Molecular simulation of a series of benzothiazole PI3Kα inhibitors: probing the relationship between structural features, anti-tumor potency and selectivity

Jinan Wang · Fangfang Wang · Zhengtao Xiao · Guowen Sheng · Yan Li · Yonghua Wang

Abstract The phosphatidylinositol 3-kinase α (PI3Kα) was genetically validated as a promising therapeutic target for developing novel anticancer drugs. In order to explore the structure-activity correlation of benzothiazole series as inhibitors of PI3Kα, comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA) were performed on 61 promising molecules to build 3D-QSAR models based on both the ligand- and receptor-based methods. The best CoMFA and CoMSIA models had a cross-validated coefficient r_{cv}^2 of 0.618 and 0.621, predicted correlation coefficient r_{pred}^2 of 0.812 and 0.83, respectively, proving their high correlative and predictive abilities on both the training and test sets. In addition, docking analysis and molecular dynamics simulation (MD) were also applied to elucidate the probable binding modes of these inhibitors at the ATP binding pocket. Based on the contour maps and MD results, some key structural factors responsible for the activity of this series of compounds were revealed as follows: (1) Ring-A has a strong preference for bulky hydrophobic or aromatic groups; (2) Electron-withdrawing groups at the para position of ring-B and hydrophilic substituents in ring-B region may benefit the potency; (3) A polar substituent like $-NHSO_2-$ between ring-A and ring-B can enhance the activity of the drug by providing hydrogen bonding interaction with the protein target. The satisfactory results obtained from this work strongly suggest that the developed 3D-QSAR models and the obtained PI3Kα inhibitor binding structures are reasonable for the prediction of the activity of new inhibitors and be helpful in future PI3Kα inhibitor design.

J. Wang · F. Wang · Z. Xiao · Y. Wang
Center of Bioinformatics, Northwest A & F University,
Yangling, Shaanxi 712100, China

G. Sheng · Y. Wang (✉)
College of Life Sciences, Northwest A & F University,
Yangling, Shaanxi 712100, China
e-mail: yh_wang@nwsuaf.edu.cn

Y. Li
Department of Materials Science and Chemical Engineering,
Dalian University of Technology,
Dalian 116023 Liaoning, China

Introduction

Phosphatidylinositol 3-kinases (PI3Ks) constitute a class of enzymes that phosphorylate the 3-hydroxyl position of phosphatidylinositol 4,5-bisphosphate, and the resulting phosphatidylinositol 3,4,5-trisphosphate (PIP3) as second messengers, can regulate a remarkably diverse array of physiological processes, including glucose homeostasis, cell growth, differentiation, and motility [1, 2]. In opposition to PI3K, the phosphatase and tensin homologue protein (PTEN) dephosphorylates the PIP3, leading to subsequent inactivation of the downstream signaling molecules and a down-regulation of the pathway. Loss of PTEN protein or function has been found in a large number of human cancers [3]. Therefore, one approach for an antitumor drug is to develop small-molecule inhibitor of PI3K to lower the PIP3 level.

The PI3Ks known to date are divided into three classes of I—III based on their sequences and substrate specific-

ities. Class I PI3Ks are further subclassified into class I_A enzymes (PI3Kα, β and δ), which have a p85 regulatory subunit and three different catalytic subunits (p110α, β, and δ), and class I_B enzyme (PI3Kγ) [4]. PI3Kα, which is activated by the receptor tyrosine kinases, undergoes a gene amplification and might also be involved in the control of endothelial cell migration of tumor cells [5–7]. Thus as a potential target for cancer therapy, PI3Kα has sparked great interests in the field of drug discovery.

To discover molecules with anticancer potential, several non-selective PI3K inhibitors have been developed, including wortmannin and LY294002 [8–10]. However, the high toxicity associated with these inhibitors have frustrated them greatly for further clinical applications. Therefore, a number of ATP-competitive PI3Kα inhibitors with improved selectivity have been identified. A natural product resveratrol found in fruits showed moderate inhibition of PI3Kα/β (IC_{50}=25 μM) [11]. Its inhibition of COX proteins initially established its chemopreventative/ chemotherapeutic potential and this potential may now also include its capability to decrease the PI3K-controlled glucose metabolism and the activatory phosphorylation of the anti-apoptotic kinase PKB (protein kinase B)/Akt. Another natural product liphagal was recently found to be a PI3Kα inhibitor (IC_{50} ~100 nM) with a 10-fold selectivity over PI3Kγ [12]. In addition, liphagal derivatives, including desformylliphagal, desmethyliphagal, and a liphagal spiro analog all have been reportedly exhibiting higher PI3Kα inhibitory values (IC_{50}=250~600 nM) [13]. A thieno[3,2-d]pyrimidine derivative was discovered as a highly selective PI3Kα inhibitor (IC_{50}=2.0 nM) with great potency of inhibiting the tumor cell proliferation in vitro [14]. However, it is not effective in vivo because of the poor pharmacokinetic profile of a short half-life<10 min. Subsequently, imidazopyridine PIK-75 has been identified as a novel potent PI3Kα inhibitor with very strong potency (IC_{50}=0.30 nM) [15]. It was not only highly selective over other PI3K isoforms (130~2800-fold increase), but also exhibited anti-proliferative activity at submicromolar concentrations in A373, HeLa, A549, MCF7 and MCF7 ADR-res cell lines as well as the potency effectively against HeLa human cervical tumor xenografts.

Experimental assessment of the activity of PI3Kα inhibitors remains a labor-intensive and time-consuming operation. Therefore, more efficient and economical alternative methods such as the in silico molecular modeling approaches should be incorporated for the purpose of predicting and prioritizing chemicals for subsequent in vitro and in vivo screening. The homology models of PI3K isoforms, i.e., p110α/β/γ and δ, showed that the overall folding pattern (β-sheets, helices, and main loops) is well conserved among the different isoforms [16]. In each isoform, the active site consists of a deep cavity bordered by conserved lipophilic residues, notably the MET804, TRP812, ILE831, ILE879, TYR867, VAL882, and MET953 (using the p110γ numbering). In addition, molecular docking and molecular dynamics (MD) simulation have been used to investigate the isoform-specific selectivity between PI3Kα and PI3Kγ [17]. The results indicate that the residues TRP780 and ASN782 in PI3Kα and the corresponding residues TRP812 and GLU814 in PI3Kγ in the solvent-accessible region confer the PI3Kα and PI3Kγ isoform specificity. A recent pharmacophore study of PI3Kα inhibitors also showed a five-point pharmacophore composed of three hydrogen bond (H-bond) acceptors, one hydrophobic group, and one aromatic ring as pharmacophore features [18].

In this work a total of 61 benzothiazole derivatives as potent PI3Kα inhibitors [19] were collected to build 3D-QSAR models using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) methods [20, 21]. Furthermore, molecular docking and MD simulation were also performed to further understand the structure properties and the probable binding modes of these inhibitors at the ATP binding pocket. All these methods applied on the inhibitors could not only help in understanding the ligand-receptor interactions but also provide useful and rational suggestions for further design of new drug candidates of PI3Kα for cancer therapy.
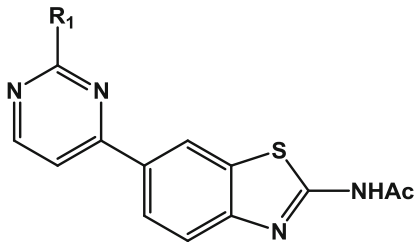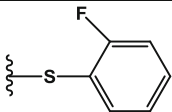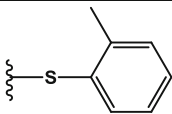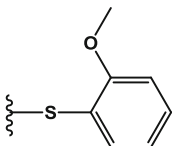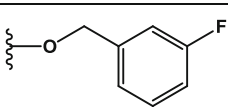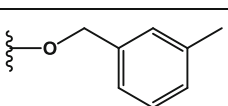
Methods and materials

Data sets and biological activity

The dataset used in the present work includes a series of 61 benzothiazole derivatives as novel PI3Kα inhibitors [19], after discarding those compounds with unspecified inhibitory activity. Their binding affinity of K_i values (in nM) to PI3Kα, was converted into pK_i (−log K_i) values and used as a dependent variable in further 3D-QSAR analyses. The whole data set was divided into two subsets: a training set of 49 compounds for generating the QSAR models and a test set of 12 compounds for validating the quality of the models. Selection of the training set and the test set was accomplished by maintaining the similar distribution of structural diversity and the range of biological activities. Table 1 depicts some representative structures and binding affinities of the molecules, and all structures, pK_i values and training/test set distribution of the whole dataset are provided in the supporting information of Tables S1 and S2.

