Bioorganic & Medicinal Chemistry Letters 18 (2008) 4972–4977

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

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

Pharmacophore modeling and virtual screening for designing potential PLK1 inhibitors

Hui-Yuan Wang [a,b], Zhi-Xing Cao [a], Lin-Li Li [b], Pei-Du Jiang [a], Ying-Lan Zhao [a], Shi-Dong Luo [a], Li Yang [a], Yu-Quan Wei [a], Sheng-Yong Yang [a,*]

[a] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, Keyuan Road 4, Chengdu, Sichuan 610041, China
[b] West China School of pharmacy, Sichuan University, Sichuan 610041, China

ARTICLE INFO

Article history:
Received 22 March 2008
Revised 24 July 2008
Accepted 9 August 2008
Available online 14 August 2008

Keywords:
Pharmacophore modeling
Virtual screening
Polo-like kinase-1
Kinase inhibitor

ABSTRACT

Pharmacophore models of Polo-like kinase-1 (PLK1) inhibitors have been established by using the HipHop and HypoGen algorithms implemented in the Catalyst software package. The best quantitative pharmacophore model, Hypo1, which has the highest correlation coefficient (0.9895), consists of one hydrogen bond acceptor, one hydrogen bond donor, one hydrophobic feature, and one hydrophobic aliphatic feature. Hypo1 was further validated by test set and cross validation method. Then Hypo1 was used as a 3D query to screen several databases including Specs, NCI, Maybridge, and Chinese Nature Product Database (CNPD). The hit compounds were subsequently subjected to filtering by Lipinski's rule of five and docking study to refine the retrieved hits and as a result to reduce the rate of false positive. Finally, a total of 20 compounds were selected and have been shifted to in vitro and in vivo studies. As far as we know, this is the first report on the pharmacophore modeling even the first publicly reported virtual screening study of PLK1 inhibitors.

Polo-like kinases (PLKs), a type of Ser/Thr protein kinases, have been identified as key regulators of cell mitosis. Dysregulation of them usually leads to some pathological changes, particularly cancer. Human Polo-like kinase subfamily includes four closely related members, that is, PLK1, PLK2, PLK3, and PLK4. Of the four members, PLK1 is the best characterized one and has been found to be overexpressed in many tumor types, including colorectal, ovarian, breast, prostate, pancreatic cancer, and melanomas.[1–6] And it has been known that the overexpression of PLK1 often correlates with poor prognosis of many cancer patients.[7] All of these indicate its involvement in carcinogenesis and its potential as a therapeutic target.

Despite the fact that PLK1 has been regarded as a validated mitotic cancer target for a number of years, very few reports of small-molecule PLK1 inhibitors have appeared to date. Fortunately, the limited number of PLK1 inhibitors still bears sufficient diversity in chemical structure and bioactivity. These provide a good basis to elucidate the structure–activity relationship of these diverse compounds and further design novel PLK1 inhibitors.

The goal of this study is to develop pharmacophore models of PLK1 inhibitors based on a series of known PLK1 inhibitors by using the HipHop[8] and HypoGen[9] modules within Catalyst software.[10] Then the best quantitative pharmacophore model generated will be used as a 3D query to screen several databases including Specs, NCI, Maybridge, and Chinese Nature Product Database (CNPD). The hit compounds will subsequently be subjected to filtering by Lipinski's rule of five[11] and docking study. Finally, potential lead compounds will be shifted to the subsequent in vitro and in vivo, as well as further structural modification studies. As far as we know, this is the first report on the pharmacophore modeling even the first publicly reported virtual screening study of PLK1 inhibitors.

Pharmacophore modeling: A total of 42 PLK1 inhibitors were collected from literature.[12–29] From which, 21 compounds were well selected to form a training set, which was based on the principles of structural diversity and wide coverage of activity range (here the IC_{50} values of the training set compounds span a range of five orders of magnitude: ranging from 0.0008 to 12.6 μM). Chemical structures as well as experimental IC_{50} values of the training set compounds are given in Chart 1 (**1**–**21**). The remaining 21 compounds, shown in Chart 2 (**22**–**42**), constitute an independent test set.

