Journal of Molecular Graphics and Modelling 37 (2012) 77–86

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

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

Pharmacophore identification, docking and "in silico" screening for novel CDK1 inhibitors

Xiaowu Dong, Jingying Yan, Lilin Du, Peng Wu, Shufang Huang, Tao Liu*, Yongzhou Hu**

ZJU-ENS Joint Laboratory of Medicinal Chemistry, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 20 September 2011
Received in revised form 4 April 2012
Accepted 17 April 2012
Available online 23 April 2012

Keywords:
CDK1 inhibitors
Pharmacophore
HypoGen
Homology modeling
LigandFit

ABSTRACT

Pharmacophore models of cyclin-dependent kinase-1 (CDK1) inhibitors were established by using the Catalyst/HypoGen. The best pharmacophore model, Hypo1, consists of one hydrogen bond acceptor (HBA), one hydrogen bond donor (HBD), one hydrophobic (HY) and one ring aromatic (RA) feature. The validation results of Hypo1 through cost analysis, test set prediction, Fisher's cross method and receiver operating characteristic (ROC) study indicated that the Hypo1 was statistically valuable and reliable in identifying structural diverse CDK1 inhibitors. It is further supported by the consistent results from molecular docking studies. Finally, the Hypo1 was used to "in silico" screen the NCI and MayBridge database. The preferable hits obtained were further docked into ATP binding site of CDK1, and nine promising compounds were retrieved as novel potential CDK1 inhibitors for further studies.

1. Introduction

Cyclin dependent kinases (CDKs), a family of serine/threonine kinases, not only play vital roles in the regulation of eukaryotic cell cycle progression but also are involved in apoptosis, neuronal cell physiology, differentiation, and transcription [1]. Regulated through phosphorylation, CDKs are activated by associating cyclin partners [2]. Thirteen CDKs (CDK1–CDK13) are identified, of which only five (CDK1, 2, 3, 4 and 6) directly intervene in the cell cycle [3]. Although it is believed that the correct sequences of the respective phases are regulated by these five CDKs, recent studies indicate that CDK1 is the only essential cell cycle CDK, because CDK1 is able to execute all the required events to drive cell division even in the absence of interphase CDKs (CDK2, 3, 4 and 6) [4–10]. These facts have raised great interests for the development of CDK1 inhibitors [11], leading to the discovery of a large number of structural diverse compounds [12–16], such as purines, pyrimidines, flavonoids, indolocarbazole analogs, with some of them are currently under clinical trials [17,18].

A rational strategy for identification of novel biologically active agents or leads with diverse chemical scaffolds involves utility of three-dimensional (3D) model generation and virtual searching. Computational techniques including pharmacophore, comparative molecular field analysis (CoMFA) and molecular docking, have been proved particularly helpful in identifying novel and potent inhibitors [19–21]. Pharmacophore stands for a set of generalized molecular features that are thought to be responsible for certain desired biological activities and very useful in virtual screening and structural optimization. Among the available approaches (such as HypoGen [22], MOE [23] and Phase [24]) involved in pharmacophore model generation, HypoGen is one of the most widely used methods which can not only predict the activities of known compounds in a test set but also select novel bioactive compounds from databases through "in silico" screening. Resorting to QSAR studies to develop CDK1 inhibitors would be a reasonable strategy since no crystallographic structure is available for complexes of CDK1 with inhibitors [25,26]. In spite of intensive studies for CDK1 inhibitors, limited results are available for the QSAR and 3D-phmacophore modeling of CDK1 inhibitors.

In the present study, a novel pharmacophore model was identified using Catalyst/HypoGen program based on a set of structural diverse CDK1 inhibitors [27–30]. The reliability and accuracy of established model (Hypo1) was seriously validated by cost analysis, Fisher's test and receiver operating characteristic (ROC). Moreover, we also describe the homology modeling and molecular dynamic and mechanical (MD/MM) simulation of a three-dimensional model of CDK1, since that the structure of CDK1 has not yet been determined. Then, the alignment of interaction mode proposed by pharmacophore model and molecular docking was performed to examine the consistency between each other. Finally, the Hypo1 was used as a 3D search query to screen database including NCI and MayBridge, and the retrieved hits were further evaluated by docking studies to check their shape fits in the binding site of protein to identify the most preferable screening hits.

