



## Studies on two types of PTP1B inhibitors for the treatment of type 2 diabetes: Hologram QSAR for OBA and BBB analogues

Yuanhua Cheng<sup>a</sup>, Mei Zhou<sup>b</sup>, Chen-Ho Tung<sup>a</sup>, Mingjuan Ji<sup>b,\*</sup>, Fushi Zhang<sup>a,\*</sup>

<sup>a</sup>Key Laboratory of Organic Optoelectronics and Molecular Engineering of Ministry of Education, Department of Chemistry, Tsinghua University, Beijing 100084, PR China

<sup>b</sup>College of Chemistry and Chemical Engineering, Graduate University of Chinese Academy of Sciences, Beijing 100049, PR China

### ARTICLE INFO

#### Article history:

Received 10 December 2009

Revised 3 March 2010

Accepted 10 April 2010

Available online 7 May 2010

#### Keywords:

HQSAR

PTP1B

Inhibitors

OBA

BBB

### ABSTRACT

Hologram quantitative structure–activity relationships (HQSAR) analysis were conducted on two series of PTP1B inhibitors, 39 2-(oxalylamino) benzoic acid (OBA) analogues and 60 benzofuran and benzothio-phenyl biphenyls (BBB) analogues. The optimal HQSAR model of the OBA analogue has  $q^2 = 0.592$  and  $r^2 = 0.940$ , while the optimal HQSAR model for the BBB analogues shows  $q^2 = 0.667$  and  $r^2 = 0.863$ . Two models were employed to predict the biological activities of two test sets. For OBA analogues, the optimal model was validated by an external test set of six compounds with satisfactory predictive  $r^2$  value of 0.786. For BBB analogues, the optimal model shows satisfactory predictive  $r^2$  value of 0.866 for an external test set of 10 compounds. The contribution maps derived from the optimal HQSAR models are consistent with the biological activities of the studied compounds. Two virtual combinatorial libraries were designed and screened by the optimal HQSAR models and potential candidates with high predictive biological activities were discovered. This work may provide valuable information for future design of more promising inhibitors for PTP1B.

© 2010 Elsevier Ltd. All rights reserved.

Protein tyrosine phosphatases (PTPs), a large family of signaling enzymes, serve as paramount regulatory components in numerous cell functions including growth, mitogenesis, motility, cell–cell interaction, metabolism, gene transcription, and immune responses.<sup>1–3</sup> In vivo, tyrosine phosphorylation is reversible and dynamic, with the phosphorylation states of proteins being governed by opposing actions of two enzyme families; protein tyrosine kinase (PTKs), which catalyze the formation of phosphotyrosyl residues in peptide and protein substrates, and the PTPs, which are responsible for dephosphorylation.<sup>4</sup> PTKs, PTPs and their associated substrates are integrated within signal transducing networks,<sup>5</sup> whose defective and unregulated operations give rise to many human diseases including cancer, diabetes and obesity.<sup>6–8</sup> Protein tyrosine phosphatases 1B (PTP1B) is a member of PTP family. It has been identified to play a major role in the dephosphorylation of insulin receptor (IR) and IR substrate thus abrogating insulin signaling.<sup>9–11</sup> Compelling data for this negative role of PTP1B were reported by two laboratories. PTP1B knockout studies in mice indicate that mice lacking PTP1B exhibit improved sensitivity to insulin and are resistant to high-fat diet induced obesity.<sup>12,13</sup> A recent study demonstrated that in vivo, the phosphorylation state of the Y1162/Y1163 site in the activation loop of the IR PTK catalytic domain is regulated by PTP1B.<sup>14</sup> Based

on these studies, PTP1B-specific inhibitors might be expected to enhance insulin sensitivity and act as effective therapeutics for the treatment of type 2 diabetes and obesity. As a result, the design of novel PTP1B inhibitor has intrigued international passion in the last few years.<sup>15–18</sup>

HQSAR is a modern 2D-QSAR technique that eliminates the need of 3D structure determination, conformational search and molecular alignment.<sup>19</sup> Compared with the 3D-QSAR technique, such as CoMFA and CoMSIA, HQSAR could also easily and rapidly generate QSAR models with high predictive value for both small and large data set.<sup>20</sup> Besides, HQSAR models could interpret both positive and negative contributions based on various atoms and structural units, which are alternatives to 3D-QSAR models. Although several papers were published for the 3D-QSAR studies of PTP1B inhibitors, the HQSAR studies of these inhibitors have rarely been reported.<sup>21,22</sup> In this Letter, two HQSAR models were generated and evaluated using two series of PTP1B inhibitors, 39 OBA analogues and 60 BBB analogues.

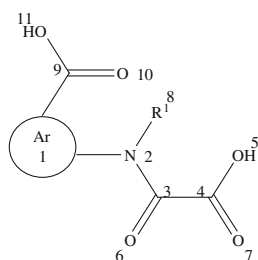
These analogues were taken from previous works.<sup>23,24</sup> Chemical structures with experimental and predicted biological activity of OBA analogues and BBB analogue are listed in Table 1 and Table 2, respectively. The binding affinity  $pK_i$  ( $-\log K_i$ ) ( $\mu\text{M}$ ) for OBA and  $pIC_{50}$  ( $-\log IC_{50}$ ) ( $\mu\text{M}$ ) for BBB were used as dependent variables in HQSAR analyses. An essential characteristic of a training set is that the molecules must be orthogonal (i.e., dissimilar from each other). Surflex-Sim is able to provide the most orthogonal and diverse set of molecules to be included in the training set.<sup>25</sup>

\* Corresponding authors. Tel.: +86 10 62782456; fax: + 86 10 62770304 (F.Z.), tel.: +86 10 88256326; fax: + 86 10 88256093 (M.J.).

E-mail addresses: [jmji@gucas.ac.cn](mailto:jmji@gucas.ac.cn) (M. Ji), [zhangfs@mail.thu.edu.cn](mailto:zhangfs@mail.thu.edu.cn) (F. Zhang).

