ORIGINAL PAPER

Pharmacophore modeling and 3D-QSAR analysis of phosphoinositide 3-kinase p110α inhibitors

Yiping Li · Yawen Wang · Fuqiang Zhang

Abstract Pharmacophore modeling studies were undertaken for a series of compounds belonging several groups of phosphoinositide 3-kinase (PI3K) p110α inhibitors: 4-morpholino-2-phenylquinazolines derivatives, pyrido [3′,2′:4,5]furo-[3,2-d]pyrimidine derivatives, imidazo[1,2-a] pyridine derivatives, sulfonylhydrazone substituted imidazo [1,2-a]pyridines, and LY294002. A five-point pharmacophore with three hydrogen bond acceptors (A), one hydrophobic group (H), and one aromatic ring (R) as pharmacophore features was developed. The pharmacophore hypothesis yielded a statistically significant 3D-QSAR model, with a correlation coefficient of $R^2=0.95$ for training set compounds. The model generated showed excellent predictive power, with a correlation coefficient of $Q^2=0.88$ and $r_{pret}^2=0.95$ for a test set of 14 compounds. Furthermore, the structure–activity relationships of PI3K p110α inhibitors were elucidated and the activity differences between them discussed. Docking studies were also carried out wherein active and inactive compounds were docked into the active site of the PI3K p110α crystal structure to analyze PI3K p110α–inhibitor interactions. The results provide insights that will aid optimization of these classes of PI3K p110α inhibitors for better activity, and may prove helpful for further lead optimization and virtual screening.

Keywords Phosphoinositide 3-kinase p110α · Pharmacophore modeling · 3D-QSAR

Y. Li (✉) · F. Zhang
Department of Pharmacy, College of Medicine,
Xi'an Jiaotong University,
Xi'an 710061, People's Republic of China
e-mail: liyipingye@sina.com

Y. Wang
First Affiliated Hospital, College of Medicine,
Xi'an Jiaotong University,
Xi'an 710061, People's Republic of China

Introduction

Phosphoinositide 3-kinases (PI3Ks) catalyze phosphorylation of the 3-OH group of phosphatidylinositol (PI) to generate phosphatidylinositol 3-phosphate (PIP), phosphatidylinositol 3,4-bisphosphate (PIP$_2$), and phosphatidylinositol 3,4,5-trisphosphate (PIP$_3$), which act as second messengers [1]. In the appropriate cellular context, these three lipids can regulate a remarkably diverse array of physiological processes, including cell growth, survival, differentiation, and chemotaxis [2]. There are three major classes of PI3Ks—classes I, II and III—based on their sequences and substrate specificities. Class I PI3Ks are further divided into class IA enzymes (PI3Kα, β, and δ), which have a p85 regulatory subunit and three different catalytic subunits (p110α, β, and δ), and a class IB enzyme (PI3Kγ) with a p101 regulatory subunit and p110γ catalytic subunit, which are activated by tyrosine kinases or G protein-coupled receptors to generate PIP$_3$. Class II and III PI3Ks play a key role in intracellular trafficking through the synthesis of PIP and PIP$_2$ [2, 3].

PI3K p110α, which is activated by receptor tyrosine kinases, undergoes gene amplification and is frequently mutated in primary tumors [4]. In addition, the lipid phosphatase PTEN, which removes the 3′ phosphate of PIP$_3$ and attenuates signaling downstream of activated PI3K, is frequently mutated or lost in primary tumors [5, 6]. Recently, it has been confirmed that angiogenesis, which is

essential for tumor growth and metastasis, selectively requires the p110α isoform of PI3K to control endothelial cell migration [7]. All these facts suggest that PI3K p110α is a potential and attractive target for cancer therapy and hence spark great interest in the discovery and development of inhibitors. To date, a few selective inhibitors of the ATP-binding site of p110α have been developed as anticancer agents [8–11]. However, these inhibitors have some inherent weaknesses, such as short half-life and low water-solubility [10, 11]. Frederick and Denny [12] have developed a homology modeled PI3K p110α structure using PI3K p110γ in order to help understand the geometry of the active site. Recently, Huang et al. [13] reported the crystal structure of the human p110α/p85α complex. In the field of computer-aided drug design, ligand-based pharmacophore modeling imparts a qualitative picture of the geometry of the active site by identifying the important features for binding of ligands and their spatial arrangements. This approach is complementary to receptor-based modeling. However, little structure–activity relationship (SAR) data is available for p110α inhibitors.

In the present work, the PHASE module of the Schrodinger suite, which is a more recently developed pharmacophore modeling package, was used to develop pharmacophore modeling and a 3D-QSAR model of p110α inhibitors. PHASE performs systematic explorations about rotatable bonds and calculates the associated conformational energies, retaining only the most reasonable conformations. It quickly locates plausible pharmacophores using a high-dimensional, tree-based partitioning algorithm in which pharmacophores from different conformations are placed in multi-dimensional boxes; each box represents a common pharmacophore only if it contains a sufficient number of active ligands. After PHASE has located reasonable alignments of active ligands, pharmacophores thus generated are evaluated according to an open, highly configurable scoring function. PHASE determines how molecular structure effects drug activity by dividing space into a fine cubic grid, encoding atom type occupation as numerical information, and performing a partial least-squares (PLS) regression, resulting in the prediction of a significant model [14–16]. The volume occluded maps generated for the pharmacophore help explain the observed variation in activity according to the variation in structural features [17–19]. Therefore, the aim of our study was to generate a pharmacophore model for p110α inhibitors, explain their mechanism of action on p110α, and to build a robust 3D-QSAR model that can be used for lead optimization and virtual screening. We also performed docking studies to determine the binding mode of p110α inhibitors and to gain further insight into their SAR. Such a model would be of great assistance in a p110α development program.

Materials and methods

PHASE 2.5 implemented in the Maestro 7.5 software package (Schrodinger, LLC, New York, NY) [14–16] was used to generate pharmacophore and 3D-QSAR models for PI3K p110α inhibitors. All docking studies were performed using a graphical user interface—Induced Fit Docking of the program Maestro 7.5 [20–22]. The protein structure of a complex human PI3K p110α/p85α was obtained from the RCSB Protein Data Bank (PDB) as entry 2RD0 (http://www.rcsb.org/pdb/explore/explore.do? structureId= 2RD0).

