Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2007) 870–884
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Discovery of new potent human protein tyrosine phosphatase inhibitors *via* pharmacophore and QSAR analysis followed by *in silico* screening

Mutasem O. Taha [a,*], Yasser Bustanji [b], Amal G. Al-Bakri [c], Al-Motassem Yousef [b], Waleed A. Zalloum [a], Ihab M. Al-Masri [a], Naji Atallah [a]

[a] *Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan*
[b] *Department of Clinical Pharmacy, Faculty of Pharmacy, University of Jordan, Amman, Jordan*
[c] *Department of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmacy, University of Jordan, Amman, Jordan*

Received 19 June 2006; received in revised form 19 August 2006; accepted 23 August 2006
Available online 3 September 2006

Abstract

A pharmacophoric model was developed for human protein tyrosine phosphatase 1B (h-PTP 1B) inhibitors utilizing the HipHop-REFINE module of CATALYST software. Subsequently, genetic algorithm and multiple linear regression analysis were employed to select an optimal combination of physicochemical descriptors and pharmacophore hypothesis that yield consistent QSAR equation of good predictive potential ($r = 0.87$, F-statistic $= 69.13$, $r^2_{BS} = 0.76$, $r^2_{LOO} = 0.68$). The validity of the QSAR equation and the associated pharmacophoric hypothesis was experimentally established by the identification of five new h-PTP 1B inhibitors retrieved from the National Cancer Institute (NCI) database.

Keywords: h-PTP 1B; Pharmacophore modeling; QSAR; *In silico* screening; Experimental validation

1. Introduction

Type II diabetes is a progressive disease characterized by insulin resistance in peripheral tissues and/or impaired insulin secretion by the pancreas. The resultant high blood glucose level generally leads to several serious complications. The World Health Organization recently warned that type II diabetes has become global pandemic [1]. Type II diabetes is strongly associated with obesity. The common link between obesity and type II diabetes is insulin resistance [2]. At the molecular level, the mechanism of insulin resistance in type II diabetes appears to involve defects in post-receptor signal transduction [3–5].

The interaction of insulin with its receptor leads to auto-phosphorylation of certain tyrosine residues within the receptor, thus activating the receptor kinase. However, protein tyrosine phosphatases (PTPases) dephosphorylate the activated insulin receptor and attenuate the tyrosine kinase activity thus reducing the metabolic action of insulin and leading to hyperglycaemia [6–8].

Among PTPases, PTP 1B has been particularly demonstrated to dephosphorylate insulin receptor and act as a negative regulator of insulin signaling. A recent study on PTP 1B knockout mice showed that the loss of PTP 1B activity enhanced sensitivity towards insulin and resistance to obesity, indicating that potent, orally active PTP 1B inhibitors could be useful for the treatment of type II diabetes and obesity [9–16].

Crystallographic structures of human (h-)PTP 1B–inhibitor complexes have uncovered distinguished features about the different binding sites and inhibitors. The active site of h-PTP 1B consists of eight amino acid residues (His-Cys-Ser-Ala-Gly-Ile-Gly-Arg) that bind to the phosphorylated tyrosine moiety of substrate proteins [14,17]. Several peptidic and non-peptidic inhibitors, directed to the active site have been developed [13,14,18,19]. A second binding site was discovered adjacent to the active site and inhibitors binding to both active site and the second binding site have been developed [20–22].

As there is no successful molecule available in the market for this potential target yet, there is an urgent need for developing therapeutically useful h-PTP 1B inhibitors. A large

* Corresponding author. Tel.: +962 6 5355000x2505; fax: +962 6 5339649.
E-mail address: mutasem@ju.edu.jo (M.O. Taha).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2006.08.008

number of patents and leading research articles have appeared in the recent past to deal with this issue [3,13,14,23–25].

However, nearly all computer-aided efforts directed towards developing new h-PTP 1B inhibitors were, and still, focused on structure-based (docking) techniques [26–29] and quantitative structure activity relationship (QSAR) methodologies, e.g., comparative molecular field and molecular similarity indices analyses (CoMFA and CoMSIA) [30–32].

However, despite the excellent potential of current docking methodologies in discovering new inhibitors, they lack the ability to predict accurately the potencies of their *in silico* hits due to their inability to evaluate binding free energies correctly [33–35]. Furthermore, docking studies rely heavily on the crystallographic structures of targeted proteins, which although considered the most reliable structural information that can be used for drug design, suffer from some serious problems such as inadequate resolution [36] and crystallization-related artifacts of the ligand–protein complex [37]. Furthermore, crystallographic structures generally ignore structural heterogeneity related to protein anisotropic motion and discrete conformational substrates [38].

On the other hand, despite the excellent predictive potential of QSAR-based methodologies (e.g., CoMFA and CoMSIA), they generally lack the ability to act as effective search queries to mine virtual three-dimensional (3D) databases for new hits [39].

Accordingly, we were prompted to develop a robust, ligand-based 3D pharmacophore integrated within a predictive QSAR model. This approach avoids the pitfalls of structure-based docking techniques; furthermore, the pharmacophore model can be effectively used as 3D search query to mine 3D libraries for new h-PTP 1B inhibitors.

We employed the HipHop-REFINE module of CATALYST software to construct plausible binding hypotheses for h-PTP 1B inhibitors [40–42]. Subsequently, genetic function approximation (GFA) algorithm and multiple linear regression (MLR) analyses were employed to achieve an optimal quantitative structure–activity relationship (QSAR) that combines a high-quality binding pharmacophore with other molecular descriptors.

The optimal pharmacophore was subsequently used as 3D search query to screen the NCI compounds database for new h-PTP 1B inhibitors.

CATALYST models drug–receptor interaction using information derived only from the drug structure [42–44]. Molecules are described as collection of chemical functionalities arranged in 3D space. The conformational flexibility of training ligands is modeled by creating multiple conformers, judiciously prepared to emphasize representative coverage over a specified energy range. HipHop-REFINE identifies a set of chemical features common to active training molecules. This 3D array of chemical features provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen bond donors and acceptors (HBDs and HBAs), aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionizable groups and aromatic planes. However, HipHop-REFINE utilizes the conformational space of inactive molecules to construct excluded volumes that represent the steric constrains of the proposed receptor [40,41]. Successful examples involving the use of CATALYST have been reported, wherein the CATALYST-derived pharmacophore has been used efficiently as a query for database searching and 3D-QSAR studies [45–48]. Nevertheless, HipHop-REFINE was only recently introduced, which explains the scarcity in reported literature about this module [74].

2. Methods

2.1. Data mining

The structures of 154 diverse h-PTP 1B inhibitors were collected from published literature [49–52]. The *in vitro* bioactivities of the collected inhibitors are expressed as the concentration of the test compound that inhibited recombinant h-PTP 1B activity by 50% (IC_{50}) or as the dissociation constants of the enzyme–inhibitor complexes (K_i values). See Supplementary information for the collected structures and corresponding bioactivities (IC_{50} or K_i values).

2.2. Modeling

2.2.1. Software and hardware

The following software packages were utilized in this project:

- ChemDraw Ultra 6.0, Cambridge Soft Corp. (www.cambridgesoft.Corn), USA;
- CATALYST (Version 4.10), Accelrys Inc. (www.accelrys.-com), USA;
- CERIUS2 (Version 4.10), Accelrys Inc. (www.accelrys.com), USA.

Modeling studies were performed using CATALYST and CERIUS2 installed on a Silicon Graphics Octane2 desktop workstation equipped with a 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system.

2.2.2. Molecular modeling

The two-dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra and saved in MDL-molfile format. Subsequently, they were imported into CATALYST, converted into corresponding standard 3D structures and energy minimized to the closest local minimum using the molecular mechanics CHARMm force field implemented in CATALYST. The resulting 3D structures were utilized as starting conformers for CATALYST conformational analysis, and were stored in SD format for calculation of variety of physicochemical properties and QSAR analysis within CERIUS2.

2.2.2.1. Conformational analysis. Molecular flexibility was taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each

inhibitor (**1–154**, seeSupplementary information) was explored adopting the ''best conformer generation'' option within CATALYST, which is based on the generalized CHARMm force field implemented in the program. Default parameters were employed in the conformation generation procedure, i.e., conformational ensembles were generated with an energy threshold of 20 kcal/mol from the local minimized structure and a maximum limit of 250 conformers per molecule. Hence, this search procedure will probably identify the best 3D arrangement of chemical functionalities explaining the activity variations among the training set [40–42,53,54].

2.2.2.2. Pharmacophore modeling. HipHop-REFINE identifies 3D spatial arrangements of chemical features that are common to active molecules in a training set. These configurations are identified by a pruned exhaustive search, starting from small sets of features and extending them until no larger common configuration is found. Active training set members are evaluated on the basis of the types of chemical features they contain, along with the ability to adopt a conformation that allows those features to be superimposed on a particular configuration. The user defines how many molecules must map completely or partially to the hypothesis *via* the Principal and MaxOmitFeat parameters. These options allow broader and more diverse hypotheses to be generated. The resultant pharmacophores are ranked as they are built. The ranking is a measure of how well the active training molecules map onto the proposed pharmacophores, as well as the rarity of the pharmacophore model. If a particular pharmacophore is ''rare'' then it will be less likely to map to an inactive compound and therefore it will be given a higher rank [42,44,55]. HipHop-REFINE uses inactive training compounds to construct excluded volumes that resemble the steric constrains of the binding pocket. It identifies spaces occupied by the conformations of inactive compounds and free from the active inhibitors. These regions are then filled with excluded volumes. HipHop-REFINE returns by default a maximum of 10 high-ranking pharmacophores from each automatic run [40,41].

A diverse subset of training compounds was selected from compounds **1–154** (see Supplementary information) for pharmacophore modeling (as shown in Table 1). HipHop-REFINE was instructed to explore up to five-featured pharmacophoric space of the following possible features: hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic and negative ionizable (NegIonizable). Furthermore, the number of features of any particular type was allowed to vary from 0 to 3, except for NegIonizable, which was restricted to 1–3.

