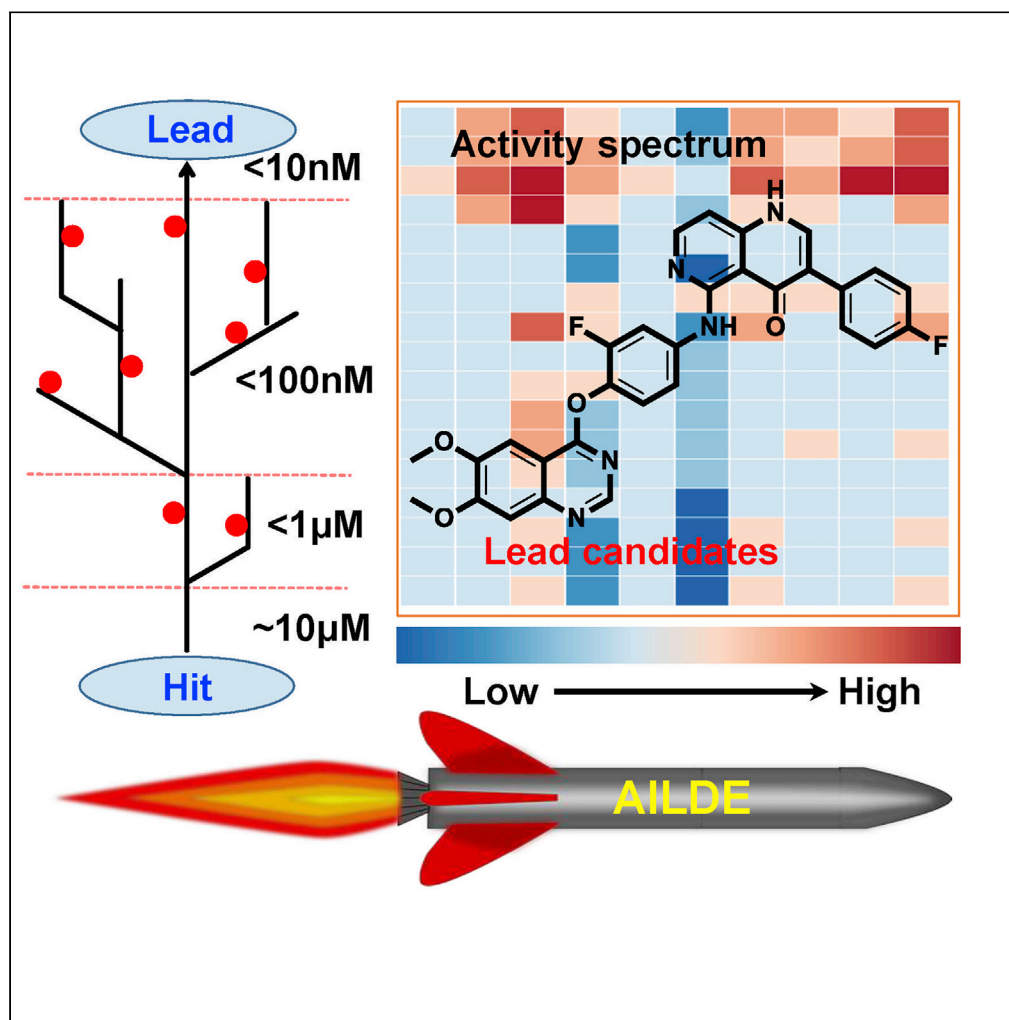## Article

Auto *In Silico* Ligand Directing Evolution to Facilitate the Rapid and Efficient Discovery of Drug Lead

Fengxu Wu,
Linsheng Zhuo,
Fan Wang, Wei
Huang, Gefei Hao,
Guangfu Yang

weihuangwuhan@126.com
(W.H.)
gefei_hao@foxmail.com (G.H.)
gfyang@mail.ccnu.edu.cn
(G.Y.)

HIGHLIGHTS
AILDE was developed for
the rapid identification of
drug leads

A potent drug lead
targeted to c-Met was
found by synthesizing only
eight compounds

iScience

Article

Auto *In Silico* Ligand Directing Evolution to Facilitate the Rapid and Efficient Discovery of Drug Lead

Fengxu Wu,[1,2,4] Linsheng Zhuo,[1,2,4] Fan Wang,[1,2] Wei Huang,[1,2,*] Gefei Hao,[1,2,5,*] and Guangfu Yang[1,2,3,*]

SUMMARY

Motivated by the growing demand for reducing the chemical optimization burden of H2L, we developed auto *in silico* ligand directing evolution (AILDE, http://chemyang.ccnu.edu.cn/ccb/server/AILDE), an efficient and general approach for the rapid identification of drug leads in accessible chemical space. This computational strategy relies on minor chemical modifications on the scaffold of a hit compound, and it is primarily intended for identifying new lead compounds with minimal losses or, in some cases, even increases in ligand efficiency. We also described how AILDE greatly reduces the chemical optimization burden in the design of mesenchymal-epithelial transition factor (c-Met) kinase inhibitors. We only synthesized eight compounds and found highly efficient compound 5g, which showed an ~1,000-fold improvement in *in vitro* activity compared with the hit compound. 5g also displayed excellent *in vivo* antitumor efficacy as a drug lead. We believe that AILDE may be applied to a large number of studies for rapid design and identification of drug leads.

INTRODUCTION

Drug discovery is a time-consuming and resource-intensive process that requires large investments by pharmaceutical industry corporations as well as governments (Paul et al., 2010). Developing a new drug from the original idea to the launch of a finished product is usually a complicated process that takes 10–15 years and costs US$0.5 billion to US$2.6 billion (Hughes et al., 2011; Zhavoronkov et al., 2019). Considerations of all possible organic molecules has led to the concept of "chemical space," which is considered to include at least 10^{60} molecules (Reymond et al., 2010). Only a small fraction of chemical space can be developed into final drugs (Medina-Franco et al., 2008). Exploring chemical space through synthesis and screening is the conventional method used when searching for new drugs, but it is also a large burden in drug discovery. Consequently, the development of strategies to efficiently explore chemical space for drug discovery is an important scientific issue.

Hit-to-lead (H2L) optimization while working in bioactive and accessible chemical space is crucial to the success of drug discovery and is thus currently a cutting edge research area in medicinal chemistry (Rebecca and Benoit, 2004; Zhao, 2007). H2L optimization involves evaluation and structural optimization of a hit compound to identify promising lead compounds (Keserű and Makara, 2006; Orita et al., 2011). The key to starting H2L optimization is to identify an appropriate scaffold with the desired basic bioactivity that can provide a sufficient number of high-quality analogues. Owing to the development of new technologies in drug design, such as virtual screening, combinatorial chemistry, and high-throughput screening, discovering an appropriate scaffold with the desired basic bioactivity is not difficult. However, 5,000–10,000 compounds still have to be synthesized in the laboratory and screened in the H2L stage to identify a clinical lead (Moridani and Harirforoosh, 2014). The large number of analogues that need to be synthesized and validated experimentally in this step make H2L, a complex multiple-property optimization process, rate-limiting in drug discovery (Hoffer et al., 2018).

To overcome this bottleneck, rational design approaches are widely used in the H2L stage of drug discovery, and many successful applications have been reported (Chéron et al., 2016; Erlanson et al., 2000, 2003; Hajduk et al., 1997; Hochguertel et al., 2002; Lin et al., 2019; Nienaber et al., 2000; Oltersdorf et al., 2005; Schulz et al., 2018; Vinkers et al., 2003). One major trend is fragment-based drug discovery (Erlanson et al., 2016;

Hao et al., 2016). Graeme et al. developed a fragment-based method to optimize the pyrazole carboxylic ester scaffold with low affinity into a potent cyclic nucleotide phosphodiesterase IV (PDE4) inhibitor by synthesizing only 21 compounds (Card et al., 2005). Another example is the discovery of the selective B cell CLL/lymphoma 2 (BCL-2) drug ABT-199 by screening fragments on two distinct regions of BCL-2 (Souers et al., 2013). Artificial intelligence (AI) is another emerging trend. In a recent work, Zhavoronkov et al. developed a deep generative model for the rapid identification of potent discoidin domain receptor 1 (DDR1) kinase drug leads by synthesizing 40 compounds (Zhavoronkov et al., 2019). These rational design approaches are used to explore key interactions and to understand the relationships between structure and activity, and they are well suited for the discovery of new ligands with improved binding, selectivity, or other pharmacological properties. However, in the H2L stage, such approaches also encounter difficulties such as targeting unreachable chemical space, poor applicability of data-driven models, and inefficient deduction and decision-making on the next leads to be synthesized, which remain formidable challenges.

We developed an efficient and general approach called auto *in silico* ligand directing evolution (AILDE) for the rapid identification of drug leads in accessible chemical space. AILDE performs minor chemical modifications on the scaffold of a hit compound, and these modifications can result in minimal losses or, in some cases, even increases in ligand efficiency. Hence, this strategy can explore the chemical space around each hit in a series of compounds and drive the evolution of hit compounds into more "clinic-ready" lead structures. The deduction and decision-making on the leads, as well as the applicability of AILDE, have been rigorously validated on 19 drug targets with 157 ligands. The predicted binding affinities were linearly correlated (R^2 = 0.82) and within 0.31 kcal mol^{-1} on average of the experimental values. We also describe how this approach was applied to discover a potent anticancer drug lead (**5g**) with surprisingly high efficiency.

Owing to their pivotal roles in signal transduction and the regulation of a range of cellular activities, mesenchymal-epithelial transition factor (c-Met) has been established as a promising drug target for the treatment of cancer (Gharwan and Groninger, 2016; Gross et al., 2015; Wu et al., 2016). A large number of c-Met kinase inhibitors have been reported over the past few decades (Basilico et al., 2013; Comoglio et al., 2008; Cui, 2014; Cui et al., 2011; Ma et al., 2005; Martens et al., 2006). However, few examples of the discovery of c-Met kinase inhibitor leads have achieved highly efficient optimization using computational design, synthesis, *in vitro* and *in vivo* assays, and cocrystallization validation. We applied AILDE to c-Met and successfully discovered a potent drug lead (**5g**) by synthesizing only eight compounds. Two steps of ligand-directed evolution were performed. Compound **5g** ultimately showed an ~1,000-fold activity improvement in the enzyme-based assay (IC$_{50}$ = 9.7 nM) compared with **5a**. **5g** also exhibited potent *in vitro* inhibition in a cell-based assay (IC$_{50}$ = 47.3 nM). Moreover, **5g** induced dose-dependent tumor growth inhibition (TGI), with a minimum effective dose (MED/ED$_{50}$, 50% TGI) of ~8.3 mg/kg. At 25 mg/kg, **5g** showed significant *in vivo* antitumor efficacy (TGI of 82%). The binding mode and interactions seen in the cocrystal structure of compound **5i** (an analogue of **5g**) with c-Met were highly consistent with our predicted result, which confirmed the reliability of our strategy. AILDE may improve the efficiency and effectiveness of the initial stages of drug discovery. We also developed a web service to allow medicinal chemists to easily use AILDE.

RESULTS AND DISCUSSION

Small Group Library

The small group library includes 47 substituents that are from two fragment-based databases, PADFrag (http://chemyang.ccnu.edu.cn/ccb/database/PADFrag/) (Yang et al., 2018) and Molinspiration (https://www.molinspiration.com/). PADFrag is a searchable web-enabled resource that combines 1,652 FDA-approved drugs, 1,259 commercial pesticides, and 5,919 generated molecular fragments. It was designed for molecule design, and several functions are included in the server, such as viewing, sorting, and fragment extracting. Molinspiration offers a database of substituents and linkers obtained by substructure analysis of a collection of current drugs, development drugs, and other molecules. About 21,000 substituents from 17,000 entries are contained in the database. It has been successfully used in the area of virtual combinatorial chemistry, generation of bioactive molecules, bioisosteric design, and so on. The selection method of the 47 substituents is shown in Transparent Methods. The structures of the substituents are shown in Figure S1.

The Computational Protocol of Ligand-Directed Evolution

A well-designed and organized computational strategy is a powerful tool for improving computational accuracy and efficiency. AILDE, which combined one-step free energy perturbation (FEP) and molecular
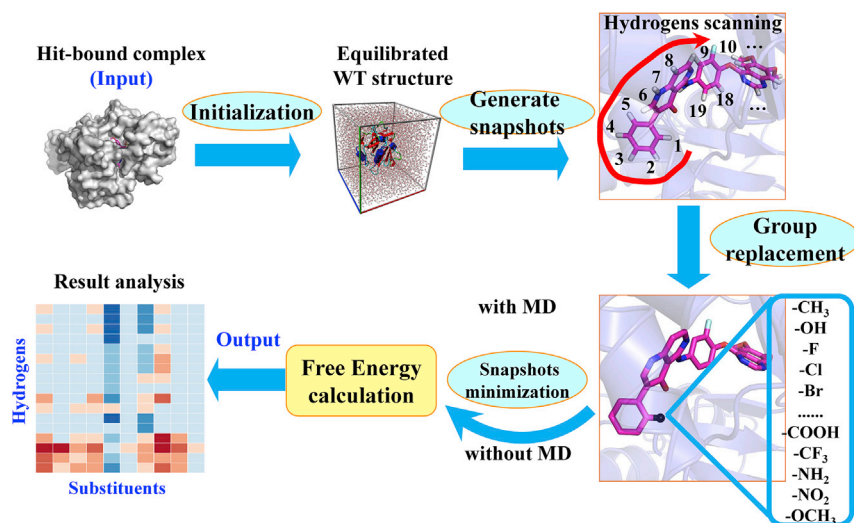
Figure 1. The Calculation Workflow of AILDE

The computational process includes initialization, MD simulation, one-step FEP scanning, and free energy calculation. In the snapshot minimization, we use two different strategies for comparison.

dynamics (MD) simulation strategies, was achieved by replacing hydrogen atoms with small groups on the hit compound to generate lead candidates with enhanced potency. MD simulation was first performed on the hit-receptor complex system. We collected 50 snapshots from the equilibrated MD trajectory to obtain a representative ensemble of the binding structures. Then, one-step FEP scanning was introduced. We replaced every hydrogen one-by-one by linking the constructed small groups in each snapshot. Each newly obtained ligand-receptor was refined by using two different minimization methods to obtain the final structures. We only performed molecular mechanics (MM) minimization with the steepest descent and conjugated gradient method in the first method. The second method was realized by adding an MD refinement step after MM minimization. Both methods were used to study the case mentioned previously. We used the MM-PBSA method to evaluate the binding free energies of the refined complex structures. The new lead compounds were finally ranked according to the binding free energy shifts between the lead-receptor and hit-receptor complexes. The details of the workflow of the computational strategy are shown in Figure 1. When evaluating a large number of changes to a parent ligand, AILDE was orders of magnitude faster than the conventional FEP-based method. The improvement in efficiency was greater when there were more changes on the ligand.

Dataset Selection

A collection of protein-ligand structures and bioactivities is prerequisite to evaluate the performance of AILDE. The diverse data to construct a test set were collected according to three criteria: (1) the cocrystal structure of the protein and hit compound is available, (2) the bioactivities of the hit compound and its analogues are available, and (3) the ligands in the dataset show high flexibility. The 19 hit-receptor complexes were collected from published papers (Aronov et al., 2009; Carbain et al., 2014; Charrier et al., 2011; Congreve et al., 2012; Crawford et al., 2014; Demont et al., 2015; Denny et al., 2017; Efremov et al., 2012; Green et al., 2015; Lawhorn et al., 2016; Liu et al., 2017; Medina et al., 2011; Mirguet et al., 2013; Narumi et al., 2013; Tsukada et al., 2010; Wang et al., 2018; Woodhead et al., 2010). There were 157 inhibitors with their experimental bioactivity data in the test dataset. These data are shown in the format of IC_{50}, K_i, K_d, or EC_{50}. These proteins included HIV1-gp120, GSK3, PDK1, and Hsp90 (Table S1). They covered multiple target types, such as viral protein, transferase, kinase, signaling protein, hydrolase, and transcription targets. The hit compounds had different sizes and properties. They were evaluated based on their molecular weight (MW), number of heavy atoms, and number of rotatable bonds (Table S1). The MWs ranged from 184.10 to 460.25 Da. Between 14 and 34 heavy atoms were present. The number of rotatable bonds ranged from 0 to 8. Hence, we considered both rigid and flexible molecules in the dataset. The structures of the hit compounds are shown in Table S1. According to the above analysis, the compounds in our testing set cover a wide range of compound sizes, properties, and target types.

Performance Evaluation and Comparison

To evaluate the impact of an additional MD simulation on the prediction accuracy, we compared the performance of two minimization methods from the perspective of qualitative and quantitative accuracy. All 19 systems were submitted to AILDE for evaluation. We obtained all the binding free energy shift ($\Delta\Delta G_{cal}$) values between the hit-receptor complexes and the lead-receptor complexes (Table S2). The related experimental binding free energy shift ($\Delta\Delta G_{exp}$) values were calculated by using the collected activity data based on Equation 4. We found that an additional MD refinement may provide a more convergent structure and a more accurate result. We have offered a detailed application example of TNNI3K-benzenesulfonamide system (see Transparent Methods and Figure S2). Finally, we also compared several published H2L strategies with AILDE, and relative to the other methods, AILDE showed a better predictive performance but lower structural diversity.

The qualitative performance of AILDE was first evaluated in terms of specificity, sensitivity, and accuracy. In the dataset, we defined the positive samples ($\Delta\Delta G_{exp} \geq 0$, indicates the derivate has lower or equal bioactivity compared with the hit compound) and negative samples ($\Delta\Delta G_{exp} < 0$, indicates the derivate has better bioactivity than the hit compound). The numerical values of sensitivity represent the probability of AILDE identifying samples that do in fact have improved bioactivity. The specificity represents the probability of AILDE recognizing samples without giving false-positive results. Accuracy is the proportion of true results, either true positive or true negative, in the total samples. Because we used two minimization methods for one-step FEP scanning, we can summarize the results as follows. (1) For the first minimization method, AILDE identified 75 of the 93 positive samples with a sensitivity (true positive rate) of 80.6% and 56 of 64 negative samples with a specificity (true negative rate) of 87.5%. A total of 131 samples were correctly classified, corresponding to an accuracy of 83.4%, and the predicted binding free energy changes were within 0.71 kcal mol^{-1} of the experimental value on average. (2) For the second minimization method, 85 of the 93 positive samples and 62 of the 64 negative samples were correctly identified, corresponding to a sensitivity of 91.4% and specificity of 96.9%, respectively. The accuracy was 93.6% based on identifying 147 samples correctly of the 157 samples in the dataset, and the predicted binding free energy changes were within 0.31 kcal mol^{-1} of the experimental values on average (Figure 2A). We observed that both methods achieved a classification accuracy over 80%. Moreover, the accuracy of the second method was over 90%. The second method had a 10% improvement in accuracy and a 0.4 kcal mol^{-1} decrease in deviation compared with the first method, which confirmed that the introduction of an MD refinement step can improve the prediction accuracy. The MD step may further optimize the snapshot toward a more reasonable and convergent state.

We then assessed the classification ability of AILDE based on the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). An ROC curve plots sensitivity versus (1-specificity) in the range of 0.0–1.0 according to different thresholds. An AUC value of 1.0 represents a perfect classifier and 0.5 represents a classifier that is no better than random. The ROC curves are shown in Figure 2B and are labeled with the AUC values. The first minimization method showed AUC = 0.875 and the second showed AUC = 0.959. The higher AUC value further confirms the better predictive performance achieved with the introduction of the MD refinement step.

A linear equation can be used to describe the relationship between two sets of variables and show how one variable changes in a linear manner as a function of changes in the other variable. For further study, the linear correlation between $\Delta\Delta G_{cal}$ and $\Delta\Delta G_{exp}$ was obtained with correlation coefficients $R^2 = 0.64$ (SD [σ] = 1.02 kcal mol^{-1}) and 0.82 (σ = 0.54 kcal mol^{-1}) (Figures 2C and 2D) for the first and second strategies, respectively. The linear relationships (R^2) of both methods are over 0.6, confirming the prediction reliability of our strategy. The second minimization method resulted in a stronger correlation. The dots in the scatterplot diagram of the second strategy (see Figures 2C and 2D) were more tightly clustered than those in the plot of the first strategy. Therefore, the introduction of an MD refinement step may improve the energy convergence result, which supports our previously mentioned conclusion.

