



Computational modeling of novel inhibitors targeting the Akt pleckstrin homology domain

Lei Du-Cuny^a, Zuohe Song^b, Sylvester Moses^b, Garth Powis^a, Eugene A. Mash^c, Emmanuelle J. Meuillet^b, Shuxing Zhang^{a,*}

^a Department of Experimental Therapeutics—Unit 36, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, USA

^b Department of Molecular and Cellular Biology, and Nutritional Sciences, University of Arizona, Tucson, AZ, USA

^c Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, USA

ARTICLE INFO

Article history:

Received 24 June 2009

Revised 3 August 2009

Accepted 5 August 2009

Available online 19 August 2009

Keywords:

Akt inhibitors

Pleckstrin homology domain

Anticancer drugs

QSAR

In silico drug design

Molecular docking/scoring

Cell permeability

Metabolism

ABSTRACT

Computational modeling continues to play an important role in novel therapeutics discovery and development. In this study, we have investigated the use of in silico approaches to develop inhibitors of the pleckstrin homology (PH) domain of AKT (protein kinase B). Various docking/scoring schemes have been evaluated, and the best combination was selected to study the system. Using this strategy, two hits were identified and their binding behaviors were investigated. Robust and predictive QSAR models were built using the *k* nearest neighbor (*k*NN) method to study their cellular permeability. Based on our in silico results, long flexible aliphatic tails were proposed to improve the Caco-2 penetration without affecting the binding mode. The modifications enhanced the AKT inhibitory activity of the compounds in cell-based assays, and increased their activity as in vivo antitumor testing.

Published by Elsevier Ltd.

1. Introduction

Akt, also known as protein kinase B, is a serine/threonine kinase that is a critical component in the PI3K/Akt survival signaling pathway. It represents an exciting target for cancer therapy development due to its key roles in cell survival, proliferation, and apoptosis.^{1,2} The kinase consists of three conserved domains: an N-terminal pleckstrin homology (PH) domain, a central kinase catalytic domain, and a C-terminal extension domain with a hydrophobic motif.³ The activation of Akt is driven by membrane translocation initiated by the binding of its PH domain to the phosphoinositides produced by PI3K. Once it is correctly positioned in the cell membrane, Akt can be activated through the phosphorylation of its kinase domain by PDK1 at Thr³⁰⁸. Akt has been found overexpressed or activated in many human cancers,⁴ and hence

it is a validated target for cancer therapy. Many attempts have been made to develop small molecule inhibitors of Akt. A majority of these are ATP-competitive inhibitors targeting the kinase domain.^{5,6} However, due to the high degree of homology in the ATP-binding pocket among different serine/threonine kinases,⁷ achieving selectivity for these inhibitors remains a major problem.⁸ Hence, to overcome these drawbacks, we have adopted a novel strategy to develop inhibitors by targeting the PH domain of Akt.⁹ This is based on the fact that the sequence identity of different PH domains is usually less than 30%, which renders the possibility of developing selective agents for different targets.⁸ The feasibility of this strategy was also demonstrated by development of the Akt inhibitor D-3-deoxyphosphatidylinositol ether lipid (DPIEL, PX-316).¹⁰ A number of lipid-based derivatives were subsequently synthesized and identified as active phosphoinositol (PI analogs) inhibitors.^{11–13} However, these compounds have limited solubility and poor pharmacokinetics.⁸

The availability of high-resolution crystal structures of human Akt PH domains¹⁴ enabled us to conduct structure-based drug design of novel Akt inhibitors using molecular docking, which is widely used in lead identification and optimization.^{15,16} Using this approach the interactions between small molecules and the Akt PH domain can be modeled and their binding affinities can be pre-

Abbreviations: Akt, protein kinase B; PI3K, phosphoinositide 3-kinase; PH, pleckstrin homology; PDK1, 3-phosphoinositide-dependent protein kinase 1; DPIEL, deoxyphosphatidylinositol ether lipid; PI, phosphoinositol; kNN, *k* nearest neighbor; QSAR, quantitative structure–activity relationship; RMSD, root mean square distance; ADMET, Absorption, Distribution, Metabolism, Excretion, and Toxicity.

* Corresponding author. Tel.: +1 713 745 2958; fax: +1 713 794 5577.

E-mail address: shuzhang@mdanderson.org (S. Zhang).

dicted *in silico*. Molecular docking mainly consists of two components: the searching algorithm and the scoring function. Briefly, the docking program creates a simplified computational description for the receptor binding site, and then the translational, rotational, and conformational space of small organic molecules within that binding site is sampled. Finally the scoring function is used to estimate the binding free energy of each pose. Although various docking programs have been developed, there is no single software that gives accurate predictions on all ligand–target systems. Frequently different combinations of searching and scoring functions render completely different results.^{17,18} Therefore, it is critical to evaluate their applicability to the system of interest before employing a docking program. The evaluation can be performed by consideration of docking accuracy (e.g., RMSD between docked and crystal structures) and scoring accuracy (e.g., discrimination of active from random compounds). In this study, a series of evaluations of available docking tools, including FLEXX,¹⁹ GOLD,²⁰ and GLIDE,²¹ led to identification of the best combination of docking and scoring methods for optimization of Akt PH domain inhibitors.

In addition to binding affinity prediction, ADMET properties are also important in lead optimization.^{22,23} Among them, absorption and bioavailability are greatly affected by cell permeability. Several *in vitro* methods are available for permeability assays,^{24,25} of which the Caco-2 cell model is the most widely used. Various *in silico* models have also been developed for prediction of Caco-2 permeability. Hou et al.²⁶ used multiple linear regressions to derive computational models with 100 compounds. Nordqvist et al.²⁷ created a statistical model using 46 collected compounds. Ekins et al.²⁸ employed 3D-QSAR to analyze the Caco-2 permeability of a series of 28 inhibitors of rhinovirus replication. In our study, we found that appropriate permeability is crucial to the activity of Akt PH domain inhibitors.²⁹ To analyze the influence of chemical modification on cell permeability, we developed robust *in silico* models using variable selection *k* nearest neighbor (*k*NN) method.³⁰ Our models achieved accurate prediction and were used to guide our design of new compounds with enhanced cell permeability and activity.

Besides permeability prediction, the elucidation of metabolic sites could be significantly helpful in designing new compounds with a better pharmacokinetic profile, as bioavailability, activity, toxicity, distribution, and final elimination may depend on metabolic biotransformations. However, experimentally this is a task that requires many techniques and consumes a considerable amount of compounds. Herein, we employed METASITE³¹ to identify possible sites of metabolism in cytochrome-mediated reactions.³² The information can be used to detect positions that should be protected in order to avoid metabolic degradation.

