



Original article

LEADOPT: An automatic tool for structure-based lead optimization, and its application in structural optimizations of VEGFR2 and SYK inhibitors

Guo-Bo Li ^{a,1}, Sen Ji ^{a,1}, Ling-Ling Yang ^{a,b,1}, Rong-Jie Zhang ^a, Kai Chen ^a, Lei Zhong ^a, Shuang Ma ^a, Sheng-Yong Yang ^{a,*}^a State Key Laboratory of Biotherapy/Collaborative Innovation Center of Biotherapy, West China Hospital, West China Medical School, Sichuan University, Sichuan 610041, China^b College of Chemical Engineering, Sichuan University, Sichuan 610041, China

ARTICLE INFO

Article history:

Received 7 September 2014

Received in revised form

3 January 2015

Accepted 12 February 2015

Available online 14 February 2015

Keywords:

Structure-based drug design (SBDD)

Lead optimization

Fragment growing

Fragment replacing

Ligand efficiency

ABSTRACT

Lead optimization is one of the key steps in drug discovery, and currently it is carried out mostly based on experiences of medicinal chemists, which often suffers from low efficiency. *In silico* methods are thought to be useful in improving the efficiency of lead optimization. Here we describe a new *in silico* automatic tool for structure-based lead optimization, termed LEADOPT. The structural modifications in LEADOPT mainly include two operations: fragment growing and fragment replacing, which are restricted to carry out in the active pocket of target protein with the core scaffold structure of ligand kept unchanged. The bioactivity of the newly generated molecules is estimated by ligand efficiency rather than a commonly used scoring function. Twelve important pharmacokinetic and toxic properties are evaluated using SCADMET, a program for the prediction of pharmacokinetic and toxic properties. LEADOPT was first evaluated using two retrospective cases, in which it showed a very good performance. LEADOPT was then applied to the structural optimizations of the VEGFR2 inhibitor, sorafenib, and the SYK inhibitor, R406. Though just several compounds were synthesized, we have obtained some compounds that are more potent than sorafenib and R406 in enzymatic and functional assays. All of these have validated, at least to some extent, the effectiveness of LEADOPT.

© 2015 Published by Elsevier Masson SAS.

1. Introduction

Discovery of drug candidates is a bottleneck in drug research and development (R&D). The whole procedure for the discovery of drug candidates often includes two phases: lead discovery and lead optimization [1]. Currently the lead discovery is no longer a big problem or at least not a key challenge due to that many new technologies for the lead discovery have been well established and applied, such as high-throughput screening, biophysical screening,

combinatorial chemistry, as well as virtual screening [2–6]. However, at present the lead optimization is carried out mostly based on experiences of medicinal chemists, which often suffers from low efficiency. The lead optimization has now become a key challenge in not only the discovery of drug candidates but also the whole process of drug R&D.

Computational methods are often thought to be helpful in improving the efficiency of lead optimization [7–10]. Thus, some *in silico* methods have been established and applied in the lead optimization. Among these methods, three-dimensional quantitative structure activity relationship (3D-QSAR) methods, for example, comparative molecular field analysis (CoMFA) [11] and comparative molecular similarity indices analysis (CoMSIA) [12], are the most widely used ones. Numerous studies have demonstrated that 3D-QSAR methods are very useful in lead optimization [13–16]. Nevertheless, the current 3D-QSAR methods as well as other commonly used *in silico* methods for lead optimization are often just based on small-molecule ligands, implying ignoring

Abbreviations: R&D, research and development; SBDD, structure-based drug design; LE, ligand efficiency; 3D-QSAR, three-dimensional quantitative structure activity relationship; ADMET, absorption, distribution, metabolism, excretion, and toxicity; PDB, Protein Data Bank; JAK1, Janus kinase 1; MPS1, monopolar spindle 1; VEGFR2, vascular endothelial growth factor receptor 2; SYK, spleen tyrosine kinase.

* Corresponding author.

E-mail address: yangsy@scu.edu.cn (S.-Y. Yang).¹ These authors contributed equally to this work.

structural information of target protein. This situation may lead to the theoretically optimized molecules having no activity due to bumps between molecules and target protein atoms. Structure-based lead optimization approaches are expected to be able to overcome these shortcomings, in which the structure of ligand is modified to enhance the binding affinity based on the known interaction mode between target protein and ligand [17–21]. Although these approaches have been applied in lead optimization for a long time, they are often used manually by medicinal chemists, which still suffer inevitably from low efficiency. We also noticed that very recently a commercial module for structure-based ligand optimization has been issued in MOE by Chemical Computing Group Inc. However, its implementation details are not known, and it likely does not include the evaluation of pharmacokinetic and toxic properties for the generated molecules, which is also important for lead optimization.

Here we present an automatic tool for structure-based lead optimization, termed LEADOPT. LEADOPT has the following characteristics that make it a practical tool in lead optimization. First, a large number of derivatives of lead compound that have the same binding mode with the active pocket of target protein can be created. These compounds can be accommodated in the active pocket without any atom bump with the target protein. Second, LEADOPT builds up new molecules using an efficient fragment-based strategy. Such a strategy can help avoid producing unreasonable molecular structures and to some extent keep the synthetic accessibility of the derived compounds. Third, ligand efficiency (LE) rather than scoring function is used as a measure to sort the newly generated molecules; LE has been considered as an effective strategy to help narrow focus to lead compounds with optimal combinations of physicochemical properties and pharmacological properties [22,23]. Fourth, a number of pharmacokinetic and toxic properties of the generated molecules are evaluated, which makes the lead candidates pharmacologically acceptable. We shall in the following describe the details for the algorithm of LEADOPT. Then, LEADOPT will be evaluated with two retrospective cases. Finally, it will be applied to the structural optimizations of sorafenib, an inhibitor of vascular endothelial growth factor receptor 2 (VEGFR2), and R406, an inhibitor of spleen tyrosine kinase (SYK).

2. Methods and materials

2.1. Fragment library construction

Since LEADOPT adopts a fragment-based approach for structural modifications to derive new molecular structures, a good fragment library is quite important for the quality of derived molecules. To establish such a fragment library, we first collected a total of 17,858 drug or drug-like molecules from CMC database, ChEMBL database [24], and DrugBank database [25]. Then, an in-house program written by us in C/C++ programming language was used to automatically build the fragment library. The process is briefly described as follows.

- (1) Each input molecule was detected for whether it contains any of the fifteen types of chemical bonds defined in Table 1; these chemical bond types were derived from common chemical reactions, which is for making the generated fragments easy to be used in later chemical synthesis [26].
- (2) Molecules that contain at least one chemical bond defined in Table 1 were cleaved into fragments at the defined chemical bonds. If a terminal fragment cleaved is only a small group, such as methyl, ethyl, propyl, or butyl, the cleaving operation will not be performed anymore.

Table 1
Bond cleavage types used for producing fragments from molecules.

Rule#	Cleavage type	Example	Cleavage marks ^a
1	Amide		
2	Ester		
3	Urea		
4	Sulphonamide		
5	Acyclic amine		
6	Cyclic amine–aliphatic carbon		
7	Lactam nitrogen–aliphatic carbon		
8	Acyclic ether		
9	Acyclic sulfide		
10	Aromatic ether		
11	Cyclic amine–aromatic carbon		
12	Aromatic nitrogen–aliphatic carbon		
13	Aromatic nitrogen–aromatic carbon		
14	Aromatic carbon–aromatic carbon		
15	Olefin		

^a *- nucleophilic mark; *+, electrophilic mark; *, aromatic mark; **, olefin mark.

- (3) Four types of connection marks, including electrophilic mark (*+), nucleophilic mark (*-), aromatic mark (*), and olefin mark (**), were used to label the cleaved terminal(s) for each fragment. Some examples are given in Supporting Information Fig. S1.
- (4) Any duplicate or fragments containing more than 14 heavy atoms or more than 3 terminal marks were removed.

We finally obtained a fragment library consisting of 6877 unique fragments, including 5106 fragments with a single terminal mark, 1587 fragments with two terminal marks, and 184 fragments with three terminal marks. The defined terminal marks will be used in the structural modification process in LEADOPT, which will be described in the next section.

2.2. Algorithms of LEADOPT

Fig. 1 schematically depicts the workflow of LEADOPT. LEADOPT is launched through inputting a ligand–receptor complex structure, which may be prepared from either experimental X-ray crystal structure or molecule docking. In LEADOPT, the ligand and protein receptor, which are hydrogenated, correspond to two separated input files. For the protein receptor, some important molecules such as metal atoms and water molecules can be kept in the binding pocket. Subsequently, LEADOPT executes the following four steps: binding site analysis, structural modification, evaluation of binding affinity and ligand efficiency, and evaluation of pharmacokinetic and toxic properties.

2.2.1. Binding site analysis

In LEADOPT, new molecules are built up within the constraints of the protein binding site. Thus the size and shape of the binding site are critical for later structural modifications. For a given protein–ligand complex, LEADOPT will define a rectangular box covering the ligand and all residues around the ligand in 6 Å (see Supporting Information Fig. S2). The rectangular box was further divided into small equal-sized square grids according to its vertex coordinates. The grid spacing was set as 0.5 Å. A hydrogen atom was used as a probe on each grid to check whether the grid is accessible. If the probe bumps with the protein, the grid will be labeled as ‘E’; if not, the grid will be labeled as ‘V’. The grid information will be used in the structural modification process.

2.2.2. Structural modification

Two main types of structural modification strategies, including ‘fragment growing’ and ‘fragment replacing’, were adopted to construct new molecules. In the structural modification process, five bond connection rules were used according to the fragment terminal marks (see Table 2). For example, the fragment terminal with a nucleophilic mark can be linked to the fragment terminal with an electrophilic mark or aromatic mark (Rule 1 and 3, Table 2),

Table 2
Five bond connection types.

Rule#	Connection type	Example	Connected bond
1	Nucleophilic–electrophilic		
2	Aromatic carbon–aromatic carbon		
3	Electrophilic–aromatic carbon		
4	Nucleophilic–aromatic carbon		
5	Olefin carbon–olefin carbon		

while it is not allowed to connect to the fragment terminal with other terminal marks such as nucleophilic mark. These defined connection rules can help avoid the generation of unreasonable molecules and to some extent improve the synthetic accessibility.

The procedure of ‘fragment growing’ is briefly described as follows (see Fig. 2A): (i) detecting whether there is certain vacancy around the reference compound according to the grid information; if no, the ‘fragment growing’ will stop; (ii) identifying possible connection points and determining their corresponding connection marks as that used for labeling the fragment terminals; (iii) connecting each fragment in the fragment library onto one of the connection points according to the five defined connection rules (see Table 2); (iv) adjusting the orientation for each fragment and checking whether the fragment bumps with the protein; (v) repeating step iii and iv for other possible connection points.

