J Mol Model (2012) 18:3087–3100
DOI 10.1007/s00894-011-1328-5

ORIGINAL PAPER

Pharmacophore modeling and virtual screening studies to identify new c-Met inhibitors

Wenting Tai · Tao Lu · Haoliang Yuan · Fengxiao Wang · Haichun Liu · Shuai Lu · Ying Leng · Weiwei Zhang · Yulei Jiang · Yadong Chen

Received: 31 October 2011 / Accepted: 5 December 2011 / Published online: 28 December 2011

Abstract Mesenchymal epithelial transition factor (c-Met) is an attractive target for cancer therapy. Three-dimensional pharmacophore hypotheses were built based on a set of known structurally diverse c-Met inhibitors. The best pharmacophore model, which identified inhibitors with an associated correlation coefficient of 0.983 between their experimental and estimated IC_{50} values, consisted of two hydrogen-bond acceptors, one hydrophobic, and one ring aromatic feature. The highly predictive power of the model was rigorously validated by test set prediction and Fischer's randomization method. The high values of enrichment factor and receiver operating characteristic (ROC) score indicated the model performed fairly well at distinguishing active from inactive compounds. The model was then applied to screen compound database for potential c-Met inhibitors. A filtering protocol, including druggability and molecular docking, were also applied in hits selection. The final 38 molecules, which exhibited good estimated activities, desired binding mode and favorable drug likeness were identified as potential c-Met inhibitors. Their novel backbone structures could be served as scaffolds for further study, which may facilitate the discovery and rational design of potent c-Met kinase inhibitors.

Keywords Cancer · c-Met inhibitors · Molecular docking · Pharmacophore modeling · Virtual screening

Introduction

Receptor tyrosine kinases (RTKs) play a key role in the regulation of many intracellular signal transduction pathways and they can coordinate a range of downstream cellular processes. Overexpression, dysregulation or inappropriate activation of RTKs can affect cell survival, proliferation, motility and lead to the development, maintenance and progression of human cancers. Therefore, members of the RTK family are attractive targets for cancer therapy as inhibition can disrupt signaling pathways that mediate tumor formation and growth. c-Met kinase is a member of this family that, together with its ligand, hepatocyte growth factor (HGF), is important for normal mammalian development [1, 2].

c-Met and HGF are widely expressed in a variety of tissues, and they are each required for normal mammalian development and have been shown to be particularly important in cell proliferation, migration, differentiation, and organization of three-dimensional tubular structures as well as cell growth and angiogenesis. However, in a number of major human cancers, both c-Met and HGF are shown to be deregulated and correlation with poor prognosis [3]. The overexpression of c-Met was found in most cancers, including brain, colorectal, gastric, lung, head and neck, stomach cancers and other carcinomas [4].

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1328-5) contains supplementary material, which is available to authorized users.

W. Tai · H. Yuan · F. Wang · H. Liu · S. Lu · Y. Leng · W. Zhang · Y. Jiang (✉) · Y. Chen (✉)
Laboratory of Molecular Design and Drug Discovery,
School of Basic Science, China Pharmaceutical University,
24 Tongjiaxiang,
Nanjing 210009, China
e-mail: jiangyulei@cpu.edu.cn
e-mail: ydchen@cpu.edu.cn

T. Lu · Y. Chen
State Key Laboratory of Natural Medicines,
China Pharmaceutical University,
24 Tongjiaxiang,
Nanjing 210009, China

The binding of HGF to c-Met leads to the activation of HGF/SF-c-Met signal transduction pathway. It induces receptor dimerisation and autophosphorylation of multiple tyrosine residues within the activation loop of the tyrosine kinase domain [5]. Activation of c-Met results in the binding and phosphorylation of adaptor proteins such as Gab-1, Grb-2, Shc, and c-Cbl, subsequent activation of signal molecules such as PI-3 K, PLC-γ, STAT, ERK1/2, FAK and their corresponding signal pathways. Besides, c-Met signaling can also interact with focal adhesion complexes and non-kinase binding partners such as β4 integrins, CD44, and semaphorins, which may further add to the complexity of regulation of cell function by this receptor [3].

Since HGF/SF-c-Met pathway plays such an important role in tumor progression, it is significant to discover novel c-Met inhibitors. Small-molecular inhibitors of c-Met compete for the adenosine triphosphate (ATP) binding site in the tyrosine kinase domain and prevent receptor transactivation and recruitment of the downstream effectors. The c-Met inhibitors are divided into two types according to their chemotypes and binding modes. Type I inhibitors (SU-11274-like) bind the c-Met kinase domain in the activated DFG-in conformation. They are characterized by a diverse range of structures, which are reasonably selective, have a compact, U-shaped conformation and show resistance to Tyr1230 mutation. In contrast to Type I inhibitors, those that bind in the inactivated DFG-out conformation are classified as Type II ligands (AM7-like). They are typically less selective and adopt an extended conformation that stretches from the kinase linker to the C-helix pocket. Currently, numerous compounds have been reported as c-Met inhibitors, but only several c-Met inhibitors have entered the clinical trials, such as PF-04217903, JNJ-38877605, AMG-208, MK-2461 and PF-02341066 (Fig. 1) [6–9]. Despite much effort in the past years, there is no inhibitor targeting c-Met approved to be used clinically so far. Therefore, the discovery of more potent and novel scaffolds against c-Met kinase is particularly important.

The pharmacophore model, which combines and visualizes critical features responsible for the inhibitory activity, can be not only applied to undertake database screening to retrieve compounds with inhibitory activity against target protein, but also provide guidance for the rational design to discover novel inhibitors. To date, there are few papers on pharmacophore studies of c-Met inhibitors [10, 11]. Here, we reported a reliable pharmacophore model based on a series of known c-Met inhibitors. In validation studies, it was demonstrated that the best pharmacophore model with strong predictive power achieved a high performance of identifying the active from inactive compounds. Then, a database screening with the validated model was performed to discovery novel scaffolds which can provide good starting points for the design of novel c-Met kinase inhibitors.

Materials and methods

Collection of dataset

For generating hypotheses, training set molecules ought to satisfy a certain set of laws like it must be broadly abided by structurally diverse representatives (minimum 16 compounds) and wrap an activity range of at least four magnitudes. All the biologically relevant data must be obtained by homogeneous processes [12]. For this study, we chose compounds from Type I inhibitors with a significantly structural diversity as training and test sets. At last, 23 structurally diverse compounds with activity values (IC_{50}) between 3 nM and 28,300 nM were selected as training set (Fig. 2) [3, 6, 13–19], which