We also compared the accuracies of published computational H2L strategies and AILDE (Carlsson et al., 2008; Chéron et al., 2016; Enyedy and Egan, 2008; Guimarães and Cardozo, 2008; Hsu et al., 2017; Montero-Torres et al., 2006; Moro et al., 2006; Pillai et al., 2005; Vyas et al., 2017; Zhou et al., 2001). The prediction results obtained with the different methods are summarized in Table S3. The accuracies are divided into quantitative and qualitative results. Quantitative accuracy is determined based on the linear correlation coefficient (R or R^2) between the experimental and calculated results. Qualitative accuracy was indicated by the AUC value or the accuracy of the model to discriminate samples into positive or negative. Compared with the ligand-based H2L methods, AILDE possesses better accuracy both quantitatively
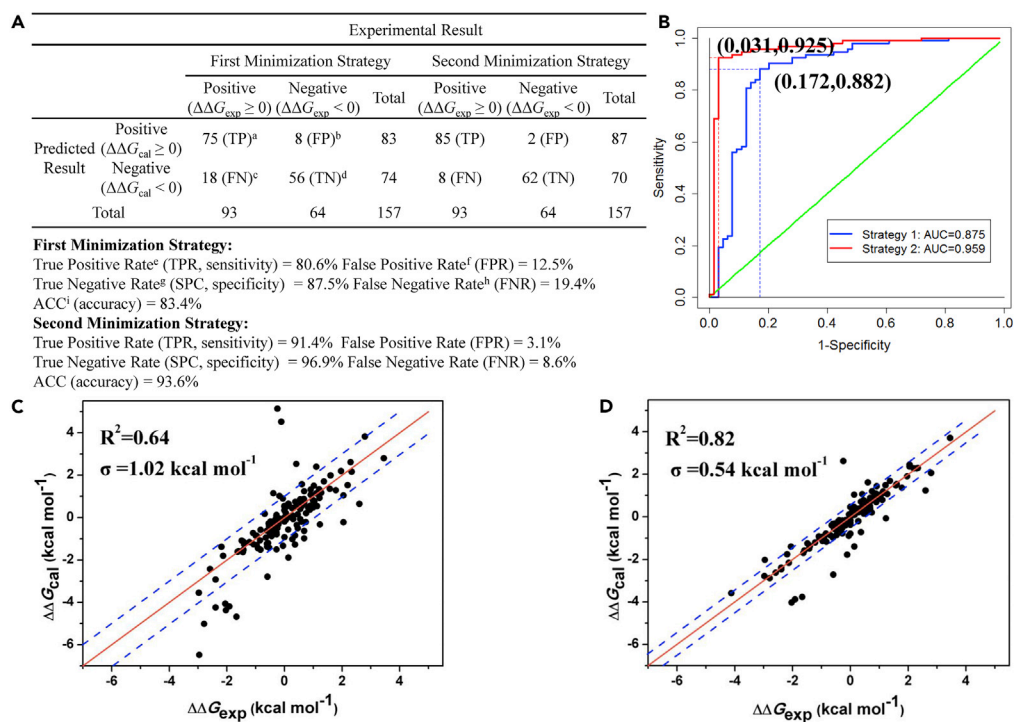
A

		Experimental Result					
		First Minimization Strategy			Second Minimization Strategy		
		Positive ($\Delta\Delta G_{exp} \geq 0$)	Negative ($\Delta\Delta G_{exp} < 0$)	Total	Positive ($\Delta\Delta G_{exp} \geq 0$)	Negative ($\Delta\Delta G_{exp} < 0$)	Total
Predicted Result	Positive ($\Delta\Delta G_{cal} \geq 0$)	75 (TP)[a]	8 (FP)[b]	83	85 (TP)	2 (FP)	87
	Negative ($\Delta\Delta G_{cal} < 0$)	18 (FN)[c]	56 (TN)[d]	74	8 (FN)	62 (TN)	70
	Total	93	64	157	93	64	157

First Minimization Strategy:
True Positive Rate[e] (TPR, sensitivity) = 80.6% False Positive Rate[f] (FPR) = 12.5%
True Negative Rate[g] (SPC, specificity) = 87.5% False Negative Rate[h] (FNR) = 19.4%
ACC[i] (accuracy) = 83.4%
Second Minimization Strategy:
True Positive Rate (TPR, sensitivity) = 91.4% False Positive Rate (FPR) = 3.1%
True Negative Rate (SPC, specificity) = 96.9% False Negative Rate (FNR) = 8.6%
ACC (accuracy) = 93.6%

Figure 2. Performance Evaluation of AILDE

(A) The details of the predictive results for the total samples under the two minimization strategies; [a] true positive, [b] false positive, [c] false negative, [d] true negative, [e] TPR = TP/(TP + FN), [f] FPR = FP/(FP + TN), [g] SPC = TN/(FP + TN), [h] FNR = FN/(TP + FN), [i] ACC = (TP + TN)/(TP + FP + FN + TN).

(B) The ROC curve showing the relationship between the sensitivity and 1-specificity under the two different minimization strategies (blue for strategy 1 and red for strategy 2) for classifying compounds with lower activity (positive samples) or better activity (negative samples) than the hit compound.

(C) Scatterplot diagram of the first strategy with correlation coefficient R^2 = 0.64 (σ = 1.02 kcal mol^{-1}) for 157 studied samples.

(D) Scatterplot diagram of the second strategy with correlation coefficient R^2 = 0.82 (σ = 0.54 kcal mol^{-1}) for 157 studied samples. The red lines in (C) and (D) correspond to ideal predictions, and the blue dashed lines mark the 1σ region. The values on the x axis and y axis represent the $\Delta\Delta G_{exp}$ value and $\Delta\Delta G_{cal}$ values, respectively.

and qualitatively. The R^2 value (0.82) is higher than the highest R^2 value (0.78) achieved with other ligand-based methods. Most importantly, our strategy is better suited for model expansion because it has a broader target selection scope and a larger number of samples than other ligand-based methods. Most structure-based H2L strategies are based on docking, MD, or MM-PB(GB)SA. The prediction accuracy of docking methods is not over 0.70, although the computational model can be applied on a larger number of systems. The accuracy of the MM-PB(GB)SA method was highest (0.81), but it was limited to the smallest number of samples. In order to compare the diversity of the structures generated by the strategies from a starting point to the potential lead compound, we calculated the Tanimoto coefficient between the initial skeleton and final most activated compound in every study (shown in Table S3) by using RDKit packages (Godden et al., 2000). The compound **5a** (hit) and **5g** (lead), discovered by us as c-Met inhibitors (seen in Figure 3), were used to calculate the Tanimoto coefficient for AILDE. It shows that AILDE has the highest Tanimoto coefficient (0.709) compared with other methods, which proves the less diversity of the structures generated by AILDE during the optimization process. By comparison, we can see that, although AILDE has higher prediction accuracy and is applicable to a considerably larger number of test samples relative to other structure-based strategies, AILDE has certain limitation compared with other methods in exploring more extensive chemical space.

Discovery of c-Met Inhibitors

We applied AILDE to the H2L optimization of c-Met inhibitors to discover a potent anticancer lead. Compound **5a** was identified in a previous virtual screening of an in-house database as a hit compound
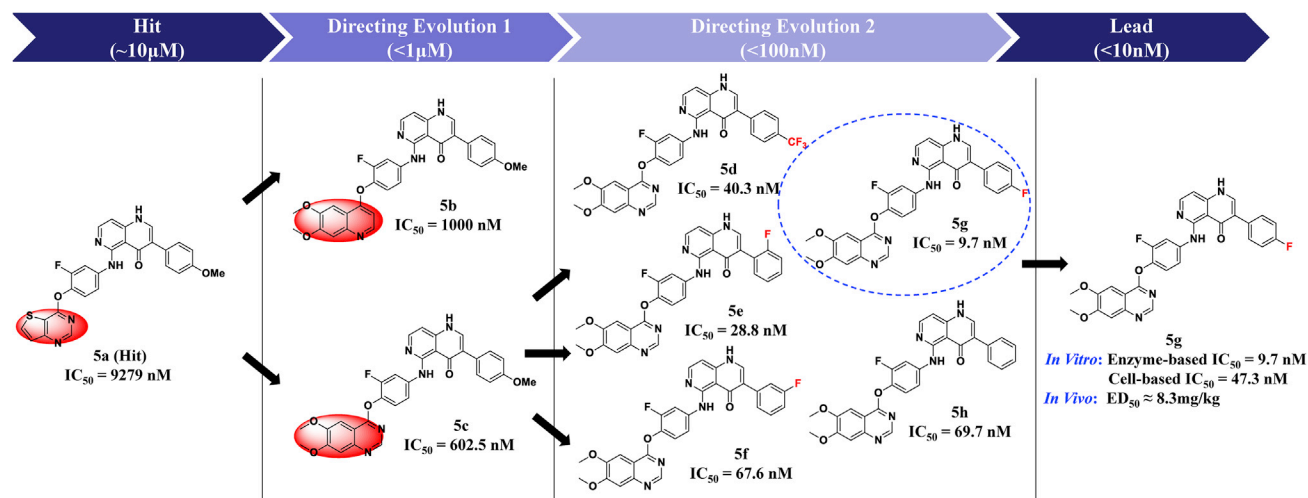
Figure 3. The Process of Structural Directed Evolution from Hit Compound 5a with Enzyme-Based IC$_{50}$ Values

targeting c-Met with weak *in vitro* activity (IC$_{50}$ = 9,279 nM). Two steps of ligand-directed evolution were performed, and the process is shown in Figure 3. In the first step, we replaced block A with other groups to improve the activity because no obvious interactions were detected between the thiophene and the surrounding residues and enough space was observed between the thieno[3,2-d]pyrimidine and the receptor (PDB: 3F82) in the docking experiment (Figure S3). The thieno[3,2-d]pyrimidin-4-yl motif (block A) in compound **5a** was replaced with a total of 9,403 medicinal fragments in PADFrag (Yang et al., 2018). The prediction result provided new compounds with different blocks, and some of the top-ranked structures are shown in Figure S4. Compounds **5b** (ΔΔG_{cal} = −1.31 kcal mol^{-1}) and **5c** (ΔΔG_{cal} = −1.45 kcal mol^{-1}) were selected for the enzyme-based activity test. **5c**, which possessed a quinazoline moiety, displayed the most promise for further modifications to improve druggability and a moderate c-Met inhibitory activity (IC$_{50}$ = 602.5 nM). **5c** showed 15-fold higher activity than **5a**.

The second step was designed to perform the small-substituent-directed evolution because large modifications may reduce the ligand efficiency. The 10 most commonly used substituents (-CH$_3$, -OH, -F, -Cl, -Br, -COOH, -CF$_3$, -NH$_2$, -NO$_2$, and -OCH$_3$) in the small group library were scanned in the R^1 position of compound **5c**. The five top-ranked and easily synthesizable compounds (**5d** ∼ **h**) were synthesized for further bioassay tests. The calculated binding free energy shifts compared with **5a** are shown in Table 1. The synthetic route is shown in Figures S5 and S6. All the synthesized compounds were structurally characterized by ^{1}H NMR and ^{13}C NMR spectroscopy and HRMS (shown in Figures S7–S22). Resulting compounds **5d–h** exhibited 8.6- to 62.1-fold increases in enzyme potency compared with **5c** and 133.1- to 956.6-fold increases compared with **5a**. In particular, compound **5g** (R^1 = 4-F) exhibited significant inhibition potency (IC$_{50}$ = 9.7 and 47.3 nM) in both the enzyme-based and cell-based assays. By using AILDE, a c-Met inhibitor with great potential (**5g**) was finally found in a highly efficient manner, and it showed an ∼1,000-fold improvement in activity compared with the hit compound (**5a**).

Compound **5g** was then subjected to pharmacokinetic (PK) evaluation and tumor growth inhibition studies in mice. As shown in Table 2, after oral administration (10 mg/kg) in rats, **5g** displayed an excellent overall PK profile with a high C$_{max}$ (12.7 μg/mL), AUC$_{0-\infty}$ (481.2 μg·h/mL), and *F* (57%). Moreover, as shown in Figure 4, **5g** induced dose-dependent TGI, with an ED$_{50}$ of 8.3 mg/kg. At 25 mg/kg, **5g** exhibited significant *in vivo* antitumor efficacy (TGI of 82%). More importantly, partial tumor regression (PR 2/6, TGI of 97%) was observed at a higher dose of **5g** (75 mg/kg), and the PKs of **5g** in beagle dogs and cynomolgus monkeys were outstanding with favorable oral bioavailabilities (*F* = 73 and 48.0%, respectively, shown in Table S4) after oral administration. In addition, the 28-day repeat-dose toxicity studies on rats showed that **5g** has a good safety performance (safety index >40-fold shown in the Transparent Methods) (shown in Figure S23). Taken together, although numerous c-Met inhibitors have been reported, compound **5g** demonstrated a favorable *in vitro* potency, *in vivo* efficacy, PK profile, and safety profile. Therefore, **5g** has been advanced to preclinical studies.
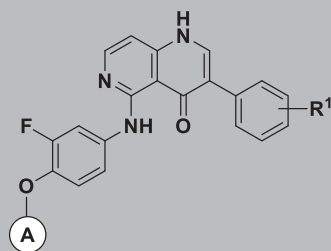
Compound	Block A	R[1]	c-Met (IC$_{50}$, nM)		$\Delta\Delta G_{cal}$ (kcal mol^{-1})
			Enzyme-Based[a]	Cell-Based[b]	
5a		4-OMe	9,279	-[c]	0
5b		4-OMe	1,000	-[c]	−1.31
5c		4-OMe	602.5	-[c]	−1.45
5d		4-CF$_3$	40.3	24,987	−3.12
5e		2-F	28.8	194.2	−3.12
5f		3-F	67.6	600.8	−2.95
5g		4-F	9.7	47.3	−3.15
5h		-H	69.7	358.4	−3.03

Table 1. Structures and c-Met Inhibitory Potencies of Compounds 5a–5h with Calculated $\Delta\Delta G_{cal}$ Values Compared with 5a

[a]*In vitro* kinase assays were performed with the indicated purified recombinant c-Met kinase domains (nM).

[b]IC$_{50}$ values (nM) for HGF-mediated autophosphoylation in MKN-45 cells.

[c]Not tested.

Cocrystal Structure of c-Met and the Inhibitor

To validate the accuracy of the simulated structure, the cocrystal structure of c-Met and the inhibitor was obtained. We have tried to cultivate the cocrystal structure of **5g** and c-Met but failed because of the slightly poor solubility of **5g**. Therefore, we synthesized compound **5i** (IC$_{50}$ = 27.1 nM, enzyme inhibition activity), which is closely related to the structure of **5g** but has better solubility, to obtain the cocrystal structure with c-Met. The

Compound	Route	Dose (mg/kg)	C_{max} (μg/mL)	T_{max} (h)	$T_{1/2}$ (h)	$AUC_{0-\infty}$ (μg* h/mL)	CL (L/h/kg)	V_z (L/kg)	F (%)
5g	p.o.	10	12.7	2.0	26.7	481.2	0.02	0.8	57
	i.v.	5	30.7	0.03	16.8	420.6	0.01	0.3	

Table 2. *In Vivo* PK Profiles of Compound 5g in Rats

Vehicle: 70% PEG400-30% water. C_{max}, maximum concentration; T_{max}, time of maximum concentration; $T_{1/2}$, half-life; $AUC_{0-\infty}$, area under the plasma concentration time curve; CL, clearance; V_Z, volume of distribution; F, oral bioavailability.
Reported data are the average of six animals.

only difference between **5i** and **5g** is that **5i** has a morpholinomethyl group linked to the methoxy moiety of **5g**. We finally determined the X-ray crystal structure of compound **5i** bound to c-Met at a resolution of 1.80 Å. As shown in Figure 5A, the quinazoline group lies in the adenine pocket and forms an H-bond with the backbone N of Met1160 in the hinge. The quinazoline-linked fluorophenoxy group forms a π-π interaction with residue Phe1223 in the DFG (Asp1222-Phe1223-Gly1224) motif. The DFG-out conformation opens a hydrophobic allosteric pocket that is occupied by the terminal 4-fluorophenyl group. In addition to forming an H-bond with residue Lys1110 in the N-lobe, the naphthyridin-4-one group also binds with an allosteric channel and directly interacts with DFG residue Asp1222 through the formation of an H-bond. The morpholine group extends into the solvent. We compared our predicted binding mode of **5g** with the X-ray crystal model of **5i** and found that the binding model of **5g** was very close to that of **5i** (Figure 5B). The same parts of their skeleton have a root-mean-square deviation (RMSD) of 0.25 Å. **5g** has the same H-bonding and hydrophobic interactions as **5i** in the crystal complex. The high consistency of the binding modes and interactions in the predicted and experimental crystal structures confirmed the reliability of our computational strategy.

AILDE Server Implementation

A freely accessible web server also named AILDE was constructed to allow medicinal chemists to easily use AILDE. The AILDE web server includes the front end and background applications. The architecture of the AILDE web server is shown in Figure S24. The front end consists of a client layer and a resulting browse layer. The client layer provides a user-friendly entry for the user to submit computational tasks. Users may also obtain comprehensive instruction about how to use AILDE from the client layer. The resulting browsing layer offers output formats for users to better understand the results. The background was coded in Python to finish the computational process. The server is managed by Maui-3.3.1 and Torque-6.0.1 on a supercomputing cluster in our laboratory (including 48 computing nodes with 40 CPUs each, 20 computing nodes with 4 GPUs each, 8 computing nodes with 40 CPUs and 256 GB large memory each, 3 storage servers, and 1 management server). The calculation time for a job depends largely on the size of the
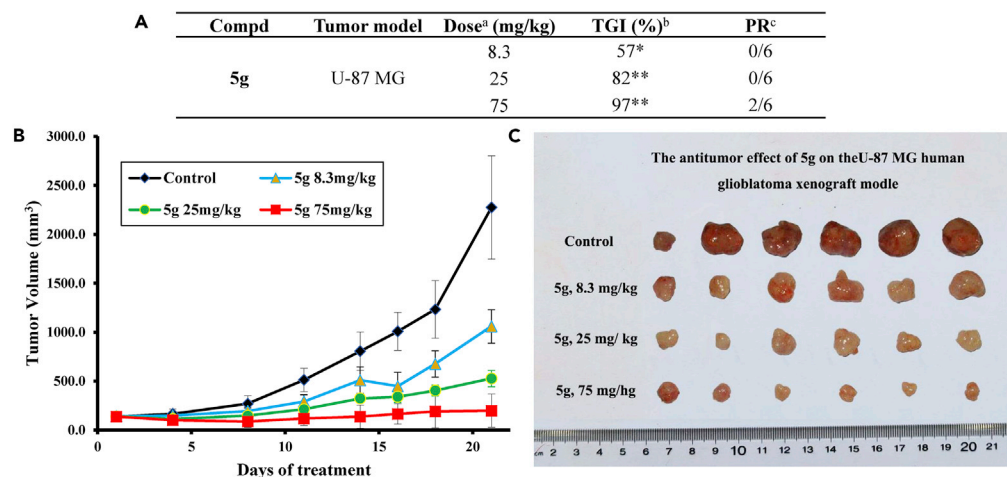

Compd	Tumor model	Dose[a] (mg/kg)	TGI (%)[b]	PR[c]
5g	U-87 MG	8.3	57*	0/6
		25	82**	0/6
		75	97**	2/6

Figure 4. *In Vivo* Bioassay of Compound 5g

(A) *In vivo* TGI activity of compound **5g** in rats. [a] Administered orally at corresponding dose once daily for 21 consecutive days (70% PEG-400/H$_2$O); [b] tumor growth inhibition value, *:p < 0.05, **:p < 0.01; [c] partial regression.
(B) Tumor growth inhibition of **5g** in the U87-MG xenograft model in mice.
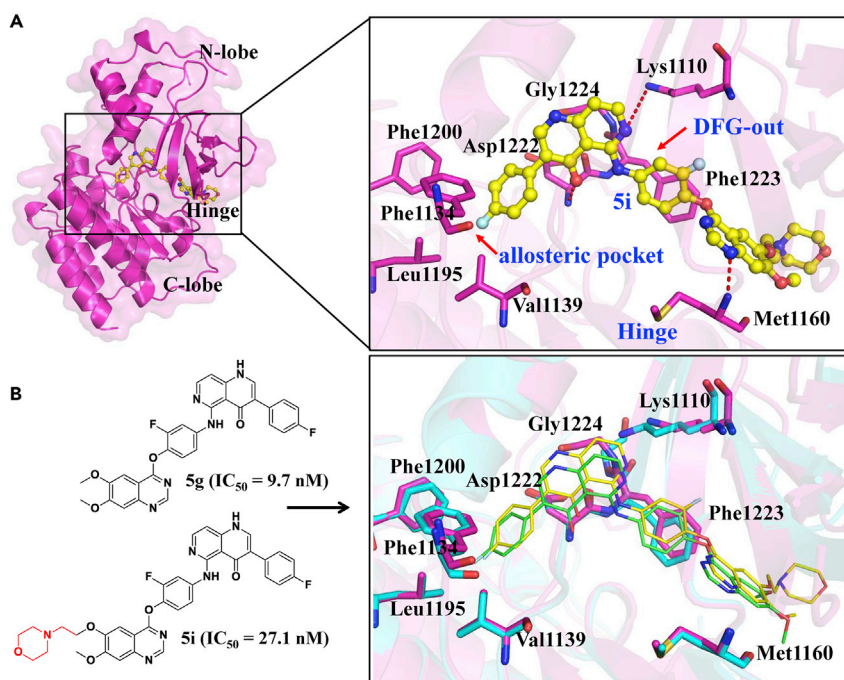(C) Tumor pictures at the end of the *in vivo* antitumor activity assay.

Figure 5. X-ray Structure of 5i and Modeled Structure of 5g Bound to c-Met

(A) The X-ray structure of **5i** bound to c-Met, **5i** is shown with yellow ball-and-stick model, residues from c-Met are shown as magenta sticks, H-bonds are shown as red, dotted lines.

(B) The superimposition of the modeled and X-ray crystal structures of c-Met in complex with ligands (**5g** and **5i**). The modeled structure of **5g** is shown in green lines, and the receptor (PDB: 3F82) is shown in blue sticks and cartoon. The X-ray structure of **5i** is given by the yellow lines, with the receptor as magenta sticks and cartoon.

complex and the hardware status. Take the example job of the server as an example, the protein has 380 residues in total, the MD simulation step (about 6 ns) takes about 1.5 h (4 GPU parallel), the hydrogen replacement and minimization step (for first minimization strategy) takes 3 min per newly obtained lead compound (25 CPUs parallel for 50 snapshots of every compound). The free energy calculation step takes 4 min per compound (4 CPUs parallel). The compound in the example has 17 hydrogens and there are 10 substituents on the server. It will generate 170 new compounds in total. The total time cost of the example job is 1.5 h + 3 min * 170 + 4 min * 170 = 21.3 h and it takes less than 1 day to finish the job. The wizard of the AILDE server is shown in Figure 6, and a detailed description of how to use the server can be found in the Transparent Methods. To enable users to use AILDE in a more private way, we have also provided an offline version of our code on GitHub (https://github.com/fwangccnu/AILDE) for scientists to download and use.

Limitations of the Study

AILDE reasonably relies on the resemblance assumption between the binding mode of hit-receptor and lead-receptor and adequacy of the conformational sampling, so it is the reliable prediction results when binding mode of the lead compound relative to the hit compound does not change much. We can also see that AILDE has certain limitation compared with other methods in exploring more extensive chemical space, although it has higher prediction accuracy and is applicable to a considerably larger number of test samples.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gefei Hao (gefei_hao@foxmail.com).

