3D-QSAR studies and molecular docking on [5-(4-amino-1*H*-benzoimidazol-2-yl)-furan-2-yl]-phosphonic acid derivatives as fructose-1,6-biphophatase inhibitors

Ping Lan · Mei-Qi Xie · Yue-Mei Yao · Wan-Na Chen · Wei-Min Chen

Abstract Fructose-1,6-biphophatase has been regarded as a novel therapeutic target for the treatment of type 2 diabetes mellitus (T2DM). 3D-QSAR and docking studies were performed on a series of [5-(4-amino-1*H*-benzoimidazol-2-yl)-furan-2-yl]-phosphonic acid derivatives as fructose-1,6-biphophatase inhibitors. The CoMFA and CoMSIA models using thirty-seven molecules in the training set gave r_{cv}^2 values of 0.614 and 0.598, r^2 values of 0.950 and 0.928, respectively. The external validation indicated that our CoMFA and CoMSIA models possessed high predictive powers with r_0^2 values of 0.994 and 0.994, r_m^2 values of 0.751 and 0.690, respectively. Molecular docking studies revealed that a phosphonic group was essential for binding to the receptor, and some key features were also identified. A set of forty new analogues were designed by utilizing the results revealed in the present study, and were predicted with significantly improved potencies in the developed models. The findings can be quite useful to aid the designing of new fructose-1,6-biphophatase inhibitors with improved biological response.

Keywords 3D-QSAR · CoMFA · CoMSIA · Docking · Fructose-1,6-biphophatase

P. Lan · M.-Q. Xie · Y.-M. Yao · W.-N. Chen · W.-M. Chen (✉)
Guangdong Province Key Laboratory of Pharmacodynamic Constituents of TCM and New Drugs Research, College of Pharmacy, Jinan University, 510632 Guangzhou, People's Republic of China
e-mail: twmchen@jnu.edu.cn

Introduction

Fructose-1,6-bisphosphatase (FBPase, EC 3.1.3.11), a well-known regulatory enzyme of the hepatic gluconeogenesis (GNG) pathway, catalyzes the irreversible reaction of hydrolysis of fructose-1,6-bisphosphate to fructose-6-phosphate [1–4]. GNG is one of the main ways responsible for the excessive hepatic glucose production (HGP) which result in increased fasting blood glucose and postprandial hyperglycemia [5–7]. Thus, FBPase inhibitors can lower blood glucose levels by reducing hepatic glucose output and are recognized as a novel therapeutic target for the treatment of type 2 diabetes mellitus (T2DM) [6–8]. FBPase is an allosteric enzyme composed of four identical chains. Regulation of FBPase enzymatic activity involves changes in quaternary structure between the active (R) and inactive (T) conformational states [8]. AMP (Fig. 1) binds to the allosteric site and inhibits the FBPase by shifting the enzyme from R to T conformation [3]. Some investigations on the design and synthesis of AMP mimetics were carried out, and many mimetics capable of retaining the key binding interactions of AMP in the allosteric binding site of FBPase and lowing the blood glucose as potent and specific FBPase inhibitors have been obtained. By using a structure-guided drug design strategy, MB05032 (Fig. 1), one of the most famous AMP mimetics, showed robust glucose lowering activity in animal models as well as patients with T2DM [1, 7, 9].

A series of [5-(4-amino-1*H*-benzoimidazol-2-yl)-furan-2-yl]-phosphonic acid derivatives that function as AMP mimetics with FBPase inhibitory activities were reported by Dang and co-workers [1]. The novel benzimidazole analogues with excellent FBPase inhibitory activities were designed and synthesized by replacing the purine scaffold with benzoimidazole. Though the preliminary
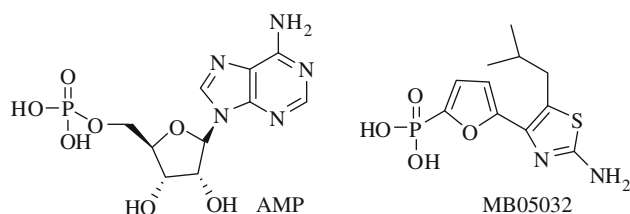
Fig. 1 Structure of AMP and MB05032

structure–activity relationship was discussed in the original paper, further investigations on the structure–activity relationship and the interaction between these inhibitors and the FBPase are still needed. In this paper, 3D-QSAR and docking approaches were applied to study this purpose. 3D-QSAR methods including comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were performed to predict the activities of these inhibitors and provided the regions in space where interactive fields may influence the activity. To evaluate the true predictive power of the 3D-QSAR models, a systemic external validation was employed. Docking was applied to investigate the protein–ligand interactions. Based on the good performance of the 3D-QSAR and docking studies, the developed models can not only helped in understanding the structure–activity relationship of these compounds but also be served as a useful guide for the design of new inhibitors with better activities. We have designed a series of novel benzimidazoles anlogues by utilizing the structure analysis results obtained from present studies, which exhibited excellent predicted activities in the CoMFA and CoMSIA models established. Meanwhile, based on the excellent performance of the external validation, the predicted activities of these newly designed derivatives may be reliable.

Materials and methods

Data sets

All the compounds and associated activity data involved in this study were obtained from literature [1]. The inhibitory activity data were reported as IC_{50} against human liver FBPase. IC_{50} was converted to pIC_{50} by taking Log $(1/IC_{50})$, the pIC_{50} values were used as the dependent variables in all the models subsequently developed. A total of 45 compounds were randomly segregated into training and test sets comprising 37 and 8 molecules, respectively. Structures and associated inhibitory activities were listed in Tables 1 and 2.

Table 1 The structures of the training and test set molecules

Compd. no.	Substituent				
	R_1	R_4	R_5	R_6	R_7
1	Me	NH_2	H	H	H
2	Et	NH_2	H	H	H
3	nPr	NH_2	H	H	H
4	iBu	NH_2	H	H	H
5	Cyclopropyl-CH_2–	NH_2	H	H	H
6	Cyclobutyl-CH_2–	NH_2	H	H	H
7	Cyclopentyl-CH_2–	NH_2	H	H	H
8	Cyclohexyl-CH_2–	NH_2	H	H	H
9	Cycloheptyl-CH_2–	NH_2	H	H	H
10	Norbonyl	NH_2	H	H	H
11	Benzyl	NH_2	H	H	H
12	4-tBu-benzyl	NH_2	H	H	H
13	4-CF_3-benzyl	NH_2	H	H	H
14	4-Ph-benzyl	NH_2	H	H	H
15	3-furnyl-CH_2–	NH_2	H	H	H
16	3-HO-benzyl	NH_2	H	H	H
17	3-thienyl-CH_2–	NH_2	H	H	H
18	iBu	NH_2	Et	H	H
19	iBu	NH_2	nPr	H	H
20	iBu	NH_2	MeO	H	H
21	iBu	NH_2	HO	H	H
22	iBu	NH_2	Cl	H	H
23	iBu	NH_2	H	H	Cl
24	iBu	NH_2	Br	H	H
25	iBu	NH_2	H	H	Br
26	iBu	NH_2	F	H	H
27	$(Et)_2CHCH_2$–	NH_2	F	H	H
28	$(Et)_2CH$–	NH_2	F	H	H
29	cPr-CH_2–	NH_2	F	H	H
30	iBu	NH_2	Br	H	Br
31	iBu	NH_2	Cl	H	Cl
32	iBu	NH_2	F	H	Br
33	iBu	NH_2	F	Cl	H
34	iBu	NH_2	F	H	Vinyl
35	iBu	NH_2	F	H	cPr
36	iBu	NH_2	F	H	Ph
37	iBu	NH_2	F	H	4-F-Ph
38	iBu	NH_2	F	H	4-Cl-Ph
39	iBu	NH_2	F	H	Et
40	iBu	NH_2	F	H	nPr
41	iBu	NH_2	F	H	$tBu(CH_2)_2$
42	iBu	NH_2	F	H	$(Me)_2CH(CH_2)_3$–
43	iBu	NH_2	F	H	$HO(CH_2)_3$–
44	iBu	NH_2	F	H	$(Me)_2N(CH_2)_3$–
45	iBu	NH_2	F	H	$Cl(CH_2)_4$–

