Original article

Protein-based alignment in 3D-QSAR of FBPase inhibitors

Ping Yi [a,b], Ying-Tong Di [b], Wei Liu [a,*], Xiao-Jiang Hao [b,**], Yong Ming [a], Du-Shu Huang [a], Jin Yang [a], Zhong-Zhou Yi [a], Zi-Jing Li [a], Rui-Dong Yang [a], Ju-Cheng Zhang [a]

[a] Key Laboratory of Natural Pharmaceutical & Chemical Biology of Yunnan Province, Honghe University, Mengzi 661100, China
[b] State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, China

ABSTRACT

Three-dimensional quantitative structure–activity relationship (3D-QSAR) studies were performed on Frusectose-1, 6-bisphosphatase (FBPase) inhibitors, based on molecular docking obtained by using GOLD and comparative molecular field analysis (CoMFA). Three random splits into training and test sets had been performed and the high leave-one-out (LOO) cross-validated correlation coefficients q^2 of 0.781, 0.725 and 0.801, respectively, revealed that the models are useful tools for the prediction of test sets as well as newly designed structures against FBPase activity. The superimposed CoMFA models on the receptor site of FBPase are guiding the design of potential inhibitory structures directed against FBPase activity.

© 2011 Elsevier Masson SAS. All rights reserved.

1. Introduction

Frusectose-1, 6-bisphosphatase (FBPase) has long been recognized as a potential therapeutic target for the treatment of type 2 diabetes mellitus (T2DM). As a rate-limiting enzyme of the gluconeogenesis (GNG) pathway, it catalyzes the hydrolysis of fructose-1, 6-bisphosphate to fructose-6-phosphate. Thus, the inhibition of gluconeogenesis is a useful approach in reducing increased blood glucose levels in patients with T2DM. FBPase inhibitors would lower blood glucose levels by reducing hepatic glucose output and are expected to be a novel class of drugs for the treatment of T2DM [1–8].

In recent years, a number of FBPase inhibitors have appeared in the patent and primary literatures [9–28]. Among them, Qun Dang et al. synthesized a series of benzimidazole-based derivatives as potent, selective FBPase inhibitors [13]. In addition, An X-ray crystal structure of a similar analogue inhibitor in complex with human liver FBPase have been determined which provide important information about the interaction with the residues of the binding site [21].

Recently many studies have been published in which the combination of receptor-based methods and 3D-QSAR. The three-dimensional structure of a target protein, along with a docking protocol is used to guide alignment selection for comparative molecular field analysis. It is quite appealing to combine the accuracy of a receptor-based alignment with the computational efficiency of a ligand-based method. Receptor structures, either experimentally resolved or obtained by homology modeling, can provide important information that is critical for an alignment in CoMFA, while QSAR can provide better prediction of binding energies [29,30].

In this study, we applied this receptor-based 3D-QSAR technique to a set of 47 FBPase inhibitors which have been recently developed [13]. The crystal structure of the catalytic domain of human liver FBPase together with an automatic docking program was used to determine the molecular alignment of the ligands. The 3D-QSAR model obtained by this way yielded a high correlation between the experimentally determined binding affinity and the calculated molecular interaction fields. It was shown that the receptor-based 3D-QSAR yields a better prediction of the binding affinity than using an interaction energy-based model or a ligand-based 3D-QSAR analysis. Encouraged by these results the receptor-based 3D-QSAR model should now be used for the screening and prediction of novel inhibitors for FBPase.

2. Computational details
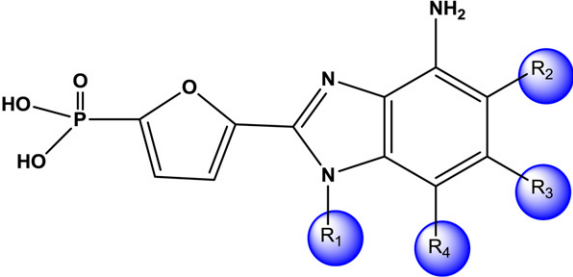
2.1. Biological data and molecular structures

A series of FBPase benzimidazole-based inhibitors, published by Qun Dang et al. in 2010 [13], were shown in Table 1. The human liver FBPase IC$_{50}$s were converted to pIC$_{50}$ ($-$logIC$_{50}$) and used as dependent variables in the CoMFA calculations.

Table 1
Inhibitory activity of miscellaneous benzimidazole-based derivatives.

	R₁	R₂	R₃	R₄	IC₅₀ (μM)	pIC₅₀
1	Me	H	H	H	6	5.2218
2	Et	H	H	H	2.25	5.6478
3	nPr	H	H	H	1.1	5.9586
4	iBu	H	H	H	1.5	5.8239
5	Cyclopropyl-CH₂−	H	H	H	0.8	6.0969
6	Cyclobutyl-CH₂−	H	H	H	0.8	6.0969
7	Cyclopentyl-CH₂−	H	H	H	1.5	5.8239
8	Cyclohexyl-CH₂−	H	H	H	2.5	5.6021
9	Cycloheptyl-CH₂−	H	H	H	3.25	5.4881
10	Norbonyl	H	H	H	1	6.0000
11	Benzyl	H	H	H	5	5.3010
12	4-tBu-Benzyl	H	H	H	9.5	5.0223
13	4-CF₃-Benzyl	H	H	H	7	5.1549
14	4-Ph-Benzyl	H	H	H	2.5	5.6021
15	3-Furnyl-CH₂−	H	H	H	4.2	5.3768
16	3-HO-benzyl	H	H	H	1.85	5.7328
17	3-Thienyl-CH₂−	H	H	H	4	5.3979
18	iBu	Et	H	H	2.5	5.6021
19	iBu	nPr	H	H	3	5.5229
20	iBu	MeO	H	H	0.7	6.1549
21	iBu	HO	H	H	0.5	6.3010
22	iBu	Cl	H	H	0.2	6.6990
23	iBu	H	H	Cl	0.9	6.0458
24	iBu	Br	H	H	0.4	6.3979
25	iBu	H	H	Br	0.4	6.3979
26	iBu	F	H	H	0.1	7.0000
27	(Et)₂CHCH₂−	F	H	H	0.15	6.8239
28	(Et)₂CH−	F	H	H	0.85	6.0706
29	cPr−CH₂−	F	H	H	0.055	7.2596
30	iBu	Br	H	Br	1	6.0000
31	iBu	Cl	H	Cl	0.45	6.3468
32	iBu	F	H	Cl	0.1	7.0000
33	iBu	F	H	Br	0.13	6.8861
34	iBu	F	Cl	H	0.225	6.6478
35	iBu	Br	Cl	Cl	10	5.0000
36	iBu	F	H	Vinyl	0.28	6.5528
37	iBu	F	H	cPr	0.06	7.2218
38	iBu	F	H	Ph	0.09	7.0458
39	iBu	F	H	4-F−Ph	0.18	6.7447
40	iBu	F	H	4-Cl−Ph	0.09	7.0458
41	iBu	F	H	Et	0.055	7.2596
42	iBu	F	H	nPr	0.1	7.0000
43	iBu	F	H	tBu(CH₂)₂−	0.21	6.6778
44	iBu	F	H	(Me)₂CH(CH₂)₃−	0.1	7.0000
45	iBu	F	H	HO(CH₂)₃−	0.08	7.0969
46	iBu	F	H	(Me)₂N(CH₂)₃−	0.055	7.2596
47	iBu	F	H	Cl(CH₂)₄−	0.07	7.1549