We defined IC_{50} (or K_i) value of 2.0 μM as the activity/inactivity cutoff. This decision was based on the following: we noticed that choosing more potent cutoffs (e.g., <0.5 μM) yielded limiting pharmacophores of poor virtual screening capabilities (i.e., low number of search hits) when employed to screen the complete list of collected inhibitors (**1–154**, see Supplementary information). On the other hand, less potent cutoffs (e.g., >5.0 μM) yielded poorly discriminating pharmacophores that lack the ability to separate potent and inactive h-PTP 1B inhibitors. Therefore, we decided to settle somewhere in the middle, i.e., 2.0 μM. Furthermore, in our opinion, selecting a 2.0 μM-bioactivity cutoff should yield pharmacophoric models capable of retrieving hits within the same bioinhibitory range, which is considered a good starting point for subsequent SAR-based optimization.

Table 1
The training list used for pharmacophore modeling of h-PTP 1B inhibitors

Compound[a]	Bioactivities (μM)		Principal value	Max omit feat
	IC_{50}	K_i		
48	0.044		2	0
49	0.18		2	0
50	0.054		2	0
52	0.1		2	0
53	0.1		2	0
55	0.052		2	0
56	0.071		2	0
58	0.029		2	0
62	0.025		2	0
63	0.17		2	0
66	0.043		2	0
69	0.054		2	0
71	0.023		2	0
74	0.17		2	0
75	0.082		2	0
79	0.16		2	0
86	1.15		2	0
89	2.34		0	1
91	1.3		2	0
104	1.2		2	0
109	1.3		2	0
110	3.05		0	0
111	0.075		2	0
115	0.034		2	0
116	0.029		2	0
121	0.032		2	0
122	0.04		2	0
124	1.16		2	0
139	0.386		2	0
142	0.083		2	0
146		750	0	2
147		39	0	2
148		21	0	2
149		240	0	2
150		110	0	2
151		54	0	1
154		2100	0	1

[a] Compounds' numbers are as in Supplementary information.

Accordingly, inhibitors of IC_{50} (or K_i) values ≤2 μM were regarded as ''actives'' and were assigned Principal and MaxOmitFeat values of 2 and 0, respectively, to ensure that all of their chemical features will be considered in building the pharmacophore space [40,41]. On the other hand, inhibitors of IC_{50} (or K_i) >2 μM were considered inactive and were assigned Principal values of zero [40,41].

However, each inactive compound was carefully evaluated to assess whether its low potency is attributable to missing one or more pharmacophoric features, i.e., compared to active compounds, or related to possible steric clashes within the

binding pocket, or due to both factors. Therefore, inactive compounds suspected of missing one or more pharmacophoric features were assigned MaxOmitFeat values of 1 or 2, respectively. Spaces occupied by conformers and/or mappings of this group of compounds and free from conformers and/or mappings of active compounds are filled with excluded volumes.

However, compounds that seem to be inactive mainly due to steric clashes within the binding pocket were assigned MaxOmitFeat value of zero (inhibitor **110** in Table 1). This value instructs HipHop-REFINE to force inactive compound(s) to map all the pharmacophoric features of the binding model, and therefore permits the software to identify spaces occupied by excess structural fragments/features of such inactive compounds and fill them with excluded volumes.

HipHop-REFINE was configured to allow a maximum of 100 exclusion spheres to be added to the generated pharmacophoric hypotheses. This represents the default value for the number of excluded volumes in HipHop-REFINE.

Table 1 shows the training compounds and their corresponding Principal and MaxOmitFeat parameters.

The followings are important HipHop-REFINE control parameters used for hypotheses generation [55]:

- Spacing: This parameter controls the minimum allowed inter-feature distance in the resulting hypotheses. The default value was employed (i.e., 3.0 Å).
- MinPoints: The default value of this parameter is 4, specifying a minimum of four individual feature components for a generated hypothesis.
- MinSubset Points: Only configurations of features in input molecules, with at least the number of points specified by this option, are considered when identifying a candidate hypothesis. The default value is 4.
- Superposition error, check superposition and tolerance factor: The three control parameters together check the superposition of compounds for hypothesis generation. All three have a default value of 1. Reducing the value from its default tightens the fit.
- Misses: This value specifies the number of compounds in the training set that do not have to map to all features in generated hypotheses. This value was modified from the default of 1–8. This value allows the eight inactive compounds to miss at least one feature in the generated hypotheses.
- Feature misses: This specifies the number of compounds allowed not to map any particular feature in a generated hypothesis. The default value is 1, however, in this case it was configured to 8, which allows a particular pharmacophoric feature to be missed by eight different training molecules (inactives).
- Complete misses: This option specifies the number of molecules in the training set that do not have to map to more than one feature per generated hypothesis. In current case this parameter was modified from the default value of zero to 2.
- Mapping coefficient: This parameter controls the importance of having compounds with similar structure map to a hypothesis in a similar way. Increasing this parameter will penalize the hypotheses that deviate from this behavior. The default value is 0.

The resulting pharmacophores (10 models) were individually evaluated as 3D search queries against the complete list of collected compounds (**1–154**, see Supplementary information). Table 2 shows the number of captured compounds (i.e., coverage) by each pharmacophoric model, as well as their corresponding pharmacophoric features and exclusion spheres.

2.2.2.3. QSAR modeling. Only compounds **1–137** (see Supplementary information) were utilized as training set for QSAR modeling as their bioactivities were determined *via* the same bioassay procedure. The logarithm of measured $1/IC_{50}$ (μM) values were used in QSAR, thus correlating the data linear to the free energy change. The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformer representations in SD format. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module of CERIUS2. The calculated descriptors included various simple and valence connectivity indices, electro-topological state indices and other molecular descriptors (e.g., logarithm of partition coefficient,

Table 2
The pharmacophoric features and coverage values (i.e., against compounds **1–154**) of the 10 hypotheses generated by the HipHop-REFINE automatic run

Hypothesis[a]	Hypothesis rank[a]	Features	Number of exclusion spheres	Coverage[b]
1	428.166	2 × HBA, 2 × Hydrophobe, NegIoniz	92	111
2	417.895	2 × HBA, 2 × Hydrophobe, NegIoniz	77	112
3	415.380	HBA, 3 × Hydrophobe, NegIoniz	84	114
4	414.142	2 × HBA, 2 × Hydrophobe, NegIoniz	100	115
5	413.198	HBA, 3 × Hydrophobe, NegIoniz	63	112
6	411.793	2 × HBA, 2 × Hydrophobe, NegIoniz	97	111
7	409.753	HBA, 3 × Hydrophobe, NegIoniz	81	112
8	405.967	2 × HBA, 2 × Hydrophobe, NegIoniz	84	110
9	403.440	HBA, 3 × Hydrophobe, NegIoniz	92	111
10	401.977	2 × HBA, 2 × Hydrophobe, NegIoniz	94	112

[a] This ranking is as provided in the log file of the HipHop-REFINE automatic run. The ranking is a measure of how well the active training molecules map onto the proposed pharmacophores, as well as the rarity of the pharmacophore models [42,44,55] (see Section 2.2.2.2 for more details).

[b] Coverage is the number of hits captured by the corresponding pharmacophoric hypothesis upon use as 3D search query against the collected dataset (compounds **1–154**, see Supplementary information for their structures).

polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, etc.) [56].

Furthermore, the training compounds were fitted against the 10 pharmacophore hypotheses returned by the HipHop-REFINE run and their fit values were added as additional descriptors. The fit value for any compound is obtained automatically *via* Eq. (1) [41]:

$$\text{fit} = \sum \text{mapped hypothesis features} \times W\left[1 - \sum \left(\frac{\text{disp}}{\text{tol}}\right)^2\right] \quad (1)$$

where '$\sum$ mapped hypothesis features' represents the number of pharmacophore features that successfully superimpose (i.e., map or overlap with) corresponding chemical moieties within the fitted compound, W is the weight of the corresponding hypothesis feature spheres. This value is fixed to 1.0 in HipHop-generated models, disp is the distance between the center of a particular pharmacophoric sphere (feature centroid) and the center of the corresponding superimposed chemical moiety of the fitted compound; tol is the radius of the pharmacophoric feature sphere (known as tolerance, equals to 1.6 Å by default). $\sum(\text{disp/tol})^2$ is the summation of $(\text{disp/tol})^2$ values for all pharmacophoric features that successfully superimpose corresponding chemical functionalities in the fitted compound.

Genetic function approximation (GFA) was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, i.e., multiple linear regression modeling (MLR). GFA techniques rely on the evolutionary operations of ''crossover and mutation'' to select optimal combinations of descriptors (i.e., chromosomes) capable of explaining bioactivity variation among training compounds from a large pool of possible descriptor combinations, i.e., chromosomes population. Each chromosome is associated with a fitness value that reflects how good it is compared to other solutions. The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF) [57].

Our preliminary diagnostic trials suggested the following optimal GFA parameters: explore linear and quadratic equations of seven terms at mating and mutation probabilities of 50%; population size = 250; number of genetic iterations = 10,000 and LOF smoothness parameter = 1.0. The optimal statistical model was cross-validated automatically employing the leave-one-out procedure implemented in the C2.QSAR module of CERIUS2 [57] QSAR modeling yielded an optimal model (Eq. (2), see Section 3.3) incorporating the fit values of the highest-ranking pharmacophore hypothesis (Hypo 1) (Figs. 1–3).

2.3. In silico screening of the NCI database for new h-PTP 1B Inhibitors

Hypol was employed as 3D search query against the NCI structural database (238,819 compounds) using the ''Best Flexible Database Search'' option within CATALYST. It captured 489 compounds. The hits were subsequently fitted against Hypol using Fast Fit approach of CATALYST, and the lowest 350 hits (of fit values equal or near to zero) were excluded. The remaining hits (139 compounds) were evaluated according to Lipinski's rule of five [58]. However, hits of a maximum of one Lipinski's violation were retained for subsequent processing. To further enhance the drug-likeness of the remaining hits (90 compounds), we excluded compounds of molecular weights higher than 450 Da. The bioactivities of the remaining hits were estimated using QSAR Eq. (2) and the highest-ranking 60 hits were requested from the NCI. However, only five were actually provided by the NCI for subsequent *in vitro* evaluation.