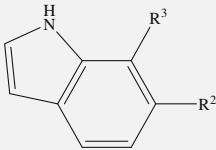
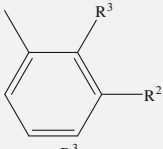
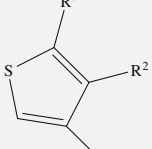
**Table 1**

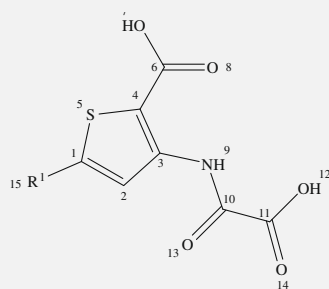
Chemical structures, experimental and predicted activities, and residuals of training set and test set of OBA analogues



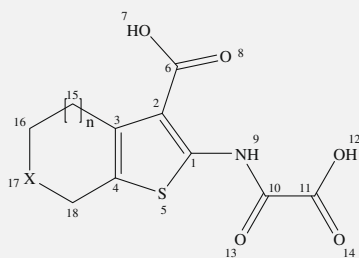
Compound	Ar <sup>a</sup>	R <sup>1</sup>	Experimental	Predicted	Residual
1		H	4.64	4.40	0.24
2		H	2.9	2.83	0.07
3		Me	2.96	2.92	0.04
4		H	3.97	4.16	−0.19
5		H	4.43	4.31	0.12
6		H	4.85	4.73	0.12
7		H	4.85	5.15	−0.30
8		H	3.80	3.86	−0.06
9		H	5.00	5.12	−0.12

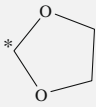
Table 1 (continued)

Compound	Ar <sup>a</sup>	R <sup>1</sup>	Experimental	Predicted	Residual
<b>10</b>		H	4.74	4.75	−0.01
<b>11</b>		H	2.77	3.06	−0.29
<b>12</b>		H	2.70	2.81	−0.11

<sup>a</sup>R<sup>2</sup> = OC(=O)C(=O)NH, R<sup>3</sup> = COOH

Compound	R <sup>1</sup>	Experimental	Predicted	Residual
<b>13</b>	Ph	4.89	5.13	−0.24
<b>14</b>	3-Thienyl	4.96	4.89	0.07
<b>15</b>	4-F-Ph	5.10	5.12	−0.02
<b>16</b>	4-( <i>i</i> -Bu)-Ph	4.89	4.91	−0.02
<b>17</b>	4-HO-Ph	5.35	5.12	0.23
<b>18</b>	4-MeO-Ph	5.05	5.17	−0.12
<b>19</b>	4-PhO-Ph	4.59	4.34	0.25
<b>20</b>	4-BnO-Ph	4.80	4.78	0.02
<b>21</b>	4-(HOOCCH <sub>2</sub> -O)-Ph	5.60	5.52	0.08
<b>22</b>	3-NO <sub>2</sub> -Ph	5.33	5.50	−0.17
<b>23</b>	3-H <sub>2</sub> N-Ph	5.22	5.38	−0.16
<b>24</b>	3,4-MeO-Ph	4.60	4.81	−0.21
<b>25</b>	3,5-MeO-Ph	5.10	5.11	−0.01



Compound	X	n	Experimental	Predicted	Residual
<b>26</b>	CH <sub>2</sub>	2	4.92	4.78	0.14
<b>27</b>	CH <sub>2</sub>	1	5.09	5.04	0.05
<b>28</b>	O	1	4.85	4.98	−0.13
<b>29</b>	S	1	5.08	5.03	0.05
<b>30</b>	SO	1	5.29	5.19	0.10
<b>31</b>	SO <sub>2</sub>	1	5.22	5.21	0.01
<b>32</b>	C=O	1	5.89	5.62	0.27
<b>33</b>		1	5.66	5.70	−0.04

Test Set					
<div></div>					
<b>34</b>		H	4.12	4.36	−0.24
<div></div>					
<b>35</b>		H	5.10	4.96	0.14
<sup>a</sup> R <sup>2</sup> = OC(=O)C(=O)NH, R <sup>3</sup> = COOH					
<div></div>					
Compound	R <sup>1</sup>	Experimental	Predicted	Residual	
<b>36</b>	4-Cl-Ph	5.00	5.22	−0.22	
<b>37</b>	3-MeO-Ph	4.92	5.10	−0.18	
<div></div>					
Compound	X	n	Experimental	Predicted	Residual
<b>38</b>	CH <sub>2</sub>	0	4.92	5.03	−0.11
<b>39</b>	CH−OH	1	5.70	5.27	0.43

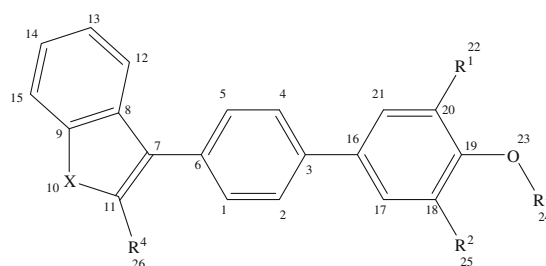
Thus, Surflex-Sim is used to select the training set and test set. For OBA analogues the 39 molecules were segregated into training and test set having 33 and 6 molecules, respectively. For BBB analogues the 60 molecules in the data set were divided into training and test set including 50 and 10 molecules, respectively. All molecular modeling studies were performed using molecular modeling package SYBYL 8.0<sup>26</sup> in a Linux system. The structures of the data sets were sketched and minimized using MMFF94 force field and

MMFF94 charges until a gradient convergence of 0.005 kcal/mol was achieved.<sup>27</sup>

In HQSAR, the molecular structures are broken down into all possible fragments (including branched, linear, and overlapping fragments) containing user defined minimum (M) and maximum (N) number of atoms (fragment size). Each fragment is defined by fragment distinction parameters, including atom (A), bonds (B), connection (C), chirality (Ch), and hydrogen (H), donor/accep-

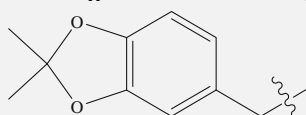
**Table 2**

Chemical structures, experimental and predicted activities, and residuals of training set and test set of BBB analogues