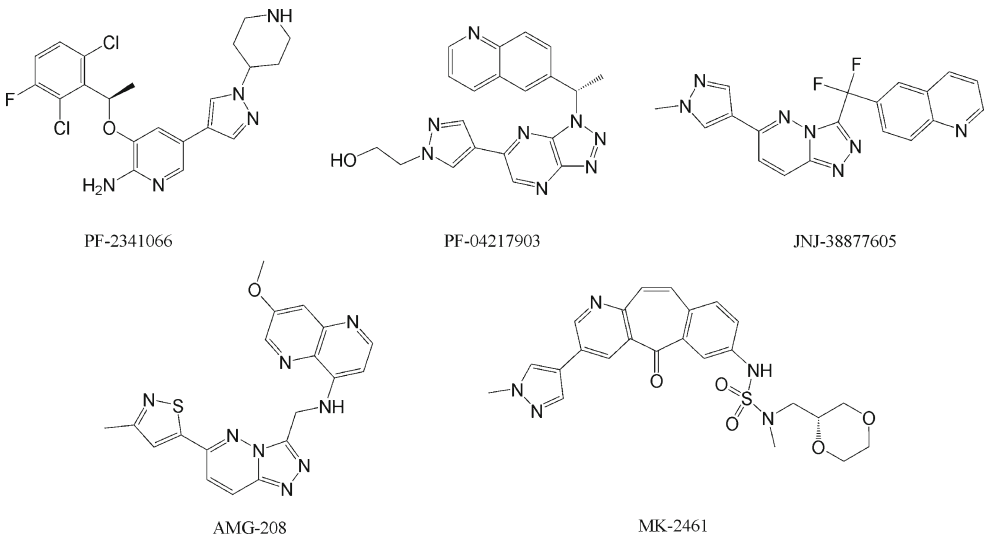

Fig. 1 Chemical structures of c-Met inhibitors in clinical trials

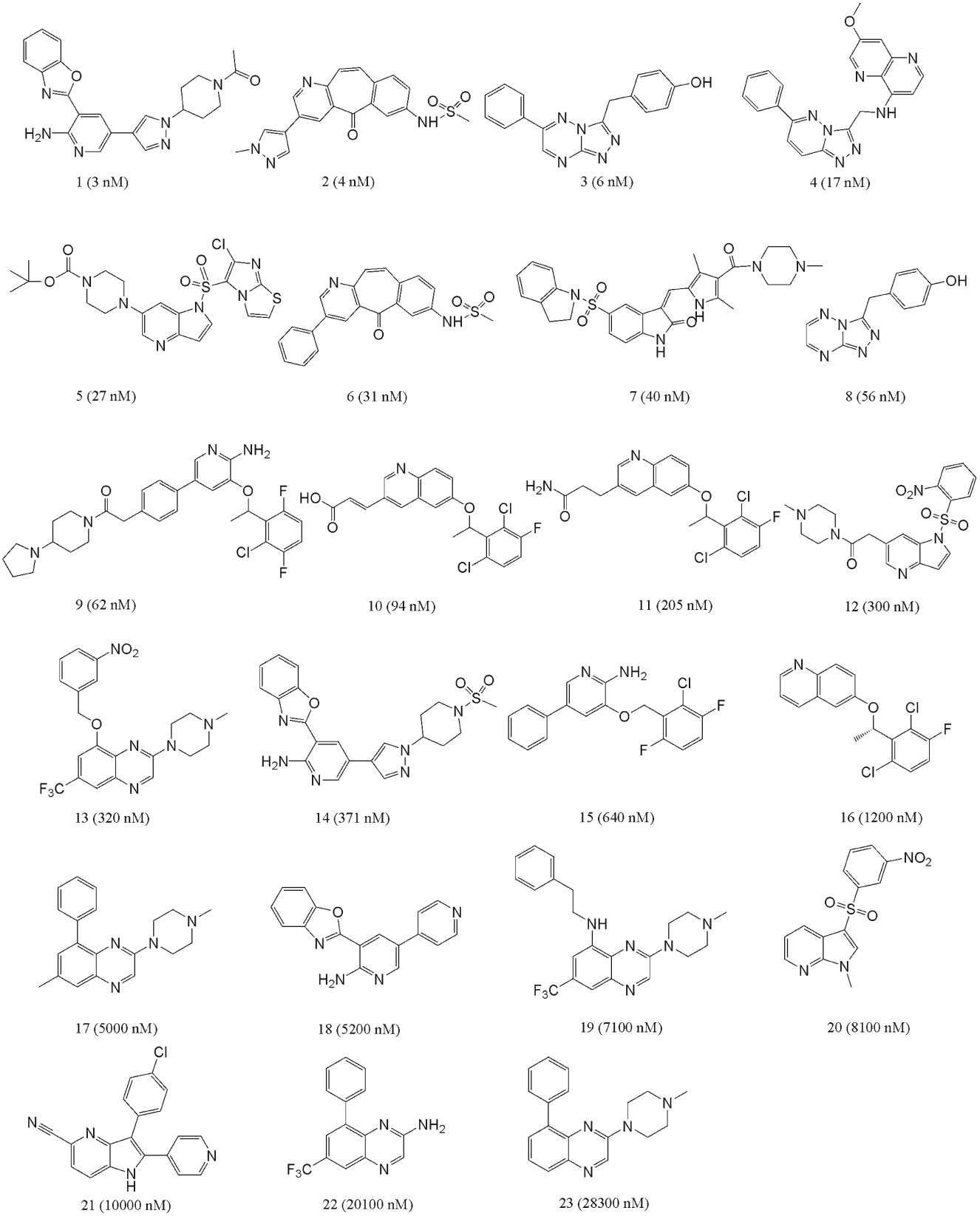

Fig. 2 Structures and IC_{50} values of all 23 training set molecules

spanned over four magnitudes. There are ten diverse scaffolds in the training set, including azaindoles, benzyloxyquinolines, quinoxalines, pyridins and other relevant structures. To validate the hypothesis, the test set was prepared using the same protocol as training set. Test set (Fig. 3) contained 31 structurally diverse compounds from the training set with a wide range of activity values. All the compounds in the training and test set were categorized into four different sets based on their activity values as high active (IC_{50} <100 nM, ++++), moderate active (100≤IC_{50} <1000 nM, +++), less active (1000≤IC_{50}< 10,000 nM, ++) and inactive (IC_{50} ≥10,000 nM, +) compounds.

Conformational analysis

All training and test set compounds were firstly minimized by Pipeline Pilot V7.5. The conformational space of each compound was extensively sampled utilizing the poling algorithm within the generation conformation module of Discovery Studio (DS) 2.5. Conformations were generated by "best conformational" option, with an energy threshold of 20 kcal mol^{-1} from the locally minimized structure and a maximum limit of 255 conformers per molecule. This step probably identifies the best spatial arrangement of chemical groups explaining the activity variations among the training set [20]. Compounds with their conformational models were then submitted to DS for generating pharmacophoric hypotheses.

Generation of pharmacophore models

Prior to quantitative pharmacophore development, we conducted a common-feature (HipHop) pharmacophore modeling study in order to identify the features necessary for potent c-Met inhibitors. Five c-Met inhibitors (PF-02341066, PF-04217903, JNJ-38877605, AMG-208, MK-2461) which have entered clinical research were used to generate HipHop pharmacophore. Every compound was treated equally by setting the principal value of 2 and the MaxOmitFeat value of 0. A five-feature pharmacophore model was constructed using this method, containing two ring aromatic (RA), two hydrophobic (HY), and one hydrogen-bond acceptor (HBA) features.

The 23 training set compounds were used to develop a HypoGen pharmacophore model. Based on the HipHop pharmacophore features, we selected HBA, HY and RA as the essential features for hypotheses generation. In addition, the hydrogen-bond donor (HBD) feature was also included, which was based on the analyzing of the chemical features in the training set. For each feature the minimum number is 0 and the maximum is 5. Default uncertainty value 3 has been changed to 2 as the activity range in the training set compounds barely spans the minimum requirement of four orders of magnitude and also to effectively correlate the training set compounds with their activity values [21]. All other parameters were kept as default. Subsequently, pharmacophore hypotheses were constructed using 3D-QSAR pharmacophore generation module and the top ten scored hypotheses were collected.

The quality of HypoGen pharmacophore hypotheses is best described by fixed cost, null cost, total cost and other statistical parameters [22]. Fixed cost represents the simplest model that fits all data perfectly, while null cost presumes that there is no relationship in the data and that the experimental activities are normally distributed around their average value [23]. For an expected pharmacophore model, the total cost should be close to the fixed cost, and there should be a significant difference (>60) between null and total cost to show the over 90% statistical significance of the model. Also, the configuration cost which depends on the complexity or the entropy of the conformational space being optimized and is constant for a given data set should have a value less than 17 [24].

