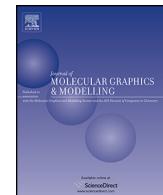




Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM



Effective virtual screening strategy focusing on the identification of novel Bruton's tyrosine kinase inhibitors

Jianhu Xiao, Shengping Zhang, Minghao Luo, Yi Zou, Yihua Zhang, Yisheng Lai*

State Key Laboratory of Natural Medicines, Jiangsu Key Laboratory of Drug Discovery for Metabolic Diseases, Center of Drug Discovery, China Pharmaceutical University, Nanjing 210009, China

ARTICLE INFO

Article history:

Received 19 March 2015

Received in revised form 5 May 2015

Accepted 7 May 2015

Available online 16 May 2015

Keywords:

BTK inhibitor

Pharmacophore model

Virtual screening

Docking

Molecular dynamics.

ABSTRACT

Dysregulation of the B-cell receptor (BCR) signaling pathway plays a vital role in the pathogenesis and development of B-cell malignancies. Bruton's tyrosine kinase (BTK), a key component in the BCR signaling, has been validated as a valuable target for the treatment of B-cell malignancies. In an attempt to find novel and potent BTK inhibitors, both ligand- and structure-based pharmacophore models were generated using Discovery Studio 2.5 and Ligandscout 3.11 with the aim of screening the ChemBridge database. The resulting hits were then subjected to sequential docking experiments using two independent docking programs, CDOCKER and Glide. Molecules displaying high glide scores and H-bond interactions with the key residue Met477 in both of the docking programs were retained. Drug-like criteria including Lipinski's rule of five and ADMET properties filters were employed for further refinement of the retrieved hits. By clustering, eight promising compounds with novel chemical scaffolds were finally selected and the top two ranking compounds were evaluated by molecular dynamics simulation. We believe that these compounds are of great potential in BTK inhibition and will be used for further investigation.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

B-cell malignancy is one of the most common cancers in the world, which is particularly highlighted by the high incidence in the western countries [1,2]. The B-cell receptor (BCR) signaling pathway acts as a key regulator in differentiation, activation and proliferation of B-lineage lymphoid cells [3,4]. Ample evidence has indicated that aberrant activation of the BCR signaling pathway is critical to the occurrence and development of B-cell malignancy [4–8]. Among various kinases involved in the BCR pathway, Bruton's tyrosine kinase (BTK), a non-receptor cytoplasmic protein tyrosine kinase that belongs to the Tec kinase family [9], is predominantly expressed in B-lymphocyte lineages and plays a central role in the BCR signaling [10,11]. BTK is initially phosphorylated at Tyr551 by its upstream kinases Syk and Lyn. Followed by the autophosphorylation of another key residue (Tyr223), BTK becomes physiologically active and then triggers the reaction of its substrates [12]. Recent studies have clearly demonstrated that the deregulation of BTK is implicated in the pathogenesis of B-cell-derived malignancies [13–15]. Therefore, inhibition of BTK has garnered significant interest of multiple pharmaceutical

companies for the treatment of B-cell lymphoma. Currently, enormous efforts have been made to discover novel BTK inhibitors.

However, until now, a limited number of promising BTK inhibitors have been designed and reported publicly (Fig. 1). Among them, ibrutinib [16] developed by Pharmacyclics and Janssen had just been approved by the US FDA for the treatment of mantle cell lymphoma (MCL) in November 2013 and chronic lymphocytic leukemia (CLL) in February 2014, respectively. Ibrutinib is a selective and irreversible BTK inhibitor that forms a covalent bond with a nonconserved cysteine residue (Cys481) proximal to the active site of BTK [17]. As an ATP-competitive inhibitor, ibrutinib binds to the ATP-binding pocket of BTK and abrogates the full activation of BTK by preventing the autophosphorylation of Tyr223, which thereby culminates the signal transduction cascade [17]. It is notable that ibrutinib is the first and only FDA-approved inhibitor of BTK till now. Despite its unprecedented success, acquired resistance to ibrutinib has already been observed in patients with CLL [18] and MCL [19]. Driven by this situation, there is an urgent need to discover BTK inhibitors with novel chemical scaffolds.

Virtual screening, a powerful tool for lead identification, has won huge reputation in recent drug development. A large number of successful examples have demonstrated the reliability of computational methods in discovering novel hits for a certain target [20–22]. Pharmacophore query, as one of virtual screening methods, is widely applied in database screening. In the past, due to the lack

* Corresponding author. Tel.: +86 025 83271015; fax: +86 025 83271015.

E-mail address: [yслai@cpu.edu.cn](mailto:yslai@cpu.edu.cn) (Y. Lai).

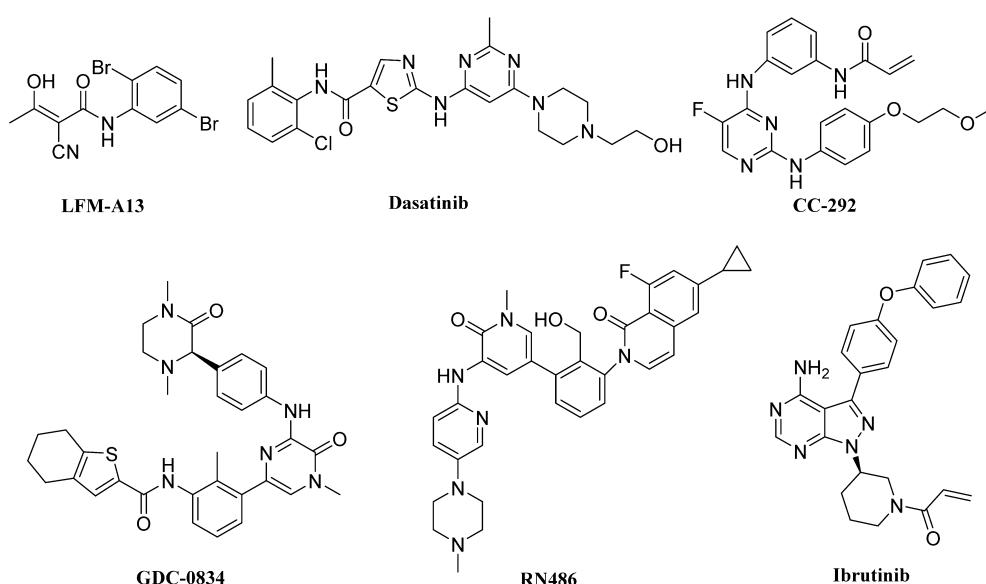


Fig. 1. Chemical structures of representative BTK inhibitors previously reported.

of protein structure information, the ligand-based pharmacophore model was utilized more frequently [23,24]. However, with large numbers of protein structures being elucidated in recent years, the application of structure-based pharmacophore has obtained much more popularity [25,26]. Since both of them have their own advantages and drawbacks [27], the combination of two pharmacophore generation strategies has already become the mainstream in computer-aided virtual screening. Moreover, it has turned out that the integration of them in one single virtual screening workflow would achieve higher hit rates [28,29].

To our best knowledge, no work has been reported about the combination of ligand- and structure-based pharmacophore screening approaches in finding novel BTK inhibitors. Herein, we first delineated an effective virtual screening campaign toward the discovery of novel and potent BTK inhibitors (Fig. 2). In our study, ligand- and structure-based pharmacophore models were generated based on the selected training set and the protein-ligand complexes, respectively. Both of them were rigidly validated through a variety of well-established methods and then applied for the screening of ChemBridge database parallelly. Compounds

passing the pharmacophore requirements were subsequently subjected to a cascade docking method. Those forming H-bond with the key residue Met477 and displaying a glide score ≥ 7 had been extracted. Eight representative compounds with novel chemical scaffolds were picked up for further study after ADMET properties prediction, structure clustering and visual inspection. Furthermore, molecular dynamics (MD) simulation was performed for the two top-ranking compounds to analyze the protein-ligand stability.

2. Materials and methods

2.1. Ligand-based pharmacophore modeling

2.1.1. Data set

Since BTK has been validated as a key component in the BCR signaling pathway which is closely associated with B-cell malignancies, a number of small molecules have been reported as potent BTK inhibitors. Among them, a total of 41 compounds with experimental inhibitory activity (IC_{50}) were carefully selected from different literature resources [16,30–34] and used in our study. It is believed that the quality of the constructed pharmacophore derived by Hypogen algorithm greatly depends on the training set, which should meet the following rules as suggested by the Accelrys Discovery Studio. (1) A minimum of 16 compounds covering large structural diversity are required in the training set. (2) Both the most active and inactive compounds should be included in the training set and with the activity range spanning at least four orders of magnitude. Based on the above criteria, we selected 23 compounds as the training set (Fig. 3) with their IC_{50} values ranging from 0.52 nM to 58162 nM, and the rest 18 compounds were used as the test set (Fig. 4) to validate the generated pharmacophore model. All the molecules were sketched in ChemDraw Ultra 14.0 and converted into 3D structures by Discovery Studio 2.5 (DS2.5) [35]. For each molecule, energy minimization process was performed with CHARMM force field [36], and a maximum of 255 conformations were generated within the energy threshold of 20.0 kcal/mol using the BEST/flexible conformation generation option implemented in DS2.5 [37].