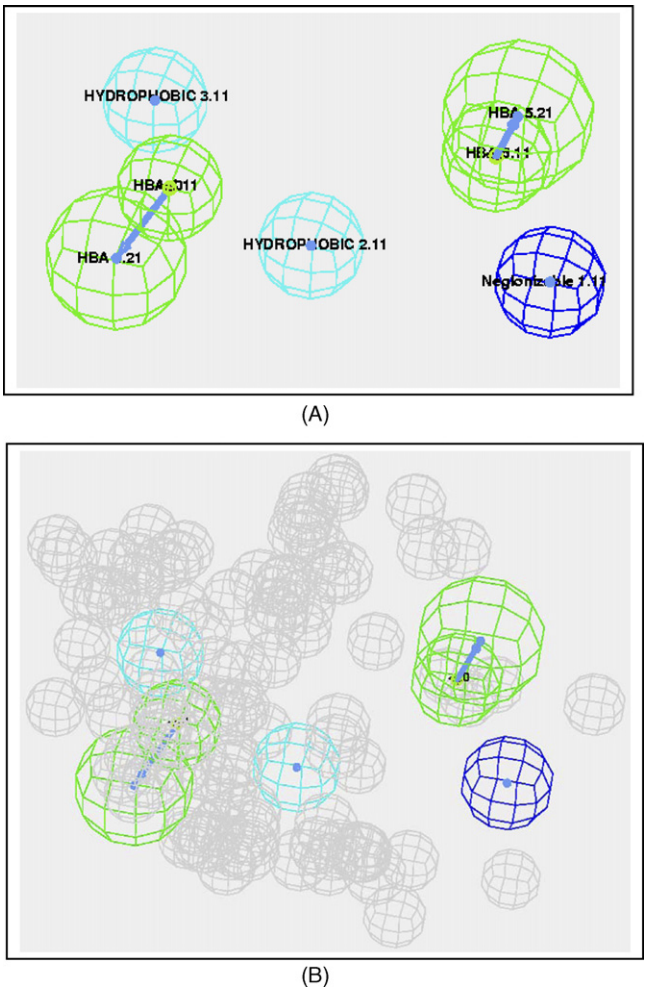

Fig. 1. (A) The pharmacophoric features of Hypo 1 (HBA: hydrogen bond acceptor, NegIonizable: negative ionizable); (B) with added excluded volumes (gray spheres).

2.4. Measurement of h-PTP 1B inhibition

The detection of free phosphate released into the medium by the phosphatase enzyme is based on the classic Malachite green ammonium molybdate method. The assay, as described previously [59] and adapted for the plate reader, was used for the nanomolar detection of liberated phosphate by recombinant h-PTP 1B.

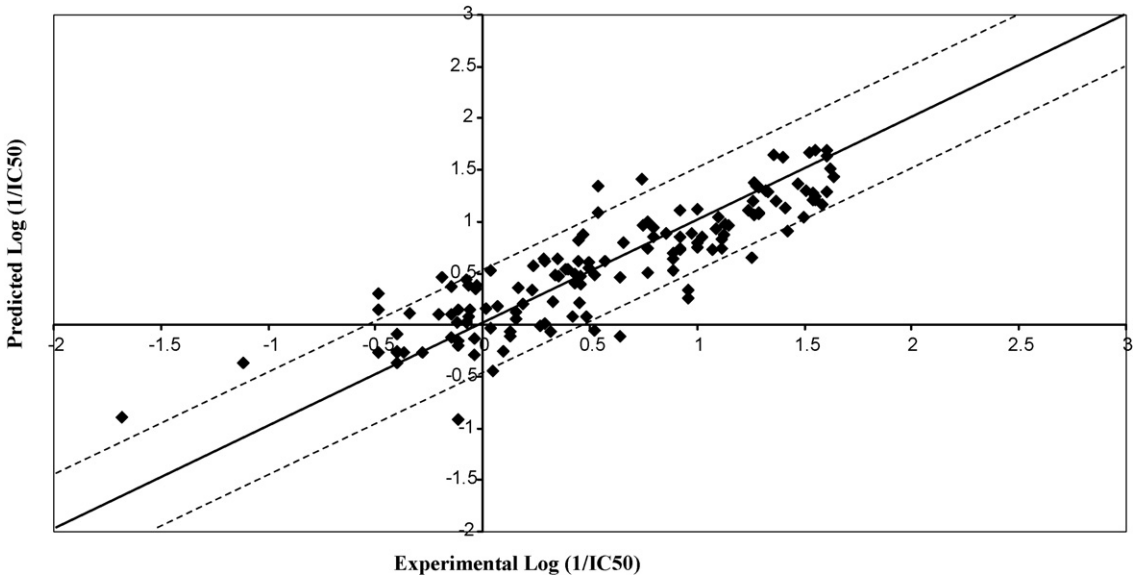

Fig. 2. Experimental *vs.* predicted bioactivities of inhibitors **1–137** as calculated from the optimal QSAR model (Eq. (2)). The solid line is the regression line for the fitted bioactivities, whereas the dotted lines indicate the ±0.5 log point error margins.

The assay used the phosphopeptide Asp-Ala-Asp-Glu-phosphoTyr-Leu-Ile-Pro-Gln-Gln-Gly as substrate. This phosphorylated peptide corresponds to the 988–998 catalytic domain of epidermal growth factor receptor (EGFR), and it is one of the most efficient peptide substrates known for h-PTP 1B [60]. The phosphopeptide was diluted with the assay buffer, pH 7.2, containing 50.0 mM HEPES, 1.0 mM DTT, 1.0 mM EDTA, and 0.05% NP-40 to obtain 150 μM substrate concentration.

The recombinant h-PTP 1B was diluted with the buffer solution pH 7.2, glycerol (50 mg/ml), and of BSA (1 mg/ml) to yield final enzymatic solution of approximate activity of 30 pmol/min. The particular hit compound was dissolved in DMSO and diluted with the assay buffer to generate from 0.020 to 5.00 μM solutions. DMSO concentration was less than 1% in all experiments and controls.

The diluted enzyme (5 μl) was pre-incubated with 35 μl of the assay buffer with or without the tested compound for 15 min at 25 °C. Then the reaction was initiated by the addition of 5 μl of peptide substrate solution to get a final concentration of 150 μM peptide substrate. The mixture was equilibrated to 25 °C and incubated for 30 min. The reaction was terminated by the addition of 100 μl of the malachite green ammonium molybdate–Tween 20 stopping reagent. Negative controls were prepared by adding the substrate after the addition of reaction terminating reagent. The controls were prepared using the pre-incubated recombinant enzyme without the addition of the particular tested compound.

A standard h-PTP 1B inhibitor, RK-682, is included as enzyme activity control for screening inhibition. RK-682 is a potent tyrosine phosphatase inhibitor isolated from *Streptomyces* sp. 88–682 [61]. The color was allowed to develop at room temperature for 30 min, and the sample absorbance was determined at 630 nm using a plate reader (Bio-Tek instruments ELx 800, USA). Samples and blanks were prepared in duplicates.

Inhibition of recombinant h-PTP 1B by a particular compound was calculated as a percent activity of the uninhibited phosphatase control:

$$\text{percent activity} = \frac{\text{absorbance with the tested compound} - \text{negative control}}{\text{absorbance without the tested compound} - \text{negative control}} \times 100\%$$

The percent activity was plotted against the logarithmic transformation of the corresponding inhibitor concentrations for determining the IC_{50} value. The standard inhibitor RK-682 was used with a final concentration of 100 μM and the h-PTP 1B has shown about 20% activity.

3. Results and discussion

3.1. Data mining and conformational coverage

An extensive literature survey was conducted to collect a large and structurally diverse set of published h-PTP 1B inhibitors (**1–154**, see Supplementary information) [49–52]. The 2D structures of the inhibitors were imported into CATALYST and converted automatically into plausible 3D single conformer representations *via* the rule-based methods implemented within CATALYST. The resulting single conformer 3D structures were later used as starting point for CATALYST's conformational analysis and in the determination of various molecular descriptors for QSAR modeling.

The conformational space of each inhibitor was extensively sampled utilizing the poling algorithm employed within the CONFIRM module of CATALYST. Poling promotes