Guided by these *in silico* predictions, lead compound Akt PH domain inhibitors were systematically modified. As a result, we have derived a better drug candidate that exhibits sub-micromolar inhibition in cell-based *in vitro* assays as well as low micromolar *in vivo* antitumor activity in a mouse xenograft model of pancreatic cancer.^{9,33}

2. Materials and methods

The whole workflow of developing novel inhibitors to target the Akt PH domain is demonstrated in Figure 1. Before the virtual screening for hit identification, three commercially available docking programs (FLEXX, GOLD, and GLIDE) were evaluated on this biological system. The best combination of the docking and scoring functions was employed to analyze the interaction between the protein and small molecules. The hits obtained from the virtual screening were validated via biological testing. Subsequently, lead optimization was performed based on combined approaches of molecular docking for binding prediction and QSAR modeling for ADME studies. Detailed methods applied in this process are described below in subsequent paragraphs.

2.1. Preparation of chemical databases for the evaluation of various docking approaches

In order to identify adequate docking and scoring functions to study the interactions between the Akt target and its inhibitors, a database was compiled for the evaluation of different combinations. The database contains 10 known Akt PH domain binders⁹ (Table 1) and 990 NCI molecules randomly chosen from the NCI diversity set³⁴ as negative decoys in our evaluation since none of the compounds showed activity in our experimental screening. The 3D structures of the known Akt PH domain inhibitors were prepared using MOE,³⁵ according to the following steps. The wash function in the software was employed to eliminate the chemical counter ions and to calculate the protonation state of ionizable groups of all 1000 ligands, at the physiological pH of 7.4. Hydrogen atoms were added and energy minimization was conducted using the MMFF94s force field and charges. During docking the ligand flexibility was considered and the programs automatically sample sufficient conformational space within the binding site using default parameters. As the starting point, the lowest energy conformation was utilized for docking.

2.2. Preparation of protein 3D structure for molecular docking

The protein crystal structure 1UNQ¹⁴ with high resolution (0.98 Å) was retrieved from the Protein Data Bank (PDB) and used for docking. In addition to 1UNQ there are several bound structure complexes available for Akt PH domains. However, the structural difference among them is very small. For instance, the RMSD for the backbone atoms of 1UNQ¹⁴ and 2UVM³⁶ was only 0.64 Å. We also investigate on the active site residues (defined as the residues 6.5 Å around Lys14, Glu17, Arg23, and Arg86) and found that the RMSD of them was only 0.58 Å. These results demonstrated that the two structures are very similar. No steric clashes were observed after merging the X-ray pose of the ligand of 2UVM³⁶ into the 1UNQ¹⁴ binding pocket. Hence, the binding site of 1UNQ¹⁴ is considered open enough to accommodate a variety of ligands, and thus can be used for the docking studies with a rigid binding pocket.

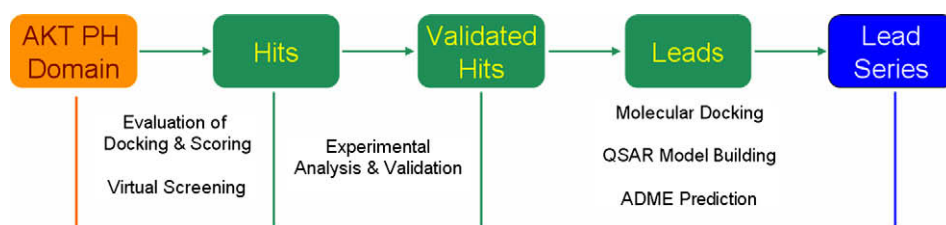
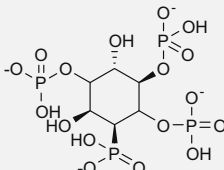
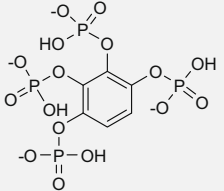
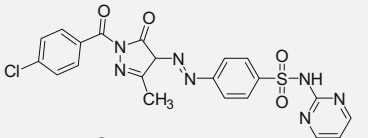
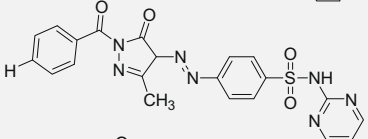
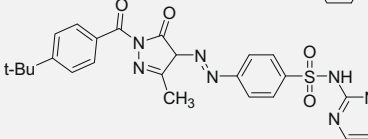
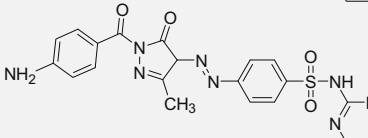
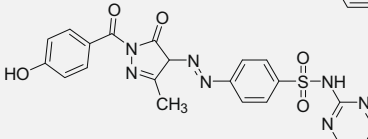
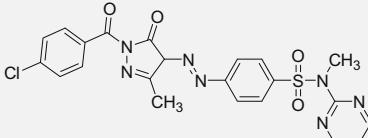
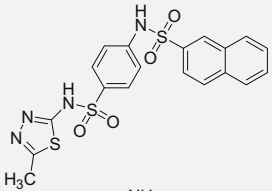
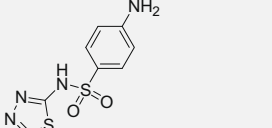


Figure 1. The whole workflow of developing novel inhibitors to target the Akt pleckstrin homology domain.

Table 1
Akt PH domain binders

Compound	Structure	Biological activity (μM)	References
1		1.50 ^a	60
2		0.08 ^b	52
3		0.37 ^a	9
4		3.66 ^a	9
5		1.37 ^a	9
6		0.51 ^a	9
7		1.35 ^a	9
8		1.62 ^a	9
9		25 ^c	33
10		0.45 ^a	33

The compound **1** is the ligand from the PDB structure 1UNQ,¹⁴ compound **2** from the PDB structure 2UVM.³⁶ The compounds **3–10** were previously reported as inhibitors targeting the PH domain of Akt.^{9,33}

^a K_D .

^b K_i .

^c IC_{50} .

SYBYL³⁷ was used to fix the protein with missing residues/atoms. All hydrogen atoms were loaded, and crystal waters and ligand were subjected to removal from the complex structure. PDB2PQR was utilized to calculate the pK_a values of protein residues to determine the residue charging status which was used in our docking.³⁸ In addition, the structure was slightly relaxed using the AMBER7 FF99 force field available in SYBYL. Based on structural analysis and literature reports,^{14,36,39} the binding pocket of the Akt PH domain was defined to include all residues within 6.5 Å around residues Lys14, Glu17, Arg23, and Arg86, which are essential for the protein–ligand interactions. These residues are involved in hydrogen bonding interactions and are responsible for the protein conformational change induced upon the binding of ligands.¹⁴

2.3. Docking methods and scoring functions

Three commercially available docking packages, FLEXX,⁴⁰ GOLD,⁴¹ and GLIDE⁴² were employed for docking studies using default parameters unless otherwise noted. No early termination was allowed in GOLD.⁴¹ The flexibility of a ligand was taken into account by GOLD⁴¹ via flipping the ring corners and hydrogen atoms of the protonated carboxylic acids. Internal hydrogen bonds of a ligand were included to restrict the flexibility. GLIDE⁴² was set to permit the conformational modification of amide bonds in order to consider docking flexibility. In all examinations, the protein was treated as a rigid body. Only the poses with the best scores were retained for further rescoring. For all ligands, docking solutions were rescored using the CScore module of SYBYL7.3³⁷ and GOLD Score in GOLD3.2.⁴¹ The CScore module comprises five scoring functions: ChemScore, D_Score, F_Score, G_Score, and PMF Score. All of these scoring functions were evaluated for the system.