2.1.2. 3D QSAR pharmacophore generation

The 3D QSAR Pharmacophore Generation module available in DS2.5 is a productive technique to generate quantitative

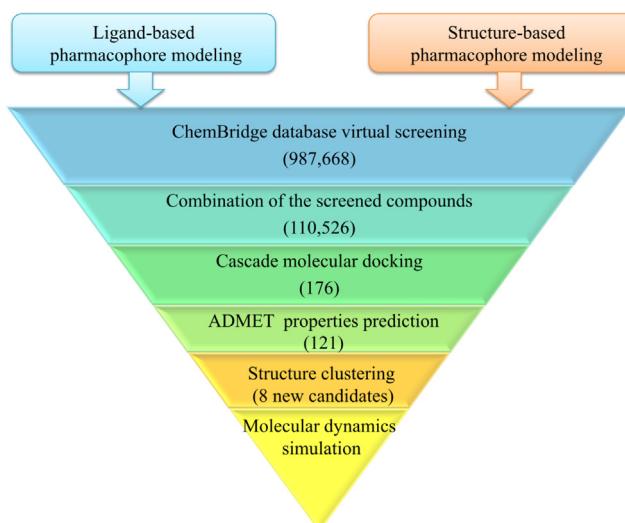
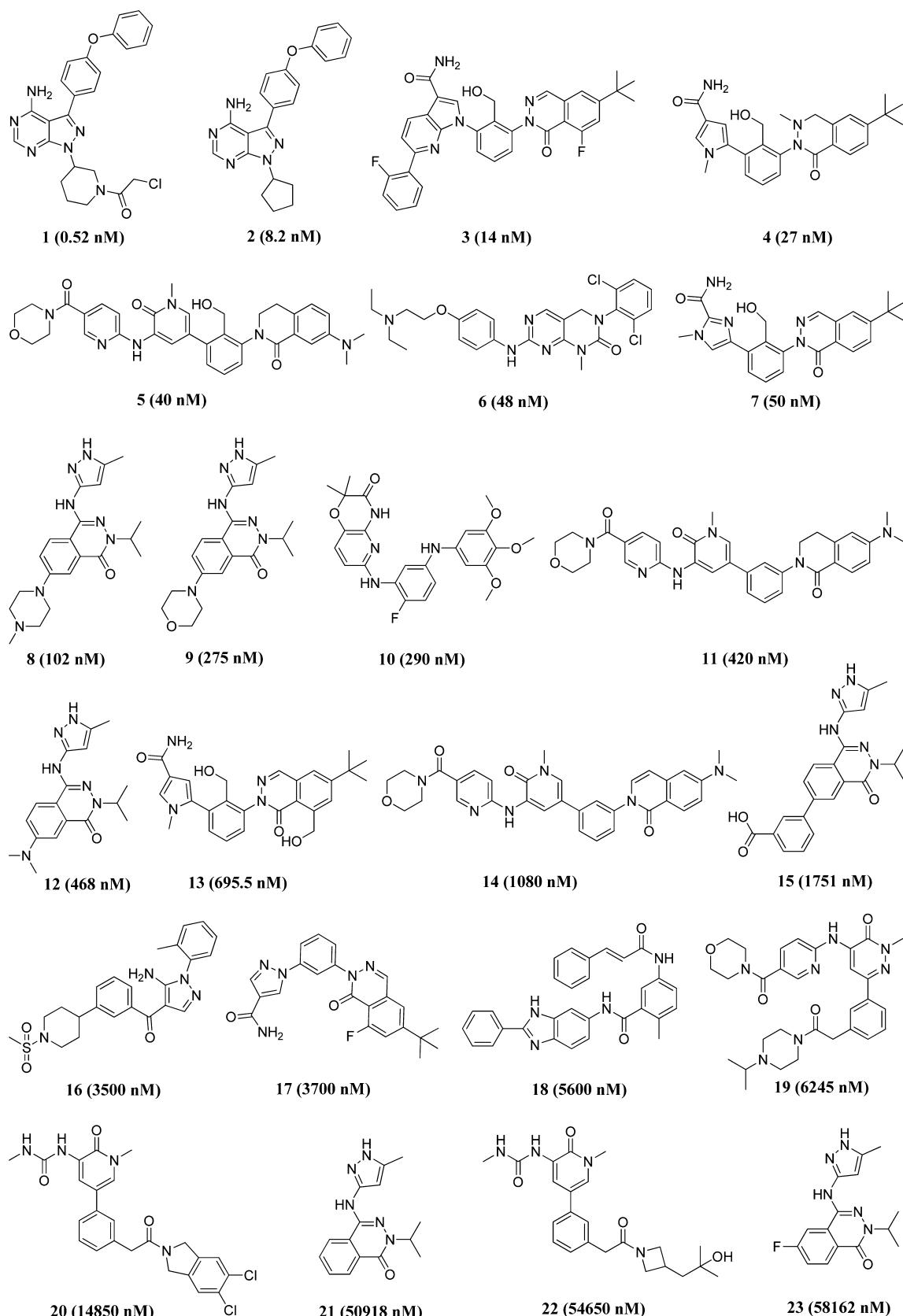
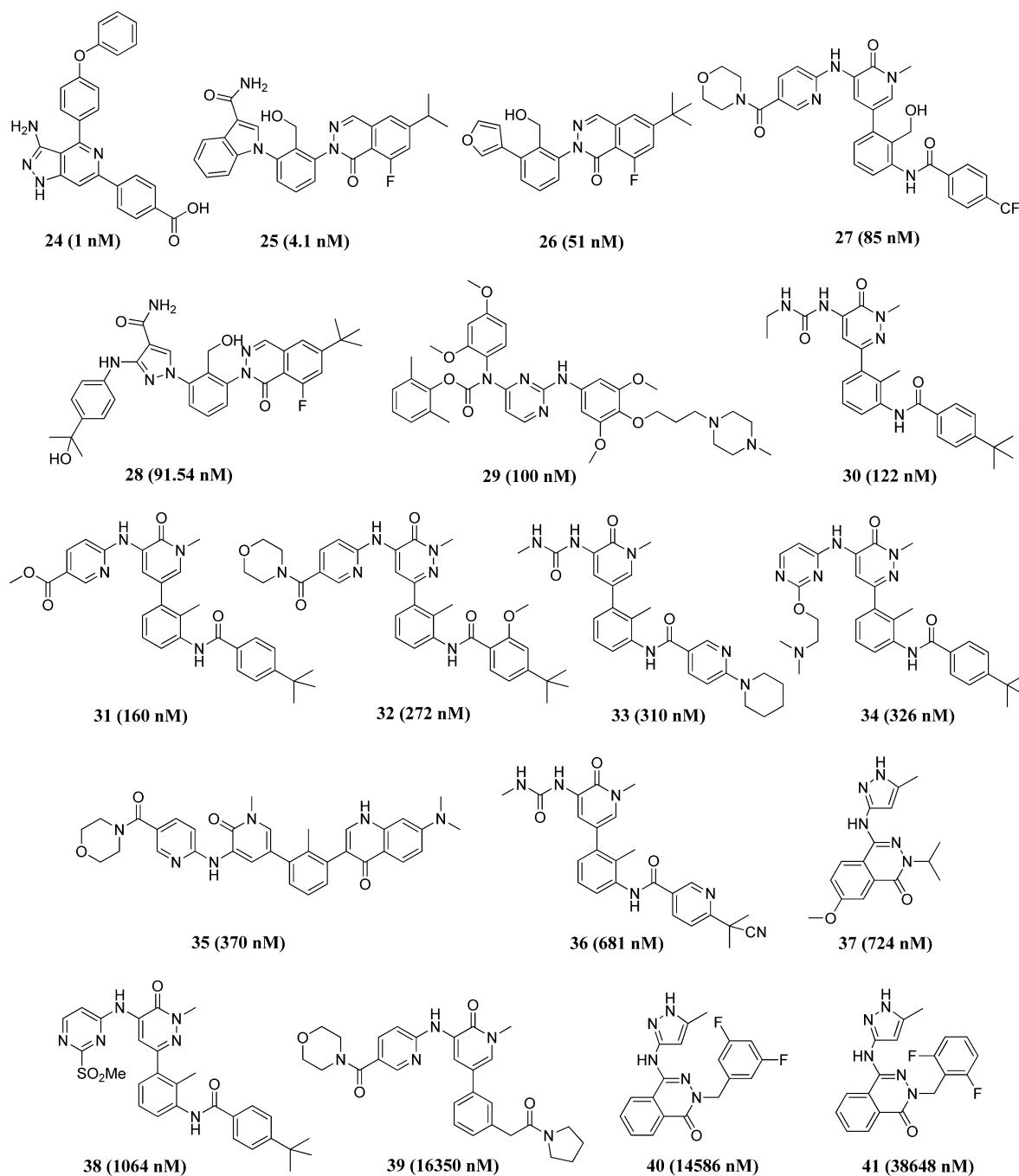


Fig. 2. Flowchart of virtual screening strategy adopted for this study.