Materials Availability

All unique/stable reagents generated in this study will be made available on request, but we may require a payment and/or a completed Materials Transfer Agreement if there is potential for commercial application.
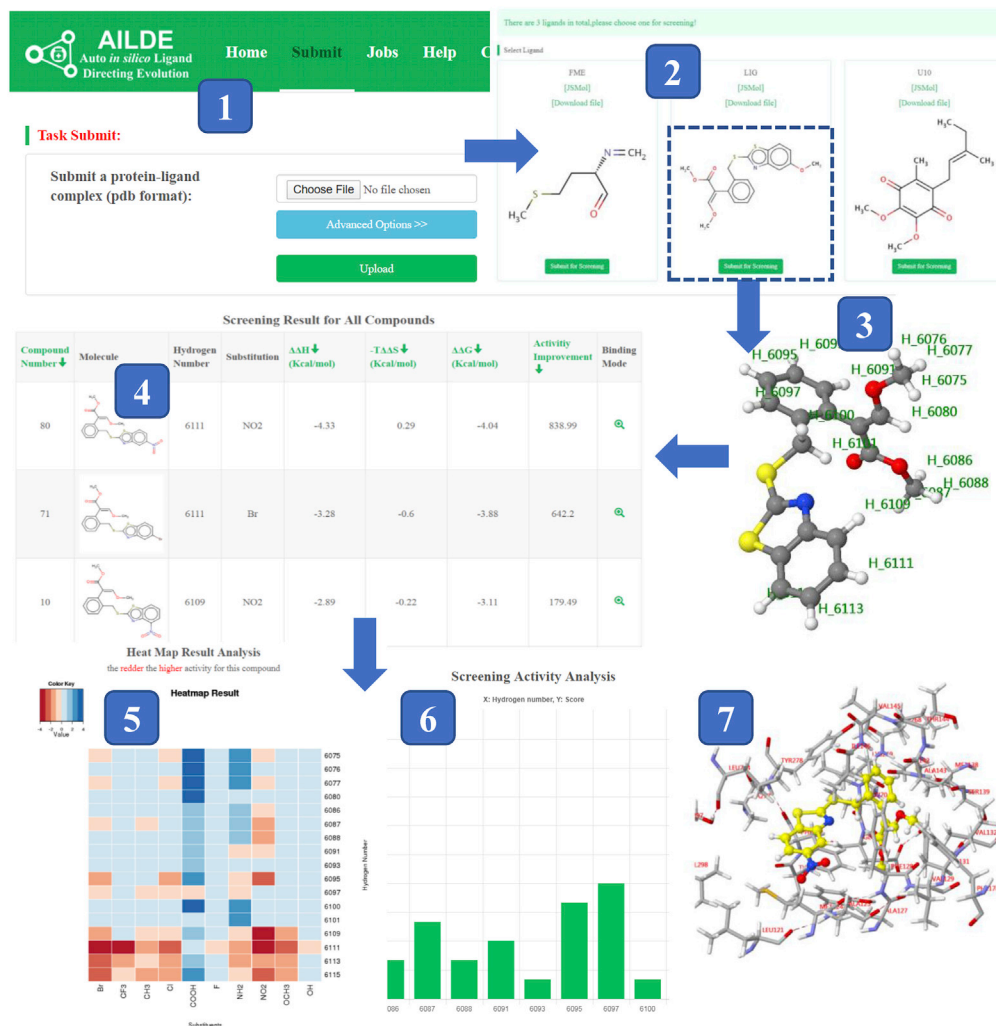
Figure 6. The Detailed Wizard of AILDE Server to Offer a Intuitive View for Users to Explore It

AILDE server wizard: (1) uploading of a hit-protein complex; (2) selection of a hit compound; (3) hydrogen serial numbers; (4) table with details about the new leads, $\Delta\Delta G$ and activity fold-changes; (5) heatmap showing the relationship matrix between the substituent positions and substituents (darker red colors indicate compounds with higher activities); (6) histogram illustrating the overall result based on substituent positions, which helps to elucidate which positions possess the most potential for transformation; (7) 3D viewer showing the binding mode of every compound with the target protein.

Data and Code Availability

The online web service of AILDE can be accessed from http://chemyang.ccnu.edu.cn/ccb/server/AILDE. The offline version of AILDE program can be downloaded from https://github.com/fwangccnu/AILDE.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101179.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

F.W. conducted molecular simulation, data analysis, and results discussion. L.Z. conducted molecular synthesis. F.W. conducted the AILDE server design. W.H., G.H., and G.Y. supervised the study, designed the experiments, and carried out the results discussion.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Aronov, A.M., Tang, Q., Martinez-Botella, G., Bemis, G.W., Cao, J.R., Chen, G.J., Ewing, N.P., Ford, P.J., Germann, U.A., Green, J., et al. (2009). Structure-guided design of potent and selective pyrimidylpyrrole inhibitors of extracellular signal-regulated kinase (ERK) using conformational control. J. Med. Chem. 52, 6362–6368.

Basilico, C., Pennacchietti, S., Vigna, E., Chiriaco, C., Arena, S., Bardelli, A., Valdembri, D., Serini, G., and Michieli, P. (2013). Tivantinib (ARQ197) displays cytotoxic activity that is independent of its ability to bind MET. Clin. Cancer Res. 19, 2381–2392.

Carbain, B., Paterson, D.J., Anscombe, E., Campbell, A.J., Cano, C., Echalier, A., Endicott, J.A., Golding, B.T., Haggerty, K., Hardcastle, I.R., et al. (2014). 8-Substituted O6-cyclohexylmethylguanine CDK2 inhibitors: using structure-based inhibitor design to optimize an alternative binding mode. J. Med. Chem. 57, 56–70.

Card, G.L., Blasdel, L., England, B.P., Zhang, C., Suzuki, Y., Gillette, S., Fong, D., Ibrahim, P.N., Artis, D.R., Bollag, G., et al. (2005). A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. Nat. Biotechnol. 23, 201–207.

Carlsson, J., Boukharta, L., and Åqvist, J. (2008). Combining docking, molecular dynamics and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. J. Med. Chem. 51, 2648–2656.

Charrier, J.D., Miller, A., Kay, D.P., Brenchley, G., Twin, H.C., Collier, P.N., Ramaya, S., Keily, S.B., Durrant, S.J., Knegtel, R.M.A., et al. (2011). Discovery and structure–activity relationship of 3-Aminopyrid-2-ones as potent and selective interleukin-2 inducible T-cell kinase (Itk) inhibitors. J. Med. Chem. 54, 2341–2350.

Chéron, N., Jasty, N., and Shakhnovich, E.I. (2016). OpenGrowth: an automated and rational algorithm for finding new protein ligands. J. Med. Chem. 59, 4171–4188.

Comoglio, P.M., Giordano, S., and Trusolino, L. (2008). Drug development of MET inhibitors: targeting oncogene addiction and expedience. Nat. Rev. Drug Discov. 7, 504–516.

Congreve, M., Andrews, S.P., Doré, A.S., Hollenstein, K., Hurrell, E., Langmead, C.J., Mason, J.S., Ng, I.W., Tehan, B., Zhukov, A., et al. (2012). Discovery of 1,2,4-Triazine derivatives as adenosine A2A antagonists using structure based drug design. J. Med. Chem. 55, 1898–1903.

Crawford, T.D., Ndubaku, C.O., Chen, H., Boggs, J.W., Bravo, B.J., DeLaTorre, K., Giannetti, A.M., Gould, S.E., Harris, S.F., Magnuson, S.R., et al. (2014). Discovery of selective 4-Amino-pyridopyrimidine inhibitors of MAP4K4 using fragment-based lead identification and optimization. J. Med. Chem. 57, 3484–3493.

Cui, J.J. (2014). Targeting receptor tyrosine kinase MET in cancer: small molecule inhibitors and clinical progress. J. Med. Chem. 57, 4427–4453.

Cui, J.J., Tran-Dube, M., Shen, H., Nambu, M., Kung, P.-P., Pairish, M., Jia, L., Meng, J., Funk, L., Botrous, I., et al. (2011). Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal-epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). J. Med. Chem. 54, 6342–6363.

Demont, E.H., Chung, C., Furze, R.C., Grandi, P., Michon, A.-M., Wellaway, C., Barrett, N., Bridges, A.M., Craggs, P.D., Diallo, H., et al. (2015). Fragment-based discovery of low-micromolar ATAD2 bromodomain inhibitors. J. Med. Chem. 58, 5649–5673.

Denny, R.A., Flick, A.C., Coe, J., Langille, J., Basak, A., Liu, S., Stock, I., Sahasrabudhe, P., Bonin, P., Hay, D.A., et al. (2017). Structure-based design of highly selective inhibitors of the CREB binding protein bromodomain. J. Med. Chem. 60, 5349–5363.

Efremov, I.V., Vajdos, F.F., Borzilleri, K.A., Capetta, S., Chen, H., Dorff, P.H., Dutra, J.K., Goldstein, S.W., Mansour, M., McColl, A., et al. (2012). Discovery and optimization of a novel spiropyrrolidine inhibitor of β-secretase (BACE1) through fragment-based drug design. J. Med. Chem. 55, 9069–9088.

Enyedy, I.J., and Egan, W.J. (2008). Can we use docking and scoring for hit-to-lead optimization? J. Comput. Aided Mol. Des. 22, 161–168.

Erlanson, D.A., Braisted, A.C., Raphael, D.R., Randal, M., Stroud, R.M., Gordon, E.M., and Wells, J.A. (2000). Site-directed ligand discovery. Proc. Natl. Acad. Sci. U S A 97, 9367–9372.

Erlanson, D.A., Fesik, S.W., Hubbard, R.E., Jahnke, W., and Jhoti, H. (2016). Twenty years on: the impact of fragments on drug discovery. Nat. Rev. Drug Discov. 15, 605–619.

Erlanson, D.A., Lam, J.W., Wiesmann, C., Luong, T.N., Simmons, R.L., DeLano, W.L., Choong, I.C., Burdett, M.T., Flanagan, W.M., Lee, D., et al. (2003). In situ assembly of enzyme inhibitors using extended tethering. Nat. Biotechnol. 21, 308–314.

Gharwan, H., and Groninger, H. (2016). Kinase inhibitors and monoclonal antibodies in oncology: clinical implications. Nat. Rev. Clin. Oncol. 13, 209–227.

Godden, J.W., Xue, L., and Bajorath, J. (2000). Combinatorial preferences affect molecular similarity/diversity calculations using binary fingerprints and Tanimoto coefficients. J. Chem. Inf. Comp. Sci. 40, 163–166.

Green, J., Cao, J., Bandarage, U.K., Gao, H., Court, J., Marhefka, C., Jacobs, M., Taslimi, P., Newsome, D., Nakayama, T., et al. (2015). Design, synthesis, and structure–activity relationships of pyridine-based Rho kinase (ROCK) inhibitors. J. Med. Chem. 58, 5028–5037.

Gross, S., Rahal, R., Stransky, N., Lengauer, C., and Hoeflich, K.P. (2015). Targeting cancer with kinase inhibitors. J. Clin. Invest. 125, 1780–1789.

Guimarães, C.R.W., and Cardozo, M. (2008). MM-GB/SA rescoring of docking poses in structure-based lead optimization. J. Chem. Inf. Model. 48, 958–970.

Hajduk, P.J., Sheppard, G., Nettesheim, D.G., Olejniczak, E.T., Shuker, S.B., Meadows, R.P., Steinman, D.H., Carrera, G.M., Marcotte, P.A., Severin, J., et al. (1997). Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. J. Am. Chem. Soc. 119, 5818–5827.

Hao, G.F., Jiang, W., Ye, Y.N., Wu, F.X., Zhu, X.L., Guo, F.B., and Yang, G.F. (2016). ACFIS: a web server for fragment-based drug discovery. Nucleic Acids Res. 44, W550–W556.

Hochguertel, M., Kroth, H., Piecha, D., Hofmann, M.W., Nicolau, C., Krause, S., Schaaf, O., Sonnenmoser, G., and Eliseev, A.V. (2002).

Target-induced formation of neuraminidase inhibitors from in vitro virtual combinatorial libraries. Proc. Natl. Acad. Sci. U S A 99, 3382–3387.

Hoffer, L., Muller, C., Roche, P., and Morelli, X. (2018). Chemistry-driven hit-to-lead optimization guided by structure-based approaches. Mol. Inform. 37, e1800059.

Hsu, H.H., Hsu, Y.C., Chang, L.J., and Yang, J.M. (2017). An integrated approach with new strategies for QSAR models and lead optimization. BMC Genomics 18, 104–112.

Hughes, J.P., Rees, S., Kalindjian, S.B., and Philpott, K.L. (2011). Principles of early drug discovery. Br. J. Pharmacol. 162, 1239–1249.

Keserű, G.M., and Makara, G.M. (2006). Hit discovery and hit-to-lead approaches. Drug Discov. Today 11, 741–748.

Lawhorn, B.G., Philp, J., Graves, A.P., Holt, D.A., Gatto, G.J., and Kallander, L.S. (2016). Substituent effects on drug–receptor H-bond interactions: correlations useful for the design of kinase inhibitors. J. Med. Chem. 59, 10629–10641.

Lin, H.Y., Chen, X., Chen, J.N., Wang, D.W., Wu, F.X., Lin, S.Y., Zhan, C.G., Wu, J.W., Yang, W.C., and Yang, G.F. (2019). Crystal structure of 4-hydroxyphenylpyruvate dioxygenase in complex with substrate reveals a new starting point for herbicide discovery. Research (Wash. D C) 2019, 2602414.

Liu, Q., Huang, F., Yuan, X., Wang, K., Zou, Y., Shen, J., and Xu, Y. (2017). Structure-guided discovery of novel, potent, and orally bioavailable inhibitors of lipoprotein-associated phospholipase A2. J. Med. Chem. 60, 10231–10244.

Ma, P.C., Schaefer, E., Christensen, J.G., and Salgia, R. (2005). A selective small molecule c-MET inhibitor, PHA665752, cooperates with rapamycin. Clin. Cancer Res. 11, 2312–2319.

Martens, T., Schmidt, N.-O., Eckerich, C., Fillbrandt, R., Merchant, M., Schwall, R., Westphal, M., and Lamszus, K. (2006). A novel one-armed anti-c-met antibody inhibits glioblastoma growth in vivo. Clin. Cancer Res. 12, 6144–6152.

Medina-Franco, J.L., Martinez-Mayorga, K., Giulianotti, M.A., Houghten, R.A., and Pinilla, C. (2008). Visualization of the chemical space in drug discovery. Curr. Comput. Aided Drugs 4, 322–333.

Medina, J.R., Becker, C.J., Blackledge, C.W., Duquenne, C., Feng, Y.H., Grant, S.W., Heerding, D., Li, W.H., Miller, W.H., Romeril, S.P., et al. (2011). Structure-based design of potent and selective 3-phosphoinositide-dependent kinase-1 (PDK1) inhibitors. J. Med. Chem. 54, 1871–1895.

Mirguet, O., Gosmini, R., Toum, J., Clement, C.A., Barnathan, M., Brusq, J.-M., Mordaunt, J.E., Grimes, R.M., Crowe, M., Pineau, O., et al. (2013). Discovery of epigenetic regulator I-BET762: lead optimization to afford a clinical candidate inhibitor of the BET bromodomains. J. Med. Chem. 56, 7501–7515.

Montero-Torres, A., García-Sánchez, R.N., Marrero-Ponce, Y., Machado-Tugores, Y., Nogal-Ruiz, J.J., Martínez-Fernández, A.R., Arán, V.J., Ochoa, C., Meneses-Marcel, A., and Torrens, F. (2006). Non-stochastic quadratic fingerprints and LDA-based QSAR models in hit and lead generation through virtual screening: theoretical and experimental assessment of a promising method for the discovery of new antimalarial compounds. Eur. J. Med. Chem. 41, 483–493.

Moridani, M., and Harirforoosh, S. (2014). Drug development and discovery: challenges and opportunities. Drug Discov. Today 19, 1679–1681.

Moro, S., Bacilieri, M., Cacciari, B., Bolcato, C., Cusan, C., Pastorin, G., Klotz, K.-N., and Spalluto, G. (2006). The application of a 3D-QSAR (autoMEP/PLS) approach as an efficient pharmacodynamic-driven filtering method for small-sized virtual library: application to a lead optimization of a human A3 adenosine receptor antagonist. Bioorg. Med. Chem. 14, 4923–4932.

Narumi, T., Arai, H., Yoshimura, K., Harada, S., Hirota, Y., Ohashi, N., Hashimoto, C., Nomura, W., Matsushita, S., and Tamamura, H. (2013). CD4 mimics as HIV entry inhibitors: lead optimization studies of the aromatic substituents. Bioorg. Med. Chem. 21, 2518–2526.

Nienaber, V.L., Richardson, P.L., Klighofer, V., Bouska, J.J., Giranda, V.L., and Greer, J. (2000). Discovering novel ligands for macromolecules using X-ray crystallographic screening. Nat. Biotechnol. 18, 1105–1108.

Oltersdorf, T., Elmore, S.W., Shoemaker, A.R., Armstrong, R.C., Augeri, D.J., Belli, B.A., Bruncko, M., Deckwerth, T.L., Dinges, J., Hajduk, P.J., et al. (2005). An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature 435, 677–681.

Orita, M., Ohno, K., Warizaya, M., Amano, Y., and Niimi, T. (2011). Chapter fifteen - lead generation and examples: opinion regarding how to follow up hits. Method. Enzymol. 493, 383–419.

Paul, S.M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lindborg, S.R., and Schacht, A.L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat. Rev. Drug Discov. 9, 203–214.

Pillai, A.D., Rani, S., Rathod, P.D., Xavier, F.P., Vasu, K.K., Padh, H., and Sudarsanam, V. (2005). QSAR studies on some thiophene analogs as anti-inflammatory agents: enhancement of activity by electronic parameters and its utilization for chemical lead optimization. Bioorg. Med. Chem. 13, 1275–1283.

Rebecca, D.P., and Benoit, D. (2004). Facts, figures and trends in lead generation. Curr. Top. Med. Chem. 4, 569–580.

Reymond, J.L., van Deursen, R., Blum, L.C., and Ruddigkeit, L. (2010). Chemical space as a source for new drugs. Medchemcomm 1, 30–38.

Schulz, R., Atef, A., Becker, D., Gottschalk, F., Tauber, C., Wagner, S., Arkona, C., Abdel-Hafez, A.A., Farag, H.H., Rademann, J., et al. (2018). Phenylthiomethyl ketone-based fragments show

selective and irreversible inhibition of enteroviral 3C proteases. J. Med. Chem. 61, 1218–1230.

Souers, A.J., Leverson, J.D., Boghaert, E.R., Ackler, S.L., Catron, N.D., Chen, J., Dayton, B.D., Ding, H., Enschede, S.H., Fairbrother, W.J., et al. (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nat. Med. 19, 202–208.

Tsukada, T., Takahashi, M., Takemoto, T., Kanno, O., Yamane, T., Kawamura, S., and Nishi, T. (2010). Structure-based drug design of tricyclic 8H-indeno 1,2-d 1,3 thiazoles as potent FBPase inhibitors. Bioorg. Med. Chem. Lett. 20, 1004–1007.

Vinkers, H.M., de Jonge, M.R., Daeyaert, F.F.D., Heeres, J., Koymans, L.M.H., van Lenthe, J.H., Lewi, P.J., Timmerman, H., Van Aken, K., and Janssen, P.A.J. (2003). SYNOPSIS: synthesize and optimize system in silico. J. Med. Chem. 46, 2765–2773.

Vyas, V.K., Shah, S., and Ghate, M. (2017). Generation of new leads as HIV-1 integrase inhibitors: 3D QSAR, docking and molecular dynamics simulation. Med. Chem. Res. 26, 532–550.

Wang, F., Jeon, K.O., Salovich, J.M., Macdonald, J.D., Alvarado, J., Gogliotti, R.D., Phan, J., Olejniczak, E.T., Sun, Q., Wang, S., et al. (2018). Discovery of potent 2-Aryl-6,7-dihydro-5H-pyrrolo[1,2-a]imidazoles as WDR5-WIN-site inhibitors using fragment-based methods and structure-based design. J. Med. Chem. 61, 5623–5642.

Woodhead, A.J., Angove, H., Carr, M.G., Chessari, G., Congreve, M., Coyle, J.E., Cosme, J., Graham, B., Day, P.J., Downham, R., et al. (2010). Discovery of (2,4-dihydroxy-5-isopropylphenyl)- 5-(4-methylpiperazin-1-ylmethyl)-1,3- dihydroisoindol-2-yl methanone (AT13387), a novel inhibitor of the molecular chaperone Hsp90 by fragment based drug design. J. Med. Chem. 53, 5956–5969.

Wu, P., Nielsen, T.E., and Clausen, M.H. (2016). Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. Drug Discov. Today 21, 5–10.

Yang, J.F., Wang, F., Jiang, W., Zhou, G.Y., Li, C.Z., Zhu, X.L., Hao, G.F., and Yang, G.F. (2018). PADFrag: a database built for the exploration of bioactive fragment space for drug discovery. J. Chem. Inf. Model. 58, 1725–1730.

Zhao, H. (2007). Scaffold selection and scaffold hopping in lead generation: a medicinal chemistry perspective. Drug Discov. Today 12, 149–155.

Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A., Veselov, M.S., Aladinskiy, V.A., Aladinskaya, A.V., Terentiev, V.A., Polykovskiy, D.A., Kuznetsov, M.D., Asadulaev, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat. Biotechnol. 37, 1038–1040.

Zhou, R., Friesner, R.A., Ghosh, A., Rizzo, R.C., Jorgensen, W.L., and Levy, R.M. (2001). New linear interaction method for binding affinity calculations using a continuum solvent model. J. Phys. Chem. B 105, 10388–10397.

Supplemental Information

Auto *In Silico* Ligand Directing

Evolution to Facilitate the Rapid

and Efficient Discovery of Drug Lead

Fengxu Wu, Linsheng Zhuo, Fan Wang, Wei Huang, Gefei Hao, and Guangfu Yang
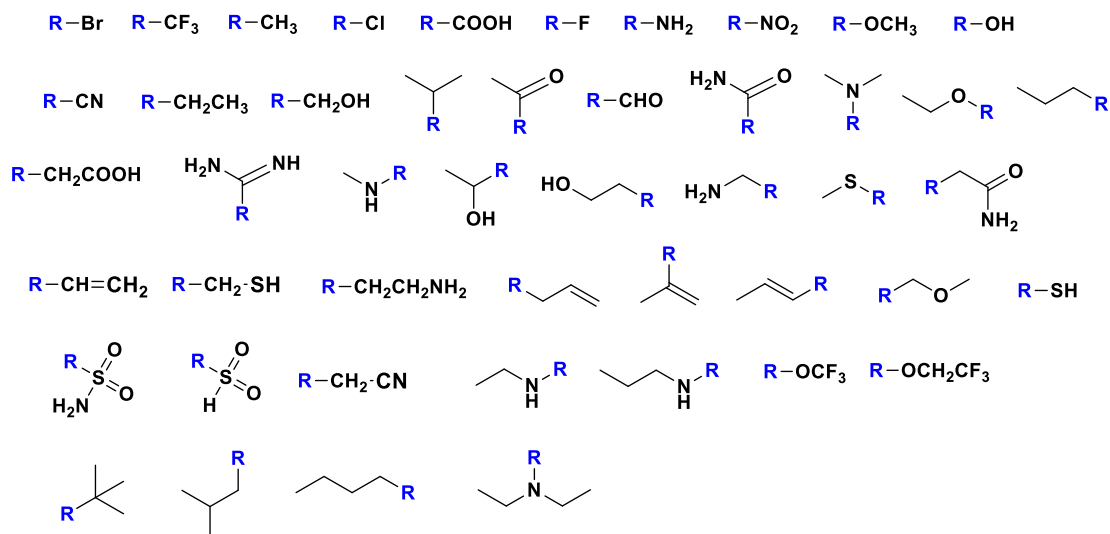
Figures and Tables

R—Br R—CF₃ R—CH₃ R—Cl R—COOH R—F R—NH₂ R—NO₂ R—OCH₃ R—OH

R—CN R—CH₂CH₃ R—CH₂OH ... R—CHO ...