Data set

Table 1 shows the structures of the 66 compounds used in this study and their observed activity (pIC$_{50}$). The reported series of compounds includes 24 4-morpholino-2-phenylquinazo-lines derivatives [8], 16 pyrido[3′,2′:4,5]furo[3,2-d] pyrimidine derivatives [9], 12 imidazo[1,2-a]pyridine derivatives [10], 13 sulfonylhydrazone substituted imidazo[1,2-a]pyridines [11], and LY294002 [8]. The in vitro biological activity data was reported as IC$_{50}$. The IC$_{50}$ values were converted to pIC$_{50}$ using the formula (pIC$_{50}$= −log IC$_{50}$). Compound selection and experimental activity (IC$_{50}$) against PI3K p110α all originated from the same laboratory. Figure 1 shows the distribution of pIC$_{50}$ for the whole data set, ranging from 4.22 to 9.59. The compounds fit roughly into four compound clusters. We divided the data set, randomly choosing 52 compounds to be in the QSAR training set and 14 compounds for the test set. The compound number and pIC$_{50}$ values of the test set are given in bold in Table 1. Care was taken to ensure a uniform distribution of structurally different compounds from each cluster with a wide range in pIC$_{50}$ for both training and test sets.

Ligand preparation

Three-dimensional (3D) conversion and minimization was performed using LigPrep (MMFFs force field [23]) incorporated in PHASE. Conformers were generated using a rapid torsion angle search approach followed by minimization of each generated structure using the MMFFs force field, with an implicit GB/SA solvent model. A maximum of 1,000 conformers were generated per structure using a preprocess minimization of 1,000 steps and postprocess minimization of 500 steps. Each minimized conformer was filtered through a relative energy window of 50 kJ mol^{-1} and a minimum atom deviation of 1.00 Å [14–19]. This value (50 kJ mol^{-1}) sets an energy threshold relative to the lowest-energy conformer. Conformers that are higher in energy than this threshold are discarded. All distances between pairs of corresponding heavy atoms must be below 1.00 Å for two conformers to be considered identical. This

Table 1 Numbering, structure, and experimental activity of the data set. The compound number and pIC$_{50}$ values of the test set are given in *bold*

Compound	Structure			pIC$_{50}$
LY294002				6.20

Compound	R$_1$	R$_2$	R$_3$	pIC$_{50}$
1	HO—	O,N (morpholine)		5.89
2	HO—	N-piperidine		4.22
3	HO—	OH piperidine		**4.92**
4	HO—	O piperidinone		4.68
5	HO—	MeO~N(H)~		4.82
6	HO—	O,N	2-OH	**5.54**
7	HO—	O,N	3-OH	7.13
8	HO—	O,N	4-OH	5.55
9	HO—	O,N	3-OMe	6.22
10	HO—	O,N	3-F	5.75
11	HO—	O,N	3-CH$_2$OH	7.00
12	HO—	O,N	4-CH$_2$OH	5.43
13	HO—	O,N	2-NH$_2$	5.41

threshold is applied only after the energy difference threshold, and only if the two conformers are within 1 kcal mol^{-1} of each other.

Creating pharmacophore sites

Each compound structure is represented by a set of points in 3D space, which coincide with various chemical features that may facilitate noncovalent binding between the compound and its target receptor. PHASE provides a built-in set of six pharmacophore features, hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively ionizable (N), positively ionizable (P), and aromatic ring (R). The rules that are applied to map the positions of pharmacophore sites are known as feature definitions, and are represented internally by a set of

SMARTS [15] patterns. Each pharmacophore feature is defined by a set of chemical structure patterns. All user-defined patterns are specified as SMARTS queries and assigned one of three possible geometries, which define physical characteristics of the site:

(1) Point: the site is located on a single atom in the SMARTS query.
(2) Vector: the site is located on a single atom in the SMARTS query, and will be assigned directionality according to one or more vectors originating from the atom.
(3) Group: the site is located at the centroid of a group of atoms in the SMARTS query. For aromatic rings, the site is assigned directionality defined by a vector that is normal to the plane of the ring.

Table 1 (continued)

14	HO structure	morpholine sulfonyl structure	3-NH$_2$	6.36
15	HO structure	morpholine sulfonyl structure	4-NH$_2$	**6.03**
16	OH structure	morpholine sulfonyl structure		4.52
17	HO structure	morpholine sulfonyl structure		5.01
18	MeO structure	morpholine sulfonyl structure		**5.68**
19	structure	morpholine sulfonyl structure		**4.85**
20	structure	morpholine sulfonyl structure	3-OH	7.25
21	N structure	morpholine sulfonyl structure	3-OH	**7.57**
22	N structure	morpholine sulfonyl structure	3-OH	7.89
23	N structure	morpholine sulfonyl structure	3-OH	**7.72**
24	S structure	morpholine sulfonyl structure	3-OH	8.60

Compound	X	Y	R	pIC$_{50}$
25	N	S	H	5.85
26	N	O	H	6.25
27	N	NH	H	4.52
28	N	S	Ph	5.77
29	N	O	Ph	**6.80**
30	C	S	H	5.62
31	C	S	Ph	4.52

Compound	R	pIC$_{50}$
32	2-OH	6.40
33	3-OH	**8.44**
34	4-OH	7.59
35	3-NH$_2$	6.85
36	3-OMe	**7.00**

A default setting having acceptor (A), donor (D), hydrophobic (H), negative (N), positive (P), and aromatic ring (R) was used to create pharmacophore sites. The morpholine ring of 4-morpholino-2-phenylquinazoline derivatives and pyrido [3′,2′:4,5]furo[3,2-d]pyrimidine derivatives, and the imidazole ring of imidazo[1,2-a]pyridine derivatives and sulfonylhydrazone-substituted imidazo[1,2-a]pyridines were defined as the hydrophobic group feature.

Active and inactive thresholds of pIC$_{50}$ of 8.50 and 5.00, respectively, were applied to the dataset to yield eight actives and nine inactives, which were used for pharmacophore generation and subsequent scoring.