**Fig. 3.** Chemical structures and IC_{50} values of the training set compounds.

**Fig. 4.** Chemical structures and IC_{50} values of the test set compounds.

pharmacophore models, and has been proved successful in discovering lead compounds with new scaffolds [24,38]. Herein, the prepared training set was submitted to the protocol, and the features including hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), ring aromatic (RA), and hydrophobic (HY) were selected to develop pharmacophore hypotheses according to the result of feature mapping. Taking the steric effect into consideration, the value for 'Maximum Excluded Volumes' was set to 4. The minimum interfeature distance option was set to 2, which meant the minimum distance between two feature points was 2 Å. The uncertainty value was changed from default 3 to 2.5, all other parameters were kept as default. Subsequently, a total of 10 pharmacophore models were constructed. The best pharmacophore hypothesis was identified on the basis of its statistical parameters and applied for further validation.

2.2. Structure-based pharmacophore modeling

2.2.1. Protein structures preparation

The protein structures reported have paved the way for structure-based pharmacophore generation. To date, a total of eight crystal structures of BTK protein in complex with different ligands are available in the RCSB Protein Data Bank (PDB) [39,40], among which six structures were selected for further modeling according to their binding modes and crystallographic resolution (PDB codes: 3GEN, 3K54, 3PJ1, 3PJ3, 3PIY, 3PIZ). These complexes were prepared by DS2.5 with all hydrogen added and water deleted. The Align Structures module implemented in DS2.5 was used to superimpose the coordinates of the six protein complexes with 3GEN as a reference structure, which has the highest crystallographic resolution.

Table 1

Statistical parameters of top 10 pharmacophore hypotheses generated by HypoGen algorithm.

Hypo no.	Total cost	Cost difference ^a	RMS ^b	Correlation	Features ^c
Hypo1	98.47	97.12	0.75	0.97	HBD HY RA RA 4EV
Hypo2	107.31	88.28	1.14	0.94	HBD HY RA RA 1EV
Hypo3	112.46	83.13	1.36	0.91	HBD HY RA RA 1EV
Hypo4	112.46	83.13	1.36	0.91	HBD HY RA RA 2EV
Hypo5	118.32	77.27	1.53	0.88	HBD HY RA RA 1EV
Hypo6	119.87	75.72	1.59	0.87	HBD HY RA RA
Hypo7	120.00	75.59	1.59	0.87	HBD HY HY HY
Hypo8	120.84	74.75	1.57	0.88	HBD HY RA RA 1EV
Hypo9	123.63	71.96	1.67	0.86	HBA HBD HY RA
Hypo10	125.69	69.90	1.70	0.85	HBD HY RA RA 1EV

^a Cost difference = null cost-total cost; null cost = 195.59; fixed cost = 90.63; configuration cost = 16.31. All cost units in bits. Configuration cost: a fixed cost which depends on the complexity of the hypothesis space being optimized.

^b RMS: the deviation of the log (estimated activities) from the log (measured activities) normalized by the log (uncertainties).

^c EV: excluded volumes.

2.2.2. Structure-based pharmacophore generation

LigandScout 3.11 [41] is a good partner for automatic pharmacophore construction. The software is capable of identifying various types of ligand-receptor interactions in a predefined binding site, such as hydrogen bond interactions, hydrophobic areas and charge transfer interactions [42]. All the protein complexes aforementioned were imported into LigandScout 3.11, and six individual pharmacophore models were generated with default parameters. Then all the pharmacophore features were clustered in DS2.5 to develop the most-frequent-feature pharmacophore model. The constraint tolerance of spheres in the multicomplex-based pharmacophore model was further refined to default values of DS2.5.

2.3. Pharmacophore model validation

The purpose of the pharmacophore validation is to evaluate the capability of the developed pharmacophore model in discriminating active molecules from the inactives and predicting their activities. Currently, there are three major methods to validate the constructed pharmacophore models.

2.3.1. Fischer's validation

Fischer's randomization method [43] is applied to checking whether there is a good statistical relevance of the proposed model. In this method, several random hypotheses are generated by scrambling the experimental activity data of the original training set

compounds with the same features and parameters of the initial hypothesis. The statistical significance is calculated using the following formula.

$$\text{Significance} = 100(1 - (1 + x)/y)$$

Here, "x" is the total number of hypotheses having a total cost lower than the best significance, and "y" is the number of all the HypoGen runs. Nineteen random spreadsheets would be generated with the confidence level be set to 95% in our study.

2.3.2. Test set validation

This method is used to confirm whether the pharmacophore model constructed could estimate the activity of the test set compounds. With the best hypothesis selected, the *Ligand Pharmacophore Mapping* module is utilized to evaluate the predictive ability of the hypothesis.

2.3.3. Decoy set validation

A good pharmacophore should not only be able to estimate the activity of the active compounds, but also have the ability to identify the active molecules from a huge database containing a

Table 2

Experimental and predicted activities of the training set compounds based on Hypo1.

Compound	Experimental IC ₅₀ (nM)	Estimated IC ₅₀ (nM)	Error ^a
1	0.52	1.38	+2.65
2	8.20	3.25	-2.52
3	14	22.36	+1.60
4	27	46.49	+1.72
5	40	52.94	+1.32
6	48	39.83	-1.21
7	50	48.91	-1.02
8	102	433.38	+4.25
9	275	431.58	+1.57
10	290	179.40	-1.62
11	420	592.48	+1.41
12	468	1220	+2.61
13	695.50	281	-2.48
14	1080	815	-1.32
15	1751	921.50	-1.90
16	3500	6013	+1.72
17	3700	1947	-1.90
18	5600	1726	-3.25
19	6245	6184	-1.01
20	14,850	33,925	+2.28
21	50,918	35,074	-1.45
22	54,650	33,889	-1.61
23	58,162	38,796	-1.50

^a Difference between the predicted and experimental values; '+' indicates that the predicted IC₅₀ is higher than the experimental IC₅₀; '-' indicates that the predicted IC₅₀ is lower than the experimental IC₅₀.

Fig. 5. The workflow of ligand-based pharmacophore generation.

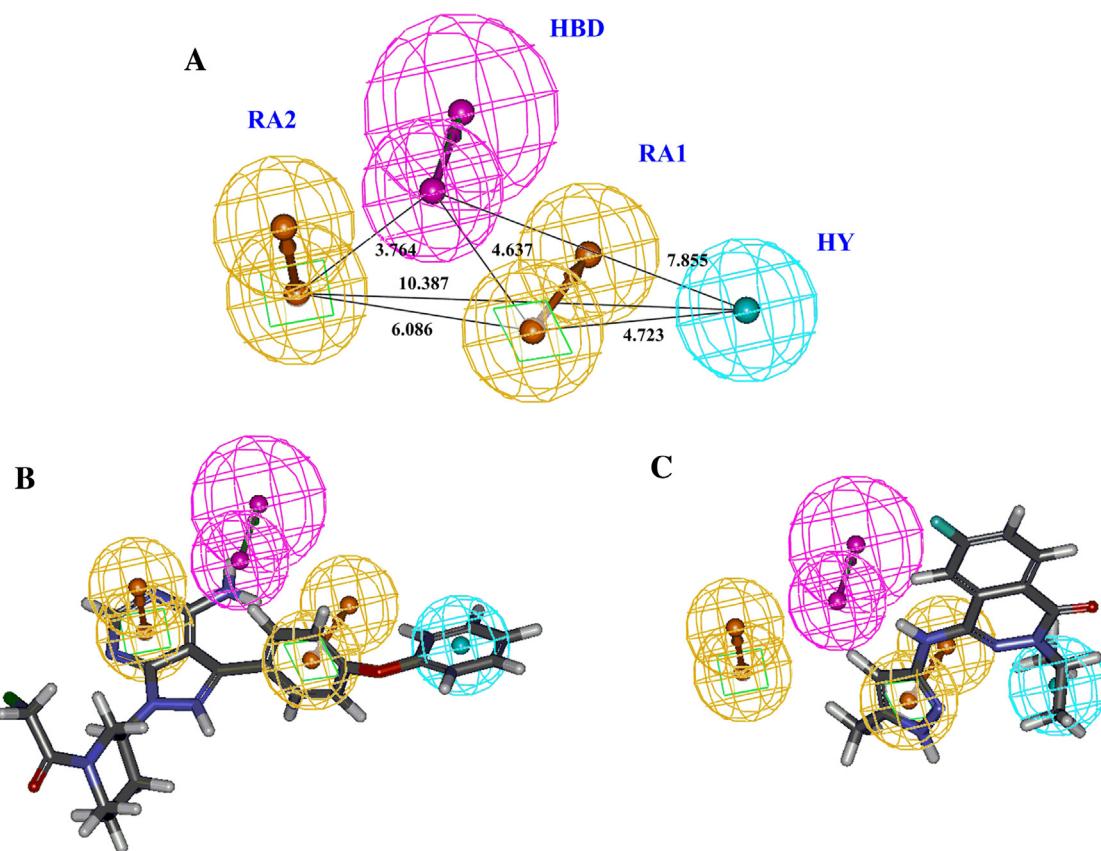


Fig. 6. 3D arrangements of the best ligand-based pharmacophore model, Hypo1. (A) 3D spatial relationship and geometric parameters of Hypo1. (B) Hypo1 mapped with the most active compound. (C) Hypo1 mapped with the least active compound. The pharmacophore features are colored with magenta (HBD), orange (RA), and cyan (HY). Excluded volumes were hidden.