The ‘fragment replacing’ procedure involves the following steps (see Fig. 2B): (i) identifying the core scaffold for the reference compound; (ii) determining the connection bonds between the

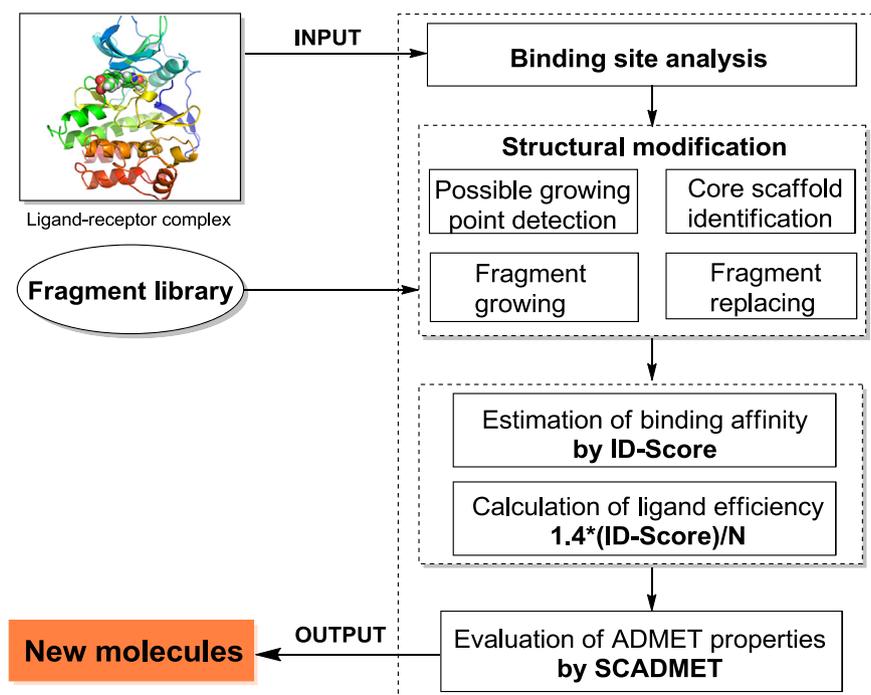


Fig. 1. The workflow of LEADOPT.

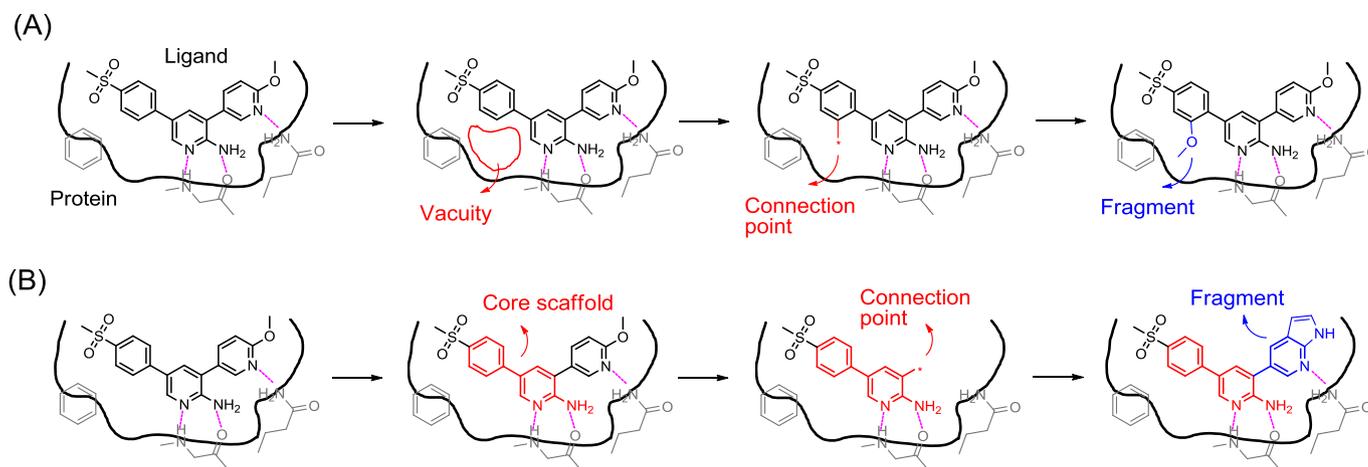


Fig. 2. Schematic illustrations for the two main types of structural modifications: (A) 'fragment growing' and (B) 'fragment replacing'.

core scaffold and other parts of the reference compound; (iii) selecting one connection bond, cleaving it, and labeling the connection point for the core scaffold; (iv) connecting each fragment in the fragment library onto one of the connection points according to the connection rules; (v) adjusting the orientation for each fragment and checking whether the fragment bumps with the protein; (vi) recording the top ranking fragments for the connection point; (vii) repeating step iii to vi for all the connection points; (viii) assembling all the top ranking fragments onto the corresponding points on the core scaffold. An example for 'fragment replacing' is given in [Supporting Information Fig. S3](#).

Besides, we introduced another structural modification strategy termed 'fragment cyclization' to create more structurally diverse molecules. This strategy relies on the 'fragment growing' and 'fragment replacing'. During a 'fragment growing' or 'fragment replacing' process, if a fragment can approach the defined core scaffold within a certain distance, it will be cyclized with the core scaffold to form a new scaffold. There have two situations for the 'fragment cyclization' operation in LEADOPT (see [Supporting Information Fig. S4](#)). First, when the fragment collides with the core scaffold by a pair of hydrogen atoms, these two hydrogen atoms will be deleted and a methyl group will be used to bridge the fragment and core scaffold. Second, when the fragment collides with the core scaffold by a pair of heavy atoms, one of the heavy atoms will be deleted and another heavy atom will be connected directly. Such a structural modification strategy could create new chemical scaffolds and generate structurally diverse derivatives. All the newly generated molecules by the modification strategies described above will be submitted to the following evaluation.

2.2.3. Evaluation of binding affinity and ligand efficiency

The protein-ligand binding affinity is assessed by using the ID-Score method. ID-Score was derived from the established comprehensive set of descriptors related to protein-ligand interactions. These descriptors cover nine categories, including hydrogen-bonding interaction, van der Waals interaction, π -system interaction, electrostatic interaction, metal-ligand bonding interaction, entropic loss effect, desolvation effect, shape matching, and surface property matching. A total of 2278 crystal protein-ligand complexes were used as the training set, and a modified support vector regression (SVR) method was used to establish the relationship between descriptors and experimental binding affinities. Details about the establishment of ID-Score please see Ref. [27]. Here we used a modified version of ID-Score. In this version of ID-Score, the descriptors regarding surface property

matching are not calculated, since they are difficult to be accurately calculated. Besides, two other descriptors, including $K_{rec-SAlnonSA}^{desolv}$ representing the ratio of the polar and nonpolar surface areas of the protein binding site and $K_{rec-SARA}^{desolv}$ describing the solvent accessible surface area of the protein binding site, were also discarded, because these two descriptors each are a constant value for one target protein. A total of 42 descriptors were finally used for the establishment of this version of ID-Score (see [Supporting Information Table S1](#)). The validation results are given in [Supporting Information Fig. S5](#).

Ligand efficiency, which is defined as the binding energy per atom of a ligand to its receptor [28,29], is then calculated based on the calculated ID-Score value.

$$LE = 1.4 * (ID - \text{Score value}) / N$$

where N is the number of non-hydrogen atoms. In LEADOPT, LE rather than scoring function is used as a measure to sort the newly generated molecules.

2.2.4. Evaluations of pharmacokinetic and toxic properties

The evaluations of pharmacokinetic and toxic properties were carried out using SCADMET, which is a prediction program of pharmacokinetic and toxic properties developed using a support vector machine (SVM) method [30–33]. In LEADOPT, 12 types of pharmacokinetic and toxic properties are predicted, including human oral bioavailability, caco-2 cell permeability, *in vivo* clearance, human intestine absorption, human plasma protein binding rate, pregnane X receptor ligand, half lethal concentration (LC₅₀), aqueous solubility (logS), mitochondria toxicity, genotoxicity, human ether-a-go-go-related gene (hERG) toxicity, and teratogenicity. The establishment and evaluation of these prediction models are well documented in Ref. [30–33]. For each of the generated molecules, LEADOPT calculates an ADMET score (ADMET-score); if a compound passes the evaluation of any of the pharmacokinetic and toxic properties, a value of 1 is assigned to the ADMET-score, otherwise 0 is assigned. The overall ADMET-score value is a sum of all the individual values for the 12 pharmacokinetic and toxic properties, which will be used to indicate the overall ADMET features of the generated derivatives.

2.2.5. Implementation of LEADOPT

The source code for the implementation of LEADOPT was written in C/C++ programming language under the LINUX operating system. The LEADOPT program is available free of charge to not-for-

profit institution upon request from the corresponding author.

2.3. Chemistry

All reagents and solvents were purchased from commercial sources (Adamas-beta Switzerland, Acros Organics USA, Sigma–Aldrich Switzerland) and used without further purification. Reactions were monitored by thin-layer chromatography (TLC) analysis (Merck, 0.2 mm silica gel 60 F₂₅₄ on glass plates). ¹H NMR and ¹³C NMR spectra were recorded on a Bruker AV-400 spectrometer at 400 and 100 MHz, respectively. Chemical shifts are reported in parts per million (ppm) relative to tetramethylsilane (TMS), and multiplicities are designed as s (singlet), d (doublet), t (triplet), m (multiplet), or br s (broad singlet). Mass spectroscopy analyses were performed with an Agilent 1100 series LC–MS instrument with UV detection at 254 nm in low-resonance electrospray mode (ESI). All of the target compounds were purified to >95% purity, as determined by high-performance liquid chromatography (HPLC). HPLC analysis was performed on a Waters 2695 HPLC system with the use of a Kramosil C18 column (4.6 mm × 250 mm, 5 μm).

2.4. Enzymatic and functional assays

2.4.1. In vitro kinase inhibitory assays

All the kinase inhibitory assays in this study were tested via the Kinase Profiler service provided by Millipore Company. The IC₅₀ values for compounds were determined from dose–response curves obtained from the assays at 10 different concentrations of each compound. All the assays were test twice and the mean values of the replicates were calculated.

2.4.2. Cell lines and cell culture

The Ramos B and MV4-11 cell lines were obtained from the American Type Culture Collection (ATCC). Ramos B cells were grown in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS; Caoyuan Ivye, Huhht, China) in 5% CO₂ at 37 °C. MV4-11 was maintained in IMDM medium according to ATCC guidelines. Human umbilical vein endothelial cells (HUVECs) were isolated from human umbilical cord veins using a standard operating procedure and grown in EBM-2 basal medium (Cat. No.cc-3156, Lonza Inc.).

