run is 119.487, the cost of the null hypothesis 309.536, and the total cost of the Hypo1 is 141.995 (Table 3).

Then, the cost range between Hypo1 and the fixed cost is 25.426, while that between the null hypothesis and Hypo1 is 163.797 (Table 3), which shows that Hypo1 has more than 90% probability of correlating the data. Noticeably, the total cost of Hypo1 was much closer to the fixed cost than to the null cost. Furthermore, a high correlation coefficient of 0.966 was observed with RMS value of 1.354 and the configuration cost of 14.536, demonstrating that we have successfully developed a reliable pharmacophore model with high predictivity.

Hypo No.	Total cost	Cost difference (Null cost-Total cost)	Error cost	RMS	Correlation (r)	Features
1	125.257	62.436	106.062	1.3258	0.911	HBA, HY, RA, RA
2	125.899	56.907	106.531	1.3398	0.927	HBA, HY, RA, RA
3	126.049	52.348	94.041	1.345	0.939	HBA, HBA, HY, RA
4	126.098	51.151	106.247	1.331	0.947	HBA, HY, RA, RA
5	126.478	50.764	107.118	1.354	0.966	HBA, HY, RA, RA
6	125.850	50.656	102.352	1.605	0.949	HBA, HY, HY, RA, RA
7	125.777	50.925	101.813	1.590	0.957	HBA, HBA, HY, HY
8	124.566	50.946	101.824	1.590	0.949	HBA, HY, HY, RA
9	125.594	49.890	100.769	1.560	0.956	HBA, HBA, HY, RA, RA
10	124.663	49.769	103.641	1.641	0.945	HBA, HY, HY, RA

<sup>a</sup>Null cost = 309.536, fixed cost = 119.487; configuration = 15.469 and weight ~ 1.224. All cost units are in bits.

<sup>b</sup>HBA, hydrogen-bond acceptor; HY, hydrophobic feature; HBD, hydrogen-bond donor; RA, ring aromatic feature.

**Table 3:** Results of pharmacophore hypothesis generated using training set against PTP-1B.

### Score hypothesis

To verify Hypo1's discriminability among PTP-1B inhibitors with different order of magnitude activity, all training set compounds were classified by their activity as highly active highly active < 0.5  $\mu$ M (highly active); ++, > 0.5-10  $\mu$ M (moderately active); +, > 10  $\mu$ M (inactive). The actual and estimated PTP-1B inhibitory activities of the 25 compounds based on Hypo1 are listed in Table 1.

The discrepancy between the actual and the estimated activity observed for the two compounds was only about one-order of magnitude, which might be an artifact of the program that uses different numbers of degrees of freedom for these compounds to mismatch the pharmacophore model. The error factor is also reported in Table 1. It shows that 20 molecules out of the 25 molecules in the training set have errors less than 10 which means that the activity prediction of these compounds falls between 10-fold greater and 1/10 of the actual activity.

These results confirm that our hypothesis is a reliable model for describing the SAR in the training set. In this study, all but one highly active compound map the hydrogen-bond acceptor (HA) feature, and one least active inhibitor do not have this feature.

### Validation of the constructed pharmacophore model

The actual activities versus estimated activity of the 125 compounds in the test are shown in Table S1 in the Supporting information. A correlation coefficient of 150 generated using the test set compounds shows a good correlation of 0.951 between the actual and the estimated activities. Detailed, 7 out of 10 highly active, 33 of 55 moderately active, and 43 of 60 inactive compounds were predicted correctly. Two highly active compounds were underestimated as moderately active; five moderately active compounds were underestimated as inactive and other seven moderately active compounds were overestimated as highly active; most of inactive compounds were overestimated as moderately active.

The most active compound 77 in the test set had a fitness score of 12.05 when mapped to the Hypo 1 as seen in Figure 2A(2) and shows that all the features are being mapped accurately.

The least active compound 110 in the test set had a fitness score of 8.16 when mapped to the Hypo 1 as seen in Figure 2B

(2) and shows that all the features are not being mapped accurately.