large number of inactive compounds. A decoy database was built, which consisted of 30 active compounds picked from different references [30–34,44,45] and 2100 inactive compounds selected from ZINC database [46]. The following protocol was applied in inactive molecules selection. The Tanimoto similarity [47] between Zinc database molecules and each of 30 active compounds was calculated in DS2.5 based on the FCFP.6 fingerprints [48], and the molecules showing Tanimoto similarity coefficient larger than 0.3 were removed. Then, the compounds having similar physical properties with the active molecules were retained by the *Find Similar Molecules by Number* property protocol searching. Finally, a non-inhibitor set was extracted from 10,030 inactives remained using the *Find Diverse Molecules* protocol. The discriminating power of pharmacophore model was verified using the *Ligand Pharmacophore Mapping* protocol, the rigid and no features omitting option was adopted for the ligand-based pharmacophore mapping, while the flexible and one maximum omitted feature search method was applied for the structure-based model.

2.4. Pharmacophore-based virtual screening

The ChemBridge database comprising 987,668 compounds was applied to virtual screening in our work. Conformations of all the compounds were generated using the *Generate Conformations* protocol in DS2.5 with the Best option. The best ligand-based pharmacophore was subjected to screening the multi-conformation database using the *Ligand Pharmacophore Mapping* protocol with rigid and no features omitting method. The compounds with fit value greater than 8.0 were retained. Compared to the ligand-based pharmacophore screening, the parameter setting of structure-based pharmacophore approach was a bit different. In accordance

with the parameters used for decoy set validation, the flexible and one maximum feature omitting method was chosen. Besides, only the compounds matching the four features including HBA, HBD1, HY1, HY3 at least and displaying fit value greater than 2.0 were retained for the sake of cutting the false positive rate. After removal of the duplicated molecules, hit compounds from the two pharmacophore screening were combined for further docking analysis.

2.5. Docking studies

2.5.1. Docking software selection

Different docking procedures would give rise to absolutely different docking results even for same protein and ligand, thus the quality of docking software is one of the decisive factors in molecular docking studies. In this study, the performance of five frequently used docking programs, CDOCKER [35], LigandFit [35], GOLD [49], Surflex-dock [50], and Glide with standard precision (SP) mode [51] was compared by conducting the native docking. The eight ligands extracting from the cocrystal were redocked into the corresponding protein structures. The docking results were evaluated by calculating the RMSD between the poses generated and the original conformation in the cocrystal complex. The one giving the smallest average RMSD was chosen as the docking software and for further cross docking analysis.

2.5.2. Cross docking analysis

With the development of proteomics, many protein-ligand complexes have been reported for a designated target. The choice of appropriate protein structure for docking study has always appeared to be an intractable task. Cross docking [52] is a popular way to solve this problem. The docking program selected in the

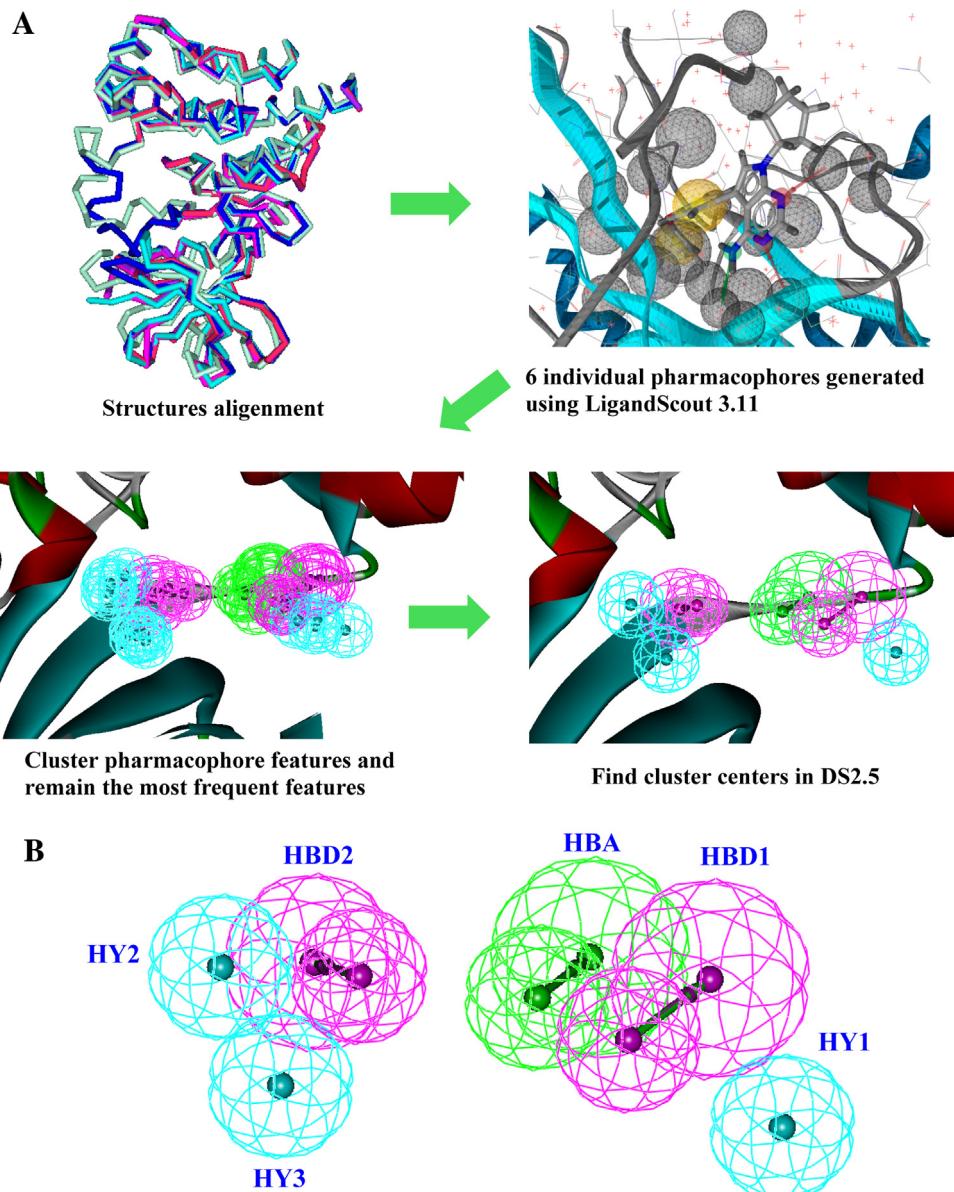


Fig. 7. (A) The basic flowchart of structure-based pharmacophore generation. (B) 3D arrangement of the modified pharmacophore, HypoS. Excluded volumes were hidden.

native docking was used and six ligands were docked into all the protein structures except the native one. The results were calculated like the native docking and the protein structure with the smallest average RMSD was selected as the working one.

2.5.3. Cascade docking

The purpose of cascade docking is to observe the interaction between the potential hits and the protein. The protein structure and the docking program selected by cross docking and native docking were employed to cascade docking study. All compounds passing the pharmacophore-based screening were docked and the compounds with good interaction patterns were retained for further analysis.

2.6. Molecular dynamics simulation

MD simulation was conducted to further confirm the binding mode determined from molecular docking using GROMACS 4.6.5 software package [53]. In the present work, the missed residues

were first repaired to ensure protein integrity. The topology files of protein and ligand were generated using pdb2gmx and the online PRODRUG server, respectively [54], and the GROMOS96 43a1 force field [55] was used. After files merging, a protein-centered regular dodecahedron box was defined with a distance of at least 10 Å between protein and the edge of the box. Then the protein-ligand complex was solvated with spc216 water and seven sodium ions were added to keep the system neutralization. The particle-mesh-Ewald (PME) method [56] was used for calculation of long-range electrostatic interactions and all bond lengths were constrained by LINCS algorithm [57]. A 50,000 steps energy minimization was performed with the steepest descent method followed by 200 picoseconds (ps) NVT and NPT equilibration phases successively. For the production phase, constraints were released and 15 nanoseconds (ns) MD simulation with a time-step of 2 femtoseconds (fs) was conducted under the equilibrated physical conditions of 300 K and 1.0 bar. Trajectories were collected and saved as individual file every 1 ps for subsequent analysis.