The 3D structures of these compounds were constructed using the molecular modeling software package Sybyl6.9 [31]. Partial atomic charges were calculated by the MMFF94 method [32–36], and energy minimizations were performed using the Tripos force field with a distance-dependent dielectric and the Powell conjugate gradient algorithm (convergence criterion of 0.001 kcal/mol Å) [37]. All compounds were generated in the protonation state under physiological condition.

2.2. Data set

The compounds fit roughly into four compound clusters according to their skeletons in the original paper [13]. In order to examine the predictive ability and robustness of the CoMFA models, we split the whole data set into two parts − the training set consists of 41 compounds and the test set, selected randomly, is comprised of six compounds. Selection of the training set and test set molecules was done by considering the fact that six compounds of the test set cover approximately all range the clusters and the IC₅₀s similar to that of the training set. Thus, the test set is the true representative of the training set. Besides, the numbers of three random splits into training and test sets and their statistical analyses were shown in Tables 2a–2c and 3a–3c.

2.3. Molecular docking

To locate the appropriate binding orientations and conformations of the benzimidazole-based inhibitors interacting with FBPase, a powerful computational searching method was needed. The advanced molecular docking program GOLD (version 4.12), with a powerful genetic algorithm (GA) method for conformational search and docking programs [38,39], was employed to generate an ensemble of docked conformations. Atomic coordinates for the human liver FBPase complex with ligand 2T4 ({[(2-amino-8H-indeno[1,2-d][1,3]thiazol-4-yl)oxy]methyl}phosphonic acid) that was used for our modeling study have been deposited in the Brookhaven Protein DataBank with a resolution of 2.45 Å (PDB ID: 3KBZ) [21]. The original ligand was removed from the coordinated set of FBPase, and five ordered water molecules within 6 Å of the active pocket were taken into account during the docking procedures. The ligands were scored based on the fitness function 'GoldScore'. GOLD was run to save up to 10 top-ranked docking solutions for the ligands.

2.4. Structural alignment

Ten conformations were obtained using GOLD for each ligand. The conformations and their alignment-the relative binding positions of the conformations in FBPase, all obtained from GOLD, were used directly in CoMFA to explore the specific contributions of electrostatic and steric effects to the molecular bioactivities. First, an original

Table 2a
Statistical analysis of the CoMFA model 1.[a]

Parameter	CoMFA		
PLS statistics	Leave-one-out	Leave-20%-out	Leave-50%-out
q^2 (CV correlation coefficient)	0.781	0.782	0.596
N (number of components)	6	6	6
SEP (standard error of prediction)	0.351	0.350	0.477
Un-cross-validated			
r^2 (correlation coefficient)	0.993		
SEE (standard error of estimate)	0.062		
F (F-ratio)	818.274		
Field distribution(%)			
Steric	33.0		
Electrostatic	67.0		
Testing set			
r^2_{pred} (correlation coefficient)	0.679		
S (standard error of prediction)	0.399		

[a] The training set of Model 1 contains compounds 1, 3–8, 10–24, 26–34, 36–38, 40–42 and 44–47. While the six randomly selected compounds 2, 9, 25, 35, 39 and 43 were used as the test set (see table 3 and section 3.2).

Table 2b
Statistical analysis of the CoMFA Model 2.[a]

Parameter	CoMFA		
PLS statistics	Leave-one-out	Leave-20%-out	Leave-50%-out
q^2 (CV correlation coefficient)	0.725	0.737	0.628
N (number of components)	6	5	6
SEP (standard error of prediction)	0.414	0.398	0.481
Un-cross-validated			
r^2 (correlation coefficient)	0.986		
SEE (standard error of estimate)	0.093		
F (F-ratio)	400.250		
Field distribution(%)			
Steric	34.9		
Electrostatic	65.1		
Testing set			
r^2_{pred} (correlation coefficient)	0.819		
S (standard error of Prediction)	0.156		

[a] The training set of Model 2 contains compounds 1–2, 4–9, 11–23, 24–33, 35–37 and 39–47. While the six randomly selected compounds 3, 10, 20, 24, 34 and 38 were used as the test set (see table 3 and section 3.2).

CoMFA model with a cross-validated q^2 value of 0.264 for two components was obtained by choosing the conformations for each ligand with the highest GoldScore. Thus, in order to increase the predictive power of the derived model, further experiments were performed. Then, we checked all conformations that GOLD generated and replaced conformations for each ligand with the ones which have the lowest residues between actual and predicted pIC_{50} forecasted by the prior CoMFA model. Finally, a new CoMFA model with a cross-validated q^2 value of 0.770 for six components of all 47 ligands was obtained by this way. The non-cross-validated PLS analysis with the optimum components of 6 resulted in a conventional r^2 of 0.985, $F = 451.888$, and an estimated standard error of 0.090.

2.5. CoMFA

Steric and electrostatic interactions were calculated using the Tripos force field with a distance-dependent dielectric constant at all intersections in a regular space (2 Å) grid taking a sp^3 carbon atom as steric probe and $a + 1$ charge as electrostatic probe. The

Table 2c
Statistical analysis of the CoMFA Model 3.[a]

Parameter	CoMFA		
PLS statistics	Leave-one-out	Leave-20%-out	Leave-50%-out
q^2 (CV correlation coefficient)	0.801	0.768	0.570
N (number of components)	6	6	4
SEP (standard error of prediction)	0.346	0.374	0.494
Un-cross-validated			
r^2 (correlation coefficient)	0.988		
SEE (standard error of estimate)	0.086		
F (F-ratio)	453.122		
Field distribution (%)			
Steric	36.5		
Electrostatic	63.5		
Testing set			
r^2_{pred} (correlation coefficient)	0.568[b]		
S (standard error of prediction)	0.203[b]		

[a] The training set of Model 3 contains Compounds 1–3, 5–7, 8–15, 17–21, 23–30, 32–39 and 41–47. While the six randomly selected Compounds 4, 8, 16, 22, 31 and 40 were used as the test set (see table 3c and section 3.2).
[b] The outlier compound 8 in Model 3 was not taken into consideration. (see table 3c and section 3.2)