2.4. Docking enrichment

Docking enrichment was evaluated to estimate the ability of different scoring functions to differentiate the known inhibitors from decoys. The enrichment was calculated using Eqs. 1 and 2, where Y specifies the percentage of true actives recovered (Eq. 1), and the number of compounds included in the database is represented by X (Eq. 2).⁴³

$$Y = \frac{L_t}{N_L} * 100 \quad (1)$$

$$X = L_t + D_t \quad (2)$$

L_t is the number of identified true positives (actives) in the subset; N_L is the total number of true actives; D_t is the number of decoys in the subset; $(L_t + D_t)$ is the total number of compounds in the subset.

2.5. Protein pharmacophore

In order to examine the pharmacophore match between the docked ligands and the protein, the protein binding site was characterized by the GRID force field.⁴⁴ GRID calculations were executed using a grid box enclosing the target with 1 Å beyond each dimension. During the GRID calculations, the GRID directive Move was set (Move = 1) to allow the flexibility of the protein side chains. The molecular interaction fields (MIFs) were computed to determine the energetically favorable binding sites for three probes: the hydrophobic (DRY), the amide nitrogen (N1, H-bond donor), and the carbonyl oxygen (O, H-bond acceptor). Local minima were identified for these three MIFs and they were selected from the GRID energy maps and used to define the protein pharmacophore features.⁴⁵

2.6. Cell permeability modeling

We have curated a dataset including 109 compounds with known experimentally determined Caco-2 permeability from various resources.^{26,46–48} The compounds with permeability are listed in the [Supplementary data](#). They are very diverse structures and consist of different type of drugs such as anticancer drugs, antibiotics, neurological agents, and so on. Using this dataset, QSAR models were developed to perform in silico prediction. MOE³⁵ was employed to generate 184 2D descriptors for the compounds. The descriptors were then normalized to avoid disproportional weighting. Eleven compounds (10%) were randomly selected as an external evaluation set, and the rest were divided into 50 training and test sets using the Sphere Exclusion (SE) algorithm as described previously.^{49,50} The dataset was treated as a collection of points in the MOE descriptor space. In brief, the SE method consisted of the following steps: (i) select randomly a compound; (ii) include it in the training set and construct a sphere around this compound; (iii) select compounds from this sphere, and include them alternatively into the test and training sets characterized by different probe sphere radii; (iv) exclude all compounds from within this sphere for further consideration. If no more compounds are left, stop. Otherwise, select a compound corresponding to the lowest d_{ij} and go to step (ii), where d_{ij} is the distances between the remaining compounds and sphere centers. The created multiple training and test sets were used to build robust and predictive models.

The kNN pattern recognition principle³⁰ and a variable selection procedure were applied to develop QSAR models for Caco-2 cell permeability predictions. Concisely, a subset of $nvar$ descriptors (number of selected variables) was selected randomly. Simulated annealing was used to sample the entire descriptor space to converge on the subset of the same size which afforded the highest value of q^2 . The descriptor subsets of different sizes were optimized using Leave-one-out (LOO) cross-validation procedure to obtain a variety of models with acceptable q^2 greater than a certain threshold (the default threshold value 0.5 was used). The training set models with acceptable q^2 were then validated on the test sets to select predictive models with R^2 exceeding 0.6. During modeling, default parameters were employed unless otherwise stated. Additionally, in order to exclude the possibility of chance correlation, Y-randomization experiments were conducted three times, as described previously,^{15,51} for the training sets but with randomized permeability values. Due to the high diversity of the dataset, stringent conditions were also employed to insure the reliability of the predictions by using a small arbitrary applicability domain ($Z = 0.1$), as published elsewhere.^{13,39}

2.7. Metabolism modeling

Ideal drug candidates should be metabolically stable. To this end, METASITE³¹ was employed to identify the potential metabolic sites of the compounds and to design analogs with improved metabolic properties. Briefly, the software uses two factors to analyze the metabolism probability of a site: the similarity between the CYP450 enzymes and the ligand, and the chemical reactivity of the substrate. The similarity evaluation of the CYP450 enzyme interaction site and the substrate is performed through the calculation of two sets of 'chemical fingerprints' descriptors: one for the CYP450 enzymes and the other one for the substrate. Additionally, the program considers the chemical reactivity of the substrate by taking into account of the activation energy required for production of reactive intermediates. The ranking for potential metabolic sites is based on the above similarity analysis and chemical reactivity.

2.8. Chemistry and biological evaluation

Synthesis of the compounds was conducted as described previously for compounds **3–8**⁹ and **9–14**,³³ respectively. K_d and K_i were measured using surface plasmon resonance (SPR) spectroscopy,³³ and IC_{50} s for cell inhibition of phospho-Akt in BxPC-3 pancreatic cancer cells were measured as previously described.⁹

3. Results and discussion

3.1. Evaluation of the accuracy of the pose prediction

1UNQ¹⁴ and 2UVM⁵² are Akt crystal structures available in the PDB,⁵³ co-crystallized with the native ligand inositol-(1,3,4,5)-tetrakisphosphate, and with benzene-1,2,3,4-tetraol tetrakisphosphate, respectively. These two complex structures are very similar with RMSD = 0.64 Å for backbone atom alignment and RMSD = 1.03 Å upon the all-atomic superimposition in the proteins. Therefore, the structure 1UNQ, which has the higher resolution, was used for docking. In order to keep the original binding mode of the ligand in the crystal structure, the X-ray pose of the ligand in 2UVM was merged into the 1UNQ binding pocket for comparing X-ray structures and docked poses, as frequently employed.^{17,18} The ability to recognize the native binding mode of a ligand to its target depends on the searching algorithm and scoring function of the docking approach. Searching algorithms are required to be able to sample the global minimum of the conformational space, and scoring functions are required to rank that pose as the best.

In order to find the appropriate combination of the scoring functions and searching algorithms, FLEXX, GOLD, and GLIDE were employed to dock the ligand crystal structures to their co-crystallized receptors. FLEXX is a flexible docking method that uses an incremental construction algorithm to place ligands into an active site and the placement of the ligand is scored on the basis of protein-ligand interactions including hydrogen bonds, salt bridges, metal contacts, and lipophilic interactions.⁴⁰ On the other hand, GOLD uses a genetic algorithm to explore the full range of ligand conformational flexibility.⁴¹ The mechanism for ligand placement is based on fitting points, which are created to consider the hydrogen bonding and hydrophobic interactions between the ligand and protein. A molecular mechanics-based scoring function is employed by GOLD to rank the docked poses. Different from these two methods, GLIDE approximates systematic searches of the conformational, orientational, and positional space of the docked ligand,⁴² where an initial rough positioning and scoring phase that dramatically narrows the search space is followed by torsionally flexible energy optimization on an OPLS-AA non-bonded potential grid. The best candidates are further refined by Monte Carlo sampling of pose conformations.