All molecular modeling calculations were performed with Sybyl 6.9 package (Tripos Associates, St. Louis, MO), where for each molecule the Gasteiger-Hückel charges were added and energy minimization was performed using

Table 1 Representative skeletons and molecular structures of benzothiazole derivatives and their binding affinity pK_i values

Skeleton type I

Compound	R$_1$	K$_i$(nM)	pK$_i$(nM)
1	-SMe	53	7.2757
2	-SPh	40	7.3979
3	-SBn	47	7.3279
4	-OPh	102	6.9914
5	-OBn	105	6.9788
6[b]	-NHPh	6100	5.2147
7	-NHBn	85	7.0706
8	-CH$_2$Ph	605	6.2182
9	-CH$_2$CH$_2$Ph	1010	5.9957
10[a]		46	7.3372
13		31	7.5086
16[a]		144	6.8416
19[a]		61	7.2147
20		47	7.3279

Tripos force field [22] with the Powell conjugate gradient minimization algorithm. The minimization would not stop until the energy gradient convergence criterion of 0.05 kcal mol^{-1} Å^{-1} was reached.

Molecular docking

An accurate 3D structure of the receptor is important for molecular docking and rational development of novel potential inhibitors. However, the X-ray structure of the PI3Kα complexed with benzothiazole based inhibitors has not yet been reported. We thus retrieved the X-ray crystal structures of apo PI3Kα (PDB code: 2RD0) [23] and the PI3Kα H1047R mutant/wortmannin complex (PDB code: 3HHM) [24] from the RCSB Protein Data Bank, thereafter the structural alignment we performed between these two proteins using the DaliLite [25] program showed that these proteins are highly consistent with each other. Therefore,
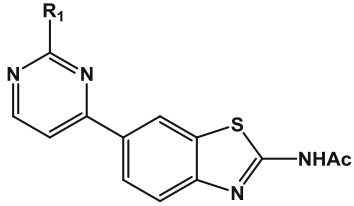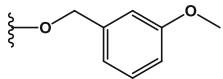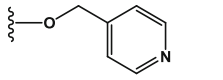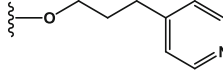
Table 1 (continued)

Compound	R$_1$	K$_i$(nM)	pK$_i$(nM)
	Skeleton type I		
21		18	7.7447
25[a]		129	6.8928
26		114	6.9431
28		32	7.4949
29		148	6.8297
30		1550	5.8097
31		333	6.4776
32		24	7.6198
33		117	6.9318
36		41	7.3872
37		65	7.1871
41[a]		38	7.4202
42		3130	5.5045

Table 1 (continued)
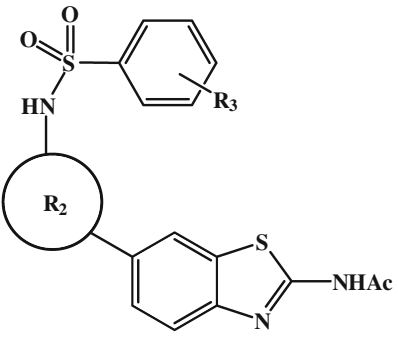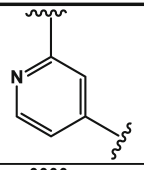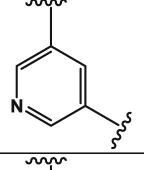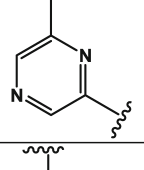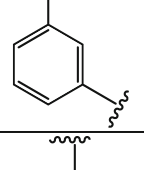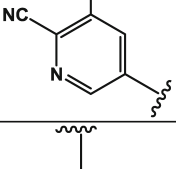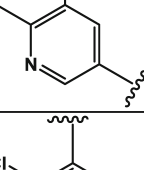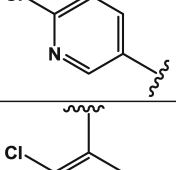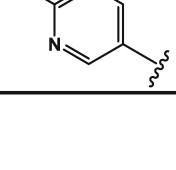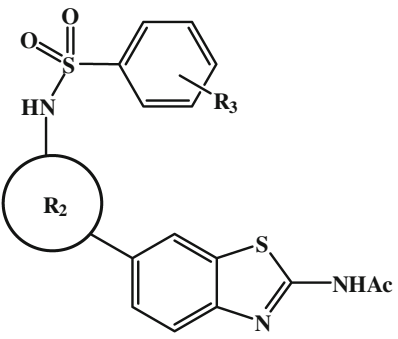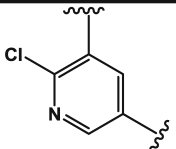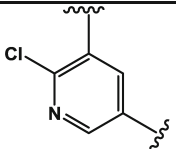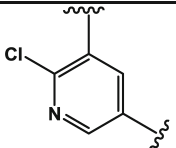
Skeleton type II

Compound	R₂	R₃	K_i(nM)	pK_i(nM)
43		4-OMe	13	7.8861
44		4-OMe	1.6	8.7959
45[a]		4-OMe	9	8.0458
46		4-OMe	48	7.3188
47[a]		4-OMe	1.4	8.8539
48		4-OMe	1.2	8.9208
49		4-OMe	0.7	9.1549
51		2-CF₃	0.8	9.0969

Table 1 (continued)

Compound	R₂	R₃	K_i(nM)	pK_i(nM)
54		2-Cl	0.9	9.0458
57		3-Bu	0.9	9.0458
59		2-F	0.9	9.0458

[a] Molecules belong to the test set.

[b] Molecule is the outlier

the 3HHM structure and the 2RD0 protein were superposed, the ligand wortmannin of 3HHM were then merged into the corresponding site of the catalytic subunit P110α of 2RD0 and used as the ligand to generate the grid file for the docking study. The side chains with missing coordinates in the crystal structure were reconstructed using the fragment library of the Biopolymer module [26]. The modified P110α was subsequently prepared using the protein preparation wizard in the Sybyl software suite.

To explore the interaction and to illustrate the accurate binding model for the active site of PI3Kα with its ligands, molecular docking was performed by using the Surflex module of Sybyl package. All parameters were set with default values in the whole process. Before docking, all water molecules were removed from the X-ray structure. Subsequently, the relative ligands were extracted with polar hydrogen atoms added. The protomol was produced using ligand-based method: ligand location in the same coordinate space in the receptor.

During docking process, two parameters, i.e., protomol_bloat and protomol_threshold, were applied to determine how far the site should extend from a potential ligand and how deep the atomic probes used to define the protomol could penetrate into the protein, respectively. Finally each conformer of all 61 inhibitors was docked into the binding site ten times with the D_score [27], G_score [28], Chemscore [29] and PMF_score [30] values further used to evaluate the docking analysis. The top ranked conformations for each molecule were extracted, then aligned together and subsequently utilized in CoMFA and CoMSIA modeling.

Conformational sampling and alignment

Molecular alignment of compounds is a crucial step in the development of CoMFA and CoMSIA models [31]. In this study, two alignment rules were employed to derive the best possible 3D-QSAR statistical model, with compound **54** with the most potent activity in the dataset used as template. One

is ligand-based alignment, where molecule **54** was used to fit all remaining compounds by using the database align function in Sybyl. The common substructure of the compounds is shown in bold in Fig. 1a and the aligned compounds are depicted in Fig. 1b. The other approach is receptor-based alignment, whose results are shown in Fig. 1c.

3D-QSAR analysis

CoMFA and CoMSIA studies were performed based on above molecular alignment rules to build predictive 3D-QSAR models. To derive the three-dimensional descriptor fields, a 3D cubic lattice with grid spacing of 2 Å in x, y and z directions was generated automatically to encompass the aligned molecules. In CoMFA, a sp^3 carbon atom was used as a probe to generate the steric (Lennard-Jones potential) field energies and a charge of +1 to generate the electrostatic (Coulombic potential) field energies [32]. The CoMFA fields generated automatically were scaled by the CoMFA-STD [33] method with default energy cut-off values of 30 kcal mol^{-1}.