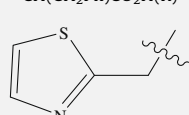


Training Set									
Compound	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	X	Experimental	Predicted	Residual	
1	H	H	H	Benzyl	O	6.04	5.96	0.08	
2	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzyl	O	6.46	6.66	-0.20	
3	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(S)	Benzyl	O	6.49	6.66	-0.17	
4	H	H	CH(CH <sub>2</sub> CH <sub>2</sub> Ph)CO <sub>2</sub> H(S)	Benzyl	O	6.66	6.50	0.16	
5	H	H	CH(Ph)CO <sub>2</sub> H(R)	Benzyl	O	6.40	6.55	-0.15	
6	H	H	CH <sub>2</sub> Ph-4-CO <sub>2</sub> H	Benzyl	O	6.44	6.50	-0.06	
7	Br	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(S)	Benzyl	O	7.25	7.21	0.04	
8	Br	Br	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(S)	Benzyl	O	7.42	7.41	0.01	
9	4-OCH <sub>3</sub> -Ph	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(S)	Benzyl	O	7.37	7.32	0.05	
10	NO <sub>2</sub>	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzyl	O	6.64	6.75	-0.11	
11	Br	Br	CH(CH <sub>2</sub> CH <sub>3</sub> )CO <sub>2</sub> H(S)	Benzyl	O	6.89	6.80	0.09	
12	CH <sub>3</sub>	CH <sub>3</sub>	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzyl	O	7.13	7.15	-0.02	
13	Cyclopentyl	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(S)	Benzyl	O	7.26	7.24	0.02	
14	Cyclopentyl	H	CH <sub>2</sub> COOH	Benzyl	O	6.77	6.68	0.09	
15	NHCH <sub>2</sub> CO <sub>2</sub> H	H	CH <sub>2</sub> CH <sub>2</sub> Ph	Benzyl	O	7.09	7.11	-0.02	
16	NHCOCH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> H	H	H	Benzyl	O	6.04	6.06	-0.02	
17	NHCOCH=CHCO <sub>2</sub> H	H	H	Benzyl	O	6.34	6.18	0.16	
18	NHCO-C <sub>6</sub> H <sub>4</sub> -2-CO <sub>2</sub> H	H	H	Benzyl	O	6.80	6.87	-0.07	
19	Br	H	H	Benzyl	S	5.97	6.30	-0.33	
20	I	I	H	Benzyl	S	6.28	6.18	0.10	
21	Br	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzyl	S	7.24	7.18	0.06	
22	Br	Br	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzyl	S	7.60	7.56	0.04	
23	4-OCH <sub>3</sub> -Ph	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzyl	S	7.28	7.46	-0.18	
24	4-Cl-Ph	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzyl	S	7.28	7.37	-0.09	
25	Br	Br	CH(CH <sub>2</sub> CH <sub>2</sub> Ph)CO <sub>2</sub> H(S)	Benzyl	S	6.54	6.64	-0.10	
26	Br	Br	CH <sub>2</sub> COOH	Benzyl	S	7.00	6.97	0.03	
27	Ph	H	CH <sub>2</sub> COOH	Benzyl	S	7.00	6.88	0.12	
28	4-OC <sub>2</sub> H <sub>5</sub> -Ph	H	CH <sub>2</sub> COOH	Benzyl	S	7.28	7.07	0.21	
29	4-OCH <sub>3</sub> -Ph	H	CH <sub>2</sub> COOH	Benzyl	S	7.10	6.92	0.18	
30	3,4,5-tri-OCH <sub>3</sub> -Ph	H	CH <sub>2</sub> COOH	Benzyl	S	7.00	6.98	0.02	
31	3-OCH <sub>3</sub> -Ph	Br	CH <sub>2</sub> COOH	Benzyl	S	7.55	7.42	0.13	
32	2,4-di-OCH <sub>3</sub> -Ph	Br	CH <sub>2</sub> COOH	Benzyl	S	7.33	7.47	-0.14	
33	4-OCH <sub>3</sub> -Ph	4-OCH <sub>3</sub> -Ph	CH <sub>2</sub> COOH	Benzyl	S	7.60	7.60	0	
34	3-OCH <sub>3</sub> -Ph	3-OCH <sub>3</sub> -Ph	CH <sub>2</sub> COOH	Benzyl	S	7.60	7.72	-0.12	
35	Br	H	CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> COOH	Benzyl	S	6.77	6.79	-0.02	
36	H	H	H	Butyl	O	6.13	5.97	0.16	
37	H	H	CH <sub>2</sub> COOH	Butyl	S	5.66	6.02	-0.36	
38	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Butyl	S	6.77	6.74	0.03	
39	H	H	CH(Ph)CO <sub>2</sub> H(R)	Butyl	S	6.96	6.74	0.22	
40	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzoyl	O	6.17	6.37	-0.20	
41	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	CH(OH)phenyl	O	6.96	6.87	0.09	
42	H	H	H	4-OH-benzyl	S	5.97	6.01	-0.04	
43	H	H	H	2,4-di-OH-benzyl	S	6.24	5.95	0.29	
44	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	4-F-benzyl	S	6.92	6.77	0.15	
45	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	4-OCH <sub>3</sub> -benzyl	S	7.11	6.84	0.27	
46	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	3,4-di-OCH <sub>3</sub> -benzyl	S	6.92	6.85	0.07	
47	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	2,4-di-OCH <sub>3</sub> -benzyl	S	7.07	6.98	0.09	
48	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	<i>b</i>	S	7.11	7.24	-0.13	
49	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	<i>c</i>	S	5.94	6.06	-0.12	
50	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	<i>d</i>	S	5.81	5.89	-0.08	

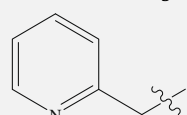
a.



b.



c.