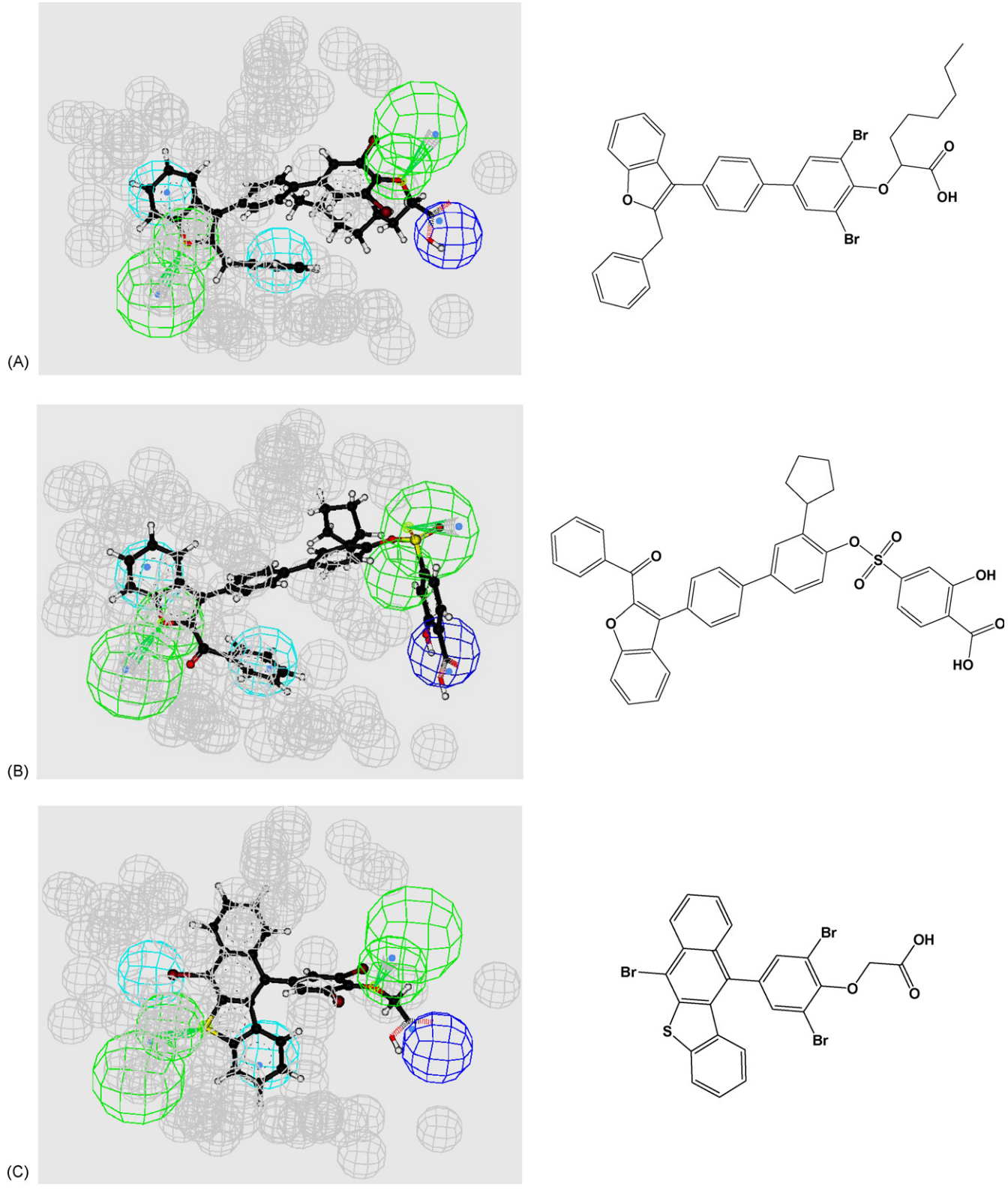

Fig. 3. Hypol mapped against training h-PTP 1B inhibitors (Table 1): (A) **71**; (B) **122**; (C) **139**.

conformational variation *via* employing molecular mechanical force field algorithm that penalizes similar conformers [53,54,64]. Conformational coverage was performed employing the ''best'' module to ensure extensive sampling of conformational space. Efficient conformational coverage guarantees minimum conformation-related noise during pharmacophore generation and validation stages. Pharmacophore generation and pharmacophore-based search procedures are known for their sensitivity to inadequate conformational sampling within the training compounds [62].

3.2. Pharmacophore modeling

CATALYST includes two separate pharmacophore-modeling modules, namely: HypoGen and HipHop. HypoGen enables automatic pharmacophore construction by using a collection of at least 16 molecules with bioactivities spanning over four orders of magnitude [41,63–66]. On the other hand, HipHop generates common feature pharmacophores regardless to the activities of the training compounds [40–42]. Still, both modules generate 3D pharmacophores that can be used as search queries to virtually screen 3D-structural libraries.

However, in the current project, we avoided using HypoGen for two reasons: (i) the collected inhibitors (**1**–**154**, see Supplementary information) originated from several literature sources of different bioassay techniques, and therefore they were not appropriate for conventional 3D-QSAR modelling; (ii) the bioactivity spread of inhibitors from any particular literature source is less than four logarithmic cycles. Therefore, it was decided to employ HipHop to generate common feature pharmacophore(s) for h-PTP 1B inhibitors. Furthermore, we employed the REFINE module to incorporate excluded volumes in the resulting binding hypotheses (see Section 2.2.2.2). Excluded volumes resemble sterically inaccessible regions within the binding site [40–42].

A structurally diverse training subset was carefully selected from the collected compounds for HipHop-REFINE modeling (Table 1, see Section 2.2.2.2). Effective pharmacophore modeling requires training sets of adequate structural diversity to elucidate the common functional features responsible for ligand–receptor affinity across wide ligand diversity. HipoHop-REFINE was instructed to explore up to five-featured pharmacophoric space. Higher-feature pharmacophores were found to be too restrictive and of limited coverage, i.e., they captured limited number of hits upon use as 3D search queries against compounds **1**–**154**. On the other hand, lower-featured pharmacophores are non-selective and tend to capture most (or all) of the 154 inhibitors without effective discrimination against poorly potent inhibitors. Furthermore, by careful evaluation of the binding pocket and active inhibitors, we realized that five-featured pharmacophores can describe the interactions of the vast majority of active inhibitors within h-PTP 1B binding pocket.

The input pharmacophore features were selected in agreement with published SAR studies and crystallographic data. For example, the fact that docking and crystallographic information suggested the involvement of positively charged arginine moieties (i.e., Arg24, Arg47 or Arg221) in salt-bridge formation with some ligands [14,67] prompted us to select negative ionizable (NegIoniz) functionality as possible pharmacophoric feature. In the same manner, the following features were fed into HipHop-REFINE: HBA, HBD, and Hydrophobic.

The software was instructed to explore pharmacophoric models incorporating from zero to three features from any particular selected feature type (i.e., HBA, HBD, and Hydrophobic), except for NegIoniz, which was limited to 1–3. Apparently, potent inhibitors require at least one negative ionizable moiety to interact with the positively charged guanidino side chains of Arg24, Arg47 or Arg221 [14,67] prompting us to force the software to select at least one negative ionizable feature.

Eventually, 10 optimal pharmacophoric hypotheses were generated. However, they all shared comparable features and coverage properties against compounds **1**–**154**. Table 2 shows the pharmacophoric features and coverage values of the generated models. The highest-ranking model (Hypol) is comprised of two hydrophobic areas, two HBAs, one NegIoniz feature and 92 excluded volumes. The exclusion volumes are arranged in such a way to create a hollow cylinder-like pouch with the NegIoniz feature positioned at the rim, as shown in Figs. 1 and 3.

Fig. 3 shows how Hypol fits three active training inhibitors. Evidently, from the figure, the hydrophobic aromatic heads of the three inhibitors (i.e., the benzothiophene of **139**, the benzofuran of **71** and **122**) reside deep into the proposed receptor where they interact with two hydrophobic pockets and one hydrogen-bonding site. On the other hand, the carboxylate moieties of the three compounds seem to interact with certain positively charged residue at the outer rim of the proposed receptor. Collectively, these interactions correlate well with certain reported docked conformers/poses of h-PTP 1B inhibitors that yielded self-consistent CoMFA models [67].

3.3. QSAR modeling

Despite the undisputed significance of pharmacophoric hypotheses in understanding ligand–macromolecule affinity and as 3D search queries, their predictive value as 3D-QSAR models is generally limited by bioactivity-enhancing or -reducing auxiliary groups [66]. Furthermore, HipHop-REFINE models, in particular, are not explicitly (quantitatively) guided by the bioactivities of their corresponding training compounds, i.e., generated by unsupervised modeling [42,44,55], which further reduces their predictive potential. Consequently, to generate a better predictive model, we were prompted to develop a classical QSAR equation incorporating high-ranking pharmacophore(s) and complemented with additional molecular descriptors *via* GFA-MLR analysis. GFA-MLR-guided analysis selects optimal descriptor combinations based on the Darwinian concept of genetic evolution whereby the statistical criteria of regression models from different descriptor combinations (chromosomes) are employed as fitness criteria [57].

However, we limited our QSAR analysis to compounds **1**–**137**, as they were evaluated *via* single bioassay approach. QSAR modeling commenced by enrolling the fit values for compounds **1**–**137**, derived from the 10 generated pharmacophores, together with various physicochemical and molecular descriptors, as independent variables (genes) in GFA-MLR analysis. The idea behind enrolling the 10 models in the genetic competition is to select the best pharmacophore capable of explaining variations in the bioactivities of the training compounds upon combination with other physicochemical descriptors.

The most significant QSAR model was achieved after 10,000 iterations of GFA-MLR analysis [49,50] Unsurprisingly, Hypol emerged in the nine highest-ranking QSAR equations produced by GFA-MLR modeling, which reflects the statistical significance and the overall good 3D qualities of this pharmacophore. The best equation was cross-validated automatically using the leave-one-out cross-validation implemented in CERIUS2 [56,57,68]. Eq. (2) shows the best QSAR model, while Fig. 2 shows the corresponding scatter plot of experimental *versus* estimated bioactivities of the training h-PTP 1B inhibitors. The 95% confidence limits are shown in brackets [±CL]:

$$\begin{aligned}\log(1/\mathrm{IC}_{50}) = {} & -2.037[\pm 0.458] + 0.071[\pm 0.042] \times \mathrm{Hypol} \\ & + 0.195[\pm 0.084] \times \mathrm{AlogP98} \\ & + 0.0253[\pm 0.011] \times \mathrm{SsOH} + 0.327[\pm 0.127] \\ & \times \mathrm{SsssCH} - 0.0067[\pm 0.003] \times \mathrm{SsCl}^2 \\ & + 0.0047[\pm 0.002] \times \mathrm{CHI\text{-}V\text{-}1}^2 \\ r = 0.87,\ & F\text{-statistic} = 69.13,\ r^2_{\mathrm{BS}} = 0.76,\ r^2_{\mathrm{LOO}} = 0.68,\ n \\ & = 137 \end{aligned} \quad (2)$$

where r^2_{LOO} is the leave-one-out correlation coefficient; r^2_{BS} the bootstrapping regression coefficient [57,68]. Hypol represents the fit values of the training compounds against the highest-ranking pharmacophore model as calculated from Eq. (1) (see Section 2.2.2.3). SsOH, SsssCH and SsCl are the sums of all topological E-state values of hydroxyl, trivalent carbon and chloro substitutions, respectively. CHI-V-1 is the first-order valence connectivity index. AlogP98 is the calculated logarithm of partition coefficient from the implementation of the atom-type-based method using the latest published set of parameters [56].