Pharmacophore model validation

The main purpose of validating a quantitative pharmacophore hypothesis is to determine whether it is capable of identifying active compounds and predicting their activities accurately [23]. The validated pharmacophore hypothesis can be used in database screening to identify novel c-Met inhibitors. In this study, the pharmacophore hypothesis were validated with three different methods, respectively test set prediction, Fisher's randomization test and enrichment factor (E) and receiver operating characteristic (ROC) curve. In the first validation, a test set containing 31 compounds that are similarly structurally diverse to the training set and with a wide range of activity values was utilized and the inhibitory activity values were estimated for every test set compound. The second validation procedure is a cross-validation based on Fischer randomization methodology with a goal to check whether there is a strong correlation between the chemical structures and the biological activity [25, 26]. This is done by randomizing the activity data associated with the training set compounds. Numbers of pharmacophore hypotheses were generated using the same parameters used to develop the original hypothesis. If any of the randomized pharmacophore hypotheses resulted with similar or better cost or correlation value than the original hypothesis, then the original hypothesis is considered to be generated by chance [21]. Finally, in order to determine the capability of the pharmacophore hypotheses to discriminate active compounds from other molecules in virtual screening [27], E value and other statistical parameters were calculated using a small database containing 168 known c-Met inhibitors and 5827 randomly sampled compounds. The randomly sampled set served as decoys, which were obtained from a collection offered by drugbank (subset of random FDA-approved small molecule drug structures without biological activities on c-Met reported) [28]. The ROC curve, which was introduced recently by Triballeau et al. [29], was also applied to evaluate

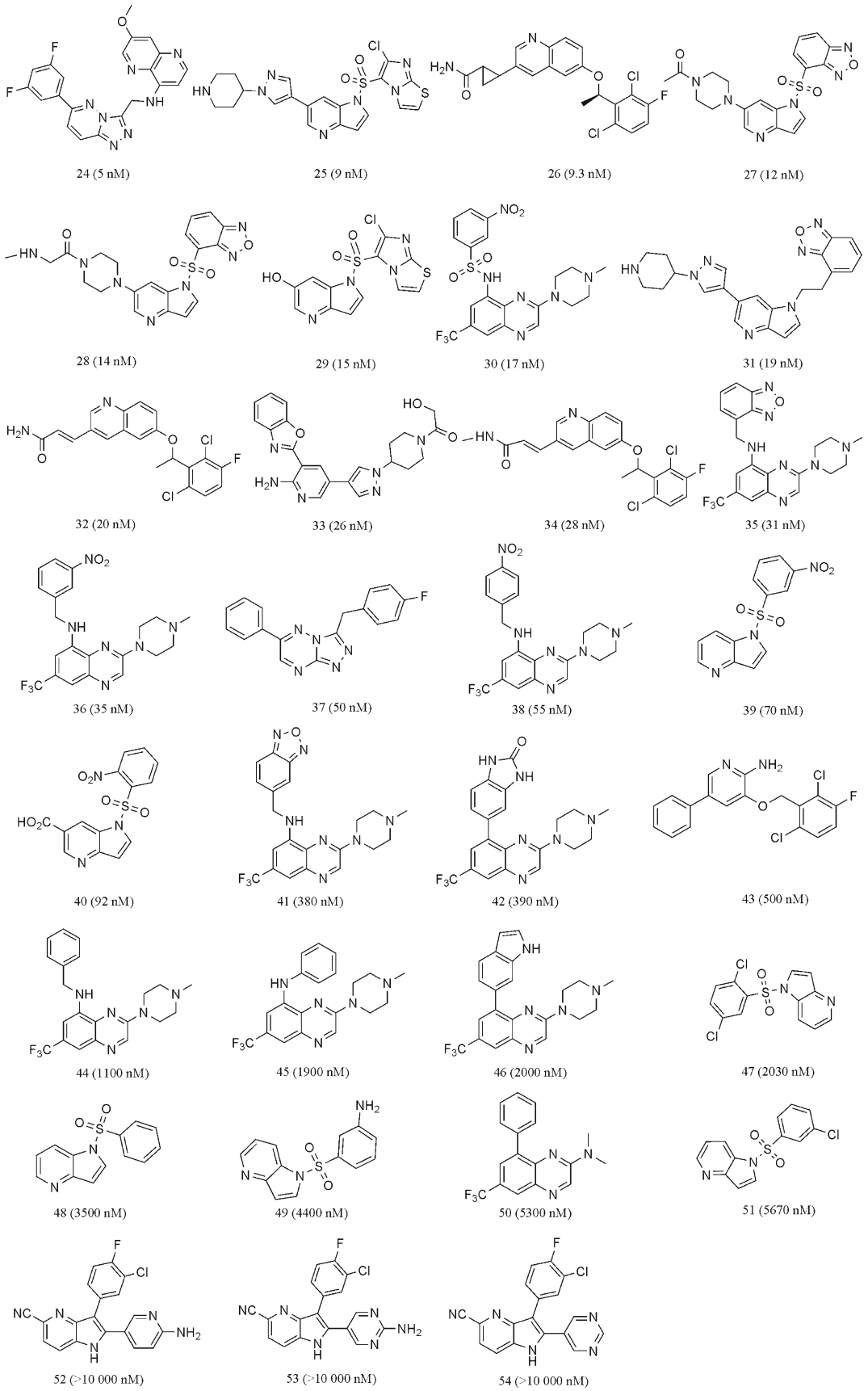

Fig. 3 Structures and IC_{50} values of all 31 test set molecules

Table 1 Results of ten top_scored pharmacophore hypotheses generated by HypoGen

Hypo	Total cost	Cost difference[a]	RMSD	Correlation	Features
1	94.962	154.2522	0.756	0.983	HBA HBA HY RA
2	114.714	134.5	1.617	0.914	HBA HBA HY RA
3	120.135	129.079	1.503	0.930	HBA HBA HY
4	123.951	125.263	1.758	0.899	HBA HBA HY
5	124.792	124.422	1.888	0.881	HBA HBA HY RA
6	125.371	123.843	1.881	0.882	HBA HBA HY RA
7	126.576	122.638	1.914	0.877	HBA HBA HY RA
8	128.108	121.106	1.978	0.868	HBA HBA HY RA
9	128.870	120.344	1.995	0.866	HBA HBA HY RA
10	130.022	119.192	1.988	0.867	HBA HBA HY RA

null cost=249.214; fixed cost=82.990; configuration cost = 15.095.

[a]Cost difference=null cost−total cost

the performance of the pharmacophore hypotheses. We subjected the pharmacophore hypotheses to ROC analysis to assess their abilities to correctly classify a list of compounds as actives or inactives [30]. In this case, the validity of a particular pharmacophore is indicated by the area under the curve (AUC) of the corresponding ROC curve.

Database screening

Database screening serves the purpose of finding diverse, potential virtual compounds suitable for further optimization and provides the preference of having easily available and/or synthesizable compounds as hits for further steps in drug development [31]. The validated pharmacophore model was used as a query for retrieving potential inhibitors from NCI compound database (260,071 compounds). Lipinski's rule of five ((1) not more than five hydrogen bond donors; (2) not more than ten hydrogen bond acceptors; (3) a molecular weight under 500 Da; and (4) a partition coefficient log P less than 5.), was firstly applied to exclude none drug-like compounds. Then, the best pharmacophore model was used to screen the qualified compounds with best/flexible searching option in DS 2.5. The compounds that satisfied all the features were retrieved as hits and considered in molecular docking studies.