In conclusion, most of the compounds in the test set were predicted correctly, which mean the hypothesis is suited for screening high active compounds from the database.

### Fisher's test

To further evaluate the statistical relevance of the model, Fisher's method was applied. With the aid of the CatScramble program, the experimental activities in the training set were scrambled randomly, and the resulting training set was used for a HypoGen run. All parameters were adopted which were used in initial HypoGen calculation. This procedure was reiterated 30 times. None of the outcome hypotheses has lower cost score than the initial hypothesis.

Finally, cross validation using the CatScramble program available in CATALYST was applied to assess the statistical confidence of Hypo1. The goal of this type of validation is to check whether there is a strong correlation between the structures and activity. CatScramble mixes up activity values of all training set compounds and creates 19 random spreadsheets (Sarma et al., 2008).

In this validation test, we select the 95% confidence level. We employed the first hypothesis (Hypo1) as 3D-search query against the NCI database using the 'fast flexible search' approach implemented within CATALYST. The pharmacophore captured 302 hits from a commercially available database of 10,458 compounds. The molecules identified included a broad range of templates that were structurally diverse from the starting molecule. The hits were subsequently fitted against the Hypo1 and the highest ranking 30 compounds were selected for being further investigated as potential new structures for design of novel PTP-1B inhibitors Sarma et al., 2008.

### Model validation and knowledge based screening

The purpose of the pharmacophore hypothesis generation is not just to predict the activity of the training set compounds accurately but also to verify whether the pharmacophore models are capable of predicting the activities of compounds of the test set series and classifying them correctly as active or inactive.

The best pharmacophore hypothesis was used initially to screen the PTP-1B inhibitors. All queries were performed using the Best

S. No	Parameter	PTP-1B
1	Total molecules in database (D)	302
2	Total Number of actives in database (A)	215
3	Total Hits (Ht)	218
4	Active Hits (Ha)	201
5	% Yield of actives [(Ha/Ht)*100]	92.2
6	% Ratio of actives [(Ha/A)*100]	93.49
7	Enrichment factor (E) [(Ha*D)/(Ht*A)]	1.3
8	False Negatives [A - Ha]	14
9	False Positives [Ht - Ha]	17
10	Goodness of Hit Score <sup>[a]</sup>	0.76

<sup>[a]</sup>-[(Ha/4HtA)(3A + Ht) - (1 - ((Ht - Ha)/(D - A)))] ; GH score of 0.6–0.7 indicates a very good model.

**Table 4:** Statistical parameters from screening test set molecules.

Flexible search databases/Spreadsheet method.

Hyporefine 1 was used to screen the known high, medium and low active inhibitors of the test set. Database mining was performed in Catalyst software using the BEST flexible searching technique.

A number of parameters such as hit list (Ht), number of active percent of yields (%Y), percent ratio of actives in the hit list (%A), enrichment factor of (E), False negatives, False positives and Goodness of hit score (GH) are calculated (Table 4) while carrying out the pharmacophore model and Virtual screening of test set molecules.

The number of molecules in the database is 302. Of these, 215 are highly active, 54 are moderately active and 36 are low active compounds. While the False positives and negatives, 16 and 12 respectively, are minimal, enrichment factor of 1.33 against a maximum value of 3.0 is a very good indication on the high efficiency of the screening. Of the 215 highly active molecules, 15 were predicted as moderately active and 4 were predicted as least active. In the 54 moderately active, 6 were predicted as low active and 3 as highly active.

The model also predicted 3 of the low active molecules as moderately active and 2 more molecules from the same set as highly active. The steric and other interaction effects might have a subtle, yet crucial role on the predicted activity.

While these additional groups may not prevent in identifying many low energy conformers or add any penalty for the total cost, but could be detrimental to fit these conformers in the active site.

Thus the features of Hyporefine 1 are relatively well optimized. However, in the case of highly active molecules, there are bulky groups present which may decrease the ability of the hyporefine to select the most highly active molecules Sarma et al., 2008.