R—CH₂COOH ... ...

R—CH=CH₂ R—CH₂-SH R—CH₂CH₂NH₂ ... R—SH

R—CH₂-CN ... R—OCF₃ R—OCH₂CF₃

Figure S1. The structures of 47 substituents. R-group represents the link point, Related to Figure 1.

Table S1. The detailed dataset information in this study. It includes 19 systems with 157 samples in total. [a] The compound serial number in the related reference. [b] The number of heavy atoms. [c] Molecular weight, Related to Figure 2.

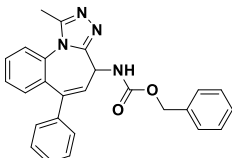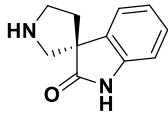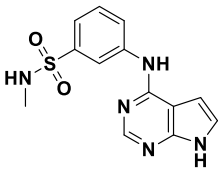| No.	Target name	PDB ID	Hit properties					Samples number
			Structures	Compound No.[a]	Heavy Atom[b]	MW[c]	Rotate bonds	
1(Narumi et al., 2013)	HIV-1 gp 120	3TGS		1	23	337.16	5	4
2(Aronov et al., 2009)	GSK3b	3I4B		2	31	413.19	8	6
3(Medina et al., 2011)	PDK1	3NUN		1	17	226.10	1	9
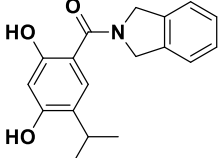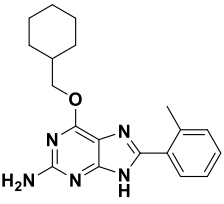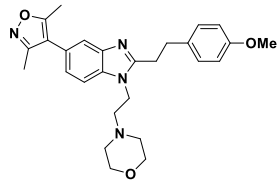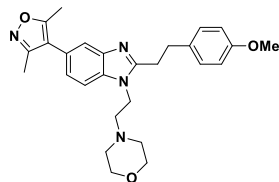
No.	Target name	PDB ID	Hit properties					Samples number
			Structures	Compound No.[a]	Heavy Atom[b]	MW[c]	Rotate bonds	
4(Woodhead et al., 2010)	HSP90	2XAB		1	22	297.14	3	8
5(Mirguet et al., 2013)	Brd4	2YEL		7	32	422.17	6	9
6(Efremov et al., 2012)	BACE	3UDH		1	14	188.09	0	6
7(Lawhorn et al., 2016)	TNNI3K	4YFI		1	21	303.08	4	13
8(Carbain et al., 2014)	CDK2	4CFM		29	25	337.19	4	4
9(Denny et al., 2017)	CREBBP	5CGP		1	34	460.25	8	5
10(Denny et al., 2017)	Brd4	5CFW		1	34	460.25	8	6

No.	Target name	PDB ID	Hit properties					Samples number
			Structures	Compound No.[a]	Heavy Atom[b]	MW[c]	Rotate bonds	
11(Congreve et al., 2012)	Adenosine A$_{2A}$	3UZA		4g	21	277.13	2	6
12(Demont et al., 2015)	ATAD2	5A5R		50	25	335.17	3	11
13(Tsukada et al., 2010)	FBPase	3KC0		10b	18	283.01	3	8
14(Charrier et al., 2011)	ITK	3QGW		7a	22	291.10	4	7
15(Liu et al., 2017)	Lp-PLA2	5YE9		21	29	406.10	7	12
16(Wang et al., 2018)	WDR5	6D9X		F-1	14	184.10	1	7
17(Wang et al., 2018)	WDR5	6DAK		4a	24	317.15	5	9
18(Crawford et al., 2014)	MAP4K4	4OBO		9	17	221.10	1	16
19(Green et al., 2015)	ROCK	4YVE		5	21	294.08	5	11
Total	-	-	-	-	-	-	-	157

Table S3. Several computational H2L strategies compared with our protocol. These strategies were divided into ligand-based and structure-based, including their detail calculation method, studying systems, sample number and prediction accuracy. [a] A linear interaction energy (LIE) method based on solvent surface generalized born (SGB) model. [b] Accuracy. [c] Area under curve, Related to Figure 2.

Year	Method	Systems	Dataset	Prediction accuracy		Tanimoto coefficient
				quantitative	qualitative	
Ligand-based method						
2005(Pillai et al., 2005)	QSAR (4 types of descriptors)	thiophene analogs on anti-inflammatory	43	R: 0.5~0.88		0.102
2006(Montero-Torres et al., 2006)	2D-QSAR (fingerprints)	antimalarials	65		acc[b]: 83.3%	0.093
2006(Moro et al., 2006)	3D-QSAR (autoMEP/PLS)	human A3 adenosine receptor antagonist	358	R: 0.82 q: 0.82		0.366
2017(Hsu et al., 2017)	QSAR (residue and atom-based interactions)	huAChE inhibitors	69	R^2: 0.78 q2: 0.82		0.627
2017(Vyas et al., 2017)	3D-QSAR (CoMFA and CoMSIA)	HIV-1 integrase	71	R^2: 0.70		0.207
Structure-based method						
2001(Zhou et al., 2001)	SGB-LIEa method	ligands of 3 targets (HIV-1 RT, thrombin, factor Xa)	35	R^2: 0.774 q^2: 0.717		0.132
2008(Carlsson et al.,	LIE with docking, MD and	Non-nucleoside inhibitors to HIV-1 RT	39	R^2: 0.70		0.179

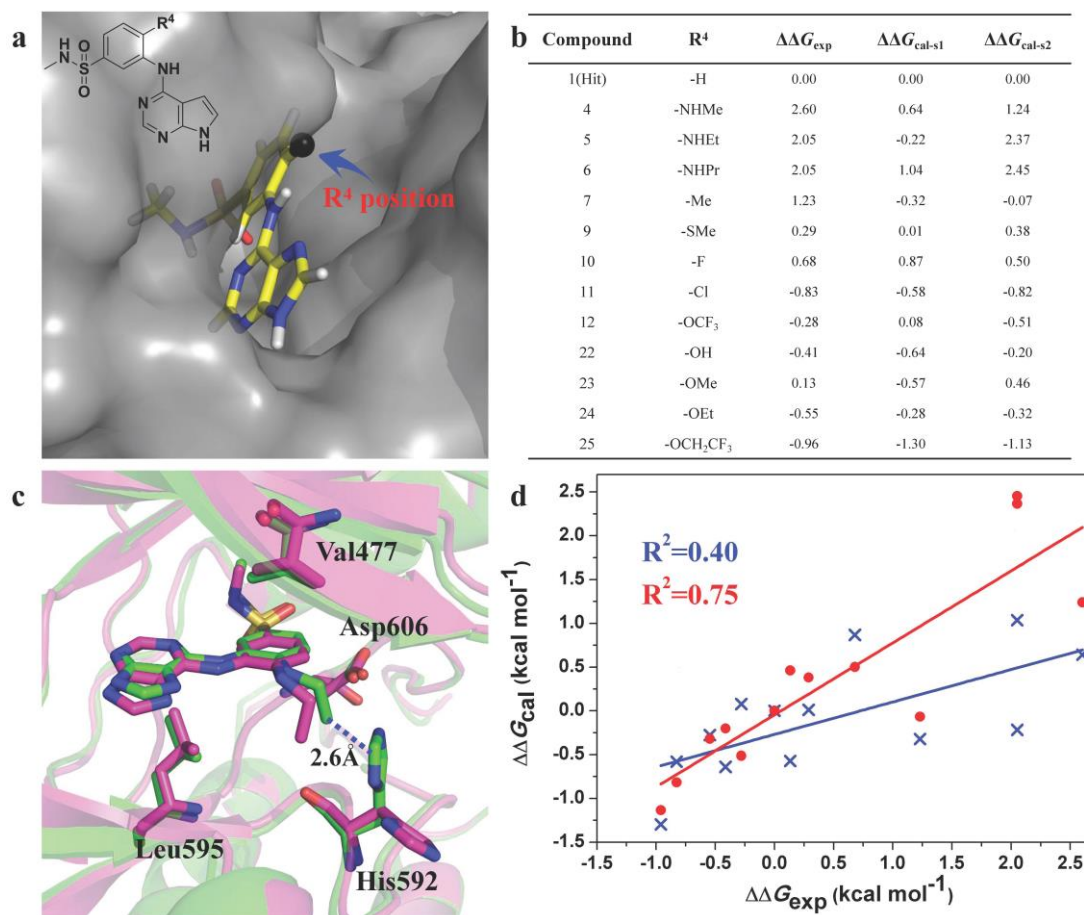| Year	Method	Systems	Dataset	Prediction accuracy		Tanimoto coefficient
				quantitative	qualitative	
2008)	scoring					
2008(Enyedy and Egan, 2008)	docking	compounds with experiments bioactivity targeted to 3 kinases (KDR, CDK2, C-ABL)	~14,300	R^2: 0.03~0.12	auc[c]: 0.59~0.71	
2008(Guimarães and Cardozo, 2008)	docking and MM/GBSA	inhibitors of 4 targets (CDK2, FactorXa, thrombin, HIV-1 RT)	75	R^2: 0.00~0.65(docking score) R^2: 0.64~0.81(MM/GBSA)		0.593~0.643
2016(Chéron et al., 2016)	OpenGrowth (group growth and MM/PBSA)	commrcial inhibitors of HIV-1 protease	9	R^2: 0.483~0.695		0.304
2019	AILDE (one-step FEP and MD)	inhibitors of 19 targets	157	R^2: 0.82	acc: 93.6% auc: 0.959	0.709

Compound	R^4	$\Delta\Delta G_{exp}$	$\Delta\Delta G_{cal-s1}$	$\Delta\Delta G_{cal-s2}$
1(Hit)	-H	0.00	0.00	0.00
4	-NHMe	2.60	0.64	1.24
5	-NHEt	2.05	-0.22	2.37
6	-NHPr	2.05	1.04	2.45
7	-Me	1.23	-0.32	-0.07
9	-SMe	0.29	0.01	0.38
10	-F	0.68	0.87	0.50
11	-Cl	-0.83	-0.58	-0.82
12	-OCF$_3$	-0.28	0.08	-0.51
22	-OH	-0.41	-0.64	-0.20
23	-OMe	0.13	-0.57	0.46
24	-OEt	-0.55	-0.28	-0.32
25	-OCH$_2$CF$_3$	-0.96	-1.30	-1.13

Figure S2. The application example of TNNI3K-benzenesulfonamide system. (a) The crystal structure of hit compound bound with TNNI3K (PDB ID: 4YFI), R^4 is the hydrogen substituent position shown with the black sphere. (b) The related compounds in the reference with the type of R^4 groups $\Delta\Delta G_{exp}$, $\Delta\Delta G_{cal-s1}$ and $\Delta\Delta G_{cal-s2}$. (c) Final structure of compound5-TNNI3K obtained by using first minimization strategy (Green) aligned with compound5-TNNI3K by using second minimization strategy (Magenta). (d) The scatter plot of $\Delta\Delta G_{exp}$-$\Delta\Delta G_{cal-s1}$ (blue cross) and $\Delta\Delta G_{exp}$-$\Delta\Delta G_{cal-s2}$ (red circle). x-axis represents the $\Delta\Delta G_{exp}$ (kcal mol^{-1}), y-axis represents the value of $\Delta\Delta G_{cal}$ (kcal mol^{-1}), Related to Figure 6.
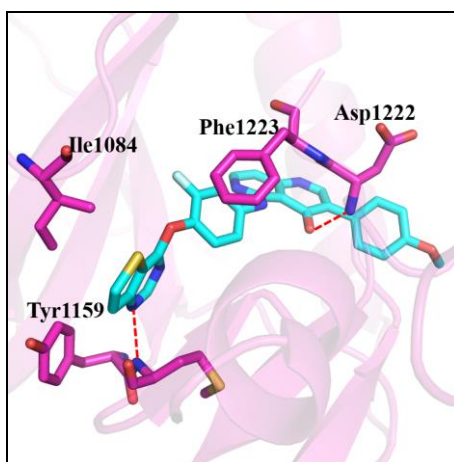

Figure S3. The binding mode of **5a** with c-Met. The hydrogen bonds were shown with red dashed, Related to Table 1 and Figure 5.
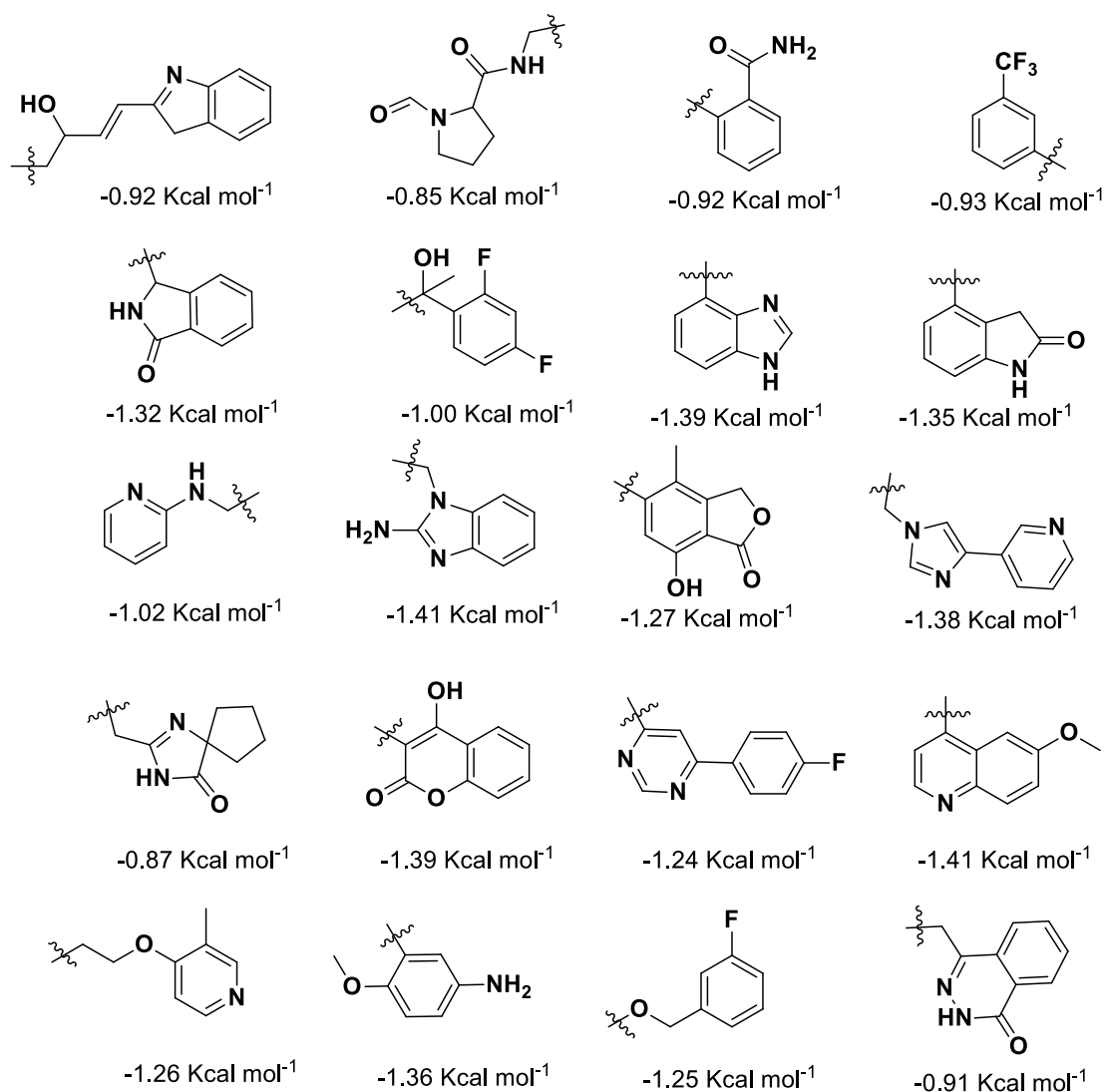
Figure S4. Some other top ranked block A structures in the first step of c-Met inhibitors directing evolution. The energy shown below was the $\Delta\Delta G_{cal}$ compared to the **5a**. Related to Table 1 and Figure 3.
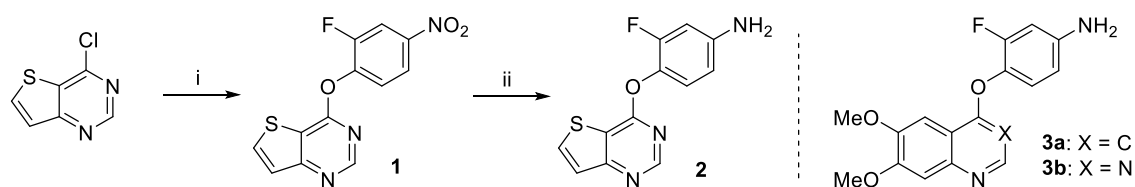

Figure S5. Reagent and Conditions: (i) DIPEA, 2-fluoro-4-nitrophenol, toluene, reflux; (ii) Fe, NH_4Cl, $EtOH/H_2O$, reflux, Related to Table 1 and Figure 3.
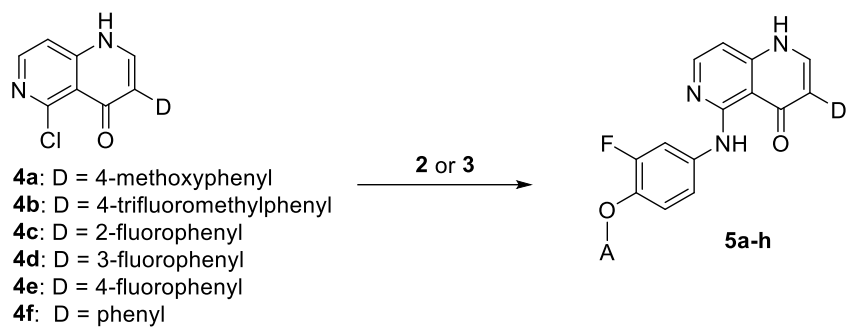
Figure S6. Reagent and Conditions: a, HCl, IPA, 90°C; b, Et₃N, DCM, r.t., Related to Table 1 and Figure 3.
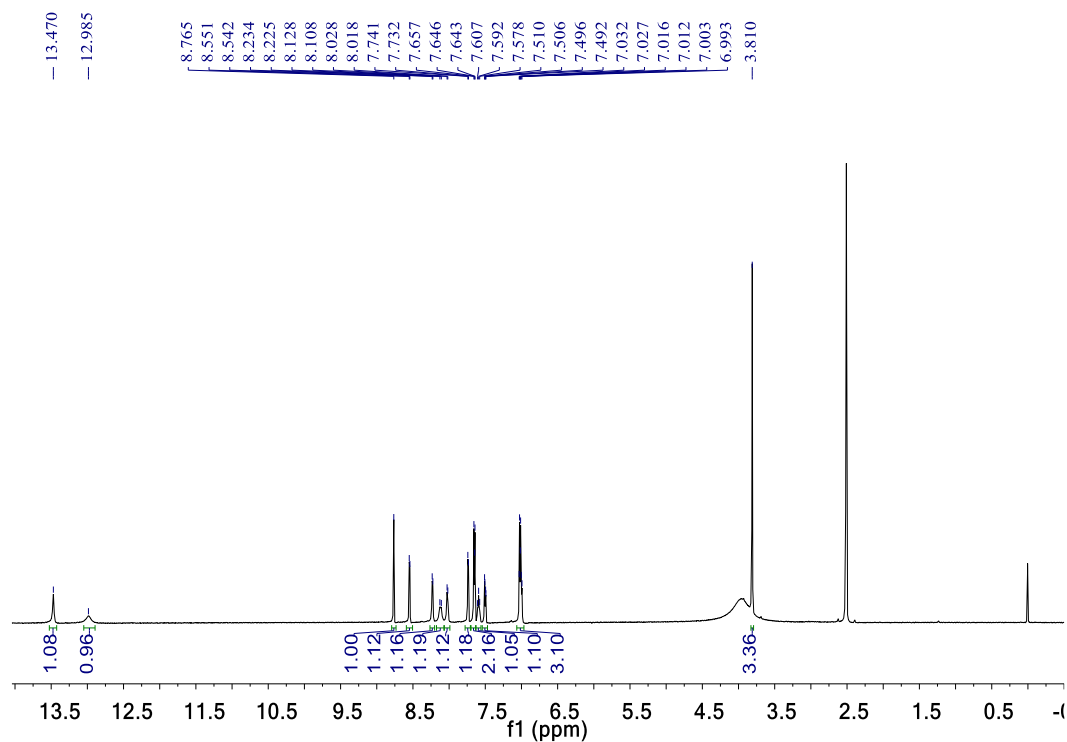

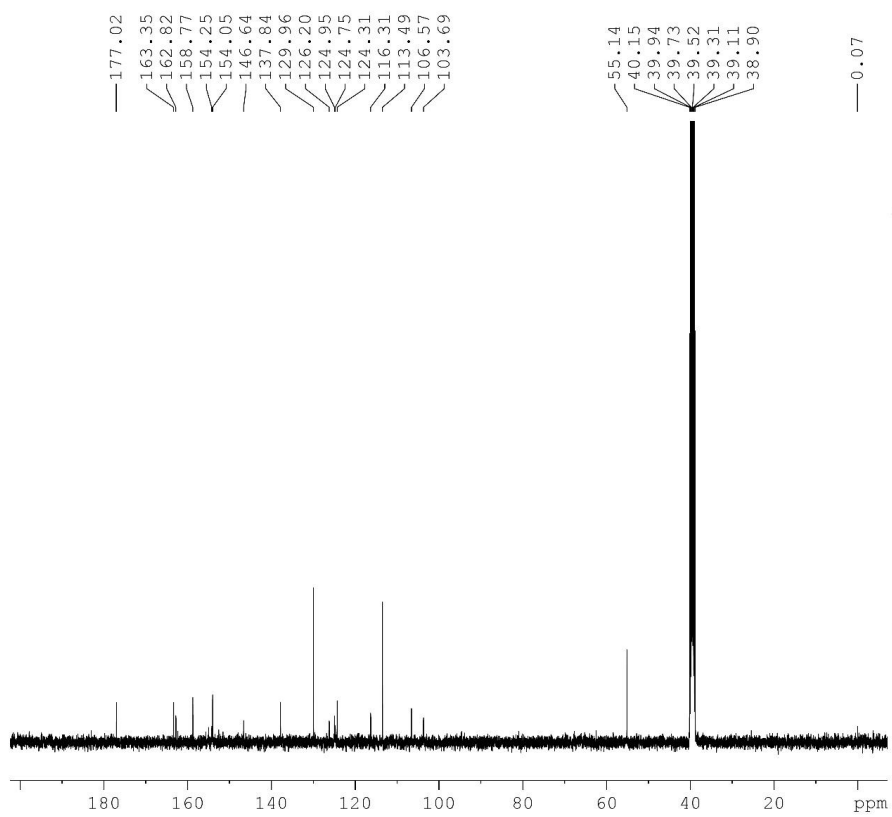**Figure S7.** 600 MHz spectrum of ¹H-NMR of compound **5a** (DMSO-d_6), Related to Table 1 and Figure 3.