Table 3a
List of actual pIC_{50}, predicted pIC_{50} (derived from LOO cross-validation), calculated pIC_{50} values and residual of each compound derived from the CoMFA Model 1.[a]

Compound number	Actual pIC_{50}	Predicted pIC_{50}	Residuals
1	5.2218	5.2982	−0.0764
2[t]	5.6478	5.4284	0.2194
3	5.9586	5.9757	−0.0171
4	5.8239	5.8664	−0.0425
5	6.0969	5.9626	0.1343
6	6.0969	6.0197	0.0772
7	5.8239	5.8296	−0.0057
8	5.6021	5.7530	−0.1509
9[t]	5.4881	5.9514	−0.4633
10	6.0000	5.9599	0.0401
11	5.3010	5.3403	−0.0393
12	5.0223	4.9321	0.0902
13	5.1549	5.1257	0.0292
14	5.6021	5.6432	−0.0411
15	5.3768	5.3728	0.0040
16	5.7328	5.7867	−0.0539
17	5.3979	5.3789	0.0190
18	5.6021	5.7272	−0.1251
19	5.5229	5.5079	0.0150
20	6.1549	6.2342	−0.0793
21	6.3010	6.2660	0.0350
22	6.6990	6.7151	−0.0161
23	6.0458	6.0274	0.0184
24	6.3979	6.3545	0.0434
25[t]	6.3979	6.1368	0.2611
26	7.0000	6.9934	0.0066
27	6.8239	6.8279	−0.0040
28	6.0706	6.0783	−0.0077
29	7.2596	7.2717	−0.0121
30	6.0000	5.9212	0.0788
31	6.3468	6.2751	0.0717
32	7.0000	6.9741	0.0259
33	6.8861	6.8506	0.0355
34	6.6478	6.5898	0.0580
35[t]	5.0000	5.6413	−0.6413
36	6.5528	6.5702	−0.0174
37	7.2218	7.2535	−0.0317
38	7.0458	7.0989	−0.0531
39[t]	6.7447	7.0051	−0.2604
40	7.0458	7.1149	−0.0691
41	7.2596	7.2894	−0.0298
42	7.0000	6.9622	0.0378
43[t]	6.6778	7.0937	−0.4159
44	7.0000	7.0374	−0.0374
45	7.0969	7.0992	−0.0023
46	7.2596	7.2059	0.0537
47	7.1549	7.1164	0.0385

[a] Compound number with "t" refers to those compounds included in the test set.

cutoff was set to 30 kcal/mol. With standard options for scaling of variables, the regression analysis was carried out.

2.6. Partial least square (PLS) analysis

To form the basis for a statistical significant model, the method of partial least squares (PLS) regression was used to analyze the inhibitors by correlating variations in their biological activities with variations in their interaction fields. The optimum number of PLS components corresponding to the smallest standard error of prediction, was determined by the leave-one-out (LOO) cross-validation procedure. Using the optimal number of components, the final PLS analysis was carried out without cross-validation to generate a predictive model with a conventional correlation coefficient. The LOO cross-validation method might lead to high q^2 values, which do not necessarily reflect a general predictive ability of a model. Therefore further cross-validation, using five and two groups of approximately the same size in which the objects were assigned randomly, was performed. In this method 80% or 50% of the

Table 3b
List of actual pIC$_{50}$, predicted pIC$_{50}$ (derived from LOO cross-validation), calculated pIC$_{50}$ values and residual of each compound derived from the CoMFA Model 2.[a]

Compound number	Actual pIC$_{50}$	Predicted pIC$_{50}$	Residuals
1	5.2218	5.2687	−0.0469
2	5.6478	5.5739	0.0739
3[t]	5.9586	5.8945	0.0641
4	5.8239	5.9116	−0.0877
5	6.0969	6.0708	0.0261
6	6.0969	5.9588	0.1381
7	5.8239	5.6595	0.1644
8	5.6021	5.7434	−0.1413
9	5.4881	5.6603	−0.1722
10[t]	6.0000	6.1366	−0.1366
11	5.3010	5.4219	−0.1209
12	5.0223	5.0252	−0.0029
13	5.1549	5.1556	−0.0007
14	5.6021	5.4938	0.1083
15	5.3768	5.2153	0.1615
16	5.7328	5.6749	0.0579
17	5.3979	5.3531	0.0448
18	5.6021	5.7264	−0.1243
19	5.5229	5.5130	0.0099
20[t]	6.1549	6.4357	−0.2808
21	6.3010	6.3479	−0.0469
22	6.6990	6.6516	0.0474
23	6.0458	6.0958	−0.0500
24[t]	6.3979	6.5186	−0.1207
25	6.3979	6.3042	0.0937
26	7.0000	7.0210	−0.0210
27	6.8239	6.8572	−0.0333
28	6.0706	6.0853	−0.0147
29	7.2596	7.3618	−0.1022
30	6.0000	5.9622	0.0378
31	6.3468	6.3602	−0.0134
32	7.0000	7.0335	−0.0335
33	6.8861	6.8735	0.0126
34[t]	6.6478	6.5381	0.1097
35	5.0000	5.1092	−0.1092
36	6.5528	6.5419	0.0109
37	7.2218	7.2778	−0.0560
38[t]	7.0458	6.8151	0.2307
39	6.7447	6.8327	−0.0880
40	7.0458	6.9748	0.0710
41	7.2596	7.2610	−0.0014
42	7.0000	7.0001	−0.0001
43	6.6778	6.8112	−0.1334
44	7.0000	6.8652	0.1348
45	7.0969	7.0551	0.0418
46	7.2596	7.1828	0.0768
47	7.1549	7.0666	0.0883

[a] Compound number with "t" refers to those compounds included in the test set.

Table 3c
List of actual pIC$_{50}$, predicted pIC$_{50}$ (derived from LOO cross-validation), calculated pIC$_{50}$ values and residual of each compound derived from the CoMFA Model 3.[a]