## Conclusions

The work presented in this study shows how chemical features of a set of compounds along with their activities ranging over several orders of magnitudes can be used to generate pharmacophore hypotheses that can successfully predict the activity. The models were capable of predicting the activities over a wide variety of scaffolds and showed distinct chemical features that may be responsible for the activity of the inhibitors.

This knowledge can be used to identify and design inhibitors with greater selectivity.

Thus, the pharmacophores generated from the PTP-1B can be used:

1. To generate Pharmacophore models as powerful search tool to be used as a 3D query to identify lead molecules from chemical databases as potential PTP-1B inhibitors.
2. To evaluate how well any newly designed compound maps on the pharmacophore before undertaking any further study including synthesis. Both these applications may help in identifying or designing compounds for further biological evaluation and optimization.

A total data set of test and training of 150 compounds of selective PTP-1B inhibitors whose chemical features along with their respective activities ranging over a wide range of magnitude is used to generate pharmacophore hypotheses to successfully and accurately predict the activity. A highly predictive pharmacophore model was generated based on 25 training set molecules, which had hydrogen-bond acceptor, hydrophobic, hydrophobic bond donor and ring aromatic as chemical features which described their activities towards PTP-1B. The validity of the model was based on 125 test set molecules, which finally showed that the model was able to accurately differentiate various classes of PTP-1B inhibitors with a high correlation coefficient of 0.851 between experimental and predicted activity.

This validated pharmacophore model, as such can be used as a query for identification of potential inhibitors of PTP-1B while it can also be used to validate the potential of the compound to inhibit the enzyme prior to taking any step regarding the synthesis. PTB 1B enzymes have proven to be exciting and promising novel targets for the treatment of obesity and cancer. In-house build Medichem database was useful as a powerful resource to identify many PTP-1B inhibitors with highly varied activities and chemotypes. These PTP-1B inhibitors have been retrieved from the resource and some of them have been used to general a Pharmacophore model while other inhibitors have been used for virtual screening to validate the model.

The best quantitative Pharmacophore model in terms of predictive value consisted of four features like one hydrogen-bond acceptor (HA), one hydrophobic aromatic (HY), and two ring aromatic (RA) features, which is further validated by using an

large set of 378 PTP-1B inhibitors and gives a  $r$  value of 0.958. The most active molecule 54 ( $IC_{50} = 0.039 \mu M$ ) in the training set fits very well with this top scoring pharmacophore hypothesis. Virtual screening produced some false positives and a few false negatives. It is being noted that concurrent use or a consensus study, which readily minimizes these errors, could be an added tool for Pharmacophore model based virtual screening in order to produce reliable true positives and negatives. This Pharmacophore model was further used to search the NCI database consisting of structurally diversified molecules, yielded 218 molecules as hits that satisfied the 3D query. The activities of those molecules were predicted using the developed Pharmacophore model and the highly active molecules are further used to design more potent lead molecules against PTP-1B inhibitors for the treatment of various types of diabetes and obesity.

Thus, we hope that the model generated will be helpful to identify novel and potential lead molecules with improved activity against PTP-1B.

## Experimental

All molecular modeling works were performed on a Silicon Graphics Octane R12000 computer running Linux 6.5.12 (SGI, 1600 Amphitheatre Parkway, Mountain View, CA 94043) Catalyst 4.11 software was used to generate Pharmacophore models.

## Acknowledgements

The authors thank Dr. J.A.R.P. Sarma, Senior Vice President and S.Vadivelan, Senior Scientist, GVK Biosciences Pvt. Ltd., Chennai for their valuable guidance, providing software facilities and a great chance to work there.