Before performing the quantitative pharmacophore modeling, the qualitative HipHop model was generated based on the five most active compounds (**1**–**5**) in training set, the purpose of which is to identify pharmacophore features necessary for potent PLK1 inhibitors. In the HipHop run, the most active compound **1** was considered as 'reference compound' specifying a 'principal' value of 2 and a 'MaxOmitFeat' value of 0. The 'principal' and 'MaxOmit-

* Corresponding author. Tel.: +86 28 85164063; fax: +86 28 85164060.
E-mail address: yangsy@scu.edu.cn (S.-Y. Yang).

0960-894X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2008.08.033

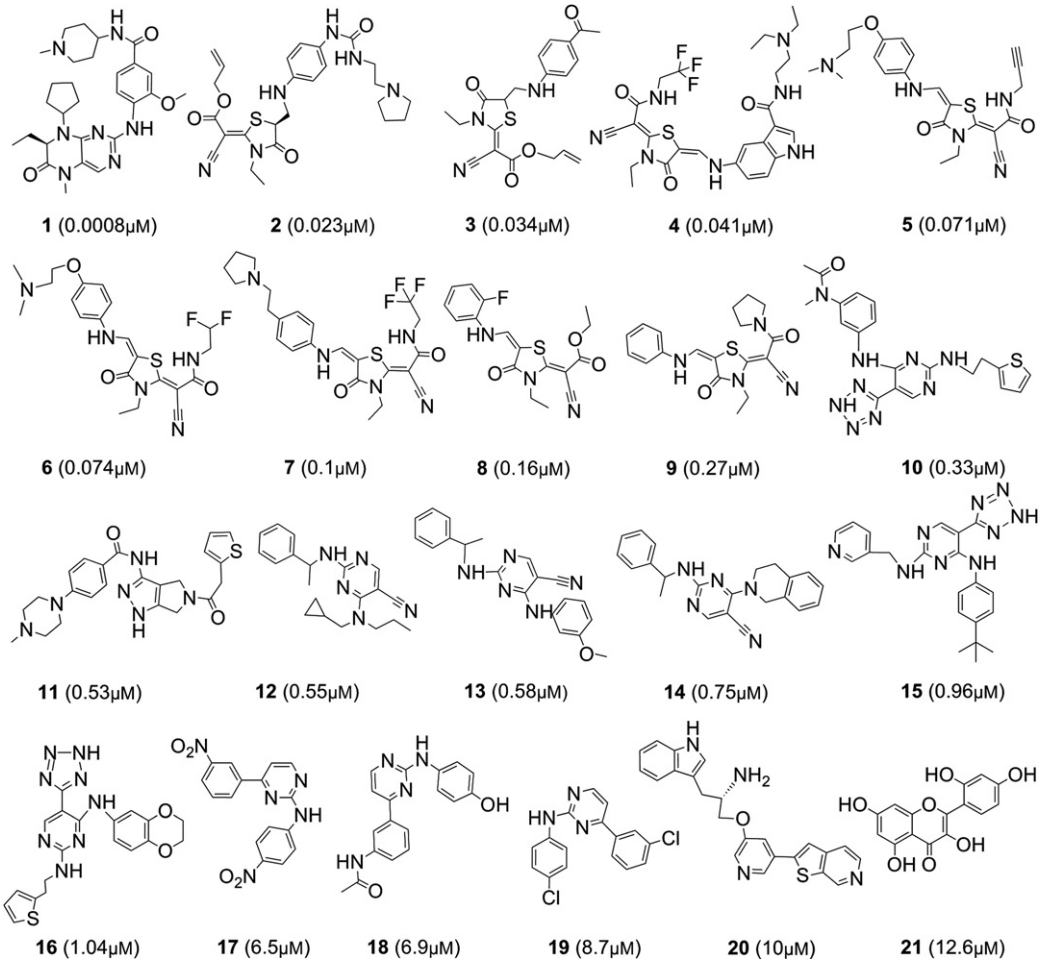

Chart 1. Chemical structures of the 21 training set compounds together with their experimental inhibitory activities (IC_{50} values, in parentheses).