Compound number	Actual pIC$_{50}$	Predicted pIC$_{50}$	Residuals
1	5.2218	5.2331	−0.0113
2	5.6478	5.5564	0.0914
3	5.9586	5.9530	0.0056
4[t]	5.8239	6.2867	−0.4628
5	6.0969	6.0846	0.0123
6	6.0969	6.0571	0.0398
7	5.8239	5.6591	0.1648
8[t]	5.6021	6.3723	−0.7702
9	5.4881	5.6836	−0.1955
10	6.0000	6.0267	−0.0267
11	5.3010	5.3896	−0.0886
12	5.0223	4.9798	0.0425
13	5.1549	5.2031	−0.0482
14	5.6021	5.4773	0.1248
15	5.3768	5.2488	0.1280
16[t]	5.7328	5.8310	−0.0982
17	5.3979	5.4457	−0.0478
18	5.6021	5.7348	−0.1327
19	5.5229	5.5013	0.0216
20	6.1549	6.2765	−0.1216
21	6.3010	6.2815	0.0195
22[t]	6.6990	6.2279	0.4711
23	6.0458	6.0932	−0.0474
24	6.3979	6.3393	0.0586
25	6.3979	6.3060	0.0919
26	7.0000	7.0401	−0.0401
27	6.8239	6.9161	−0.0922
28	6.0706	6.1353	−0.0647
29	7.2596	7.3213	−0.0617
30	6.0000	5.9563	0.0437
31[t]	6.3468	6.2862	0.0606
32	7.0000	6.9771	0.0229
33	6.8861	6.9242	−0.0381
34	6.6478	6.6394	0.0084
35	5.0000	5.0761	−0.0761
36	6.5528	6.5478	0.0050
37	7.2218	7.2412	−0.0194
38	7.0458	7.0296	0.0162
39	6.7447	6.7676	−0.0229
40[t]	7.0458	6.6064	0.4394
41	7.2596	7.2296	0.0300
42	7.0000	7.0147	−0.0147
43	6.6778	6.7615	−0.0837
44	7.0000	6.8289	0.1711
45	7.0969	7.0324	0.0645
46	7.2596	7.1562	0.1034
47	7.1549	7.1876	−0.0327

[a] Compound number with "t" refers to those compounds included in the test set.

compounds were randomly selected and a model is generated, which is then used to predict the remaining compounds (leave-20%-out, leave-50%-out). This cross-validation technique, especially the leave-50%-out procedure, has been shown to yield better indices for the robustness of a model than the normal LOO procedure [29,30].

2.7. Predictive correlation coefficient (r^2_{pred})

The predictive ability of the 3D-QSAR models were determined from a set of six compounds that were excluded during model development. The optimization and alignment of these test sets were the same as that of the training set compounds as described above, and their activities were predicted using the model produced by the training set. The predictive correlation coefficient (r^2_{pred}), based on the test set molecules, is computed using

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

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

3. Results and discussion

3.1. Docking study

In a preliminary docking study, we tested whether GOLD is able to correctly reproduce the position of the co-crystallized inhibitor. Docking of inhibitor into the FBPase binding site, using default settings and with considering the co-crystallized water molecules, showed good agreement with the experimentally derived binding mode (RMSD 0.507 Å).

Due to the structural similarity of the analyzed data set with the co-crystallized inhibitor, it is likely that all active compounds show the same orientation at the binding pocket of the FBPase. We employed default settings of the GOLD program to end up with a consistent alignment of all 47 benzimidazole-based derivatives as described in detail in the Methods Section. Forty-seven conformational binding
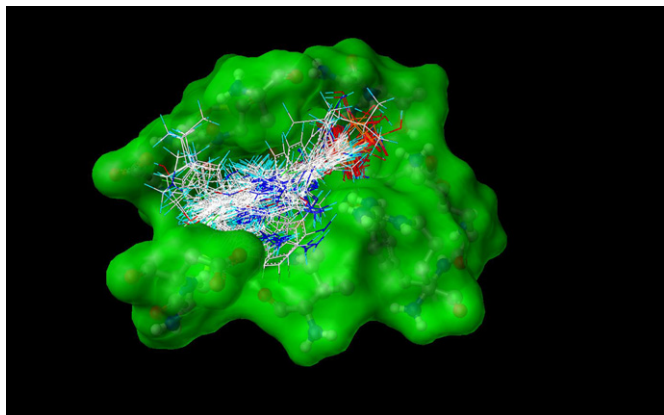
Fig. 1. Superimposition of 47 benzimidazoles at the binding site of FBPase for 3D-QSAR studies.

modes for each ligand were generated at the active site. Moreover, it was observed that the orientations of all 47 inhibitors as extracted from the GOLD docking are quite similar. When they were superimposed, a small but distinct shift of the benzimidazole scaffold was found (Fig. 1). Fig. 2, which was generated by LIGPLOT (version.4.5.3) [40], represents the interaction model of the docked inhibitor 41 with FBPase. Inhibitor 41 binds to the active sites and makes several interactions with the hinge-binding region of the enzyme. Beside the six hydrogen bonds of the phosphonate group of inhibitor 41 to Glu28, Thr27, Lys112, Tyr113 and Leu30, the most inhibitors including compound 41 form an additional hydrogen bond between the N3-amino group of the benzimidazoles and the side chain COOH of Asp178 (Fig. 2). On the other hand, the R$_1$, R$_2$, R$_3$ and R$_4$ groups of the inhibitors were located near the hydrophobic surface area formed by hydrophobic residues of Val17, Glu20, Gly21 and Met177. In addition, the 5-(2-phosphono)furanyl moieties were found near the hydrophobic surface area formed by hydrophobic residues of Ala24, Gly26
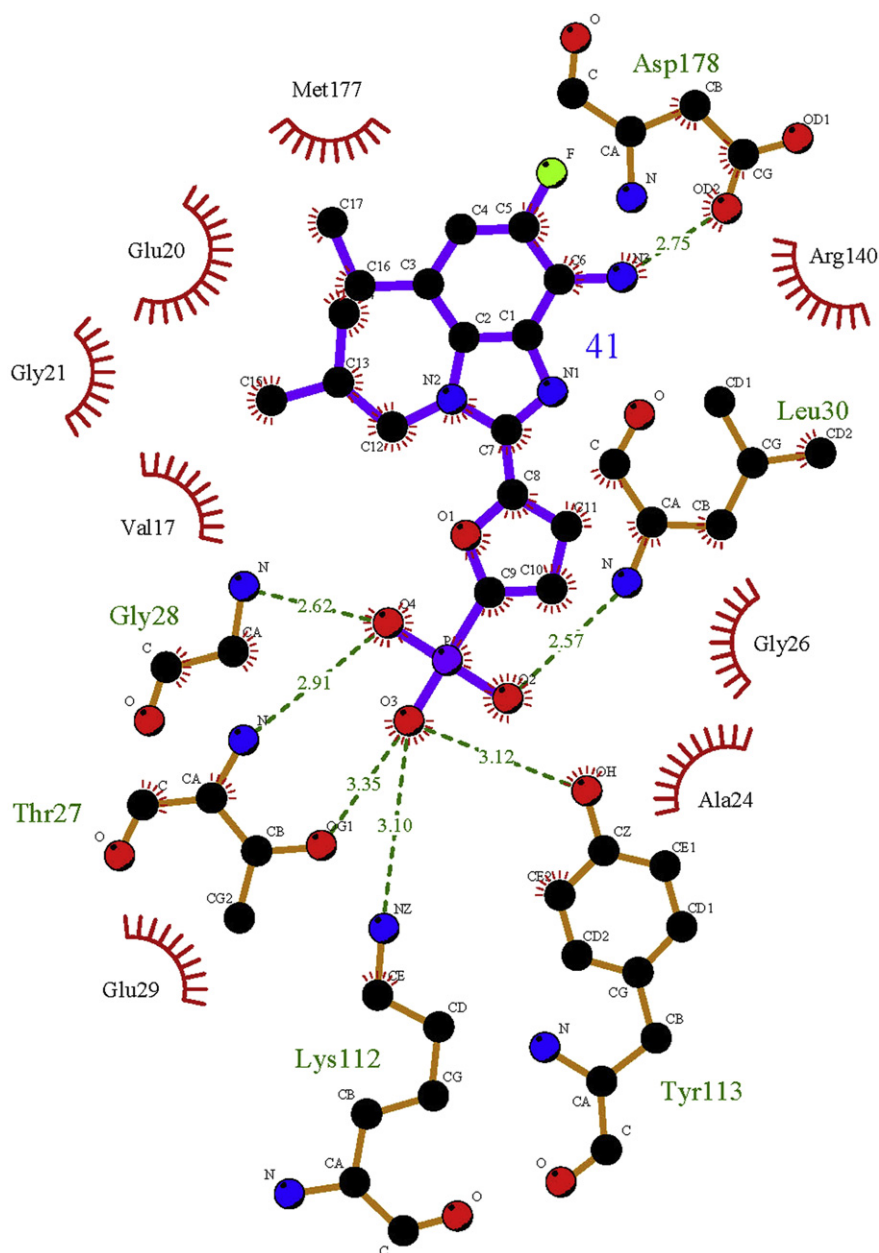
Fig. 2. Interaction model of the docked inhibitor 41 with FBPase.