Table 2 The actual pIC$_{50}$s, predicted pIC$_{50}$s (Pred.), their residuals (Res.) and docking score values of the training and test set molecules

| No.	pIC$_{50}$ Actual	CoMFA Pred.	CoMFA Res.	CoMSIA Pred.	CoMSIA Res.	Docking Score
1	5.222	5.394	−0.172	5.485	−0.263	4.26
2	5.648	5.739	−0.091	5.720	−0.072	4.68
3*	5.959	5.712	0.247	5.580	0.379	6.49
4*	5.824	6.078	−0.254	6.038	−0.214	4.67
5	6.097	5.952	0.145	6.103	−0.006	5.11
6	6.097	6.080	0.017	6.014	0.083	4.20
7	5.824	5.751	0.073	5.828	−0.004	4.10
8	5.602	5.716	−0.114	5.770	−0.168	4.95
9	5.488	5.621	−0.133	5.516	−0.028	4.76
10	6.000	5.849	0.151	5.832	0.168	5.15
11*	5.301	5.487	−0.186	5.443	−0.142	5.56
12	5.022	4.908	0.114	5.063	−0.041	3.95
13	5.155	5.118	0.037	5.207	−0.052	6.60
14	5.602	5.545	0.057	5.308	0.294	5.59
15	5.377	5.324	0.053	5.257	0.120	6.23
16	5.733	5.865	−0.132	5.728	0.005	5.12
17	5.398	5.332	0.066	5.554	−0.156	6.04
18	5.602	5.679	−0.077	5.896	−0.294	5.81
19	5.523	5.427	0.096	5.473	0.050	4.64
20	6.155	6.127	0.028	6.023	0.132	5.92
21*	6.301	6.457	−0.156	6.225	0.076	4.85
22	6.699	6.487	0.212	6.483	0.216	5.28
23	6.046	6.275	−0.229	6.179	−0.133	5.04
24*	6.398	6.402	−0.004	6.562	−0.164	5.09
25*	6.398	6.277	0.121	6.347	0.051	6.39
26	7.000	6.585	0.415	6.636	0.364	5.68
27	6.824	6.946	−0.122	6.939	−0.115	5.36
28	6.071	6.379	−0.308	6.129	−0.058	5.55
29	7.260	7.146	0.114	7.292	−0.032	4.65
30	6.000	5.999	0.001	6.001	−0.001	2.91
31	6.347	6.632	−0.285	6.468	−0.121	3.82
32	6.886	6.957	−0.071	6.955	−0.069	3.97
33	6.648	6.754	−0.106	7.136	−0.488	4.56
34	6.553	6.735	−0.182	6.576	−0.023	4.98
35	7.222	7.131	0.091	6.851	0.371	4.78
36	7.046	6.838	0.208	6.718	0.328	6.42
37	6.745	6.822	−0.077	6.694	0.051	5.86
38*	7.046	6.850	0.196	6.637	0.409	5.78
39*	7.260	6.969	0.291	6.832	0.428	4.13
40	7.000	6.987	0.013	6.889	0.111	5.73
41	6.678	6.709	−0.031	6.993	−0.315	5.30
42	7.000	6.909	0.091	6.830	0.170	4.83
43	7.097	7.013	0.084	7.149	−0.052	4.89
44	7.260	7.026	0.234	7.221	0.039	6.19
45	7.155	7.322	−0.167	7.160	−0.005	6.12

* Test set molecules

Molecular modeling and alignment

Structures of entire sets of [5-(4-amino-1*H*-benzoimidazol-2-yl)-furan-2-yl]-phosphonic acid derivatives were built using SYBYL 8.1 program package of Tripos, Inc. [10]. 3D structures of all compounds were constructed using the Sketch Molecule module. Structural energy minimization was performed using the standard Tripos molecular mechanics force field and Gasteiger-Hückel charge [11]. All of the structures were aligned into a lattice box by fitting with benzimidazole group as a common structure using compound **44** as a template, which was one of the most active compounds. The aligned molecules were shown in Fig. 2.

CoMFA and CoMSIA setup

The CoMFA and CoMSIA descriptor fields was obtained by a 3D cubic lattice with grid spacing of 1 Å and extending to 4 Å units beyond the aligned molecules in all directions [11]. The Van Der Waals potentials and Coulombic terms were calculated by using Tripos force field. In CoMFA method a sp^3 hybridized carbon atom with a charge of 1e served as the probe atom to calculate steric and electrostatic fields, in which their energy values were truncated at 30 kcal/mol [12]. The CoMSIA method with the same lattice box that was used in CoMFA, incorporating steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields, were carried out using a probe atom with radius 1.0 Å, +1.0 charge, and hydrophobic and hydrogen bond properties of +1. A default value of 0.3 was used as the attenuation factor [13].
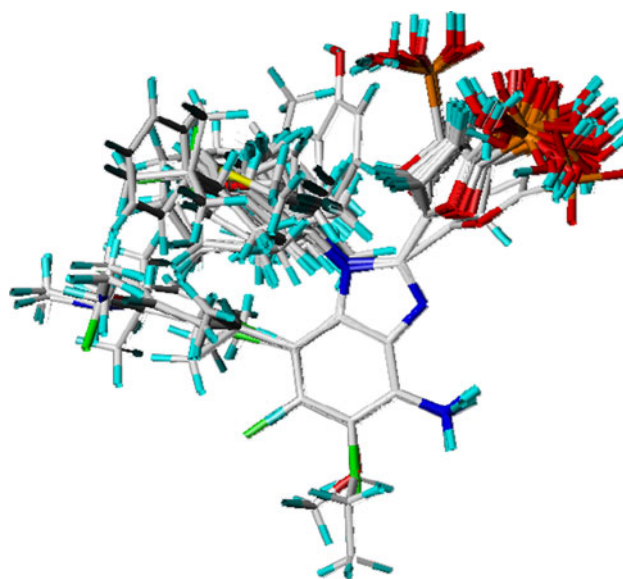
Fig. 2 Alignment of the compounds used in the training set

Regression analysis and models validation

Partial Least Squares (PLS) was used to linearly correlate the CoMFA and CoMSIA fields to the pIC_{50} values. The performance of both the CoMFA and CoMSIA models were evaluated using the leave-one-out (LOO) method [14]. PLS was conjunct with the cross-validation option to determine the optimum number of components (ONC) which were then used in deriving the final CoMFA and CoMSIA model without cross-validation. The ONC was the number of components resulted in highest cross-validated correlated correlation coefficient (r_{cv}^2) [15]. Column filtering was used at the default value of 2.0 kcal/mol in the cross-validation part. The final models were developed with ONC by using non-cross-validated analysis equal yielded the highest correlation coefficient (r^2) [16].

Predictive correlation co-efficient (r_{pred}^2)

The predictive abilities were determined from a test set of eight compounds that were not included in the training set. These molecules were aligned to the template and their pIC_{50} values were predicted. The predictive correlation coefficient (r_{pred}^2), based on the molecules of test set, was defined as follows: $r_{pred}^2 = $ (SD-PRESS)/SD.

SD is the sum of the squared deviations between the inhibitory activities of the test set and mean activities of the training molecules and PRESS is the sum of squared deviations between predicted and actual activity values for each molecule in the test set [17–20].

External validation

It was revealed in the previous researches that a high cross-validated correlation coefficient (r_{cv}^2) may be the necessary condition for a 3D-QSAR model to have a high predictive power. Nevertheless, it is not a sufficient condition. In fact, the low values of r_{cv}^2 and r_{pred}^2 can serve as an indicator of a low predictive ability of a model, however, the opposite is not necessarily true [21]. In many cases, a model with high r_{cv}^2 and r_{pred}^2 values can be proved to be inaccurate. Even though a model may exhibit a good predictive ability based on the statistics for the test set, it is not always sure that the model will perform well on a new set of data [22]. The only way to estimate the true predictive power of a model is to test it on an external validation. To evaluate the true predictive abilities of the established models, both the CoMFA and CoMSIA models were subjected to rigorous external validation process. Several statistics such as r_m^2, r_0^2, R and k were employed according to literature [21].