The statistical significance of individual descriptors in Eq. (2) is evident from their corresponding *F*-values (Fischer statistic) and *p*-values (probability of significance at 95% confidence) (shown in Table 3). *F*-values reflect the significance of individual descriptors in independently explaining variations in the response variable; while on the other hand, *p*-values reflect their statistical significances when combined together within an overall collective QSAR model (i.e., descriptors' interactions) [73]. Explanatory variables (i.e., descriptors) of *F*-values above 4 and/or *p*-values below 0.05 (95% confidence) are usually considered statistically significant in a particular model, which means that their influence on the response variable (bioactivity) is not merely by chance [73]. In some instances, a particular descriptor might be individually insignificant in explaining the response variable, however, the same descriptor can gain statistical significance upon interaction with other descriptors within a collective QSAR model [73]. For example, the term SsssCH in Eq. (2) is not singly statistically significant, as evident from its corresponding *F*-value (1.01, Table 3), while it became statistically significant upon interaction with other descriptors in Eq. (2), as evident from its corresponding *p*-value (1.18×10^{-6}, Table 3).

Eq. (2) and Table 3 illustrate an evident role played by the partition coefficient in ligand binding to h-PTP 1B. The significant positive contribution of log(*P*) in $\log(1/\mathrm{IC}_{50})$ suggests that hydrophobic ligands exhibit superior inhibitory activities. This trend is probably related to the fact that hydrophobic ligands favor migration from water to hydrophobic pockets, such as h-PTP 1B binding site, irrespective to their ability to optimally fit the corresponding pharmacophoric model.

Similarly, Eq. (2) and Table 3 suggest significant role played by the first-order valence connectivity index (CHI-V-1) in ligand/h-PTP 1B binding affinity. This descriptor is strongly related to the molecular length of the ligands [56]. Therefore, the significant positive contribution of CHI-V-1 in $\log(1/\mathrm{IC}_{50})$ suggests that extended long ligands exhibit superior inhibitory activities, probably due to the longitudinal nature of h-PTP 1B binding pocket (Fig. 5). Emergence of additional non-pharmacophoric descriptors in the optimal QSAR model emphasizes the fact that ligand–receptor is modulated by additional factors in excess to pharmacophore recognition.

Finally, it remains to be mentioned that the overall statistical qualities of Eq. (2) is comparable to those reported in two key QSAR studies conducted for h-PTP 1B inhibitors. In one instance, structure-based CoMFA modeling was attempted for 92 h-PTP 1B inhibitors employing a combination of CoMFA descriptors and log *P* [30]. This attempt resulted in an optimal model of $r^2 = 0.84$ and $r^2_{\mathrm{LOO}} = 0.75$. In a second instance, several docking/scoring combinations were explored to achieve optimal protein-aligned CoMFA models for 110 h-PTP 1B inhibitors. The optimal docking/scoring configuration furnished a CoMFA model of $r^2 = 0.80$ and $r^2_{\mathrm{LOO}} = 0.65$ [67].

Nevertheless, the fact that Eq. (2) incorporates a pharmacophoric component (Hypo 1 fit values), which can be used as 3D search query, is a significant advantage over the two aforementioned studies.

Table 3
The statistical criteria of individual descriptors in QSAR Eq. (2)

Descriptor	*F*-value[a]	*p*-Value[b]
Hypol (fit)	34.48	1.12×10^{-3}
AlogP98	133.71	1.12×10^{-5}
SsOH	10.50	8.70×10^{-6}
SsssCH	1.01	1.18×10^{-6}
SsCl^2	9.40	4.34×10^{-5}
$\mathrm{CHI\text{-}V\text{-}1}^2$	219.27	1.51×10^{-6}

[a] Determined for regression models incorporating each individual descriptor separately against $\log(1/\mathrm{IC}_{50})$.
[b] The probability of significance values at 95% confidence determined for different descriptors as they interact within Eq. (2).

3.4. In silico screening of NCI database and subsequent in vitro evaluation

Hypol was employed as 3D search query to screen the NCI database (238,819 compounds). The search process took nearly 3 h and it captured 489 hits. Hits are defined as those

compounds that possess chemical groups that spatially overlap (map) with corresponding features within the pharmacophoric model. The hits were subsequently fitted against Hypol (see the fit Eq. (1) in Section 2.2.2.3), and those of fit values ≈0 were excluded from subsequent processing. A low fit value indicates that although the chemical functions of the particular hit(s) overlap with the corresponding pharmacophoric features, the centers of the functional groups (of the hit) are displaced from the centers of the corresponding pharmacophoric features, such that the term $\sum(\text{disp/tol})^2$ in Eq. (1) (Section 2.2.2.3)

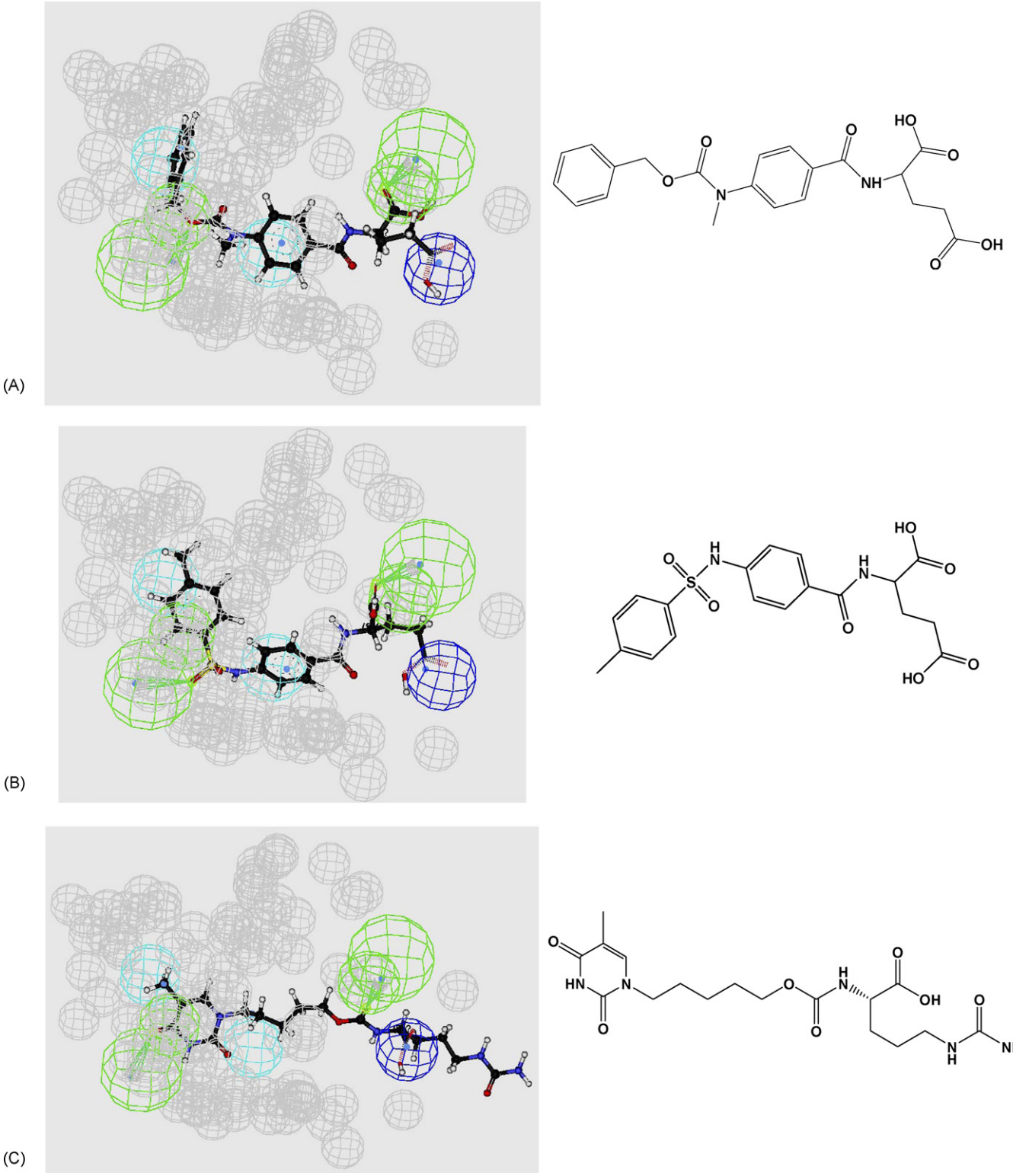

Fig. 4. Hypol mapped against captured hits (**A**) **155** (IC_{50} = 3.30 μM); (B) **156** (IC_{50} = 3.17 μM); (C) **157** (IC_{50} = 0.76 μM); (D) **158** (IC_{50} = 0.47 μM); (**E**) **159** (IC_{50} = 1.86 μM).