Finding common pharmacophores and scoring hypotheses

Common pharmacophore features were identified from a set of variants—a set of feature types that define a possible pharmacophore—using a tree-based partitioning algorithm with a maximum tree depth of five. The terminal size of the pharmacophore box was 1 Å, which governs the tolerance

Table 1 (continued)

37	3-(Piperidinoethoxy)			6.77
38	3-(Morpholinoethoxy)			7.80
39	3-(Thiomorpholinoethoxy)			7.10
40	3-OCH$_2$COOH			7.62

Table 1.(*continued*)

Compound	Structure

Compound	R$_1$	R$_2$	X	R$_3$	pIC$_{50}$
41	H	Me	N	4-F	6.17
42	Cl	Me	N	4-F	6.12
43	Cl	Me	N	4-NO$_2$	4.52
44	Cl	Me	N	3- NO$_2$	6.55
45	Cl	Me	N	2-Me,5-NO$_2$	8.55
46	Br	Me	N	2-Me,5-NO$_2$	8.51
47	Br	H	N	2-Me,5-NO$_2$	8.75
48	Cl	H	CH	2-Me,5-NO$_2$	**7.00**

Compound	R$_1$	X	R$_2$	pIC$_{50}$
49	Cl	S	Me	7.09
50	Cl	SO	Me	7.51
51	Cl	SO$_2$	Me	8.55
52	Cl	SO	CH$_2$OH	7.70

Table 1.(*continued*)

Compound	Structure

Compound	R$_1$	R$_2$	R$_3$	pIC$_{50}$
53	Me	Me	H	6.40
54	H	Me	H	7.77
55	H	H	H	7.68
56	Me	Me	Me	**6.77**
57	H	Me	Me	7.09
58	H	H	Me	9.52

Compound	R$_1$	pIC$_{50}$
59	Cl	**9.11**
60	F	**8.28**
61	CF$_3$	7.85
62	Me	8.22
63	CN	9.59
64	CO$_2$Et	6.47
65	CONH$_2$	6.11
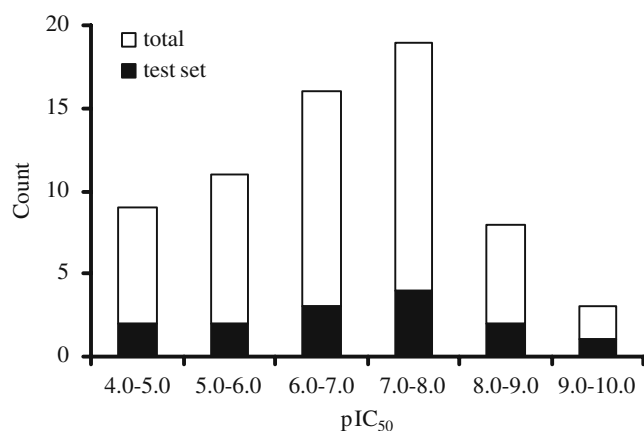
Fig. 1 Distribution of pIC_{50} for the data set

on matching; the smaller the box size, the more closely pharmacophores must match. Common pharmacophore hypotheses were generated by systematically varying the number of sites (n_{sites}) and the number of matching active compounds (n_{act}). With $n_{act}=n_{act\_tot}$ initially (n_{act_tot} is the total number of active compounds in the training set), n_{sites} was varied from seven to three until at least one hypothesis was found and scored successfully; if this failed, n_{act} was decreased by one and the n_{sites} cycle was repeated. The common pharmacophore hypotheses were scored by setting the root mean square deviation (RMSD) value below 1.0, the vector score value to 0.5 and weighting to include consideration of the alignment of inactive compounds using default parameters.

Building the 3D-QSAR model

All the common pharmacophore hypotheses successfully generated and scored in the above manner were used to generate atom-based 3D-QSAR models by correlating the observed and estimated activity for the set of 52 training molecules using PLS analysis. The PLS regression was carried out using PHASE with a maximum of $N/5$ PLS factors (N=number of ligands in training set, and a grid spacing of 1.0Å). All models were validated by predicting activity of the set of 14 test molecules.

Docking

The protein structure of the human PI3K p110α/p85α(2RD0) complex was obtained from the RCSB PDB. The protein structure was treated to add missing hydrogens, assign appropriate bond orders and formal charges, and to delete water molecules that are more than 5Å from the heterogeneous groups. Ligand preparation was accomplished on all 66 molecules using the Ligprep module of PHASE to clean the structure and generate tautomers as described above.

The center of binding site, defined as a Glide enclosing box, was set at the centroid of residues 851 and 802, while the size was set to a default value of 26Å. No H bonding constraint was defined.

The Induced Fit Docking protocol, also described elsewhere [20–22], is intended to circumvent the inflexible binding site requirement of grid-based docking through use of post-docking refinement steps. In "initial Glide docking", the protein structure was prepared and refined to minimize the default RMSD value of 0.18Å. Each ligand was then docked into the binding site of the PI3K p110α catalytic subunit using the standard precision (SP) scoring mode of Glide. The van der Waals radii of the ligand and protein were scaled down by 50% to compensate for the lack of protein flexibility during docking, and 20 poses were generated for each ligand in order to sample a wide range of possible docking modes. Using "Prime induced fit" mode, refinement of the predicted protein–ligand complexes was then performed using Prime. First, all side chains within a 5.0Å radius of each docked ligand pose were identified using Prime's side-chain sampling algorithms. This was followed by energy minimization of each docked protein–ligand complex using the OPLS2001 force field and Prime's implicit solvent model [24]. In "Glide redocking", the minimized ligand used in the first docking step is redocked using the Extra Precision (XP) scoring function of Glide with default settings into each of the 20 protein structures produced in the protein refinement step. The top scoring docked poses of each ligand were superimposed with the ligand alignments obtained using the pharmacophore hypotheses above.

Results and discussion

Determination of pharmacophore and 3D-QSAR models

Eight compounds with highest activity from the four diverse classes were selected for common pharmacophore hypotheses generation. Using a tree-based partition algorithm requiring that all eight active compounds must match, 11 five-featured probable common pharmacophore hypotheses were generated from the list of variants. No common pharmacophore hypotheses were obtained for seven and six common features. On applying the scoring function for five-featured common pharmacophore hypotheses using default values, three common pharmacophore hypotheses survived, belonging to the types AAAHR. Training set compounds were aligned on these common pharmacophore hypotheses and analyzed by PLS analysis described in PHASE with ten PLS factors. Over-fitting of results was observed when more than six PLS factors were utilized. The predictivity of each hypothesis was analyzed by the test set compounds. A summary of the statistical data of the three common

pharmacophore hypotheses labeled as Models 1–3, together with their survival scores, is listed in Table 2.