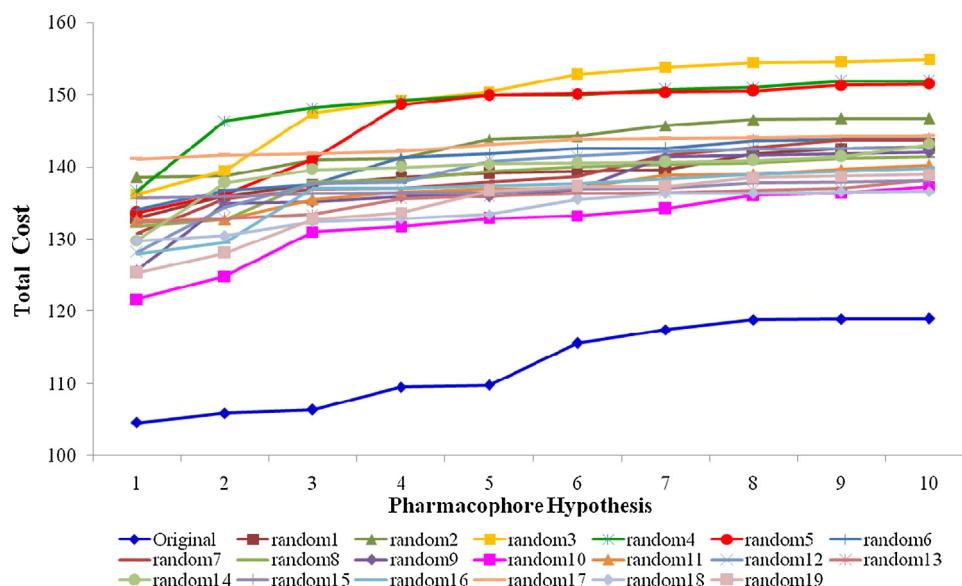


Fig. 8. Results of Fisher's randomization test for 95% confidence level.

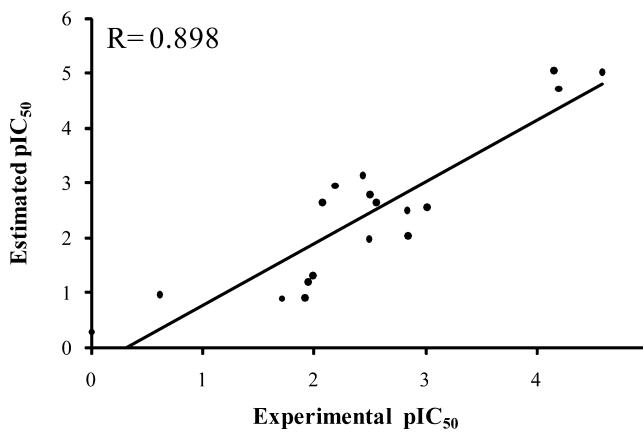


Fig. 9. Correlation (R) graph between the experimental activity and the predicted activity by Hypo1 for the test set.

3. Results and discussion

3.1. Generation of ligand-based pharmacophore

The pharmacophore model was constructed based on the training set we collected. The overall workflow of ligand-based pharmacophore generation was presented in Fig. 5.

Top 10 hypotheses were exported finally using 3D QSAR pharmacophore Generation protocol implemented in DS2.5. The result was shown in Table 1. Hypo1 with the highest cost difference (97.12), lowest RMSD (0.75), and best correlation coefficient (0.97) was considered as the best one. It consisted of four features: one HBD, one HY, two RA as well as four excluded volumes (Fig. 6A). The most active compound mapped very well with Hypo1 (Fig. 6B), while the least potent compound overlapped badly with it (Fig. 6C). The estimated activity of the training set and the errors were displayed in Table 2.

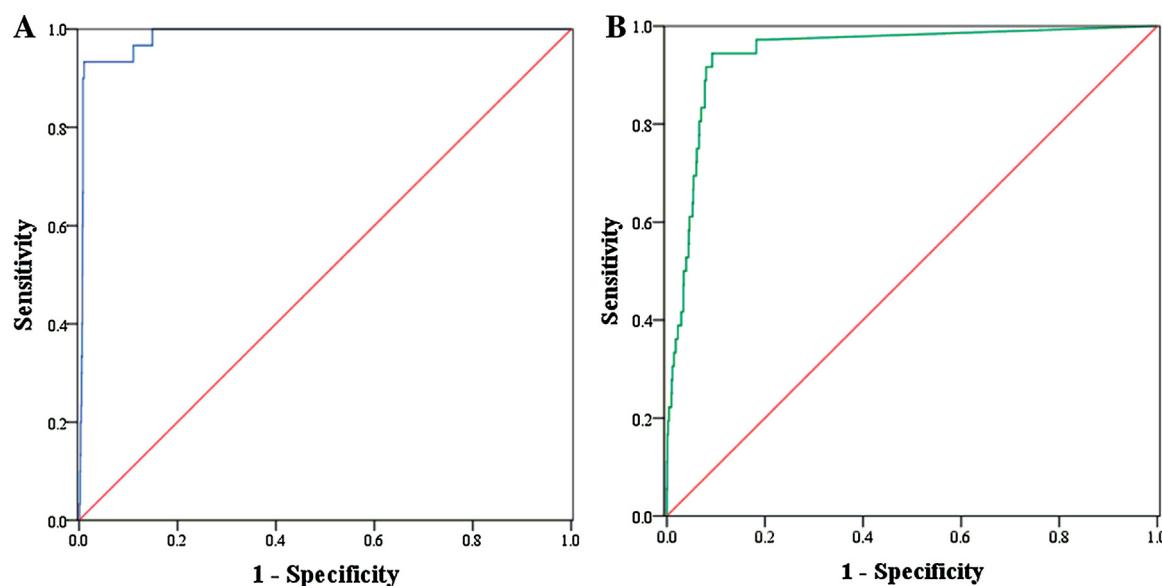


Fig. 10. ROC curves of the decoy set. (A) ROC curve obtained by Hypo1. (B) ROC curve obtained by Hypo5.

Table 3

RMSD of protein-ligand complexes used for native docking.

PDB (complex)	CDOCKER	LigandFit	GOLD	Surflex-Dock	Glide (SP)
3GEN	0.2567	0.3671	0.3374	0.5547	1.2907
3K54	1.0592	0.6169	1.1577	2.5984	0.7963
3OCS	1.0082	0.2146	0.4602	0.7332	1.2063
3PJ1	1.6826	1.8653	11.0924	2.5355	1.9593
3PIX	0.9077	8.0538	2.1161	1.4900	1.1274
3PIY	0.6866	6.2395	5.9909	1.5996	0.4177
3PIZ	1.4018	3.7818	0.8131	1.6543	1.7208
3PJ3	0.4107	0.4334	0.5097	0.7705	0.8612
av ^a	0.9267	2.6970	2.8100	1.4920	1.1720
std ^b	0.4772	3.0270	3.8330	0.7861	0.4990

^a Average RMSD values of native ligand poses referring to their native poses.^b Standard deviation of these RMSD values.**Table 4**

RMSD of ligands in cross docking using CDOCKER.

Protein	3GEN	3K54	3PJ1	3PIX	3PIY	3PIZ	3PJ3	av ^a	Std ^b
3GEN	0.2567	2.6181	1.2864	0.9356	1.6413	2.2341	1.4654	1.491	0.7874
3K54	2.0227	1.0592	1.7767	0.8820	4.4965	9.349	1.3637	2.993	3.0530
3PJ1	1.0448	2.0294	1.6826	0.7612	4.3366	8.7033	12.249	4.401	4.4280
3PIX	1.9359	0.7518	1.7146	0.9077	1.524	1.2967	3.2017	1.619	0.8150
3PIY	1.2198	1.1277	0.9744	0.8434	0.6866	1.0361	1.3884	1.039	0.2343
3PIZ	1.0731	1.0872	1.3960	0.7153	2.0859	1.4018	12.026	2.827	4.0790
3PJ3	1.0312	2.1806	1.3964	1.1472	9.2998	1.1811	0.4107	2.378	3.0970

^a Average RMSD values of native ligand poses referring to their native poses.^b Standard deviation of these RMSD values.

3.2. Generation of structure-based pharmacophore

The basic flowchart of structure-based pharmacophore generation was depicted in Fig. 7A. Six pharmacophore models were generated based on their own protein-ligand complexes. After the combination of six models, a comprehensive model was established by keeping back the features and excluded volume spheres showing the occurrence higher than four times. Refinement of the model by modifying the radius of the features to default values embedded in DS2.5. Pharmacophore features of the model were directed towards important amino acids previously reported like Thr474 and Met477 [34,58]. Hence, the pharmacophore model HypoS (Fig. 7B) could be considered as a good model to discover potent BTK inhibitors.

3.3. Pharmacophore model evaluation

3.3.1. Fischer's randomization method

Fisher's randomization test was adopted to evaluate the statistical significance of Hypo1. Nineteen random spreadsheets were created in order to achieve the confidence level of 95%. The results were shown in Fig. 8. The total cost values of the randomly 19 pharmacophore models were much higher than the original one, which clearly indicated there was a strong significance of our constructed model Hypo1.