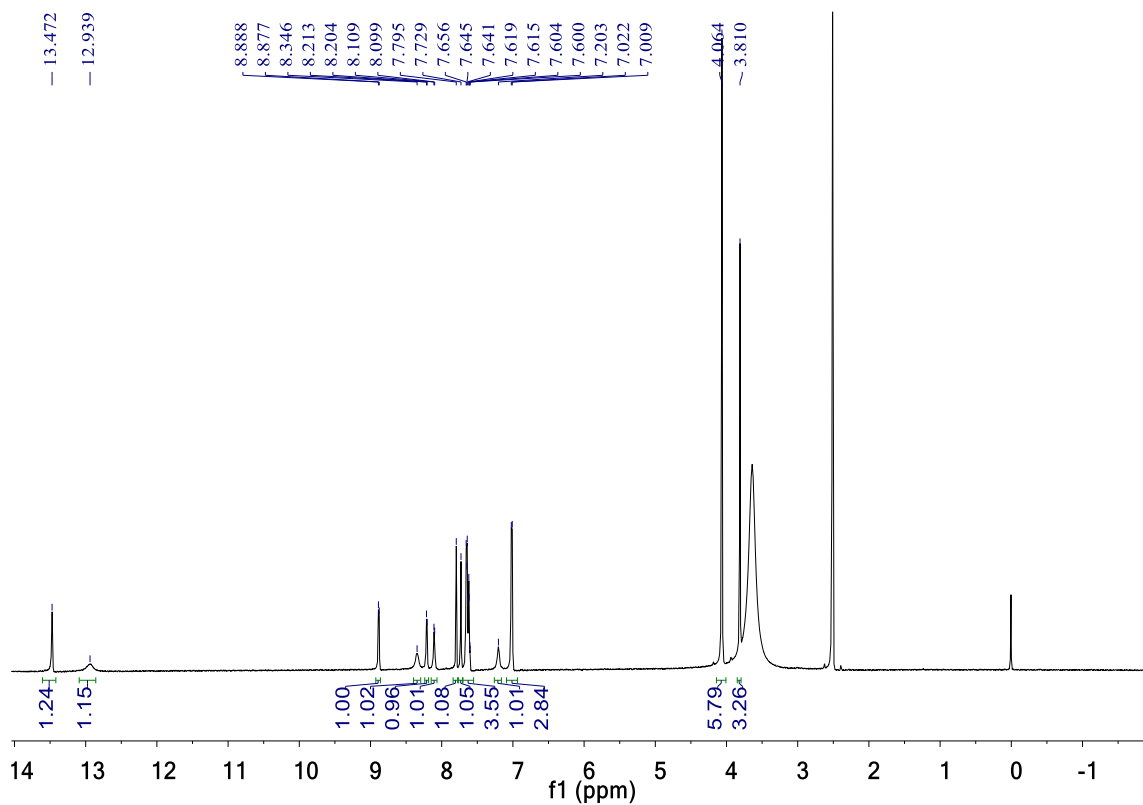**Figure S8.** 100 MHz spectrum of ^{13}C-NMR of compound **5a** (DMSO-d_6), Related to Table 1 and Figure 3.

Figure S9. 600 MHz spectrum of ^{1}H-NMR of compound **5b** (DMSO-d_6), Related to Table 1 and Figure 3.
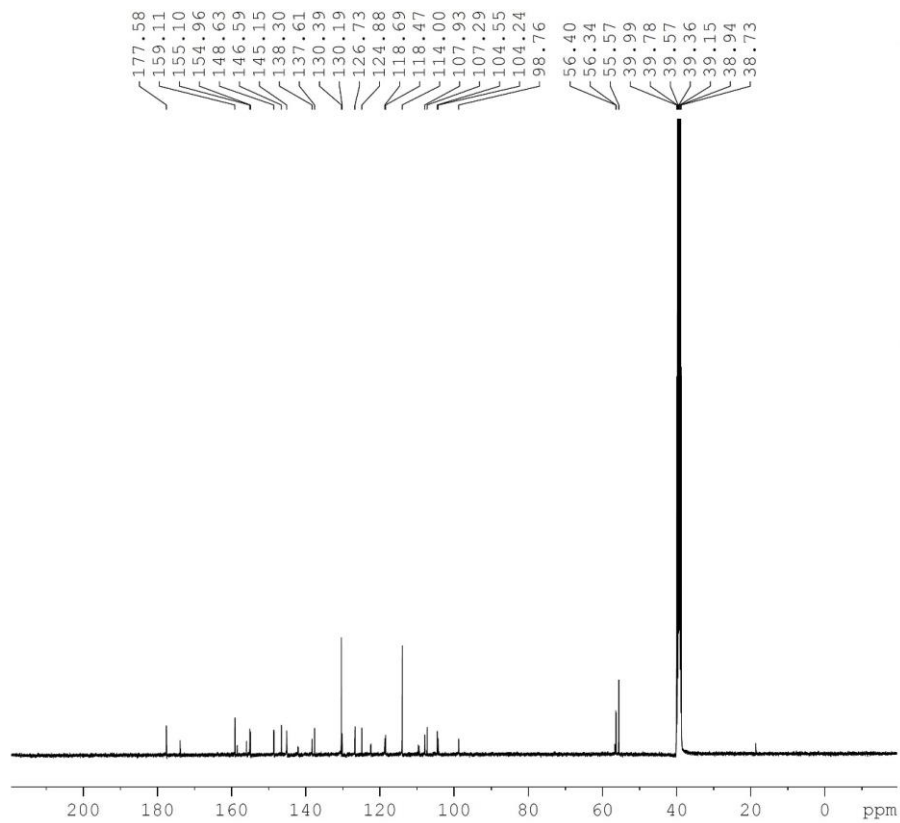

Figure S10. 100 MHz spectrum of ^{13}C-NMR of compound **5b** (DMSO-d_6), Related to Table 1 and Figure 3.

Figure S11. 400 MHz spectrum of ^{1}H-NMR of compound **5c** (DMSO-d_6), Related to Table 1 and Figure 3.

Figure S12. 100 MHz spectrum of ^{13}C-NMR of compound **5c** (DMSO-d_6), Related to Table 1 and Figure 3.

Figure S13. 400 MHz spectrum of ^{1}H-NMR of compound **5d** (DMSO-d_6), Related to Table 1 and Figure 3.

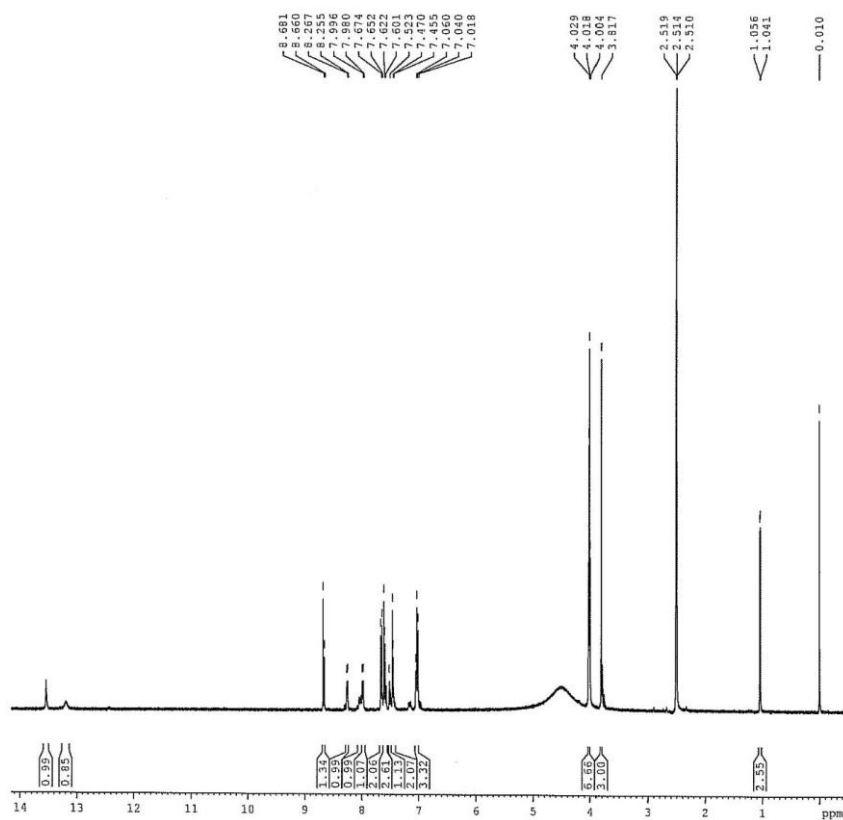**Figure S14.** 100 MHz spectrum of ^{13}C-NMR of compound **5d** (DMSO-d_6), Related to Table 1 and Figure 3.

Figure S15. 400 MHz spectrum of ^{1}H-NMR of compound **5e** (DMSO-d_6), Related to Table 1 and Figure 3.
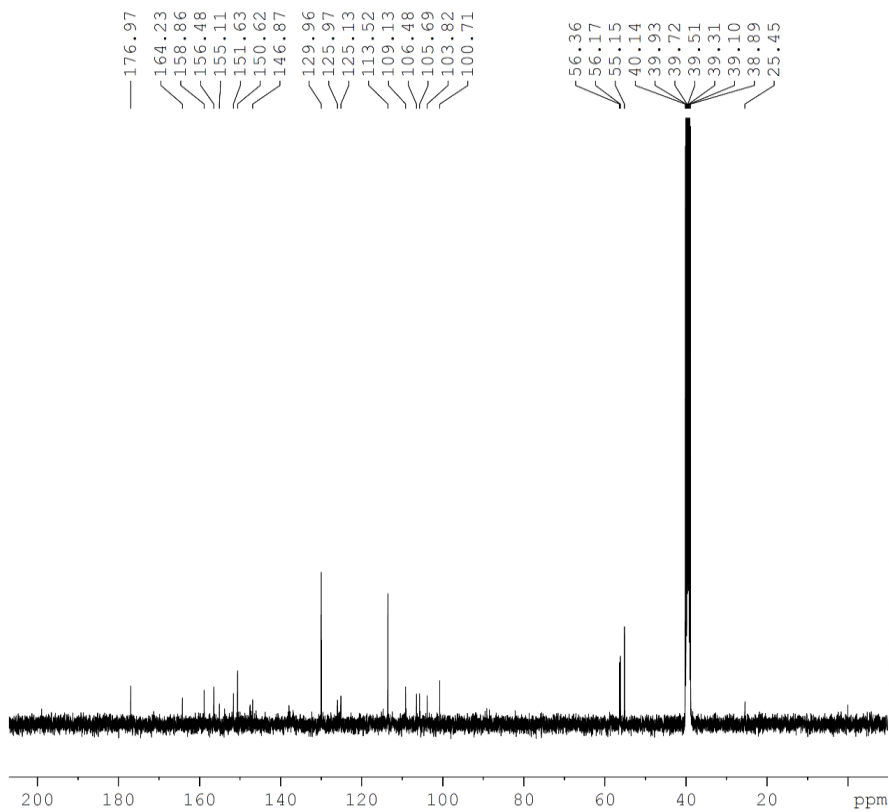

Figure S16. 100 MHz spectrum of ^{13}C-NMR of compound **5e** (DMSO-d_6), Related to Table 1 and Figure 3.

Figure S17. 400 MHz spectrum of ^{1}H-NMR of compound **5f** (DMSO-d_6), Related to Table 1 and Figure 3.
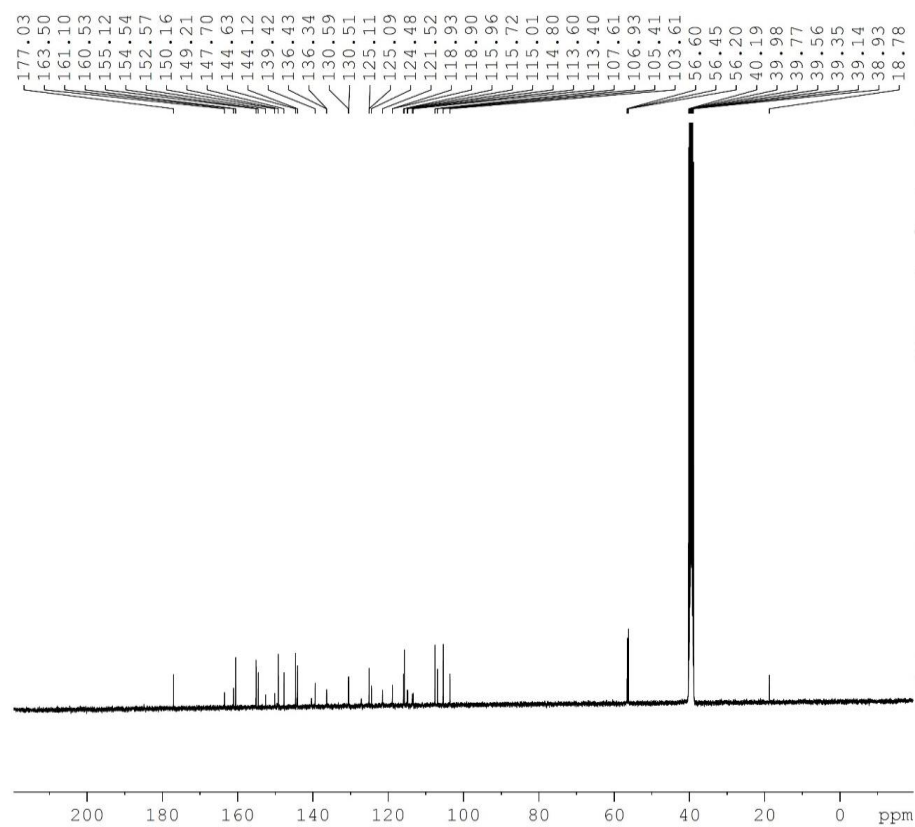
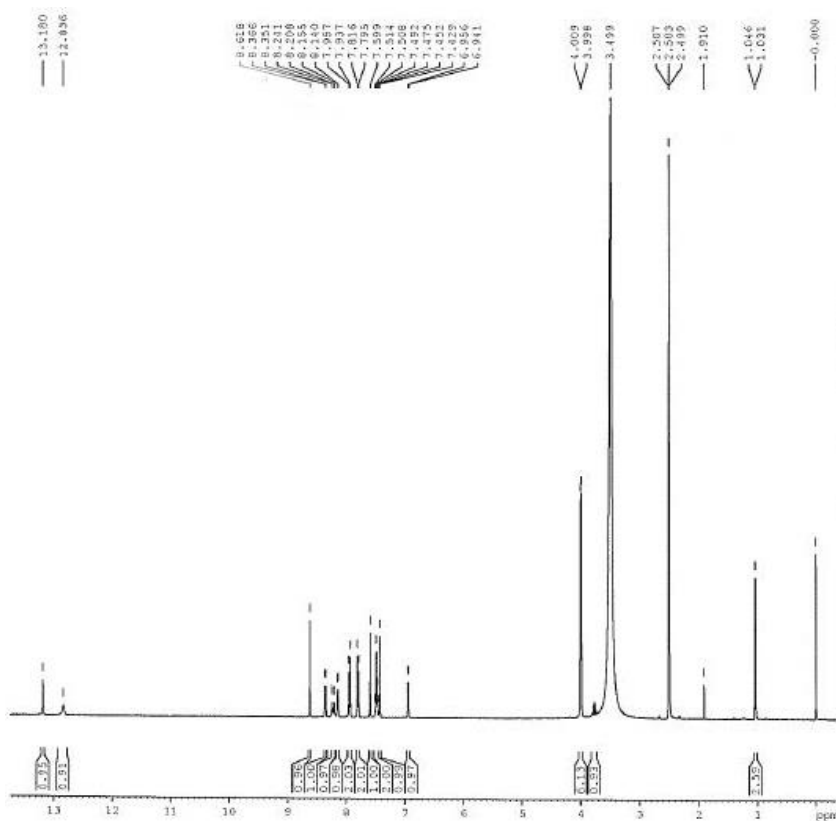**Figure S18.** 100 MHz spectrum of ^{13}C-NMR of compound **5f** (DMSO-d_6), Related to Table 1 and Figure 3.
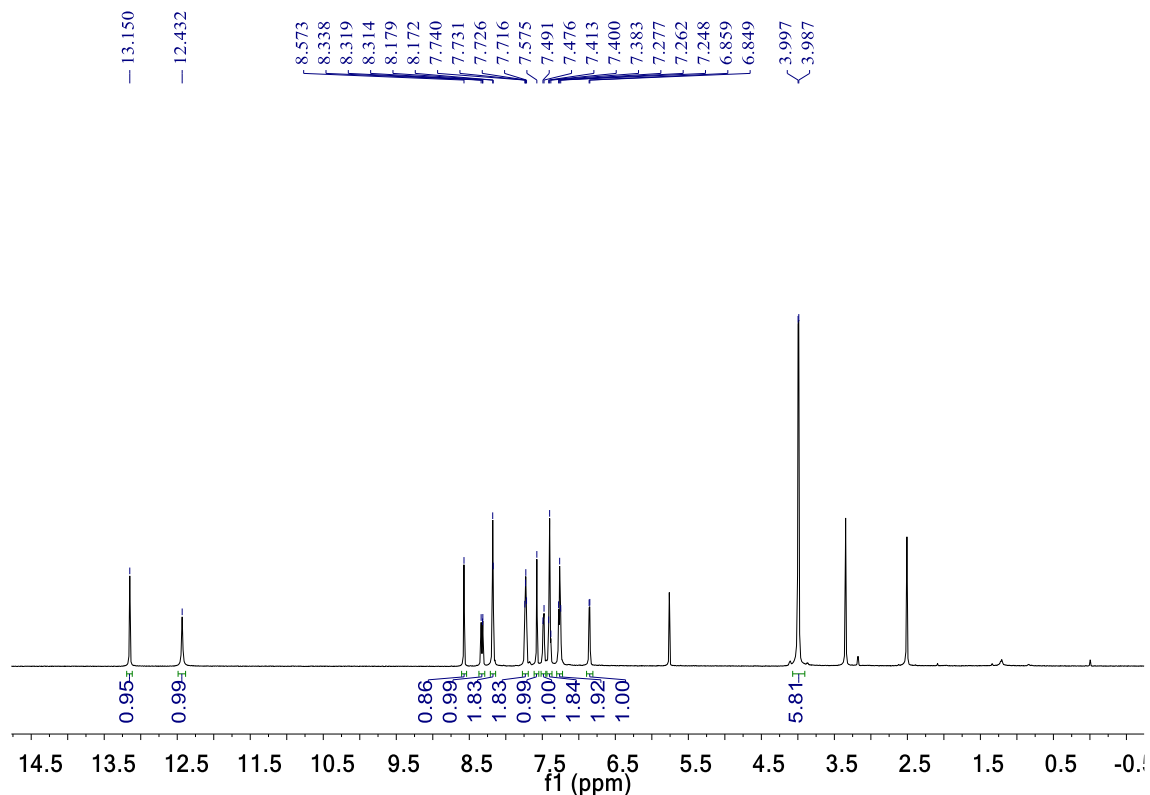
Figure S19. 400 MHz spectrum of ^{1}H-NMR of compound **5g** (DMSO-d_6), Related to Table 1 and Figure 3.

Figure S20. 100 MHz spectrum of ^{13}C-NMR of compound **5g** (DMSO-d_6), Related to Table 1 and Figure 3.