We defined $\tilde{y}_i$ and y_i as the predicted and observed pIC_{50} values of the test set molecules, respectively. If we plot y versus $\tilde{y}$ for the ideal QSAR model, the regression line

will bisect the angle formed by positive directions of the orthogonal axes $\tilde{y}$ and y. The regression line can be described by expression $y^r = a\tilde{y} + b$, where

$$a = \frac{\sum (y_i - \bar{y}_o)(\tilde{y}_i - \bar{y}_p)}{\sum (\tilde{y}_i - \bar{y}_p)^2}$$

and

$$b = \bar{y}_o - a\bar{y}_p$$

In these equations, $\bar{y}_o$ and $\bar{y}_p$ are the average values of the observed and predicted pIC_{50} values of the test set molecules. For the ideal model, the slope a is equal to 1, intercept b is equal to 0, and correlation coefficient:

$$R = \frac{\sum (y_i - \bar{y}_o)(\tilde{y}_i - \bar{y}_p)}{\sqrt{\sum (y_i - \bar{y}_o) \sum (\tilde{y}_i - \bar{y}_p)^2}}$$

for the regression is equal to 1. A 3D-QSAR model may have a high predictive power, if it is close to the ideal one. Meanwhile, the regression of y against $\tilde{y}$ through the origin: $y_i^{r0} = k\tilde{y}_i$, should be characterized by k close to 1. Slope k is calculated as follow:

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2}$$

Another essential parameter r_m^2 was defined as follow:

$$r_m^2 = r^2 \left(1 - \sqrt{|r^2 - r_0^2|} \right)$$

where, the r^2 was the non-cross-validated correlation coefficient obtained from the PLS process, and the r_0^2 was calculated as follows:

$$r_0^2 = 1 - \frac{\sum (\tilde{y}_i - y_i^{r0})^2}{\sum (\tilde{y}_i - \bar{y}_p)^2}$$

where, the y_i^{r0} was obtained by this formula:

$$y_i^{r0} = k\tilde{y}_i$$

According to literature [21–24], 3D-QSAR models were considered acceptable if they satisfy all of the following conditions:

$$r_{cv}^2 > 0.5, \ r^2 > 0.6, \ [(r^2 - r_0^2)/r^2] < 0.1, \ 0.85 \leq k \leq 1.15$$
$$\text{and } r_m^2 > 0.5.$$

Molecular docking

The Surflex-Dock using an empirical scoring function and a patented search engine to dock ligands into a protein's binding site was applied to study molecular docking [17]. Crystal structure of FBPase was retrieved from RCSB Protein Data Bank (PDB entry code: 1FTA). The FBPase structure was utilized in subsequent docking experiments

without energy minimization. The ligands were docked into corresponding protein's binding site by an empirical scoring function and a patented search engine in Surflex-Dock [10]. All ligands and water molecules have been removed and the polar hydrogen atoms were added. Protomol, an idealized representation of a ligand that makes every potential interaction with the binding site was used to guide molecular docking. Establishment of protomol supplies three manners: (a) Automatic: Surflex-Dock finds the largest cavity in the receptor protein; (b) Ligand: a ligand in the same coordinate space as the receptor; (c) Residues: specified residues in the receptor [10]. In present work, the automatic docking was applied. To visualize the binding mode between the protein and ligand, the MOLCAD (Molecular Computer Aided Design) program was employed. MOLCAD calculates and displays the surfaces of channels and cavities, as well as the separating surface between protein subunits [10]. MOLCAD program provides several types to create a molecular surface. In this paper, the fast Connolly method which uses a marching cube algorithm to generate the surface was applied, meanwhile the MOLCAD Robbin and Multi-Channel surfaces program displayed with several potentials were established. Other parameters were established by default in software.

Results and discussions

CoMFA model

The statistical parameters associated in CoMFA model were listed in Table 3. The CoMFA model of a series of [5-(4-amino-1*H*-benzoimidazol-2-yl)-furan-2-yl]-phosphonic acid derivatives gave a good cross-validated correlation coefficient (r_{cv}^2) of 0.614 (>0.5) with an optimized component of 6, which suggested that the model should be a useful tool for predicting the IC_{50} values. A high non-cross-validated correlation coefficient (r^2) of 0.950 with a low standard error estimate (SEE) of 0.169, F value of 95.176 and predictive correlation coefficient (r_{pred}^2) of 0.826 were obtained. Contributions of steric and electrostatic fields were 0.686 and 0.314, respectively. The actual and predicted pIC_{50} values of the training set and test set by the model were given in Table 2, and the graph of actual activity versus predicted pIC_{50} of the training set and test set was illustrated in Fig. 3.

CoMSIA model

The CoMSIA model using steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields gave a cross-validated correlation coefficient (r_{cv}^2) of 0.598 (>0.5) with

Table 3 PLS results of CoMFA and CoMSIA models

|	CoMFA	CoMSIA
PLS statistics		
r_{cv}^{2a}	0.614	0.598
r^{2b}	0.950	0.928
ONC[c]	6	6
SEE[d]	0.169	0.203
F value	95.176	64.175
r_{pred}^{2e}	0.826	0.658
Field contribution		
Steric	0.686	0.203
Electrostatic	0.314	0.149
Hydrophobic	–	0.251
H-bond Donor	–	0.287
H-bond Acceptor	–	0.109

[a] Cross-validated correlation coefficient

[b] Non-cros-validated coefficient

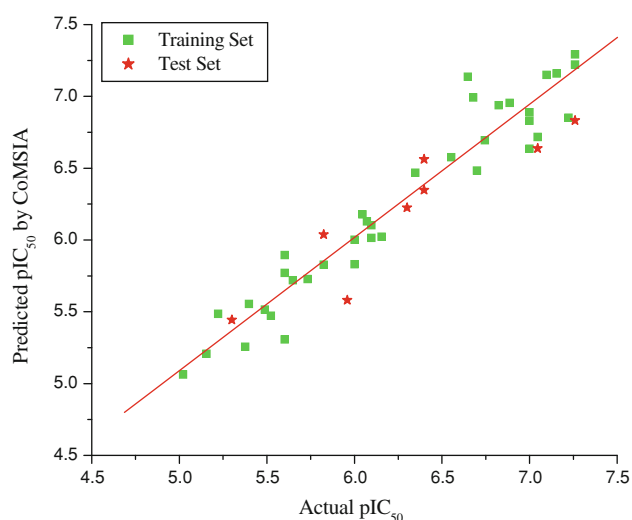[c] Optimal number of components

[d] Standard error of estimate

[e] Predictive correlation coefficient
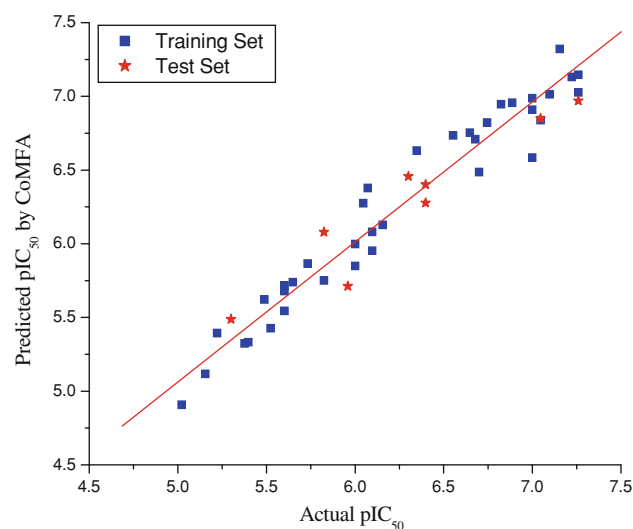
Fig. 3 Graph of actual versus predicted pIC_{50} of the training set and the test set using CoMFA

an optimized component of 6. A high non-cross-validated correlation coefficient (r^2) of 0.928 with a standard error estimate (SEE) of 0.203, F value of 64.175 and predictive correlation coefficient (r_{pred}^2) of 0.658 was obtained. Contributions of steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields were 0.203, 0.149, 0.251, 0.287 and 0.109, respectively. The actual and predicted pIC_{50} values and residual values for training set and test set compounds were given in Table 2. The relationship between actual and predicted pIC_{50} of the training set and test set compounds was illustrated in Fig. 4.

Fig. 4 Graph of actual versus predicted pIC_{50} of the training set and the test set using CoMSIA

Analysis of external validation

The results of the external validation were shown in Table 4. The established CoMFA model using eight molecules in the test set, gave a slope a value of 1.186 (close to 1), intercept b value of -1.136 (close to 0), an excellent r_m^2 value of 0.751 (>0.5) as well as high slope of regression lines through the origin (k) value of 0.994 ($0.85 \leq k \leq 1.15$), and the correlation coefficient (R) values of 0.955 (close to 1), the calculated $[(r^2-r_0^2)/r^2]$ values of -0.046 (<0.1) were obtained. The CoMSIA model also using eight compounds in the test set, gave a slope a value of 1.174 (close to 1), intercept b value of -0.977 (close to 0), an excellent r_m^2 value of 0.690 (>0.5) as well as high slope of regression lines through the origin (k) value of 0.994 ($0.85 \leq k \leq 1.15$), and the correlation coefficient (R) values of 0.915 (close to 1), the calculated $[(r^2-r_0^2)/r^2]$ values of -0.071 (<0.1) were obtained. The results of the external validation indicated that both the CoMFA and CoMSIA models established by present study possessed a

Table 4 Results of the external validation for CoMFA and CoMSIA models

Parameters	CoMFA	CoMSIA
Slope a	1.186	1.174
Intercept b	-1.136	-0.977
Correlation coefficient R	0.955	0.915
Slope k	0.994	0.994
r_0^2	0.994	0.994
r_m^2	0.751	0.690
$[(r^2-r_0^2)/r^2]$	-0.046	-0.071

high accommodating capacity, they may be reliable for being used to predict the activities of new derivatives.