2.4.3. Wound healing assay

HUVECs were cultured to confluence in 24 well plates and wounded by using a sterilized yellow pipette tip to make a straight scratch. Cells were rinsed gently with sterile PBS, and then PBS was replaced with EGM-2 medium (EBM-2 basal medium plus recombinant human VEGF) containing vehicle, compound **A4** or sorafenib (Selleck Inc). After 18 h, pictures were taken using an OLYMPUS digital camera attached to a light microscope.

2.4.4. Cell viability assays

Cell viability was measured using MTT assay. The Ramos B cells and MV4-11 cells were seeded in a 96-wellplate at (2–5) × 10⁴ cells per well in RPMI-1640 and IMDM medium, respectively, with 10% FBS and incubated overnight. Different concentrations of compound **B4** or R406 (Selleck Inc.) were added to the cells and incubated at 37 °C for 72 h. Then, 20 μL of 5 mg/mL MTT solution was added to each well. After the incubation for 2–4 h at 37 °C, 50 μL of acidified SDS (20%, w/v) was added to lyse the oxidative product. Finally, the light absorption of the dissolved cells was measured at 570 nm on Multiskan MK3 (Thermo Scientific, USA). Each assay was performed in triplicate. The IC₅₀ values were calculated using GraphPad Prism software.

2.4.5. Western blot analysis

After treatment with different concentrations of compound **B4** or R406 for 6 h at 37 °C, MV4-11 cells were harvested, washed with ice-cold physiological saline, and lysed with RIPA lysis buffer (Beyotime, China) containing 1% cocktail (Sigma–Aldrich). Then, the cell lysates were separated by SDS–PAGE and electrotransferred onto PVDF membranes (Millipore). The PVDF membranes were incubated with each antibody and detected according to the immunoblot analysis principle. All the antibodies used in this immunoblot analysis were purchased from Cell Signaling Technology, with the exception of the anti-FLT3 antibody, which was obtained from Abcam.

For the HUVEC immunoblot studies, the cells were serum starved overnight in EBM-2 medium. After that, the cells were incubated with vehicle, compound **A4** or sorafenib for 2 h, followed by treatment with 50 ng/mL recombinant human VEGF (Lonza Inc.) for 10 min. The cells were harvested and lysed in RIPA buffer with 1% cocktail. Proteins were separated by gel electrophoresis on 5%–10% SDS–PAGE gels and probed with antibodies. All of the antibodies were used at a 1:1000 dilution, and the horseradish peroxidase-coupled secondary antibodies (Zhong Shan Golden Bridge Biotechnology, China) were used at 1:5000.

2.4.6. In vivo live fluorescent zebrafish assay

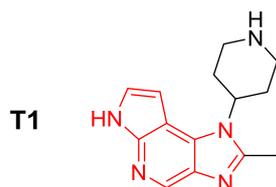
Transgenic zebrafish (FLK-1:EGFP) embryos were grown and maintained in accordance to the same protocols as given in Ref. [34]. Compound **B4** and R406 were prepared initially as 10 mmol/L stock solution in dimethyl sulfoxide (DMSO), and then were diluted in different assay concentrations with fish water, and 0.3% DMSO served as control. The embryos were distributed to 24-well plates (10 embryos in each well). Then the embryos were exposed to compound solution and included at 28.5 °C from 15 h postfertilization to 31 h postfertilization. In this end, the zebrafishes were anesthetized with 0.01% tricaine and imaged using a fluorescence microscope (Carl Zeiss Microimaging Inc.) equipped with an AxioCam MRc5 digital CCD camera (Carl Zeiss Microimaging Inc.).

3. Results and discussion

3.1. Development of LEADOPT

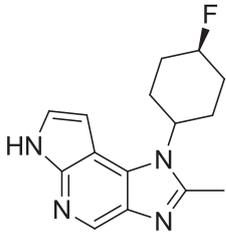
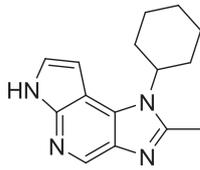
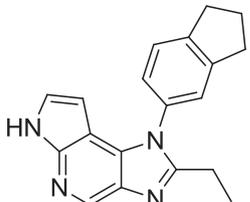
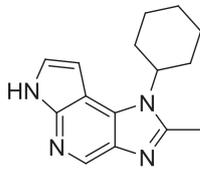
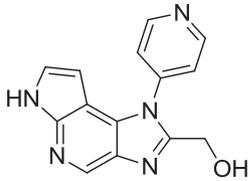
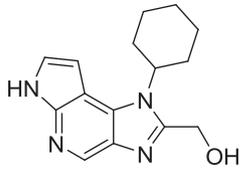
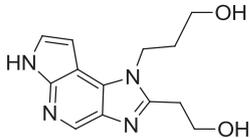
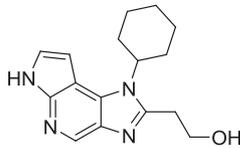
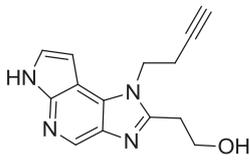
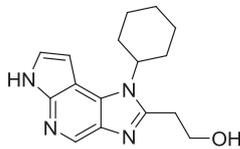
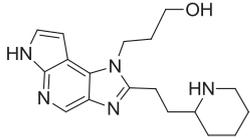
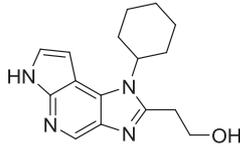
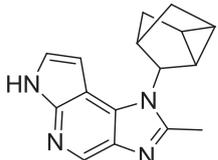
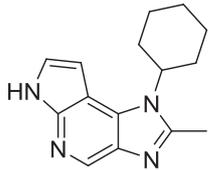
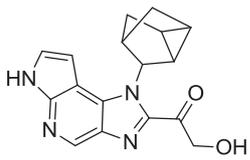
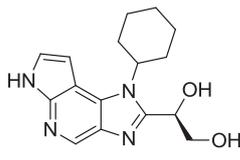
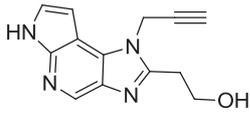
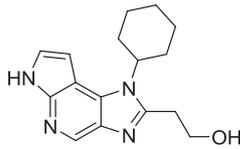
LEADOPT is an automatic tool developed for structure-based lead optimization. The overall workflow of LEADOPT is schematically depicted in Fig. 1. A detailed description of the algorithms of LEADOPT has been given in the Methods and Materials section. Here we just make a short summary for the workflow of LEADOPT. LEADOPT is started through input of the structure of ligand–receptor complex, which can be from either experimental X-ray crystal structure or molecule docking. Firstly, LEADOPT performs an analysis to the binding site of receptor, which includes identifying the active pocket, partitioning the active pocket into small equal-sized square grids, and checking which grids are accessible with a probe. Secondly, LEADOPT carries out the structural modification, which is the most crucial step in the whole lead optimization process. A fragment library containing 6877 fragments that were obtained through cleaving known drug molecules was established in advance. Based on the established fragment library, two types of structural modification operations, fragment growing and fragment replacing, were used to generate new derivatives. Finally, LEADOPT performs evaluations of bioactivity and LE, and evaluations of pharmacokinetic and toxic properties of the new derivatives. The bioactivity evaluation was performed by estimating the binding affinity of ligand with the target protein, which was carried out using a modified version of ID-Score. LE was

Table 3
20 representative derivatives generated by LEADOPT in the structural optimization of JAK1 inhibitor (**T1**) together with their most similar compounds of the known JAK1 inhibitors that are more potent than **T1**.



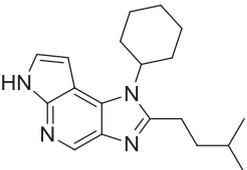
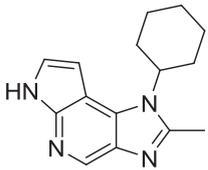
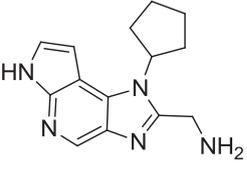
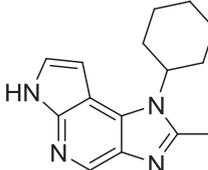
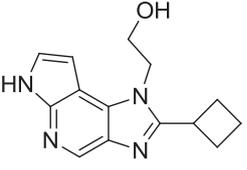
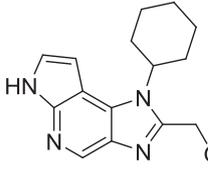
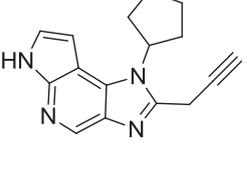
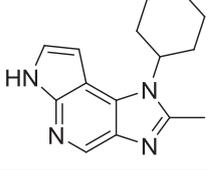
ID	Compounds generated by LEADOPT			The most similar JAK1 inhibitors that are more potent than T1		
	Chemical structure	Calculated LE	ADMET Score	Chemical structure	MeasuredLE ^a	Tc ^b
1		0.437244	7		0.64	1
2		0.436997	7		0.61	0.979592
3		0.462091	7		0.64	0.953488
4		0.492473	7		0.61	0.94
5		0.446427	8		0.64	0.931818
6		0.437396	9		0.64	0.930233
7		0.434816	10		0.61	0.924528

Table 3 (continued)

ID	Compounds generated by LEADOPT			The most similar JAK1 inhibitors that are more potent than T1		
	Chemical structure	Calculated LE	ADMET Score	Chemical structure	MeasuredLE ^a	Tc ^b
8		0.447304	7		0.64	0.913043
9		0.437135	9		0.64	0.911111
10		0.439435	9		0.61	0.9
11		0.484464	9		0.55	0.8909
12		0.434759	9		0.55	0.8908
13		0.435165	10		0.55	0.8905
14		0.52583	7		0.64	0.883721
15		0.453018	8		0.55	0.881356
16		0.449766	7		0.55	0.872727

(continued on next page)

Table 3 (continued)

ID	Compounds generated by LEADOPT			The most similar JAK1 inhibitors that are more potent than T1		
	Chemical structure	Calculated LE	ADMET Score	Chemical structure	MeasuredLE ^a	Tc ^b
17		0.445717	8		0.64	0.87234
18		0.460586	7		0.64	0.866667
19		0.44639	8		0.61	0.867925
20		0.445855	7		0.64	0.866667

^a The experimentally measured LE values were selected from Ref. [35].

^b The Tanimoto coefficients were calculated with the aid of 'MDLPublicKeys' fingerprints implemented in Discovery Studio 3.1.

then calculated and used for sorting the generated molecules. The evaluations of pharmacokinetic and toxic properties were performed using SCADMET, which is a prediction program of pharmacokinetic and toxic properties developed using the support vector machine method. After the evaluations of bioactivity/LE and pharmacokinetics/toxicity properties, the program will output a number of derivatives of the lead compound together with their calculated LEs and ADMET-scores.