The differences between the X-ray and docked poses of the ligand are listed in Table 2. For both 1UNQ and 2UVM ligands, FLEXX and GOLD delivered excellent docking accuracy. The whole ligand was correctly docked except the slight deviation of the phosphate moieties (Fig. 2). This may be due to the fact that the phosphate group is ionized and thus all oxygen atoms are equivalent and barely differentiable to the docking programs. In comparison with

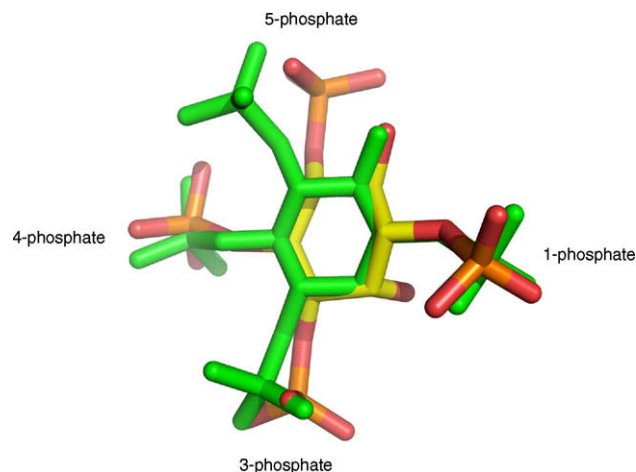


Figure 2. The X-ray crystal structure (green) and the GOLD (yellow) docked pose of the ligand (inositol-(1,3,4,5)-tetrakisphosphate) in AKT PH domain (using PDB structure 1UNQ).

FLEXX and GOLD docking results, GLIDE did not accurately reproduce the binding mode found in the crystal structures. Therefore, only the best poses obtained from FLEXX and GOLD were further rescored using different scoring functions.

3.2. Evaluation of the accuracy of scoring and ranking

The enrichment plots obtained with different scoring functions are displayed in Figure 3 for FLEXX and GOLD. As illustrated, the percentage of identified actual binders (Y-axis) was plotted against the number of compounds screened (X-axis). The ideal curve (black on the leftmost in Fig. 2) is demonstrated for the case where all true actives can be recovered in the top 10 hits, and the gray dashed line represents the random screening. GOLD fitness was found to be more robust as its enrichment curve is closer to the ideal one. When the GOLD pose was used for scoring with GOLD score, the first five real actives were identified within the top 30 hits and all 10 actives ranked within the top 90 compounds. However, the ranking by the PMF scoring function, based on the FLEXX or GOLD poses, is not as good as the others. For example, six real binders were ranked on the top 733, which is even worse than the random screening (Fig. 1B). It is notable that for the GOLD poses, scoring functions including strong hydrogen bonding terms, such as GOLD score, FlexX score, and ChemScore, provided good enrichment results. All of the 10 real active compounds could be found within the top 500 compounds. D_Score takes the charge and van der Waals interactions between the protein and ligand into account and it performed better than PMF (based on a set of Helmholtz free energies of interactions for protein–ligand atom pairs). We also found that hydrogen bonding and charge–charge interactions are particularly important for Akt PH domain and its ligands. This conclusion agrees with the fact that the Akt binding site is highly positively charged and possesses several residues, such as Lys14, Arg23, and Arg86, involved in the ligand binding.

Based on our evaluations, GOLD docking and GOLD scoring were found to be the best combination for this ligand–receptor system. The enrichment experiment demonstrated how efficiently the known binders were identified from the 1000 compound pool compared to a random screening. As indicated in Figure 3B, the GOLD docking and scoring (colored in deep blue) is the only combination for which all of the 10 inhibitors were ranked in the top 83 of 1000 compounds. This combination was thus employed in further structure-based lead optimization efforts.

Table 2
Molecular docking of ligands from 1UNQ and 2UVM to crystal structure 1UNQ

Ligands	RMSD of docked versus crystal structure (Å)		
	FLEXX	GOLD	GLIDE
1UNQ ligand	1.39	1.84	4.81
2UVM ligand	2.00	1.93	3.73

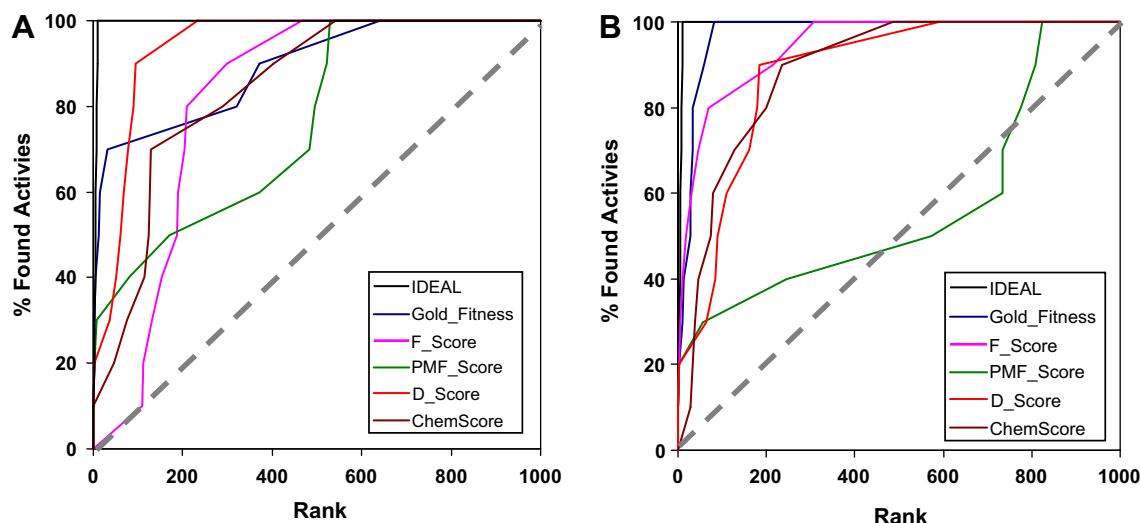


Figure 3. The enrichment of Akt PH domain inhibitors obtained from two docking programs and five scoring functions. (A) FLEX docking, (B) GOLD docking. Enrichments achieved by different scoring functions are denoted using legend in different colors. The gray dash line is for random screenings. X-axis reflects the ranking (position in the scoring list) of each compounds proposed by the five scoring functions from the obtained docking poses. Y-axis reflects the percentage of true actives recovered using the scoring functions.