and Glu29. Finally, the amino group of inhibitor 41 was found near the polar side chain of Arg140, suggesting that the electrostatic interactions might be happened in the area.

3.2. CoMFA model

Based on three random splits into training and test sets, three CoMFA models were obtained and named as Model 1, Model 2 and Model 3 respectively. Moreover their Statistical analyses were shown in Tables 2a−c and 3a−c. Forty-one benzimidazole-based inhibitors were randomly included in a training set for constructing CoMFA models, and the remaining 6 were used as a test set for each model validation. PLS analysis was carried out for the 41 binding conformations of each model, and the results are listed in Tables 2a−c respectively.

Model 1 showed that a CoMFA model with a leave-one-out cross-validated q^2 of 0.781 for six components was obtained. The non-cross-validated PLS analysis with the optimum components of 6 resulted in a conventional r^2 of 0.993, $F = 818.274$, and an estimated standard error of 0.062. The steric field descriptors explain 33.0% of the variance, while the electrostatic descriptors explain 67.0%. The predicted activities for the 47 inhibitors versus their experimental activities with their residues are listed in Table 3a, and the correlation between the predicted activities and the experimental activities depicted in Fig. 3a. Tables 2a and 3a demonstrate that the activities predicted by the constructed CoMFA model are in good agreement with the experimental data, suggesting that a reliable CoMFA model was successfully constructed.

In addition, further cross-validation, using five and two groups of approximately the same size in which the objects were assigned randomly, was performed. Thus, the q^2 of the leave-20%-out and the leave-50%-out were determined and it yielded 0.782 and 0.596 for leave-20%-out and leave-50%-out, respectively. The high q^2 value of the leave-20%-out and leave-50%-out, which are in the same range as the LOO value, indicates the robustness of the derived CoMFA model.

Models 2 and 3 using the different test set from Model 1 shown similar Statistical analyses in Tables 2b, c and 3b, c respectively. And the correlation between the predicted activities and the experimental activities for two models depicted in Figs. 3b and c respectively. These statistical analyses indicate the robustness and high predictive potential of their derived models.

3.3. CoMFA graphical contour plot

Since the contours of three CoMFA models are very similar, we chose the plots of Model 1 to describe our contour plot research. The CoMFA steric and electrostatic contours are shown in Figs. 4 and 5, respectively. The steric interaction is represented by green and yellow contours, in which green-coloured regions indicate areas where increased steric bulk is associated with enhanced activity, and yellow regions suggest areas where increased steric bulk is unfavourable. Electrostatic interaction is indicated by red and blue contours, among which blue-coloured regions show areas where more positively charged groups are favoured, and red regions highlight areas where increased negatively charged groups are favoured. In addition, the CoMFA model plots were mapped back to the binding sites of FBPase, to get a better understanding of vital interactions between the inhibitors and the FBPase enzyme.

Two large and one small yellow contours are observed near Ala24, Arg140, Val160 and Met177 where the amino group and R2-subsitutent of the benzimidazole ring of potent inhibitors are blocked by the kinase, which can explain the low activity of compounds 18 and 19. In addition, five small sterically disfavored yellow regions are close to the phosphonate groups of the inhibitors, which are blocked by the
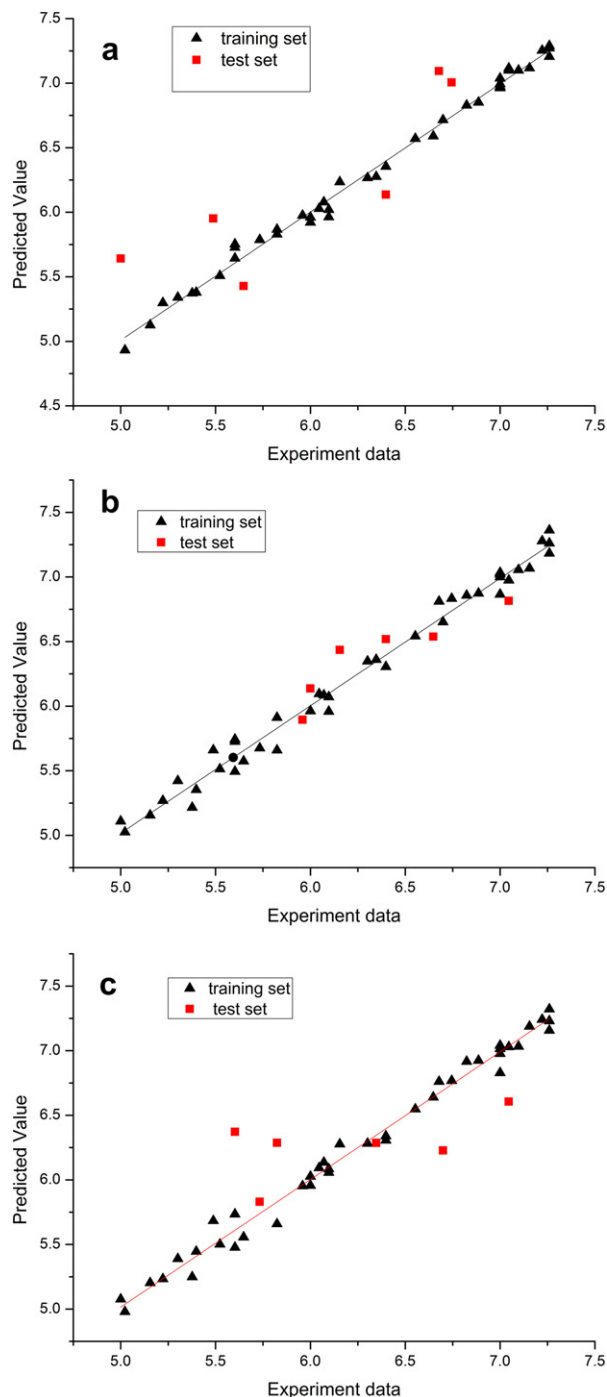