Molecular docking

Molecular docking is a computational technique that generates and scores putative protein–ligand complexes according to their calculated binding affinities. It has been successfully used for identifying active compounds by filtering out those that do not fit into the binding sites [32–34]. In this study, a high resolution (2.20 Å) co-crystal complex structure of c-Met (PDB code: 2RFS) was selected for molecular docking.

In order to ensure which program was suitable for docking c-Met inhibitors accurately, the co-crystal ligand was redocked with three different programs GOLD, CDOCKER and Glide. The RMSD values between the docked and crystal conformations were correspondingly 1.31, 0.24 and 0.20, which revealed that Glide performed best in reproducing experimental binding conformation of the ligand. Therefore, Glide was selected for the molecular docking study. The crystal complex structure was prepared with the protein preparation wizard workflow. The receptor-grid file was generated with an enclosing box, which was defined as the ligand-binding site by centering on the co-crystal ligand and similar in size. Other parameters were kept as default. All compounds were prepared with the LigPre module and then flexibly docked into the binding site with the extra precision (XP) docking mode selected [11].

Results and discussion

Pharmacophore generation

Ten pharmacophore hypotheses were generated. The cost values, correlation coefficients, RMSD, and pharmacophore

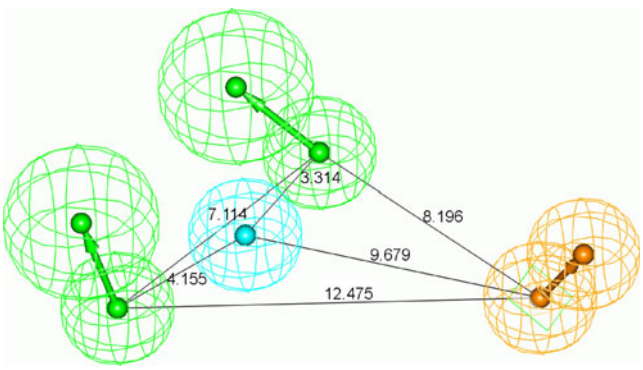

Fig. 4 The best pharmacophore model Hypo1 with distance constraints. Features are color-coded as follows: HBA, green; HY, cyan; and RA, orange

Table 2 Experimental IC_{50} vs. estimated IC_{50} values of training set compounds based on Hypo1

Compounds	IC_{50}(nM)		Error[a]	Activity scale[b]	
	Experimental	Estimated		Experimental	Estimated
1	3	2.41	−1.2	++++	++++
2	4	7.82	2.0	++++	++++
3	6	8.56	1.4	++++	++++
4	17	27.96	1.6	++++	++++
5	27	51.72	1.9	++++	++++
6	31	35.33	1.1	++++	++++
7	40	52.44	1.3	++++	++++
8	56	49.31	−1.1	++++	++++
9	62	30.03	−2.1	++++	++++
10	94	110.48	1.2	++++	+++
11	205	280.36	1.4	+++	+++
12	300	348.24	1.2	+++	+++
13	320	174.43	−1.8	+++	+++
14	371	351.59	−1.1	+++	+++
15	640	559.87	−1.1	+++	+++
16	1200	2303.48	1.9	++	++
17	5000	11,402.20	2.3	++	+
18	5200	5624.96	1.1	++	++
19	7100	2147.51	−3.3	++	++
20	8100	6531.67	−1.2	++	++
21	10,000	9943.28	−1.0	+	++
22	20,100	11,624.50	−1.7	+	+
23	28,300	10,679.10	−2.7	+	+

[a] Positive value indicates that the estimated activity is higher than experimental activity and negative value indicates that the estimated activity is lower than experimental activity.

[b] Activity scale: high active, ++++ (IC50<100 nM); moderate active, +++ (100≤ IC50<1000 nM); less active, ++ (1000≤IC50<10,000 nM); inactive , + (IC50≥10,000 nM).

features were listed in Table 1. The top ranked hypothesis (Hypo1), consisting of spatial arrangement of four chemical features (Fig. 4), including two HBA, one HY, and one RA features, was identified as the best model. It had the lowest total cost (94.962), the least difference between total and fixed cost (11.972), the highest cost difference between null cost and total cost (154.252), the least RMSD (0.756), and a strong correlation coefficient (0.983) between experimental and estimated activities. And the configuration cost of Hypo1 was 15.095, which did not exceed the maximum limit value of

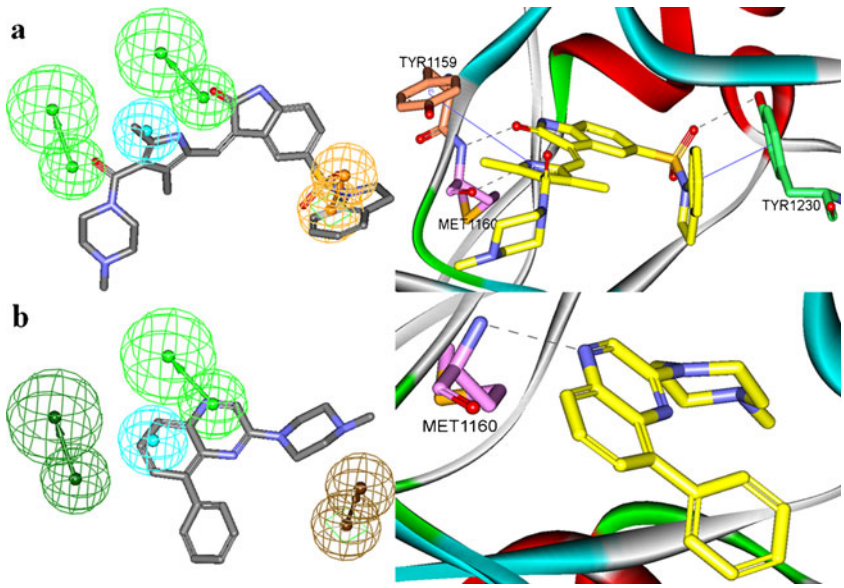

Fig. 5 Comparision of pharmacophore mapping with the docked binding models of two representative compounds. (**a**) compound 7, (**b**) compound 23

17 and could guarantee the entire conformation space sampled during the pharmacophore generation.

One method of judging a pharmacophore's merit is its ability to predict the activities of individual compounds in the training set [23–25]. As is shown in Table 2, all the compounds were correctly estimated by Hypo1 with the error values less than 10, which means the differences between the experimental IC_{50} and predicted IC_{50} values are less than one order magnitude. At the same time, an analysis regarding mapping of the training set compounds on Hypo1 was performed. The analysis revealed that one of the most active compounds 7 in the training set mapped on all four pharmacophoric features of Hypo1. Two amide groups mapped well onto the two HBA features, respectively. The pyrrole ring mapped on the HY feature and the indoline group at the end of the molecule mapped on the RA feature (Fig. 5a); whereas the least active compound 23 mapped only on two of the four features, missed one HBA and the RA feature (Fig. 5b). The fit values of the most and least active compounds are 7.592 and 4.369, respectively. This might be used to explain why compound 7 was more active than compound 23. Binding modes of the two compounds were in good agreement with their pharmacophore mappings (Fig. 5).