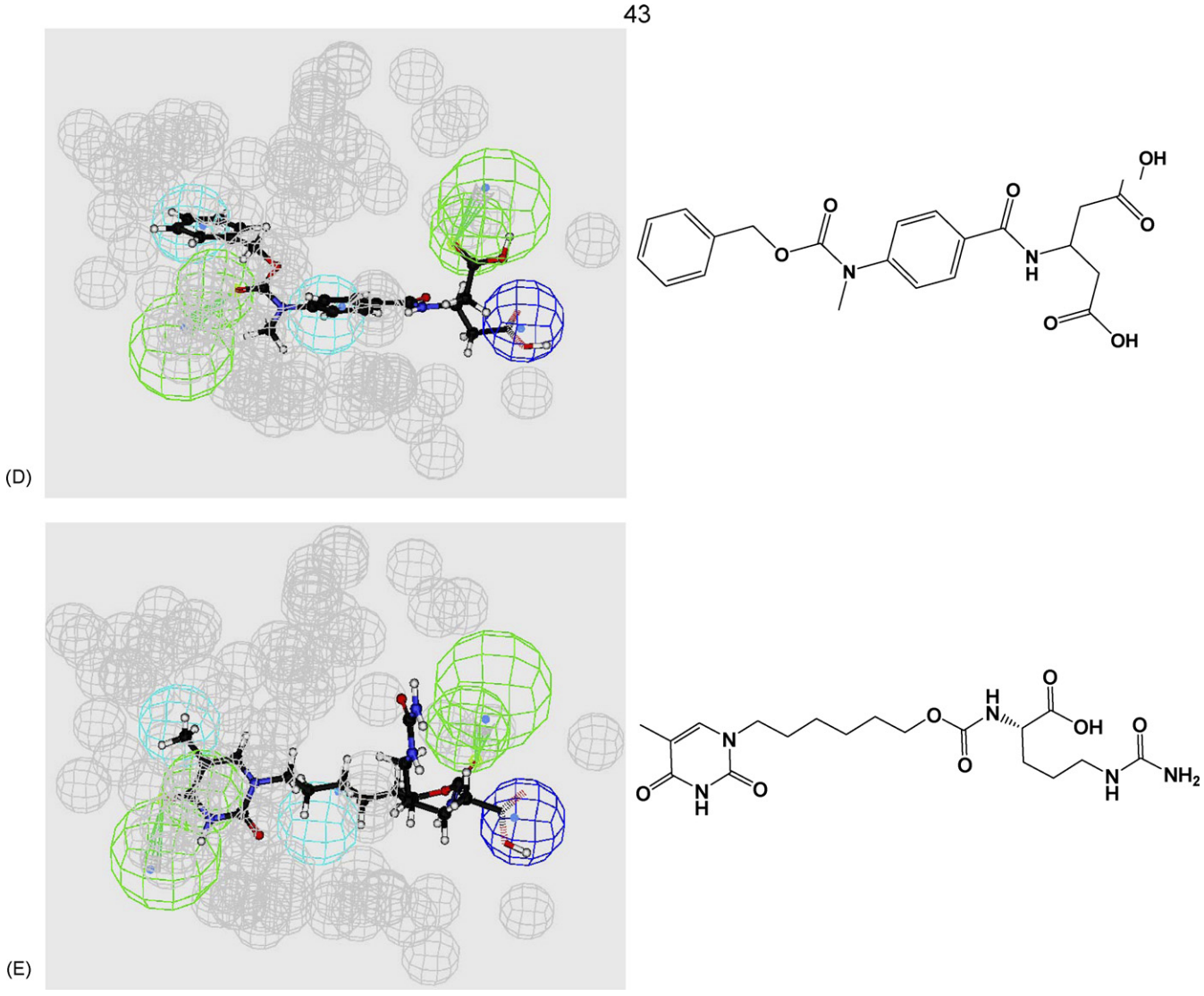

Fig. 4. (*Continued*).

approaches a value of 1.0 and the overall fit value approaches 0. Therefore, excluding poor fitters should improve the success rates of the retained hits. The remaining hits (139 compounds) were filtered according to Lipinski's "rule of five" to retain drug-like molecules [58]. However, we allowed a maximum of one Lipinski's violation to enrich the chemical diversity of the selected hits, particularly under the limiting effects of the 92 exclusion volumes in Hypol. This selection retained 90 hit molecules. Subsequently, the bioactivities of remaining molecules were estimated using QSAR Eq. (2). The highest-ranking hits (60 compounds) were requested from the NCI. However, only five compounds were provided for subsequent *in vitro* testing. Interestingly, all five compounds were found to possess potent *in vitro* inhibitory actions against h-PTP 1B. The fact that the tested hits ranked (based on their predicted bioactivities) within the lowest third of the requested list of NCI compounds (Table 4) suggests that the remaining unavailable hits should have potent inhibitory activities, which highlights the success of our overall strategy of pharmacophore modeling, QSAR analysis and *in silico* screening.

Table 4 shows the NCI codes of the tested hits and their corresponding estimated and experimental anti h-PTP 1B bioactivities. However, Fig. 4 shows how Hypol fits the tested compounds.

From Table 4, one can quickly conclude that Eq. (2) underestimated the anti h-PTP 1B bioactivities of the tested

Table 4
Active hits captured by Hypol and their corresponding QSAR estimates (from Eq. (2)) and *in vitro* bioactivities

Compound	NCI code	Predicted $\log(1/IC_{50})$	Rank of Hit[a]	Experimental[b]	
				$\log(1/IC_{50})$	IC_{50} (μM)
155	112,855	−1.186	40	−0.52	3.30
156	211,015	−1.230	42	−0.50	3.17
157	691,234	−1.765	60	0.12	0.76
158	112,854	−1.127	38	0.33	0.47
159	691,235	−1.638	56	−0.27	1.86

[a] Rank order of hit molecules among the best 60 hits requested from the NCI (see Sections 2.3 and 3.4).
[b] The IC_{50} values are the average of two trials.

hits. This error is probably related to two factors: (i) the hits were evaluated using a different bioassay technique from that used for assessing the bioactivities of the training inhibitors (i.e., **1**–**137**, see Supplementary information). In particular, we tested the hits using a method that relies on a phosphorylated peptide corresponding to the 988–998 catalytic domain of epidermal growth factor receptor (EGFR) as h-PTP 1B substrate [60]. On other hand, the training inhibitors (**1**–**137**) were evaluated employing a dodecaphosphopeptide corresponding to the 1142–1153 catalytic domain of the insulin receptor [49,50]. (ii) Generally, QSAR equations have limited extrapolatory predictive performance, i.e., when used to predict bioactivities of scaffolds outside the chemical space of their training compounds [69].

3.5. Comparison of Hypol with the binding site of h-PTP 1B

Despite the problems associated with crystallographic structures (see the introduction) [36–38], the pharmacophore features obtained by HipHop-REFINE can be compared with

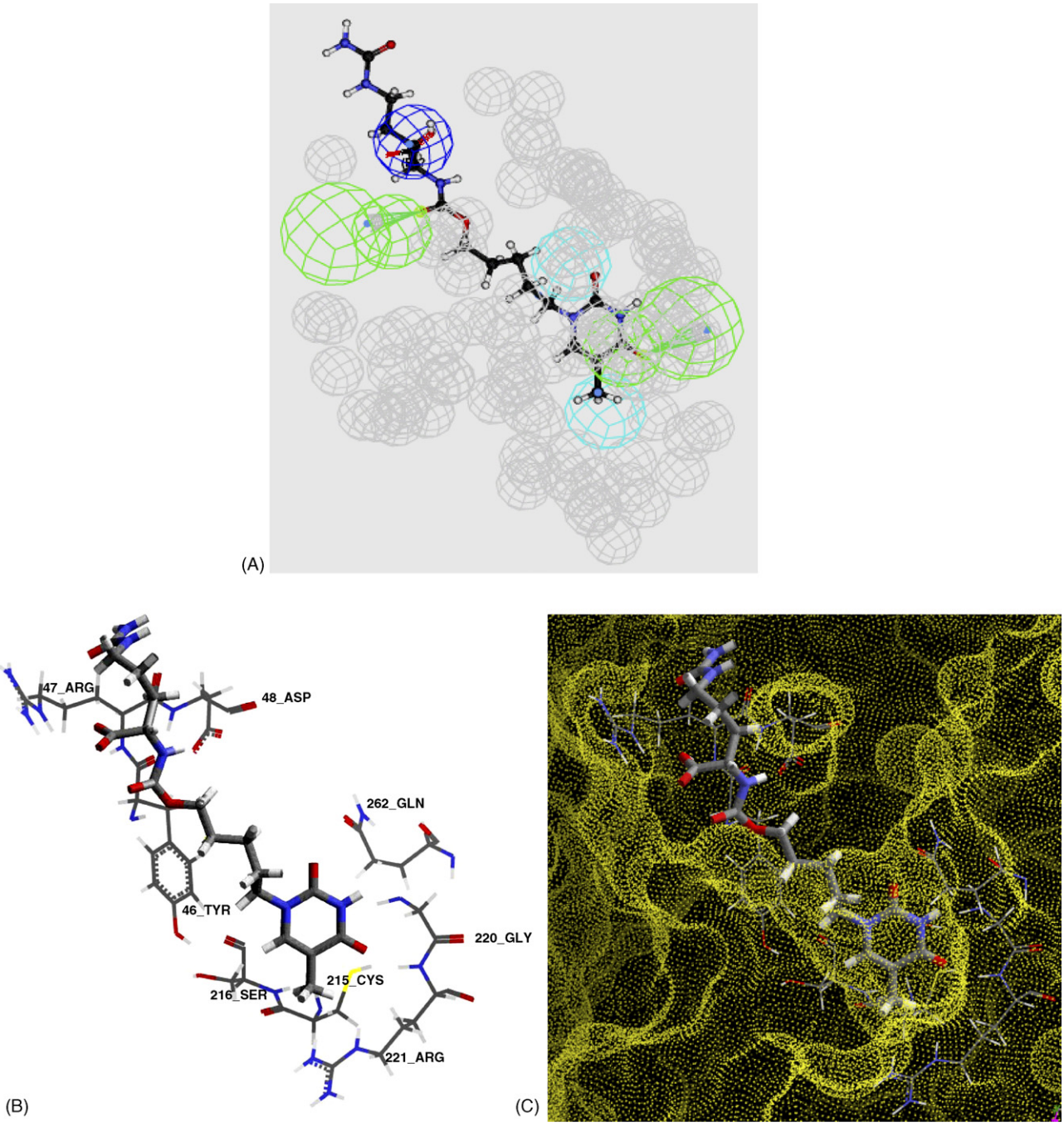

Fig. 5. (A) Hypol mapped against **157**; (B) **157** docked into the binding site of h-PTP 1B (PDB code: 1G7F, resolution = 1.8 Å). The docking experiment was performed employing Cerius2.LIGANDFIT docking engine (from Accelrys Inc.) and PLP1 scoring function. Hydration water molecules and protein hydrogen atoms were hidden for clarity. (C) Connolly's water-accessible surface (dotted yellow surface) and its spatial relationship to the docked inhibitor 157.

the structure of h-PTP 1B binding site to identify probable residues important for inhibition. The features in Hypol, as well as the alignment of compound **157** as proposed by Hypol, were compared with the structure of **157** as it docks into the binding site of h-PTP 1B (Fig. 5). The docking experiment was performed employing LIGANDFIT docking engine [70,71] and PLP1 scoring function [67,72] and *via* default parameters.