Fig. 3. a. Correlation between predicted activities by the CoMFA model 1 and the experimental pIC50 values of training and test sets. Black triangles represent predictions for the training set, while red filled rectangles represent predictions for the test set. b. Correlation between predicted activities by the CoMFA model 2 and the experimental pIC50 values of training and test sets. Black triangles represent predictions for the training set, while red filled rectangles represent predictions for the test set. c. Correlation between predicted activities by the CoMFA model 3 and the experimental pIC50 values of training and test sets. Black triangles represent predictions for the training set, while red filled rectangles represent predictions for the test set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Met18, Gly26, Thr27, Gly28, Thr31, Glu29, Lys112 and Tyr113, indicating that the binding positions are important for high activity. On the other hand, three green contours are found near the R1, R3 and R4-subsitutents of the benzimidazole ring of potent inhibitors, which
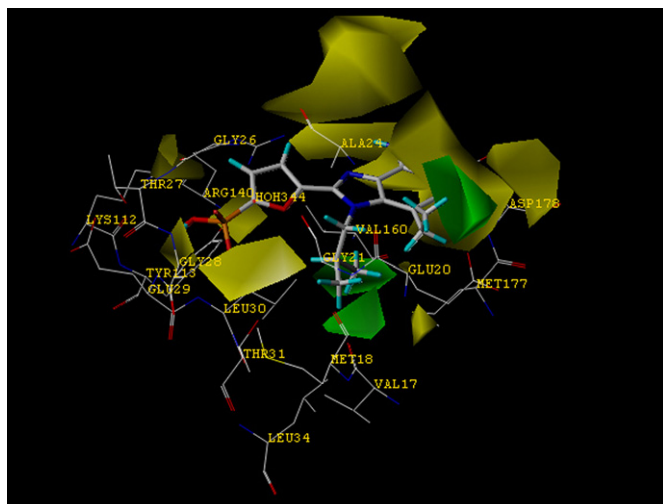
Fig. 4. CoMFA steric field contour maps in combination with inhibitor 41 within the active site of the complex structure of FBPase. Sterically favoured areas are in green, and sterically unfavored areas are in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

are surrounded by Val17, Glu20, Gly21 and Met177 that form a moderately lipophilic binding pocket in FBPase enzyme. Compounds which contain middle sized hydrophobic substituents such as iBu, cPr–CH2–, Ph and (Me)$_2$CH(CH$_2$)$_3$– may have well hydrophobic interaction with these residues. This is a possible reason why compounds 26, 27, 29, 41 and 44 have high potency.

Two red electrostatic contours are located around R$_2$-subsitutents of the benzimidazole ring of potent inhibitors, indicating that more negatively charged groups can favourably interact with the surrounding residue Asp178. The docking of the potent inhibitor 41 showed that the amino group is making a hydrogen bond to Asp178 (Fig. 2). Natively electron-withdrawing R$_2$-subsitutents such as –F, and –Cl can make the neighbor amino group more positive, So a more strong hydrogen bond may be formed to the side chain carboxyl group of Asp178. This observation is in agreement with the experimental data: compounds 26, 27, 32, 33, 37, 38, 41, 42, 44, 45, 46 and 47 bearing a partial positive charge at neighbor amino group show high pIC$_{50}$ values. In addition, three red contours can be observed near Thr27, Gly28, Leu30 and Lys112, respectively. One big blue contour is found around Thr27 and
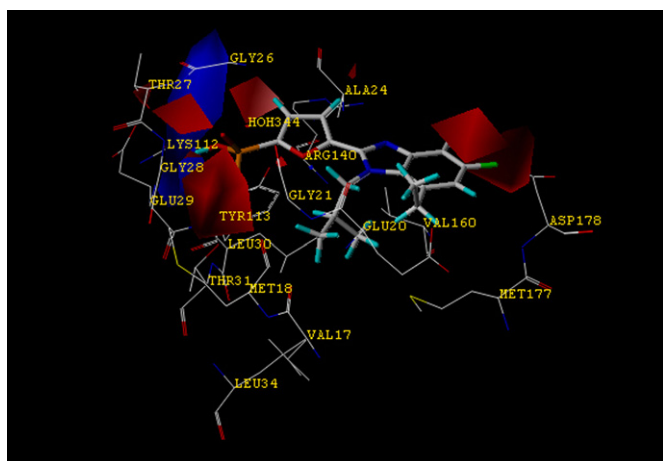
Tyr113. As shown in Fig. 2, the phosphonate group of highly active inhibitors 26, 27, 29, 32, 33, 37, 38, 40, 41, 42, 44, 45, 46 and 47, is making hydrogen bonds with the near residues (Fig. 2). These results imply that the phosphonate group of FBPase inhibitors can play as key role in the interaction with FBPase [41–43].

3.4. Validation of the 3D-QSAR models

Test sets are generally used to evaluate the external predictive capabilities of QSAR models. Based on three random splits into training and test sets, three CoMFA models were obtained and named as Models 1, 2 and 3 respectively. The six randomly selected compounds were used as a test set for each model to verify the constructed CoMFA models. The calculated results are listed in Tables 3a–c and displayed in Figs. 3a–c (red rectangles). The predicted pIC$_{50}$s with the QSAR models were in good agreement with the experimental data within a statistically tolerable error range (−0.5 to plus;0.5). In addition, in the CoMFA models, three predictive correlation coefficient of $r^2_{pred} = 0.679$, $r^2_{pred} = 0.819$ and $r^2_{pred} = 0.568$ were found for each CoMFA model respectively (Tables 2a–c). The test results indicated that the CoMFA models would be reliable for use in new FBPase inhibitor design.