3.3. QSAR model for the prediction of Caco-2 permeability

Our QSAR process produced 41 models with q^2 values for the training set greater than 0.5 and R^2 values for the test set compounds greater than 0.6. The best models were selected based on multiple modeling parameters, including the value of q^2 , R^2 , the number of selected variables as descriptors ($nvar$), the number of compounds in the test set (n) and several others, as described previously.^{15,51,54,55} Generally, a good model was expected to have high values of q^2 , R^2 , n , and a low value of $nvar$. For instance, as demonstrated in Figure 4, the best QSAR model was obtained using five descriptors to predict 51 compounds in the training set with $q^2 = 0.95$ (black filled circles), 47 compounds in the test set with $R^2 = 0.64$ (gray triangles) and 11 compounds in the external evalu-

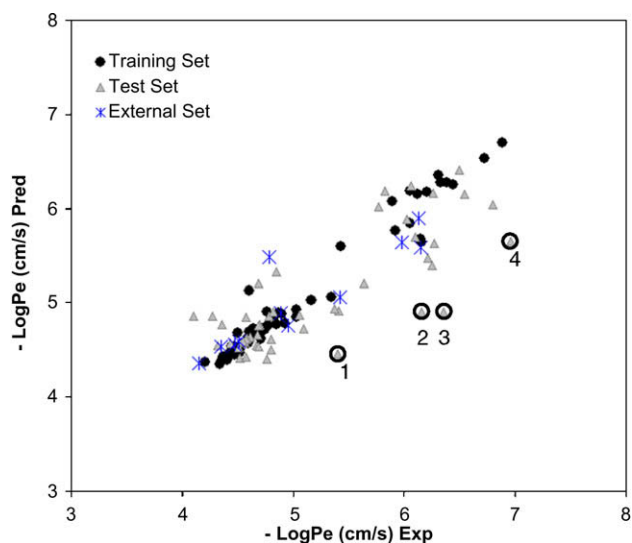


Figure 4. Experimental versus predicted Caco-2 permeability using one of the QSAR models. This model has 51 compounds (black dots) in the training set, 47 in the test set (gray triangles), and 11 in the external set (blue stars). Five descriptors were selected for the calculation. $q^2 = 0.95$, $R^2_{\text{test}} = 0.64$, and $R^2_{\text{external}} = 0.81$. The circled compounds are regarded as outliers. (1) artesunate; (2) methyl scopolamine; (3) pirenzepine; (4) olsalazine.

ation set with $R^2 = 0.81$ (blue stars). Y-randomization experiments indicated that no acceptable model was derived based on our requirement. Further inspection showed that the q^2 for all random models was always lower than 0.15 for the training sets (see Supplementary data Fig. 1), and thus no R^2 for test set compounds was calculated. This result excludes possibility of chance correlations.

The goal of our modeling is to design better Akt PH domain inhibitors. To this end, we evaluated the relationship between the Caco-2 permeability of the compounds and the descriptors selected in our models. The descriptors selected by the model were TPSA (Topological Polar Surface Area), opr_nring (Oprea Ring Count), GCUT_PEOp_0 (PEOE Charge GCUT (0/3)), lip_don (Lipinski Donor Count) and vdW_area (van der Waals surface area). We found that the molecular flexibility contributed positively to the Caco-2 permeability. For instance, the opr_nring is an indicator of molecular flexibility. Lower values of opr_nring mean higher molecular flexibility, and thus better Caco-2 permeability. However, other properties, including the total polar surface area, molecular charges and hydrogen donors, had negative contributions. One of the examples is cerftriaxone, which has higher values of TPSA, opr_nring and lip_don, but lower values of GCUT_PEOp_0 (higher charge) than lidocaine, and consequently its Caco-2 permeability is lower (1.30×10^{-7} cm/s vs 6.17×10^{-5} cm/s). Another example is ibuprofen, which has lower values of opr_nring, TPSA, and lip_don, but higher GCUT_PEOp_0 values, compared to doxorubicin, and thus it has higher Caco-2 permeability (5.25×10^{-5} cm/s vs 1.60×10^{-7} cm/s).

In addition to interpreting the relationship among descriptors and bioactivities, it is also critical to analyze the outliers with poor predictions by our models. Four outliers are highlighted in Figure 4. They are artesunate (labeled as 1), methyl scopolamine (labeled as 2), pirenzepine (labeled as 3) and olsalazine (labeled as 4). All of them were predicted to have much higher permeability than the experimental value. One of the possible reasons could be the wrong assignment of molecular charges. Charges have strong impacts on permeability of these compounds and their charged forms have poorer permeability than their neutral forms. To test our hypothesis, the structures of these compounds were modified to their ionized forms. Charges were also reassigned, and then descriptors were recalculated based on the new structures. Indeed our prediction gained dramatic improvement. For

instance, sodium artesunate had its Caco-2 permeability predicted as 2.51×10^{-5} cm/s and pirenzepine 1.20×10^{-6} cm/s, while their experimental values were 3.98×10^{-6} cm/s and 4.37×10^{-7} cm/s, respectively. The errors were less than one log unit after the structural modification. These *in silico* permeability models were employed in our Akt inhibitor lead optimization as described in later sections.

3.4. Pharmacophore characterization of the Akt PH domain

The active site of Akt PH domain was characterized with the GRID force field and virtualized using GView. The GRID isovolumes⁴⁴ are displayed in Figure 5 for the hydrophobic probe (energy cutoff = -0.8 kcal/mol) in orange and for the hydrogen bond acceptor (energy cut off = -10 kcal/mol) in blue. No isovolume was identified for the hydrogen bond donor probe when threshold was set to -10 kcal/mol. Our analysis also demonstrated that Tyr18 and Trp80 were specified as the preferential area for the interaction with a hydrophobic moiety. Lys14 and Arg25 are favorable places on the protein binding site to interact with hydrogen bond acceptor. This is in agreement with the crystal structure that showed Lys14 and Arg25 forming hydrogen bonding interactions with the phosphate on the third position of the inositol ring of IP4 (inositol(1,3,4,5)tetrakisphosphate).¹⁴ Hence, these relevant residues can be employed as protein pharmacophores to filter the docking poses of ligands.

3.5. Molecular docking of Akt hits to the PH domain

As described in our previous reports,⁹ we have identified 23 hits for Akt PH domain. Two of them, compounds **9** and **10**, were experimentally tested and confirmed to be active with IC_{50} of 20 μ M and 25 μ M, respectively.⁹ As GOLD docking/scoring was shown to be the best combination for the system, it was employed to study the binding of the compounds to the Akt PH domain. The docking results are shown in Figure 6, along with the crystal structure of

the original ligand, inositol(1,3,4,5)P₄ (IP4).¹⁴ The docking poses were selected based on the consideration of the docking scores and the population of the conformational clusters, together with molecular visualization of the interaction between the ligand and the protein pharmacophores. Based on all these criteria, the binding pose with the lowest estimated binding free energy was chosen for compound **10**, while the binding pose for the compound **9** with the lowest energy in the cluster with the second largest population was selected.

Compound **10** is a ‘needle fragment’⁵⁶ of compound **9**, but exhibits better inhibition of Akt phosphorylation³³ and its small size allows high potential for structural modification and optimization. According to our docking study, six hydrogen bonds were observed between the sulfonyl moiety of compound **10** as hydrogen acceptors and the Akt PH domain residue Arg23, Arg25, and Lys14 as hydrogen donors. In addition, the nitrogen atoms in the thiadiazolyl group were observed to strongly interact with residue Glu17 via hydrogen bonding (Fig. 6). This is consistent with the report by Carpten et al.³⁶ who demonstrated that Glu17 is found frequently mutated in human breast, colorectal, and ovarian cancers. The mutation of this residue seems to alter the electrostatic property of the pocket and may affect the activation of Akt. Our analysis of docked poses also showed that the 4-position of the phenyl ring pointed away from the binding site, and hence the modification at this position was predicted not to affect the binding. However, based on our QSAR modeling, substitution with a long aliphatic tail could enhance the permeability of the compound, and thus increase its cellular bioactivity (inhibition of Akt phosphorylation). This will be discussed below.