CoMFA versus CoMSIA

Compared with CoMSIA, the CoMFA model displayed better r_{cv}^2, r^2 and r_{pred}^2 values, which indicated that the CoMFA model possessed higher predictive power than CoMSIA. In the external validation analysis, although both the CoMFA and CoMSIA showed excellent parameters, the CoMFA model was found to be slightly more reliable, it had better R and r_m^2 values than the CoMSIA model.

CoMFA contour maps

To visualize the information content of the derived 3D-QSAR model, CoMFA contour maps were generated to rationalize the regions in 3D space around the molecules where changes in the steric and electrostatic fields were predicted to increase or decrease the activity. The CoMFA steric and electrostatic contour maps were shown in Fig. 5.

The steric contour map of CoMFA was shown in Fig. 5a. Compound **44** was selected as a reference molecule. The steric field is represented by green and yellow contours, in which green contours (80% contribution) indicate regions where bulky group would be favorable, while the yellow contours (20% contribution) represent regions where bulky group would decrease the activity. As shown in Fig. 5a, the N-1 position (R_1) was surrounded by four yellow contours, which suggested a bulky group at this region would decrease the inhibitory activity. This may explain why compounds **5–17** bearing a bulky group (e.g. cyclohexyl-CH_2–, norbonyl, 4-tBu-benzyl, 3-thienyl-CH_2–) at R_1 showed significant decreased activities than other compounds with a minor substituent at R_1. For example, compounds **6–9** had an order for the potency of **6 > 7 > 8 > 9**, with the corresponding R_1 substituent cyclobutyl-CH_2–, cyclopentyl-CH_2–, cyclohexyl-CH_2–, cycloheptyl-CH_2–, respectively. Additionally, compounds **11–17** with benzyl, furnyl, thienyl groups at the R_1 position were the most inactive compounds. The presence of the yellow contour around the C-5 (R_5) position also suggested a bulky group at this region would be unfavorable. By checking up all the compounds modified at C-5, it was found that derivatives **19–21** have the activity order of **21** (R_5 = OH) > **20** (R_5 = OMe) > **19** (R_5 = nPr); compounds **26**, **24**, **22** also have the activity order of **26** (R_5 = F) > **24** (R_5 = Cl) > **22** (R_5 = Br). This is satisfactory according to the contour map. The green contour at the middle of the C-7 (R_7) side chain indicated that bulky groups at this position may increase potency. This may explain that compound **25** (R_7 = Br) was more potential than **23** (R_7 = Cl), while compound **23** was more active

Fig. 5 Std* coeff contour maps of CoMFA analysis with 2 Å grid spacing in combination with compound **44**. **a** Steric fields: *green and yellow* contours indicate regions where bulky groups increase or decrease the activity, and **b** Electrostatic fields: *blue and red* contours indicate regions where electron-donating or electron-withdrawing groups increase the potency

than **4** (R_7 = H). Compared compound **34** (R_7 = vinyl) with **35** (R_7 = cPr), as well as **43** (R_7 = –(CH$_2$)$_3$OH) with **44** (R_7 = –(CH$_2$)$_3$N(Me)$_2$), it can be easily found that their activity discrepancies can also be explained by this green contour.

The electrostatic field contour map of CoMFA was shown in Fig. 5b. Compound **44** was selected as a reference molecule again. The electrostatic field is indicated by blue (80% contribution) and red (20% contribution) contours, which demonstrated the regions where electron-donating group and electron-withdrawing group would be favorable respectively. In the CoMFA electrostatic field, a huge blue contour around the N-1 (R_1) side chain revealed the electron-donating substituent was essential for the inhibitory activity. Taking the compound **13** (R_1 = 4-CF$_3$-benzyl) for an example, the strong electron-withdrawing –CF$_3$ group at the terminal of N-1 side chain resulted in significant decreased activity. The red contour near C-5 (R_5) position suggested the electron-withdrawing groups on this position benefited potency, this may be the reason why compounds **26–29**, **32–35** and **36–45** possessing electron-withdrawing groups (–F, –Cl) at R_5 had higher potencies than the rest compounds with H atom or electron-donating groups such as compound **18** (R_5 = Et), **19** (R_5 = nPr), **20** (R_5 = OMe). For instance, compounds **26**, **24**, **22** had an activity order of **26** > **24** > **22**, with the corresponding R_5 substituent –F, –Cl, –Br respectively. Moreover, the activity discrepancies of compounds **30–32** can also be explained by this red contour. The small blue contour near the R_7 position suggested that an electron-donating substituent would increase the activity. Derivatives **39** (R_7 = Et) and **44** (R_7 = (Me)$_2$N(CH$_2$)$_3$–) with electron-donating groups at this site, were the most active compounds; meanwhile, compounds **23** (R_7 = Cl), **25** (R_7 = Br), **36** (R_7 = Ph), **37** (R_7 = 4-F-Ph) and **38** (R_7 = 4-Cl-Ph) possessing electron-withdrawing substituent at this position exhibited decreased activities.

CoMSIA contour maps

Figure 6a–e provided the steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor contours plots for compound **44** of the CoMSIA model. The CoMSIA steric and electrostatic contour maps were shown in Fig. 6a and b which were almost the same to the corresponding CoMFA steric and electrostatic contour maps. In hydrophobic field, white (20% contribution) and yellow (80% contribution) contours highlighted areas where hydrophilic and hydrophobic properties were favored. In hydrogen bond donor field, the cyan (80% contribution) and purple (20% contribution) contours indicated favorable and unfavorable hydrogen bond donor groups. In hydrogen bond acceptor field, the magenta (80% contribution) and red (20% contribution) contours identified favorable and unfavorable positions for hydrogen bond acceptors.

In Fig. 6c, two yellow contours around R_1 position indicated that a hydrophobic substituent would benefit the potency. All of the derivatives involved in this study possessed a hydrophobic group at this site, which revealed the extreme importance of the hydrophobic substituent. A huge white contour near the R_7 site suggested that a hydrophilic group may be favored. This may explain why derivative **44** with a relative more hydrophilic N atom at this position exhibited better potencies than compounds **36–38** as well as **40–42**. Another yellow contour around R_5 position demonstrated that a hydrophobic substituent would be favorable. Most of the compounds possessed a hydrophobic substituent at this site, except that compound **21** with a hydrophilic -OH at R_5 which displayed lower activity than compound **26**. Two white contours near the phosphonic position indicated the extreme importance of this hydrophilic phosphonic group. Without it, the compounds would decrease their inhibitory potency.

In Fig. 6d, the purple contour near the C-5 (R_5) position revealed that hydrogen bond acceptor groups may benefit