3.2. Retrospective case studies

To test the performance of LEADOPT, we carried out retrospective case studies on two selected targets: JAK1 and MPS1; these targets were chosen because crystal structural information of these targets and structure–activity relationships (SARs) of their ligands were all publicly available.

JAK1 is a key intracellular mediator of helical cytokine signaling pathways, which plays a central role in inflammation, immune function and hematopoiesis [35]. Very recently, Zak et al. reported a series of selective JAK1 inhibitors and the crystal structure of JAK1 in complex with an inhibitor (**T1**), 2-methyl-1-(piperidin-4-yl)-1,6-dihydroimidazo[4,5-d]pyrrolo[2,3-b]pyridine (PDB code: 4EHZ) [35]. In this case study, we tried to use LEADOPT to perform structural modifications for **T1**. The 1,6-dihydroimidazo[4,5-d]pyrrolo[2,3-b]pyridine motif of **T1** was chosen as the core scaffold (highlighted in red in Table 3). The running of LEADOPT led to the generation of a total of 40,961 derivatives. Among them, 4898 derivatives have an LE value larger than 0.4343 (an LE value of **T1**) and an ADMET-score value larger than 7 (an ADMET-score value of **T1**).

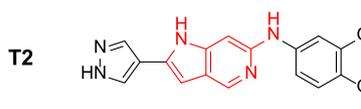
Then, we performed a structural similarity analysis through calculating the Tanimoto coefficients (Tc) between the generated derivatives and known JAK1 inhibitors that are more potent than **T1** using MDLPublicKeys fingerprints implemented in Discovery Studio 3.1. The results showed that 55 derivatives have a Tc value larger than 0.85, and, particularly, one derivative has a Tc value of 1; two compounds bearing a large value of Tc are generally thought to have a very similar bioactivity. Table 3 lists the top 20 representative derivatives generated by LEADOPT according to the Tc value.

MPS1 is a central component of the spindle assembly checkpoint signal, which has been linked to many human diseases especially cancers. Naud et al. recently identified a series of 1H-Pyrrolo[3,2-c]pyridine derivatives as inhibitors of MPS1 through structure-based drug design, and obtained a crystal structure of MPS1 with one inhibitor (**T2**), N-(3,4-dimethoxyphenyl)-2-(1H-pyrrolo[3,2-c]pyridin-6-yl)-1H-pyrrolo[3,2-c]pyridin-6-amine (PDB code: 4C4E) [36]. Here the 1H-pyrrolo[3,2-c]pyridin-6-amine moiety of **T2** was defined as the core scaffold (Table 4). After running LEADOPT, 95,419 derivatives were generated, of which 8643 derivatives have an LE value larger than 0.4150 (an LE value of **T2**) and an ADMET-score value larger than 11 (an ADMET-score value of **T2**). Again, we performed a similarity analysis between these derivatives and known MPS1 inhibitors whose experimentally measured LE values larger than **T2**. A total of 86 derivatives have a Tc value larger than 0.85, and two completely identical compounds with known MPS1 inhibitors were generated (Tc = 1). Table 4 lists 20 representative derivatives for MPS1 generated by LEADOPT.

The above two retrospective case studies demonstrated, to a certain degree, the effectiveness of LEADOPT. In the following

Table 4

20 representative derivatives generated by LEADOPT in the structural optimization of MPS1 inhibitor (T2) together with their most similar compounds in the known JAK1 inhibitors that are more potent than T2.



ID	Compounds generated by LEADOPT			The most similar MPS1 inhibitors that are more potent than T2		
	Chemical structure	Calculated LE	ADMET Score	Chemical structure	Measured LE ^a	Tc ^b
1		0.471995	11		0.4627	1
2		0.446064	12		0.4491	1
3		0.464795	11		0.4806	0.975
4		0.447191	11		0.4491	0.958333
5		0.44145	11		0.4806	0.95122
6		0.453121	11		0.4806	0.95
7		0.417374	11		0.3751	0.938776
8		0.442617	11		0.4806	0.928571
9		0.451198	11		0.4806	0.926829
10		0.456824	11		0.4806	0.926829
11		0.492806	12		0.4806	0.926829
12		0.488157	11		0.4806	0.926829
13		0.417272	11		0.4627	0.916667
14		0.428762	11		0.4627	0.914894
15		0.418037	11		0.4491	0.914894
16		0.415137	12		0.4806	0.906977
17		0.434185	11		0.4806	0.904762

(continued on next page)

Table 4 (continued)

ID	Compounds generated by LEADOPT			The most similar MPS1 inhibitors that are more potent than T2		
	Chemical structure	Calculated LE	ADMET Score	Chemical structure	Measured LE ^a	Tc ^b
18		0.439537	11		0.4806	0.904762
19		0.904762	11		0.4806	0.904762
20		0.449756	12		0.4806	0.902439

^a The experimentally measured LE was calculated by: $LE = (-1.4 \log(\text{experimental } IC_{50})) / (n \text{ heavy atoms})$, where experimental IC_{50} values were selected from Ref. [36].

^b The Tanimoto coefficients were calculated with the aid of 'MDLPublicKeys' fingerprints implemented in Discovery Studio 3.1.

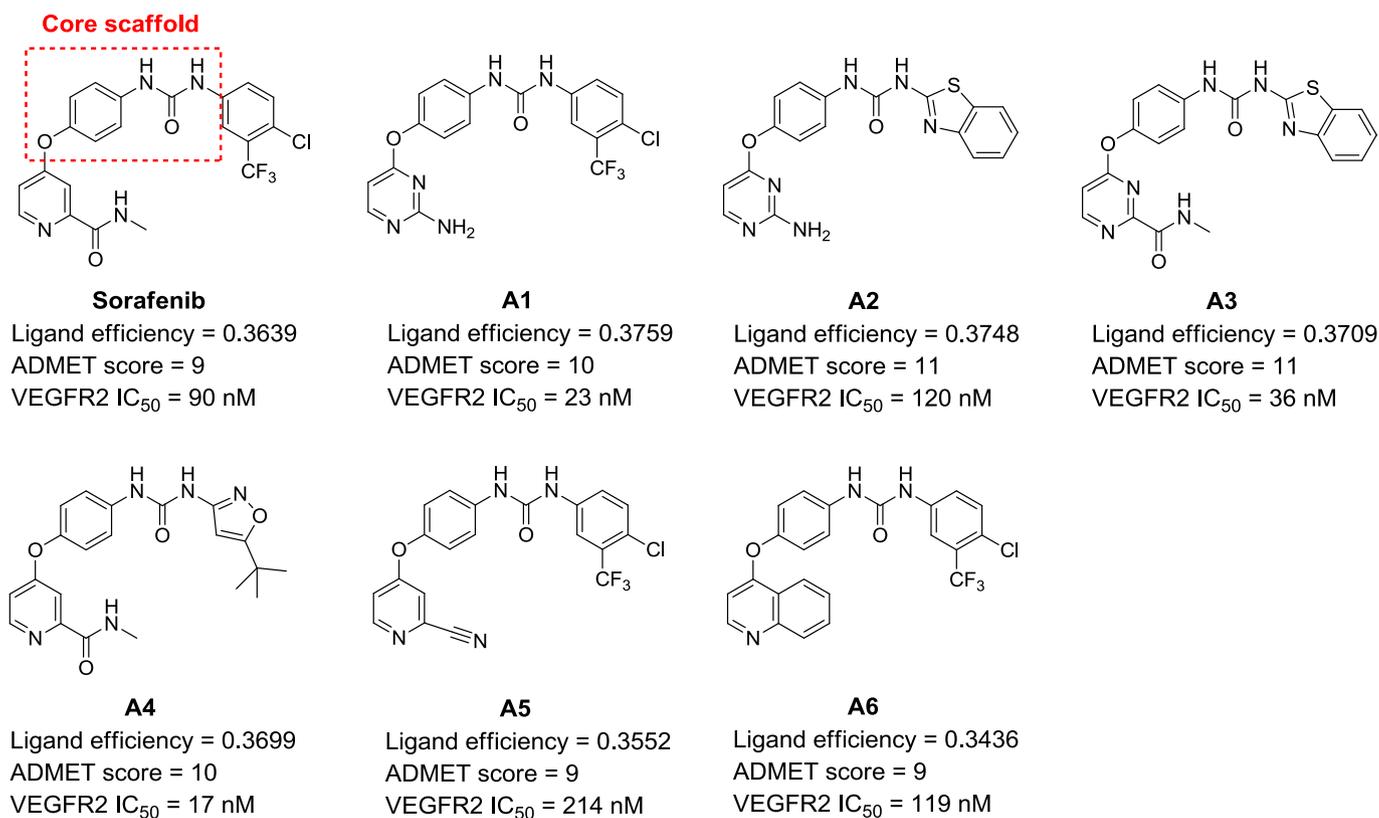


Fig. 3. The chemical structures and inhibitory activities of sorafenib and compounds A1–A6 against VEGFR2.

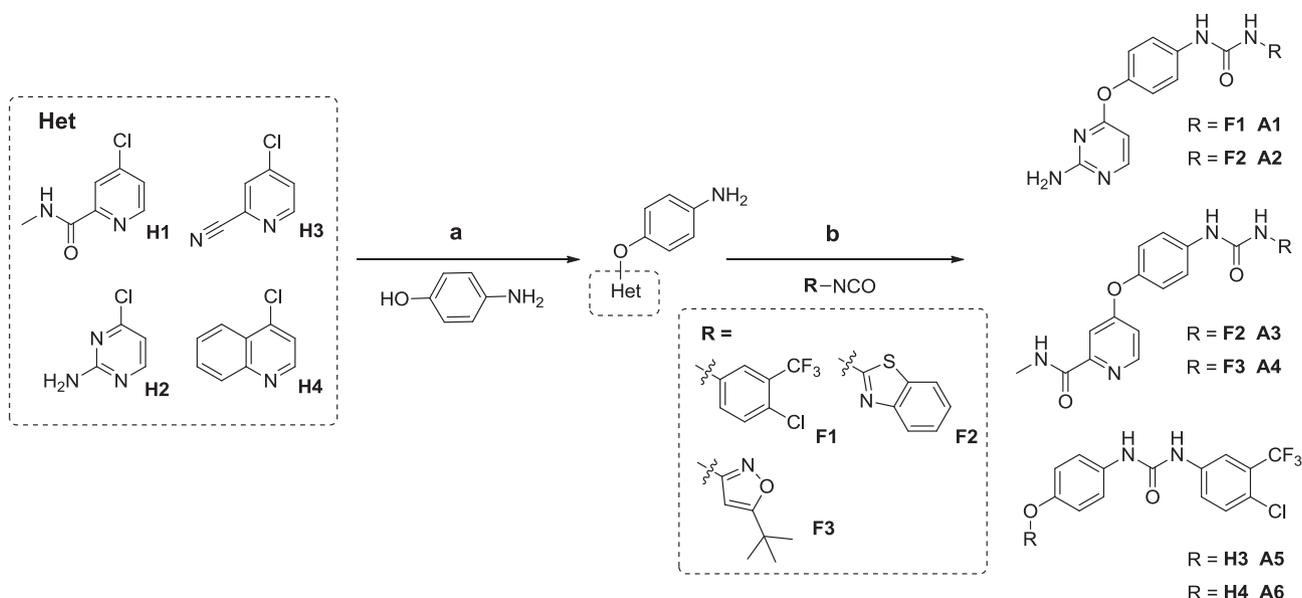
sections, we shall apply LEADOPT to the structural optimizations of VEGFR2 inhibitor sorafenib, and SYK inhibitor R406. The aim of which is mainly to further evaluate the performances of LEADOPT by these two studies. In addition, we also hope to obtain more potent VEGFR2 and SYK inhibitors due to their potential therapeutic values in treating related diseases.