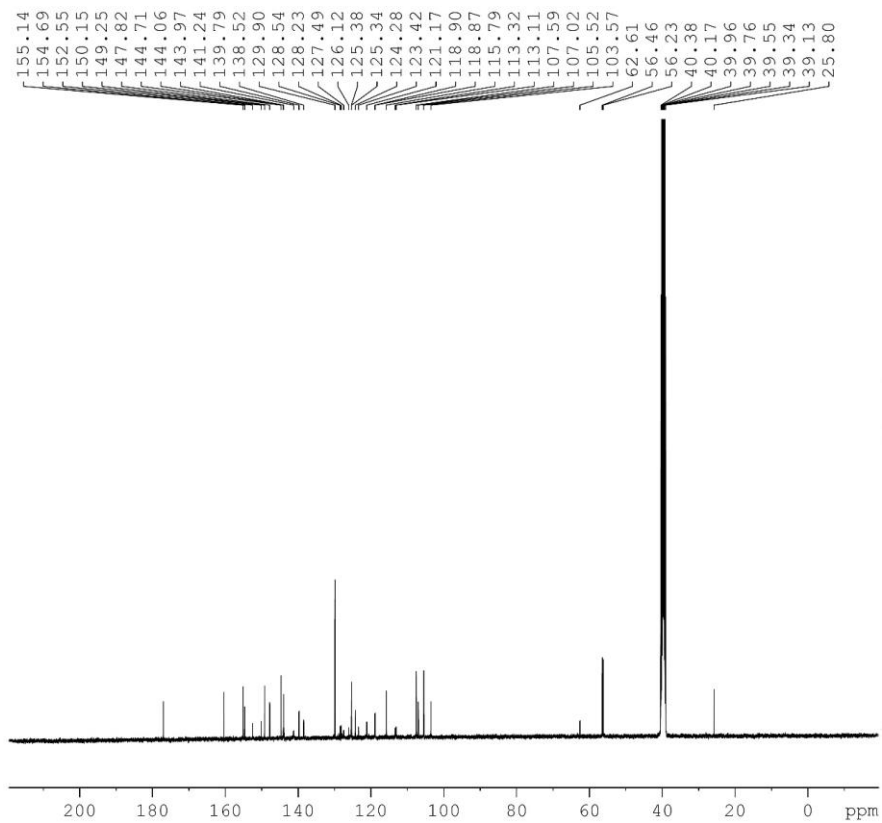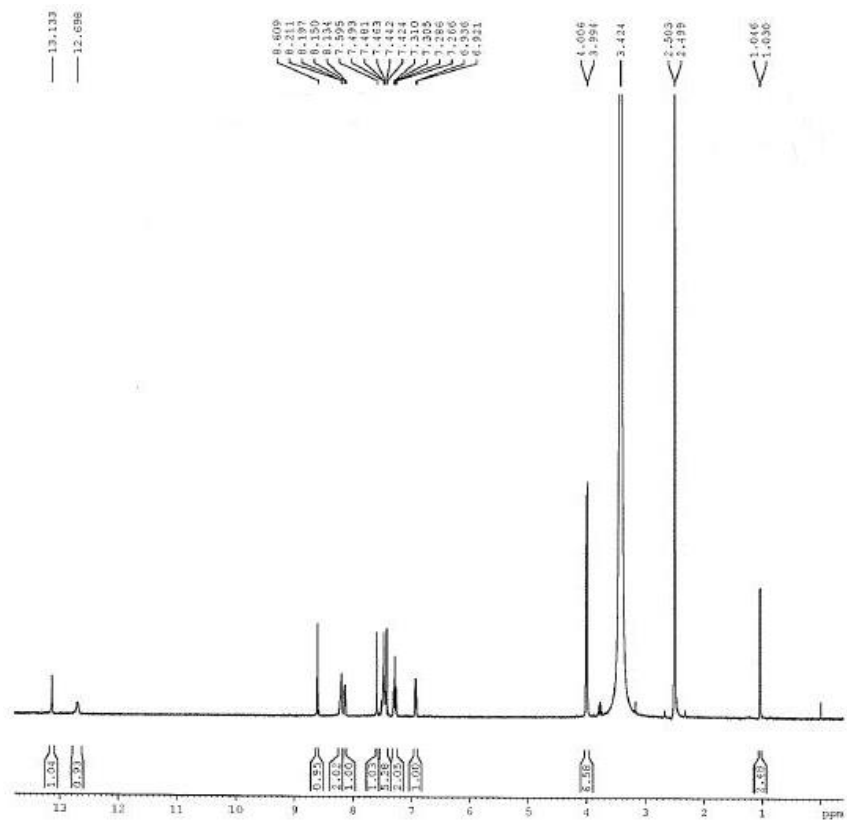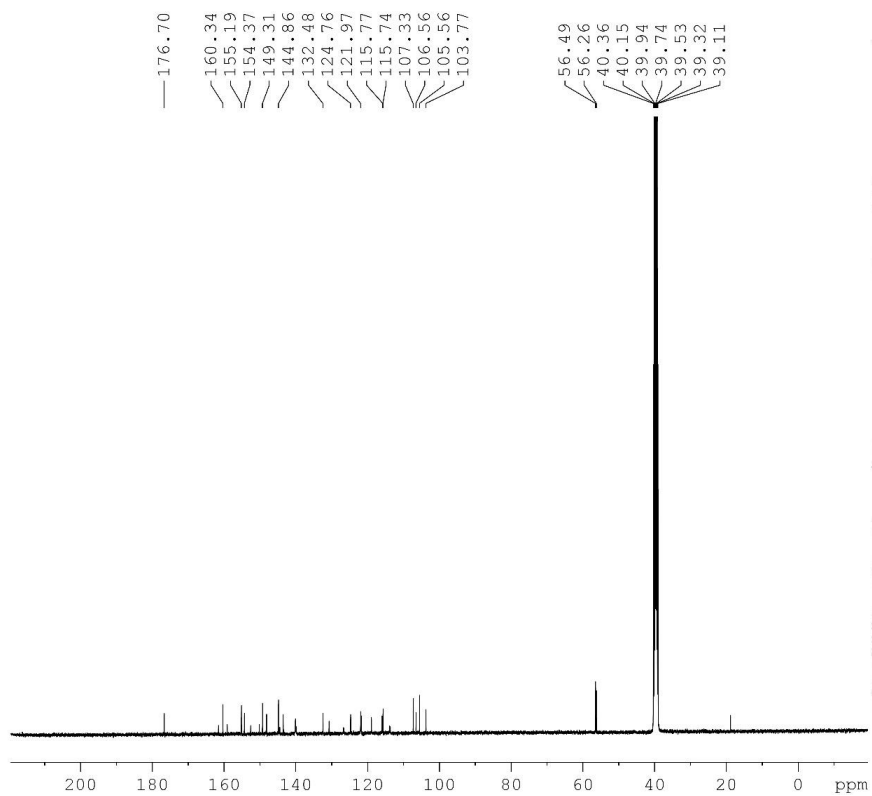**Figure S21.** 400 MHz spectrum of ^{1}H-NMR of compound **5h** (DMSO-d_6), Related to Table 1 and Figure 3.

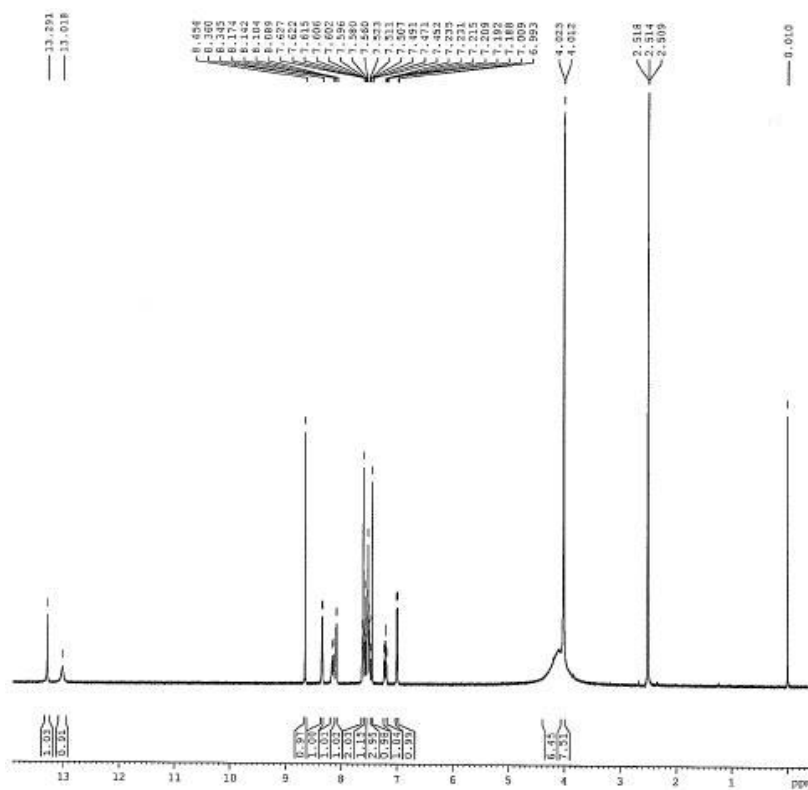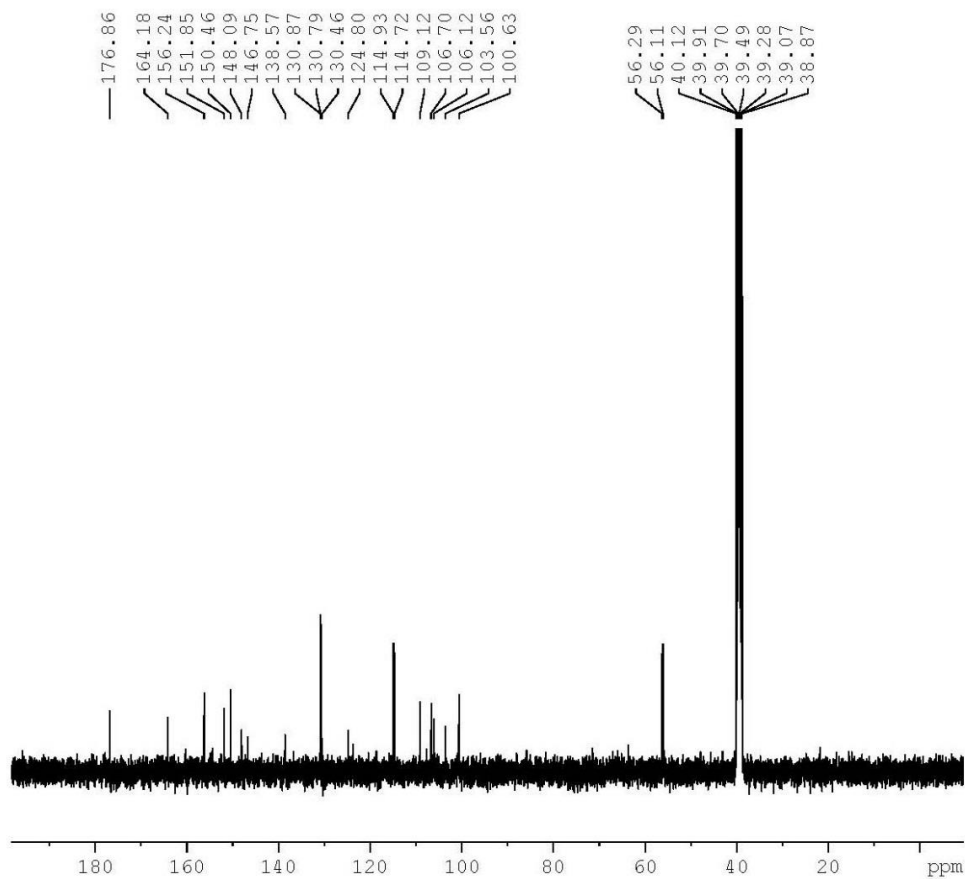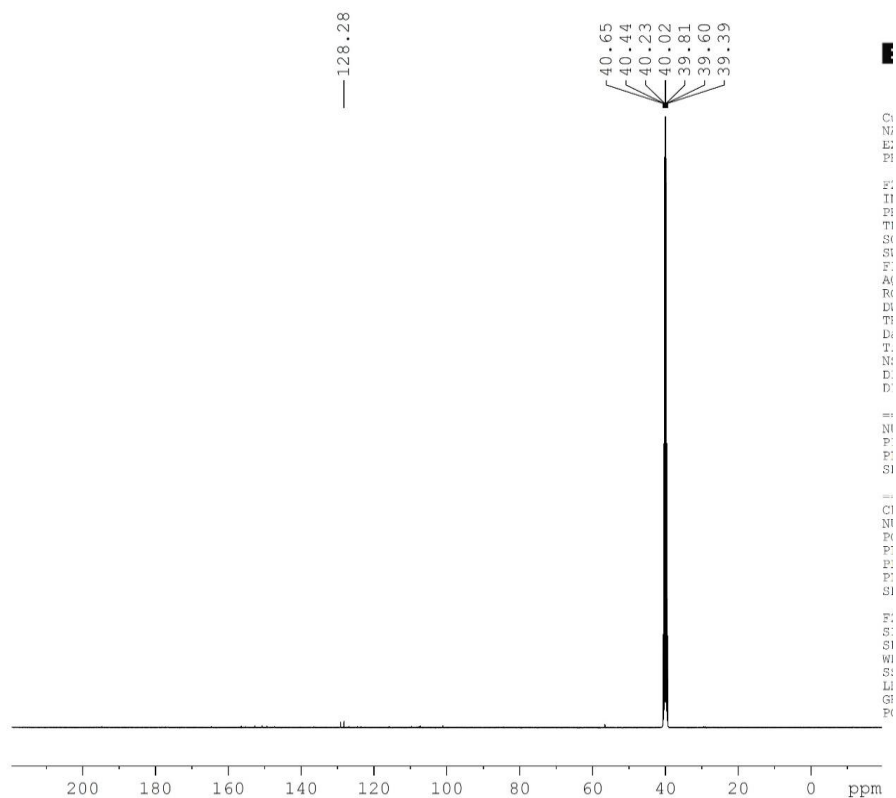**Figure S22.** 100 MHz spectrum of ^{13}C-NMR of compound **5h** (DMSO-d_6), Related to Table 1 and Figure 3.

Table S4. In Vivo PK Profiles of Compound **5g** in Dog and Monkey[a], Related to Table 2.

parameter	dog[b]	monkey[c]
p.o. dose (mg/kg)	4	3
i.v. dose (mg/kg)	2	1.5
C_{max}(μg/mL), p.o.	5.67	2.02
T_{max}(h) , p.o.	2.33	4.00
AUC_{last}(h*μg/mL), p.o.	40.11	24.54
$t_{1/2}$ (h), i.v.	8.25	13.07
CL (L/h/kg), i.v.	0.068	0.041
Vz(L/kg), i.v.	0.69	0.80
F (%)	72.6	48.0

[a] Vehicle: 70% PEG400-30% water; [b] Data reported as the average of six animals; [c] Data reported as the average of four animals.
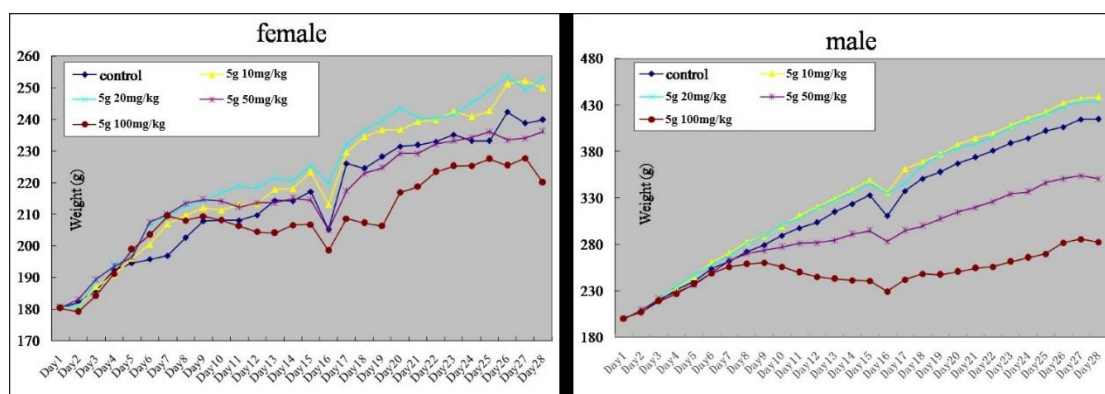
Figure S23. Repeat-dose toxicity studies of **5g** on grouped rats, Related to Table 2.
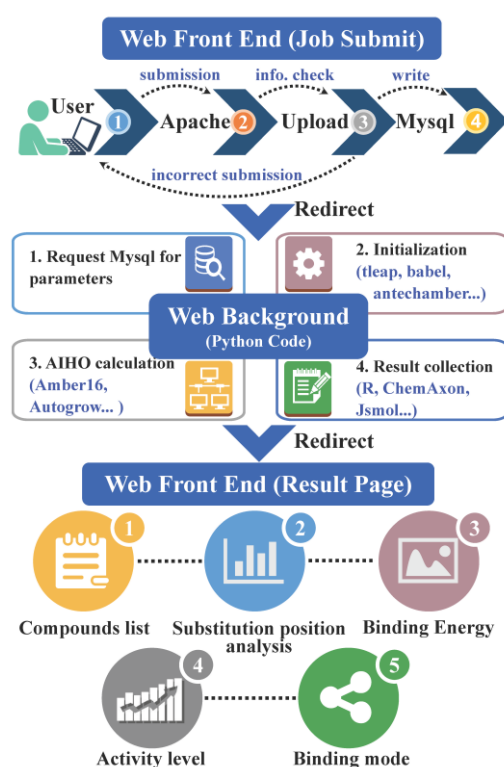
Figure S24. Architecture of AILDE web server. The AILDE is mainly make up of front end and background. The front end includes client layer for submitting jobs and result browse layer for checking jobs. The background is coded by Python for executing calculation process. Related to Figure 6.

Transparent Methods

Small group library construction

The small group library was constructed based on PADFrag (http://chemyang.ccnu.edu.cn/ccb/database/PADFrag/) (Yang et al., 2018) and Molinspiration (https://www.molinspiration.com/), two fragment-based databases. The fragments were derived from

commercial drugs or pesticides. We selected substituents according to three criteria: (i) their frequency of occurrence ranks in the top 50, (ii) their molecular weight (MW) is less than 100 Daltons, (iii) and the substituent is acyclic. Finally, we selected 47 substituents. The three-dimensional structure of every substituent was prepared by using Sybyl2.0. Then, the steepest descent and conjugate gradient methods were used for minimization. Finally, the prepared substituents were saved in PDB format.

MD simulation

We used the hit-receptor 3D complex structure as the starting point for the MD simulation. The structure can be downloaded from Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank (PDB, https://www.rcsb.org/) (Rose et al., 2017) or from a docking result. AM1-BCC charge were assigned for the hit compound by using antechamber module in Amber 16 packages (Case and Kollman, 2016; Jakalian et al., 2000; Jakalian et al., 2002). Then, the complex coordinate and topology parameters were constructed by using Leap module in Amber16 packages. We used AMBER ff14SB as the force filed for the amino acids and general AMBER force field (GAFF) for the hit compound. Cl^- or Na^+ was added to the system to neutralize the system (Maier et al., 2015; Wang and Merz, 2006; Wang et al., 2004). The complex was finally solvated with the TIP3P waters (Price and Brooks, 2004).

The complex was subjected to three-step energy minimization using the Sander module of Amber16 before the MD simulation. First, only waters, ions and hydrogens were allowed to move, and the solute was kept fixed with a constraint of 100 kcal mol^{-1} $Å^{-2}$. Second, we fixed the backbone atoms of the protein and allowed other atoms to move. Third, all the atoms of the system were free to move. The 2000-step steepest descent method and the 2000-step conjugated gradient method were used in the three-step minimization processes. The MD simulation was performed employing the periodic boundary condition with the NPT ensemble. First, a 10-ps simulation was performed on the solvent molecules and ions to obtain an equilibrated solvent environment. Second, the system temperature was gradually heated from 0 to 298 K over 50 ps. The systems were finally maintained at 298 K with a constant pressure until the MD reached 6 ns. The SHAKE algorithm(Ryckaert et al., 1977) was employed to constrain all bonds involving hydrogens. The particle mesh Ewald (PME) algorithm (Darden et al., 1993; Essmann et al., 1995) was used to deal with long-range electrostatic interactions and van der Waals (vdW) energy terms with a cutoff distance of 10 Å. The time step was 2.0 fs, and the coordinates were collected every 1 ps during the MD simulation.

One-step FEP

We extracted 50 snapshots from the last 500 ps of the MD trajectory with a time step of 10 ps by using the Cpptraj module in the Amber16 program. We introduced different small groups to replace hydrogen atoms in the snapshots by using a modified version of AutoGrow (Durrant et al., 2009). Because the growth algorithm started from the hit compound placed in the binding site and was based on the binding conformation, a series of new compounds bound to the receptor were generated as lead compounds.

We tried two different minimization strategies to obtain the final structures of the newly generated lead-receptor complexes. (i) First, we minimized all side chain atoms of the receptor and fixed the receptor backbone and lead compound. Second, we minimized the complex, allowing all atoms to move. Both steps were the combination of the 2000-step steepest descent method and the 2000-step conjugated gradient method with a convergence criterion of 0.2 kcal mol^{-1}. (ii) Based on the final

minimized structure from the first strategy, the side chain atoms were subjected to MD simulation refinement. The only difference between the two strategies was an additional MD simulation in the second strategy compared with the first. Each snapshot of the hit-receptor complex was dealt with using the same minimization strategies to maintain parallel processes.

Free energy calculation

The MM-PBSA method was used to calculate the binding free energy (ΔG_{bind}) between the receptor and ligand (Hou et al., 2011). This value was obtained by calculating the differences in free energies between the ligand-receptor complex (G_{cpx}) and the unbound receptor (G_{rec}) and ligand (G_{lig}) as follows:

$$\Delta G_{bind} = G_{cpx} - (G_{rec} + G_{lig}) \tag{1}$$

The ΔG_{bind} consists of the molecular mechanical (MM) gas-phase binding energy (ΔE_{MM}), solvation free energy (ΔG_{sol}) and entropic contribution ($-T\Delta S$):

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \tag{2}$$

The ΔE_{MM} includes two parts, the electrostatic energies (ΔE_{ele}) and van der Waals interaction (ΔE_{vdW}):

$$\Delta E_{MM} = \Delta E_{ele} + \Delta E_{vdW} \tag{3}$$

The ΔG_{sol} is made up of electrostatic contribution (ΔG_{PB}) and nonelectrostatic contribution (ΔG_{np}) to the solvation free energy. ΔG_{PB} is calculated by the Poisson-Boltzman (PB) method using MM_PBSA module in amber16 program. ΔG_{np} is determined by the solvent accessible surface area (Doree Sitkoff et al., 1994).

$$\Delta G_{sol} = \Delta G_{PB} + \Delta G_{np} \tag{4}$$

For the entropic contribution, an empirical method (Pan et al., 2008) was used, it consists two subitems, the solvation entropy change (ΔS_{sol}) and conformational entropy change (ΔS_{conf}):

$$\Delta S = \Delta S_{sol} + \Delta S_{conf} \tag{5}$$

The ΔS_{sol} is obtained by the tendency of water molecules to minimize their contacts with hydrophobic groups in protein, ΔS_{conf} is related to the change of the number of rotatable bonds during the binding process. The entropic contribution is evaluated by using the procedure developed by Dr. Zhan (Pan et al., 2008), and the conformational entropy change is proportional to the number (ΔN_{rot}) of the lost rotatable bonds during the binding:

$$-T\Delta S_{conf} = w(\Delta N_{rot}) \tag{6}$$

in which w is a scaling factor which was set to be 1 Kcal/mol for the binding energy calculation. Thus, the equation (1) can be written as:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{PB} + \Delta G_{np} - T\Delta S_{sol} + w(\Delta N_{rot}) \tag{7}$$

all other parameters in the energy calculation are the standard parameters or the default values of the Amber16 program.

The binding free energy shift ($\Delta\Delta G$) between the hit-receptor and lead-receptor complex is defined as follows:

$$\Delta\Delta G = \Delta G_{bind}(Lead) - \Delta G_{bind}(Hit) \tag{8}$$

$\Delta\Delta G$ was used to evaluate the change in the binding affinity after group replacement. A positive $\Delta\Delta G$ value indicated a decrease in the binding affinity, and a negative value represents an increase in the binding affinity. The experimental bioactivity fold (F) value was defined as follows:

$$F = \frac{K_i(\text{Lead})}{K_i(\text{Hit})} \tag{9}$$

in which K_i(Lead) and K_i(Hit) are the dissociation constants for the lead-receptor and hit-receptor complex, respectively. The experimental binding free energy shifts ($\Delta\Delta G_{exp}$) can be calculated according to the following equation:(Ho et al., 1995; Smith et al., 1998; Wang et al., 2001)

$$\Delta\Delta G_{exp} = -RT\ln \frac{K_i(\text{Lead})}{K_i(\text{Hit})} = -RT\ln F \approx -RT\ln \frac{\text{IC}_{50}(\text{Lead})}{\text{IC}_{50}(\text{Hit})} \tag{10}$$

Therefore, we built the relationship between F and $\Delta\Delta G_{exp}$ according to the activity values of the compound, such as IC_{50}, K_i and K_d.