3.6. Modification on the aniline moiety to improve cell permeability

As described above, in order to improve the cellular permeability of compound **10**, three analogs were initially suggested based on our molecular docking and Caco-2 QSAR models, followed by synthesis and experimental evaluation.⁹ The docking results for these three compounds are illustrated in Figure 7, and their experimental values of binding affinity, Akt inhibition, and other properties are summarized in Table 3. The GOLD docking showed that the compound with a PEGylated tail (compound **12**) exhibited a completely different binding mode to compound **10**, and no binding poses with high population were obtained. Consistently, no bind-

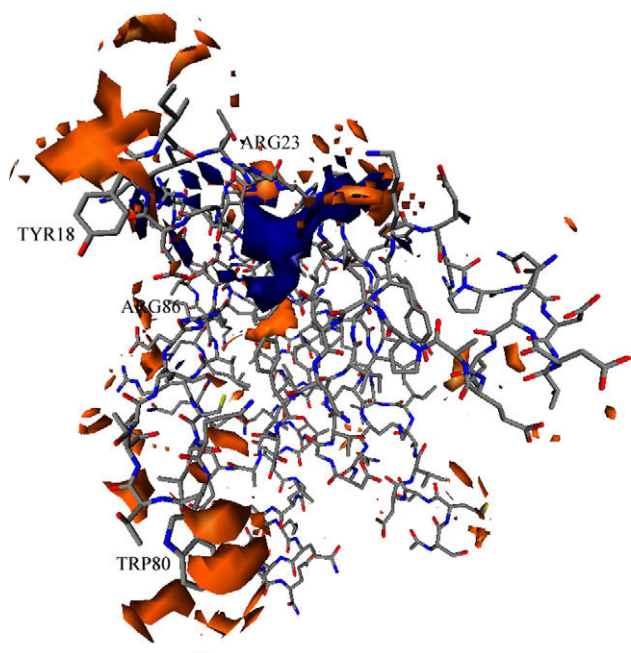


Figure 5. Isovolumes of Akt PH domain generated using GRID and displayed in orange and blue surface for the hydrophobic (-0.8 kcal/mol) and the H-bond acceptor (-10 kcal/mol) probes, respectively. The critical residues are labeled around the pharmacophores.

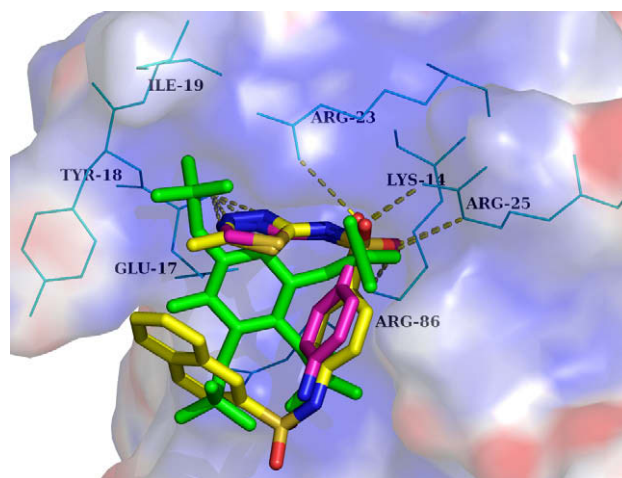


Figure 6. The docked poses with GOLD for compounds **9** (yellow) and **10** (magenta) together with the X-ray structure of inositol(1,3,4,5)tetrakisphosphate (green) of 1UNQ. The dash lines represent hydrogen bonds and the electrostatic surface is for the protein.

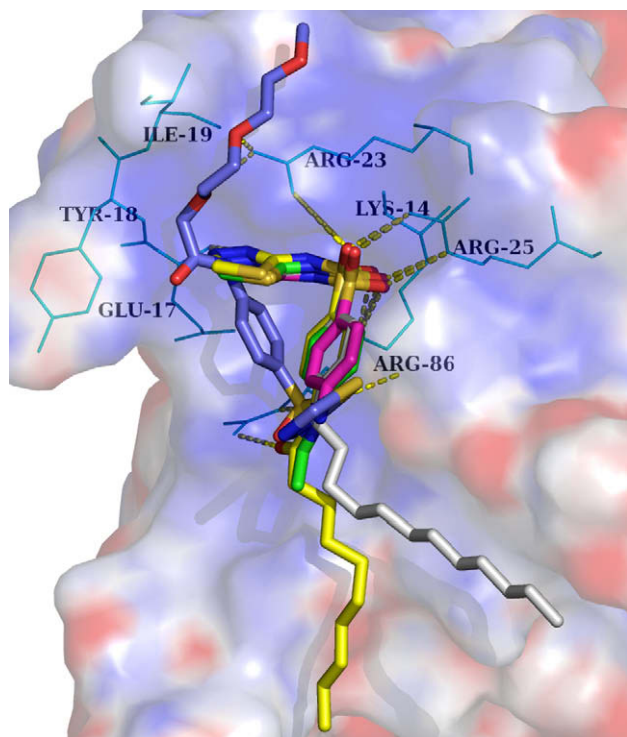


Figure 7. The docked poses with GOLD for compounds **10–14** in the PH domain of Akt. Compounds **10** (magenta), **11** (green), **12** (blue), **13** (yellow), **14** (gray). The dash lines represent hydrogen bonds and the electrostatic surface is for the protein.

ing was observed experimentally for this compound (Table 3). We also found that the compounds with acetyl (compound **11**) and decanoyl (compound **13**) tails have comparable binding and similar binding modes to compound **10** (Fig. 7). However, the decanoyl derivative was found to be the most active Akt PH domain inhibitor, while no activity was obtained for the acetyl analog. This partially could be due to the high Caco-2 cell permeability of the decanoyl derivative (4.80×10^{-6} cm/s) (Table 3), as predicted by our QSAR models. With the hydrophobic tail, the percentage of the TPSA of the compound is lower. According to the QSAR models aforementioned, Caco-2 cell permeability increases with lower percentage of TPSA.

Since an amide functional group can potentially be hydrolyzed in vivo, a non-amide analog **14**, supposedly more metabolically stable, was further proposed for synthesis and testing. It was predicted to have better Caco-2 permeability using our QSAR models. Although compound **14** was observed with lower binding affinity ($K_D = 40.8 \pm 2.5$ μ M) than compound **13** ($K_D = 19.6 \pm 4.9$ μ M), it exhibited higher Akt inhibition ($IC_{50} = 6.3 \pm 0.9$ μ M vs $IC_{50} = 10$ μ M) (Table 3).⁹ The exact mechanism is not yet clear, but our docking study revealed that the carbonyl moiety in the decanoyl tail of compound **13** formed hydrogen bonds with Arg86. This might be one of the reasons of its stronger binding (Fig. 7). However, the decanoyl tail of compound **13** might be cleaved in the cell through the hydrolysis of the amide moiety. Additionally, the hydrophobic dodecyl tail of compound **14** is more flexible and less constrained; therefore it could enhance the binding by interacting with the membrane, as some models have suggested.^{57,58} This may potentially increase its concentration around the membrane where PIP3 binding and AKT activation occurs.