Feat' values were set to 1 for the remaining four compounds. The 'spacing' value was set to 120. Except for these, all the rest HipHop parameters were kept at their default values. The initial features, which were specified based on an overview of all the training set molecules, include hydrogen-bond acceptor, hydrogen-bond donor, hydrophobic feature, hydrophobic aromatic feature, and hydrophobic aliphatic moiety. Figure 1A shows the best HipHop pharmacophore hypothesis, which contains seven features: three hydrogen-bond acceptors, one hydrogen-bond donor, two hydrophobic aliphatic moieties, and one hydrophobic aromatic feature. Figure 1B presents the mapping of the pharmacophore model onto the most active compound **1** in the training set.

The HipHop pharmacophore hypothesis clearly indicates the importance of hydrogen-bond acceptor, hydrogen-bond donor, hydrophobic aliphatic moiety, and hydrophobic aromatic feature. According to this information, hydrogen-bond acceptor, hydrogen-bond donor, and hydrophobic aliphatic and hydrophobic aromatic feature were selected as the initial pharmacophore

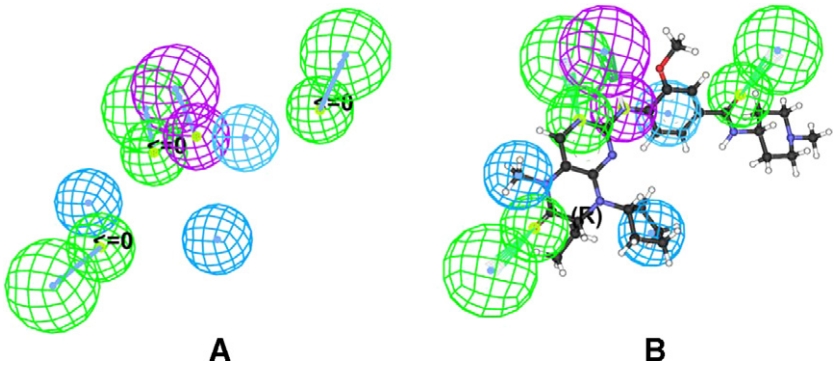

Figure 1. HipHop pharmacophore model for PLK1 inhibitors. (A) The best HipHop pharmacophore model. (B) The best HipHop model mapped with the most active compound **1** in the training set. The features are color coded with green, hydrogen-bond acceptor; magenta, hydrogen-bond donor; light blue, hydrophobic aliphatic feature; light blue, hydrophobic aromatic feature.