Test Set								
Compound	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	X	Experimental	Predicted	Residual
<b>51</b>	H	H	CH(CH <sub>3</sub> )CO <sub>2</sub> H(R)	Benzyl	O	5.88	6.10	−0.22
<b>52</b>	Br	Br	CH[CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub> ] <sub>2</sub> CO <sub>2</sub> H(R)	Benzyl	O	7.27	7.34	−0.07
<b>53</b>	NHCH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> H	H	CH <sub>2</sub> CH <sub>2</sub> Ph	Benzyl	O	6.85	7.05	−0.20
<b>54</b>	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	Benzyl	S	7.02	6.81	0.21
<b>55</b>	Br	Br	H	Benzyl	S	6.35	6.61	−0.26
<b>56</b>	Br	H	CH <sub>2</sub> COOH	Benzyl	S	6.44	6.61	−0.17
<b>57</b>	2,3-di-OCH <sub>3</sub> -Ph	H	CH <sub>2</sub> COOH	Benzyl	S	7.15	7.04	0.11
<b>58</b>	4-OCH <sub>3</sub> -Ph	Br	CH <sub>2</sub> COOH	Benzyl	S	7.54	7.35	0.19
<b>59</b>	H	H	H	Butyl	S	6.16	6.05	0.11
<b>60</b>	H	H	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	2,4-di-OH-benzyl	S	6.92	6.65	0.27

tor (D). Then, each unique fragment is assigned a specific large positive integer by means of a cyclic redundancy check (CRC) algorithm. Each of these integers corresponds to a bin in an integer array of fixed length *L*. Bin occupancies are incremented according to the fragments generated. Thus, all generated fragments are hashed into array bins in the range from 1 to *L*. This array is called a molecular hologram, and the bin occupancies are the descriptor.<sup>28</sup> HQSAR module provides 12 default hologram length (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 and 401), which are prime numbers, in order to minimize the possibility of fragment collision. Statistical significant HQSAR models are derived from a training set and produced by Partial least squares (PLS) analysis.<sup>29</sup> Cross-validation was done by means of Leave-One-Out (L-O-O) procedure. The optimum number of components is obtained by performing L-O-O analysis with SAMPLS.<sup>30</sup> Through this procedure, each compound is systematically excluded once from the data set, after which its predicted activity is obtained by the model derived from the remaining compounds. The predicted activities of the 'left out' compounds are used to calculate the *q*<sup>2</sup> and cross-validated standard error.

For both training sets of the OBA analogues and the BBB analogues, the default fragment size (4–7) was employed to obtain the statistical values with different combinations of fragment distinction parameters. The statistical results for the OBA analogues and the BBB analogues are listed in Table 3 and Table 4, respectively. In order to improve the statistical results, we applied different fragment sizes with the fragment distinctions of those models having highest statistical values, which are shown in Table 3 and 4 in bold font. The statistical results with different fragment sizes for the OBA analogues and the BBB analogues are summarized in Table 5 and Table 6, respectively. Finally, the optimal HQSAR model for the OBA analogues generated using atoms, bonds, connections, hydrogens and donor/acceptors as fragment distinction and 5–8 as fragment size displaying *q*<sup>2</sup> = 0.592 and *r*<sup>2</sup> = 0.940. The optimal HQSAR model for the BBB analogues developed using atoms, bonds and hydrogens as fragment distinction and 4–7 as fragment size showing *q*<sup>2</sup> = 0.667 and *r*<sup>2</sup> = 0.863.

To evaluate the derived HQSAR models, biological activities of an external test set were predicted using the models derived from the training set.<sup>31</sup> The predictive ability of the HQSAR models is expressed with predictive *r*<sup>2</sup> value, which is similar to cross-validated *r*<sup>2</sup> (*q*<sup>2</sup>), and is defined using the (Eq. (1)):

$$r_{\text{pred}}^2 = \frac{\text{SD} - \text{PRESS}}{\text{SD}} \quad (1)$$

where SD is the sum of squared deviations between the biological activity of the test set and the mean activity of the training set molecules and the PRESS is the sum of squared deviations between the observed and the predicted activity values for every molecule in the test set.<sup>32</sup> For the OBA analogues, the optimal model was validated by an external test set of six compounds with satisfactory predictive *r*<sup>2</sup> value of 0.786. Similarly, for the BBB analogues, the optimal model

also shows satisfactory predictive *r*<sup>2</sup> value of 0.866 for an external test set of 10 compounds. The high predictive correlation coefficients authenticate the reliability and good predictable capability of these HQSAR models. The correlation between the experimental and the predicted activities of both the training set and the test set for the OBA analogues and the BBB analogues with the optimal HQSAR models are showed in Figure 1 and Figure 2, respectively.

**Table 3**

Results of HQSAR analyses for OBA analogues for various fragment distinction combinations on the key statistical parameters using fragment size (4–7)

Model	Fragment distinction	<i>q</i> <sup>2</sup>	<i>r</i> <sup>2</sup>	SEE	HL	N
1	A/B	0.388	0.507	0.577	61	1
2	A/B/C	0.396	0.511	0.575	71	1
3	A/B/H	0.404	0.852	0.339	199	5
4	A/B/D	0.404	0.79	0.403	61	5
5	A/B/Ch	0.389	0.508	0.576	61	1
6	A/B/C/D	0.257	0.469	0.599	307	1
7	A/B/C/H	0.405	0.706	0.461	353	3
8	A/B/D/H	0.472	0.881	0.303	97	5
9	A/B/C/Ch	0.396	0.511	0.575	71	1
10	A/B/D/Ch	0.386	0.741	0.44	151	4
11	A/B/H/Ch	0.384	0.761	0.328	199	5
<b>12</b>	<b>A/B/C/D/H</b>	<b>0.504</b>	<b>0.889</b>	<b>0.296</b>	<b>61</b>	<b>5</b>
13	A/B/C/D/Ch	0.256	0.469	0.599	307	1
14	A/B/C/H/Ch	0.405	0.706	0.461	353	3
15	A/B/D/H/Ch	0.412	0.87	0.317	97	5
16	A/B/C/D/H/Ch	0.423	0.831	0.356	97	4

*q*<sup>2</sup>, cross-validated correlation coefficient; *r*<sup>2</sup>, noncross-validated correlation coefficient; SEE, standard estimated error; HL, hologram length; N, optimal number of components. Fragment distinction: A, atom; B, bond; C, connections; H, hydrogen atom; Ch, chirality; D, donor and acceptor.