Pharmacophore validation

Test set prediction

A test set of 31 c-Met inhibitors was used to validate the best pharmacophore model Hypo1. Experimental and estimated activities of the test set compounds are shown in Table 3, as we can see, most IC_{50} values of the compounds in the test set

Table 3 Experimental IC_{50} vs. estimated IC_{50} values of test set compounds based on Hypo1

Compounds	IC_{50}(nM)		Error[a]	Activity scale[b]	
	Experimental	Estimated		Experimental	Estimated
24	5	3.25	−1.5	++++	++++
25	9	8.36	−1.1	++++	++++
26	9.3	5.45	−1.7	++++	++++
27	12	7.76	−1.5	++++	++++
28	14	7.63	−1.8	++++	++++
29	15	20.61	1.4	++++	++++
30	17	18.28	1.1	++++	++++
31	19	28.21	1.5	++++	++++
32	20	21.37	1.1	++++	++++
33	26	10.02	−2.6	++++	++++
34	28	9.16	−3.1	++++	++++
35	31	52.31	1.7	++++	++++
36	35	20.17	−1.7	++++	++++
37	50	58.54	1.2	++++	++++
38	55	22.12	−2.5	++++	++++
39	70	74.96	1.1	++++	++++
40	82	39.89	−2.1	++++	++++
41	380	151.38	−2.5	+++	+++
42	390	337.28	−1.2	+++	+++
43	500	291.84	−1.7	+++	+++
44	1100	1179.27	1.1	++	++
45	1900	6185.68	3.3	++	++
46	2000	1951.00	−1.0	++	++
47	2030	5430.81	2.7	++	++
48	3500	6133.65	1.8	++	++
49	4400	5989.14	1.4	++	++
50	5300	2435.29	−2.2	++	++
51	5670	6188.64	1.1	++	++
52	>10,000	5522.95	−1.8	+	++
53	>10,000	24,228.40	2.4	+	+
54	>10,000	27,057.40	2.7	+	+

[a] Positive value indicates that the estimated activity is higher than experimental activity and negative value indicates that the estimated activity is lower than experimental activity.

[b] Activity scale: high active, ++++ (IC_{50}<100 nM); moderate active, +++ (100≤IC_{50}<1000 nM); less active, ++ (1000≤IC_{50}<10,000 nM); inactive , + (IC_{50}≥10,000 nM).

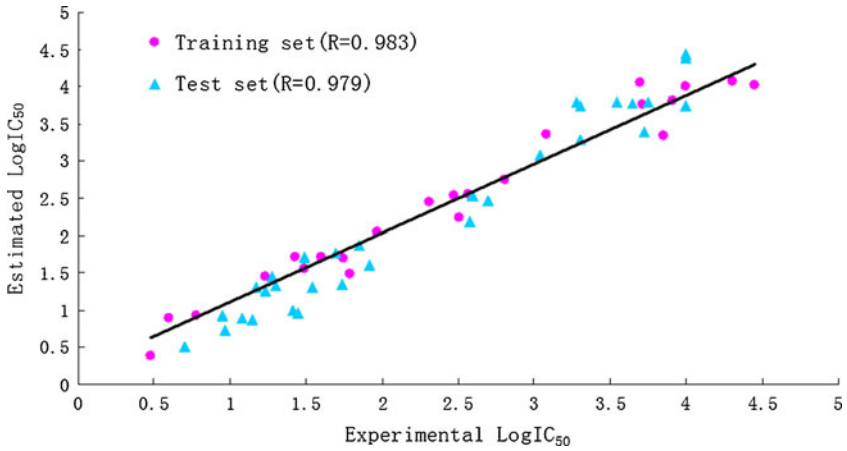

Fig. 6 The correlations between experimental and estimated activities of training and test set compounds based on Hypo1

were estimated properly, with the error values mainly around 2. Moreover, the regression analysis achieved a high correlation coefficient of 0.979 (Fig. 6), which suggested a good correlation between the experimental and estimated activities.

Fischer's randomization method

In order to further validate Hypo1, cross-validation based on Fischer's randomization was carried out. A confidence level of 95% was selected, and then 19 random spreadsheets were generated. Results of the validation are shown in Fig. 7, obviously, all values generated after randomization produced models with no predictive values similar or near to that of Hypo1. Compared to the corresponding 19 runs, the correlation coefficients of the original hypotheses were the highest while their total cost values were the lowest, which proved that Hypo1 was dependable and not generated by chance.

Enrichment factor and ROC curve

A database of 5995 compounds (D) including 168 known inhibitors (A) was used to further validate the best pharmacophore

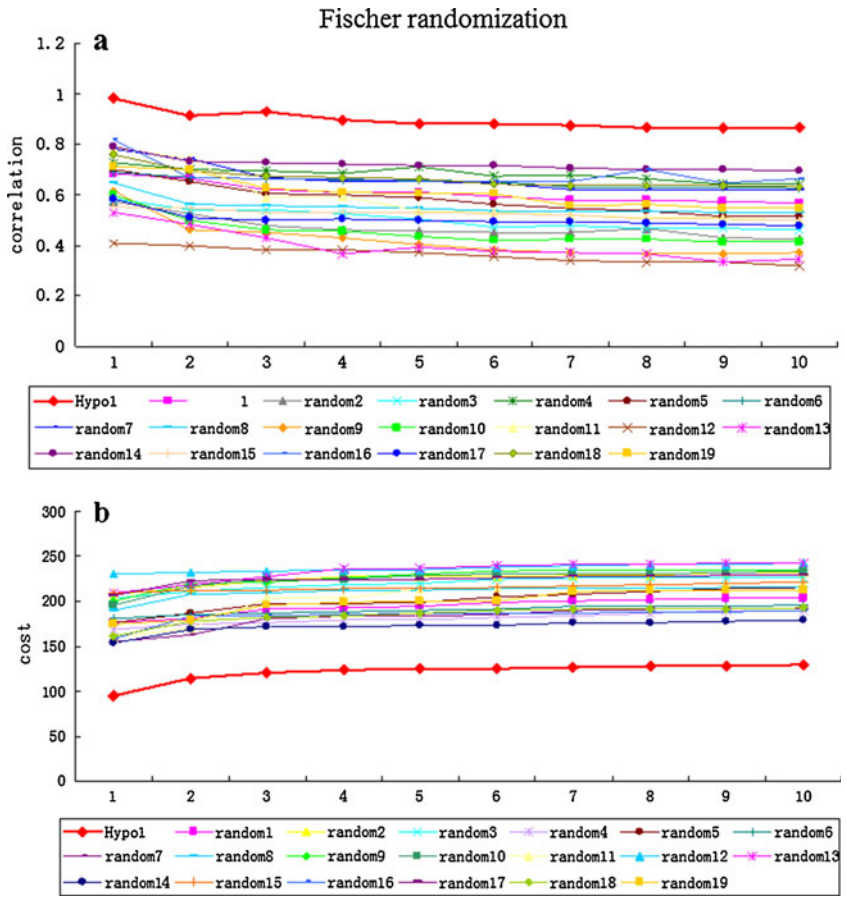

Fig. 7 Results of Fischer randomization test for 95% confidence level. (**a**) indicates Hypo1 has significantly higher correlation coefficient values than that of all 19 random hypotheses; (**b**) shows Hypo1 has significantly lower cost values than that of all 19 random hypotheses

Table 4 Enrichment factor and goodness of hit score validation for Hypo1

Parameter	Values
Total molecules in database (D)	5995
Total number of actives in database (A)	168
Total hits (Ht)	249
Active hits (Ha)	147
% Yield of actives [(Ha/Ht)×100]	59.04
% Ratio of actives [(Ha/A)×100]	87.50
Enrichment factor (E) [(Ha×D)/(Ht×A)]	21.07
False negatives [A−Ha]	21
False positives [Ht−Ha]	102
Goodness of hit score (GH)[a]	0.65

[a] [(Ha/4HtA) (3A+Ht))×(1−((Ht−Ha)/(D−A))]; GH score of >0.6 indicates a very good model.

in this study. Using the selected pharmacophore model, Hypo1, 249 compounds (Ht) were retrieved as hits from the database screening. Among these hits, 147 (Ha) compounds were from the 168 known inhibitors. Therefore, the enrichment factor was calculated to be 21.07, which means that it is 21.07 times more probable to pick active compounds from the database than expected by chance. The calculated goodness of hit score (GH) value was 0.65, greater than 0.5, which is significant for any pharmacophore hypothesis (Table 4).