3.3. Prospective study 1: the optimization of VEGFR2 inhibitor sorafenib

The VEGF receptor 2 (VEGFR2), a type III receptor tyrosine kinase, is a key mediator of angiogenesis. It has been recognized as an important target for diseases associated with angiogenesis,

particularly solid tumors [37]. Several VEGFR2 inhibitors have been approved by the US Food and Drug Administration (FDA) for treatment of solid tumors [38]. Among them, sorafenib is the first marketed small molecule VEGFR2 inhibitor [39]. However, its potency against VEGFR2 is relatively poorer compared with drugs approved recently (a half maximal inhibitory concentration (IC_{50}) of sorafenib against VEGFR2 is 90 nM) [39]. Here we utilized LEADOPT to perform structural optimization to improve its potency. The structure of sorafenib-VEGFR2 complex was taken from the Protein Data Bank (PDB code: 4ASD) [38]. The 1-(4-hydroxyphenyl)urea motif of sorafenib was defined as the core scaffold (see Fig. 3), which was kept unchanged in the optimization process.

LEADOPT generated a total of 94,071 molecules. Among them,



Scheme 1. The general synthetic route to compounds **A1–A6**. Reagents and conditions: (a) *t*-BuOK, DMF, 80–110 °C, 38% to quantitative; (b) for F1, F3: acetonitrile, reflux; for F2: toluene, reflux, quantitative.

64,412 compounds have an LE value larger than 0.3639 (an LE value of sorafenib) and an ADMET-score value larger than 9 (an ADMET-score value of sorafenib). From them, we selected four compounds (Fig. 3, **A1**, **A2**, **A3**, and **A4**) to synthesize. Besides, we also selected another two compounds (Fig. 3, **A5** and **A6**) that have an LE value less than 0.3639. The six compounds were selected because they have never been reported before and can be very easily synthesized without too much synthetic efforts in our current experimental conditions.

3.3.1. Chemical synthesis for compounds A1–A6

The target compounds **A1–A6** were synthesized following the procedures illustrated in Scheme 1. Firstly, 4-aminophenol (10 mmol, 1.0 equiv) forms 4-aminophenolate in a potassium tert-butyl alcohol solution (12 mmol, 1.2 equiv, potassium tert-butyl alcohol in 5 ml/10 mmol dimethyl formamide) during 0.5 h. Selective nucleophilic substitution reaction of the oxygen anion of 4-aminophenolate with the chloride of heterocyclic compounds (**H1–H4**, 1.0 equiv) resulted in the corresponding intermediates containing an ether bond at 80 °C–110 °C with the yields ranging from 38% to quantitative, since the nucleophilicity of the oxygen anion is relatively stronger than that of nitrogen in 4-aminophenolate. Secondly, the intermediates were respectively reacted with the isocyanates (**F1–F3**, 1.1 equiv) that were readily synthesized using a similar method as that described previously [40], producing the target compounds **A1–A6** with excellent yields. For details about compounds **A1–A6** see Supporting Information.

3.3.2. Inhibitory activities of compounds A1–A6 against VEGFR2

The inhibitory activities of compounds **A1–A6** were measured with gold-standard ^{33}P radio labeled technology. The results are present in Fig. 3. Three of the four compounds that have an LE value larger than 0.3639 (**A1**, **A3**, and **A4**) have indeed shown a higher potency against VEGFR2 than sorafenib. The most active compound, **A4**, showed a 5-fold improvement of VEGFR2 inhibition compared with sorafenib (IC_{50} : 17 nM of **A4** vs 90 nM of sorafenib). Fig. 4A shows the interaction mode of the most active compound, **A4**, with the VEGFR2 kinase. The introduced fragment by LEADOPT, 5-(tert-butyl)isoxazole, forms very intensive hydrophobic

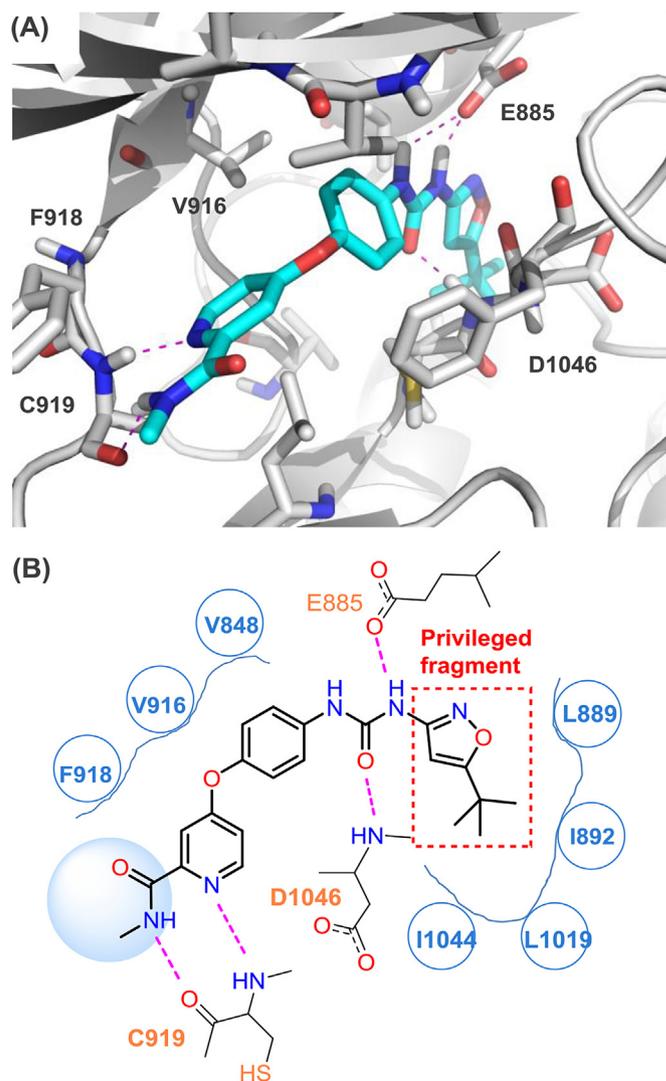


Fig. 4. The binding mode of compound **A4** with VEGFR2.

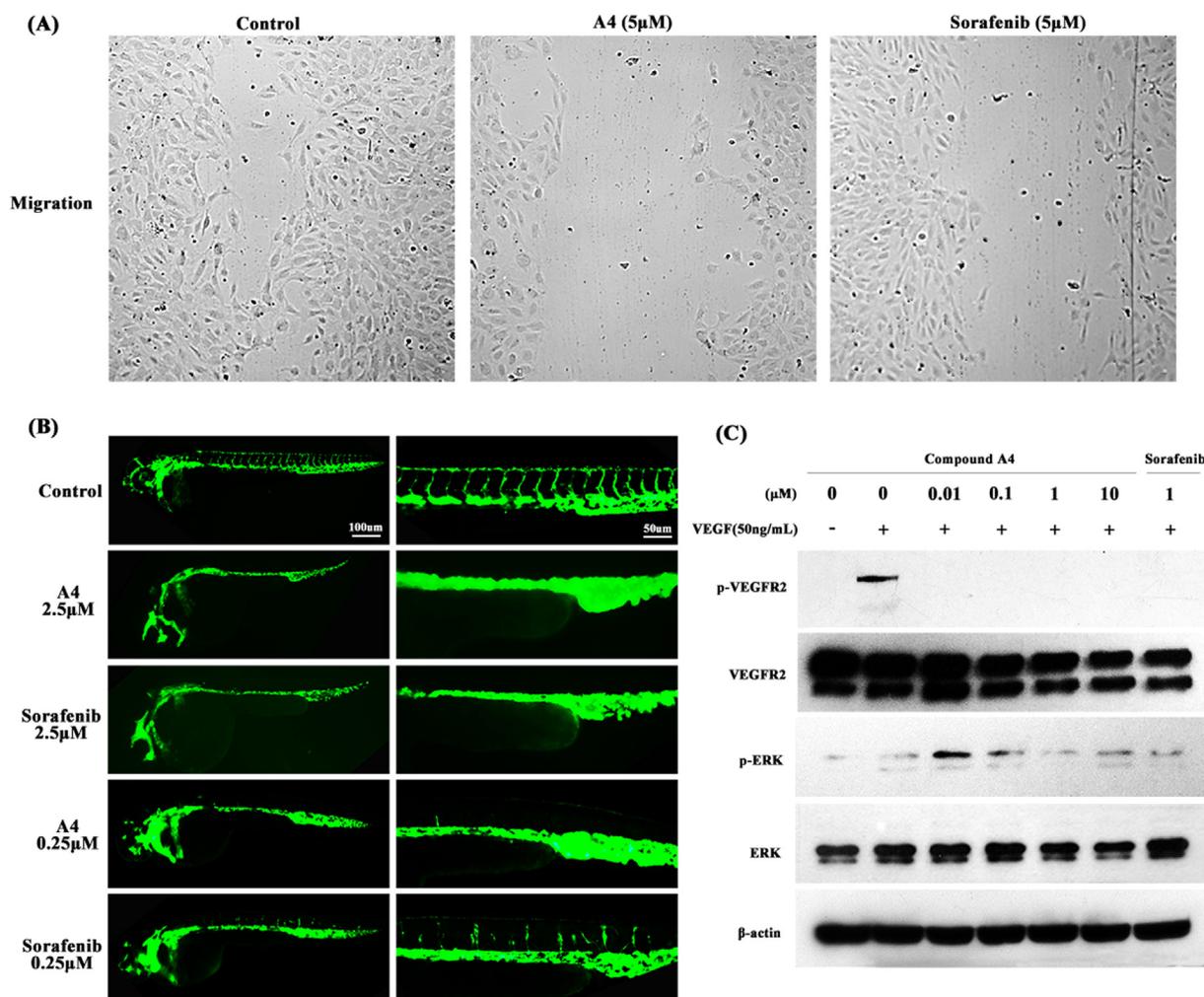


Fig. 5. The anti-angiogenesis effects of compound **A4** in HUVECs and transgenic zebrafish embryos assays and its anti-angiogenesis mechanism of action. (A) Compound **A4** inhibited HUVEC cells migration in wound healing assay; (B) Compound **A4** inhibited zebrafish embryonic angiogenesis; (C) Compound **A4** inhibited the phosphorylation of VEGFR2 and ERK1/2 in VEGF-stimulated HUVECs.

interactions with Ile1044, Leu1019, Ile892, and Leu889, which is the most important factor contributing to the high potency of compound **A4** against VEGFR2 (see Fig. 4B). Compounds **A5** and **A6**, which have a lower LE value than sorafenib, indeed showed a lower inhibitory activity against VEGFR2. Collectively, these results validated, at least to some extent, the effectiveness of LEADOPT.