3.3.2. Test set validation

The method is applied to examining whether the pharmacophore model is capable of predicting the activities of external compounds. Here, a test set consisting of 18 structurally diverse compounds outside the training set were mapped by Hypo1, which gave a good correlation coefficient of 0.898 (Fig. 9), confirming the universality of Hypo1.

3.3.3. Decoy set method

Discovery of a small fraction of bioactive compounds from a large database is the most important point of our work, thus the pharmacophore model must have the ability to distinguish the actives from numerous background inactives. The discriminating power of a pharmacophore model was quantitatively characterized by the AUC area under a receiver operating characteristics (ROC) curve. In our study, the AUC area of Hypo1 and HypoS was 0.986 and 0.945, respectively (Fig. 10), which further authenticated the validity of the two pharmacophore models for identification of novel BTK inhibitors.

3.4. Native and cross docking results

According to the results of native docking (Table 3), CDOCKER and Glide outperformed the other three programs with all six RMSD values lower than 2.0. It also showed those top two programs exhibited similar average RMSD and standard derivation, although

Table 5

RMSD of ligands in cross docking using Glide (SP).

Protein	3GEN	3K54	3PJ1	3PIX	3PIY	3PIZ	3PJ3	av ^a	Std ^b
3GEN	1.2907	3.7624	2.1139	1.0765	2.8000	2.9060	1.6643	2.230	0.9727
3K54	1.5232	0.7963	1.5845	1.5121	3.3599	3.3269	1.8286	1.990	0.9769
3PJ1	1.0393	1.6546	1.9593	1.4070	3.0109	2.9734	1.4384	1.926	0.7791
3PIX	1.2749	3.0612	1.0285	1.1274	1.3525	1.9992	2.3700	1.745	0.7591
3PIY	1.4692	1.7974	2.2813	1.1168	0.4177	0.9269	2.0827	1.441	0.6661
3PIZ	1.7867	1.6204	2.1346	1.3012	2.0742	1.7208	2.4635	1.871	0.3827
3PJ3	1.9623	1.6277	2.0599	1.3123	2.9536	3.0035	0.8612	1.969	0.7982

^a Average RMSD values of native ligand poses referring to their native poses.^b Standard deviation of these RMSD values.

Table 6

Structures and docking scores of the final 8 lead compounds.

ZINC database ID	Structure	CDOCKER_ENERGY(kcal/mol)	Glide score
ZINC00098100(A)		-43.95	-9.648
ZINC04759768(B)		-42.18	-9.020
ZINC12054044(C)		-43.56	-8.699
ZINC14985329(D)		-42.19	-8.597
ZINC11870137(E)		-42.40	-8.384
ZINC72117812(F)		-48.46	-8.069
ZINC14887663(G)		-43.70	-7.967
ZINC01192184(H)		-42.02	-7.935

the truth was that CDOCKER was little better than Glide. Therefore, a sequential docking methodology was developed using both CDOCKER and Glide for docking screening. During the cross docking validation, the cocrystallized ligands were docked into the rest of the six proteins using CDOCKER and Glide, respectively. Obviously, docking using 3PIY with different ligands performed much better than other proteins by CDOCKER, as depicted in Table 4, the protein can predict the active conformation of every ligand successfully (RMSD < 2.0). Undoubtedly, 3PIY was selected as the cooperative protein with CDOCKER in docking-based virtual screening. When referring to Glide, 3PIY was also favorable with the smallest average RMSD, though the standard derivation was a bit higher than docking using 3PIZ. Considering the success ratio (RMSD < 2.0), 3PIY can predict four ligands successfully, while 3PIZ was only three (Table 5). Therefore, we also chose 3PIY as the receptor in Glide docking.

3.5. Virtual screening results

Virtual screening is a versatile technique to identify novel and potent lead compounds for a particular target. The validated pharmacophore models Hypo1 and HypoS were used as queries to search the ChemBridge database. A hit list of 105,070 and 33,283 compounds were remained after pharmacophore mapping, respectively. Merging the compounds obtained and removing the reduplicate molecules, 110,526 compounds were retained for further docking study.

3.6. Docking analysis

The docking protocol validated before was employed to refine the retrieved hits and reduce the false positive rate through receptor-ligand interaction analysis. A total of 110,526 compounds

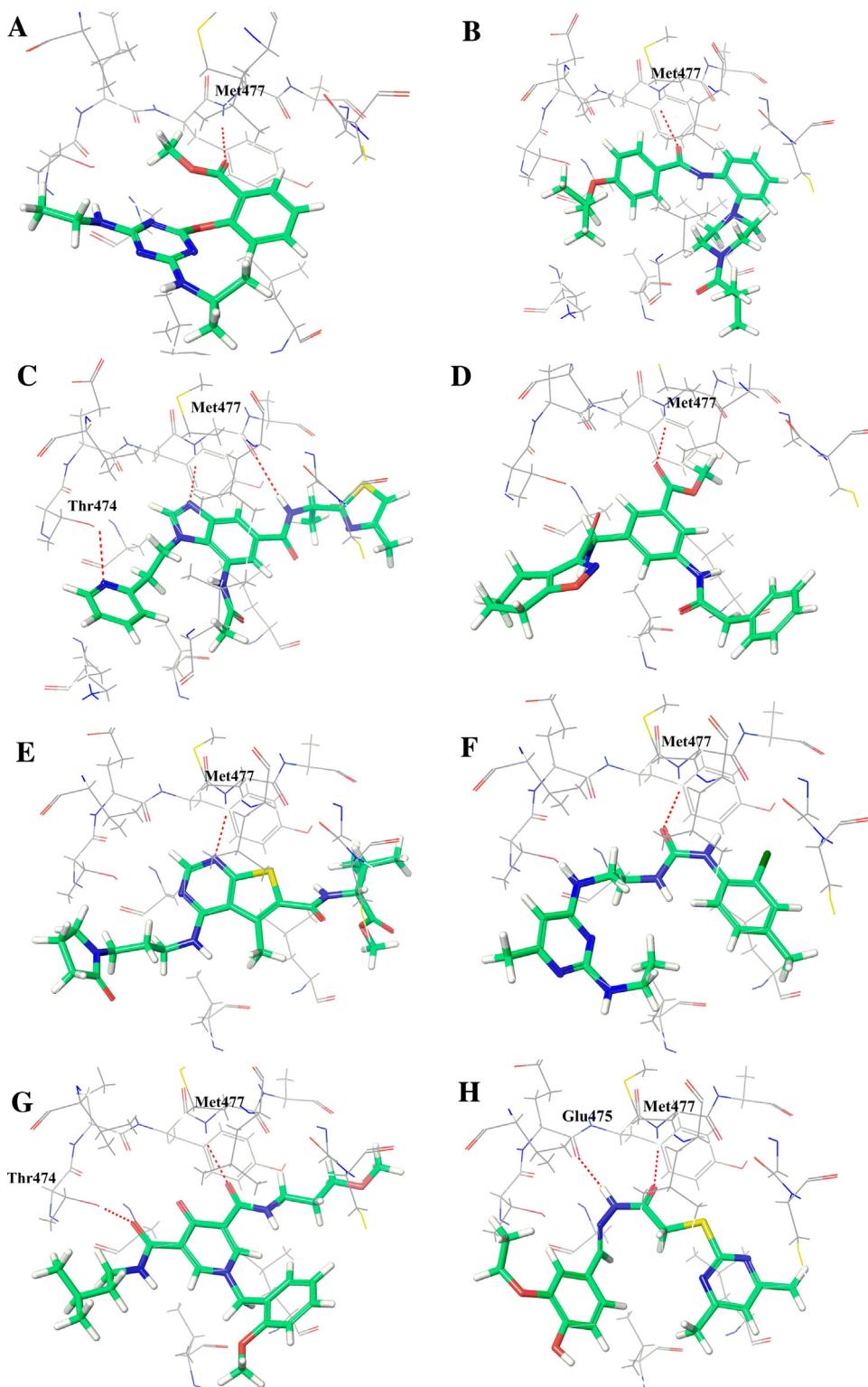


Fig. 11. The binding modes of the final eight hits based on docking analysis. The hydrogen bonds are labeled in red dash line. The hits are represented in sticks and colored by elements. The residues in 5 Å are showed in sticks and colored by elements (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

derived from virtual screening were docked to protein 3PIY binding site using CDOCKER. Hits with negative interaction energy above 42 and H-bond interaction with Met477 were selected. Next, these 421 hits were docked to 3PIY again by Glide to confirm the interaction patterns, among which 176 compounds retained hydrogen bond with Met477 and showed a glide score larger than 7.0, and

were therefore considered as actives. The reference negative interaction energy and glide score were selected based on the threshold of protein 3PIY and its native ligand redocking result. The ADMET properties prediction was further determined and 55 compounds conflicting with the criteria were rejected. The remaining 121 compounds were classified in eight categories based on their structural