The ROC curve was also used to estimate the performance of Hypo1. It can be used to help understand the tradeoff between model sensitivity (ability to discover true positives) and specificity (ability to avoid false positives) [28]. The fit property of the compounds, which indicates the degree of consistency with Hypo1, was calculated. Thus, the 5995 compounds in the validation database were ranked according to their fit values. The ROC score (the area under the ROC curve, AUC) provides a practical way to measure the overall performance of the model. The closer the ROC score is to 1.0, the better the model is at distinguishing good samples from bad ones. ROC curve analysis of Hypo1 yielded the ROC score of 0.896 (Fig. 8), which means that in eight out of ten cases, a randomly selected active c-Met inhibitor is ranked higher than an inactive one. These three validation procedures provided strong confidence on Hypo1. Based on these validation results, Hypo1 was capable enough to be used in database searching to identify novel leads.

Evaluation of the pharmacophore with the crystal complex

As the crystal structure of c-Met in complex with compound 34 was resolved (PDB code: 3A4P) and the protein–ligand interaction mode was revealed [16], it can be utilized to further evaluate the best pharmacophore Hypo1. As is shown in Fig. 9a, the scaffold anchored into the ATP pocket by two π-π stacking interactions and two crucial hydrogen bonds from hinge region which is located in the binding site of c-Met and usually hydrophobic: the quinoline accepted one H-bond from hinge Met1160 while also formed π-π stacking interaction with Tyr1159, the distal amide accepted one H-bond from hinge Lys1161, as well as one co-planar π-π stacking interaction of the halogenated benzene ring with Tyr1230.

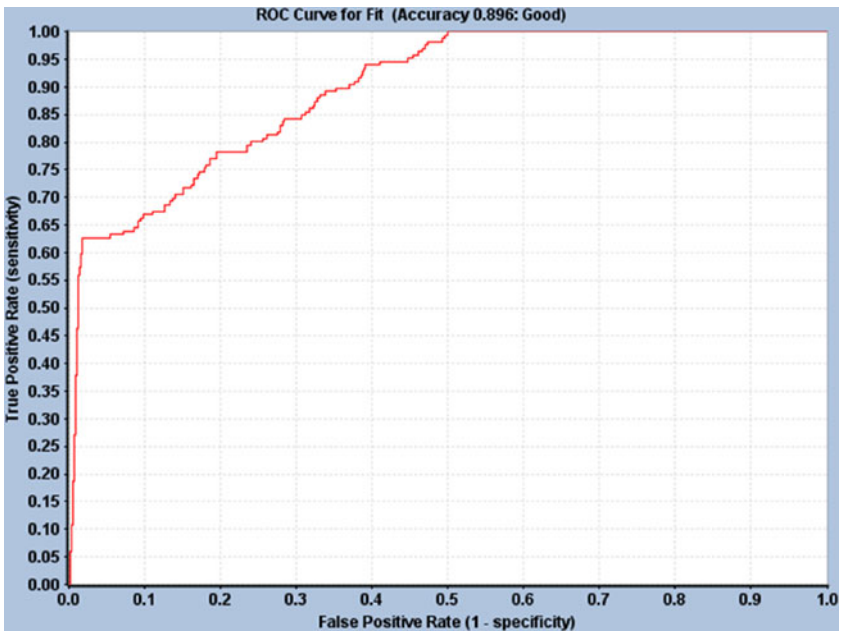

Fig. 8 ROC curve generated from screening the database

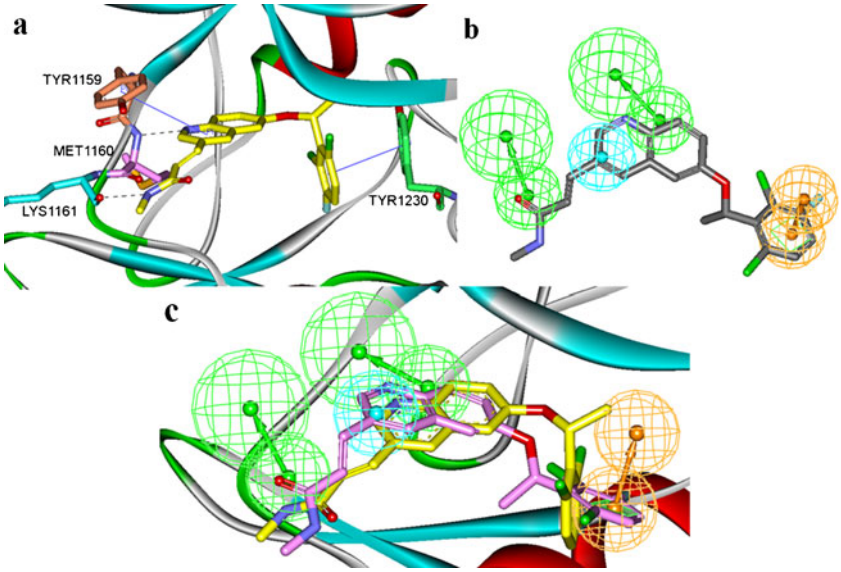

Fig. 9 (**a**) Binding mode of compound 34 in crystal complex (PDB code: 3A4P); (**b**) Pharmacophore mapping of compound 34; (**c**) Superimposed of the mapping conformation (pink) with Hypo1 and the original bound conformation (yellow)

When mapped onto Hypo1 by choosing best/flexible searching option, compound 34 from the crystal structure fit quite well with all the features of the pharmacophore (Fig. 9b), the fit value was as high as 8.698. Then the mapping of Hypo1 with compound 34 was superimposed to the active site of the crystal complex. The RMSD value between the mapping conformation of Hypo1 and the original bound conformation was 1.679, which was low and acceptable. Figure 9c clearly shows that the essential parts of two conformations could be generally superimposed, from which we can also see that the two hydrogen-bond acceptor (HBA) features were respectively mapped on the two hydrogen bonds accepted from hinge region, the hydrophobic (HY) features were mapped on the π-π stacking interaction and consistent with the character of hinge region, while the ring aromatic (RA) mapped on the benzene ring which could form π-π stacking interaction with Tyr1230. These analyses revealed that Hypo1 contained all the essential features the crystallographic studies covered, which were consistent with the interaction mechanism of c-Met inhibitors. And the results also suggested Hypo1 was reasonable and reliable enough to retrieve compounds that fit all the features of the model from a chemical database.

Database screening

The validated pharmacophore model, Hypo1, was used to do virtual screening over the NCI database comprised of 260,071 compounds. Based upon Lipinski's rule of 5, 203,965 drug-like compounds from NCI database were selected for virtual screening by the option of search 3D database protocol with the best search method. A hit list of 9256 compounds matching the pharmacophore model was obtained, which included some compounds structurally similar to the existing c-Met kinase inhibitors and novel scaffolds also emerged. Moreover, the IC_{50} value of each compound was predicted. To sample a sufficient chemical space and increase hit-rate, compounds with estimated IC_{50} below 1 μM were considered as active new hits. A set of 1623 hits satisfied the specified cutoff value was preceded for further evaluation.