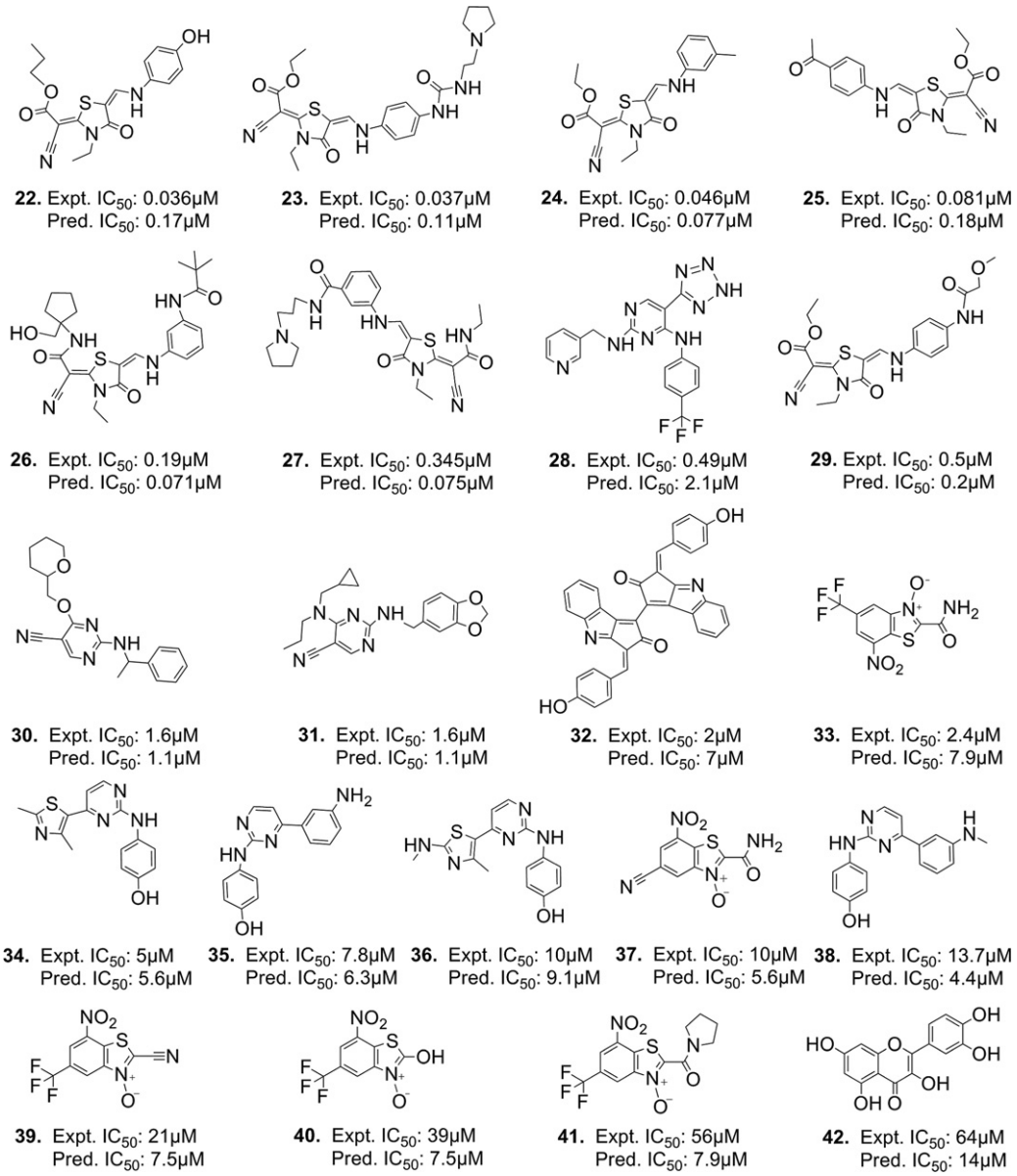

Chart 2. Chemical structures of the 21 test set molecules together with their experimental and predicted inhibitory activities (IC_{50} values).

features in the subsequent quantitative pharmacophore modeling. In addition, hydrophobic feature was also included, which was based on the results of several trials made by us. In these trials, we found that involving a more 'general' hydrophobic feature can further improve the quality of the developed pharmacophore model. The generated HypoGen models were evaluated according to Debnath[30] in terms of cost functions and statistical parameters, which were calculated by HypoGen module during hypothesis generation. A good pharmacophore model should have a high correlation coefficient, lowest total cost and rmsd values, and the total cost should be close to the fixed cost and away from the null cost. The ranked top 10 hypotheses as well as their statistical parameters are presented in Table 1. The best pharmacophore model (Hypo1, Fig. 2A), which was characterized by the lowest total cost value (82.4009), the highest cost difference (100.0901), the lowest RMSD (0.4995), and the best correlation coefficient (0.9895), contains four features, namely, one hydrogen bond acceptor, one hydrogen bond donor, and one hydrophobic and one hydrophobic aliphatic feature. The fixed cost and null cost are 78.7938 and 182.491 bits, respectively. The 3D space and distance constraints of these pharmacophore features are shown in Figure 2B. Figure 2C and D present the Hypo1 aligned with the most active compound **1** (IC_{50}: 0.0008 μM) and the least active compound **21** (IC_{50}: 12.6 μM) in the training set, respectively.

Then the training set compounds were roughly classified into three categories: highly active ($IC_{50} \leqslant 0.1$ μM, +++), moderately active (0.1 μM < $IC_{50} \leqslant 1$ μM, ++), and low active ($IC_{50} > 1$ μM, +). Table 2 shows the experimental and estimated inhibitory activities of the 21 training set compounds. Obviously, all compounds were correctly predicted except one compound (**15**), which originally is moderately active but is predicted as low active.

Validation of the pharmacophore model: An independent test set which contains 21 external compounds (Chart 2, **22**–**42**) was used to validate the established model (Hypo1). The experimental and predicted activities of the test set compounds are shown in Chart 2. The correlation coefficient is 0.894, indicating that Hypo1 has a good predictive ability.