CoMSIA similarity index descriptors were derived using the same lattice boxes as those used in CoMFA calculations. In addition to steric and electrostatic fields, hydrophobic, H-bond donor and acceptor descriptors were

calculated using the standard settings that probe with charge +1, radius 1 Å and hydrophobicity +1, H-bond donating +1, H-bond accepting +1, attenuation factor α of 0.3 and grid spacing 2 Å. CoMSIA similarity indices (A_F) for a molecule j with atoms i at a grid point q are calculated by Eq. 1 as follows:

$$A_{F,k}^q(j) = -\sum \omega_{probe,k}\omega_{ik}e^{-\alpha r_{iq}^2} \tag{1}$$

In this equation, k represents the steric, electrostatic, hydrophobic, or H-bond donor or acceptor descriptors. $\omega_{probe,k}$ is the probe atom with radius 1.0 Å, charge +1, hydrophobicity +1, H-bond donating +1, H-bond accepting +1; ω_{ik} is the actual value of physicochemical property k of atom i; α is the attenuation factor with the default value of 0.3 used. A Gaussian type distance dependence was used between the grid point q and each atom i of the molecule. This can avoid singularities at the atomic positions and the dramatic changes of potential energy due to grids in the proximity of the surface [34].

The CoMFA/CoMSIA fields combined with observed biological activities (pK_i) were included in a molecular spreadsheet and partial least square (PLS) methods [35] were applied to generate 3D-QSAR models. The cross-validation analysis was performed using leave-one-out (LOO) method. The cross-validated r_{cv}^2 which was used

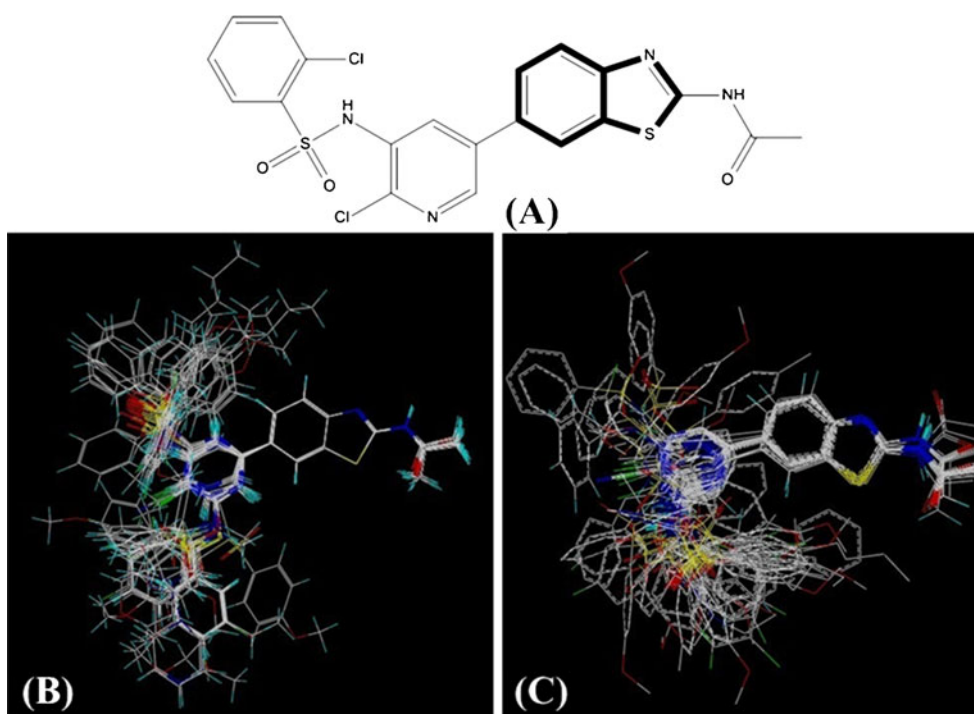**Fig. 1** Alignment of the compounds used in the training and test sets. (**a**) Compound **54** used as a template for ligand-based alignment. The common substructure is shown in bold. Ligand- and receptor-based alignments of all the compounds are shown in panels (**b**) and (**c**), respectively. Molecules are colored in white for common C, blue for N, red for O, yellow for S, cyan for H atoms, respectively

to generate the optimum number of components and lowest standard error of prediction were taken. Then the optimum number of components obtained from the cross-validation analysis was used to calculate the conventional r_{ncv}^2. The predictive ability of the 3D-QSAR model was determined using the compounds which were not included in the training set. The predictive correlation coefficient r_{pred}^2, based on the test set molecules, was calculated using the Eq. 2:

$$r_{pred}^2 = (SD - PRESS)/SD \qquad (2)$$

where SD is the sum of the squared deviation between the actual pIC_{50} values of the compounds in the test set and the mean pIC_{50} value of the training set ones. PRESS is the sum of squared deviation between predicted and actual activities of the test set compounds.

Molecular dynamics simulations

The docked complex with compound **54** served as a starting structure for MD simulations using Amber10 [36]. Chloridion ions were added to the system by random replacement of water molecules until neutrality was achieved. We select the standard AMBER99SB force field for proteins [37], meanwhile ligand charges and parameters were determined with the antechamber module of Amber 10 based on the AM1-BCC charge scheme [38] and the general atom force field (GAFF) [39]. The complex was solvated in a rectangular box of TIP3P water [40], keeping a minimum distance of 12 Å between the solute and each face of the box ($99.07 \times 80.13 \times 71.66$ Å^3). The total number of the atoms in the simulation system was 173,845 including the complex and waters. The cutoff distance was kept to 8 Å to compute the nonbonded interactions. All simulations were performed under periodic boundary conditions. To remove possible bad contacts, 1500 steps of energy minimization were performed (with 500 steps using the steepest descent and the other ones with the conjugate gradient algorithm). Constant volume dynamics with a cutoff of 8 Å was chosen. The SHAKE algorithm [41] was employed for all covalent bonds containing a hydrogen atom and the time increment was set to 2 fs.

Firstly, the minimized systems were gradually heated to 300 K at a constant force of 2.0 kcal mol^{-1} Å^{-2}. The second step consisted of a 50 ps pressure-constant period to raise the density while still keeping the complex atoms constrained. The third step was a 500 ps Langevin dynamics calculation with a collision frequency of 1 ps^{-1}, which was performed with a 2 fs time step in the NPT ensemble at a constant temperature of 300 K. Finally, the production phase was run for 4 ns with a 2 fs time step. The calculation of electrostatic forces utilized the particle-mesh-Ewald method [42] with default values.

Results and discussion

CoMFA and CoMSIA statistical results

To obtain an effective 3D-QSAR model, a number of statistical parameters, i.e., the cross-validated correlation coefficient (r_{cv}^2), non-cross-validated correlation coefficient (r_{ncv}^2), standard error estimate (SEE) and F-statistic values (F) should be analyzed. In view of the central role that the structural alignment of compounds plays in the development of successful 3D-QSAR models [31], all compounds of the dataset were aligned according to two rules (both ligand- and docking-based) to derive the CoMFA and CoMSIA models. The optimal results obtained from both alignment rules using the same training set are summarized in Table 2.

As seen from this table, all optimal CoMFA models are constructed from steric and electrostatic descriptor fields, and the CoMSIA ones are built by varying the electrostatic, hydrophobic, and H-bond acceptor descriptor fields. In addition, obviously both the ligand-based CoMFA and CoMSIA models exhibited better statistical results than corresponding receptor-based CoMFA and CoMSIA ones, where the ligand-based modeling yielded r_{cv}^2 =0.618, r_{ncv}^2= 0.897, r_{pred}^2=0.812 for CoMFA, r_{cv}^2 =0.621, r_{ncv}^2 =0.934, r_{pred}^2=0.833 values for CoMSIA models respectively, while the receptor-based modeling gave r_{cv}^2 =0.426, r_{ncv}^2=0.914, r_{pred}^2=0.475 for CoMFA, r_{cv}^2 =0.54 and r_{ncv}^2=0.890, r_{pred}^2=0.859 for CoMSIA models, respectively.