* Corresponding author. Tel.: +86 571 88208458; fax: +86 571 88208458.
** Corresponding author. Tel.: +86 571 88208460; fax: +86 571 88208460.
E-mail addresses: lt601@zju.edu.cn (T. Liu), huyz@zju.edu.cn (Y. Hu).

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.04.003

2. Materials and methods

2.1. General methodology

All pharmacophore models generation and Hypo1-based virtual screening were performed using the HypoGen implemented in Catalyst (Catalyst 4.1, Molecular Simulations Inc., San Diego, CA). Homology model and Flexidock were conducted using SYBYL 6.9 software (TRIPOS Inc., Saint Louis, MO). Lipinski filtration was performed using Pipeline Pilot Studio (SciTegic, Inc., San Diego, CA). LigandFit docking studies were achieved using Discovery Studio 2.1/LigandFit software (Accelrys Inc., San Diego, CA). All of the calculations were performed on Dell Precision 380 workstation.

2.2. Data preparation

For the pharmacophore modeling studies, twenty-seven structural diverse CDK1 inhibitors (Fig. 1, Table 1) covering five orders of magnitude spanning 1 nM to 100,000 nM were selected from literatures [27–30] and used as a training set. Another seventy CDK1 inhibitors (Fig. S1, Table 2) were used as a test set of external data. For estimation purposes, the activity values were classified based on a CDK1 inhibitory activity scale: highly active (+++, $IC_{50} < 250$ nM); moderately active (++, 250 nM < IC_{50} < 10,000 nM), and weakly active (+, $IC_{50} > 10{,}000$ nM). Beside, 5000 negatives compounds in receiver operating characteristic (ROC) study were retrieved from Available Chemical Directory (ACD) database (Symyx Technologies, Santa Clara, CA) using "Random Percent Filter protocol" by Pipeline Pilot software. All of these compounds were optimized in Discovery Studio 2.0 software (Accelrys, Inc., San Diego, CA) using the CHARMm-like force field [31]. The conformation of these molecules was generated using the Poling Algorithm [32] with an energy range of 20 kcal/mol of the calculated potential energy minimum. Specify 250 as the maximum number of conformers for each molecule to ensure maximum coverage of the conformational space.

2.3. Pharmacophore hypotheses generation

Taking into account the chemical nature of the compounds considered in this work, a minimum of 0 to a maximum of 5 features involving the following four chemical feature types were selected to carry on the pharmacophore hypothesis generation process: hydrogen-bond acceptor (HBA), hydrogen-bond donor (HBD), hydrophobic group (HY), and ring aromatic (RA). Based on all of these features of twenty-seven compounds in the training set, 10 pharmacophore models were generated using HypoGen program with the default uncertainty value of 3. The uncertainty value represents a ratio range of uncertainty in the activity value based on the expected statistical irregularities of biological data collection.

2.4. Pharmacophore model evaluation

2.4.1. Cost analysis

Catalyst provides certain theoretical costs (represented in bit units) to assess the validity of the hypothesis. The first is the cost of an ideal hypothesis (fixed cost), which represents the simplest model that fits all data perfectly. The second is the cost of the null hypothesis (null cost), which represents the highest cost of a pharmacophore with no features and which estimates activity to be the average of the activity data of the training set molecules. If a returned cost (total cost) differs from the null cost by 40–60 bits, it is highly probable that the generated hypothesis has 75–90% chance representing the true correlation of the data. If the difference becomes more than 60 bits, the probability of representing the true correlation of the data are even higher. In addition, two other parameters that also determine the quality of any pharmacophore hypothesis with possible predictive values are the configuration cost, which depends on the complexity of the pharmacophore hypothesis space and should have a value <17, and the error cost, which is dependent on the correlation coefficient and root mean square (RMS) differences between the estimated and the actual activities of the training set molecules.

2.4.2. Fisher's cross validation [33]

The activity data associated with the training set compounds were reshuffled to generate pharmacophore hypotheses using the same features and parameters as had been used to develop the original one. This procedure was reiterated 19 times to achieve a 95% confidence level. If the pharmacophore generated with the randomized data had similar or better cost values, RMSD or correlations, the original hypothesis would be considered to be generated by chance, otherwise the obtained model is reliable.