3.3.3. Functional evaluations of compound **A4**

It has been widely accepted that VEGFR2 is the most crucial regulator of angiogenesis. We thus assessed the effects of compound **A4** on angiogenesis using *in vitro* functional assays and *in vivo* transgenic zebrafish assays. Firstly, wound healing assays were performed to evaluate the influence of compound **A4** on the migration ability of human umbilical vein endothelial cells (HUVEC). The results showed that compound **A4** could dose-

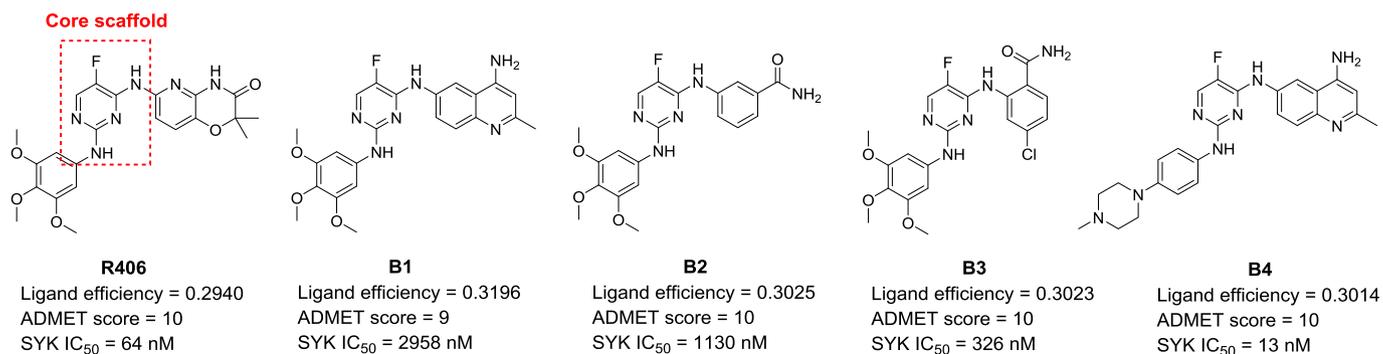
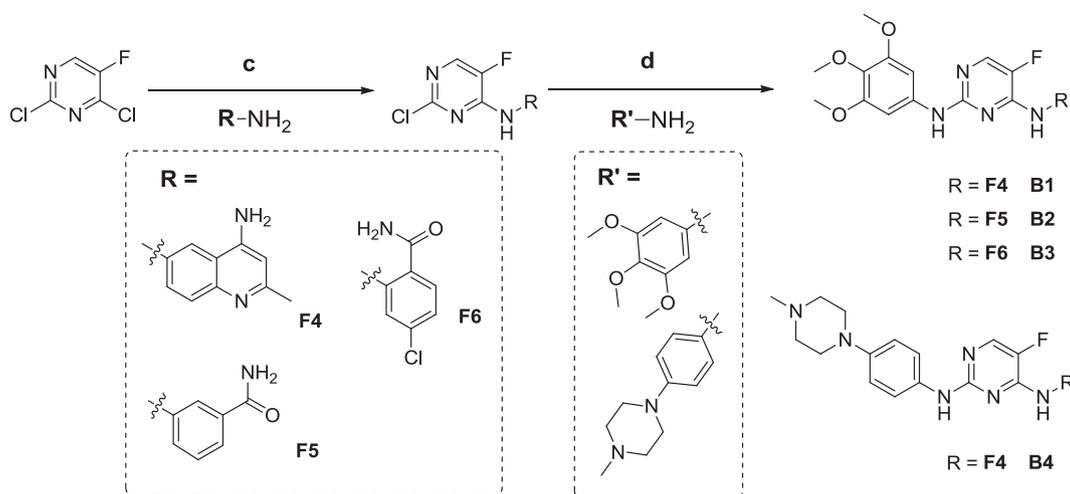


Fig. 6. The chemical structures and inhibitory activities of compounds **B1–B4** against SYK.



Scheme 2. The general synthetic route to compounds **B1–B4**. Reagents and conditions: (c) for F4, F5, F6: DIPEA, ethyl alcohol, 90–100 °C, 64%–91%; (d) for 3,4,5-trimethoxyaniline: DIPEA, n-butyl alcohol, 90–120 °C, 70%–95%; for 4-(4-methylpiperazin-1-yl)aniline: HCl, dimethyl formamide, 120–125 °C, 32%–61%.

independently prevent HUVECs migration (see [Supporting Information Fig. S6](#)) and significantly reduce the number of migrating cells at a concentration of 5 μM . By comparison, sorafenib showed a slightly weak suppression effects in the wound healing assay. Secondly, transgenic zebrafish assays were used to directly evaluate the anti-angiogenesis ability of compound **A4**. The results showed that compound **A4** could almost entirely inhibit the growth of the intersegmental blood vessel of zebrafish at a concentration of 0.25 μM ([Fig. 5B](#)), while 0.25 μM sorafenib just led to a suppression of about 60%. Finally, western blot analysis was adopted to examine the ability of compound **A4** to inhibit the phosphorylation of VEGFR2 and downstream ERK1/2 in HUVECs. As shown in [Fig. 5C](#), compound **A4** could effectively inhibit the phosphorylation of VEGFR2 and ERK1/2. Taken together, all the results demonstrated that compound **A4** could effectively inhibit angiogenesis and block the VEGFR2 signaling in intact cells.

3.4. Prospective study 2: the optimization of SYK inhibitor R406

Spleen tyrosine kinase (SYK), a non-receptor tyrosine kinase, is a key integrator of intracellular signals triggered by activated immune receptors, including B cell receptors (BCR) and Fc receptors, which are important for the development and function of lymphoid cells [41]. SYK has been thought an attractive drug target for the treatment of autoimmune disorders, such as rheumatoid arthritis, allergic asthma, and multiple sclerosis [42,43]. Though some SYK inhibitors have been reported, there is no SYK inhibitor used in clinic so far. R406 is a typical SYK inhibitor, which is the active moiety of R788, an SYK inhibitor that is in clinical trials [43]. Nevertheless, it just has moderate activity against SYK with IC_{50} of 64 nM. Here we shall use LEADOPT to optimize its potency.

The structure of R406-SYK complex was taken from the Protein Data Bank (PDB code: 3FQS) [44]. The 5-fluoropyrimidine-2,4-diamine motif in R406 was defined as the core scaffold ([Fig. 6](#)). After LEADOPT running, 82,690 molecules were generated. Among them, 67,975 molecules have an LE value larger than 0.2940 (an LE value of R406) and an ADMET-score value larger than 9. We selected four compounds (**B1–B4**), which have an LE value larger than 0.2940, to synthesize. Again, these compounds were selected because they have never been reported before and can be very easily synthesized without too much synthetic efforts in our current experimental conditions.

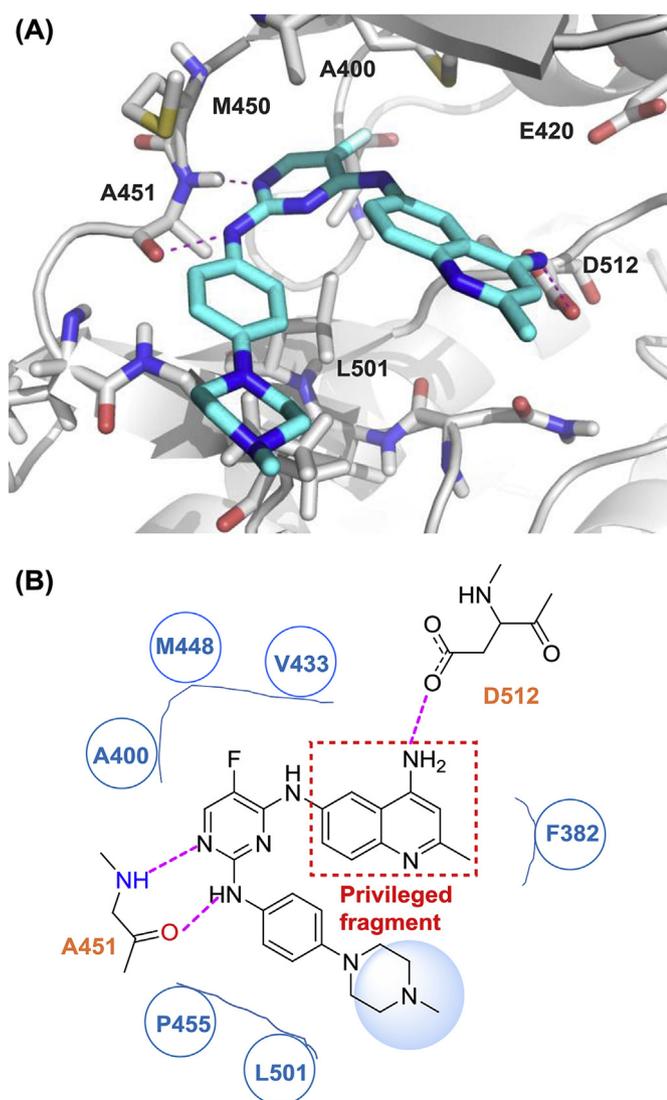


Fig. 7. The binding mode of compound **B4** with SYK.