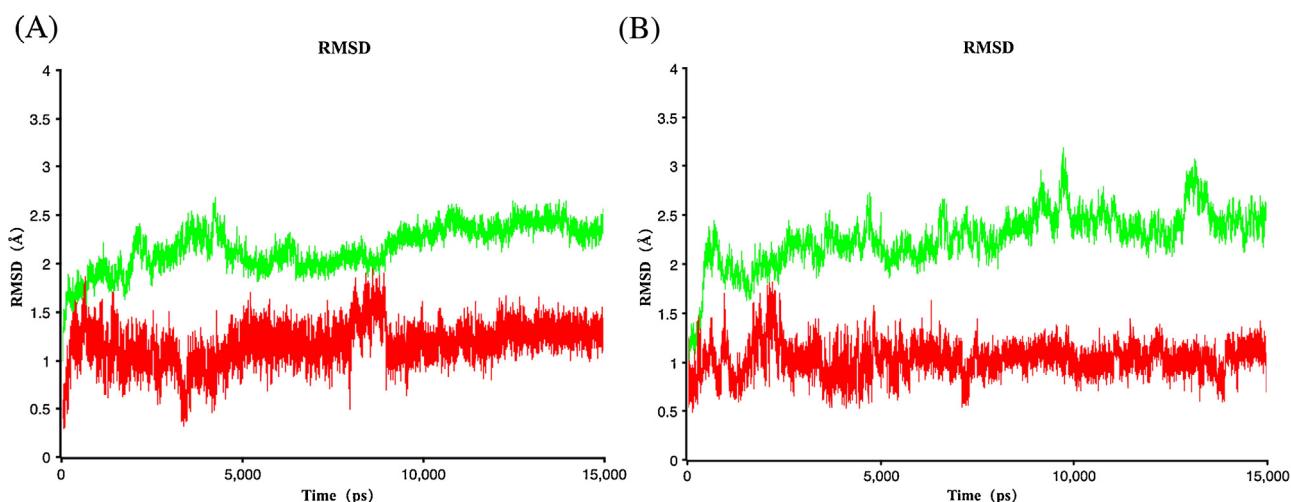


Fig. 12. RMSD plot of ZINC00098100 and ZINC04759768 during the 15 ns MD simulation. (A) The RMSD of backbone atoms of the protein and ligand ZINC00098100. (B) The RMSD of backbone atoms of the protein and ligand ZINC04759768. The RMSD profile for backbone atoms of protein and ligand is colored in green and red, respectively (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

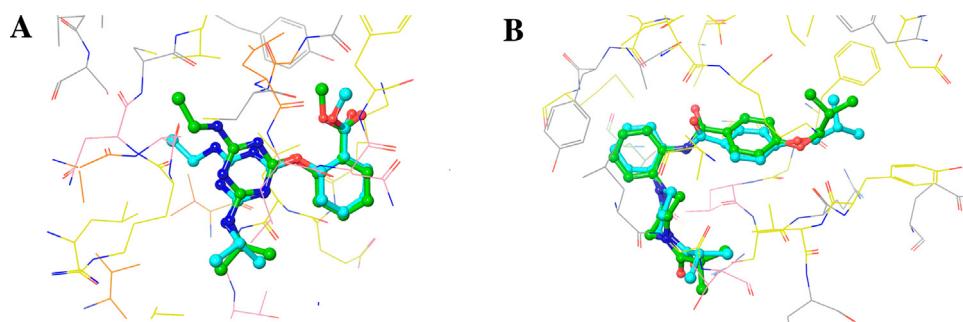


Fig. 13. The superimposition of the initial and final structures of ZINC00098100 (A) and ZINC04759768 (B) are obtained after 15 ns MD simulation. The initial structures of ZINC00098100 and ZINC04759768 are colored in green, and the final structures of ZINC00098100 and ZINC04759768 are colored in blue (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

similarity using the *Cluster Ligands* module in DS2.5. Eventually, the top hit from each category with novel chemical scaffolds was revealed for further *in vitro* study (Table 6). The docking poses of the top eight hits were shown in Fig. 11. Similar binding modes were obtained in the complexes and all of hits formed the hydrogen bond with the key residue Met477. Apart from engaging hydrogen bond with Met477, ZINC12054044, ZINC14887663 and ZINC01192184 also showed an additional hydrogen bond with residue Thr474 or Glu475 nearby.

3.7. Molecular dynamics simulation study

The binding stability of the two molecules with the top glide scores was further evaluated using 15 ns MD simulation. The results demonstrated that the two representative hits reached equilibrium in a few ps after MD simulation and displayed small fluctuation in the whole process, which indicated that they could bind stably to the protein (Fig. 12). The superimposition was made between the structure extracting from the last ps trajectory and the original structure before the MD simulation (Fig. 13). Obviously, both of the two hits before and after the MD simulation exhibited the similar binding conformation, suggesting the presumed active conformations resulting from the docking experiments were reliable. In summary, the MD simulation study provided substantial evidence of the reliability of molecular docking and the binding stability of protein-ligand complex.

4. Conclusions

Clinical studies have shown BTK is an effective target in the treatment of B-cell malignancies. The purpose of our work is to discover novel and potent BTK inhibitors by *in-silico* screening method. The ligand- and structure-based pharmacophore models were developed successfully. After systematically validation, Hypo1 and HypoS were considered as reliable models to screen the ChemBridge database. The optimal docking program and protein structure were selected carefully via native docking and cross docking. All compounds obtained from pharmacophore-based virtual screening were filtered by cascade docking studies to improve the reliability and accuracy of virtual screening. Eight representative hits were picked up for further *in vitro* BTK inhibition study according to molecular scaffold dissimilarity and ADMET properties prediction, and two of them were selected as an example to judge the protein-ligand stability by MD simulation study.

In conclusion, we integrated a parallel pharmacophore-based virtual screening with molecular docking, ADMET analysis and MD simulation for identification of novel BTK inhibitors. We believe our approach will accelerate the discovery of potent BTK inhibitors with novel scaffolds.

Acknowledgments

We sincerely thank Dr. Cheng Fang for his professional advices in drafting the manuscript. This work was supported by grants

from the National Natural Science Foundation of China (no. 21472244), the Natural Science Foundation of Jiangsu Province (no. BK2011626), the Six Major Talent Peak Project of Jiangsu Province (no. 2011-YY-004), and the Fundamental Research Funds for the Central Universities (no. 2015021).