There's one more thing to consider. Analyzing the test set of Model 1, we found that ligand 35 have high residual value of actual pIC$_{50}$ vs. predicted pIC$_{50}$ (−0.6413), indicating that the ligand is a truly ligand outlier. We tried to find an explanation for that. First, we checked the residual values of ligand 35 in the training sets of Model 2, 3(−0.1092 in Model 2 and −0.0761 in Model 3) and found they were within the tolerable error range, indicating that the high residual value of ligand 35 in the test set of Model 1 is a statistical problem. Then we discovered that Compounds 30−47 belong to the same compound cluster according to their skeleton and all of them have high pIC$_{50}$s except the pIC$_{50}$ of ligand 35 (5). Finally we drew the conclusion that in case of compound 35, which is the only low pIC$_{50}$ value among its cluster and out of the training set, the 3D-QSAR Model 1 is not able to cover this difference. In the same way, the similar reason might suit Compound 8, the ligand outlier in the test set of the CoMFA Model 3, which is the only compound with a unique cyclohexyl substituent group conformation among its cluster.

4. Conclusion

In summary, 3D-QSAR analyses have been performed on 47 compounds using molecular docking method and CoMFA. Satisfactory CoMFA models were obtained to predict the activities of test set structures. The steric and electrostatic recognition sites of the peripheral binding site as well as the superimposed CoMFA contours on the receptor sites will guide the design of novel structures which demonstrate optimal binding to and inhibition of FBPase.

Acknowledgments

Fig. 5. CoMFA electrostatic field contour maps in combination with inhibitor 41 within the active site of the complex structure of FBPase. Positive potential favoured areas are in blue; positive potential unfavored areas are in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

References

[1] S.J. Pilkis, M.R. El-Maghrabi, J. Pilkis, T. Claus, Inhibition of fructose-1,6-bisphosphatase by fructose 2,6-bisphosphate, J. Biol. Chem. 256 (1981) 3619−3622.

[2] S.J. Benkovic, M.M. deMaine, Mechanism of action of fructose 1,6-bisphosphatase, Adv. Enzymol. Relat. Areas Mol. Biol. 53 (1982) 45–82.

[3] M. Gidh-Jain, Y. Zhang, P.D. van Poelje, J.Y. Liang, S. Huang, J. Kim, J.T. Elliott, M.D. Erion, S.J. Pilkis, M. Raafat el-Maghrabi, et al., The allosteric site of human liver fructose-1,6-bisphosphatase. Analysis of six AMP site mutants based on the crystal structure, J. Biol. Chem. 269 (1994) 27,732–27,738.

[4] Y. Zhang, J.Y. Liang, S. Huang, W.N. Lipscomb, Toward a mechanism for the allosteric transition of pig kidney fructose-1,6-bisphosphatase, J. Mol. Biol. 244 (1994) 609–624.

[5] M.D. Erion, P.D. van Poelje, Q. Dang, S.R. Kasibhatla, S.C. Potter, M.R. Reddy, K.R. Reddy, T. Jiang, W.N. Lipscomb, MB06322 (CS-917): a potent and selective inhibitor of fructose 1,6-bisphosphatase for controlling gluconeogenesis in type 2 diabetes, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 7970–7975.

[6] C.V. Iancu, S. Mukund, H.J. Fromm, R.B. Honzatko, R-state AMP complex reveals initial steps of the quaternary transition of fructose-1,6-bisphosphatase, J. Biol. Chem. 280 (2005) 19737–19745.

[7] M.D. Erion, Q. Dang, M.R. Reddy, S.R. Kasibhatla, J. Huang, W.N. Lipscomb, P.D. van Poelje, Structure-guided design of AMP mimics that inhibit fructose-1,6-bisphosphatase with high affinity and specificity, J. Am. Chem. Soc. 129 (2007) 15480–15490.

[8] P.D. van Poelje, Q. Dang, M.D. Erion, Discovery of fructose-1,6-bisphosphatase inhibitors for the treatment of type 2 diabetes, Curr. Opin. Drug Discov. Develop. 10 (2007) 430–437.

[9] J.Y. Choe, S.W. Nelson, K.L. Arienti, F.U. Axe, T.L. Collins, T.K. Jones, R.D. Kimmich, M.J. Newman, K. Norvell, W.C. Ripka, S.J. Romano, K.M. Short, D.H. Slee, H.J. Fromm, R.B. Honzatko, Inhibition of fructose-1,6-bisphosphatase by a new class of allosteric effectors, J. Biol. Chem. 278 (2003) 51176–51183.

[10] Q. Dang, B.S. Brown, Y. Liu, R.M. Rydzewski, E.D. Robinson, P.D. van Poelje, M.R. Reddy, M.D. Erion, Fructose-1,6-bisphosphatase inhibitors. 1. Purine phosphonic acids as novel AMP mimics, J. Med.Chem. 52 (2009) 2880–2898.

[11] Q. Dang, S.R. Kasibhatla, T. Jiang, K. Fan, Y. Liu, F. Taplin, W. Schulz, D.K. Cashion, K.R. Reddy, P.D. van Poelje, J.M. Fujitaki, S.C. Potter, M.D. Erion, Discovery of phosphonic diamide prodrugs and their use for the oral delivery of a series of fructose-1,6-bisphosphatase inhibitors, J. Med. Chem. 51 (2008) 4331–4339.

[12] Q. Dang, S.R. Kasibhatla, K.R. Reddy, T. Jiang, M.R. Reddy, S.C. Potter, J.M. Fujitaki, P.D. van Poelje, J. Huang, W.N. Lipscomb, M.D. Erion, Discovery of potent and specific fructose-1,6-bisphosphatase inhibitors and a series of orally-bioavailable phosphoramidase-sensitive prodrugs for the treatment of type 2 diabetes, J. Am. Chem. Soc. 129 (2007) 15491–15502.

[13] Q. Dang, S.R. Kasibhatla, W. Xiao, Y. Liu, J. Dare, F. Taplin, K.R. Reddy, G.R. Scarlato, T. Gibson, P.D. van Poelje, S.C. Potter, M.D. Erion, Fructose-1,6-bisphosphatase inhibitors. 2. Design, synthesis, and structure–activity relationship of a series of phosphonic acid containing benzimidazoles that function as 5′-adenosinemonophosphate (AMP) mimics, J. Med. Chem. 53 (2010) 441–451.

[14] Q. Dang, S. Mazumdar, S.J. Anderson, P.R. Houck, C.F. Reynolds, Using trajectories from a bivariate growth curve as predictors in a Cox regression model, Stat. Med. 26 (2007) 800–811.

[15] P. Hebeisen, B. Kuhn, P. Kohler, M. Gubler, W. Huber, E. Kitas, B. Schott, J. Benz, C. Joseph, A. Ruf, Allosteric FBPase inhibitors gain 10(5) times in potency when simultaneously binding two neighboring AMP sites, Bioorg. Med. Chem. Lett. 18 (2008) 4708–4712.

[16] C. Lai, R.J. Gum, M. Daly, E.H. Fry, C. Hutchins, C. Abad-Zapatero, T.W. von Geldern, Benzoxazole benzenesulfonamides as allosteric inhibitors of fructose-1,6-bisphosphatase, Bioorg. Med. Chem. Lett. 16 (2006) 1807–1810.