The model chosen for analysis is highlighted in bold fonts.

**Table 4**

Results of HQSAR analyses for BBB analogues for various fragment distinction combinations on the key statistical parameters using fragment size (4–7)

Model	Fragment distinction	<i>q</i> <sup>2</sup>	<i>r</i> <sup>2</sup>	SEE	HL	N
1	A/B	0.621	0.726	0.282	53	3
2	A/B/C	0.608	0.771	0.258	83	3
<b>3</b>	<b>A/B/H</b>	<b>0.667</b>	<b>0.863</b>	<b>0.204</b>	<b>199</b>	<b>5</b>
4	A/B/D	0.639	0.743	0.273	53	3
5	A/B/Ch	0.616	0.720	0.285	53	3
6	A/B/C/D	0.606	0.831	0.224	257	4
7	A/B/C/H	0.665	0.862	0.205	59	5
8	A/B/D/H	0.598	0.887	0.211	307	6
9	A/B/C/Ch	0.616	0.777	0.254	83	3
10	A/B/D/Ch	0.625	0.749	0.270	199	3
11	A/B/H/Ch	0.596	0.846	0.216	401	5
12	A/B/C/D/H	0.582	0.762	0.265	71	4
13	A/B/C/D/Ch	0.643	0.855	0.207	353	4
14	A/B/C/H/Ch	0.648	0.846	0.216	59	5
15	A/B/D/H/Ch	0.599	0.793	0.211	307	5
16	A/B/C/D/H/Ch	0.596	0.772	0.260	71	4

**Table 5**

Influence of fragment size for OBA analogues on the statistical parameters using the fragment distinction combinations (A/B/C/D/H)

Fragment size	$q^2$	$r^2$	SEE	HL	N
1–4	0.435	0.782	0.411	71	5
2–5	0.349	0.731	0.441	199	3
3–6	0.380	0.700	0.465	97	3
4–7	0.504	0.889	0.296	61	5
<b>5–8</b>	<b>0.592</b>	<b>0.940</b>	<b>0.217</b>	<b>61</b>	<b>5</b>
6–9	0.491	0.844	0.345	61	3
7–10	0.480	0.856	0.288	401	5

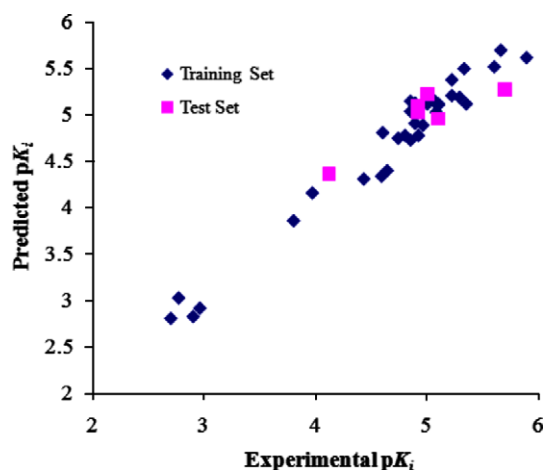
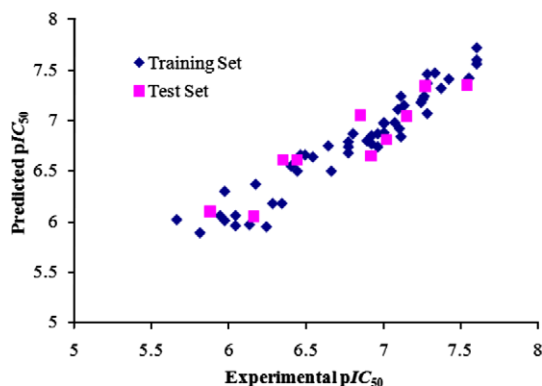
**Table 6**

Influence of fragment size for OBA analogues on the statistical parameters using the fragment distinction combinations (A/B/H)

Fragment size	$q^2$	$r^2$	SEE	HL	N
1–4	0.654	0.813	0.238	97	5
2–5	0.661	0.834	0.225	257	3
3–6	0.645	0.847	0.215	401	5
<b>4–7</b>	<b>0.667</b>	<b>0.863</b>	<b>0.204</b>	<b>199</b>	<b>5</b>
5–8	0.592	0.814	0.235	97	4
6–9	0.642	0.851	0.205	59	5
7–10	0.622	0.849	0.219	257	5

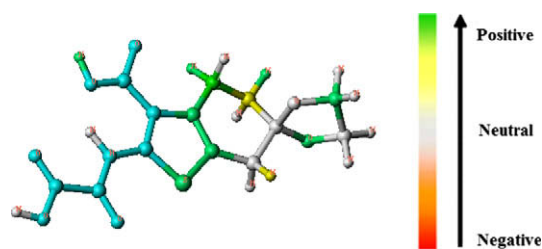
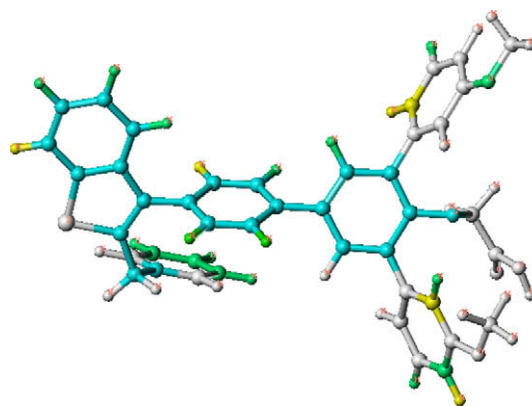
The result of the HQSAR analysis can be graphically displayed as a color-coded structure diagram, in which the color coding of each atom reflects the contribution of that atom to the molecule's overall activity. The red end of the spectrum (red, red orange, and orange) represents the unfavorable or negative contribution to the activity, while the green end (yellow, green blue, and blue) denotes favorable or positive contribution to the activity. Atoms with intermediate contribution to the activity are colored white. The common backbones (maximal common structures) are colored cyan.<sup>20</sup> The individual atomic contributions of the most active OBA analogue (compound **33**) and BBB analogue (compound **34**) are displayed in Figure 3 and Figure 4, respectively.