Table 1
Statistical parameters of the top 10 pharmacophore models generated by the HypoGen

Hypo No.	Total cost[a]	Cost diff.[b]	RMSD	Correlation (r)	Features[c]
1	82.4009	100.0901	0.4995	0.9895	ADLH
2	82.5119	99.9791	0.5441	0.9873	ADLH
3	87.329	95.162	0.9012	0.9643	AADL
4	88.6378	93.8532	0.9678	0.9587	ADLL
5	88.7181	93.7729	0.9695	0.9585	ADLLH
6	88.7265	93.7645	0.9721	0.9583	ADLLH
7	89.6013	92.8897	0.9049	0.9642	ADLLH
8	91.5851	90.9059	1.1028	0.9460	ADLH
9	91.6144	90.8766	1.1044	0.9459	ADLL
10	91.8167	90.6743	1.1111	0.9452	ADDL

[a] The total cost value of a hypothesis is calculated by summing three cost factors, a weight cost (data not shown), an error cost (data not shown), and a configuration cost (a constant among all the hypotheses).
[b] The difference between the total cost of a hypothesis and that of the null hypothesis, roughly correlates with significance. The larger the difference, the greater the significance of the hypothesis. A true correlation in the data will very likely be estimated by models that exhibit a cost difference (Null cost − Total cost) (fixed cost = 78.7938, configuration cost = 16.7049, and null cost = 182.491). All cost values are in bits.
[c] A, hydrogen bond acceptor; D, hydrogen bond donor; H, hydrophobic feature; and L, hydrophobic aliphatic moiety.

Furthermore, Fischer randomization test[31] method was used to evaluate the statistical relevance of Hypo1 by using the CatScramble program implemented in Catalyst. The confidence level was set to 95%. Thereby CatScramble program generated 19 random spreadsheets to construct hypotheses using exactly the same conditions as used in generating the original pharmacophore hypothesis. The statistical parameters of pharmacophore models obtained in the 19 HypoGen runs as well as the original HypoGen run are presented in Table 3. From Table 3, one can see that out of the 19 runs, only one trial had a correlation value more than 0.8, but the RMSD value and total cost are much higher than the original pharmacophore hypothesis. The results of CatScramble clearly demonstrate that the original hypothesis is far more superior to those of the 19 randomization produced hypotheses, which provide confidence on our pharmacophore model.

Virtual screening: The validated hypothesis Hypo1 was used as a 3D structural query for retrieving potential inhibitors from chemical databases including Specs, NCI, Maybridge and CNPD. A total of 53,337 molecules show very good mapping with the Hypo1. Then these compounds were further screened by using the Lipinski's rule of five to make them more drug-like and 17,307 molecules passed this filtration.

Docking study: To further refine the retrieved hits, the 17,307 compounds were docked into the inhibitor binding site of PLK1 by using LigandFit within Cerius2 program package.[32] 3D structure of the PLK1 was taken from the crystal structure of PLK1 combined with compound **11** (PDB entry: 2owb).[17,33] Several scoring functions are available within LigandFit, including LigScore1, LigScore2,[34] PLP1, PLP2,[35] Jain,[36] Ludi,[37,38] PMF[39] as well as a consensus score.[40] Since there is no generally applicable scoring function so far, a solution to this problem is the calculation of a consensus score,[40] which makes use of the merits of different scoring functions by combining their results. The compounds were then ranked based on the consensus score. Finally we selected a total of 20 compounds from the ranked top 100 compounds, which was based on a visual look as well as on our experience. These selected compounds have been shifted to in vitro and in vivo studies.

In conclusion, chemical feature-based pharmacophore models of PLK1 inhibitors have been developed with the aid of HipHop and HypoGen modules in Catalyst program packet. The best quantitative pharmacophore model Hypo1, which was characterized by the lowest rmsd (0.4995) and the highest correlation coefficient (0.9895), consists of one hydrogen bond acceptor, one hydrogen bond donor, and one hydrophobic aliphatic and one hydrophobic feature. Hypo1 was further validated by test set and Fischer randomization test method. Results obtained from the test set method show a fairly good correlation between the experimental and estimated IC_{50} values (correlation coefficient of 0.894), indicating a good predictive power. And the results of Fischer randomization test by using CatScramble program within Catalyst further