Although the hits mapped all the features of Hypo1 and had high estimated active values, they might not fit into the

Table 5 Analyses of critical amino acid residues for c-Met inhibition from nine co-crystal structures deposited in protein data bank

PDB code	Met1160	Asp1222	Pro1158	Met1161	Tyr1230	Tyr1159
2RFS	√	√			√	
3CCN	√	√			√	
3CD8	√	√			√	√
2WD1	√	√			√	√
3A4P	√			√	√	√
3I5N	√	√			√	√
3DKF	√	√			√	
2WGJ	√		√		√	√
2WKM	√	√			√	

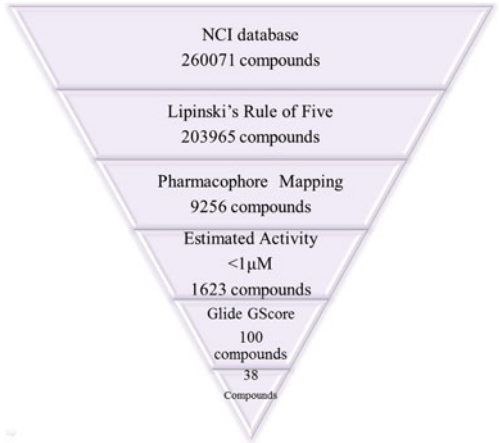

Fig. 10 Flowchart of virtual screening procedure applied in this paper

binding site of c-Met kinase. Thus, molecular docking was applied to prioritize the compounds by identifying their capability to interact with the receptor. The results of molecular docking were demonstrated based on both Glide GScore and the favorable interactions formed between ligands and active site.

In this study, with the purpose of further refining the retrieved hits and also excluding the false positives, all the 1623 drug-like hits were docked into the active site of c-Met crystal structure using Glide. The scoring functions (GScore) implemented in Glide were also used for ranking the molecules. We obtained conformations with GScore presented in ascending order and top 100 compounds were selected for detailed investigation. In order to ensure whether the compounds did interact with the crucial residues, nine co-crystal structures of c-Met bound with ligands were analyzed. By summarizing the interactions between the receptor residues and the ligands in the co-crystal structures of c-Met, hydrogen bond interactions with Met1160, Asp1222 and π-π stacking interactions with Tyr1230 and Tyr1159 were prevalently found. Table 5 clearly indicates that Met1160 and Tyr1230 were particularly crucial for c-Met binding.

The hit compounds whose docking conformations satisfied the binding model and formed key hydrogen bonds and π-π stacking interactions with crucial residues in the active site of c-Met were considered as candidates for further study. The compounds having similar structures with known inhibitors were rejected for retrieving new scaffolds. Finally, a small set of 38 hits with favorable drug-like properties, high estimated IC_{50} values, good GScores, and desired docking poses were obtained. Abundant structure diversity of the final hits demonstrated that hits with similar chemical features but novel scaffolds could be retrieved with the pharmacophore model. Figure 10 lists the steps and outcomes of the database screening and molecular docking procedure.

Among these hits, compounds NSC48870 and NSC614530, which are different in their chemical scaffolds, were identified as promising novel leads against c-Met kinase with high estimated IC_{50} value of 0.263 μM and 0.483 μM, respectively, and acceptable GScores. They mapped all the features of Hypo1 by choosing best/flexible searching option (Fig. 11), and the docked conformations formed hydrogen bonds and π-π stacking interactions with key residues in the active site of c-Met (Fig. 11). As they also satisfied all the drug-like properties, the two lead compounds would be focalized for further refining and optimizing to discover novel inhibitors with potent activity against c-Met kinase.

Conclusions

In this work, 3D-QSAR pharmacophore models were successfully generated with 23 known c-Met inhibitors. The

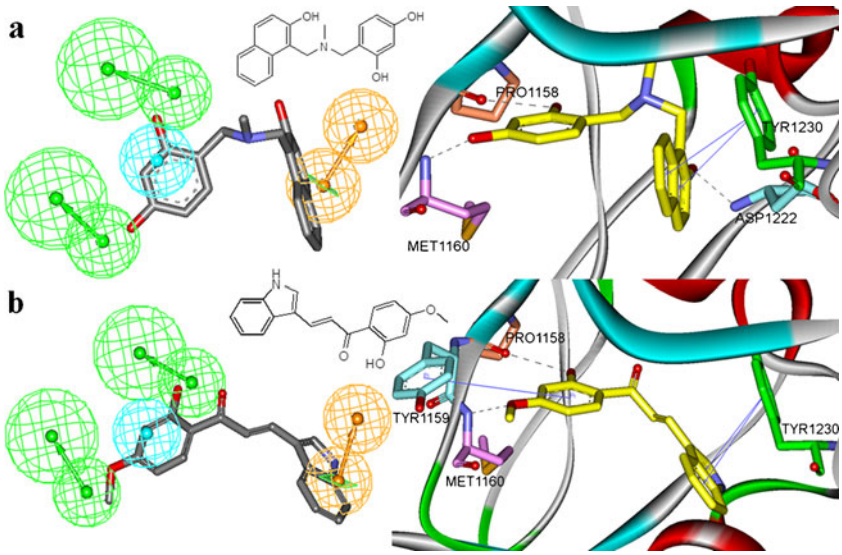

Fig. 11 Chemical structures, pharmacophore mapping and docked binding models of the two representative compounds. (**a**) NSC48870, (**b**) NSC614530

best pharmacophore model, Hypo1, consisted of two HBA, one HY, and one RA features. A good predictive power of the pharmacophore model was validated by test set prediction and Fischer randomization methods. Enrichment factor of 21.07 and ROC score of 0.896 indicated that Hypo1 performed fairly well at identifying active from inactive compounds. Moreover, compared with the binding site of c-Met kinase, Hypo1 represented the essential strucrural features for the c-Met inhibitors. The results suggested Hypo1 pharmacophore can serve as a reliable tool for the discovery of novel c-Met inhibitors. Database screening with Hypo1 finally retrieved 38 compounds in total, which satisfied all the drug-like properties, had good estimated active values and formed crucial interactions. Especially two compounds, NSC48870 and NSC614530, could serve as novel scaffolds for further refining and optimizing.

Acknowledgments The work was supported financially by the National Natural Science Foundation of China (30973609) and Qing Lan Project of Jiangsu Province, China.