## References

1. Ahn JH, Cho SY, Ha JD, Chu SY, Jung SH, et al. (2002) Synthesis and PTB1B Inhibition of 1,2-Naphthoquinone Derivatives as Potent Anti-Diabetic Agents. *Bioorg Med Chem Lett* 12: 1941-1946. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
2. Bharatham K, Bharatham N, Lee KW (2007) Pharmacophore Modeling for Protein Tyrosine Phosphatase 1B Inhibitors. *Arch Pharm Res* 30: 533-542. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
3. Black E, Breed J, Breeze AL, Embrey K, Garcia R, et al. (2005) Structure-based design of protein tyrosine phosphatase-1B inhibitors. *Bioorg Med Chem Lett* 15: 2503-2507. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
4. Catalyst 4.11, Accelrys, Inc., San Diego, CA, 2005, <http://www.accelrys.com>.
5. Cao S, Foster C, Brisson M, Lazo JS, Kingston DG (2005) Halenaquinone and xestoquinone derivatives, inhibitors of Cdc25B phosphatase from a Xestospongia sp. *Bioorg Med Chem* 13: 999-1003. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
6. Chen RM, Hu LH, An TY, Li J, Shen Q (2002) Natural PTP1B Inhibitors from *Broussonetia papyrifera*. *Bioorg Med Chem Lett* 12: 3387-3390. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
7. Cho SY, Baek JY, Han SS, Kang SK, Ha JD, (2006) PTP-1B inhibitors: Cyclopenta[d][1,2]-oxazine derivatives. *Bioorg Med Chem Lett* 16: 499-502. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
8. Cui L, Na M, Oh H, Bae EY, Jeong DG, et al. (2006) Protein tyrosine phosphatase 1B inhibitors from *Morus* root bark. *Bioorg Med Chem Lett* 16: 1426-1429. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
9. Dewang PM, Hsu NM, Peng SZ, Li WR (2005) Protein Tyrosine Phosphatases and their Inhibitors. *Curr Med Chem* 12: 1-22. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
10. Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, et al. (2002) Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. *J Med Chem* 45: 2213-2221. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
11. Furstner A, Feyen F, Prinz H, Waldmann H (2004) Synthesis and evaluation of the antitumor agent TMC-69-6H and a focused library of analogs. *Tetrahedron* 60: 9543-9558. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
12. Gao Y, Voigt J, Zhao H, Pais GC, Zhang X, et al. (2001) Utilization of a Peptide Lead for the Discovery of a Novel PTP1B-Binding Motif. *J Med Chem* 44: 2869-2878. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
13. Gasteiger J, Marsilli M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 36: 3219-3228. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
14. Huang P, Ramphal J, Wei J, Liang C, Jallal B, et al. (2003) Structure-Based Design and Discovery of Novel Inhibitors of Protein Tyrosine Phosphatases. *Bioorg Med Chem* 11: 1835-1849. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
15. Jia Z, Ye Q, Dinaut AN, Wang Q, Waddleton D, et al. (2001) Structure of Protein Tyrosine Phosphatase 1B in Complex with Inhibitors Bearing Two Phosphotyrosine Mimetics. *J Med Chem* 44: 4584-4594. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
16. Lazo JS, Aslan DC, Southwick EC, Cooley KA, Ducruet AP, et al. (2001) Discovery and Biological Evaluation of a New Family of Potent Inhibitors of the Dual Specificity Protein Phosphatase Cdc25. *J Med Chem* 44: 4042-4049. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
17. Lazo JS, Nunes R, Skoko JJ, Queiroz de Oliveira PE, Vogt A, et al. (2006) Novel benzofuran inhibitors of human mitogen-activated protein kinase phosphatase-1. *Bioorg Med Chem* 14: 5643-5650. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
18. Leung C, Grzyb J, Lee J, Meyer N, Hum G, et al. (2002) The Difluoromethylenesulfonic Acid Group as a Monoanionic Phosphate Surrogate for Obtaining PTP1B Inhibitors. *Bioorg Med Chem* 10: 2309-2323. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
19. Liu DG, Gao Y, Voigt JH, Lee K, Nicklaus MC, et al. (2003) Acylsulfonamide-Containing PTP1B Inhibitors Designed to Mimic an Enzyme-Bound Water of Hydration. *Bioorg Med Chem Lett* 13: 3005-3007. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
20. Lyon MA, Ducruet AP, Wipf P, Lazo JS (2002) Dual-Specificity Phosphatases As Targets For Antineoplastic Agents, *Nat. Rev. Drug Discov* 1: 961-970. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
21. Maccari R, Paoli P, Ottanà R, Jacomelli M, Ciurleo R, et al. (2007) 5-Arylidene-2,4-thiazolidinediones as inhibitors of protein tyrosine phosphatases. *Bioorg Med Chem* 15: 5137-5149. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
22. Malamas MS, Sredy J, Gunawan I, Mihan B, Sawicki DR, et al. (2000a) New Azolidinediones as Inhibitors of Protein Tyrosine Phosphatase 1B with Antihyperglycemic Properties. *J Med Chem* 43: 995-1010. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
23. Malamas MS, Sredy J, Moxham C, Katz A, Xu W, et al. (2000b) Novel Benzofuran and Benzothiophene Biphenyls as Inhibitors of Protein Tyrosine Phosphatase 1B with Antihyperglycemic Properties. *J Med Chem* 43: 1293-1310. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
24. Mao SC, Guo YW, Shen X (2006) Two novel aromatic valerenane-type sesquiterpenes from the Chinese green alga *Caulerpa taxifolia*. *Bioorg Med Chem Lett* 16: 2947-2950. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
25. Na M, Oh WK, Kim YH, Cai XF, Kim S, et al. (2006a) Inhibition of protein tyrosine phosphatase 1B by diterpenoids isolated from *Acanthopanax koreanum*. *Bioorg Med Chem Lett* 16: 3061-3064. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
26. Na M, Cui L, Min BS, Bae K, Yoo JK, et al. (2006b) Protein tyrosine phosphatase 1B inhibitory activity of triterpenes isolated from *Astilbe koreana*. *Bioorg Med Chem Lett* 16: 3273-3276. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)