From Figure 3, it can be seen that the fragment of spiro [1',3']dioxolane substituted cyclohexyl is strongly associated with the biological activity of this compound. The color green and yellow indicate their favorable contribution to the activity enhancement, which is consistent with previous study about the binding style of OBA inhibitors with PTP1B.<sup>21</sup> According to that research, the spiro[1',3']dioxolane substituted cyclohexyl is on the 'mouth'

**Figure 1.** Plot of experimental versus predicted values of  $pK_i$  of the training set and test set molecules for OBA analogues.**Figure 2.** Plot of experimental versus predicted values of  $pIC_{50}$  of the training set and test set molecules for BBB analogues.

of binding pocket, thus enlarging molecular size here is crucial to increase binding affinity and improve activity. Compounds **26–32**, which possess fused ring (cyclohexyl etc.), also show valuable biological activities. Previous publications pointed out that fused ring systems markedly increase the overall potency against PTP1B due to additional van der Waals interactions to the phenyl phosphate binding pocket.<sup>23</sup> In addition, when an additional phenyl ring was introduced into the core structure (compound **13–25**), green or yellow of the phenyl ring displays their positive contribution to biological activity, which can be interpreted by the van der Waals interaction between the phenyl ring and two residues, Ile219 and Val49.<sup>23</sup>

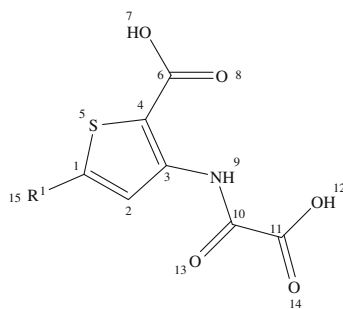
As what suggests in Figure 4, the *ortho-ortho*-disubstituted groups (methoxyl-phenyl group) of the phenyllactic head piece and the benzyl on the benzofuran portion are strongly related to the biological activity of this compound. This result is consistent with previous study which shows the *ortho-ortho*-disubstituted

**Figure 3.** The HQSAR contribution map of compound **33** for OBA analogues.**Figure 4.** The HQSAR contribution map of compound **34** for BBB analogues.



**Table 7**

Chemical structures and predicted activities of designed molecules for OBA analogues



Compound	R <sup>1</sup>	Predicted	
		HQSAR	FlexX (kJ mol <sup>-1</sup> )
<b>D1</b>	4-N(CH <sub>3</sub> ) <sub>2</sub> -Ph	5.944	−36.16
<b>D2</b>	4-NHOH-Ph	5.861	−37.24
<b>D3</b>	3-NH <sub>2</sub> -4-OCH <sub>2</sub> OH-Ph	6.225	−38.12
<b>D4</b>	3-SH-4-NH <sub>2</sub> -Ph	5.983	−37.38
<b>D5</b>	3,4-di-NH <sub>2</sub> -Ph	6.023	−37.71

analogues have great activity since they occupy a large hydrophobic region at the vicinity of the catalytic site. This site is defined by Phe182, Asp181 and the loop containing residues 215–221.<sup>24</sup> Similarly, the compounds **9**, **23** and **29–33**, which possess one or two *ortho*-phenyl fragments of the phenyllactic head part, all have high biological activities. Consequently, the structural unit of *ortho*-phenyl should be included in a new design of BBB analogues for PTP1B inhibitors.

Finally, based on above suggestions, two virtual libraries were designed by Legion in SYBYL and screened by the optimal HQSAR models. Legion is a technique for creating virtual combinatorial libraries for efficient storage in a UNITY database, for retrieving them from such databases, and for generating lists of the individual compounds included in the combinatorial structure.<sup>26</sup> Thus, Legion is used to generate new OBA analogues and BBB analogues. First, a database of more than three hundreds of mono-valent substitutes was generated. Then two virtual combinatorial libraries of the OBA analogues and the BBB analogues were created by replacing the mono-valent substitutes at different positions. Subsequently, the

virtual libraries for the OBA analogues and BBB analogues were screened by the optimal HQSAR models. Chemical structures of the designed molecules for the OBA analogues and the BBB analogues, which have high predicted activities are listed in Table 7 and Table 8, respectively. From these tables, it can be seen that all designed compounds for OBA analogues had an additional phenyl ring and for BBB analogues had the structural unit of *ortho*-phenyl, which are consistent with the HQSAR graphical result of individual atomic contribution.

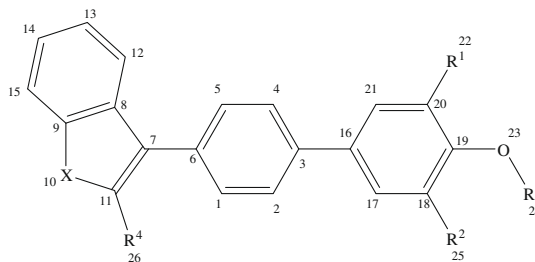
For the purpose to further validate the rationality of designed molecules, docking studies were performed by FlexX on the OBA analogues as an example. FlexX is a flexible automated docking program that the flexibility of the ligand is considered while the protein or biomolecule is considered as a rigid structure<sup>33</sup> (see Supplementary data for detailed information about FlexX docking). The predicted binding free energies (total scores) of designed molecules for the OBA analogues are listed in Table 7. From Table 7, it can be seen that all designed OBA analogues (except compound **D1**) had satisfactory predicted binding free energies which are consistent with the predicted HQSAR activities. In order to explain it, the binding modes of designed molecules of the OBA analogues with PTP1B have been intensively studied.