Hence, our attention was mainly focused on the ligand-based alignment model due to its more satisfactory statistical data. The CoMFA model (see Table 2) has an F value of 130.743 and an SEE value of 0.324. The CoMSIA model has an F value of 155.171 and an SEE value of 0.263. The high r_{cv}^2, r_{ncv}^2, and F values along with the low SEE values suggest that the models are reasonable and should have good internal predictive ability. The results also reveal that the electrostatic and hydrophobic and H-bond acceptor features play important roles in determining the biological activity of these inhibitors. For the CoMFA model the contributions from the steric and electrostatic fields were found to be 62% and 38% respectively, while for the CoMSIA one the electrostatic, hydrophobic and H-bond acceptor fields contribute about 28%, 39% and 33% contributions, respectively.

The external predictive abilities of these models were determined by the test set composed of 12 compounds independent from the training set ones. The predicted r_{pred}^2 values from the CoMFA and CoMSIA models were found to be 0.812 and 0.833, respectively (Table 2). The plots of actual versus predicted activities for training and test set molecules of the two models are depicted in Fig. 2, showing that there is no systematic error in the methods, i.e., the predicted activities are almost as accurate as the experimental

Table 2 Summary of the best CoMFA and CoMSIA results

Parameters	Ligand-based model		Receptor-based model	
	CoMFA	CoMSIA	CoMFA	CoMSIA
r_{cv}^2	0.618	0.621	0.426	0.54
r_{ncv}^2	0.897	0.934	0.914	0.890
SEE	0.324	0.263	0.295	0.334
F	130.743	155.171	159.631	121.288
r_{pred}^2	0.812	0.833	0.475	0.859
SEP	0.624	0.628	0.763	0.683
Nc	3	4	3	3
Field contribution				
S	0.620	-	0.463	-
E	0.380	0.277	0.537	0.347
H	-	0.392	-	0.328
D	-	-	-	-
A	-	0.331	-	0.326

r_{cv}^2 =Cross-validated correlation coefficient using leave-one-out method

r_{ncv}^2 =Non-cross-validated correlation coefficient; SEE=Standard error of estimate

F=Ratio of r_{ncv}^2 explained to unexplained=$r_{ncv}^2/(1-r_{ncv}^2)$

r_{pred}^2 =Predicted correlation coefficient for the test set of compounds

SEP=Standard error of prediction; Nc=Optimal number of principal components

S=Steric; E=Electrostatic; H=Hydrophobic; D=H-bond donor; A=H-bond acceptor

data. In addition, as seen from this figure the data points are rather uniformly distributed around the regression line, indicating the reasonability of the obtained models.

3D-QSAR contour maps

One of the attractive features of 3D-QSAR modeling is the visualization of the information content of the derived models by the contour maps generated to rationalize the regions in 3D space around the molecules where changes in the fields were predicted to increase or decrease the activity [43]. After consideration of both the internal and external

predictive powers, the best ligand-based CoMFA and CoMSIA models are selected for each conformation to construct the stdev*coeff contour maps to view the field effects on the target features.

CoMFA contour maps

Figure 3 shows the contour maps derived from the CoMFA PLS model. The most potent analogue, compound **54**, was embedded in the maps to demonstrate its affinity for the steric and electrostatic regions of inhibitors. In the steric field (Fig. 3b), the green (sterically favorable) and yellow

Fig. 2 Graphs of the predicted pK_i versus the actual pK_i values the optimal models. (**a**) Result from CoMFA. (**b**) Result from CoMSIA. The solid lines are the regression lines for the fitted and predicted bioactivities of the dataset
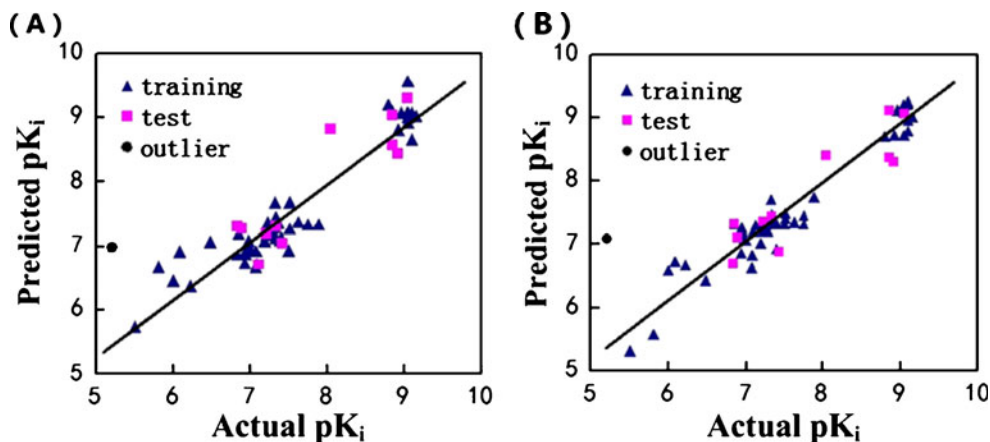
Fig. 3 CoMFA StDev*Coeff contour plots. (**a**) Compound **54** is shown in sticks as a reference. (**b**) Steric fields: green and yellow contours indicate regions where bulky groups increase or decrease the activity. (**c**) Electrostatic fields: Red contours indicate regions where negative charges increase activity; blue contours indicate regions where positive charges increase activity
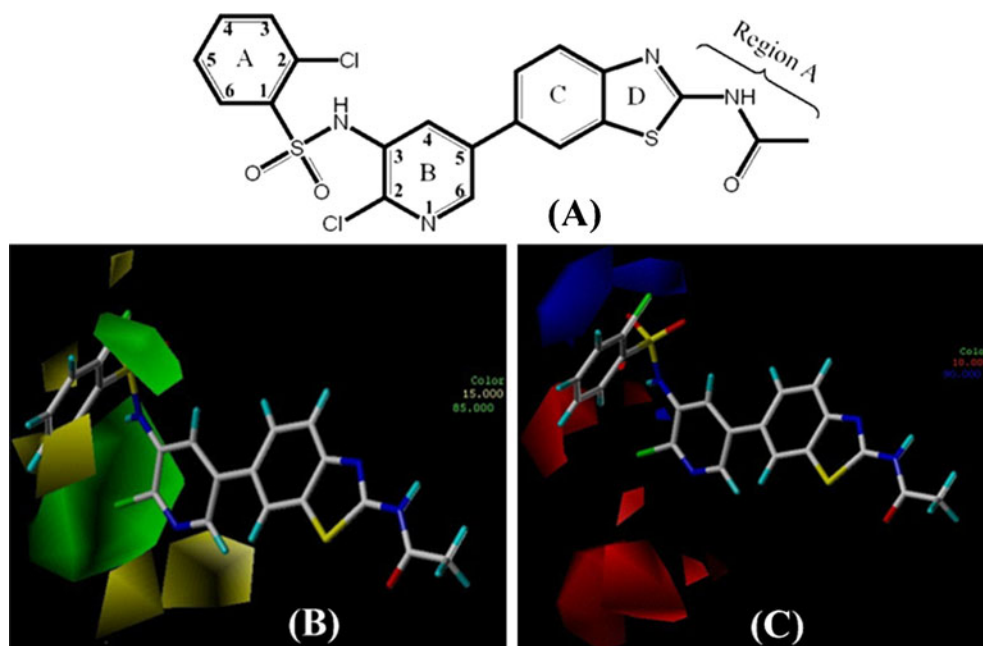
(sterically unfavorable) contours represent 85% and 15% level contributions, respectively. A small sterically favorable (green) contour is observed between ring-A and ring-C. In addition, several small sterically unfavorable (yellow) contours appear around ring-A and ring-B. This indicates that molecules carrying bulkier substituents at ring-A are more active than those compounds having small substituent at ring-A or reaching yellow contours. Hence, an addition of bulky groups at ring-A near the green contour might aid in an improved potency of the inhibitors. For example, in the high active molecules (**44**, **48**, **49**, **51**, **54∼61**), all their substituted ring-As fall into this green contour showing that this may be a suitable site for bulky group. Another green contour appears adjacent to ring-B (position-2), therefore, bulky substituents at this site may improve the activity. For example, substituents of 2-Me (**48**), 2-CN (**47**), 2-H (**44**) all located at the large green regions, cause compound **48** to exhibit stronger inhibitory activity than compounds **47** and **44** due to its bulkier substituent.