27. Ripka WC (2000) Protein Tyrosine Phosphatase Inhibition. *Annu Rep Med Chem* 35: 231-250. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
28. Sarma R, Sinha S, Ravikumar M, Kumara MK, Mahmood SK (2008) Pharmacophore modeling of diverse classes of p38 MAP kinase inhibitors. *Eur J Med Chem* 43: 2870-2876. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
29. Shrestha S, Shim YS, Kim KC, Lee KH, Cho H (2004) Evans Blue and other dyes as protein tyrosine phosphatase inhibitors. *Bioorg Med Chem Lett* 14: 1923-1926. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
30. Shim YS, Kim KC, Lee KA, Shrestha S, Lee KH, et al. (2005) Formylchromone derivatives as irreversible and selective inhibitors of human protein tyrosine phosphatase 1B. Kinetic and modeling studies. *Bioorg Med Chem* 13: 1325-1332. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
31. Taha MO, AlDamen MA (2005) Effects of Variable Docking Conditions and Scoring Functions on Corresponding Protein-Aligned Comparative Molecular Field Analysis Models Constructed from Diverse Human Protein Tyrosine Phosphatase 1B Inhibitors. *J Med Chem* 48: 8016-8034. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
32. Wang Q, Huang Z, Ramachandran C, Dinaut AN, Taylor SD (1998) Naphthalene bis[alpha, alpha-Difluoromethylenephosphonates] as Potent Inhibitors of Protein Tyrosine Phosphatases. *Bioorg Med Chem Lett* 8: 345-350. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
33. Wipf P, Aslan DC, Southwick EC, Lazo JS (2001) Sulfonylated Aminothiazoles as New Small Molecule Inhibitors of Protein Phosphatases. *Bioorg Med Chem Lett* 11: 313-317. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
34. Wrobel J, Sredy J, Moxham C, Dietrich A, Li Z, et al. (1999) PTP1B Inhibition and Antihyperglycemic Activity in the ob/ob Mouse Model of Novel 11-Arylbenzo[b]naphtho[2,3-d]furans and 11-Arylbenzo[b]naphtho[2,3-d]thiophenes. *J Med Chem* 42: 3199-3202. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
35. Zhang S, Zhang ZY (2007) PTP-1B as a drug target: recent developments in PTP-1B inhibitor discovery. *Drug Discov Today* 12: 373-381. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)