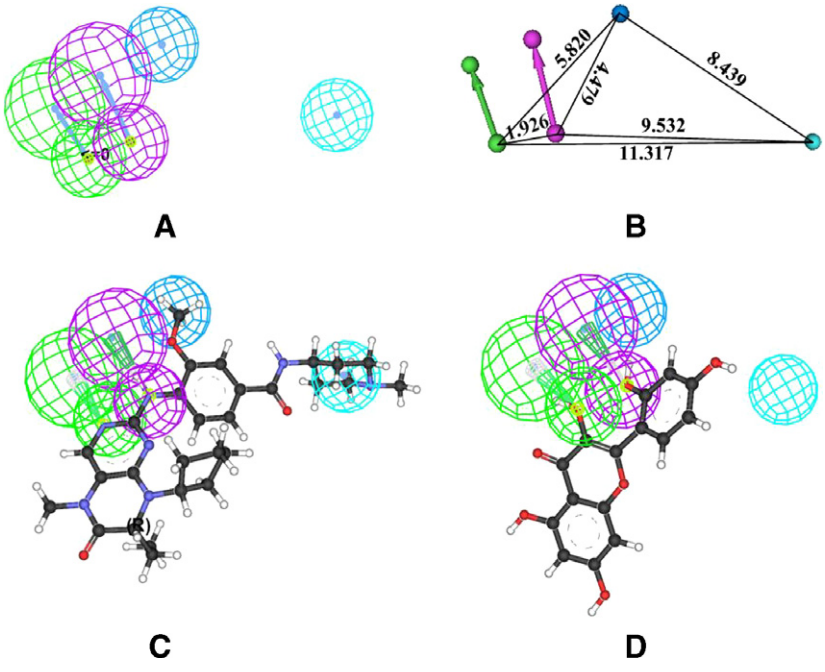

Figure 2. HypoGen pharmacophore model for PLK1 inhibitors. (A) The best HypoGen pharmacophore model Hypo1. (B) 3D spatial relationship and the distance constraints of the Hypo1. (C) Hypo1 mapped with the most active compound **1**. (D) Hypo1 mapped with the least active compound **21**. The features are color coded with green, hydrogen-bond acceptor; magenta, hydrogen-bond donor; cyan, hydrophobic feature, light blue, hydrophobic aliphatic moiety.

Table 2
Experimental and predicted IC_{50} values of the training set compounds

Molecule	Exptl. IC_{50} (μm)	Estimated IC_{50} (μm)	Error[a]	Fit value[b]	Exptl scale[c]	Estimated scale
1	0.0008	0.00073	−1.1	8.73	+++	+++
2	0.023	0.03	1.3	7.12	+++	+++
3	0.034	0.041	1.2	6.98	+++	+++
4	0.041	0.071	1.7	6.74	+++	+++
5	0.071	0.077	1.1	6.7	+++	+++
6	0.074	0.081	1.1	6.68	+++	+++
7	0.1	0.063	−1.6	6.79	+++	+++
8	0.16	0.12	−1.3	6.5	++	++
9	0.27	0.5	1.9	5.89	++	++
10	0.33	0.49	1.5	5.9	++	++
11	0.53	0.39	−1.4	6	++	++
12	0.55	0.71	1.3	5.74	++	++
13	0.58	0.48	−1.2	5.91	++	++
14	0.75	0.68	−1.1	5.76	++	++
15	0.96	1.7	1.8	5.35	++	+
16	1	0.68	−1.5	5.76	++	++
17	6.5	5.6	−1.2	4.85	+	+
18	6.9	5.6	−1.2	4.85	+	+
19	8.7	6.4	−1.4	4.79	+	+
20	10	5.5	−1.8	4.85	+	+
21	13	14	1.1	4.46	+	+

[a] + Means that the estimated IC_{50} is higher than the experimental IC_{50}; − means that the estimated IC_{50} is lower than the experimental IC_{50}; a value of 1 indicates that the estimated IC_{50} is equal to the experimental IC_{50}.
[b] Fit value indicates how well the features in the pharmacophore map the chemical features in the molecule.
[c] Activity scale: +++, $IC_{50} \leqslant 0.1$ μM (highly active); ++, 0.1 μM < $IC_{50} \leqslant 1$ μM (moderately active); and +, $IC_{50} > 1$ μM (low active).