Fig. 6 Std* coeff contour maps of CoMSIA analysis with 2 Å grid spacing in combination with compound **44**. **a** Steric contour map. *Green and yellow* contours refer to sterically favored and unfavored regions. **b** Electrostatic contour map. *Blue and red* contours refer to regions where electron-donating and electron withdrawing groups are favored. **c** Hydrophobic contour map. *White and yellow* contours refer to regions where hydrophilic and hydrophobic substituent are favored. **d** Hydrogen bond donor contour map. The cyan and purple contours indicate favorable and unfavorable hydrogen bond donor groups. **e** Hydrogen bond acceptor contour map. The *magenta and red* contours demonstrated favorable and unfavorable hydrogen bond acceptor groups
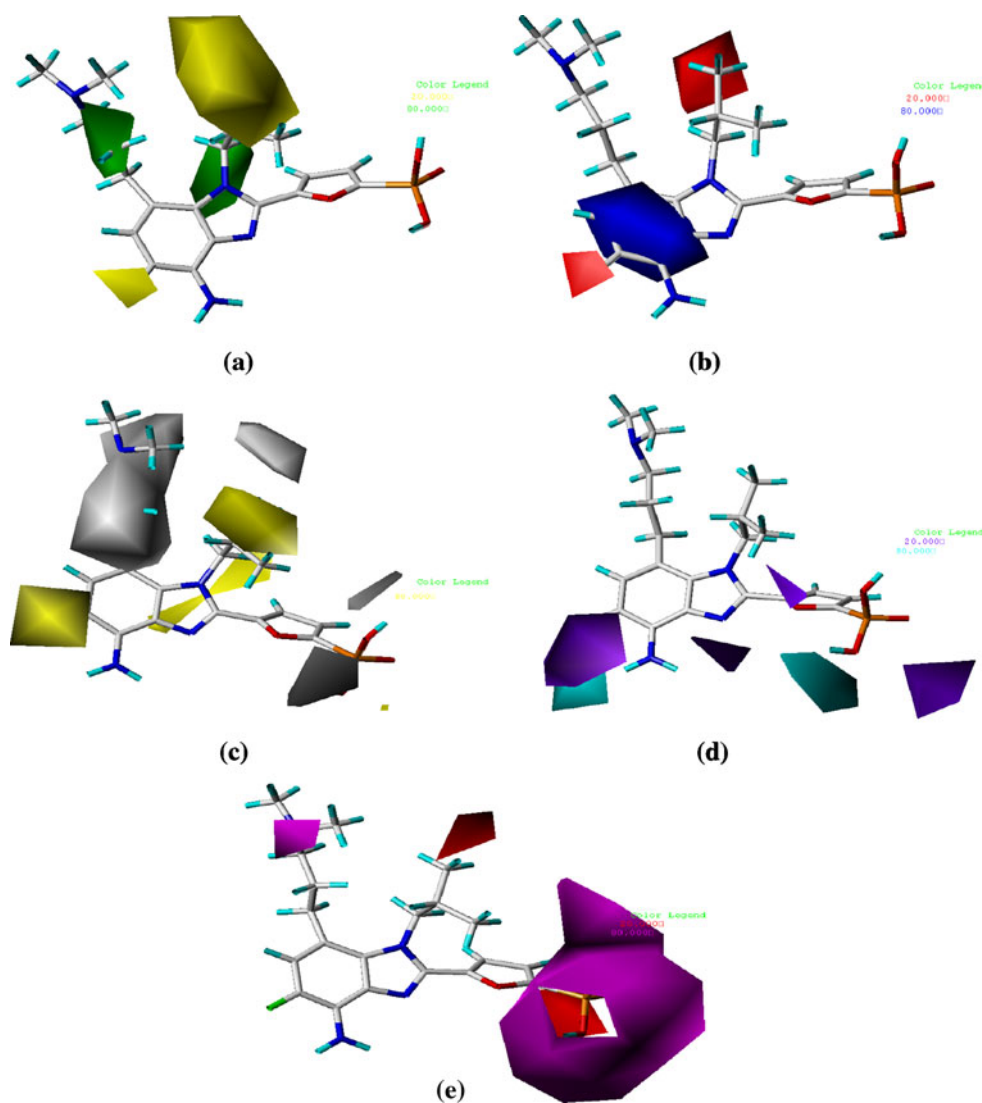
the potency. In fact the –F atom at this position acted as hydrogen bond acceptor with residue Tyr244 of the allosteric site. This may explained the reason why compounds **26–29**, **32–35** and **36–45** which possessed –F at C-5 showed relative better activities. Two big cyan and two purple contours near the two hydroxyl groups of phosphonic group revealed that they acted as hydrogen bond donors and acceptors by forming hydrogen bonds with residues of the AMP allosteric site. Lack of them, the compounds may decrease their inhibitory activity.

In Fig. 6e, the purple contour around the N atom of the R_7 side chain in compound **44** indicated that a hydrogen bond acceptor substituent at this site would increase the activity. Compared compound **44** with **42**, as well as **43** with **41**, their activity discrepancies can be explained by this red contour. The huge purple contour along with a red contour around the phosphonic group suggested that they acted as hydrogen bond donors and acceptors by forming

hydrogen bonds with residues of the AMP-binding site. Without them, the compounds may decrease their inhibitory potency.

Docking analysis

Docking was employed to explore the binding mode between these benzimidazoles and the receptor, furthermore, to examine the stability of 3D-QSAR models previous established.

Since the crystal structure of FBPase was known, we replaced AMP with the benzimidazoles derivatives to examine their bound to FBPase. The calculated docking scores of the entire database ranging from 2.91 to 6.89 were listed in Table 2. In general, compounds **1**, **2**, **4**, **6–9**, **12** and **19** with the score ranging from 3.95 to 4.95 were less active than compounds **20–29** with the score ranging from 4.65 to 6.39; the most potential derivatives **36–38**, **40**,

41 and **43–45** had a score value ranging from 5.30 to 6.89. Since the docking score represented the binding affinities and energies between these inhibitors and the pocket, it could be inferred that the enhanced affinities may resulted in improved inhibitory potencies.

The most potential compound **44** was selected for more detailed researches. Figure 7 showed the hydrogen bond binding between the selected compound **44** and the residues of allosteric site of FBPase, the key residues and hydrogen bonds were labeled. As shown in Fig. 7, the –F at C-5 position acted as the hydrogen bond acceptor and formed H-bond with the -OH group of Tyr244 residue. The N atom of the -(CH$_2$)$_3$N(Me)$_2$ substituent in compound **44** formed H-bond with -NH$_2$ group of Asn125 residue. The two hydroxyl groups of phosphonic group formed six H-bonds with four residues of the allosteric site. One of the two hydroxyl groups served as hydrogen bond donor and acceptor by forming three H-bonds with the –NH group of Gly122 and –NH as well as –OH groups of Ser123 residues. The other also acted as H-bond donor and acceptor by forming three H-bonds with the two oxygen atoms of Asp118 and the -NH$_2$ group of Arg276. The results confirmed the observation from the CoMSIA hydrogen bond donor and acceptor contour maps.

In order to visualize secondary structure elements, the MOLCAD Robbin surfaces program was applied. Figure 8 showed the secondary structure of compound **44** in complex with the AMP pocket, alpha helices were shown as helices or cylinders, while beta sheets were shown as arrows and the loop regions as tubes. The key residues and hydrogen bonds were labeled.
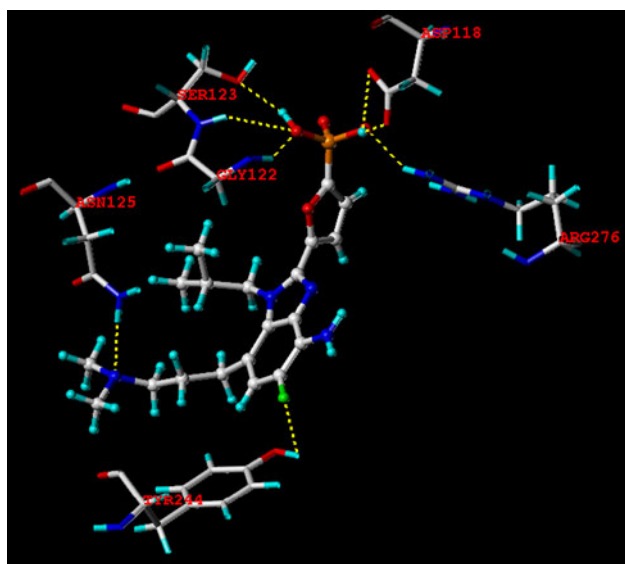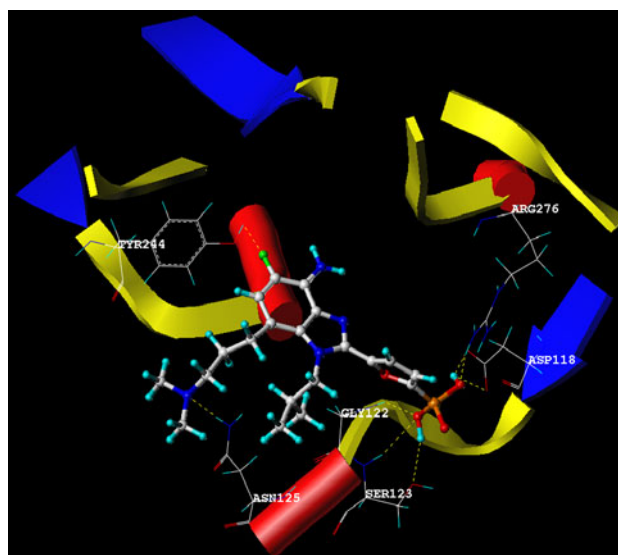
Fig. 8 MOLCAD Robbin surfaces structure of compound **44** in complex with the allosteric site of FBPase (PDB code 1FTA). Key residues and hydrogen bonds were labeled. The alpha helices were shown as *helices or cylinders*, while beta sheets were shown as *arrows* and the loop regions as *tubes*

The MOLCAD surface of allosteric site was also developed and displayed with cavity depth (CD), electrostatic potential (EP), lipophilic potential (LP) and hydrogen bond site (HB) to further explore the interaction between these inhibitors and the receptor. These potentials on a protein surface can be used to find the sites that act attractively on ligands by matching opposite colors.