Biochemical kinase assays

The inhibitory activities of the compounds against various kinases (Invitrogen) were measured using an HTRF KinEASE TK kit (Cisbio) following the manufacturer's instructions. Briefly, after incubating the kinase with the compound for 5 minutes at 25 to 30 °C, the kinase reaction was initiated by adding 2 μL of a mixed substrate solution (mixture of biotin substrate and ATP (Sigma)). The final kinase concentration was EC_{80}, and the total reaction volume was 8 μL. The plates were incubated at 30 °C for 30 to 60 minutes. The reactions were then quenched by adding 8 μL of the detection mixture (streptavidin-XL665 and antibody-encrypted antibody in buffer). The fluorescence was measured at 665 nm and 620 nm using a PHERAstar FS plate reader (BMG) with a time delay of 50 μs. We performed all the kinase assays with kinase-specific biotinylated substrate peptides and ATP concentrations below the enzyme K_{mapp}.

We plotted the dose-response curves as inhibition rate versus $\log_{10}$ of the drug concentration, and the IC_{50} values were calculated by nonlinear regression using GraphPad Prism 5. The inhibition rates were calculated with the following formula: $100 \times [1 - (U_1 - C_2) / (C_1 - C_2)]$, where U is the emission ratio of 620 nm and 665 nm of the test sample; C_1 is the average value obtained for the solvent control (2% DMSO), and C_2 is the average value obtained for the negative control.

Cellular kinase phosphorylation ELISA assays

MKN-45 cells were seeded in 24-well plates in complete medium [RPMI-1640 (Gibco) + 10% fetal bovine serum (FBS, Gibco)]. The cells were starved with serum-free medium [RPMI-1640 + 0.05% bovine serum albumin (BSA)] overnight. The next day, the cells were treated with the test compounds at various concentrations for 1 h, and then 80 ng/mL HGF (R&D system) or vehicle was added, and the plates were incubated for an additional 10 minutes. Then, the cells were lysed, and the phosphorylation of c-Met was assessed by a sandwich ELISA developed in-house. c-Met capture antibody (R&D System) was used to coat 96-well plates. The cell lysate was added to the antibody-coated plate and incubated at 30 °C for 2 h. At the end of the incubation period, the plates were washed with PBST (PBS+0.05% Tween 20) three times, and then the samples were incubated with a detection antibody (4G10, 1:2000, Upstate) specific for phosphorylated tyrosine residues for 1 h. The detection antibody was washed away with PBST three times. Then, a secondary antibody (horseradish peroxidase–conjugated anti-mouse IgG) was loaded onto the plates, and they were incubated at 30 °C for 1 h. Finally, the secondary antibody was washed away, and 3,3,5,5-tetramethyl benzidine peroxidase substrate (Pierce) was added to the plates to initiate a colorimetric reaction. The reaction was stopped by adding 2 M H_2SO_4, and the light absorbance at 450 nm was measured with an Infinite M200 spectrophotometer (Tecan). IC_{50} values were calculated as described in the kinase assay method.

Pharmacokinetic profiles in SD rats

The pharmacokinetic parameters of compound **5g** were subjected to PK studies on male SD rats (provided by Shanghai Slake Experimental Animal Co., Ltd.) weighing between 180 and 200 g with four animals in each group. The tested compound **5g** (a solution of 70% PEG400 + 30% water) administered to male SD rats at a dose of 5 mg/kg (i.v.) or 10 mg/kg (p.o.). Blood samples (0.3 mL) were collected at the point including 2 min, 5 min, 15 min, 30 min, 1 h, 2 h, 4 h, 6 h, 8 h, 10 h and 24 h (i.v.) or 15 min, 30 min, 45 min, 1 h, 2 h, 4 h, 6 h, 8 h, 10 h and 24 h (p.o.) after administration and centrifuged at 8000 rpm for 5 min at 4 °C, and then analyzed after protein precipitation. To obtain the best sensitivity and selectivity of the analyte, the LC/MS/MS analysis of compound **5g** was carried out under optimized conditions in SRM (selected reaction monitoring) mode containing an internal standard. Plasma concentration-time data were measured by a noncompartmental approach using the software WinNonlin Enterprise, version 5.2 (Pharsight Co., Mountain View, CA).

Pharmacokinetic profiles in beagle dogs

The pharmacokinetic parameters of compound **5g** were subjected to PK studies on male beagle dogs (batch: 20110729-2, Beijing Marshall Biotechnology Co., Ltd.) weighing 10kg with three animals in each group. The tested compound **5g** (a solution of 70% PEG400 + 30% water) was administered to male beagle dogs at dose of 2 mg/kg (i.v.) or 4 mg/kg (p.o.). Blood samples (1 mL) were collected before the dosage and at 5 min, 10 min, 30 min, 1 h, 2 h, 4 h, 6 h, 8 h, 10 h and 24 h (i.v.) or 15 min, 30 min, 1 h, 2 h, 3 h, 4 h, 6 h, 8 h, 10 h and 24 h (p.o.) after administration and centrifuged to separate plasma. The separated plasma was transferred into 96-well plates and kept frozen (<-60°C) until LC/MS/MS detection.

Pharmacokinetic profiles in cynomolgus monkeys

The pharmacokinetic parameters of compound **5g** were subjected to PK studies on male cynomolgus monkeys (Provided by Zhaoqing Chuangyao Biological Technology Co., Ltd.) weighing 4kg with two animals in each group. The tested compound **5g** (a solution of 70% PEG400 + 30% water) was administered to male cynomolgus monkeys at a dose of 1.5 mg/kg (i.v.) or 3 mg/kg (p.o.). Blood samples (0.5 mL) were collected before the dosage and at 5 min, 10 min, 30 min, 1 h, 2 h, 4 h, 6 h, 8 h, 12 h and 24 h (i.v.) or 15 min, 30 min, 1 h, 2 h, 3 h, 4 h, 6 h, 8 h, 12 h and 24 h (p.o.) after administration and centrifuged to separate plasma. The separated plasma was transferred into 96-well plates and kept frozen (<-60°C) until LC/MS/MS detection.

***In vivo* antitumor activity assay**

The *in vivo* activities of the compounds were evaluated in a mouse xenograft model according to institutional ethical guidelines for animal care. Tumor cells were inoculated into the flanks of athymic nude mice (2×10^6 cells/mouse). When the tumor volume reached 100-200 mm^3, the mice were randomly divided into compound **5g**-treated and vehicle-treated groups. Mice in the vehicle-treated groups were given vehicle alone, and mice in the compound **5g**-treated groups received compound **5g** via p.o. administration at 8.3, 25, and 75 mg/kg every day for 3 weeks. The sizes of the tumors were measured 3 times per week using a microcaliper. The tumor volume (TV = (length $\times$width^2)/2) for the indicated days is the median tumor volume in each group (+ SEM). The percentage of tumor growth inhibition (%TGI = [1-(TV$_{5g\_treated\_final\ day}$ −TV$_{5g\_treated\ day\ 1}$)/(TV$_{vehicle\_treated\_final\ day}$−TV$_{vehicle\_treated\_final\ 1}$)]$\times$100%) was used to evaluate the antitumor efficacy of the compound. Tumor volumes were

statistically analyzed using Student's paired *t*-tests.

Web server configuration

The web server was implemented on an Apache server running on a CentOS7 enterprise Linux-based operating system. The web interface was coded in JavaScript and PHP. The calculation modules and accessory scripts were coded in Python2.7 and R-3.0.0. They were manipulated to run by Maui-3.3.1 and Torque-6.0.1. Job information and calculation result data were stored in the database by using MySQL (version 14.12).

Application example

We used TNNI3K (Troponin I-interacting protein kinase)-benzenesulfonamide system as an example to describe the detailed calculation process of AILDE web server. TNNI3K is a tyrosine-like kinase attracts broader concerns in recent studies. Benzenesulfonamides is one kind of its inhibitors. The co-crystal structure (PDB ID: 4YFI) of a benzenesulfonamide inhibitor bounded with TNNI3K is downloaded from Protein Data Bank (PDB, https://www.rcsb.org/). The activity data of the hit compound and its substituted analogs (13 compounds in total) were collected from reference.(Lawhorn et al., 2016) Firstly, the complex structure was submitted to AILDE for a MD simulation. Then the snapshots were obtained from equilibrated trajectory and we performed the one-step FEP at the R^4 site of the hit compound (shown in Figure S2, a) to obtain the initial structures of other 12 analogs bound with the TNNI3K. The two minimization strategies were carried out on both hit-receptor and analogs-receptor complexes to obtain the final conformations. Finally, two groups of $\Delta\Delta G_{exp}$ was acquired for first strategy ($\Delta\Delta G_{cal-s1}$) and second strategy ($\Delta\Delta G_{exp-s2}$) by performing the MM/PBSA calculation (shown in Figure S2, b). The $\Delta\Delta G_{exp}$ was obtained by using the experimental IC_{50} value according to equation (4). The scatter plots of $\Delta\Delta G_{exp}$-$\Delta\Delta G_{cal-s1}$ and $\Delta\Delta G_{exp}$-$\Delta\Delta G_{cal-s2}$ were shown in Figure S2, d with the $R^2 = 0.40$ and 0.75 respectively. The $\Delta\Delta G_{exp}$ ranges from -0.96 to 2.60 Kcal/mol, $\Delta\Delta G_{cal-s1}$ ranges from -1.30 to 1.31 Kcal/mol, and $\Delta\Delta G_{cal-s2}$ ranges from -1.13 to 2.43 Kcal/mol. We note that the $\Delta\Delta G_{cal-s2}$ (red circle on Figure S2, d) has a more fitting range to $\Delta\Delta G_{exp}$ than $\Delta\Delta G_{cal-s1}$ (blue cross on Figure S2, d) does. 12 out of 13 samples were identified correctly to be positive or negative samples for the second minimization strategy. Only compound 7 was slightly underestimated on $\Delta\Delta G_{cal-s2}$ (-0.07 Kcal/mol) compared with $\Delta\Delta G_{exp}$ (1.23 Kcal/mol). Take compound 5 as an example, the $\Delta\Delta G_{exp}$ is 2.05 kcal mol^{-1}, it feedbacks a predicted value -0.22 kcal mol^{-1} by using the first minimization strategy. It is a big deviation. But the second minimization strategy output a preferable value (2.37 kcal mol^{-1}). We aligned the final complex structures of compound 5-TNNI3K from two minimization strategies (shown in Figure S2, c). We found the replaced R^4 group has an electrostatic clash with the side chain of His592 in the first strategy's structure (green sticks in Figure S2, c) that leads the unreasonable binding mode. However, the side chain of His592 has a rotation to keep away from the electrostatic clash after a MD refinement step (second strategy) (magenta sticks on Figure S2, c). The more reasonable structure lead to the better predicted result. Thus, we consider AILDE to be strongly of the predictive power in the design of TNNI3K inhibitors with quantitative accuracy $R^2=0.75$.

Docking experiment

The 3D strcture of compound **5a** was constructed and primarily optimized by using Sybyl 2.0 software. The crystal structure of c-Met was obtained from Protein Data Bank (PDB ID: 3F82). The

hydrogens of the receptor structure were added by using Discovery Studio 4.0. The original ligand in the protein was used as the reference to define the active site. The GOLD program was used to perform the molecular docking. The radius of active site was set to be 10Å. Genetic algorithm was used for the conformation searching. Other parameters were set to be default. Finally, 100 conformation was generated and top ranked conformation was selected as the binding conformation. The binding mode of 5a with c-Met was shown in Figure S3.

Synthetic chemistry

A series of target compounds and their intermediates were synthesized according to the pathways described in Figure S5-S6 (Tasler et al., 2009). As shown in Figure S5, commercially available 4-chlorothieno[3,2-d]pyrimidine was converted to **1** by nucleophilic substitution with 2-fluoro-4-nitrophenol under DIPEA in toluene, subsequent reduction of nitro easily afforded intermediate **2**. Similarly, intermediates **3** also was obtained from relevant commercially available chlorinated derivatives by the same method as **2**. The diverse chlorinated 1,6-naphthyridone **4** could be easily obtained by our reported methods (Huang et al., 2013). With intermediate **2**, **3** and **4** in hand, compounds **5a-h** were successfully synthesized through the direct reaction of **4** with **2** or **3** in a hydrochloric acid-isopropanol system and subsequent alkalize reaction in Et_3N-DCM system.

General procedures for the synthesis of intermediates

4-(2-fluoro-4-nitrophenoxy)thieno[3,2-d]pyrimidine (**1**)

A solution of 4-chlorothieno[3,2-d]pyrimidine (500 mg, 2.93 mmol), 2-fluoro-4-nitrophenol (3.5 mmol) and DIPEA (8.79 mmol) in toluene (20 mL) was refluxed for 8 hour, then cooled to room temperature. The reaction mixture was concentrated under reduced pressure, and purified by chromatography (PE/EA = 30:1) to yield **1** as a yellow solid (725 mg, 85%).

3-fluoro-4-(thieno[3,2-d]pyrimidin-4-yloxy)aniline (**2**)

A mixture of iron powder (595 mmol), ammonium chloride (255 mmol), **1** (700 mg, 2.4 mmol), ethanol (15 ml) and water (5 ml) was refluxed for 3 hours. The mixture was filtered through celite and washed with EtOAc. The organic layer was washed with water and Sat. NaCl, dried over Na_2SO_4, and concentrated to afford the product (577 mg, 92%).

General procedures for the synthesis of targets 5a-h

5-((3-fluoro-4-(thieno[3,2-d]pyrimidin-4-yloxy)phenyl)amino)-3-(4-methoxyphenyl)-1,6-naphthyridin-4(1H)-one (**5a**)

A solution of **4a** (200 mg, 0.7 mmol), **2** (0.7 mmol) and HCl (20 mmol%) in isopropanol (10 mL) was heated to 90 °C under nitrogen for 2 h. The mixture was filtered, and the solid was dissolved in ethyl acetate. The solution was stirred with K_2CO_3 (0.5 mmol) at r.t. for 1h and filtered. The filtrate was concentrated in vacuum and purified by flash chromatography (CH_2Cl_2/MeOH = 20:1) to yield title compound as yellow solid (242 mg, 68%). mp 253-254 °C ^{1}H-NMR (600 MHz, DMSO-d_6) δ 13.47 (s, 1H), 12.99 (s, 1H), 8.76 (s, 1H), 8.55 (d, J = 5.4 Hz, 1H), 8.27 – 8.21 (m, 1H), 8.17 – 8.08 (m, 1H), 8.06 – 7.98 (m, 1H), 7.74 (d, J = 5.4 Hz, 1H), 7.69 – 7.63 (m, 2H), 7.62 – 7.56 (m, 1H), 7.55 – 7.48 (m, 1H), 7.07 – 6.96 (m, 3H), 3.81 (s, 3H). ^{13}C-NMR (100 M, DMSO-d6) δ 177.0, 163.5, 162.8, 158.8, 154.3, 154.0, 146.6, 137.8, 129.9, 126.2, 124.9, 124.7, 124.3, 116.3, 113.5, 106.6, 103.7, 55.1. HR-MS (EI) m/z calcd for $C_{27}H_{18}FN_5O_3S$, 511.1114; found 512.1187 $[M+H]^+$.

5-((4-((6,7-dimethoxyquinolin-4-yl)oxy)-3-fluorophenyl)amino)-3-(4-methoxyphenyl)-1,6-naphth yridin-4(1H)-one (**5b**)

Prepared according to the procedure for the preparation of **5a**, from **4a** and **3a**, to yield title compound as yellow solid (60%). mp 263-264 °C ^{1}H NMR (600 MHz, DMSO-d_6) δ 13.47 (s, 1H), 12.94 (s, 1H), 8.88 (d, J = 6.6 Hz, 1H), 8.42 – 8.28 (m, 1H), 8.26 – 8.17 (m, 1H), 8.16 – 8.04 (m, 1H), 7.87 – 7.57 (m, 6H), 7.28 – 7.13 (m, 1H), 7.10 – 6.95 (m, 3H), 4.06 (br, 6H), 3.81 (s, 3H). ^{13}C-NMR (100 M, DMSO-d6) δ 177.6, 159.1, 155.1, 154.9, 148.6, 146.6, 145.1, 138.3, 137.6, 130.4, 130.2, 126.7, 124.9, 118.7, 118.5, 114.0, 107.9, 107.3, 104.5, 104.2, 98.8, 56.4, 56.3, 55.6. HR-MS (EI) m/z calcd for $C_{32}H_{25}FN_4O_5$, 564.1809; found 565.1882 [M+H]$^+$.

5-((4-((6,7-dimethoxyquinazolin-4-yl)oxy)-3-fluorophenyl)amino)-3-(4-methoxyphenyl)-1,6-naph thyridin-4(1H)-one (**5c**)

Prepared according to the procedure for the preparation of **5a**, from **4a** and **3b**, to yield title compound as yellow solid (64%). mp 240-241 °C. ^{1}H-NMR (400 M, DMSO-d_6) δ 13.61 (s, 1H), 12.82 (bs, 1H), 8.48 (d, J = 12.0 Hz, 1H), 8.39 (d, J = 7.2 Hz, 1H), 8.17-8.22 (m, 2H), 7.64-7.75 (m, 5H), 7.01-7.03 (m, 3H), 6.79 (d, J = 7.6 Hz, 1H), 6.59 (s, 1H), 3.95 (s, 3H), 3.82 (s, 3H), 3.75 (s, 3H); ^{13}C-NMR (100 M, DMSO-d_6) δ 177.6, 172.5, 159.1, 158.5, 155.1, 154.9, 148.7, 146.6, 145.2, 141.2, 138.3, 137.6, 130.4, 130.2, 126.7, 124.9, 123.0, 118.7, 118.5, 114.0, 109.0, 107.9, 107.3. 104.5, 104.2, 98.8, 56.4, 56.3, 55.6. MS (ESI). 565.2 [M]$^+$. HR-MS (EI) m/z calcd for $C_{31}H_{24}FN_5O_5$, 565.1761; found 566.1839 [M+H]$^+$.

5-((4-((6,7-dimethoxyquinazolin-4-yl)oxy)-3-fluorophenyl)amino)-3-(4-(trifluoromethyl)phenyl)-1,6-naphthyridin-4(1H)-one (**5d**)

Prepared according to the procedure for the preparation of **5a**, from **4b** and **3b**, to yield title compound as yellow solid (68%). mp 218-219 °C. 1H-NMR (400 M, DMSO-d6) δ 13.18 (s, 1H), 12.84 (bs, 1H), 8.62 (s, 1H), 8.36 (d, J = 6.0 Hz, 1H), 8.22 (d, J = 13.2 Hz, 1H), 8.15 (d, J = 6.0 Hz, 1H), 7.95 (d, J = 8.0 Hz, 2H), 7.81 (d, J = 8.0 Hz, 2H), 7.60 (s, 1H), 7.45-7.51 (m, 2H), 7.43 (s, 1H), 6.95 (d, J = 6.0 Hz, 1H), 4.01 (s, 3H), 3.99 (s, 3H); ^{13}C-NMR (100 M, DMSO-d6) δ 177.1, 160.5, 155.1, 154.7, 152.6, 150.2, 149.3, 147.8, 144.7, 144.0, 141.2, 139.8, 138.5, 129.9, 126.1, 125.4, 124.3, 123.4, 118.9, 115.8, 113.3, 107.6, 107.0. 105.5, 103.6, 56.4, 56.2. MS (ESI). 604.2 [M+H]+. HR-MS (EI) m/z calcd for C31H21F4N5O4, 603.1530; found 604.1617 [M+H]$^+$.

5-((4-((6,7-dimethoxyquinazolin-4-yl)oxy)-3-fluorophenyl)amino)-3-(2-fluorophenyl)-1,6-naphth yridin-4(1H)-one (**5e**)

Prepared according to the procedure for the preparation of **5a**, from **4c** and **3b**, to yield title compound as yellow solid (73%). mp 270-271 °C. ^{1}H-NMR (400 M, DMSO-d_6) δ 13.29 (s, 1H), 13.02 (bs, 1H), 8.66 (s, 1H), 8.35 (d, J = 6.0 Hz, 1H), 8.15 (d, J = 12.0 Hz, 1H), 8.09 (d, J = 6.0 Hz, 1H), 7.56-7.63 (m, 3H), 7.47-7.52 (m, 3H), 7.45 (s, 1H), 7.19-7.24 (m, 1H), 7.00 (d, J = 6.4 Hz, 1H), 4.01 (s, 6H); ^{13}C-NMR (100 M, DMSO-d_6) δ 177.0, 163.5, 161.1, 160.5, 155.1, 154.5, 152.6, 149.2, 147.7, 144.6, 144.1, 140.2, 139.4, 136.4, 130.6, 127.6, 125.1, 124.5, 121.5, 118.9, 116.0, 115.0, 114.8, 113.6, 107.6, 106.9. 105.4, 103.6, 56.6, 56.4. MS (ESI). 554.2 [M+H]$^+$. HR-MS (EI) m/z calcd for $C_{30}H_{21}F_2N_5O_4$, 553.1562; found 554.1660 [M+H]$^+$.

5-((4-((6,7-dimethoxyquinazolin-4-yl)oxy)-3-fluorophenyl)amino)-3-(3-fluorophenyl)-1,6-naphth

yridin-4(1H)-one (**5f**)

Prepared according to the procedure for the preparation of **5a**, from **4d** and **3b**, to yield title compound as yellow solid (73%). mp 262-263 °C. ^{1}H-NMR (400 M, DMSO-d_6) δ 13.13 (s, 1H), 12.70 (bs, 1H), 8.61 (s, 1H), 8.19-8.21 (m, 2H), 8.14 (d, J = 6.4 Hz, 1H), 7.59 (s, 1H),7.42-7.49 (m, 5H), 7.26-7.31 (t, J = 8.0 Hz, 2H), 6.93 (d, J = 6.0 Hz, 1H), 4.01 (s, 6H); ^{13}C-NMR (100 M, DMSO-d_6) δ 176.7, 161.3, 160.3, 158.3, 155.2, 154.4, 150.5, 149.3, 148.1, 144.9, 143.5, 140.4, 132.5, 130.7, 125.8, 124.8, 122.0, 119.2, 115.8, 115.7, 113.4, 107.3, 106.6. 105.6, 103.8, 56.5, 56.2. MS (ESI). 554.2 [M+H]$^+$. HR-MS (EI) m/z calcd for $C_{30}H_{21}F_2N_5O_4$, 553.1562; found 554.1654 [M+H]$^+$.