2.4.3. Receiver operating characteristic (ROC) analysis [34]

The ROC allowed quick calculation of sensitivity (Se) and specificity (Sp) from a comparison between experimental and calculated data.

Sensitivity (Se): the likelihood that an event will be detected if that event is present.

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},$$

where TP is the number of true positives and FN is the number of false negatives.

Specificity (Sp): the likelihood that the absence of an event will be detected.

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}},$$

where FP is the number of false positive and TN is the number of true negatives.

The ROC curve is a graphical plot of the true positive rate (Se) versus false positive rate ($1 - \mathrm{Sp}$), and the area under the ROC curve (AUC) is the important way of measuring the performance of the test.

$$\mathrm{AUC} = \sum_{x=2}^{N} \mathrm{Se}(x)[(1 - \mathrm{Sp})(x) - (1 - \mathrm{Sp})(x - 1)],$$

where $\mathrm{Se}(x)$ is the percent of the true positives versus the total positives at rank position x, $(1 - \mathrm{Sp})(x)$ is the percent of the false positives versus the total negatives at rank position x.

2.5. Homology modeling

The amino acidic sequences of CDK1 were obtained from Swiss-Prot/TrEMBL database (http://expasy.org/). Sequence analysis reveals that the CDK1 and CDK2 proteins share an identity of 64.1% and a similarity of 80.9%. For this reason, the CDK sequences can be aligned readily and we used the pairwise sequence alignment program CLUSTALW [35] to align them. In present study, COMPOSER [36] implemented in SYBYL6.9 software (Tripos Inc., St. Louis, MO) was applied to build a three-dimensional model of the CDK1 structure. The X-ray structure of CDK2 obtained at 2.70 Å resolution and deposited in the protein data bank as 2CCI [37], was used as a template. Then, the ATP-bound CDK2 complex was further optimized by molecular dynamic and mechanical (MD/MM) calculations [38] using SYBYL6.9 software: (a) the minimization was initially run with Tripos force-field and Gasteiger–Huckel charges for 500 iterations of steepest descents, followed by a conjugate gradients optimization until the maximum derivative of