Figure 9a and b depicted the MOLCAD Robbin and Multi-Channel cavity depth potential surfaces structure of the binding site within the compound **44**. The cavity depth color ramp ranges from blue (low depth values = outside of the pocket) to light red (high depth values = cavities deep inside the pocket). In both Fig. 9a and b, the –F (R$_5$) and –NH$_2$ (R$_4$) position of compound **44** were found in light red area which indicated that the two position were anchored deep inside the AMP pocket. The phosphonic group and the *t*-Bu (R$_1$) were in a yellow region which indicated that they were anchored in less deep inside of the pocket. The (Me)$_2$N(CH$_2$)$_3$– substituent (R$_7$) was located in blue area, indicating that this site was anchored outside of the pocket. It can be easily inferred from Fig. 9 that minor substituent at R$_4$ and R$_5$ positions would be favorable, for the two sites were deeply imbedded in the binding site; and bulky group at R$_7$ site may be favored, since this position was solvent exposed. The inference obtained by Fig. 9 satisfactorily matched the corresponding CoMFA and CoMSIA steric contour maps.

Figure 10 demonstrated the MOLCAD electrostatic potential surface of the binding region, the color ramp for EP ranges from red (most positive) to purple (most

Fig. 7 The binding mode between compound **44** and the allosteric site of FBPase (PDB code 1FTA). Key residues and hydrogen bonds were labeled
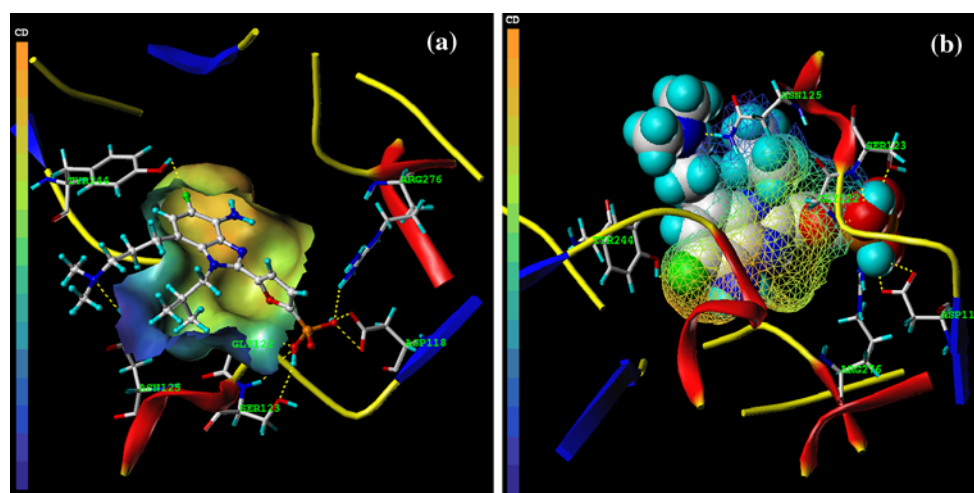
Fig. 9 The MOLCAD Robbin and multi-channel surfaces structure displayed with cavity depth potential of the allosteric site within the compound **44**. Key residues and hydrogen bonds were labeled. The cavity depth color ramp ranges from *blue* (low depth values = outside of the pocket) to *light red* (high depth values = cavities deep inside the pocket)
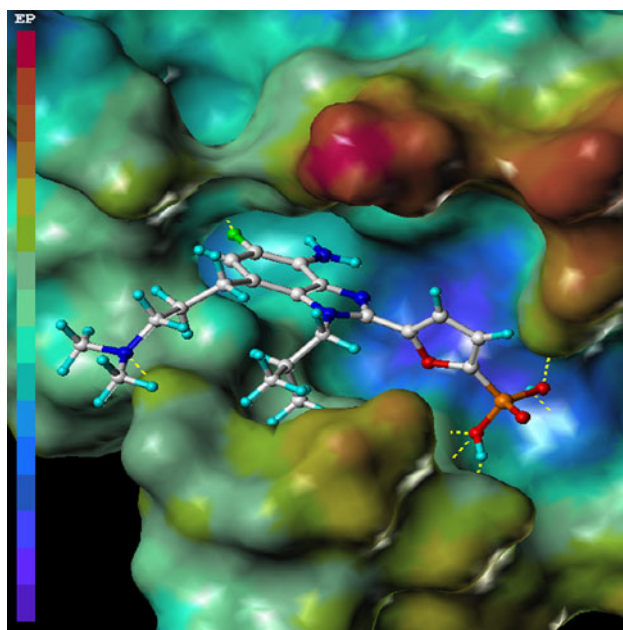
Fig. 10 The MOLCAD electrostatic potential surface of the allosteric site of FBPase within the compound **44**. The color ramp for EP ranges from *red* (most positive) to *purple* (most negative)

Fig. 11 The MOLCAD lipophilic potential surface of the allosteric site of FBPase within the compound **44**. The color ramp for LP ranges from *brown* (highest lipophilic area of the surface) to *blue* (highest hydrophilic area)

negative). In Fig. 10, the R_1 position was found in a blue area, which indicated that electron-donating properties at this site were essential for the potency; the R_5 site was in a yellow area, which suggested that electron-withdrawing properties would be favored; the R_7 position was anchored in a cyan area which suggested that an electron-donating substituent at this position would be favored. The observations obtained from this electrostatic potential surface satisfactorily matched the corresponding CoMFA and CoMSIA electrostatic contour maps.
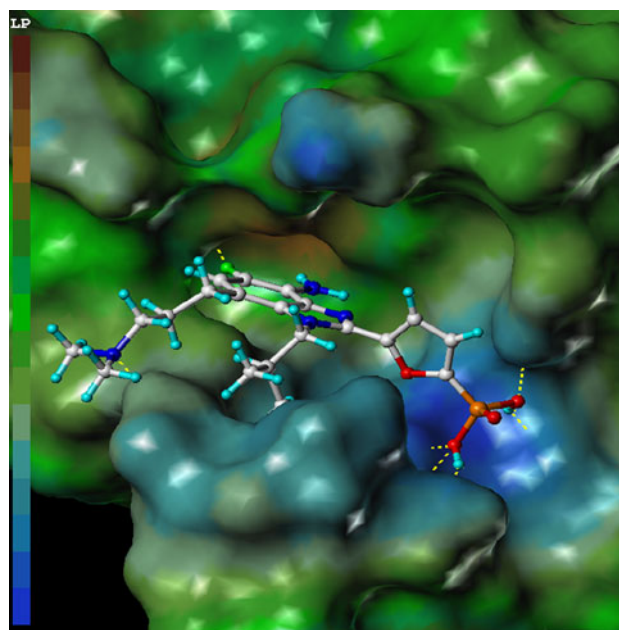
Figure 11 showed the MOLCAD lipophilic potential surface of the binding area, the color ramp for LP ranges from brown (highest lipophilic area of the surface) to blue (highest hydrophilic area). The R_5 position was oriented to a brown region, suggesting that a hydrophobic substituent may be favored; the N atom of the $(Me)_2N(CH_2)_3-$ substituent (R_7) was oriented to a white area which indicated that a hydrophilic group would be favorable; the phosphonic group was surrounded by a blue surface which

demonstrated that a hydrophilic group would benefit the potency. The observations taken from Fig. 11 satisfactorily matched those of the CoMSIA hydrophobic contour map.