Table 3
Results of Fischer's randomization test using CatScramble implemented in Catalyst software

Validation No.	Total cost	Fixed cost	RMSD	Correlation (r)
Hypo1	82.4009	78.7938	0.4995	0.9895
Results for scrambled				
Trial1	109.584	78.947	1.6962	0.867
Trial2	142.839	74.7415	2.4961	0.6801
Trial3	145.081	80.3603	2.4764	0.6858
Trial4	140.337	70.9407	2.5692	0.6556
Trial5	144.248	77.9207	2.5005	0.6787
Trial6	131.109	77.0637	2.2687	0.7452
Trial7	138.34	78.9096	2.3252	0.7305
Trial8	124.883	77.3403	2.0867	0.7901
Trial9	132.576	77.1285	2.2884	0.7401
Trial10	145.76	79.2811	2.5093	0.6754
Trial11	153.997	78.7725	2.6271	0.6361
Trial12	150.283	76.173	2.6442	0.63
Trial13	155.4	79.0369	2.6833	0.6151
Trial14	153.931	75.3303	2.7002	0.6091
Trial15	124.522	72.1496	2.2245	0.7567
Trial16	172.358	79.0476	2.9604	0.4938
Trial17	144.732	78.7831	2.497	0.6794
Trial18	129.606	76.6739	2.2243	0.7569
Trial19	149.381	77.053	2.6142	0.6405

confirmed the statistical confidence of Hypo1. Then Hypo1 was used as a 3D query to screen several databases including Specs, NCI, Maybridge, and CNPD. The hit compounds were subsequently subjected to filtering by Lipinski's rule of five and docking study to refine the retrieved hits. These refined hit compounds have been shifted to the subsequent in vitro and in vivo studies, the results of which will be reported in the near future.

Acknowledgments

This work was supported by the 863 Hi-Tech Program (2006AA020400), the National Natural Science Foundation of China (30772651), and the Youth Foundation of Sichuan Province.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2008.08.033.

References and notes

1. Weichert, W.; Schmidt, M.; Jacob, J.; Gekeler, V.; Langrehr, J.; Neuhaus, P.; Bahra, M.; Denkert, C.; Dietel, M.; Kristiansen, G. *Pancreatology* **2005**, *5*, 259.
2. Weichert, W.; Kristiansen, G.; Winzer, K. J.; Schmidt, M.; Gekeler, V.; Noske, A.; Muller, B. M.; Niesporek, S.; Dietel, M.; Denkert, C. *Virchows Arch.* **2005**, *446*, 442.
3. Weichert, W.; Schmidt, M.; Gekeler, V.; Denkert, C.; Stephan, C.; Jung, K.; Loening, S.; Dietel, M.; Kristiansen, G. *Prostate* **2004**, *60*, 240.
4. Weichert, W.; Denkert, C.; Schmidt, M.; Gekeler, V.; Wolf, G.; Kobel, M.; Dietel, M.; Hauptmann, S. *Br. J. Cancer* **2004**, *90*, 815.
5. Takahashi, T.; Sano, B.; Nagata, T.; Kato, H.; Sugiyama, Y.; Kunieda, K.; Kimura, M.; Okano, Y.; Saji, S. *Cancer Sci.* **2003**, *94*, 148.
6. Kneisel, L.; Strebhardt, K.; Bernd, A.; Wolter, M.; Binder, A.; Kaufmann, R. *J. Cutan. Pathol.* **2002**, *29*, 354.
7. Takai, N.; Hamanaka, R.; Yoshimatsu, J.; Miyakawa, I. *Oncogene* **2005**, *24*, 287.
8. Clement, O. O.; Mehl, A. T. In *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, 2000; pp 69–84.
9. Li, H.; Sutter, J.; Hoffman, R. In *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, 2000; pp 171–189.
10. CaTALYST 4.11; Accelrys Inc., San Diego, CA, 2005, http://www.accelrys.com.
11. Lipinski, C. A.; Lombardo, F.; Domiay, B. W.; Freeney, P. J. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3.
12. Moriarty, K. J.; Koblish, H.; Johnson, D. L.; Galemmo, R. A., Jr. *Top. Med .Chem.* **2007**, *1*, 207.
13. McInnes, C.; Mezna, M.; Fischer, M. P. *Curr. Top. Med. Chem.* **2005**, *5*, 181.
14. McInnes, C.; Mazumdar, A.; Mezna, M.; Meades, C.; Midgley, C.; Scaerou, F.; Carpenter, L.; Mackenzie, M.; Taylor, P.; Walkinshaw, M.; Fischer, P. M.; Glover, D. *Nat. Chem. Biol.* **2006**, *2*, 608.
15. Lansing, T. J.; McConnell, R. T.; Duckett, D. R.; Spehar, G. M.; Knick, V. B.; Hassler, D. F.; Noro, N.; Furuta, M.; Emmitte, K. A.; Gilmer, T. M.; Mook, R. A., Jr.; Cheung, M. *Mol. Cancer Ther.* **2007**, *6*, 450.
16. Johnson, E. F.; Stewart, K. D.; Woods, K. W.; Giranda, V. L.; Luo, Y. *Biochemistry* **2007**, *46*(95), 51.
17. Kothe, M.; Kohls, D.; Low, S.; Coli, R.; Cheng, A. C.; Jacques, S. L.; Johnson, T. L.; Lewis, C.; Loh, C.; Nonomiya, J.; Sheils, A. L.; Verdries, K. A.; Wynn, T. A.; Kuhn, C.; Ding, Y. –H. *Biochemistry* **2007**, *46*(59), 60.
18. Strebhardt, K.; Ullrich, A. *Nat. Rev. Cancer* **2006**, *6*, 321.
19. Schulze, V.; Eis, K.; Wortmann, L.; Schwede, W.; Siemeister, G.; Briem, H.; Schneider, H.; Eberspacher, U.; Hess-Stumpp, H. WO 2005/042505 A1.