A marked similarity was observed between the features and steric constrains proposed by the pharmacophore model and the ligand binding features in the docked structure. The docking study suggests that the pyrimidinedione head of **157** is hydrogen-bonded to the amidic N–H of Arg221 *via* its 4-oxo functionality (Fig. 5b). This interaction perfectly corresponds to a HBA feature in Hypol mapping the same carbonyl group (Fig. 5a). Furthermore, the methyl substituent of the pyrimidinedione is docked at close proximity to the thiol moiety of Cys216 (Fig. 5b), which is consistent with mapping the methyl group by the innermost hydrophobic feature in Hypol (Fig. 5a).

The pentylene chain of **157** is fitted against one of the hydrophobic features in Hypol (Fig. 5a), which seems to correspond to a proposed hydrophobic/van der Waals' interaction between the five-carbon chain and the aromatic ring of Tyr46 in the docked model (Fig. 5b and c).

The mapping of the carboxylic acid moiety of **157** against the NegIoniz feature of Hypol is highly consistent with a proposed electrostatic attraction between the carboxylate of **157** and the ionized guanidino side chain of Arg47 in the docking model. Finally, mapping of the carbamate carbonyl of **157** by a HBA feature in Hypol suggests the existence of corresponding hydrogen bonding to certain feature in the binding pocket. However, this proposition seems not to be supported by any analogous interaction in docking model. Still, we believe that the flexibility of Arg47 side chain allows the guanidine group to approach the carbamate carbonyl of **157** causing their mutual hydrogen bonding.

4. Conclusion

This work includes pharmacophore modeling of h-PTP 1B inhibitors utilizing HipHop-REFINE. The best binding hypothesis was incorporated into a statistically significant QSAR equation, and it was subsequently used as 3D search query to screen the NCI database for new h-PTP 1B inhibitors. The resulting hits were filtered according to Lipinski's rule of five and evaluated using the QSAR equation. Five of the high-ranking hits were acquired and were found to possess nanomolar to low micromolar inhibitory IC_{50} values against h-PTP 1B. The features and steric properties of the pharmacophoric model have striking resemblance to the binding site of h-PTP 1B. Pharmacophore models have the advantages of avoiding docking to crystallographic protein structures and shorter search times compared to docking techniques.

Acknowledgments

This project was sponsored by the Deanship of Scientific Research at the University of Jordan (grant no. 19/2003-2004). The authors wish to thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for providing funds towards purchasing O2 and Octane2 Sgi workstations and CATALYST and CERIUS2 software packages. The authors also wish to thank the National Cancer Institute for freely providing hit compounds for experimental validation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2006.08.008.

References

[1] H. King, R. Aubert, W. Herman, Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections, Diabetes Care 21 (1998) 1414–1431.
[2] P. Zimmet, K.G.M.M. Alberti, J. Shaw, Global and societal implications of the diabetes epidemic, Nature 414 (2001) 782–787.
[3] J. Montalibet, B.P. Kennedy, Therapeutic strategies for targeting h-PTP 1B in diabetes, Drug Discov. Today: Therap. Strat. 2 (2005) 129–135.
[4] J.F. Youngren, I.D. Goldfine, The molecular basis of insulin resistance, Sci. Med. 4 (1997) 18–27.
[5] R.K. Vats, V. Kumar, A. Kothari, A. Mital, U. Ramachandran, Emerging targets for diabetes, Curr. Sci. 88 (2005) 241–249.
[6] K.A. Kenner, E. Anyanwu, J.M. Olefsky, J. Kusari, Protein-tyrosine phosphatase 1B is a negative regulator of insulin- and insulin-like growth factor-I-stimulated signaling, J. Biol. Chem. 271 (1996) 19810–19816.
[7] B.J. Goldstein, F. Ahmad, W. Ding, P.-M. Li, W.-R. Zhang, Regulation of the insulin signaling pathway by cellular protein–tyrosine phosphatases, Mol. Cell. Biochem. 182 (1998) 91–99.
[8] B.P. Kennedy, C. Ramachandran, Protein tyrosine phosphatase-1B in diabetes, Biochem. Pharmacol. 60 (2000) 877–883.
[9] M. Elchebly, P. Payette, E. Michaliszyn, W. Cromlish, S. Collins, A.L. Loy, D. Normandin, A. Cheng, J.H. Hagen, C.-C. Chan, C. Ramachandran, M.J. Gresser, M.L. Tremblay, B.P. Kennedy, Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene, Science 283 (1999) 1544–1548.
[10] M. Na, W.K. Oh, Y.H. Kim, X.F. Cai, S.H. Kim, B.Y. Kim, J.S. Ahn, Inhibition of protein tyrosine phosphatase 1B by diterpenoids isolated from *Acanthopanax koreanum*, Bioorg. Med. Chem. Lett. 16 (2006) 3061–3064.
[11] L.D. Klaman, O. Boss, O.D. Peroni, J.K. Kim, J.L. Martino, J.M. Zabolotny, N. Moghal, M. Lubkin, Y.-B. Kim, A.H. Sharpe, A. Stricker-Krongrad, G.I. Shulman, B.G. Neel, B.B. Kahn, Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein–tyrosine phosphatase 1B-deficient mice, Mol. Cell. Biol. 20 (2000) 5479–5489.
[12] J.N. Andersen, O.H. Mortensen, G.H. Peters, P.G. Drake, L.F. Iversen, O.H. Olsen, P.G. Jansen, H.S. Andersen, N.K. Tonks, N.P.H. Moller, Structural and evolutionary relationships among protein tyrosine phosphatase domains, Mol. Cell. Biol. 21 (2001) 7117–7136.
[13] A.P. Combs, E.W. Yue, M. Bower, P.J. Ala, B. Wayland, B. Douty, A. Takvorian, P. Polam, Z. Wasserman, W. Zhu, M.L. Crawley, J. Pruitt, R. Sparks, B. Glass, D. Modi, E. McLaughlin, L. Bostrom, M. Li, L. Galya, K. Blom, M.I. Hillman, L. Gonneville, B.G. Reid, M. Wei, M. Becker-Pasha, R. Klabe, R. Huber, Y. Li, G. Hollis, T.C. Burn, R. Wynn, P. Liu, B. Metcalf, Structure-based design and discovery of protein tyrosine phosphatase inhibitors incorporating novel isothiazolidinone heterocyclic phosphotyrosine mimetics, J. Med. Chem. 48 (2005) 6544–6548.
[14] T.O. Johnson, J. Ermolieff, M.R. Jirousek, Protein tyrosine phosphatase 1B inhibitors for diabetes, Nat. Rev. Drug Discov. 1 (2002) 696–709.
[15] R. Hooft van Huijsduijnen, S. Walchli, M. Ibberson, A. Harrenga, Protein tyrosine phosphatases as drug targets: PTP1B and beyond, Expert Opin. Ther. Targets 6 (2002) 637–647.