Figure 5 displayed the binding mode of compound **D3** of OBA analogue with PTP1B. From this figure, it can be seen compound **D3** extends deep into the binding site and forms a hydrogen bonding network with key residues in the catalyst site, such as Lys120, Tyr46, Asp181, Ser216, Gly220 and Arg221. The thiophene ring is sandwiches between Tyr46 and Phe182, with which can form strong  $\pi$ – $\pi$  stacking interactions to further stabilize the binding. The additional phenyl ring forms a set of van der Waals interactions with Tyr46, Val49 and Ile219. Furthermore, the –OCH<sub>2</sub>OH group and the –NH<sub>2</sub> group on this phenyl ring could form additional hydrogen bonding interactions with Asp48 and Gln262, respectively. Compared to the binding mode of compound **21** with PTP1B (see Supplementary data), it is worthwhile to note that compound **D3** had two additional hydrogen bonding interactions with Gln262 in PTP1B. This might interpret the reason that compound **D3** has better predicted binding free energy than compound **21**. The predicted binding free energies of compounds **D3** and **21** are −38.12 and −37.13 kJ/mol, respectively.

The binding modes of the rest designed molecules of OBA analogues were shown in Figure S4. As what suggests in Figure S4,

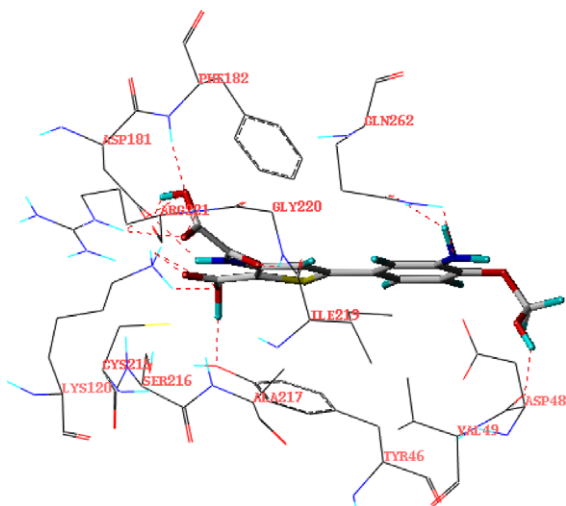
**Table 8**

Chemical structures and predicted activities of designed molecules for BBB analogues



Compound	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	X	Predicted
<b>1</b>	3-OCH <sub>2</sub> CH <sub>3</sub> -Ph	Benzyl	CH <sub>2</sub> COOH	4-OH-benzyl	S	8.205
<b>2</b>	Benzyl	Benzyl	CH <sub>2</sub> COOH	3,4-di-OCH <sub>3</sub> -benzyl	S	8.305
<b>3</b>	2,4-di-OCH <sub>3</sub> -Ph	Benzyl	CH <sub>2</sub> COOH	2,4-di-OCH <sub>3</sub> -benzyl	S	8.126
<b>4</b>	CH(OH)-Ph	Br	CH(CH <sub>2</sub> Ph)CO <sub>2</sub> H(R)	2,4-di-OCH <sub>3</sub> -benzyl	S	8.227
<b>5</b>	4-OH-benzyl	4-COOH-Ph	CH <sub>2</sub> COOH	CH(OH)-Phenyl	S	8.038
<b>6</b>	3-Br-Ph	4-OH-Ph	CH <sub>2</sub> COOH	3,4-di-OCH <sub>3</sub> -benzyl	S	8.102
<b>7</b>	4-OH-Ph	Benzyl	CH <sub>2</sub> COOH	3,4-di-OCH <sub>3</sub> -benzyl	O	8.064
<b>8</b>	CH(OH)-Ph	4-COOH-Ph	CH <sub>2</sub> COOH	4-OCH <sub>3</sub> -benzyl	O	8.186
<b>9</b>	Benzyl	Benzyl	CH <sub>2</sub> COOH	CH(OH)-Phenyl	O	8.112
<b>10</b>	4-OH-benzyl	CH(OH)-Ph	CH <sub>2</sub> COOH	4-Br-Benzyl	O	8.034





**Figure 5.** FlexX-predicted binding mode of compound D3 of OBA analogues with PTP1B active site. For clarity, only the polar hydrogens are shown. Hydrogen bonds are shown as dotted red lines. Active site amino acid residues are represented as lines with the atoms colored as (carbon: black, hydrogen: cyan, nitrogen: blue, oxygen: red and sulfur: yellow) while the inhibitor is shown as stick with same color scheme as above except carbons are represented in white.

these compounds display the similar binding manners as those previously described for compound **D3** except the hydrogen bonding interactions with Asp48 and Gln262. For compound **D5**, it could be found that the two  $\text{-NH}_2$  groups form two additional hydrogen bonding interactions with Asp48 and Gln262, respectively. Similarly, both the  $\text{-NHOH}$  group of compound **D2** and the  $\text{-NH}_2$  group of compound **D4** could form an additional hydrogen bonding interaction with Asp48. This might be the reason that they have the similar predicted binding free energies to compound **D1**. The predicted binding free energies of compounds **D2**, **D4** and **D5** are  $-37.24$ ,  $-37.38$  and  $-37.71$  kJ/mol, respectively. However, after carefully investigating the binding mode of compound **D1**, no additional hydrogen bonding interaction with Asp48 and Gln262 had been found, which might interpret the reason that the predicted binding free energy of compound **D1** only equaled  $-36.16$  kJ/mol. Thus, based on the HQSAR prediction and the FlexX docking result of OBA analogues, compounds **D2–D5** might be chosen as potential candidates for future work. Similarly, for the designed compounds of BBB analogues, some of them may have good values of binding free energies which are consistent with the predicted HQSAR activities.

In summary, we have successfully generated two HQSAR models for PTP1B inhibitors. For OBA analogues, the statistical result of the optimal HQSAR model displays  $q^2 = 0.592$  and  $r^2 = 0.940$ . The optimal HQSAR model for the BBB analogues shows  $q^2 = 0.667$  and  $r^2 = 0.863$ . The reliability and the predictable capability of two models were verified by the predictive  $r^2$ , which for the OBA analogues and the BBB analogues are 0.786 and 0.866, respectively. Moreover, the contribution maps derived from the optimal HQSAR models are consistent with the biological activities of the studied compounds. Finally, two virtual combinatorial libraries were designed and screened by the optimal HQSAR models and the potential candidates with high predictive biological activities were discovered. This work may provide valuable information for future design of more promising inhibitors for PTP1B.