5-((4-((6,7-dimethoxyquinazolin-4-yl)oxy)-3-fluorophenyl)amino)-3-(4-fluorophenyl)-1,6-naphth yridin-4(1H)-one (**5g**)

Prepared according to the procedure for the preparation of **5a**, from **4e** and **3b**, to yield title compound as yellow solid (78%). mp 280-281 °C. ^{1}H-NMR (400 M, DMSO-d_6) δ 13.41 (s, 1H), 13.23 (bs, 1H), 8.68 (s, 1H), 8.31 (d, J = 5.6 Hz, 1H), 8.08 (d, J = 12.0 Hz, 1H), 8.03 (d, J = 5.6 Hz, 1H), 7.75 (t, J = 8.8 Hz, 2H), 7.61 (s, 1H), 7.55-7.61 (m, 2H), 7.46 (s, 1H), 7.29 (t, J = 9.2 Hz, 2H), 7.04 (d, J = 6.8 Hz, 1H), 4.01 (s, 6H); ^{13}C-NMR (100 M, DMSO-d_6) δ 176.9, 164.2, 160.6, 156.3, 154.2, 152.6, 151.9, 150.5, 148.1, 146.8, 144.9, 138.6, 136.2, 130.9, 130.5, 124.8, 123.5, 114.9, 114.7, 109.1, 108.0, 106.7, 106.1, 104.0, 100.6, 59.3, 56.1. MS (ESI). 554.1 [M+H]$^+$. HR-MS (EI) m/z calcd for $C_{30}H_{21}F_2N_5O_4$, 553.1562; found 554.1659 [M+H]$^+$.

5-((4-((6,7-dimethoxyquinazolin-4-yl)oxy)-3-fluorophenyl)amino)-3-phenyl-1,6-naphthyridin-4(1 H)-one (**5h**)

Prepared according to the procedure for the preparation of **5a**, from **4f** and **3b**, to yield title compound as yellow solid (55%). mp 252-253 °C. 1H-NMR (400 M, DMSO-d6) δ 13.59 (s, 1H), 12.88 (bs, 1H), 8.57 (s, 1H), 8.32-8.35 (m, 1H), 8.15 (s, 1H), 8.07 (d, J = 6.4 Hz, 1H), 7.69 (d, J = 5.6 Hz, 2H),7.58 (s, 1H), 7.44-7.46 (m, 1H), 7.35-7.41 (m, 1H), 7.28-7.30 (m, 1H), 3.99 (s, 6H); 13C-NMR (100 M, DMSO-d6) δ 177.6, 164.6, 156.3, 156.2, 155.2, 152.8, 152.7, 150.7, 149.4, 148.9, 146.5, 139.6, 138.8, 135.3, 133.8, 129.3, 128.4, 127.6, 124.7, 124.3, 116.2, 109.6, 108.3, 107.4, 107.2, 103.6, 101.0, 56.6, 55.4. MS (ESI). 536.2 [M+H]+. HR-MS (EI) m/z calcd for C30H22FN5O4, 535.1656; found 558.1544 [M+Na]$^+$.

Cocrystallization method

1. Protein purification and crystallization

The kinase domain (1038-1346 aa) of recombinant human c-Met was generated based on the protocols of Wang *et al.* with certain modifications (Wang et al., 2006). The cDNA fragment was cloned into the vector pET28a and the protein was co-expressed with catYopH subcloned in pET15b (164-468AA) (Seeliger et al., 2005). The expressed c-Met kinase domain was passed through a Ni-NTA column (Qiagen) and further purified by QHP ion exchange column (GE) which eluted with 25 mM Tris pH 8.5, 100 mM NaCl, 10% glycerol, 1 mM DTT. The protein was concentrated to ~10 mg/mL for further crystallization.

Cocrystallization of the c-Met kinase domain with compound **5i** was carried out by mixing a solution of the protein-ligand complex with an equal volume of precipitant solution (100 mM Hepes pH 7.5, 16% PEG 4000, 8% isopropanol, 3 mM TCEP). The protein-ligand complex was prepared by adding the compound to the protein solution to a final concentration of 1 mM of **5i**. Cocrystallization

utilized the vapour-diffusion method in hanging drops. Crystals were flash frozen in liquid nitrogen in the presence of well solution supplemented with 25% PEG 4000.

2. Structure determination and refinement

Data were collected at 100 K on beamline BL17U at the Shanghai Synchrotron Radiation Facility (SSRF), and were processed with the XDS software packages (Kabsch, 2010). The structure was solved by molecular replacement, using the program PHASER (Afinogenov, 1997) with the search model of PDB ID 3U6H. The structure was refined with PHENIX (Adams et al., 2010). With the aid of the program Coot (Emsley et al., 2010), compound **5i**, water molecules were fitted into to the initial Fo-Fc map.

Repeat-dose toxicity studies

1. Study design

Seventy SD rats were divided into vehicle control and 4 test article groups in a randomizedmanner to achieve similar group mean body weights. Each group was composed of 7 male and 7 female rats. Vehicle control group was dosed orally with 0.5% CMC-Na solution while test article groups were administrated at levels of 10, 20, 50 and 100 mg/kg **5g** once daily for 28 days. During the study, mortality, body weight, blood chemistry, hematology and pharmacokinetics parameters were analyzed to evaluated gross toxicologic effects on rats.

2. Results

(1) Mortality and gross pathology

All groups' rats survived until the end of sacrifice except the 50 and 100 mg/kg group in male with 1 and 2 rats died respectively. Male rats in treated with 20 mg/kg **5g** were found minimal ascites and abdominal fluid retention were moderate and severe in 50 and 100 mg/kg groups respectively.

(2) Body weight

Body weight were measured upon arrival of animals, before group assignment and once weekly thereafter. There were no treatment-related or statistically significant differences in the mean body weight of female and 10 and 20 mg/kg male groups compared to vehicle control. The growth rate of body weight in 50 and 100 mg/kg group decrease to 32.2% and 70.0% versus vehicle control.

(3) Clinical pathology

There were no apparent changes in **5g** treated group in hematology parameters. Referring to clinical chemistry, 10 mg/kg group were not affected whereas the indexes of liver and kidney of 20, 50 and 100 mg/kg group were influenced in a linear dose-response relationship.

(4) Pharmacokinetics

The pharmacokinetic properties of **5g** following oral administration in rats have been evaluated at dose from 20 to 100 mg/kg. Both C_{max} and AUC parameters did not increase in a linear dose-response relationship while both of them increased in Day 28 than Day 1 except AUC of 100 mg/kg at the end.

3. Conclusion

Collectively, these data illustrate that NOAEL of this study is 10 mg/kg/day, MTD is 20 mg/kg/day ($AUC_{MTD} = 12903$ h·μg/mL). This may provide sufficient safety margin (> 40) when comparing with the AUC of ED_{50} 8.3 mg/kg ($AUC_{MED} = 271$ h·μg/mL).

The usage of AILDE server

A user-friendly web page is important for users to have a convenient and fast way to use the web server. Three important AILDE web pages are introduce here, 'Submit', 'Jobs' and 'Help'. They were coded by using PHP and JavaScript. The 'Submit' page is used to submit jobs. To minimize the difficulties of submitting jobs, a protein-ligand complex with PDB format is only required to upload to minimize the difficulties of submitting jobs. Other parameters are optional, including task name, E-mail, password, and RMSF (root-mean-square-fluctuation). E-mail is used to receive notice about the job status. Password option keeps your job private. The RMSF parameter provides alternative minimization strategy for hit-receptor and lead-receptor complexes at the one-step FEP scanning. If default, first minimization strategy will be performed on both hit-receptor and lead-receptor complexes. If a value (should be between 0 and 1) is assigned, the second minimization strategy will be carried out. The background program will read the parameters and carry out the whole calculation procedure including system initialization, MD simulation, one-step FEP scanning, free energy evaluation and result collection. Considering computational time cost, we only provide 10 most used substitution groups ($-CH_3$, $-Br$, $-CF_3$, $-Cl$, $-F$, $-COOH$, $-NH_2$, $-NO_2$, $-OCH_3$, $-OH$) on AILDE web server. The 'Jobs' page is used to check job status. It offers an entrance for user to browse the job results once the task is finished. Firstly, a list of newly generated lead compounds ranked by activity fold change with the decreasing order is returned on the result web page. Detailed results were also provided for every lead, with the 2D image of compound structure, substituent position, substitution group, energy shift value. The search function may be used to find specific compound by indexing group type or substituent position. The server also supplies the 3D visualization of the binding mode for every candidate compound with the receptor by using JSmol applet (http://www.jmol.org/).(Hanson, 2016) A heat map reflecting the relationship matrix between substituent positions and groups was also shown for a better visual understanding. A histogram illustrates the overall result counted based on substituent positions. It may help users to conclude which substituents position are most potential to be transformed for getting more activity compounds. The 'Help' page provides a detail description and explanation about how to submit jobs and the meaning of every output format. To summarize, we designed a user-friendly web interface to make it convenient to use.

Supplemental References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Structural Biology *66*, 213-221.

Afinogenov, I.I. (1997). History of the pharmacy of the Main Hospital. Voenno-meditsinskii zhurnal *318*, 57-60.

Aronov, A.M., Tang, Q., Martinez-Botella, G., Bemis, G.W., Cao, J.R., Chen, G.J., Ewing, N.P., Ford, P.J., Germann, U.A., Green, J.*, et al.* (2009). Structure-Guided Design of Potent and Selective Pyrimidylpyrrole Inhibitors of Extracellular Signal-Regulated Kinase (ERK) Using Conformational Control. J. Med. Chem. *52*, 6362-6368.

Carbain, B., Paterson, D.J., Anscombe, E., Campbell, A.J., Cano, C., Echalier, A., Endicott, J.A., Golding, B.T., Haggerty, K., Hardcastle, I.R.*, et al.* (2014). 8-Substituted O6-Cyclohexylmethylguanine

CDK2 Inhibitors: Using Structure-Based Inhibitor Design to Optimize an Alternative Binding Mode. J. Med. Chem. *57*, 56-70.

Carlsson, J., Boukharta, L., and Åqvist, J. (2008). Combining Docking, Molecular Dynamics and the Linear Interaction Energy Method to Predict Binding Modes and Affinities for Non-nucleoside Inhibitors to HIV-1 Reverse Transcriptase. J. Med. Chem. *51*, 2648-2656.

Case, D.A., and Kollman, P.A. (2016). AMBER 2016.

Charrier, J.D., Miller, A., Kay, D.P., Brenchley, G., Twin, H.C., Collier, P.N., Ramaya, S., Keily, S.B., Durrant, S.J., Knegtel, R.M.A.*, et al.* (2011). Discovery and Structure−Activity Relationship of 3-Aminopyrid-2-ones as Potent and Selective Interleukin-2 Inducible T-Cell Kinase (Itk) Inhibitors. J. Med. Chem. *54*, 2341-2350.

Chéron, N., Jasty, N., and Shakhnovich, E.I. (2016). OpenGrowth: An Automated and Rational Algorithm for Finding New Protein Ligands. J. Med. Chem. *59*, 4171-4188.

Congreve, M., Andrews, S.P., Doré, A.S., Hollenstein, K., Hurrell, E., Langmead, C.J., Mason, J.S., Ng, I.W., Tehan, B., Zhukov, A.*, et al.* (2012). Discovery of 1,2,4-Triazine Derivatives as Adenosine A2A Antagonists using Structure Based Drug Design. J. Med. Chem. *55*, 1898-1903.

Crawford, T.D., Ndubaku, C.O., Chen, H., Boggs, J.W., Bravo, B.J., DeLaTorre, K., Giannetti, A.M., Gould, S.E., Harris, S.F., Magnuson, S.R.*, et al.* (2014). Discovery of Selective 4-Amino-pyridopyrimidine Inhibitors of MAP4K4 Using Fragment-Based Lead Identification and Optimization. J. Med. Chem. *57*, 3484-3493.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. J. Chem. Phys. *98*, 10089-10092.

Demont, E.H., Chung, C., Furze, R.C., Grandi, P., Michon, A.-M., Wellaway, C., Barrett, N., Bridges, A.M., Craggs, P.D., Diallo, H.*, et al.* (2015). Fragment-Based Discovery of Low-Micromolar ATAD2 Bromodomain Inhibitors. J. Med. Chem. *58*, 5649-5673.

Denny, R.A., Flick, A.C., Coe, J., Langille, J., Basak, A., Liu, S., Stock, I., Sahasrabudhe, P., Bonin, P., Hay, D.A.*, et al.* (2017). Structure-Based Design of Highly Selective Inhibitors of the CREB Binding Protein Bromodomain. J. Med. Chem. *60*, 5349-5363.

Doree Sitkoff, Kim A. Sharp, and Honig, B. (1994). Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. J. Phys. Chem. *98*, 1978-1988.

Durrant, J.D., Amaro, R.E., and McCammon, J.A. (2009). AutoGrow: a novel algorithm for protein inhibitor design. Chem. Biol. Drug. Des. *73*, 168-178.

Efremov, I.V., Vajdos, F.F., Borzilleri, K.A., Capetta, S., Chen, H., Dorff, P.H., Dutra, J.K., Goldstein, S.W., Mansour, M., McColl, A.*, et al.* (2012). Discovery and Optimization of a Novel Spiropyrrolidine Inhibitor of β-Secretase (BACE1) through Fragment-Based Drug Design. J. Med. Chem. *55*, 9069-9088.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D *66*, 486-501.

Enyedy, I.J., and Egan, W.J. (2008). Can we use docking and scoring for hit-to-lead optimization? J. Comput. Aid. Mol. Des. *22*, 161-168.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Hsing, L., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. J. Chem. Phys. *103*, 8577-8593.

Green, J., Cao, J., Bandarage, U.K., Gao, H., Court, J., Marhefka, C., Jacobs, M., Taslimi, P., Newsome, D., Nakayama, T.*, et al.* (2015). Design, Synthesis, and Structure–Activity Relationships of Pyridine-Based Rho Kinase (ROCK) Inhibitors. J. Med. Chem. *58*, 5028-5037.

Guimarães, C.R.W., and Cardozo, M. (2008). MM-GB/SA Rescoring of Docking Poses in Structure-Based Lead Optimization. J. Chem. Inf. Model. *48*, 958-970.

Hanson, R.M. (2016). Jmol SMILES and Jmol SMARTS: specifications and applications. J. Cheminformatics *8*, 50.

Ho, W., Kukla, M.J., Breslin, H.J., Ludovici, D.W., Grous, P.P., Diamond, C.J., Miranda, M., Rodgers, J.D., Ho, C.Y., Declercq, E., *et al.* (1995). Synthesis and anti-HIV-1 activity of 4,5,6,7-tetrahydro-5-methylimidazo 4,5,1-jk 1,4 benzodazepin-2(1H)-one (TIBO) derivatives .4. J. Med. Chem. *38*, 794-802.

Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. J. Chem. Inf. Model. *51*, 69-82.

Hsu, H.H., Hsu, Y.C., Chang, L.J., and Yang, J.M. (2017). An integrated approach with new strategies for QSAR models and lead optimization. BMC genomics *18*, 104-112.

Huang, W., Zhao, X.E., Cai, J.F., Wang, J., Cong, X., and Ye, J. (2013). Fused heterocyclic derivative and application thereof. Patent WO2013097753A1,.

Jakalian, A., Bush, B.L., Jack, D.B., and Bayly, C.I. (2000). Fast, efficient generation of high-quality atomic charges. AM1-BCC model. I. Method. J. Comput. Chem. *21*, 132-146.

Jakalian, A., Jack, D.B., and Bayly, C.I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J. Comput. Chem. *23*, 1623-1641.

Kabsch, W. (2010). XDS. Acta Crystallogr. D *66*, 125-132.

Lawhorn, B.G., Philp, J., Graves, A.P., Holt, D.A., Gatto, G.J., and Kallander, L.S. (2016). Substituent Effects on Drug–Receptor H-bond Interactions: Correlations Useful for the Design of Kinase Inhibitors. J. Med. Chem. *59*, 10629-10641.

Liu, Q., Huang, F., Yuan, X., Wang, K., Zou, Y., Shen, J., and Xu, Y. (2017). Structure-Guided Discovery of Novel, Potent, and Orally Bioavailable Inhibitors of Lipoprotein-Associated Phospholipase A2. J. Med. Chem. *60*, 10231-10244.

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. *11*, 3696-3713.

Medina, J.R., Becker, C.J., Blackledge, C.W., Duquenne, C., Feng, Y.H., Grant, S.W., Heerding, D., Li, W.H., Miller, W.H., Romeril, S.P., *et al.* (2011). Structure-Based Design of Potent and Selective 3-Phosphoinositide-Dependent Kinase-1 (PDK1) Inhibitors. J. Med. Chem. *54*, 1871-1895.

Mirguet, O., Gosmini, R., Toum, J., Clement, C.A., Barnathan, M., Brusq, J.-M., Mordaunt, J.E., Grimes, R.M., Crowe, M., Pineau, O., *et al.* (2013). Discovery of Epigenetic Regulator I-BET762: Lead Optimization to Afford a Clinical Candidate Inhibitor of the BET Bromodomains. J. Med. Chem. *56*, 7501-7515.

Montero-Torres, A., García-Sánchez, R.N., Marrero-Ponce, Y., Machado-Tugores, Y., Nogal-Ruiz, J.J., Martínez-Fernández, A.R., Arán, V.J., Ochoa, C., Meneses-Marcel, A., and Torrens, F. (2006). Non-stochastic quadratic fingerprints and LDA-based QSAR models in hit and lead generation through virtual screening: theoretical and experimental assessment of a promising method for the discovery of new antimalarial compounds. Eur. J. Med. Chem. *41*, 483-493.

Moro, S., Bacilieri, M., Cacciari, B., Bolcato, C., Cusan, C., Pastorin, G., Klotz, K.-N., and Spalluto, G. (2006). The application of a 3D-QSAR (autoMEP/PLS) approach as an efficient pharmacodynamic-driven filtering method for small-sized virtual library: Application to a lead

optimization of a human A3 adenosine receptor antagonist. Bioorgan. Med. Chem. *14*, 4923-4932.

Narumi, T., Arai, H., Yoshimura, K., Harada, S., Hirota, Y., Ohashi, N., Hashimoto, C., Nomura, W., Matsushita, S., and Tamamura, H. (2013). CD4 mimics as HIV entry inhibitors: lead optimization studies of the aromatic substituents. Bioorgan. Med. Chem. *21*, 2518-2526.

Pan, Y.M., Gao, D.Q., and Zhan, C.G. (2008). Modeling the catalysis of anti-cocaine catalytic antibody: Competing reaction pathways and free energy barriers. J. Am. Chem. Soc. *130*, 5140-5149.

Pillai, A.D., Rani, S., Rathod, P.D., Xavier, F.P., Vasu, K.K., Padh, H., and Sudarsanam, V. (2005). QSAR studies on some thiophene analogs as anti-inflammatory agents: enhancement of activity by electronic parameters and its utilization for chemical lead optimization. Bioorgan. Med. Chem. *13*, 1275-1283.

Price, D.J., and Brooks, C.L., 3rd (2004). A modified TIP3P water potential for simulation with Ewald summation. J. Chem. Phys. *121*, 10096-10103.

Rose, P.W., Prlic, A., Altunkaya, A., Bi, C., Bradley, A.R., Christie, C.H., Di Costanzo, L., Duarte, J.M., Dutta, S., Feng, Z*., et al.* (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic. Acids. Res. *45*, D271-D281.

Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. *23*, 327-341.

Seeliger, M.A., Young, M., Henderson, M.N., Pellicena, P., King, D.S., Falick, A.M., and Kuriyan, J. (2005). High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. Protein Sci. *14*, 3135-3139.

Smith, R.H., Jr., Jorgensen, W.L., Tirado-Rives, J., Lamb, M.L., Janssen, P.A., Michejda, C.J., and Kroeger Smith, M.B. (1998). Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method. J. Med. Chem. *41*, 5272-5286.

Tasler, S., Müller, O., Wieber, T., Herz, T., Pegoraro, S., Saeb, W., Lang, M., Krauss, R., Totzke, F., Zirrgiebel, U*., et al.* (2009). Substituted 2-arylbenzothiazoles as kinase inhibitors: Hit-to-lead optimization. Bioorgan. Med. Chem. *17*, 6728-6737.

Tsukada, T., Takahashi, M., Takemoto, T., Kanno, O., Yamane, T., Kawamura, S., and Nishi, T. (2010). Structure-based drug design of tricyclic 8H-indeno 1,2-d 1,3 thiazoles as potent FBPase inhibitors. Bioorg. Med. Chem. Lett. *20*, 1004-1007.

Vyas, V.K., Shah, S., and Ghate, M. (2017). Generation of new leads as HIV-1 integrase inhibitors: 3D QSAR, docking and molecular dynamics simulation. Med. Chem. Res. *26*, 532-550.

Wang, B., and Merz, K.M. (2006). A fast QM/MM (Quantum Mechanical/Molecular Mechanical) approach to calculate nuclear magnetic resonance chemical shifts for macromolecules. J. Chem. Theory Comput. *2*, 209-215.

Wang, F., Jeon, K.O., Salovich, J.M., Macdonald, J.D., Alvarado, J., Gogliotti, R.D., Phan, J., Olejniczak, E.T., Sun, Q., Wang, S*., et al.* (2018). Discovery of Potent 2-Aryl-6,7-dihydro-5H-pyrrolo[1,2-a]imidazoles as WDR5-WIN-Site Inhibitors Using Fragment-Based Methods and Structure-Based Design. J. Med. Chem. *61*, 5623-5642.

Wang, J., Morin, P., Wang, W., and Kollman, P.A. (2001). Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J. Am. Chem. Soc. *123*, 5221-5230.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing

of a general amber force field. J. Comput. Chem. *25*, 1157-1174.

Wang, W.R., Marimuthu, A., Tsai, J., Kumar, A., Krupka, H.I., Zhang, C., Powell, B., Suzuki, Y., Nguyen, H., Tabrizizad, M*., et al.* (2006). Structural characterization of autoinhibited c-Met kinase produced by coexpression in bacteria with phosphatase. P. Natl. Acad. Sci. U. S. A. *103*, 3563-3568.

Woodhead, A.J., Angove, H., Carr, M.G., Chessari, G., Congreve, M., Coyle, J.E., Cosme, J., Graham, B., Day, P.J., Downham, R*., et al.* (2010). Discovery of (2,4-Dihydroxy-5-isopropylphenyl)-5-(4-methylpiperazin-1-ylmethyl)-1,3- dihydroisoindol-2-yl methanone (AT13387), a Novel Inhibitor of the Molecular Chaperone Hsp90 by Fragment Based Drug Design. J. Med. Chem. *53*, 5956-5969.

Yang, J.F., Wang, F., Jiang, W., Zhou, G.Y., Li, C.Z., Zhu, X.L., Hao, G.F., and Yang, G.F. (2018). PADFrag: A Database Built for the Exploration of Bioactive Fragment Space for Drug Discovery. J. Chem. Inf. Model. *58*, 1725-1730.

Zhou, R., Friesner, R.A., Ghosh, A., Rizzo, R.C., Jorgensen, W.L., and Levy, R.M. (2001). New Linear Interaction Method for Binding Affinity Calculations Using a Continuum Solvent Model. J. Phys. Chem. B *105*, 10388-10397.