In the CoMFA electrostatic field (Fig. 3c), the blue (electropositive charge favorable) and red (electronegative charge favorable) contours represent 90% and 10% level contributions, respectively. There is a blue contour presented in the side of the -NHSO$_2$- (between ring-A and ring-B), suggesting that positively charged groups may significantly enhance the affinity between PI3Kα and its inhibitors (Fig. 3c). This can be well illustrated by the higher potency of compound **33** having -NHSO$_2$- at this region than compound **42** with -NHCO- at the same position. A red contour close to the blue one indicates that electronegative groups in this region (ring-B) are favored for higher activity, which can be exemplified by com-

pounds **44**, **47**, **48** and **49∼61** (having substituted pyridine) that are more active than **46** (having substituted benzene). Another red contour presented in the side of ring-B and near ring-A indicates that electronegative groups at these regions will enhance the activity. Therefore, the fact that compounds **51∼61** are more active than compounds **43**, **44**, **45**, **47**, **48∼50** can be explained as the former molecules all have an electronegative substituted benzene which reaches the red contour area while the latter have not.

CoMSIA contour maps

Compared to standard CoMFA, the major advantage of CoMSIA is a better ability to visualize and interpret the obtained correlation in terms of the field contributions. The CoMSIA electrostatic, hydrophobic and H-bond acceptor field contours with compound **54** are presented in Fig. 4. In the hydrophobic contour map (Fig. 4b), the yellow (hydrophobic favorable) and white (hydrophobic unfavorable) contours represent 85% and 15% level contributions, respectively. In the H-bond acceptor field (Fig. 4c), the purple (H-bond acceptor favorable) and red (H-bond acceptor unfavorable) contours represent 90% and 10% level contributions, respectively.

In the case of the CoMSIA contour, the electrostatic field contours are closer to the ligand (shown in Fig. 4a). Besides, another red contour which does not appear on the CoMFA contour map covers -NHCOCH$_3$ (region A), indicates that the electronegative groups at this region would increase the inhibitory activity. The fact is that all derivatives involved in the present study possess -NHCOCH$_3$ substituents at this site.

Fig. 4 CoMSIA StDev*Coeff contour plots. (**a**) Electrostatic contour map. Blue and red contours refer to regions where electron-donating and electron-withdrawing groups are favored. (**b**) Hydrophobic contour map. White and yellow contours refer to regions where hydrophilic and hydrophobic substituents are favored. (**c**) H-bond acceptor contour map. The magenta and red contours demonstrated favorable and unfavorable H-bond acceptor groups. Compound **54** is shown in sticks as a reference
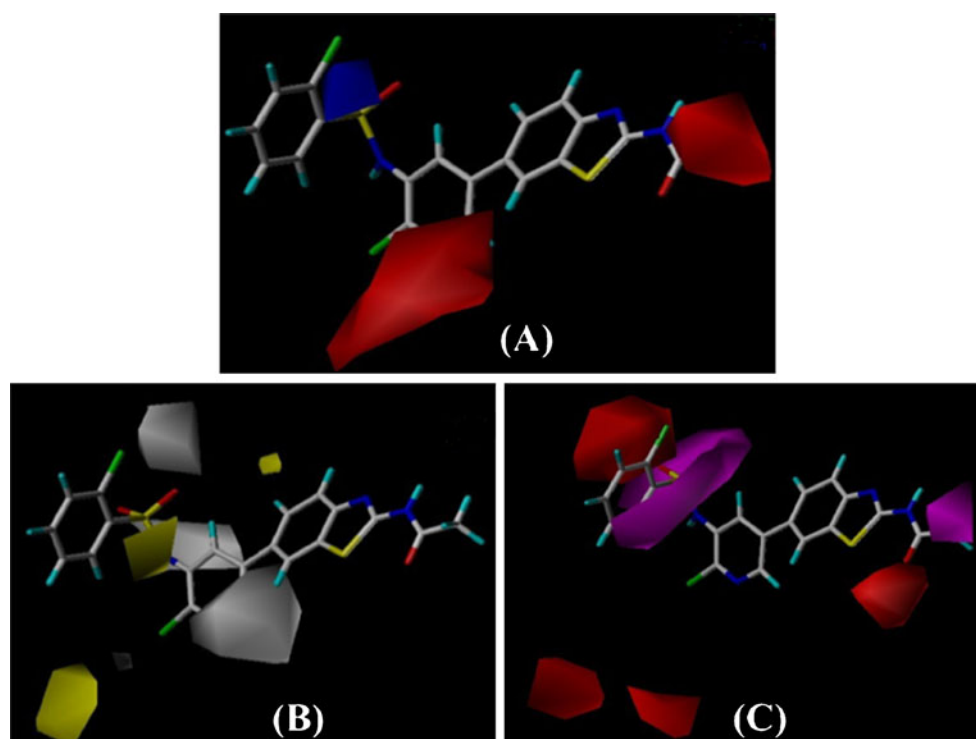
In the hydrophobic contour map (Fig. 4b), two yellow contours near the chlorobenzene substituent (ring-A) indicate that a hydrophobic group at this site would be favorable. It is consistent with the fact that compound **28** bearing a benzene substituent had higher potency than compounds **29**, **30** with a morpholine substituent and compounds **25**, **26**, **27** possessing a pyridine substituent at ring-A. Two white contours around -NHSO$_2$- between ring-A and ring-B indicate that a hydrophilic substituent at this site would be favorable. A large white contour observed appearing adjacent to pyridine ring (ring-B), especially adjoined to the N atom, indicates that a hydrophilic substituent at this site would increase the activity. For example, compounds **43**, **44**, **47~61** (having pyridine ring), **45** (having pyrazine ring) exhibited higher potency than compound **46** which has a benzene ring at this site.

In the H-bond acceptor contour map (Fig. 4c), a purple cricoid polyhedron is observed revolving around -NHSO$_2$- between ring-A and ring-B, indicating that H-bond acceptor substituents may increase the activity. However, a red contour close to this position implies that H-bond acceptor groups have a negative effect on the inhibitory activity. This phenomenon can be explained by the fact that the -NH- of -NHSO$_2$- as H-bond donor and -SO$_2$- as H-bond acceptor would increase the activity of inhibitors. Additionally, another medium sized red contour was observed around -NHCOCH$_3$ (region A) suggesting that H-bond acceptor groups at this position are unfavorable for the inhibitory activity. However, a purple contour close to this position

implies that H-bond acceptor groups have a positive effect on the inhibitory activity. This is consistent with the fact that the -NH- of -NHCOCH$_3$ as H-bond donor and -CO- as H-bond acceptor.

Docking analysis and comparison with 3D contour maps

Molecular docking for all 61 inhibitors was performed to find the probable binding conformation of the ligands in the active sites of PI3Kα, and to predict the binding affinity of the molecules, as well as to complement the 3D-QSAR studies for the rational design of drugs [44]. To elucidate the interaction mechanism, compound **54**, one of the most potent inhibitors among the whole dataset, was selected for detailed analysis. The ligand core is anchored in the binding site via seven H-bonds with the protein which are depicted in Fig. 5. The analysis shows that N atom of pyridine ring forms two H-bonds with the hydrogen atom of -NH of ASP933 (2.79 Å, 135.5°) and -OH of TYR836 (2.85 Å, 102.6°), respectively, while the benzo-thiazole ring forms a H-bond with SER854 (1.93 Å, 122.4°). The hydrogen atom of -NHCO- (region-A) forms two H-bonds with VAL851 (2.56 Å, 176°) and SER854 (1.82 Å, 149.6°), respectively. In addition, the -NHSO$_2$- group forms a H-bond with LYS776 (1.98 Å, 148°) and a H-bond with ASP933 (2.04 Å, 116.0°).