References

- [1] S.G. Fisher, R.I. Fisher, The epidemiology of non-Hodgkin's lymphoma, *Oncogene* 23 (2004) 6524–6534.
- [2] R. Siegel, D. Naishadham, A. Jemal, Cancer statistics, CA. Cancer J. Clin. 63 (2013) 11–30.
- [3] J.M. Dalport, S.B. Gauld, K.T. Merrell, D. Mills, A.E. Pugh-Bernard, J. Cambier, B cell antigen receptor signaling 101, *Mol. Immunol.* 41 (2004) 599–613.
- [4] F.K. Stevenson, S. Krysov, A.J. Davies, A.J. Steele, G. Packham, B-cell receptor signaling in chronic lymphocytic leukemia, *Blood* 118 (2011) 4313–4320.
- [5] F. Suarez, O. Lortholary, O. Hermine, M. Leclerc, Infection-associated lymphomas derived from marginal zone B cells: a model of antigen-driven lymphoproliferation, *Blood* 107 (2006) 3034–3044.
- [6] R. Küppers, Mechanisms of B-cell lymphoma pathogenesis, *Nat. Rev. Cancer* 5 (2005) 251–262.
- [7] T.J. Kipps, The B-cell receptor and ZAP-70 in chronic lymphocytic leukemia, *Best Pract. Res. Clin. Haematol.* 20 (2007) 415–424.
- [8] C. Pighi, T.L. Gu, I. Dalai, S. Barbi, C. Parolini, A. Bertolaso, et al., Phosphoproteomic analysis of mantle cell lymphoma cells suggests a pro-survival role of B-cell receptor signaling, *Cell Oncol. (Dordr.)* 34 (2011) 141–153.
- [9] D. Vetrici, I. Voiechovsky, P. Sideras, I. Holland, A. Davies, F. Flinter, The gene involved in X-linked agammaglobulinemia is a member of the *src* family of protein-tyrosine kinases, *Nature* 361 (1993) 226–233.
- [10] C. Smith, B. Baskin, P. Humire-Greiff, J.N. Zhou, P. Olsson, H. Maniar, Expression of Bruton's agammaglobulinemia tyrosine kinase gene, BTK, is selectively down-regulated in T lymphocytes and plasma cells, *J. Immunol.* 152 (1994) 557–565.
- [11] A.J. Mohamed, L. Yu, C.M. Bäckesjö, L. Vargas, R. Faryal, A. Aints, Bruton's tyrosine kinase, *Immunol. Rev.* 228 (2009) 58–73.
- [12] H. Park, M.I. Wahl, D.E. Afar, C.W. Turck, D.J. Rawlings, C. Tam, et al., Regulation of Btk function by a major autophosphorylation site within the SH3 domain, *Immunity* 4 (1996) 515–525.
- [13] M. Cinar, F. Hamedani, Z. Mo, B. Cinar, H.M. Amin, S. Alkan, Bruton tyrosine kinase is commonly overexpressed in mantle cell lymphoma and its attenuation by Ibrutinib induces apoptosis, *Leuk. Res.* 37 (2013) 1271–1277.
- [14] R.E. Davis, V.N. Ngo, G. Lenz, P. Tolar, R.M. Young, P.B. Romesser, et al., Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma, *Nature* 463 (2010) 88–92.
- [15] S.E. Herman, A.L. Gordon, E. Hertlein, A. Ramanunni, X. Zhang, S. Jaglowski, et al., Bruton tyrosine kinase represents a promising therapeutic target for treatment of chronic lymphocytic leukemia and is effectively targeted by PCI-32765, *Blood* 117 (2011) 6287–6296.
- [16] Z. Pan, H. Scheerens, S.J. Li, B.E. Schultz, P.A. Sprengeler, L.C. Burrill, Discovery of selective irreversible inhibitors for Bruton's tyrosine kinase, *ChemMedChem* 2 (2007) 58–61.
- [17] L.A. Honigberg, A.M. Smith, M. Sirisawad, E. Verner, D. Loury, B. Chang, et al., The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy, *Proc. Natl. Acad. Sci. U.S.A.* 107 (2010) 13075–13080.
- [18] J.A. Woyach, R.R. Furman, T.M. Liu, H.G. Ozer, M. Zapata, A.S. Ruppert, Resistance mechanisms for the Bruton's tyrosine kinase inhibitor ibrutinib, *N. Engl. J. Med.* 370 (2014) 2286–2294.
- [19] D. Chiron, M. Di Liberto, P. Martin, X. Huang, J. Sharman, P. Blecua, Cell-cycle reprogramming for PI3K inhibition overrides a relapse-specific C481S BTK mutation revealed by longitudinal functional genomics in mantle cell lymphoma, *Cancer Discov.* 4 (2014) 1022–1035.
- [20] A.K. Gupta, K. Varshney, A.K. Saxena, Toward the identification of a reliable 3D QSAR pharmacophore model for the CCK2 receptor antagonism, *J. Chem. Inf. Model.* 52 (2012) 1376–1390.
- [21] M.A. Khanfar, M.O. Taha, Elaborate ligand-based modeling coupled with multiple linear regression and k nearest neighbor QSAR analyses unveiled new nanomolar mTOR inhibitors, *J. Chem. Inf. Model.* 53 (2013) 2587–2612.
- [22] J. Wang, L. Chen, S.H. Sinha, Z. Liang, H. Chai, S. Muniyan, Pharmacophore-based virtual screening and biological evaluation of small molecule inhibitors for protein arginine methylation, *J. Med. Chem.* 55 (2012) 7978–7987.
- [23] H.J. Kim, M.R. Doddareddy, H. Choo, Y.S. Cho, K.T. No, W.K. Park, New serotonin 5-HT6 ligands from common feature pharmacophore hypotheses, *J. Chem. Inf. Model.* 48 (2008) 197–206.
- [24] J.X. Ren, L.L. Li, J. Zou, L. Yang, J.L. Yang, S.Y. Yang, Pharmacophore modeling and virtual screening for the discovery of new transforming growth factor-beta type I receptor (ALK5) inhibitors, *Eur. J. Med. Chem.* 44 (2009) 4259–4265.
- [25] Z. Chen, G. Tian, Z. Wang, H. Jiang, J. Shen, W. Zhu, Multiple pharmacophore models combined with molecular docking: a reliable way for efficiently identifying novel PDE4 inhibitors with high structural diversity, *J. Chem. Inf. Model.* 50 (2010) 615–625.
- [26] R. Kurczab, A.J. Bojarski, New strategy for receptor-based pharmacophore query construction: a case study for 5-HT(7) receptor ligands, *J. Chem. Inf. Model.* 53 (2013) 3233–3243.
- [27] S.Y. Yang, Pharmacophore modeling and applications in drug discovery: challenges and recent advances, *Drug Discov. Today.* 15 (2010) 444–450.
- [28] M. Arooj, S. Sakkiah, S. Kim, V. Arulalapperumal, K.W. Lee, A combination of receptor-based pharmacophore modeling & QM techniques for identification of human chymase inhibitors, *PLoS One* 8 (2013) e63030.
- [29] G. Zhang, H. Ge, Q. Gu, J. Xu, Predicting hiCE inhibitors based upon pharmacophore models derived from the receptor and its ligands, *Sci. China Chem.* 56 (2013) 1402–1412.
- [30] R.J. Billedeau, R.K. Kondu, F.J. Lopez-Tapia, Y. Lou, T.D. Owens, Y. Qian, Inhibitors of bruton's tyrosine kinase, WO2012156334, 2012.
- [31] N.J. Dewdney, J. Kennedy-Smith, R.K. Kondu, B.E. Loc, Y. Lou, J. McIntosh, Inhibitors of Bruton's tyrosine kinase, US20090306041, 2009.
- [32] D.M. Goldstein, M. Rueth, Methods of inhibiting BTK and SYK protein kinases, US20070209195, 2007.
- [33] N.J. Dewdney, R.K. Kondu, Y. Lou, M. Soth, T. Gabriel, Btk protein kinase inhibitors, US20090105209, 2009.
- [34] A. Kuglstatter, A. Wong, S. Tsing, S.W. Lee, Y. Lou, A.G. Villasenor, Insights into the conformational flexibility of Bruton's tyrosine kinase from multiple ligand complex structures, *Protein Sci.* 20 (2011) 428–436.
- [35] Discovery Studio Version 2.5, Accelrys Inc., San Diego, CA, 2012.
- [36] B.R. Brooks, R.E. Brucoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM. A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.* 4 (1983) 187–217.
- [37] A. Smellie, S.L. Teig, P. Towbin, Poling: promoting conformational variation, *J. Comput. Chem.* 16 (1995) 171–187.
- [38] D. Dube, V. Periwal, M. Kumar, S. Sharma, T.P. Singh, P. Kaur, 3D-QSAR based pharmacophore modeling and virtual screening for identification of novel pteridine reductase inhibitors, *J. Mol. Model.* 18 (2012) 1701–1711.
- [39] (<http://www.rcsb.org/pdb/home/home.do>).
- [40] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, The protein data bank, *Nucleic Acids Res.* 28 (2000) 235–242.
- [41] ligandscout 3.11, Inte:Ligand GmbH, Clemens-Maria-Hofbauer-G.6, 2344, Maria Enzersdorf, Austria, 2010.
- [42] G. Wolber, T. Langer, LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters, *J. Chem. Inf. Model.* 45 (2005) 160–169.
- [43] E. Neria, S. Fischer, M. Karplus, Simulation of activation free energies in molecular systems, *J. Chem. Phys.* 105 (1996) 1902–1921.
- [44] G. Adjabeng, E. Baum, N. Bifulco, R.G. Davis-Ward, S.H. Dickerson, K.H. Donaldson, Thiazole sulfonamide and oxazole sulfonamide kinase inhibitors, US20110319392, 2011.
- [45] C. Liu, K. Leftheris, A.J. Tebben, Fused heterocyclic compounds useful as kinase modulators, WO2010011837, 2010.
- [46] J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman, ZINC: a free tool to discover chemistry for biology, *J. Chem. Inf. Model.* 52 (2012) 1757–1768.
- [47] E.J. Martin, J.M. Blaney, M.A. Siani, D.C. Spellmeyer, A.K. Wong, W.H. Moos, Measuring diversity: experimental design of combinatorial libraries for drug discovery, *J. Med. Chem.* 38 (1995) 1431–1436.
- [48] D. Rogers, R.D. Brown, M. Hahn, Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up, *J. Biomol. Screen.* 10 (2005) 682–686.
- [49] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.* 267 (1997) 727–748.
- [50] A.N. Jain, Surfflex: fully automatic flexible molecular docking using a molecular similarity-based search engine, *J. Med. Chem.* 46 (2003) 499–511.
- [51] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, Glide: a new approach for rapid, accurate docking and scoring, 1. Method and assessment of docking accuracy, *J. Med. Chem.* 47 (2004) 1739–1749.
- [52] R. Thilagavathi, R.L. Mancera, Ligand-protein cross-docking with water molecules, *J. Chem. Inf. Model.* 50 (2010) 415–421.
- [53] H.J. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun.* 91 (1995) 43–56.
- [54] A.W. Schuettelkopf, D.M. Van Aalten, PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, *Acta Crystallogr. D Biol. Crystallogr.* 60 (2004) 1355–1363.
- [55] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, *J. Comput. Chem.* 26 (2005) 1701–1718.
- [56] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089.
- [57] B. Hess, H. Bekker, H.J. Berendsen, J.G. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [58] D.J. Marcotte, Y.T. Liu, R.M. Arduini, C.A. Hession, K. Miatkowski, C.P. Wildes, et al., Structures of human Bruton's tyrosine kinase in active and inactive conformations suggest a mechanism of activation for TEC family kinases, *Protein Sci.* 19 (2010) 429–439.