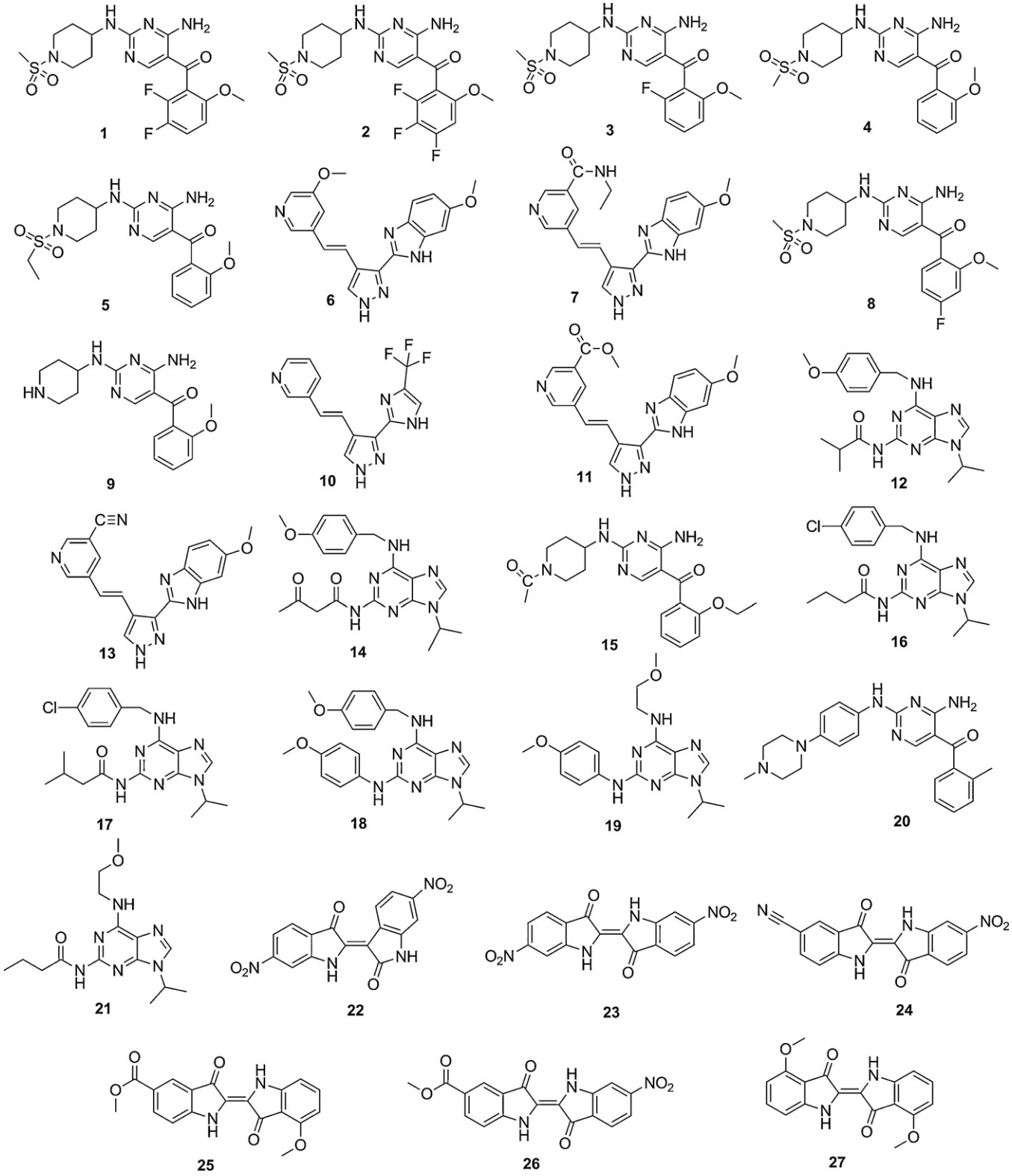

Fig. 1. Chemical structures of the 27 compounds forming the training set used to generate HypoGen pharmacophore model.

energy became less than 0.1 kcal/mol Å^{-1}; (b) an NVT ensemble molecular dynamic simulations in vacuum with Tripos force-field and Gasteiger–Huckel charges was then performed at a constant temperature of 373 K with a time step of 1 fs for 0.5 ns equilibration and 1 ns production, while atomic constraints were applied to retain the backbone atoms; (c) 200 representative ATP-bound CDK2 complexes were retrieved from the simulations. Each of these complexes was further minimized with 500 iterations of steepest descent without the constraints defined in the MD simulation and subsequently minimized using a conjugate gradient method until the maximum derivative of the total energy was less than 0.1 kcal/mol Å^{-1}. The other parameters

Table 1
The experimental and Hypo1-predicted IC_{50} values of compounds **1–27** in training set.

No.	IC_{50} (nM)		Error[a]	Activity scale[b]		No.	IC_{50} (nM)		Error[a]	Activity scale[b]	
	Exp.	Est.		Exp.	Est.		Exp.	Est.		Exp.	Est.
1	1	2.1	2.1	+++	+++	**15**	2500	1700	−1.4	++	++
2	5	3.7	−1.3	+++	+++	**16**	2900	2200	−1.3	++	++
3	17	4.1	−4.1	+++	+++	**17**	3000	4700	1.6	++	++
4	28	170	5.9	+++	+++	**18**	4600	3800	−1.2	++	++
5	43	85	2	+++	+++	**19**	10,000	1200	−8.4	++	++
6	47	260	5.5	+++	++	**20**	10,000	7400	−1.4	++	++
7	110	710	6.5	+++	++	**21**	13,000	10,000	−1.3	+	++
8	193	250	1.3	+++	+++	**22**	15,000	25,000	1.7	+	+
9	278	130	−2.1	++	+++	**23**	65,000	130,000	2.1	+	+
10	480	2100	4.3	++	++	**24**	70,000	40,000	−1.8	+	+
11	890	680	−1.3	++	++	**25**	93,000	41,000	−2.3	+	+
12	940	1300	1.4	++	++	**26**	100,000	41,000	−2.4	+	+
13	1000	1100	1.1	++	++	**27**	100,000	40,000	−2.5	+	+
14	2400	940	−2.5	++	++						

[a] The difference between estimated activity values and experimental activity values is represented as error (ratio between the estimated and experimental activity), with a negative sign if the actual activity is higher than the estimated.
[b] Activity scale: +++ (0–250 nM, highly active), ++ (250–10,000 nM, moderately active), + (>10,000 nM, weakly active).

of MD/MM simulation were maintained at their SYBYL default configuration. Among the MD/MM simulated conformers, the lowest-energy 3D structural model of the ATP-bound CDK2 complex was finally chosen for later structure-based virtual screening.