In this docking result, the conserved lipophilic adenine region of PI3Kα is made up of ILE848, ILE932, VAL850, VAL851, MET922 and PHE930 residues (shown in

Fig. 5 Docking results.
(a) Surface of the binding site surrounding the compound **54**.
(b) Docked conformation derived for compound **54** in complex to the active site of p110α. The ligands are colored in magenta and key amino acid residues in white labels around compound **54** within 4 Å. H-bonds are shown in green dashed lines
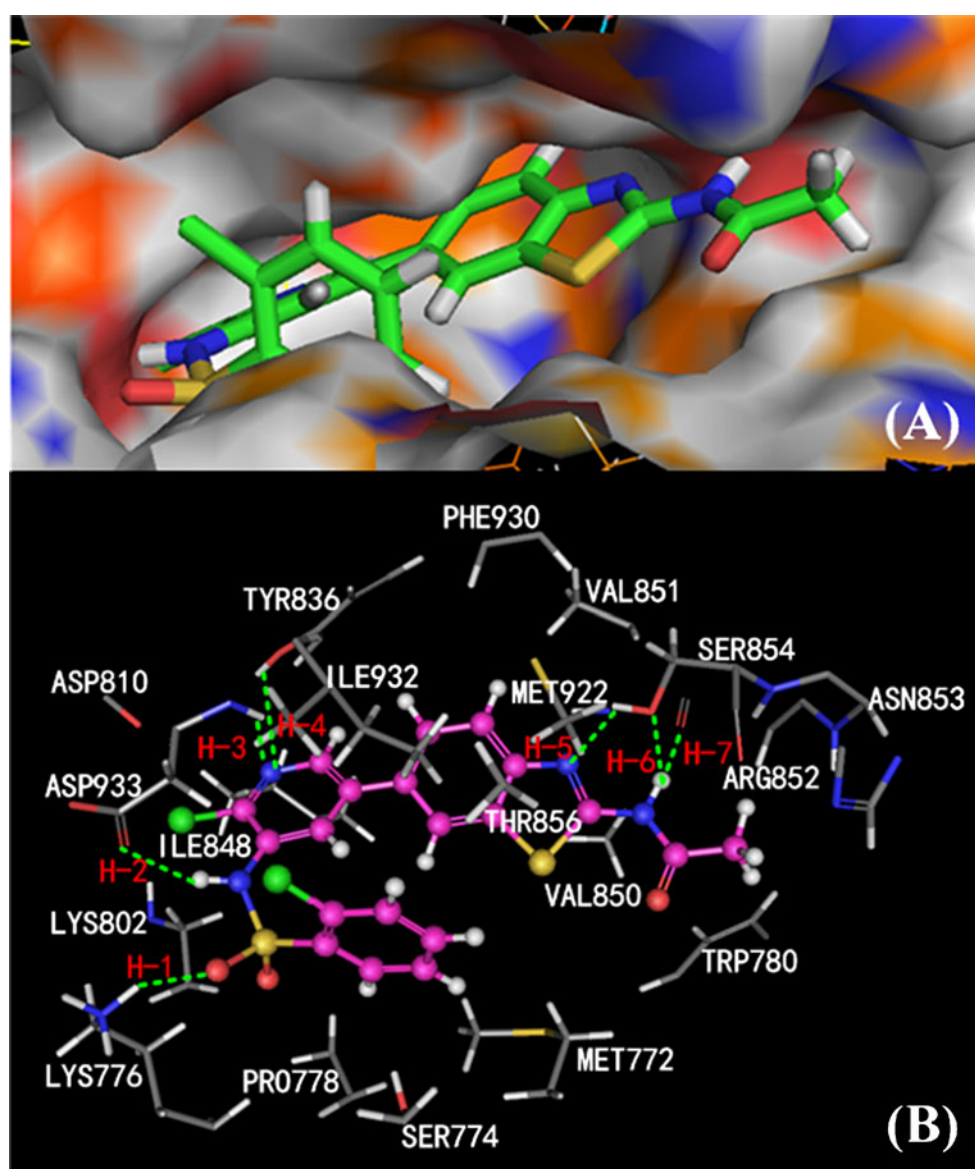

Fig. 5b) and accommodates the flat benzothiazole ring and pyridine ring of the compounds. Of note, the two hydrophilic rings present here are a contradiction to the hydrophobic environment, and besides the H-bonding analysis demonstrates that there are three H-bonds formed between these two rings and the side-chains of residues ASP933, TYR836 and SER854, which are not observed in the H-bond donor contour map of the CoMSIA model. These issues will be discussed in Comparison of the results of docking and MD simulation.

We proposed that a hydrophobic substituent in the benzene ring (ring-A) would be beneficial for the inhibitory activity because it would project into a hydrophobic groove comprised of ILE932, PRO778 and MET772. This is consistent with the smaller green (sterically favorable) contour of CoMFA steric field contour map (Fig. 3b) and

the yellow (hydrophobic favorable) contour of CoMSIA hydrophobic contour map (Fig. 4b).

The polar interaction of the -NHSO$_2$- or other groups with polar amino acids results in the flat aromatic scaffold (ring-A) extending into the hydrophobic groove, which is helpful for increasing the binding affinity. It is also consistent with the CoMFA electrostatic field contour map (Fig. 3c) and H-bond acceptor field contour map (Fig. 4c) of the CoMSIA model.

The -NHCOCH$_3$ group (region-A) which is linked with ring-D, forms a H-bond with the side chain of VAL851 and SER854. Since SER854, ASN853 and TRP780 are not located in the highly conserved adenine, sugar, or phosphate regions but in the solvent-accessible region of the ATP binding pocket [45], the solvent-accessible region can be exploited to increase the binding affinity

and helps the amino acid residues lock the ligand in the ATP binding pocket. This is consistent with the red contour of the electrostatic field contour map (Fig. 4a) and H-bond acceptor field contour map (Fig. 4c) of the CoMSIA model.

The results indicate that the benzothiazole derivatives are of an ideal length (about 14.7 Å) for forming tight H-bonds with ASP933 at one end of the pocket and with SER854 at the other end. In general, the results of the docking studies and its comparison with 3D contour maps can complement and validate each other, indicating that the 3D-QSAR model which we developed is reasonable and can offer constructive suggestions to the further rectification and modification of inhibitors of PI3Kα.

Comparison of the results of docking and MD simulation

Although docking analysis can provide a good starting point for further calculations with the purpose of predicting the binding modes, the solvent effect on the ligands and kinases and the potential ligand-induced conformational changes are not fully taken into account. Therefore, MD simulations were undertaken in the state close to natural conditions with the aim of checking the stability of the complex in aqueous solution.

The 10 ns simulations of the docked complex structure of modified P110α with ligand 54 were performed to obtain a dynamical picture of the conformational changes that occur in aqueous solution, with the main purpose of exploring the conformational alterations that take place in the ligand 54 and the P110α. The root mean square deviation (RMSD) of the receptors and the ligands is shown in Fig. 6. The RMSD of the trajectory with respect to their initial structure ranges from 0.9 to 4.1 Å. After 4 ns, the RMSD of the complex reaches about 3.5 Å, and retains this value throughout the simulation, which indicates that the docked complexes can reach metastable conformation
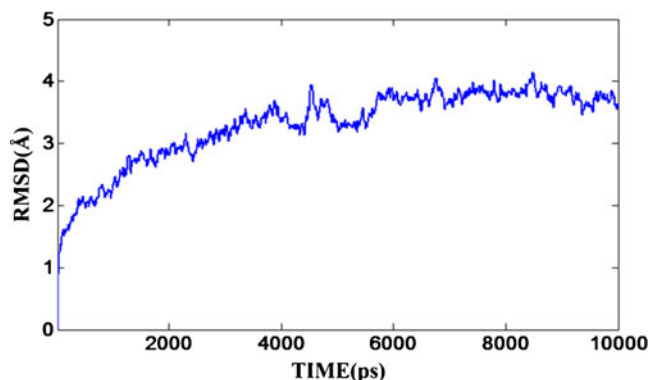

Fig. 6 Plot of the root-mean-square deviation (RMSD) of docked complex versus the MD simulation time in the MD-simulated structures

after 4 ns of simulation. From the average structure of ensemble for the last 500 ps (Fig. 7 a and b), we can see that it mainly forms four H-bonds. The H-bond formed between ligand and ASP933 (1.83 Å, 118°) also exists in the docking result. The O atom of -NHSO$_2$- of the ligand forms a H-bond with LYS802 (2.97 Å, 114.7°), which is similar to the docking result (forms a H-bond with LYS776). Another oxygen atom of -NHSO$_2$- forms a H-bond with ASN920 (3.21 Å, 151.8°). In addition, the O atom of -NHCOCH$_3$ of the ligand forms a H-bond with GLN859 (2.35 Å, 166.6°). During docking simulation, seven H-bonds were formed between compound 54 and P110α. However, after MD simulations, only four H-bonds were preserved, and the amino acids involved in forming the H-bonds also changed as described above.