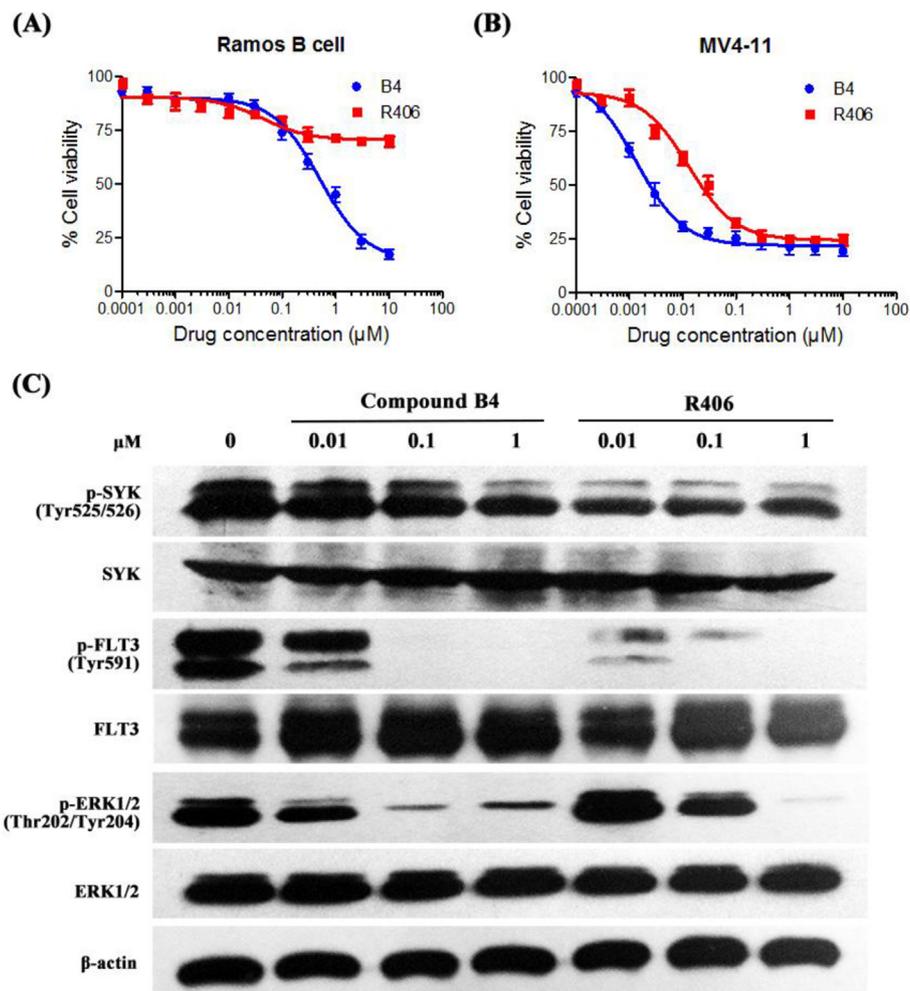


Fig. 8. *In vitro* growth inhibitory profile of compound **B4** against (A) Ramos B cells and (B) MV4-11 cells using the MTT assay. (C) The inhibitory effects of compound **B4** on phosphorylation and downstream signals of SYK in MV4-11 cells was examined using the western blot analysis.

3.4.1. Chemical synthesis for compounds **B1**–**B4**

The general synthetic route chosen for the synthesis of compounds **B1**–**B4** is depicted in Scheme 2. Compounds **B1**–**B4** were all synthesized from the initial compound, 2,4-dichloro-5-fluoropyrimidine, which contains two regioselective chlorine atoms. Firstly, commercially available or readily synthesized amines (**F4**–**F6**, 1.0 equiv) were reacted regioselectively with the 4-chlorine of 2,4-dichloro-5-fluoropyrimidine (1.2 equiv) at 90 °C–100 °C in the presence of diisopropylethylamine (DIPEA) to afford the corresponding 2-chloro-5-fluoro-N-substituted-pyrimidin-4-amine intermediates. Subsequently, the nucleophilic displacement of 2-chlorine of the intermediates by 3,4,5-trimethoxyaniline (1.0 equiv) or 4-(4-methylpiperazin-1-yl)aniline (1.0 equiv) led to the target compounds **B1**–**B4** with the yields ranging from 32% to 95%, which required conditions of a temperature of 90 °C or 125 °C and the catalysis of acid (hydrochloric acid for 4-(4-methylpiperazin-1-yl)aniline) or base (DIPEA for 3,4,5-trimethoxyaniline). Details about compounds **B1**–**B4** are given in Supporting Information.

3.4.2. Inhibitory activities of compounds **B1**–**B4** against SYK

The inhibitory activities of compounds **B1**–**B4** are present in Fig. 6. The four compounds all exhibited potency against SYK with IC_{50} values in the range of 2958 nM–13 nM. However only compound **B4** showed a higher potency (IC_{50} : 13 nM) than R406 (IC_{50} :

64 nM). Fig. 7A depicts the interaction mode between compound **B4** and SYK. Compound **B4** perfectly interacts with the residues in the binding site of SYK. Apparently, the assembled fragment, 2-methylquinolin-4-amine, forms a strong hydrogen-bond interaction with Asp512 and aromatic interactions with Phe382, which reflect that this fragment is very suitable to the hydrophobic pocket surrounded by residues of Phe382, Val433 (see Fig. 7B).

3.4.3. Functional evaluations of compound **B4**

We first assessed the inhibitory effect of compound **B4** on tumor cell viability by MTT assays. Here, two tumor cell lines, Ramos B [45] and MV4-11 [46], were used; the two tumor cell lines were chosen because their viabilities have been reported to be regulated by SYK [45,46]. Compound **B4** displayed a good anti-viability potency against both Ramos B and MV4-11 cell lines with IC_{50} values of 0.51 μM and 0.003 μM, respectively (Fig. 8A and B). For both cell lines, compound **B4** showed a higher potency than R406. Then, we used western blot analysis to examine the effect of compound **B4** on the SYK signaling in intact MV4-11 cells. As shown in Fig. 8C, compound **B4** significantly inhibited the phosphorylation of FLT3 and ERK1/2, which are SYK downstream signaling proteins. All of these demonstrated that compound **B4** has a good functional activity.

Though a more potent SYK inhibitor was obtained here, the performance of LEADOPT was apparently poorer than that in the

case study of VEGFR2 inhibitor optimization. In fact, because we just chose a very limited number of simple compounds to synthesize and measured their potency in the two case studies, it is hard to make a performance comparison from the results. Of course, we also acknowledge the possibility that LEADOPT has a different performance on different targets.

4. Concluding remarks

In this investigation, we developed an automatic tool for structure-based lead optimization, termed LEADOPT. LEADOPT starts its work from an input structure of ligand–receptor complex, which can come from either x-ray crystal structure or molecular docking. The structural modifications in LEADOPT mainly include two operations: fragment growing and fragment replacing, which are restricted to carry out in the active pocket of target protein with the core scaffold structure of ligand kept unchanged. LE rather than scoring function is used as a measure to sort the newly generated molecules. Twelve important pharmacokinetic and toxic properties are also evaluated using SCADMET. LEADOPT was evaluated with two retrospective cases. The results showed that LEADOPT could generate derivatives that are very similar with known potent compounds. LEADOPT was finally applied to structural optimizations of VEGFR2 inhibitor sorafenib, and SYK inhibitor R406. Though just several compounds were synthesized, we have obtained some compounds that are more potent than sorafenib and R406. All of these have validated, at least to some extent, the effectiveness of LEADOPT.

Finally, we have to mention that the binding affinities of the generated molecules are estimated directly by ID-Score without further sophisticated molecular docking because the main purpose of LEADOPT is to provide a rapid automatic tool for lead optimization. In addition, to make the generated molecules easy to synthesize, atom mutations are not allowed in LEADOPT. It is also necessary to point out that, although pharmacokinetic and toxic properties are evaluated in LEADOPT, we did not validate these issues in this investigation mainly due to our very limited experimental resource. Even so, it does not hamper LEADOPT as a useful tool for experimental medicinal chemists.

Acknowledgments

This work was supported by the 973 Program (2013CB967204), the 863 Hi-Tech Programs (2012AA020301, 2012AA020308), the National Natural Science Funds for Distinguished Young Scholar (81325021), the National S&T Major Project (2012ZX09501001-003), and partly by PCSIRT (NO: IRT13031).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.ejmech.2015.02.019>.