Reliable predictions can only come from statistically valid QSAR models. There are several statistical parameters, such as R^2, Q^2, SD, RMSE, and F, that can be used to evaluate the robustness of a QSAR model. Consider first which models are best based on high R^2. As seen from Table 2, the R^2 values of all three models are greater than 0.95, the SD values are lower than 0.3, and the F-test values are very high. This shows that these three models are interpret the SAR of this series of training set compounds satisfactorily. According to Tropsha [25], high R^2 is a necessary but not sufficient condition for a predictive QSAR model. Besides the consideration of high R^2, the best QSAR model should be chosen based on its predictive ability. So next consider which models are best based on high Q^2 and r_{pret}^2. The r_{pret}^2 value of the model generated is calculated by $r_{pret}^2 = 1 - PRESS/SD$, where SD is the sum of squared deviations between the biological activities of each molecule and the mean activity of the molecules in the training set, and PRESS is the sum of squared deviations between the predicted and actual activity values for every molecule in the test set. For each model, Tables 2 and 3 list Q^2, r_{pret}^2 and the residuals for each compound in the test set. The Q^2 value of Model 1 is much higher than that of Models 2 and 3. The r_{pret}^2 value of Model 1 is higher than that of Models 2 and 3. In fact, Model 1 has the highest Q^2 and r_{pret}^2 value of all the models we considered, suggesting that Model 1 is the best model.

Additionally, Model 1 has the lowest RMSE value and highest Pearson-R value of all the models. Further, five different combinations of training and test sets were generated and analyzed by PHASE PLS analysis. Good and consistent predictivity were observed for Model 1 for

each combination compared to the other models (see Electronic supplementary material). Considering the flexibility of all the molecules, the predictive qualities of the QSAR models are satisfactory. Overall, based on R^2, Q^2, r_{pret}^2, SD, and RMSE, as well as on the highest value on the Pearson-R, the best model was Model 1. The predicted and observed activity of the training and test set compounds of Model 1 are shown in Fig. 2.

Hence, Model 1 was selected as the pharmacophore model of PI3K p110α inhibitor and for the QSAR analysis of the PI3K p110α inhibitor below. This pharmacophore model includes two hydrophobic parts located at both ends, with one being a hydrophobic group and the other an aromatic ring. Two H-bond acceptors are located at both ends of the two hydrophobic parts, respectively, and one H-bond acceptor is located between the two hydrophobic parts. The pharmacophore model and the alignment of active compounds are shown in Fig. 3. The distances and angles between the five features are shown in Table 4.
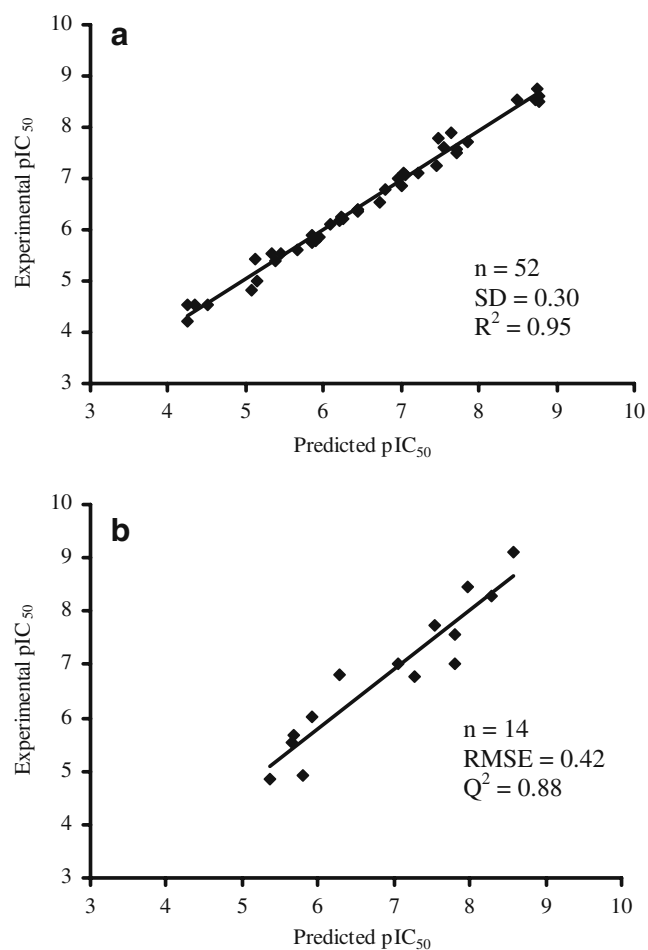
Analysis of atom-based PHASE 3D-QSAR model

The contribution maps obtained from Model 1 show how 3D-QSAR methods can identify features important for the interaction between ligands and their target protein. Such maps allow identification of those positions that require a particular physicochemical property to enhance the bioactivity of a ligand. A pictorial representation of the contours generated is shown in Fig. 4a–h. In these representations, gray cubes indicate favorable regions while e black cubes indicate unfavorable regions for activity.

Electron-withdrawing (N, O; includes hydrogen-bond acceptors)

Figure 4a–d compares the most significant favorable and unfavorable hydrogen bond acceptor and electron withdrawing features that arise when the QSAR model is applied to the more active compounds 58 and 24, and to the less active compounds 64 and 65. As seen from Fig. 4a,b, gray regions are observed near the N atom of the imidazo [1,2-a]pyridine ring, the O atom of sulfonyl moiety, and the O atom of the nitro group in the reference ligand— compound 58. Gray regions are also observed near the O atom of the morpholine ring, the N atom of the quinazoline ring, and the oxygen atom of the hydroxy group in compound 24. All the electron-withdrawing atoms in these positions are indicated to increase activity, and are also mapped as H-bond acceptors in the pharmacophore model features. The H-bond acceptor, N atom of imidazo[1,2-a] pyridine ring and the O atom of morpholine ring are especially significantly critical features. Due to the loss of this H-bond acceptor, compounds 2, 3, 4, and 5 are