As seen from Fig. 7b, in the solvent-accessible region, the H-bonds between the -NH group of -NHCOCH$_3$ and SER854 in the initial simulation structure broke within a very short time and the O atom forms a H-bond with GLN859. This phenomenon could be interpreted as the solvent affects the stability of the H-bond, meanwhile, it demonstrates that H-bond acceptor and donor groups are all quite essential at this position to the binding affinity increase. And this also explains why there are two opposite contour areas at the same position in H-bond acceptor contour map of CoMSIA model (Fig. 4c).

In the docking result we found that there are two hydrophilic rings present in the hydrophobic environment, besides, there are three H-bonds formed between these two rings and the sidechains of polar amino acid residues which are not observed in the H-bond donor contour map of the CoMSIA model. It is because the docking procedure ignores both the flexibility of the protein and the effect of water solvation. However, MD simulations treat both the ligand and the protein in a flexible way, allowing for an induced fit of the receptor-binding site around the ligand. Therefore, these two hydrophilic rings extend out of the hydrophobic pocket slightly and can not form H-bonds with residues. These are consistent with the hydrophobic field and H-bond acceptor field counter maps of CoMSIA model. So, the conformations obtained after molecular dynamics are more reasonable than the docked conformations.

In the hydrophobic groove which is comprised of ILE932, PRO778 and MET772, the benzene substituents of the ligand happen to flip to be closer to polar amino acids such as ASN920 and SER774, which indicates that both the hydrophobic effect and the electrostatic force keep the structure more stable. Furthermore, the H-bonds between the ligand and ASP933 and LYS802 provide stability in addition to the H-bond with the ASN920. The docking result and MD simulation both show that residue ASP933 in PI3Kα is a key residue to confer the inhibitory activity. It can be seen, that the residues in this binding

Fig. 7 MD results. (**a**) Surface of the binding site surrounding the compound **54**. (**b**) Plot of the MD-simulated structures of the binding site with ligand **54**. H-bonds are shown in green dashed lines, active site amino acid residues are represented as sticks, the inhibitor is shown as stick and ball model
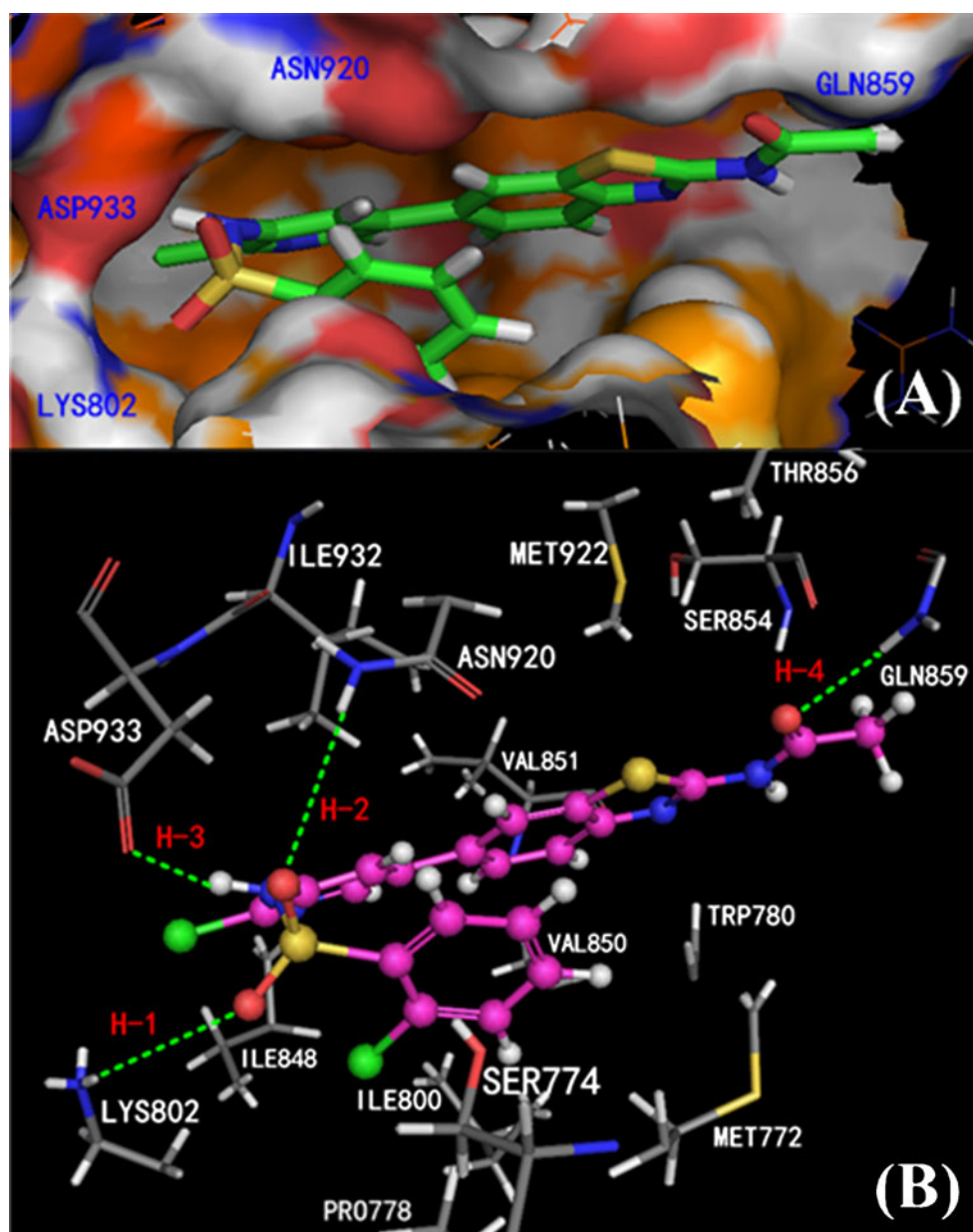
pocket were hydrophilic and some are important H-bond acceptors or donors, which are consistent with the hydrophobic field, H-bond acceptor field and electrostatic field counter maps of the CoMSIA model. The analysis described above suggests that there are no significant differences between the docked model of the complex and the stable structure extracted from MD simulations, which indicates that the docking model is rational and valid.

Conclusions

Understanding intermolecular interactions of benzothiazole derivatives with PI3Kα was achieved by performing molecular 3D-QSAR, docking, and molecular dynamics. The QSAR models exhibit relatively high r_{cv}^2, r_{pred}^2 and small SEE values, along with further testing, indicating that the obtained models were valuable in predicting the inhibitory activity of benzothiazole derivatives against the protein target. Furthermore, the 3D contour maps produced by the best CoMFA and CoMSIA models along with the docking results offered useful information to understand the structure-activity relationship and identified the structural features influencing the inhibitory activity. To validate the binding mode and elucidate the effects of ligand binding on the receptor conformation, MD simulation is performed. A set of 3D contour maps reveal that moderate bulky and hydrophobic

substituents at ring-A region are favored, which fits with the hydrophobic region that consists of hydrophobic residues in the active pocket. The moderate bulky, electron-withdrawing group on the para position of ring-B and hydrophilic substituents at ring-B region may benefit the potency. Besides, the polar substituents between ring-A and ring-B are preferential, indicating that there is a counterpart "receptor" favorable region for hydrophilic substituents, as well as a "receptor" favorable region for H-bonding interactions at this place. Similarly, moderate length and weak electronegative substituent at region-A might enhance the activity. The good consistency between the 3D-QSAR, the docking and MD modeling results implies the robustness of the 3D-QSAR models. Therefore, these models are useful in predicting the activity of a new PI3Kα inhibitor and can offer guidelines for the further modification of ligand design.