## Acknowledgments

The project was supported by the Natural Science Foundation of China (No. 20773077) and the National Science and Technology Major Special Project of China (No. 2009ZX09501-011).

We thank Professor Weiren Xu in Tianjin Institute of Pharmaceutical Research for providing access to the SYBYL molecular simulation package.

## Supplementary data

Supplementary data (FlexX docking) associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2010.04.033.

## References and notes

- Hunter, T. *Cell* **1995**, *80*, 225.
- Östman, A.; Böhmer, F.-D. *Trends Cell. Biol.* **2001**, *11*, 258.
- Evans, J. L.; Jallal, B. *Expert Opin. Invest. Drugs* **1999**, *8*, 139.
- Zhang, Z.-Y. *Curr. Opin. Chem. Biol.* **2001**, *5*, 416.
- Karin, M.; Hunter, T. *Curr. Biol.* **1995**, *5*, 747.
- Lu, J.; Li, Q.; Xie, H.; Chen, Z.-J.; Borovitskaya, A. E.; Maclaren, N. K.; Notkins, A. L.; Lan, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2307.
- Hunter, T. *Cell* **2000**, *100*, 113.
- Chan, A. C.; Desai, D. M.; Weiss, A. *Ann. Rev. Immunol.* **1994**, *12*, 555.
- Tobin, J. F.; Tam, S. *Curr. Opin. Drug Disc. Dev.* **2002**, *5*, 500.
- Liu, G.; Trevillyan James, M. *Curr. Opin. Invest. Drugs* **2002**, *3*, 1608.
- Johnson, T. O.; Ermolieff, J.; Jirousek, M. R. *Nat. Rev. Drug Disc.* **2002**, *1*, 696.
- Elchebly, M.; Payette, P.; Michaliszyn, E.; Cromlish, W.; Collins, S.; Loy, A. L.; Normandin, D.; Cheng, A.; Himms-Hagen, J.; Chan, C. C.; Ramachandran, C.; Gresser, M. J.; Tremblay, M. L.; Kennedy, B. P. *Science* **1999**, *283*, 1544.
- Klaman, L. D.; Boss, O.; Peroni, O. D.; Kim, J. K.; Martino, J. L.; Zabolotny, J. M.; Moghal, N.; Lubkin, M.; Kim, Y. B.; Sharpe, A. H.; Stricker-Krongrad, A.; Shulman, G. I.; Neel, B. G.; Kahn, B. B. *Mol. Cell. Biol.* **2000**, *20*, 5479.
- Galic, S.; Hauser, C.; Kahn, B. B.; Haj, F. G.; Neel, B. G.; Tonks, N. K.; Tiganis, T. *Mol. Cell. Biol.* **2005**, *25*, 819.
- Holmes, C. P.; Li, X.-F.; Pan, Y.; Xu, C.; Bhandari, A.; Moody, C. M.; Miguel, J. A.; Ferla, S. W.; Francisco, M. D.; Irvine, J. D.; Grove, J. R. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2719.
- Shrestha, S.; Bhattarai, B. R.; Lee, K.-H.; Cho, H. *Bioorg. Med. Chem.* **2007**, *15*, 6535.
- Combs, A. P.; Yue, E. W.; Bower, M.; Ala, P. J.; Wayland, B.; Douth, B.; Takvorian, A.; Polam, P.; Wasserman, Z.; Zhu, W.; Crawley, M. L.; Pruitt, J.; Sparks, R.; Glass, B.; Modi, D.; McLaughlin, E.; Bostrom, L.; Li, M.; Galya, L.; Blom, K.; Hillman, M.; Gonneville, L.; Reid, B. G.; Wei, M.; Becker-Pasha, M.; Klabbe, R.; Huber, R.; Li, Y.; Hollis, G.; Burn, T. C.; Wynn, R.; Liu, P.; Metcalf, B. J. *Med. Chem.* **2005**, *48*, 6544.
- Forghieri, M.; Laggner, C.; Paoli, P.; Lauger, T.; Manao, G.; Camici, G.; Bondioli, L.; Prati, F.; Costantino, L. *Bioorg. Med. Chem.* **2009**, *17*, 2658.
- Tong, W.; Lewis, D. R.; Perkins, R.; Chen, Y.; Welsh, W. J.; Goddette, D. W.; Heritage, T. W.; Sheehan, D. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669.
- Tripos Sybyl, HQSAR manual.
- Zhou, M.; Ji, M.-J. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5521.
- Pan, Y.-M.; Ji, M.-J.; Ye, X.-Q.; Kuang, P.-X. *Chin. J. Org. Chem.* **2003**, *23*, 167.
- Anderson, H. S.; Olsen, O. H.; Iversen, L. F.; Sorensen, A. L.; Mortensen, S. B.; Christensen, M. S.; Branner, S.; Hansen, T. K.; Lau, J. F.; Jeppesen, L.; Moran, E. J.; Su, J.; Bakir, F.; Judge, L.; Shahbaz, M.; Collins, T.; Vo, T.; Newman, M. J.; Ripka, W. C.; Moller, N. P. *J. Med. Chem.* **2002**, *45*, 4443.
- Malamas, M. S.; Sredy, J.; Moxham, C.; Katz, A.; Xu, W.-X.; McDevitt, R.; Adebayo, F. O.; Sawicki, D. R.; Seestaller, L.; Sullivan, D.; Taylor, J. R. *J. Med. Chem.* **2000**, *43*, 1293.
- Jain, A. N. *J. Med. Chem.* **2004**, *47*, 947.
- SYBYL 8.0, Tripos Inc.
- Halgren, T. J. *Comput. Chem.* **1999**, *20*, 720.
- Flower, D. R. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 379.
- Word, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindberg, W.; Sioström, M.; Kowalski, Reidel, Dordrecht, Netherlands, 1984, p 17.
- Bush, B. L.; Nachbar, R. B. *J. Comput. Aided Mol. Des.* **1993**, *7*, 587.
- Golbraikh, A.; Tropsha, A. *Mol. Divers.* **2000**, *5*, 231.
- Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1998**, *110*, 5959.
- Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.