[17] B.E. Maryanoff, A.B. Reitz, G.F. Tutwiler, S.J. Benkovic, P.A. Benkovic, S.J. Pilkis, Stereoselective synthesis and biological activity of beta- and alpha-D-arabinose 1,5-diphosphate: analogs of a potent metabolic regulator, J. Am. Chem. Soc. 106 (1984) 7851–7853.

[18] A. Rudnitskaya, K. Huynh, B. Torok, K. Stieglitz, Novel heteroaromatic organofluorine inhibitors of fructose-1,6-bisphosphatase, J. Med. Chem. 52 (2009) 878–882.

[19] T. Tsukada, O. Kanno, T. Yamane, J. Tanaka, T. Yoshida, A. Okuno, T. Shiiki, M. Takahashi, T. Nishi, Discovery of potent and orally active tricyclic-based FBPase inhibitors, Bioorg. Med. Chem (2010).

[20] T. Tsukada, M. Takahashi, T. Takemoto, O. Kanno, T. Yamane, S. Kawamura, T. Nishi, Synthesis, SAR, and X-ray structure of tricyclic compounds as potent FBPase inhibitors, Bioorg. Med. Chem. Lett. 19 (2009) 5909–5912.

[21] T. Tsukada, M. Takahashi, T. Takemoto, O. Kanno, T. Yamane, S. Kawamura, T. Nishi, Structure-based drug design of tricyclic 8H-indeno[1,2-d][1,3]thiazoles as potent FBPase inhibitors, Bioorg. Med. Chem. Lett. 20 (2010) 1004–1007.

[22] T. Tsukada, K. Tamaki, J. Tanaka, T. Takagi, T. Yoshida, A. Okuno, T. Shiiki, M. Takahashi, T. Nishi, A prodrug approach towards the development of tricyclic-based FBPase inhibitors, Bioorg. Med. Chem. Lett. 20 (2010) 2938–2941.

[23] P.D. van Poelje, S.C. Potter, V.C. Chandramouli, B.R. Landau, Q. Dang, M.D. Erion, Inhibition of fructose 1,6-bisphosphatase reduces excessive endogenous glucose production and attenuates hyperglycemia in Zucker diabetic fatty rats, Diabetes 55 (2006) 1747–1754.

[24] T.W. von Geldern, C. Lai, R.J. Gum, M. Daly, C. Sun, E.H. Fry, C. Abad-Zapatero, Benzoxazole benzenesulfonamides are novel allosteric inhibitors of fructose-1,6-bisphosphatase with a distinct binding mode, Bioorg. Med. Chem. Lett. 16 (2006) 1811–1815.

[25] S.W. Wright, A.A. Carlo, M.D. Carty, D.E. Danley, D.L. Hageman, G.A. Karam, C.B. Levy, M.N. Mansour, A.M. Mathiowetz, L.D. McClure, N.B. Nestor, R.K. McPherson, J. Pandit, L.R. Pustilnik, G.K. Schulte, W.C. Soeller, J.L. Treadway, I.K. Wang, P.H. Bauer, Anilinoquinazoline inhibitors of fructose 1,6-bisphosphatase bind at a novel allosteric site: synthesis, in vitro characterization, and X-ray crystallography, J. Med. Chem. 45 (2002) 3865–3877.

[26] S.W. Wright, A.A. Carlo, D.E. Danley, D.L. Hageman, G.A. Karam, M.N. Mansour, L.D. McClure, J. Pandit, G.K. Schulte, J.L. Treadway, I.K. Wang, P.H. Bauer, 3-(2-carboxyethyl)-4,6-dichloro-1H-indole-2-carboxylic acid: an allosteric inhibitor of fructose-1,6-bisphosphatase at the AMP site, Bioorg. Med. Chem. Lett. 13 (2003) 2055–2058.

[27] S.W. Wright, D.L. Hageman, L.D. McClure, A.A. Carlo, J.L. Treadway, A.M. Mathiowetz, J.M. Withka, P.H. Bauer, Allosteric inhibition of fructose-1,6-bisphosphatase by anilinoquinazolines, Bioorg. Med. Chem. Lett. 11 (2001) 17–21.

[28] T. Yoshida, A. Okuno, M. Izumi, K. Takahashi, Y. Hagisawa, J. Ohsumi, T. Fujiwara, CS-917, a fructose 1,6-bisphosphatase inhibitor, improves postprandial hyperglycemia after meal loading in non-obese type 2 diabetic Goto-Kakizaki rats, Eur. J. Pharmacol. 601 (2008) 192–197.

[29] W. Sippl, Receptor-based 3D QSAR analysis of estrogen receptor ligands–merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods, J. Comput. Aided Mol. Des. 14 (2000) 559–572.

[30] W. Sippl, J.M. Contreras, I. Parrot, Y.M. Rival, C.G. Wermuth, Structure-based 3D QSAR and design of novel acetylcholinesterase inhibitors, J. Comput. Aided Mol. Des. 15 (2001) 395–410.

[31] SYBYL6.9, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO 63144, USA.

[32] T.A. Halgren, Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, J. Comput. Chem. 17 (1996) 490–519.

[33] T.A. Halgren, Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules, J. Comput. Chem. 17 (1996) 616–641.

[34] T.A. Halgren, Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions, J. Comput. Chem. 17 (1996) 520–552.

[35] T.A. Halgren, Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94, J. Comput. Chem. 17 (1996) 553–586.

[36] T.A. Halgren, R.B. Nachbar, Merck molecular force field. IV. conformational energies and geometries for MMFF94, J. Comput. Chem. 17 (1996) 587–615.

[37] M. Clark, N.V. Opdenbosch, Validation of the general purpose tripos 5.2 force field, J. Comput. Chem. 10 (1989) 982–1012.

[38] G. Jones, P. Willett, R.C. Glen, Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation, J. Mol. Biol. 245 (1995) 43–53.

[39] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.

[40] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, Protein Eng. 8 (1995) 127–134.

[41] J. Yan, L. Sun, G. Wu, P. Yi, F. Yang, L. Zhou, X. Zhang, Z. Li, X. Yang, H. Luo, M. Qiu, Rational design and synthesis of highly potent anti-acetylcholinesterase activity huperzine A derivatives, Bioorg. Med. Chem. 17 (2009) 6937–6941.

[42] P. Yi, X. Fang, M. Qiu, 3D-QSAR studies of Checkpoint Kinase Weel inhibitors based on molecular docking, CoMFA and CoMSIA, Eur. J. Med. Chem. 43 (2008) 925–938.

[43] P. Yi, M. Qiu, 3D-QSAR and docking studies of aminopyridine carboxamide inhibitors of c-Jun N-terminal kinase-1, Eur. J. Med. Chem. 43 (2008) 604–613.