Table 2 Summary of partial least-squares (PLS) analysis results for Models 1–3 with survival scores. *SD* Standard deviation of the regression, R^2 value of R^2 for the regression, *F* variance ratio, *P* significance level of variance ratio, *RMSE* root mean-square error, Q^2 value of Q^2 for the predicted activity, *Pearson R* correlation between the predicted and observed activity for the test set, r_{pret}^2 value of r_{pret}^2 for predicted activity

	Model 1	Model 2	Model 3
Survival score	3.35	3.09	3.00
SD	0.30	0.25	0.23
R^2	0.95	0.96	0.97
F	147.50	233.40	265.90
P	1.82e-27	1.14e-30	7.10e-32
RMSE	0.42	0.67	0.73
Q^2	0.88	0.73	0.67
Pearson R	0.95	0.86	0.82
r_{pret}^2	0.95	0.90	0.88

Table 3 Residuals of test set prediction for Models 1–3

Compound	Experimental pIC_{50}	Residuals from different models		
		Model 1	Model 2	Model 3
3	4.92	−0.87	−0.82	−1.16
6	5.54	−0.11	−0.17	−1.01
15	6.03	0.12	0.64	0.16
18	5.68	0	0.57	−0.1
19	4.85	−0.51	−0.33	0
21	7.57	−0.24	−0.26	0.28
23	7.72	0.19	0.32	0.38
29	6.80	0.51	0	−0.28
33	8.44	0.46	1.1	1
36	7.00	−0.05	0.22	−0.21
48	7.00	−0.81	−0.45	−1.04
56	6.77	−0.51	−1.65	−1.1
59	9.11	0.53	0.21	0.23
60	8.28	−0.01	0.5	1.19

inactive. In addition, this H-bond has been verified as a highly conserved interaction with the adenine pocket of the ATP-binding site of PI3K isoforms [3, 12, 26, 27]. The H-bond made by the O atom of the nitro group in compound 58, and the O atom of the hydroxy group is also critical for activity, contributing to the very low activity of compounds 1–5, 16–19, 25–31, and 41, which lacks this H-bond acceptor. Also, the nitro and hydroxy groups in the *meta* position on the terminal phenyl ring will improve activity, for example in compounds 24, 33, 46 and 58, while compound 43 with a nitro group in the *para* position on the phenyl ring, and compounds 34, 8, 32, and

Fig. 2 Scatter plots for the experimental and predicted pIC_{50} values for Model 1 applied to the training set (**a**) and the test set (**b**)
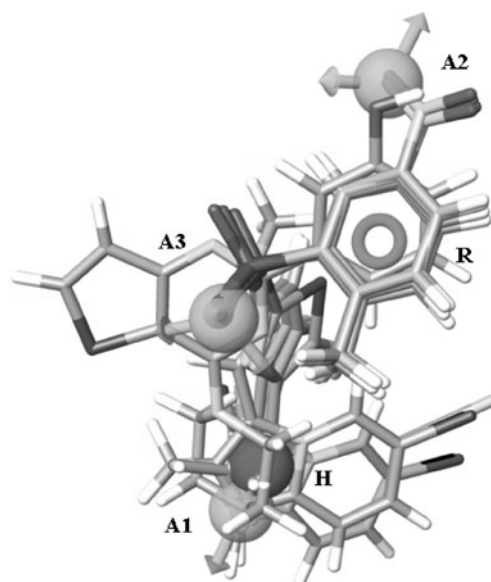
Fig. 3 The common pharmacophore model based alignment of active compounds. *Spheres with vectors A1, A2, A3* Acceptor feature, *Sphere R* aromatic ring feature, *Sphere H* hydrophobic

Table 4 Distances and angles of pharmacophore hypothesis Model 1

Site1	Site2	Distance (Å)[a]	Site1	Site2	Site3	Angle (degree)[b]	Site1	Site2	Site3	Angle (degree)[b]	Site1	Site2	Site3	Angle (degree)[b]
A1	A2	11.277	A2	A1	A3	27.5	A2	H	R	14.6	A1	A2	R	29.6
A1	A3	6.070	A2	A1	R	12.4	A3	H	R	32.2	A3	A2	H	26.7
A1	H	1.191	A2	A1	H	10.8	A1	A3	A2	127.1	A3	A2	R	32.0
A1	R	8.319	A3	A1	R	26.4	A1	A3	H	8.1	H	A2	R	30.2
A2	A3	6.523	A3	A1	H	38.0	A1	A3	R	110.5	A2	R	A1	138.0
A2	H	10.110	H	A1	R	20.1	A2	A3	H	119.0	A2	R	A3	118.9
A2	R	3.624	A1	H	A2	168.0	A2	A3	R	29.1	A2	R	H	133.1
A3	H	5.185	A1	H	A3	133.8	H	A3	R	103.5	A1	R	A3	43.1
A3	R	3.947	A1	H	R	156.7	A1	A2	A3	25.4	A1	R	H	3.2
H	R	7.212	A2	H	A3	34.4	A1	A2	H	1.3	A3	R	H	44.3