References

1. Leevers SJ, Vanhaesebroeck B, Waterfield MD (1999) Signalling through phosphoinositide 3-kinases: the lipids take centre stage. Curr Opin Cell Biol 11:219–225
2. Stein RC, Waterfield MD (2000) PI3-kinase inhibition: a target for drug development? Mol Med Today 6:347–357
3. Chang HW, Aoki M, Fruman D, Auger KR, Bellacosa A, Tsichlis PN, Cantley LC, Roberts TM, Vogt PK (1997) Transformation of chicken cells by the gene encoding the catalytic subunit of PI 3-kinase. Science 276:1848–1850
4. Katso R, Okkenhaug K, Ahmadi K, White S, Timms J, Waterfield MD (2001) Cellular function of phosphoinositide 3-kinases: implications for development, homeostasis, and cancer. Annu Rev Cell Dev Biol 17:615–675
5. Jiang BH, Liu LZ (2008) PI3K/PTEN signaling in tumorigenesis and angiogenesis. BBA-Proteins Proteom 1784:150–158
6. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ, Willson JK, Markowitz S, Kinzler KW, Vogelstein B, Velculescu VE (2004) High frequency of mutations of the PIK3CA gene in human cancers. Science 304:554
7. Graupera M, Guillermet-Guibert J, Foukas LC, Phng LK, Cain RJ, Salpekar A, Pearce W, Meek S, Millan J, Cutillas PR, Smith AJ, Ridley AJ, Ruhrberg C, Gerhardt H, Vanhaesebroeck B (2008) Angiogenesis selectively requires the p110alpha isoform of PI3K to control endothelial cell migration. Nature 453:662–666
8. Schultz RM, Merriman RL, Andis SL, Bonjouklian R, Grindey GB, Rutherford PG, Gallegos A, Massey K, Powis G (1995) In vitro and in vivo antitumor activity of the phosphatidylinositol-3-kinase inhibitor, wortmannin. Anticancer Res 15:1135–1139
9. Wymann MP, Bulgarelli-Leva G, Zvelebil MJ, Pirola L, Vanhaesebroeck B, Waterfield MD, Panayotou G (1996) Wortmannin inactivates phosphoinositide 3-kinase by covalent modification of Lys-802, a residue involved in the phosphate transfer reaction. Mol Cell Biol 16:1722–1733
10. Vlahos CJ, Matter WF, Hui KY, Brown RF (1994) A specific inhibitor of phosphatidylinositol 3-kinase, 2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one (LY294002). J Biol Chem 269:5241–5248
11. Frojdo S, Cozzone D, Vidal H, Pirola L (2007) Resveratrol is a class IA phosphoinositide 3-kinase inhibitor. Biochem J 406:511–518
12. Marion F, Williams DE, Patrick BO, Hollander I, Mallon R, Kim SC, Roll DM, Feldberg L, Van Soest R, Andersen RJ (2006) Liphagal, a Selective inhibitor of PI3 kinase alpha isolated from the sponge akacoralliphaga: structure elucidation and biomimetic synthesis. Org Lett 8:321–324
13. Pat. WO2006081659, 2006-08-10
14. Hayakawa M, Kaizawa H, Moritomo H, Koizumi T, Ohishi T, Okada M, Ohta M, Tsukamoto S, Parker P, Workman P, Waterfield M (2006) Synthesis and biological evaluation of 4-morpholino-2-phenylquinazolines and related derivatives as novel PI3 kinase p110alpha inhibitors. Bioorg Med Chem 14:6847–6858
15. Hayakawa M, Kawaguchi K, Kaizawa H, Koizumi T, Ohishi T, Yamano M, Okada M, Ohta M, Tsukamoto S, Raynaud FI, Parker P, Workman P, Waterfield MD (2007) Synthesis and biological evaluation of sulfonylhydrazone-substituted imidazo[1,2-a]pyridines as novel PI3 kinase p110alpha inhibitors. Bioorg Med Chem 15:5837–5844
16. Frederick R, Denny WA (2008) Phosphoinositide-3-kinases (PI3Ks): combined comparative modeling and 3D-QSAR to rationalize the inhibition of p110alpha. J Chem Inf Model 48:629–638
17. Han M, Zhang JZ (2010) Class I phospho-inositide-3-kinases (PI3Ks) isoform-specific inhibition study by the combination of docking and molecular dynamics simulation. J Chem Inf Model 50:136–145
18. Li Y, Wang Y, Zhang F (2010) Pharmacophore modeling and 3D-QSAR analysis of phosphoinositide 3-kinase p110alpha inhibitors. J Mol Model 16:1449–1460
19. D'Angelo ND, Kim TS, Andrews K, Booker SK, Caenepeel S, Chen K, D'Amico D, Freeman D, Jiang JA, Liu LB, McCarter JD, Miguel TS, Mullady EL, Schrag M, Subramanian R, Tang J, Wahl RC, Wang L, Whittington DA, Wu TA, Xi N, Xu Y, Yakowec P, Yang K, Zalameda LP, Zhang N, Hughes P, Norman MH (2011) Discovery and Optimization of a Series of Benzothiazole Phosphoinositide 3-Kinase (PI3K)/Mammalian Target of Rapamycin (mTOR) Dual Inhibitors. J Med Chem 54:1789–1811
20. Cramer RD III, Bunce JD, Patterson DE, Frank IE (1988) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. Quant Struct Act Rel 7:18–25
21. Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146
22. Matthew C, Richard DC III, Van Nicole O (1989) J Comput Chem 10:982–1012
23. Huang CH, Mandelker D, Schmidt-Kittler O, Samuels Y, Velculescu VE, Kinzler KW, Vogelstein B, Gabelli SB, Amzel LM (2007) The structure of a human p110α/p85α complex elucidates the effects of oncogenic PI3Kα mutations. Science 318:1744–1748
24. Mandelker D, Gabelli SB, Schmidt-Kittler O, Zhu J, Cheong I, Huang CH, Kinzler KW, Vogelstein B, Amzel LM (2009) A frequent kinase domain mutation that changes the interaction between PI3Kα and the membrane. Proc Natl Acad Sci USA 106:16996–17001
25. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16:566–567
26. Wang Y, Li Y, Ma Z, Yang W, Ai C (2010) Mechanism of microrna-target interaction: Molecular dynamics simulations and thermodynamics analysis. PLoS Comput Biol 6:e1000866
27. Kuntz I, Blaney J, Oatley S, Langridge R, Ferrin T (1982) A geometric approach to macromolecule-ligand recognitions. J Mol Biol 161:269–288
28. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking1. J Mol Biol 267:727–748

29. Eldridge M, Murray C, Auton T, Paolini G, Mee R (1997) The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 11:425–445

30. Muegge I, Martin YC (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 42:791–804

31. AbdulHameed MDM, Hamza A, Liu J, Zhan CG (2008) Combined 3D-QSAR modeling and molecular docking study on indolinone derivatives as inhibitors of 3-phosphoinositide-dependent protein kinase-1. J Chem Inf Model 48:1760–1772

32. Cramer RD III, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967

33. Nilsson J (1998) Multiway Calibration in 3D QSAR. University of Groningen, Applications to Dopamine Receptor Ligands

34. Uddin R, Yuan H, Petukhov PA, Choudhary MI, Madura JD (2008) Receptor-based modeling and 3D-QSAR for a quantitative production of the butyrylcholinesterase inhibitors based on genetic algorithm. J Chem Inf Model 48:1092–1103

35. Wold S, Ruhe A, Wold H, Dunn W III (1984) The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM J Sci Stat Comput 5:735–743

36. Case D, Darden T, Cheatham Iii T, Simmerling C, Wang J, Duke R, Luo R, Crowley M, Walker RC, Zhang W (2008) AMBER 10. University of California, San Francisco

37. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999–2012

38. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 23:1623–1641

39. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174

40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

41. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341

42. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593

43. Wang F, Ma Z, Li Y, Zhu S, Xiao Z, Zhang H, Wang Y Development of in silico models for pyrazoles and pyrimidine derivatives as cyclin-dependent kinase 2 inhibitors. J Mol Graph Model 30:67–81

44. Liu J, Zhang H, Xiao Z, Wang F, Wang X, Wang Y (2011) Combined 3D-QSAR, Molecular Docking and Molecular Dynamics Study on Derivatives of Peptide Epoxyketone and Tyropeptin-Boronic Acid as Inhibitors Against the β Subunit of Human 20S Proteasome. Int J Mol Sci 12:1807–1835

45. Vulpetti A, Bosotti R (2004) Sequence and structural analysis of kinase ATP pocket residues. II Farmaco 59:759–765