Table 3
Experimental binding affinity and Akt inhibition values measured for compounds **10–14** together with predicted Caco-2 permeability (Pe_{Caco2}) using our QSAR models and lipophilicity predicted using MOE

Compound	Structure	Pe_{Caco2} (10^{-6} cm/s)	Log P	K_D (μ M)	IC_{50} (μ M)
10		0.38	0.90	0.45	20
11		0.39	0.86	NB	>50
12		0.58	−0.01	NB	>50
13		3.43	4.43	19.6	10
14		4.80	6.76	40.8	6.3

NB denotes no binding.

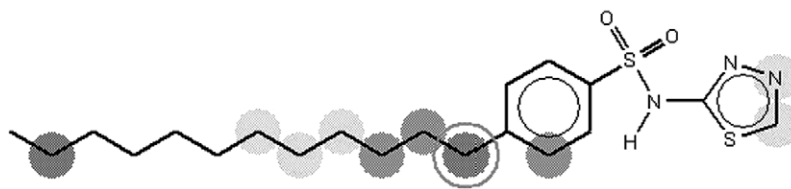


Figure 8. Plot of MetaSite prediction for the metabolism of compound **14**. The groups metabolic are marked in shaded gray circles. The darker the color, the higher the probability of metabolism is to occur.

3.7. In silico prediction of the metabolism of our compounds

In addition to molecular docking and QSAR modeling, an analysis of the potential metabolism of our compounds was also conducted. The cytochrome-mediated metabolically labile positions of these molecules were studied using the program METASITE.³¹ Default parameters and all CYP models in the software (CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4) were applied. In the case of compound **10**, the fifth carbon atom of the 1,3,4-thiadiazole ring has the highest potential to be metabolized according to all CYP models in METASITE. By adding the dodecyl tail, the potential for metabolism on this position was minimized, although some carbon atoms in the dodecyl tail can be hydroxylated. (Fig. 8) The experimental analysis of the metabolism of the compounds will be published in forthcoming papers. To date, in addition to its high cellular activity, in vivo experiments have shown that compound **14** has significant antitumor activity with cessation of tumor growth.⁹ A single dose caused significant inhibition of tumor Akt measured as phospho-Ser⁴⁷³-Akt with up to 70% inhibition at 6 h and 50% inhibition at 12 h, as published elsewhere.⁹

4. Conclusion

This study was focused on the development of novel Akt PH domain inhibitors. Molecular docking and in silico ADMET studies were employed to guide chemical design and lead optimization. As there is no single docking/scoring program which can work universally on all ligand–receptor systems, a critical assessment of various combinations of docking and scoring methods for our target system was conducted. As a result, the GOLD docking and scoring function were found to be the best combination to analyze the interactions between the inhibitors and the Akt PH domain. According to the docking results, an aliphatic chain was proposed to enhance the interactions but maintain the binding mode.

Based on the QSAR study and metabolism predictions, the modified hit with a dodecyl tail had the highest Caco-2 permeability within this series of compounds, and thus enhanced cellular uptake. In addition, the thiadiazole warhead involved in binding was predicted to be metabolically stable via cytochrome-mediated mechanisms. The optimized inhibitor was experimentally validated with significant in vitro and in vivo antitumor activity. In order to unambiguously identify the drug–receptor binding and further guide our design of better inhibitors, crystallographic studies are in progress. Moreover, the discovery of novel inhibitor scaffolds is also underway with high-throughput docking and QSAR-based virtual screening.

We believe that development of novel Akt PH domain inhibitors for targeted cancer therapy is promising and shall result in more selective and specific anticancer agents. We also suggest that our recent successes^{9,16,33,51,55,59} in identifying novel active anticancer compounds by a combined application of molecular docking, rigorous QSAR modeling, and ADMET prediction positions our comprehensive design approach as a general methodology for computer-aided cancer therapeutics development.

Acknowledgment

This work was supported by the MD Anderson Cancer Center Faculty Startup to S.Z. and the NIH/NCI grants RO1 CA061015 and P30 CA23074 to G.P. We also thank Dr. A. Tropsha for providing the kNN package.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2009.08.022.

References and notes

- Hennessy, B. T.; Smith, D. L.; Ram, P. T.; Lu, Y.; Mills, G. B. *Nat. Rev. Drug Disc.* **2005**, *4*, 988.
- Manning, B. D.; Cantley, L. C. *Cell* **2007**, *129*, 1261.
- Cheng Jin, Q.; Lindsley Craig, W.; Cheng George, Z.; Yang, H.; Nicosia Santo, V. *Oncogene* **2005**, *24*, 7482.
- Bellacosa, A.; Testa, J. R.; Moore, R.; Larue, L. *Cancer Biol. Ther.* **2004**, *3*, 268.
- Breitenlechner, C. B.; Wegge, T.; Berillon, L.; Graul, K.; Marzenell, K.; Friebe, W.-G.; Thomas, U.; Schumacher, R.; Huber, R.; Engh, R. A.; Masjost, B. *J. Med. Chem.* **2004**, *47*, 1375.
- Reuveni, H.; Livnah, N.; Geiger, T.; Klein, S.; Ohne, O.; Cohen, I.; Benhar, M.; Gellerman, G.; Levitzki, A. *Biochemistry* **2002**, *41*, 10304.
- Yang, J.; Cron, P.; Thompson, V.; Good Valerie, M.; Hess, D.; Hemmings Brian, A.; Barford, D. *Mol. Cell* **2002**, *9*, 1227.
- Kumar, C. C.; Madison, V. *Oncogene* **2005**, *24*, 7493.
- Mahadevan, D.; Powis, G.; Mash, E. A.; George, B.; Gokhale, V. M.; Zhang, S.; Shakalya, K.; Du-Cuny, L.; Berggren, M.; Ali, M. A.; Jana, U.; Ihle, N.; Moses, S.; Franklin, C.; Narayan, S.; Shirahatti, N.; Meuillet, E. *J. Mol. Cancer Ther.* **2008**, *7*, 2621.
- Powis, G.; Aksoy, I. A.; Melder, D. C.; Aksoy, S.; Eichinger, H.; Fauq, A. H.; Kozikowski, A. P. *Cancer Chemother. Pharmacol.* **1991**, *29*, 95.
- Meuillet, E. J.; Mahadevan, D.; Vankayalapati, H.; Berggren, M.; Williams, R.; Coon, A.; Kozikowski, A. P.; Powis, G. *Mol. Cancer Ther.* **2003**, *2*, 389.
- Meuillet, E.; Stratton, S.; Cherukuri, D. P.; Goulet, A.-C.; Kagey, J.; Porterfield, B.; Nelson, M. A. *J. Cell. Biochem.* **2004**, *91*, 443.
- Kozikowski, A. P.; Sun, H.; Brognard, J.; Dennis, P. A. *J. Am. Chem. Soc.* **2003**, *125*, 1144.
- Milburn, C. C.; Deak, M.; Kelly, S. M.; Price, N. C.; Alessi, D. R.; van Aalten, D. M. F. *Biochem. J.* **2003**, *375*, 531.
- Zhang, S.; Golbraikh, A.; Tropsha, A. *J. Med. Chem.* **2006**, *49*, 2713.
- Zhang, S.; Kumar, K.; Jiang, X.; Wallqvist, A.; Reifman, J. *BMC Bioinf.* **2008**, *9*, 126.
- Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759.
- Warren Gregory, L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert Millard, H.; Lindvall, M.; Nevins, N.; Semus Simon, F.; Senger, S.; Tedesco, G.; Wall Ian, D.; Woolven James, M.; Peishoff Catherine, E.; Head Martha, S. *J. Med. Chem.* **2006**, *49*, 5912.
- FLEXx [1.20.1], BioSolveIT GmbH: Sankt Augustin, Germany, 2007.
- GOLD [3.2], CCDC: Cambridge, UK, 2007.
- GLIDE [4.5], Schrödinger: Portland, OR, 2007.
- Du-Cuny, L.; Huwyler, J.; Wiese, M.; Kansy, M. *Eur. J. Med. Chem.* **2008**, *43*, 501.
- Stoner, C. L.; Troutman, M. D.; Laverty, C. E. *Cancer Drug Des. Disc.* **2008**, *131*.
- Artursson, P. *Crit. Rev. Ther. Drug Carrier Syst.* **1991**, *8*, 305.
- Kansy, M.; Senner, F.; Gubernator, K. *J. Med. Chem.* **1998**, *41*, 1007.
- Hou, T. J.; Zhang, W.; Xia, K.; Qiao, X. B.; Xu, X. J. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1585.
- Nordqvist, A.; Nilsson, J.; Lindmark, T.; Eriksson, A.; Garberg, P.; Kihlen, M. *QSAR Comb. Sci.* **2004**, *23*, 303.
- Ekins, S.; Durst, G. L.; Stratford, R. E.; Thorner, D. A.; Lewis, R.; Loncharich, R. J.; Wikel, J. H. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1578.
- Sun, D.; Yu, L. X.; Hussain, M. A.; Wall, D. A.; Smith, R. L.; Amidon, G. L. *Curr. Opin. Drug Disc. Dev.* **2004**, *7*, 75.
- Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemical Analysis, Vol. 82: Chemometrics* 1986; p 332.