[a] Distance between Site1 and Site2

[b] Angle degree of Site1-Site2-Site3

6 with hydroxy groups in the *para* or *ortho* position on the phenyl ring have low activity, indicating the importance of the spatial position of this H-bond acceptor for activity. The H-bond made by the O atom of the sulfonyl group and the N atom of the quinazoline ring is beneficial for activity, contributing to the lower activities of compound 49 and LY294002, which lack this H-bond receptor. Therefore, it can be inferred that there are three H-bond interactions
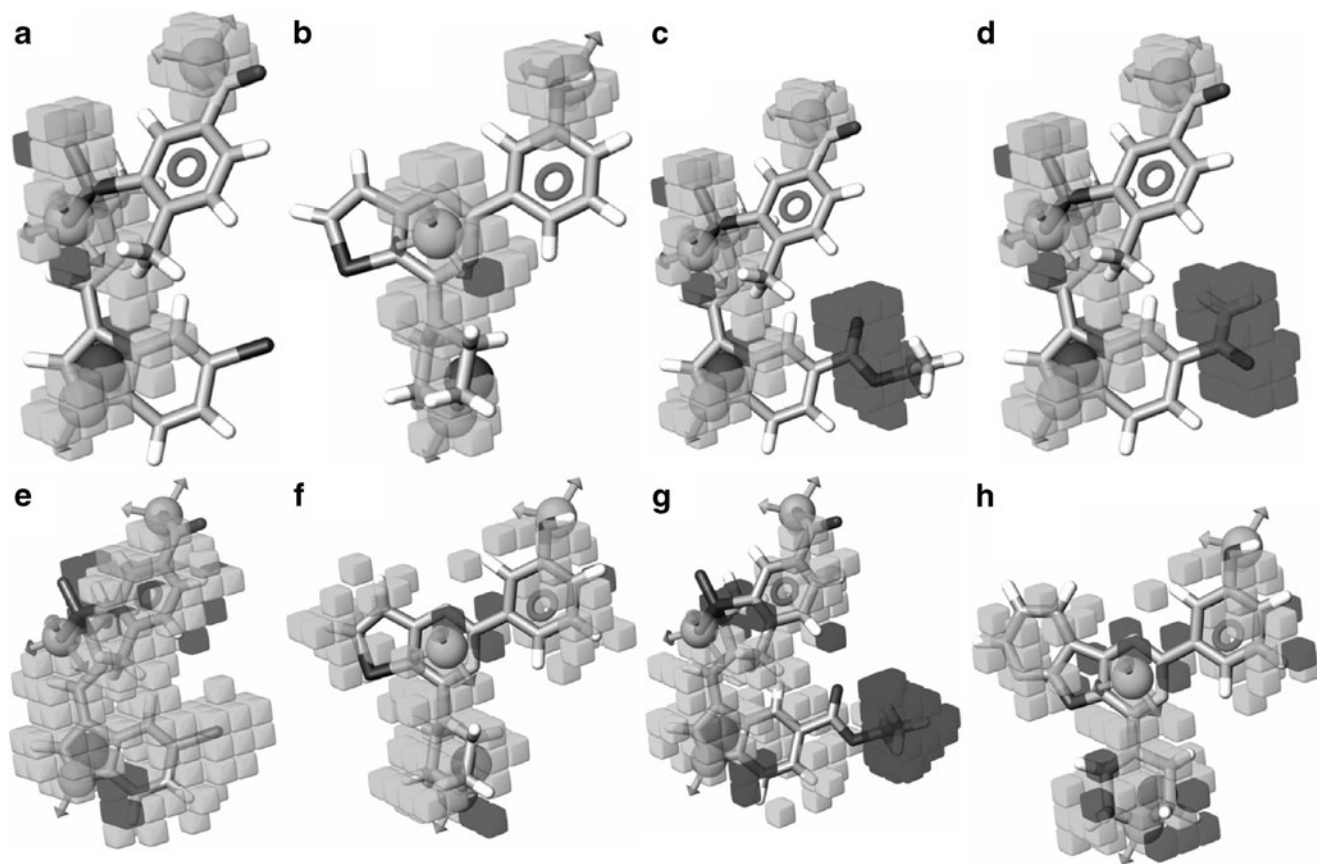
Fig. 4a–h Pictorial representation of the contours generated using the QSAR model. Hydrogen bond acceptor property and electron withdrawing features that arise when the QSAR model is applied to compounds 58 (**a**), 24 (**b**), 64 (**c**), and 65 (**d**). Hydrophobic interactions features that arise when the QSAR model is applied to compounds 58 (**e**), 24 (**f**), 64 (**g**), and 33 (**h**). *Gray cubes* favorable regions for activity, *black cubes* unfavorable region for activity

between PI3K p110α and its inhibitors, while the two hydrogen bond acceptors located at the terminii of the molecule are necessary for activity, and the intermediate H-bond acceptor may further stabilize this spatial arrangement.

As seen from Fig. 4c,d, black regions are observed near the two oxygen atoms of the ethyl carboxylate group in compound 64, and the nitrogen and oxygen atoms of the carboxamide group in compound 65, contributing to the low activity of these two compounds. This indicates that the 6-position of the imidazo[1,2-a]pyridine ring should not have electron-withdrawing atoms in order to improve activity. In addition, the low activity of compound 64 is also attributed to its steric and hydrophobic properties (see below).

Hydrophobic/non polar interactions

Figure 4e–h illustrates the significant favorable and unfavorable hydrophobic interactions that arise when the QSAR model is applied to the reference ligand compound 58—the most active compound—and compounds 24, 64, and 33. As seen from Fig. 4e, f, gray regions are observed near the phenyl ring and the imidazo[1,2-a]pyridine ring in compound 58, and the phenyl and morpholine rings in compound 24, indicating their importance for activity. Of these, two hydrophobic regions are also mapped as pharmacophore model features. In Fig. 4e, gray regions are observed near the Br atom in the 6-position of the imidazo[1,2-a]pyridine scaffold, which appears to be one of the pharmacophore features. This hydrophobic feature is present in most of the active compounds, e.g., active compounds 45–47, 58–60, and 63. Additionally, the activity of active compound 24, which lacks this hydrophobic feature, is lower than those active compounds 58, 59, and 63 with this feature, indicating the importance of this hydrophobic feature for activity despite the fact that this feature is missing from the developed pharmacophore model. However, compound 64, which has the same molecular scaffold as compound 58, has low activity. As seen from Fig. 4g, gray regions are observed near the carboxylate group in the 6-position of the imidazo[1,2-a]pyridine scaffold in compound 64. However, black regions are observed near the ethyl moiety, contributing to the low activity of compound 64. Additionally, the ethyl carboxylate group of compound 64 protrudes significantly out from the Br atom of the reference ligand, justifying the possibility of reduced activity due to steric factors. These findings show that a small hydrophobic group in the 6-position of the imidazo[1,2-a]pyridine scaffold is beneficial to activity.

In Fig. 4f,h, gray regions are observed near the thiophene ring in compound 24 and the pyrido[3′,2′:4,5] furane ring in compound 33, contributing to the higher activity of these two compounds.

Docking studies

As a measure of docking reliability, the docking results were evaluated through their RMSD values by comparison of the docked poses of the ligands and ligand alignments obtained from Model 1 of the pharmacophore hypotheses. The average RMSD value of active compounds was found to be 1.18 Å, with the most active compound showing a RMSD of 0.73 Å, the average RMSD value of inactive compounds being 1.06 Å, and the least active showing 1.01 Å (see Electronic supplementary material).