References

- [1] T. Kenakin, Predicting therapeutic value in the lead optimization phase of drug discovery, *Nat. Rev. Drug Discov.* 2 (2003) 429–438.
- [2] G.M. Keseru, G.M. Makara, The influence of lead discovery strategies on the properties of drug candidates, *Nat. Rev. Drug Discov.* 8 (2009) 203–212.
- [3] C.W. Murray, D.C. Rees, The rise of fragment-based drug discovery, *Nat. Chem.* 1 (2009) 187–192.
- [4] L.M. Mayr, D. Bojanic, Novel trends in high-throughput screening, *Curr. Opin. Pharmacol.* 9 (2009) 580–588.
- [5] J.P. Kennedy, L. Williams, T.M. Bridges, R.N. Daniels, D. Weaver, C.W. Lindsley, Application of combinatorial chemistry science on modern drug discovery, *J. Comb. Chem.* 10 (2008) 345–354.
- [6] C. McInnes, Virtual screening strategies in drug discovery, *Curr. Opin. Chem. Biol.* 11 (2007) 494–502.
- [7] U. Rester, From virtuality to reality – virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective, *Curr. Opin. Drug Discov. Devel.* 11 (2008) 559–568.
- [8] V.M. Popov, W.A. Yee, A.C. Anderson, Towards in silico lead optimization: scores from ensembles of protein/ligand conformations reliably correlate with biological activity, *Proteins* 66 (2007) 375–387.
- [9] W.L. Jorgensen, Efficient drug lead discovery and optimization, *Acc. Chem. Res.* 42 (2009) 724–733.
- [10] L.-L. Yang, G.-B. Li, H.-X. Yan, Q.-Z. Sun, S. Ma, P. Ji, Z.-R. Wang, S. Feng, J. Zou, S.-Y. Yang, Discovery of N6-phenyl-1H-pyrazolo[3,4-d]pyrimidine-3,6-diamine derivatives as novel CK1 inhibitors using common-feature pharmacophore model based virtual screening and hit-to-lead optimization, *Eur. J. Med. Chem.* 56 (2012) 30–38.
- [11] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *J. Am. Chem. Soc.* 110 (1988) 5959–5967.
- [12] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, *J. Med. Chem.* 37 (1994) 4130–4146.
- [13] C.L. Kuo, H. Assefa, S. Kamath, Z. Brzozowski, J. Slawinski, F. Saczewski, J.K. Buolamwini, N. Neamati, Application of CoMFA and CoMSIA 3D-QSAR and docking studies in optimization of mercaptobenzenesulfonamides as HIV-1 integrase inhibitors, *J. Med. Chem.* 47 (2004) 385–399.
- [14] B.M. Baron, R.J. Cregge, R.A. Farr, D. Friedrich, R.S. Gross, B.L. Harrison, D.A. Janowick, D. Matthews, T.C. McCloskey, S. Meikrantz, P.L. Nycze, R. Vaz, W.A. Metz, CoMFA, synthesis, and pharmacological evaluation of (E)-3-(2-carboxy-2-arylvinyl)-4,6-dichloro-1H-indole-2-carboxylic acids: 3-[2-(3-aminophenyl)-2-carboxyvinyl]-4,6-dichloro-1H-indole-2-carboxylic acid, a potent selective glycine-site NMDA receptor antagonist, *J. Med. Chem.* 48 (2005) 995–1018.
- [15] A. Politi, S. Durdagi, P. Moutevelis-Minakakis, G. Kokotos, M.G. Papadopoulos, T. Mavromoustakos, Application of 3D QSAR CoMFA/CoMSIA and in silico docking studies on novel renin inhibitors against cardiovascular diseases, *Eur. J. Med. Chem.* 44 (2009) 3703–3711.
- [16] B. Wendt, M. Mulbaier, S. Wawro, C. Schultes, J. Alonso, B. Janssen, J. Lewis, Toluidinesulfonamide hypoxia-induced factor 1 inhibitors: alleviating drug-drug interactions through use of PubChem data and comparative molecular field analysis guided synthesis, *J. Med. Chem.* 54 (2011) 3982–3986.
- [17] S. Sartini, V. Coviello, A. Bruno, V. La Pietra, L. Marinelli, F. Simorini, S. Taliari, S. Salerno, A.M. Marini, A. Fioravanti, P. Orlandi, A. Antonelli, F. Da Settimo, E. Novellino, G. Bocci, C. La Motta, Structure-based optimization of tyrosine kinase inhibitor CLM3. Design, synthesis, functional evaluation and molecular modeling studies, *J. Med. Chem.* 57 (2014) 1225–1235.
- [18] M. Mori, A. Massaro, V. Calderone, M. Fragai, C. Luchinat, A. Mordini, Discovery of a new class of potent MMP inhibitors by structure-based optimization of the Arylsulfonamide scaffold, *ACS Med. Chem. Lett.* 4 (2013) 565–569.
- [19] J. Jacobs, V. Grum-Tokars, Y. Zhou, M. Turlington, S.A. Saldanha, P. Chase, A. Egglar, E.S. Dawson, Y.M. Baez-Santos, S. Tomar, A.M. Mielech, S.C. Baker, C.W. Lindsley, P. Hodder, A. Mesecar, S.R. Stauffer, Discovery, synthesis, and structure-based optimization of a series of n-(tert-Butyl)-2-(n-arylamido)-2-(pyridin-3-yl) Acetamides (ML188) as potent Noncovalent small molecule inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Protease, *J. Med. Chem.* 56 (2013) 534–546.
- [20] J.D. Bauman, D. Patel, S.F. Baker, R.S.K. Vijayan, A. Xiang, A.K. Parhi, L. Martinez-Sobrido, E.J. LaVoie, K. Das, E. Arnold, Crystallographic fragment screening and structure-based optimization yields a new class of influenza endonuclease inhibitors, *ACS Chem. Biol.* 8 (2013) 2501–2508.
- [21] O. Nicolotti, I. Giangreco, T.F. Miscioscia, A. Carotti, Improving quantitative structure-activity relationships through multiobjective optimization, *J. Chem. Inf. Model.* 49 (2009) 2290–2302.
- [22] C. Abad-Zapatero, J.T. Metz, Ligand efficiency indices as guideposts for drug discovery, *Drug Discov. Today* 10 (2005) 464–469.
- [23] C. Abad-Zapatero, Ligand efficiency indices for effective drug discovery, *Expert. Opin. Drug Discov.* 2 (2007) 469–488.
- [24] A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J.P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.* 40 (2012) D1100–D1107.
- [25] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets, *Nucleic Acids Res.* 36 (2008) D901–D906.
- [26] X.Q. Lewell, D.B. Judd, S.P. Watson, M.M. Hann, RECAP-retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry, *J. Chem. Inf. Comput. Sci.* 38 (1998) 511–522.
- [27] G.B. Li, L.L. Yang, W.J. Wang, L.L. Li, S.Y. Yang, ID-Score: a new empirical scoring function based on a comprehensive set of descriptors related to protein–ligand interactions, *J. Chem. Inf. Model.* 53 (2013) 592–600.
- [28] I.D. Kuntz, K. Chen, K.A. Sharp, P.A. Kollman, The maximal affinity of ligands, *Proc. Natl. Acad. Sci. USA* 96 (1999) 9997–10002.
- [29] S.H. Reynolds, B.A. Toung, S.D. Bembenek, Ligand binding efficiency: trends, physical basis, and implications, *J. Med. Chem.* 51 (2008) 2432–2438.
- [30] S.Y. Yang, Q. Huang, L.L. Li, C.Y. Ma, H. Zhang, R. Bai, Q.Z. Teng, M.L. Xiang, Y.Q. Wei, An integrated scheme for feature selection and parameter setting in the support vector machine modeling and its application to the prediction of

- pharmacokinetic properties of drugs, *Artif. Intell. Med.* 46 (2009) 155–163.
- [31] H. Zhang, Q.Y. Chen, M.L. Xiang, C.Y. Ma, Q. Huang, S.Y. Yang, In silico prediction of mitochondrial toxicity by using GA-CG-SVM approach, *Toxicol. Vitro* 23 (2009) 134–140.
- [32] H. Zhang, M.L. Xiang, C.Y. Ma, Q. Huang, W. Li, Y. Xie, Y.Q. Wei, S.Y. Yang, Three-class classification models of logS and logP derived by using GA-CG-SVM approach, *Mol. Divers.* 13 (2009) 261–268.
- [33] C.Y. Ma, S.Y. Yang, H. Zhang, M.L. Xiang, Q. Huang, Y.Q. Wei, Prediction models of human plasma protein binding rate and oral bioavailability derived by using GA-CG-SVM method, *J. Pharm. Biomed. Anal.* 47 (2008) 677–682.
- [34] T.C. Tran, B. Sneed, J. Haider, D. Blavo, A. White, T. Aiyejorun, T.C. Baranowski, A.L. Rubinstein, T.N. Doan, R. Dingleline, E.M. Sandberg, Automated, quantitative screening assay for Antiangiogenic compounds using transgenic zebrafish, *Cancer Res.* 67 (2007) 11386–11392.
- [35] M. Zak, C.A. Hurley, S.I. Ward, P. Bergeron, K. Barrett, M. Balazs, W.S. Blair, R. Bull, P. Chakravarty, C. Chang, Identification of C-2 hydroxyethyl imidazopyrrolopyridines as potent JAK1 inhibitors with favorable physicochemical properties and high selectivity over JAK2, *J. Med. Chem.* 56 (2013) 4764–4785.
- [36] S. Naud, I.M. Westwood, A. Faisal, P. Sheldrake, V. Bavetsias, B. Atrash, K.M. Cheung, M. Liu, A. Hayes, J. Schmitt, A. Wood, V. Choi, K. Boxall, G. Mak, M. Gurden, M. Valenti, A. de Haven Brandon, A. Henley, R. Baker, C. McAndrew, B. Matijssen, R. Burke, S. Hoelder, S.A. Eccles, F.I. Raynaud, S. Linardopoulos, R.L. van Montfort, J. Blagg, Structure-based design of orally bioavailable 1H-pyrrolo[3,2-c]pyridine inhibitors of mitotic kinase monopolar spindle 1 (MPS1), *J. Med. Chem.* 56 (2013) 10045–10065.
- [37] N. Ferrara, R.S. Kerbel, Angiogenesis as a therapeutic target, *Nature* 438 (2005) 967–974.
- [38] M. McTigue, B.W. Murray, J.H. Chen, Y.L. Deng, J. Solowiej, R.S. Kania, Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors, *Proc. Natl. Acad. Sci. USA* 109 (2012) 18281–18289.
- [39] S. Wilhelm, C. Carter, M. Lynch, T. Lowinger, J. Dumas, R.A. Smith, B. Schwartz, R. Simantov, S. Kelley, Discovery and development of sorafenib: a multikinase inhibitor for treating cancer, *Nat. Rev. Drug Discov.* 5 (2006) 835–844.
- [40] L.L. Yang, G.B. Li, S. Ma, C. Zou, S. Zhou, Q.Z. Sun, C. Cheng, X. Chen, L.J. Wang, S. Feng, L.L. Li, S.Y. Yang, Structure-activity relationship studies of pyrazolo [3,4-d]pyrimidine derivatives leading to the discovery of a novel multi-kinase inhibitor that Potently inhibits FLT3 and VEGFR2 and evaluation of its activity against acute myeloid leukemia in vitro and in vivo, *J. Med. Chem.* 56 (2013) 1641–1655.
- [41] A. Mocsai, J. Ruland, V.L.J. Tybulewicz, The SYK tyrosine kinase: a crucial player in diverse biological functions, *Nat. Rev. Immunol.* 10 (2010) 387–402.
- [42] M. Ulanova, F. Duta, L. Puttagunta, A.D. Schreiber, A.D. Befus, Spleen tyrosine kinase (Syk) as a novel target for allergic asthma and rhinitis, *Expert Opin. Ther. Targets* 9 (2005) 901–921.
- [43] M.E. Weinblatt, A. Kavanaugh, M.C. Genovese, T.K. Musser, E.B. Grossbard, D.B. Magilav, An oral spleen tyrosine kinase (Syk) inhibitor for rheumatoid arthritis, *New. Engl. J. Med.* 363 (2010) 1303–1312.
- [44] A.G. Villaseñor, R. Kondru, H. Ho, S. Wang, E. Papp, D. Shaw, J.W. Barnett, M.F. Browner, A. Kuglstatler, Structural insights for design of potent spleen tyrosine kinase inhibitors from crystallographic analysis of three inhibitor complexes, *Chem. Biol. Drug Des.* 73 (2009) 466–470.
- [45] F. Padilla, N. Bhagirath, S. Chen, E. Chiao, D.M. Goldstein, J.C. Hermann, J. Hsu, J.J. Kennedy-Smith, A. Kuglstatler, C. Liao, W. Liu, L.E. Lowrie, K.C. Luk, S.M. Lynch, J. Menke, L. Niu, T.D. Owens, C. O-Yang, A. Railkar, R.C. Schoenfeld, M. Slade, S. Steiner, Y.-C. Tan, A.G. Villaseñor, C. Wang, J. Wanner, W. Xie, D. Xu, X. Zhang, M. Zhou, M.C. Lucas, Pyrrolopyrazines as selective spleen tyrosine kinase inhibitors, *J. Med. Chem.* 56 (2013) 1677–1692.
- [46] A. Puissant, N. Fenouille, G. Alexe, Y. Pikman, C.F. Bassil, S. Mehta, J. Du, J.U. Kazi, F. Luciano, L. Ronnstrand, A.L. Kung, J.C. Aster, I. Galinsky, R.M. Stone, D.J. DeAngelo, M.T. Hemann, K. Stegmaier, SYK is a critical regulator of FLT3 in acute myeloid leukemia, *Cancer Cell* 25 (2014) 226–242.