Figure 12 displayed the MOLCAD hydrogen bonding sites of the binding surfaces, ligands can be docked to proteins by matching the patterns displayed on the surface, the color ramp for HB ranges from red (hydrogen donors) to blue (hydrogen acceptors). In Fig. 12, the –F (R_5) site was oriented to a blue surface, which suggested that the surface of this site were hydrogen bond acceptors, and a hydrogen bond donor substituent would be favorable; the N atom of the $(Me)_2N(CH_2)_3$– substituent (R_7) was anchored to a red area, which indicated that the surface of this region were hydrogen bond donors, and a hydrogen bond acceptor property may be favored; both of the two hydrogen atoms of hydroxyl groups in the phosphonic group were oriented to blue regions, which suggested that the surface of this site

were hydrogen bond acceptors, and a hydrogen bond donor substituent would be favorable; and the two oxygen atoms of hydroxyl groups in the phosphonic group were oriented to red surfaces, which indicated that the surface of this region were hydrogen bond donors, and a hydrogen bond acceptor property may be favored. The observations taken from this hydrogen bonding sites satisfactorily matched the corresponding CoMSIA hydrogen bond donor and acceptor contour maps.
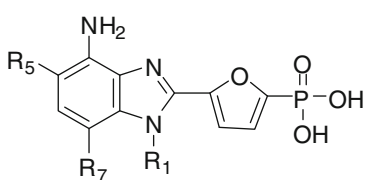
Summary of structure–activity relationship

The structure–activity relationship revealed by 3D-QSAR and molecular docking studies were illustrated in Fig. 13. In detail, the minor, electron-donating and hydrophobic groups at R_1 position are favorable; the minor, electron-withdrawing, hydrophobic and hydrogen bond acceptor substituent at R_5 position would increase the activity; the bulky, electron-donating, hydrophilic and hydrogen bond acceptor groups at R_7 position may benefit the potency. A phosphonic group was essential for binding to the allosteric site of FBPase.

Design for novel derivatives

By utilizing the structure–activity relationship revealed by this study as well as the synthetic availability of these derivatives, we have designed a series of novel derivatives based on the synthetic availability of these compounds. These molecules were aligned to the database and their activities were predicted by the CoMFA and CoMSIA models previously established. The chemical structures and predicted pIC_{50} values of these compounds were shown in Table 5, and the graph of their predicted pIC_{50} values versus the most active compounds 29, 39 and 44 was shown in Fig. 14. Most of the molecules exhibited better predicted pIC_{50} values than compounds 29, 39 and 44 in the CoMFA and CoMSIA models, especially those with a 4-methylsulfonylpiperidinyl group at R_7 position (D1–D8). Molecules D1, D2, D5, D15–D17, D32–D34 and D36 which possessed an minor, electron-withdrawing and

Fig. 12 The MOLCAD hydrogen binding sites surface of the allosteric site of FBPase within the compound 44. The color ramp for HP ranges from *red* (hydrogen bond donors) to *blue* (hydrogen bond acceptors)

Fig. 13 Structure-activity relationship revealed by 3D-QSAR and docking studies

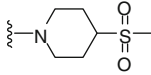**Table 5** Structures and predicted pIC$_{50}$ values of newly designed derivatives
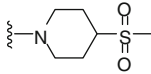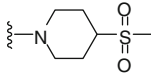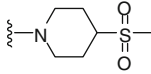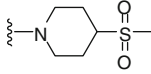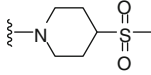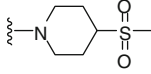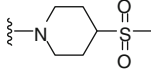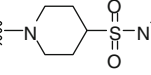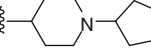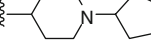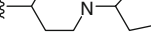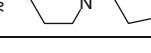
No.	Subtituent			Predicted pIC$_{50}$	
	R$_1$	R$_5$	R$_7$	CoMFA	CoMSIA
D1	-(CH$_2$)$_3$N(CH$_2$)$_2$	-CF$_3$		7.794	7.947
D2	-(CH$_2$)$_3$N(CH$_2$)$_2$	-CN		7.793	7.834
D3	-(CH$_2$)$_3$N(CH$_2$)$_2$	-OMe		7.273	7.444
D4	-(CH$_2$)$_3$N(CH$_2$)$_2$	-OEt		7.160	7.314
D5	-(CH$_2$)$_3$N(CH$_2$)$_2$	-F		7.801	7.827
D6	-(CH$_2$)$_3$N(CH$_2$)$_2$	-SO$_3$H		7.411	7.520
D7	-(CH$_2$)$_3$N(CH$_2$)$_2$	-COOH		7.674	7.611
D8	-(CH$_2$)$_3$N(CH$_2$)$_2$	-OH		7.695	7.823
D9	*t*-Bu	-F		7.208	7.191
D10	Et	-NO$_2$		7.398	7.072
D11	Et	-CF$_3$		7.205	7.117
D12	Et	-CN		7.309	7.004
D13	Et	-F		7.106	7.359

Table 5 continued
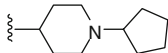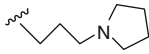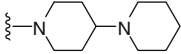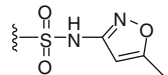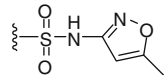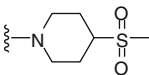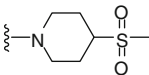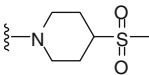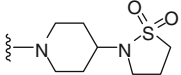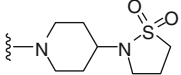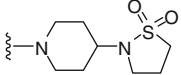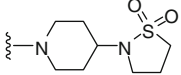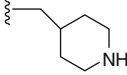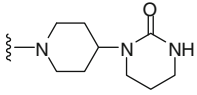
D14	Et	-OH		6.972	7.289
D15	Me	-NO$_2$	Et	7.534	6.838
D16	Me	-CF$_3$	Et	7.435	6.890
D17	Me	-CN	Et	7.572	6.861
D18	Me	-OCH$_3$	Et	7.368	6.719
D19	Me	-SO$_3$H	Et	7.236	7.260
D20		-CF$_3$		6.829	7.607
D21		-CN		6.903	7.292
D22	*i*-Pr	-CN	-(CH$_2$)$_3$N(Me)$_2$	7.273	7.116
D23	Me	-CF$_3$		7.211	6.986
D24	Me	-CN		7.324	6.924
D25		-CF$_3$		7.341	7.020
D26		-CN		7.343	6.903
D27		-F		7.324	6.997
D28	-CH$_2$C(Cl)$_3$	-CF$_3$		6.786	7.408
D29	-CH$_2$C(Cl)$_3$	-CN		6.789	7.293
D30	-CH$_2$C(Cl)$_3$	-F		6.754	7.382
D31	-CH$_2$C(Cl)$_3$	-OH		6.622	7.376
D32		-CF$_3$		7.072	7.622

Table 5 continued

|					
D33	[structure]	-CN	[structure]	7.089	7.507
D34	[structure]	-F	[structure]	7.067	7.608
D35	[structure]	-COOH	[structure]	6.933	7.394
D36	[structure]	-OH	[structure]	6.924	7.633
D37	*t*-Bu	-CF₃	[structure]	6.501	7.488
D38	*t*-Bu	-F	[structure]	6.657	7.407
D39	*t*-Bu	-CF₃	[structure]	6.735	7.315
D40	*t*-Bu	-F	[structure]	6.822	7.508

hydrogen bond acceptor substituent (–F, –CN, –CF₃, –OH) at R₅ position were the most potential designed molecules. Compared molecules **D1**, **D2** and **D5** with **D25**, **D26** and **D27**, it could be found that the electron-donating substituent at R₁ site was essential for the potency. The results validated the structure–activity relationship obtained by this study. Since the structure–activity relationship obtained from docking approach satisfactorily matched the CoMFA and CoMSIA models, for instance, the docking cavity depth (CD) matched the steric contour maps; electrostatic potential (EP) matched the electrostatic contour maps; lipophilic potential (LP) matched the hydrophobic contour map; hydrogen bond site (HB) matched the hydrogen bond donor and acceptor contour maps, it could be inferred that these favorable steric, electrostatic, hydrophobic and hydrogen bond binding interactions between the designed analogs and the pocket resulted in enhanced binding affinity as well as potent predicted activities.

As shown in Table 5 and Fig. 14, there were significant discrepancies between the CoMFA and CoMSIA predicted pIC₅₀ values of molecules **D15–D18** as well as **D20–D40**.

In general, both the PLS and external validation analyses indicated that the CoMFA model may be more reliable than the CoMSIA model. It can be inferred that the CoMFA predicted pIC₅₀ values of **D15–D18** as well as **D20–D40** were more credible. In case of **D15–D18**, the CoMSIA hydrogen bond donor and acceptor fields have no significant effect on R₇ position (R₇ = Et), resulted in better CoMFA results. In case of **D20–D40**, compared to the CoMFA model, the CoMSIA hydrophobic, hydrogen bond donor and acceptor fields possessed more effects on these complicated substituent in R₁ and R₇ sites (e.g. 4-methylsulfonylpiperidinyl, N-(5-methylisoxazol-3-yl)sulfonic amidyl, 1,1,1-trichloroethyl, 4-tetrahydropyrimidin-2(1H)-one-1-piperidinyl), yielded better CoMSIA predicted results.