[16] X. Espanel, S. Walchli, R. Pescini-Gobert, M. El Alama, M.-L. Curchod, N. Gullu-Isler, R. Hooft van Huijsduijnen, Pulling strings below the surface. Hormone receptor signaling through inhibition of protein tyrosine phosphatases, Endocrine 15 (2001) 19–28.
[17] D. Barford, A.J. Flint, N.K. Tonks, Crystal structure of human protein tyrosine phosphatase 1B, Science 263 (1994) 1397–1404.
[18] M.R. Groves, Z.-J. Yao, P.P. Roller, T.R. Burke Jr., D. Barford, Structural basis for inhibition of the protein tyrosine phosphatase 1B by phosphotyrosine peptide mimetics, Biochemistry 37 (1998) 17773–17783.
[19] P. Huang, J. Ramphal, J. Wei, C. Liang, B. Jallal, G. McMahon, C. Tang, Structure-based design and discovery of novel inhibitors of protein tyrosine phosphatases, Bioorg. Med. Chem. 11 (2003) 1835–1849.
[20] Y.A. Puius, Y. Zhao, M. Sullivan, D.S. Lwarence, S.C. Almo, Z.Y. Zhang, Identification of a second aryl phosphate-binding site in protein–tyrosine phosphatase 1B: a paradigm for inhibitor design, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 13420–13425.
[21] G. Scapin, S.B. Patel, J.W. Becker, Q. Wang, C. Desponts, D. Waddleton, K. Skorey, W. Cromlish, C. Bayly, M. Therien, J.Y. Gauthier, C.S. Li, C.K. Lau, C. Ramachandran, B.P. Kennedy, E. Asante-Appiah, The structural basis for the selectivity of benzotriazole inhibitors of PTP 1B, Biochemistry 42 (2003) 11451–11459.
[22] E. Asante-Appiah, S. Patel, C. Dufresne, P. Roy, Q. Wang, V. Patel, R.W. Friesen, C. Ramachandran, J.W. Becker, Y. Leblanc, B.P. Kennedy, G. Scapin, The structure of PTP-1B in complex with a peptide inhibitor reveals an alternative binding mode for bisphosphonates, Biochemistry 41 (2002) 9043–9051.
[23] M.A. Blaskovich, H.-O. Kim, Recent discovery and development of protein tyrosine phosphatase inhibitors, Expert. Opin. Ther. Patent 12 (2002) 871–905.
[24] S.D. Taylor, Inhibitors of protein tyrosine phosphatase IB (PTP1B), Curr. Top. Med. Chem. 3 (2003) 759–782.
[25] G. Liu, Protein tyrosine phosphatase 1B inhibition: opportunities and challenges, Curr. Med. Chem. 10 (2003) 1407–1421.
[26] E. Black, J. Breed, A.L. Breeze, K. Embrey, R. Garcia, T.W. Gero, L. Godfrey, P.W. Kenny, A.D. Morley, C. Minshull, A.D. Pannifer, J. Read, A. Rees, D.J. Russell, D. Toader, J. Tucker, Structure-based design of protein tyrosine phosphatase-1B inhibitors, Bioorg. Med. Chem. Lett. 15 (2005) 2503–2507.
[27] I.K. Lund, H.S. Andersen, L.F. Iversen, O.H. Olsen, K.B. Møller, A.K. Pedersen, Y. Ge, D.D. Holsworth, M.J. Newman, F.U. Axe, N.P.H. M0ller, Structure-based design of selective and potent inhibitors of protein–tyrosine phosphatase β, J. Biol. Chem. 279 (2004) 24226–24235.
[28] T.N. Doman, S.L. McGovern, B.J. Witherbee, T.P. Kasten, R. Kurumbail, W.C. Stallings, D.T. Connolly, B.K. Shoichet, Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B, J. Med. Chem. 45 (2002) 2213–2221.
[29] P. Huang, J. Ramphal, J. Wei, C. Liang, B. Jallal, G. McMahon, C. Tang, Structure-based design and discovery of novel inhibitors of protein tyrosine phosphatases, Bioorg. Med. Chem. 11 (2003) 1835–1849.
[30] V.S. Murthy, V.M. Kulkarni, 3D-QSAR CoMFA and CoMSIA on protein tyrosine phosphatase 1B inhibitors, Bioorg. Med. Chem. 10 (2002) 2267–2282.
[31] M.E. Sobhia, P.V. Bharatam, Comparative molecular similarity indices analysis (CoMSIA) studies of 1,2-naphthoquinone derivatives as PTP1B inhibitors, Bioorg. Med. Chem. 13 (2005) 2331–2338.
[32] W. Sippl, Development of biologically active compounds by combining 3D QSAR and structure-based design methods, J. Comput.-Aid. Mol. Des. 16 (2002) 825–830.
[33] R. Wang, Y. Lu, S. Wang, Comparative evaluation of 11 scoring functions for molecular docking, J. Med. Chem. 46 (2003) 2287–2303.
[34] E.M. Krovat, T. Langer, Impact of scoring functions on enrichment in docking-based virtual screening: an application study on renin inhibitors, J. Chem. Inf. Comput. Sci. 44 (2004) 1123–1129.
[35] M. Feher, Consensus scoring for protein–ligand interactions, Drug Discov. Today 11 (2006) 421–428.
[36] N.R.A. Beeley, C. Sage, GPCRs: an update on structural approaches to drug discovery, Targets 2 (2003) 19–25.
[37] B. Waszkowycz, in: A.L. Harvey (Ed.), Advances in Drug Discovery Techniques, John Wiley & Sons, Chichester, 1998, pp. 150–153.
[38] M.A. DePristo, P.I.W. de Bakker, T.L. Blundell, Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography, Structure 12 (2004) 831–838.
[39] M. Akamatsu, Current state and perspectives of 3D-QSAR, Curr. Top. Med. Chem. 12 (2002) 1381–1394.
[40] in: Proceedings of the 9th European CATALYST User Group Meeting, Advanced Seminars in CATALYSTFrankfurt, Germany, March 24, 2006, Accelrys Inc., San Diego, CA, 2006.
[41] Catalyst User Guide, Accelrys Software Inc., San Diego, 2005.
[42] O.O. Clement, A.T. Mehl, in: F.O. Guner (Ed.), Pharmacophore Perception, Development, and Use in Drug Design—IUL Biotechnology Series, International University Line, La Jolla, California, 2000, pp. 71–84.
[43] P.W. Sprague, R. Hoffmann, in: H. Van de Waterbeemd, B. Testa, G. Folkers (Eds.), Computer Assisted Lead Finding and Optimization—Current Tools for Medicinal Chemistry, VHCA, Basel, 1997, pp. 230–240.
[44] D. Barnum, J. Greene, A. Smellie, P. Sprague, Identification of common functional configurations among molecules, J. Chem. Inf. Comput. Sci. 36 (1996) 563–571.
[45] J. Singh, C.E. Chuaqui, P.A. Boriack-Sjodin, W.-C. Lee, T. Pontz, M.J. Corbley, H.-K. Cheung, R.M. Arduini, J.N. Mead, M.N. Newman, J.L. Papadatos, S. Bowes, S. Josiah, L.E. Ling, Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFβ receptor kinase (TβRI), Bioorg. Med. Chem. Lett. 13 (2003) 4355–4359.
[46] M.O. Taha, A.M. Qandil, D.D. Zaki, M.A. AlDamen, Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling, Eur. J. Med. Chem. 40 (2005) 701–727.
[47] P.A. Keller, M. Bowman, K.H. Dang, J. Garner, S.P. Leach, R. Smith, A. McCluskey, Pharmacophore development for corticotropin-releasing hormone: new insights into inhibitor activity, J. Med. Chem. 42 (1999) 2351–2357.
[48] R.G. Karki, V.M. Kulkarni, A feature based pharmacophore for *Candida albicans* MyristoylCoA: protein *N*-myristoyltransferase inhibitors, Eur. J. Med. Chem. 36 (2001) 147–163.
[49] M.S. Malamas, J. Sredy, C. Moxham, A. Katz, W. Xu, R. McDevitt, F.O. Adebayo, D.R. Sawicki, L. Seestaller, D. Sullivan, J.R. Taylor, Novel benzofuran and benzothiophene biphenyls as inhibitors of protein tyrosine phosphatase 1B with antihyperglycemic properties, J. Med. Chem. 43 (2000) 1293–1310.
[50] M.S. Malamas, J. Sredy, I. Gunawan, B. Mihan, D.R. Sawicki, L. Seestaller, D. Sullivan, B.R. Flam, New azolidinediones as inhibitors of protein tyrosine phosphatase 1B with antihyperglycemic properties, J. Med. Chem. 43 (2000) 995–1010.
[51] M. Sarmiento, L. Wu, Y. Keng, L. Song, Z. Luo, Z. Huang, G. Wu, A.K. Yuan, Z. Zhang, Structure-based discovery of small molecule inhibitors targeted to protein tyrosine phosphatase 1B, J. Med. Chem. 43 (2000) 146–155.
[52] J. Wrobel, J. Sredy, C. Moxham, A. Dietrich, Z. Li, D.R. Sawicki, L. Seestaller, L. Wu, A. Katz, D. Sullivan, C. Tio, Z. Zhang, 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 (1999) 3199–3202.
[53] A. Smellie, S.D. Kahn, S.L. Teig, Analysis of conformational coverage. Part 1. Validation and estimation of coverage, J. Chem. Inf. Comput. Sci. 35 (1995) 285–294.
[54] A. Smellie, S. Teig, P. Towbin, Poling: promoting conformational variation, J. Comp. Chem. 16 (1995) 171–187.
[55] HipHop User Guide Version 3.1, Catalyst 4.10, Accelrys Inc., San Diego, CA, 2005 (http://www.accelrys.com/doc/life/catalyst410/help/hipHop/HipHop_23TOC.doc.htm).
[56] Cerius 2 QSAR Users' Manual, Accelrys Inc., San Diego, CA, 2005, pp. 43–88.
[57] Cerius 2 QSAR Users' Manual, Accelrys Inc., San Diego, CA, 2005, pp. 221–235.
[58] C.A. Lipinski, Lead- and drug-like compounds: the rule-of-five revolution, Drug Discov. Today: Technol. 1 (2004) 337–341.

[59] P.A. Lanzetta, L.J. Alvarez, P.S. Reinach, O.A. Candia, An improved assay for nanomole amounts of inorganic phosphate, Anal. Biochem. 100 (1979) 95–97.
[60] Z.Y. Zhang, A.M. Thieme-Sefler, D. Maclean, D.J. McNamara, E.M. Dobrusin, T.K. Sawyer, J.E. Dixon, Substrate specificity of the protein tyrosine phosphatases, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 4446–4450.
[61] T. Hamaguchi, T. Sudo, H. Osada, RK-682, a potent inhibitor of tyrosine phosphatase, arrested the mammalian cell cycle progression at G1 phase, FEBS Lett. 372 (1995) 54–58.
[62] R.P. Sheridan, S.K. Kearsley, Why do we need so many chemical similarity search methods? Drug Discov. Today 7 (2002) 903–911.
[63] Y. Kurogi, O.F. Güner, Pharmacophore modeling and three-dimensional database searching for drug design using catalyst, Curr. Med. Chem. 8 (2001) 1035–1055.
[64] H. Li, J. Sutter, R. Hoffmann, in: O.F. Güner (Ed.), Pharmacophor Perception, Development, and Use in Drug Design, International University Line, California, 2000, pp. 173–189.
[65] J. Sutter, O. Giiner, R. Hoffmann, H. Li, M. Waldman, in: O.F. Giiner (Ed.), Pharmacophor Perception, Development, and Use in Drug Design, International University Line, California, 2000, pp. 501–511.
[66] I.B. Bersuker, S. Bahceci, J.E. Boggs, in: O.F. Giiner (Ed.), Pharmacophor Perception, Development, and Use in Drug Design, International University Line, California, 2000, pp. 457–473.
[67] M.O. Taha, M.A. AlDamen, 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. 10 (2005) 8016–8034.
[68] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, Quant. Struct.–Act. Relat. Comb. Sci. 22 (2003) 69–77.
[69] M.T.D. Cronin, T.W. Schultz, Pitfalls in QSAR, J. Mol. Struct. (Theochem) 622 (2003) 39–51.
[70] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites, J. Mol. Graph. Modell. 21 (2003) 289–307.
[71] CERIUS2 4.10 LigandFit User Manual, Accelrys Inc., San Diego, CA, 2005, pp. 3–48.
[72] D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel, S.T. Freer, Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming, Chem. Biol. 2 (1995) 317–324.
[73] L.F. Ramsey, W.D. Schafer, The Statistical Sleuth, 1st ed., Wadsworth Publishing Company, USA, 1997.
[74] M.O. Taha, A.G. Al-Bakri, W.A. Zalloum, Discovery of potent inhibitors of pseudomonal quorum sensing via pharmacophore modeling and in-silico screening. Bioorg. Med. Chem. Lett., in press.