References

1. Dancey J, Sausville EA (2003) Issues and progress with protein kinase inhibitors for cancer treatment. Nat Rev Drug Discov 2:296–313
2. Baselga J (2006) Targeting tyrosine kinases in cancer: the second wave. Science 312:1175–1178
3. Christensen JG, Burrows J, Salgia R (2005) c-Met as a target for human cancer and characterization of inhibitors for therapeutic intervention. Cancer Lett 225:1–26
4. Qian CN, Guo X, Cao B, Kort EJ, Lee CC, Chen J, Wang LM, Mai WY, Min HQ, Hong MH, Vande Woude GF, Resau JH, Teh BT (2002) Met protein expression level correlates with survival in patients with late-stage nasopharyngeal carcinoma. Cancer Res 62:589–596
5. Canadas I, Rojo F, Arumi-Uria M, Rovira A, Albanell J, Arriola E (2010) C-MET as a new therapeutic target for the development of novel anticancer drugs. Clin Transl Oncol 12:253–260
6. Underiner TL, Herbertz T, Miknyoczki SJ (2010) Discovery of small molecule c-Met inhibitors: Evolution and profiles of clinical candidates. Anticancer Agents Med Chem 10:7–27
7. Dussault I, Bellon SF (2009) From concept to reality: the long road to c-Met and RON receptor tyrosine kinase inhibitors for the treatment of cancer. Anticancer Agents Med Chem 9:221–229
8. Porter J (2010) Small molecule c-Met kinase inhibitors: a review of recent patents. Expert Opin Ther Pat 20:159–177
9. Liu X, Newton RC, Scherle PA (2009) Developing c-MET pathway inhibitors for cancer therapy: progress and challenges. Trends Mol Med 16:37–45
10. Liang Z, Zhang D, Ai J, Chen L, Wang H, Kong X, Zheng M, Liu H, Luo C, Geng M, Jiang H, Chen K (2011) Identification and synthesis of N′-(2-oxoindolin-3-ylidene)hydrazide derivatives against c-Met kinase. Bioorg Med Chem Lett 21:3749–3754
11. Yuan H, Lu T, Ran T, Liu H, Lu S, Tai W, Leng Y, Zhang W, Wang J, Chen Y (2011) Novel strategy for three-dimensional fragment-based lead discovery. J Chem Inf Model 51:959–974
12. Sakkiah S, Thangapandian S, John S, Kwon YJ, Lee KW (2010) 3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors. Eur J Med Chem 45:2132–2140
13. Boezio AA, Berry L, Albrecht BK, Bauer D, Bellon SF, Bode C, Chen A, Choquette D, Dussault I, Fang M, Hirai S, Kaplan-Lefko P, Larrow JF, Lin MH, Lohman J, Potashman MH, Qu Y, Rex K, Santostefano M, Shah K, Shimanovich R, Springer SK, Teffera Y, Yang Y, Zhang Y, Harmange JC (2009) Discovery and optimization of potent and selective triazolopyridazine series of c-Met inhibitors. Bioorg Med Chem Lett 19:6307–6312
14. Albrecht BK, Harmange JC, Bauer D, Berry L, Bode C, Boezio AA, Chen A, Choquette D, Dussault I, Fridrich C, Hirai S, Hoffman D, Larrow JF, Kaplan-Lefko P, Lin J, Lohman J, Long AM, Moriguchi J, O'Connor A, Potashman MH, Reese M, Rex K, Siegmund A, Shah K, Shimanovich R, Springer SK, Teffera Y, Yang Y, Zhang Y, Bellon SF (2008) Discovery and optimization of triazolopyridazines as potent and selective inhibitors of the c-Met kinase. J Med Chem 51:2879–2882
15. Porter J, Lumb S, Franklin RJ, Gascon-Simorte JM, Calmiano M, Riche KL, Lallemand B, Keyaerts J, Edwards H, Maloney A, Delgado J, King L, Foley A, Lecomte F, Reuberson J, Meier C, Batchelor M (2009) Discovery of 4-azaindoles as novel inhibitors of c-Met kinase. Bioorg Med Chem Lett 19:2780–2784
16. Nishii H, Chiba T, Morikami K, Fukami TA, Sakamoto H, Ko K, Koyano H (2010) Discovery of 6-benzyloxyquinolines as c-Met selective kinase inhibitors. Bioorg Med Chem Lett 20:1405–14099
17. Porter J, Lumb S, Lecomte F, Reuberson J, Foley A, Calmiano M, le Riche K, Edwards H, Delgado J, Franklin RJ, Gascon-Simorte JM, Maloney A, Meier C, Batchelor M (2009) Discovery of a novel series of quinoxalines as inhibitors of c-Met kinase. Bioorg Med Chem Lett 19:397–400
18. Koolman H, Heinrich T, Bottcher H, Rautenberg W, Reggelin M (2009) Syntheses of novel 2,3-diaryl-substituted 5-cyano-4-azaindoles exhibiting c-Met inhibition activity. Bioorg Med Chem Lett 19:1879–1882
19. Cho SY, Han SY, Ha JD, Ryu JW, Lee CO, Jung H, Kang NS, Kim HR, Koh JS, Lee J (2010) Discovery of aminopyridines substituted with benzoxazole as orally active c-Met kinase inhibitors. Bioorg Med Chem Lett 20:4223–4227
20. Thangapandian S, John S, Sakkiah S, Lee KW (2011) Pharmacophore-based virtual screening and Bayesian model for the identification of potential human leukotriene A4 hydrolase inhibitors. Eur J Med Chem 46:1593–1603
21. John S, Thangapandian S, Sakkiah S, Lee KW (2010) Identification of potent virtual leads to design novel indoleamine 2,3-dioxygenase inhibitors: pharmacophore modeling and molecular docking studies. Eur J Med Chem 45:4004–4012
22. Debnath AK (2002) Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of Mycobacterium avium complex dihydrofolate reductase. J Med Chem 45:41–53
23. Li HF, Lu T, Zhu T, Jiang YJ, Rao SS, Hu LY, Xin BT, Chen YD (2009) Virtual screening for Raf-1 kinase inhibitors based on pharmacophore model of substituted ureas. Eur J Med Chem 44:1240–1249
24. Thangapandian S, John S, Sakkiah S, Lee KW (2010) Docking-enabled pharmacophore model for histone deacetylase 8 inhibitors and its application in anti-cancer drug discovery. J Mol Graph Model 29:382–395
25. Macdougall IJ, Griffith R (2008) Pharmacophore design and database searching for selective monoamine neurotransmitter transporter ligands. J Mol Graph Model 26:1113–1124
26. Bharatham N, Bharatham K, Lee KW (2007) Pharmacophore identification and virtual screening for methionyl-tRNA synthetase inhibitors. J Mol Graph Model 25:813–823
27. Fang C, Xiao Z, Guo Z (2011) Generation and validation of the first predictive pharmacophore model for cyclin-dependent kinase 9 inhibitors. J Mol Graph Model 29:800–808

28. Chen XM, Lu T, Lu S, Li HF, Yuan HL, Ran T, Liu HC, Chen YD (2010) Structure-based and shape-complemented pharmacophore modeling for the discovery of novel checkpoint kinase 1 inhibitors. J Mol Model 16:1195–1204
29. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO (2005) Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. J Med Chem 48:2534–2547
30. Zou J, Xie HZ, Yang SY, Chen JJ, Ren JX, Wei YQ (2008) Towards more accurate pharmacophore modeling: multicomplex-based comprehensive pharmacophore map and most-frequent-feature pharmacophore model of CDK2. J Mol Graph Model 27:430–438
31. Thangapandian S, John S, Sakkiah S, Lee KW (2011) Potential virtual lead identification in the discovery of renin inhibitors: application of ligand and structure-based pharmacophore modeling approaches. Eur J Med Chem 46:2469–2476
32. Vangrevelinghe E, Zimmermann K, Schoepfer J, Portmann R, Fabbro D, Furet P (2003) Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. J Med Chem 46:2656–2662
33. Tikhonova IG, Sum CS, Neumann S, Engel S, Raaka BM, Costanzi S, Gershengorn MC (2008) Discovery of novel agonists and antagonists of the free fatty acid receptor 1 (FFAR1) using virtual screening. J Med Chem 51:625–633
34. Rollinger JM, Steindl TM, Schuster D, Kirchmair J, Anrain K, Ellmerer EP, Langer T, Stuppner H, Wutzler P, Schmidtke M (2008) Structure-based virtual screening for the discovery of natural inhibitors for human rhinovirus coat protein. J Med Chem 51:842–851