Conclusion

AMP mimetics and their derivatives have long been regarded as a source of FBPase inhibitors. Here, for the
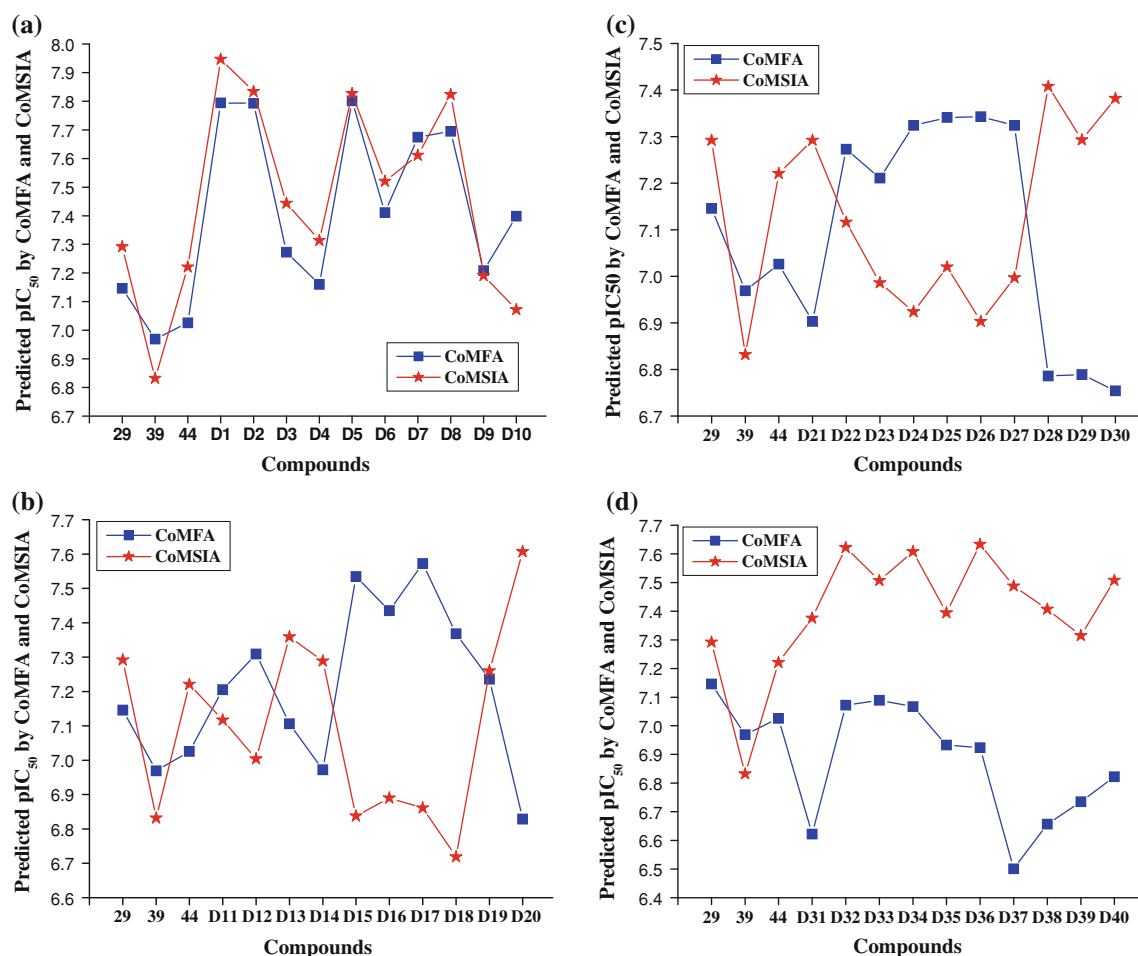
Fig. 14 Graph of the predicted pIC$_{50}$ of the designed molecules versus compound **29**, **39** and **44**

first time, we employed 3D-QSAR and docking methods to explore the structure–activity relationship of a series of [5-(4-amino-1H-benzoimidazol-2-yl)-furan-2-yl]-phosphonic acid derivatives as FBPase inhibitors. The excellent predictive ability of CoMFA and CoMSIA observed for the test set of compounds indicated that these models could be successfully used for predicting the IC$_{50}$ values. Furthermore, the CoMFA and CoMSIA contour maps along with the docking results offered enough information to understand the structure–activity relationship and identified structural features influencing the inhibitory activity. A set of novel derivatives were designed by utilizing the structure–activity relationship taken from present study, based on the excellent performance of the external validation, the predicted activities of these newly designed molecules may be reliable. The correlation of the results obtained from 3D-QSAR and docking studies successfully explored the primitive structure–activity relationship and the molecular interaction between these derivatives and the FBPase. The findings can be quite useful to aid the designing of new FBPase inhibitors with improved biological response.

References

1. Dang Q, Kasibhatla SR, Xiao W, Liu Y, DaRe J, Taplin F, Reddy KR, Scarlato GR, Gibson T, Poelje PD, Potter SC, Erion MD (2010) J Med Chem 53:441–451
2. Kitas E, Mohr P, Kuhn B, Hebeisen P, Wessel HP, Haap W, Ruf A, Benz J, Joseph C, Huber W, Sanchez RA, Paehler A, Benardeau A, Gubler M, Schott B, Tozzo E (2010) Bioorg Med Chem Lett 20:594–599
3. Heng S, Harris KM, Kantrowitz ER (2010) Eur J Med Chem 45:1478–1484
4. Gizak A, Zarzycki M, Rakus D (2009) Biochi Et Bioph Act 1793:871–877
5. Rudnitskaya A, Huynh K, Torok B, Stieglitz K (2009) J Med Chem 52:878–882
6. Dang Q, Brown BS, Liu Y, Rydzewsky RM, Robinson ED, Poelje PD, Reddy MR, Erion MD (2009) J Med Chem 52: 2280–2898
7. Dang Q, Kasibhatla SR, Jiang T, Fan K, Liu Y, Taplin F, Schulz W, Cashion DK, Reddy KR, Poelje PD, Fujitake JM, Potter SC, Erion MD (2008) J Med Chem 51:4331–4339

8. Heng S, Gryncel KR, Kantrowitz ER (2009) Bioorg Med Chem 17:3916–3922
9. Hebeisen P, Kuhn B, Kohler P, Gubler M, Huber W, Kitas E, Schott B, Benz J, Joseph C, Ruf A (2008) Bioorg Med Chem Lett 18:4708–4712
10. SYBYL 8.1, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA
11. Lan P, Huang ZJ, Sun JR, Chen WM (2010) Int J Mol Sci 11:3357–3374
12. Politi A, Durdagi S, Moutevelis-Minakakis P, Kokotos G, Papadopoulos MG, Mavromoustakos T (2009) Eur J Med Chem 44:3703–3711
13. Chung JY, Chung HW, Cho SJ, Hah JM, Cho AE (2010) J Comput Aided Mol Des 24:385–397
14. Liu HY, Liu SS, Qin LT, Mo LY (2009) J Mol Model 15:837–845
15. Pan X, Tan N, Zeng G, Huang H, Yan H (2010) Eur J Med Chem 45:667–681
16. Lu XY, Chen YD, You QD (2010) Chem Biol Drug Des 75:195–203
17. Murumkar PR, Zambre VP, Yadav MR (2010) J Comput Aided Mol Des 24:143–156
18. Tsukada T, Takahashi M, Takemoto T, Kanno O, Yamane T, Kawamura S, Nishi T (2010) Bioorg Med Chem 20:1004–1007
19. Pasha FA, Muddassar M, Srivastava AK, Cho SJ (2010) J Mol Model 16:263–277
20. Durdagi S, Mavromoustakos T, Papadopoulos MG (2008) Bioorg Med Chem Lett 18:6283–6289
21. Golbraikh A, Tropsha A (2002) J Mol Graph Model 20:269–276
22. Roy PP, Roy K (2008) QSAR Comb Sci 27:302–313
23. Lu P, Wei X, Zhang R (2010) Eur J Med Chem 45:3413–3419
24. Basu A, Jasu K, Jayaprakash V, Mishra N, Ojha P, Bhattacharya S (2009) Eur J Med Chem 44:2400–2407