The binding mode of the most active compound (58) is shown in Fig. 5a. In its binding mode, the imidazo[1,2-a] pyridine scaffold is observed to be inserted deeply in the cavity, interacting with Tyr836 and Val851 through T-Shape and H-bond contacts, respectively. This deep cavity is very well conserved in all PI3K isoforms, and coincides with the ATP-binding site region. This binding orientation of the imidazo[1,2-a]pyridine scaffold is consistent with the results of Frederick and Denny [12]. The hydrophobic ring and H-bond acceptor of the imidazo[1,2-a]pyridine scaffold, which are beneficial to activity according to the 3D-QSAR model, are well complemented by the Tyr836 and Val851 residues through T-Shape interactions and H-bond contacts. The bromine atom in the 6-position of the imidazo[1,2-a] pyridine scaffold is located in a small lipophilic cavity bordered by residues Ile848, Lys802, Asp933, Ile932, and Tyr 836, supporting the benefit of a small, lipophilic substituent in that position in the 3D-QSAR model.

Two H-bonds, between the O atom of the nitro and Arg770 and between the O atom of the sulfonyl group and Gln859, are observed, indicating the important role of these H-bonds. The phenyl ring of compound 58 exhibited interactions with Met772 and Ser774, explaining why the phenyl ring is beneficial for activity in the 3D-QSAR model. Two H-bond acceptors of the nitro and sulfonyl groups, which are beneficial for activity in the 3D-QSAR model, are also well complemented by the Arg770 and Gln859 residues through H-bond contacts. All these interactions highlight the good complementarity between the 3D-QSAR model and the proposed ligand–protein docking model.

Compared with docking results obtained using the PI3K p110α homology modeling method mentioned by Frederick and Denny [12], the predicted binding orientation of compound 58 was similar to that of Frederick and Denny. But the latter study found the two H-bonds made by the O atom of the nitro and sulfonyl groups to be positioned at Ser773 and His855, respectively, while our findings indicated Arg770 and Gln859. Carefully comparison of the active site of human PI3K p110α (2RDO) with that of the homology model proposed by Frederick and Denny revealed a critical difference between them. Ser773 was orientated to stretch
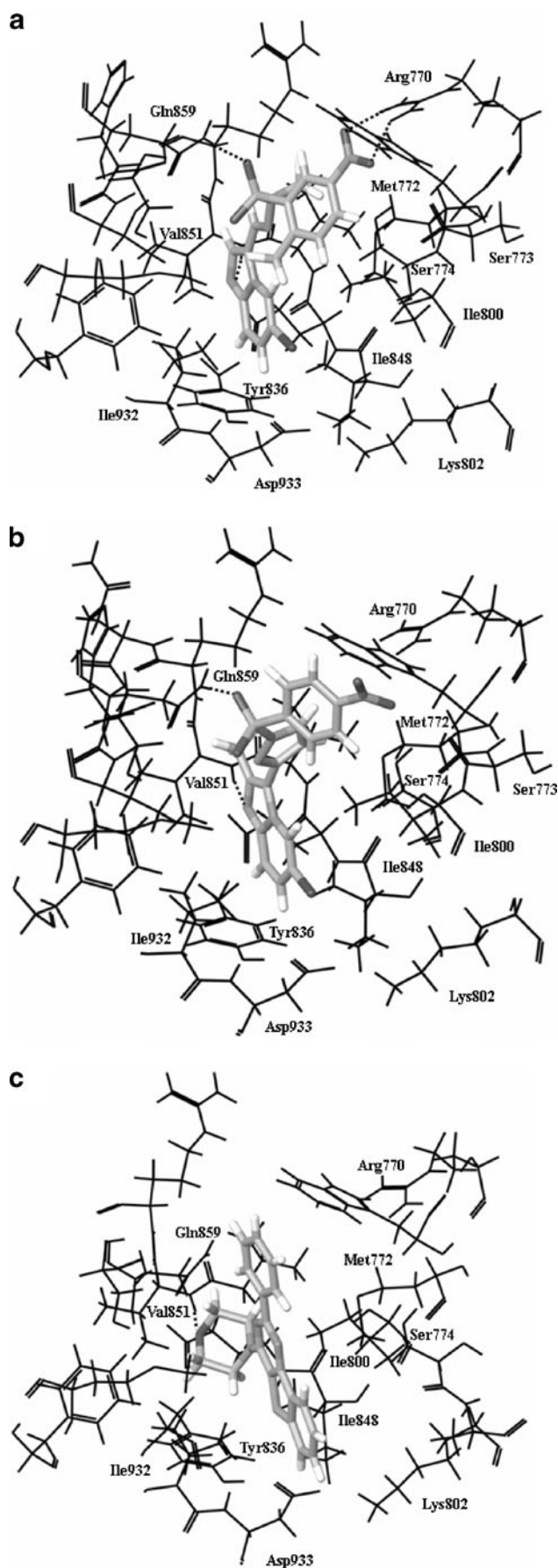
Fig. 5a–c Stereo-view of compounds 58 (**a**), 43 (**b**), and 31 (**c**) and their binding to active site amino acids. *Dotted lines* H-bonds

out to the active site of human PI3K p110α (2RDO), while, on the contrary, Ser773 was orientated to extend to the active site in the homology model (Fig. 5a,b). This explains why the H-bond made by the O atom of the nitro cannot be positioned to Ser773 in our docking model. The reason that Ser773 was orientated to extend to the active site of the homology model is related to the protein structure of the human PI3K p110γ that was used as a template in homology modeling. Ala805 in p110γ, which corresponds to residue Ser773 in p110α, was orientated to extend to the active site of p110γ. According to Amzel et al. [28], the residues lining the ATP-binding pockets of p110α and p110γ are highly conserved and have similar 3-D structures; however, the loop between residues 771 and 779 (IMSSAKRPL) in p110α adopted a conformation different from that in the corresponding loop in p110γ (VMASKKKPL, residues 803–811). It is further demonstrated that the orientation of Ser773 in p110α and Ala805 in p110γ is very different. Therefore, the docking model here should be more helpful in discovering new high activity p110α inhibitors.