20. Umehara, H.; Yamashita, Y.; Tsujita, T.; Arai, H.; Hagihara, K.; Machii, D. WO2004/043936.
21. Wang, S.; McLachlan, J.; Gibson, D.; Causton, A.; Turner, N.; Fischer, P. M. UK WO2005/012262 A1, WO 2005012262.
22. Drewry, D. H. M.; Mook, R. A. Jr; Salovich, J. M.; Schoenen, F. J.; Wagner, D. S.; Wagner, R. W. WO 2005/019193.
23. Cheng, M. K.; King, N. P.; Kunts, K. W.; Mook, R. A.; Pobanz, M. A.; Salovich, J. M.; Wilson, B. J. WO 2004/087652.
24. Emmitte, K. A. WO 2005/075470.
25. Davis-Ward, D.; Mook, R. A. J.; Neeb, M. J.; Salovich, J. M. WO 2004/074244.
26. McInnes, C.; Meades, C.; Mezna, M.; Fischer, P. M. WO 2004/067000.
27. Andrews, C. W. I.; Cheng, M.; Davis-Ward, R. G.; Drewry, D. H.; Emmitte, K. A.; Hubbard, R. D.; Kunts, K. W.; Linn, J. A.; Mook, R. A.; Smith,G. K.; Veal, J. M. WO 2004/014899.
28. Jacobs, R. S.; Stevenson, C. S.; Gerwick, W. H.; Marshall, L. A. WO 2001/62900 A1.
29. McInnes, C.; McLachan, J.; Mezna, M.; Fischer, P. M. WO 2005/047526 A2.
30. Debnath, A. K. J. *Med. Chem.* **2002**, *45*, 41.
31. Fischer, R. A. *The Design of Experiments*; Hafner Publishing: New York, 1966 (Chapter 2).
32. Cerius2 4.11; Accelrys Inc., San Diego, CA, 2005, http://www.accelrys.com.
33. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalv, I. N.; Bonrne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
34. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. J. *Mol. Graph. Model.* **2003**, *21*, 289.
35. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. R.; Fogel, L. J.; Freer, S. T. *Chem. Biol.* **1995**, *2*, 317.
36. Jain, A. N. J. *Comput. Aid. Mol. Des.* **1996**, *10*, 427.
37. Böhm, H. J. J. *Comput. Aid. Mol. Des.* **1994**, *8*, 243.
38. Böhm, H. J. J. *Comput. Aid. Mol. Des.* **1998**, *12*, 309.
39. Muegge, I.; Martin, Y. C. J. *Med. Chem.* **1999**, *42*, 791.
40. Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. *J. Mol. Graph. Model.* **2002**, *20*, 281.