31. METASITE [2.7.5], Molecular Discovery Ltd: Pinner, Middlesex, UK, 2007.
32. Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. *J. Med. Chem.* **2005**, *48*, 6970.
33. Moses, S.; Ali, M. A.; Zuohe, S.; Du-Cuny, L.; Zhou, L. L.; Lemos, R.; Ihle, N.; Sillman, G.; Zhang, S.; Mash, E. A.; Powis, G.; Meuillet, E. *Cancer Res.* **2009**, *69*, 5073.
34. http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html.
35. MOE [2006.08], Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2006.
36. Carpten, J. D.; Faber, A. L.; Horn, C.; Donoho, G. P.; Briggs, S. L.; Robbins, C. M.; Hostetter, G.; Boguslawski, S.; Moses, T. Y.; Savage, S.; Uhlik, M.; Lin, A.; Du, J.; Qian, Y.-W.; Zeckner, D. J.; Tucker-Kellogg, G.; Touchman, J.; Patel, K.; Mousses, S.; Bittner, M.; Schevitz, R.; Lai, M.-H. T.; Blanchard, K. L.; Thomas, J. E. *Nature* **2007**, *448*, 439.
37. SYBYL [7.3], Tripos: St. Louis, MO, USA, 2007.
38. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32*, W665.
39. Thomas, C. C.; Deak, M.; Alessi, D. R.; van Aalten, D. M. F. *Curr. Biol.* **2002**, *12*, 1256.
40. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
41. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
42. Friesner Richard, A.; Banks Jay, L.; Murphy Robert, B.; Halgren Thomas, A.; Klicic Jasna, J.; Mainz Daniel, T.; Repasky Matthew, P.; Knoll Eric, H.; Shelley, M.; Perry Jason, K.; Shaw David, E.; Francis, P.; Shenkin Peter, S. *J. Med. Chem.* **2004**, *47*, 1739.
43. Jain, A. N. *J. Comput. Aided Mol. Des.* **2008**, *22*, 201.
44. GRID [22b], Molecular Discovery Ltd: Oxford, UK, 2008.
45. Pirard, B.; Brendel, J.; Peukert, S. *J. Chem. Inf. Model.* **2005**, *45*, 477.
46. Gleevec FDA Information. <http://www.fda.gov/CDER/DRUG/infopage/gleevec/default.htm>.
47. Kamath, A. V.; Wang, J.; Lee, F. Y.; Marathe, P. H. *Cancer Chemother. Pharmacol.* **2008**, *61*, 365.
48. Rakhit, A.; Pantze, M. P.; Fettner, S.; Jones, H. M.; Charoin, J.-E.; Riek, M.; Lum, B. L.; Hamilton, M. *Eur. J. Clin. Pharmacol.* **2008**, *64*, 31.
49. Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.-D.; Lee, K.-H.; Tropsha, A. *J. Comput. Aided Mol. Des.* **2003**, *17*, 241.
50. Zheng, W.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185.
51. Zhang, S.; Wei, L.; Bastow, K.; Zheng, W.; Brossi, A.; Lee, K.-H.; Tropsha, A. *J. Comput. Aided Mol. Des.* **2007**, *21*, 97.
52. Mills, S. J.; Komander, D.; Trusselle, M. N.; Safrany, S. T.; van Aalten, D. M. F.; Potter, B. V. L. *ACS Chem. Biol.* **2007**, *2*, 242.
53. Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. *Nucleic Acids Res.* **2007**, *35*, D301.
54. Oloff, S.; Zhang, S.; Sukumar, N.; Breneman, C.; Tropsha, A. *J. Chem. Inf. Model.* **2006**, *46*, 844.
55. Zhang, S.; Golbraikh, A.; Oloff, S.; Kohn, H.; Tropsha, A. *J. Chem. Inf. Model.* **2006**, *46*, 1984.
56. Boehm, H.-J.; Boehringer, M.; Bur, D.; Gmuender, H.; Huber, W.; Klaus, W.; Kostrewa, D.; Kuehne, H.; Luebbbers, T.; Meunier-Keller, N.; Mueller, F. *J. Med. Chem.* **2000**, *43*, 2664.
57. Lemmon, M. A.; Ferguson, K. M.; Abrams, C. S. *FEBS Lett.* **2002**, *513*, 71.
58. Lemmon, M. A. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 99.
59. Zhang, S.; Du-Cuny, L. *Int. J. Bioinform. Res. Appl.* **2009**, *5*, 269.
60. DiNitto, J. P.; Lambright, D. G. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2006**, *1761*, 850.