According to Hayakawa [11], compound 58 is the most selective p110α inhibitor reported to date. In its binding mode, two H-bonds between residues Arg770 and Gln859 and compound 58 are particularly important for its selectivity. The Arg770 and Gln859 residues in p110α are different from the corresponding residues in other PI3K isoforms: Arg770 in p110α is Lys777 in p110β, Thr750 in p110δ, and Lys802 in p110γ. The Gln859 in p110α is Asp862 in p110β, Asn836 in p110δ, and Lys890 in p110γ. Therefore, these two H-bond interactions might account for the selectivity profile.

The binding mode of the least active compound (43) is shown in Fig. 5b. In its binding mode, two H-bonds between the N atom of imidazo[1,2-a]pyridine scaffold and Val 851, and the O atom of the sulfonyl group and Gln859 are observed, which have the same orientation as that of the most active compound (58). However, the nitro group in the *para* position on the phenyl ring, which is in the *meta* position on the phenyl ring of compound 58, cannot interact with Arg770 through H-bond contact, which can account for the low activity of compound 43. And, on the other hand, this result indicates that the H-bond interacting with Arg770 is critical to the activity of the ligand, which is consistent with the 3D-QSAR model. As seen from Fig. 5c, the morpholine ring scaffold is observed to interact with Val851 through H-bond contacts, which have the same orientation as that of the active compound 24. But compound 31 does not interact with Gln859 and Arg770 through H-bond contacts, which can account for the low activity of compound 31.

Conclusions

Different pharmacophore hypotheses of PI3K p110α inhibitors were developed using PHASE, and the alignment based on these pharmacophores was used as the input for the development of 3D-QSAR models. A five-point pharmacophore with three hydrogen bond acceptors (A), one lipophilic/hydrophobic group (H), and one aromatic ring (R) as pharmacophore features was associated with a 3D-QSAR model with good statistical significance and good predictive ability. Model 1 was significantly more accurate than other models, with $R^2 = 0.95$, $Q^2 = 0.88$ and $r_{pret}^2 = 0.95$. Furthermore, visualization of the 3D-QSAR model in the context of the molecules under study provided details of the relationship between structure and activity, and thus provides explicit indications for the design of better analogues. In addition, a molecular docking approach was applied to verify the pharmacophore and 3D-QSAR model developed in this paper. The docking results highlight the good complementarity between the 3D-QSAR model and the proposed ligand–protein docking model.

In summary, the ligand-based 3D-QSAR model presented in this study based on pharmacophore conformations could be very useful for lead optimization. Moreover, the 3D-QSAR and docking model developed can be used as a 3D-query for virtual database screening in order to find new PI3K p110α inhibitors.

References

1. Fruman DA, Meyers RE, Cantley LC (1998) Annu Rev Biochem 67:481–507
2. Katso R, Okkenhaug K, Ahmadi K, White S, Timms J, Waterfield MD (2001) Annu Rev Cell Dev Biol 17:615–675
3. Djordjevic S, Driscoll PC (2002) Trends Biochem Sci 27:426–432
4. Jiang BH, Liu LZ (2008) BBA Proteins Proteomics 1784:150–158
5. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ (2004) Science 304:554–571
6. Cantley LC, Neel BG (1999) Proc Natl Acad Sci USA 96:4240–4245
7. Graupera M, Guillermet-Guibert J, Foukas LC, Phng LK, Cain RJ, Salpekar A, Pearce W, Meek S, Millan J, Cutillas PR, Smith AJH, Ridley AJ, Ruhrberg C, Gerhardt H, Vanhaesebroeck B (2008) Nature 453:662–666
8. Hayakawa M, Kaizawa H, Moritomo H, Koizumi T, Ohishi T, Okada M, Ohta M, Tsukamoto S, Parker P, Workman P, Waterfield M (2006) Bioorg Med Chem 14:6847–6858
9. Hayakawa M, Kaizawa H, Moritomo H, Koizumi T, Ohishi T, Yamano M, Okada M, Ohta M, Tsukamoto S, Raynaud FI, Workman P, Waterfield MD, Parker P (2007) Bioorg Med Chem Lett 17:2438–2442
10. Hayakawa M, Kaizawa H, Kawaguchi K, Ishikawa N, Koizumi T, Ohishi T, Yamano M, Okada M, Ohta M, Tsukamoto S, Raynaud FI, Waterfield MD, Parker P, Workman P (2007) Bioorg Med Chem 15:403–412
11. 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) Bioorg Med Chem 15:5837–5844
12. Frederick R, Denny WA (2008) J Chem Inf Model 48:629–639
13. Huang CH, Mandelker D, Schmidt-Kittler O, Samuels Y, Velculescu VE, Kinzler KW, Vogelstein B, Gabelli SB, Amzel LM (2007) Science 318:1744–1748
14. Phase 1.0 (2005) User manual. Schrodinger, New York
15. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA (2006) J Comput Aided Mol Des 20:647–671
16. Dixon SL, Smondyrev AM, Rao SN (2006) Chem Biol Drug Des 67:370–372
17. Evans DA, Doman TN, Thorner DA, Bodkin MJ (2007) J Chem Inf Model 47:1248–1257
18. Narkhede SS, Degani MS (2007) QSAR Comb Sci 26:744–753
19. Tawari NR, Bag S, Degani MS (2008) J Mol Model 14:911–921
20. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) J Med Chem 47:1750–1759
21. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) J Med Chem 47:1739–1749
22. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R (2006) J Med Chem 49:534–553
23. Halgren TA (1996) J Comput Chem 17:520
24. MacroModel 2.0 (2006) User manual. Schrodinger, New York
25. Tropsha A (2005) In: Oprea TI (ed) Chemoinformatics in drug discovery. Wiley, Weinheim, pp 437–455
26. Knight ZA, Gonzalez B, Feldman ME, Zunder ER, Goldenberg DD, Williams O, Loewith R, Stokoe D, Balla A, Toth B, Balla T, Weiss WA, Williams RL, Shokat KM (2006) Cell 125:733–747
27. Zvelebil MJ, Waterfield MD, Shuttleworth SJ (2008) Arch Biochem Biophys 477:404–410
28. Amzel LM, Huang CH, Mandelker D, Lengauer C, Gabelli SB, Vogelstein B (2008) Nat Rev Cancer 8:665–669