2.6. Molecular docking

2.6.1. LigandFit [39]

The docking site was derived from the position of the ATP bound in the binding site of proposed CDK1 structure using the LigandFit

Table 2
The experimental and Hypo1-predicted IC_{50} values of compounds **28–97** in test set.

No.	IC_{50} (nM)		Error[a]	Activity scale[b]		No.	IC_{50} (nM)		Error[a]	Activity scale[b]	
	Exp.	Est.		Exp.	Est.		Exp.	Est.		Exp.	Est.
28	14	240	17	+++	+++	**63**	2770	2000	−1.4	++	++
29	14	180	13	+++	+++	**64**	3000	11,000	3.7	++	+
30	38	590	15	+++	++	**65**	4000	660	−6.1	++	++
31	60	710	12	+++	++	**66**	4400	5900	1.3	++	++
32	133	230	1.7	+++	+++	**67**	5300	600	−8.9	++	++
33	160	1200	7.5	+++	++	**68**	6615	260	−25	++	++
34	181	290	1.6	+++	++	**69**	7000	2100	−3.3	++	++
35	246	350	1.4	+++	++	**70**	7200	2100	−3.4	++	++
36	344	1600	4.7	++	++	**71**	7500	25,000	3.3	++	+
37	439	4300	9.8	++	++	**72**	8300	39,000	4.7	++	+
38	540	1500	2.7	++	++	**73**	8500	17,000	2	++	+
39	687	1200	1.7	++	++	**74**	9000	24,000	2.7	++	+
40	781	2500	3.2	++	++	**75**	9600	8100	−1.2	++	++
41	808	1200	1.5	++	++	**76**	10,000	39,000	3.9	+	+
42	971	990	1	++	++	**77**	10,000	38,000	3.8	+	+
43	1040	410	−2.5	++	++	**78**	10,000	38,000	3.8	+	+
44	1060	1000	−1	++	++	**79**	10,000	4600	−2.2	+	+++
45	1108	450	−2.5	++	++	**80**	15,000	5800	−2.6	+	++
46	1108	1200	1.1	++	++	**81**	26,000	120,000	4.7	+	+
47	1140	720	−1.6	++	++	**82**	40,000	3800	−11	+	++
48	1200	7400	6.2	++	++	**83**	50,000	5300	−9.5	+	++
49	1240	1100	−1.1	++	++	**84**	50,000	45,000	−1.1	+	+
50	1240	16,000	13	++	+	**85**	50,000	4800	−11	+	++
51	1259	240	−5.3	++	+++	**86**	50,000	8800	−5.7	+	++
52	1500	2800	1.9	++	++	**87**	50,000	21,000	−2.4	+	+
53	1530	240	−6.3	++	+++	**88**	50,000	55,000	1.1	+	+
54	1540	15,000	9.8	++	+	**89**	50,000	49,000	−1	+	+
55	1610	950	−1.7	++	++	**90**	50,000	6700	−7.5	+	++
56	1638	21,000	13	++	+	**91**	50,000	9500	−5.2	+	++
57	1667	190	−8.7	++	++	**92**	50,000	26,000	−1.9	+	+
58	1700	960	−1.8	++	++	**93**	50,000	7600	−6.5	+	+
59	1770	6700	3.8	++	++	**94**	50,000	17,000	−2.9	+	+
60	2300	950	−2.4	++	++	**95**	50,000	9900	−5	+	++
61	2300	1500	−1.5	++	++	**96**	100,000	19,000	−5.3	+	+
62	2400	1500	−1.6	++	++	**97**	100,000	100,000	−1	+	+

[a] The difference between estimated activity values and experimental activity values is represented as error (ratio between the estimated and experimental activity), with a negative sign if the actual activity is higher than the estimated.
[b] Activity scale: +++ (0–250 nM, highly active), ++ (250–10,000 nM, moderately active), + (>10,000 nM, weakly active).

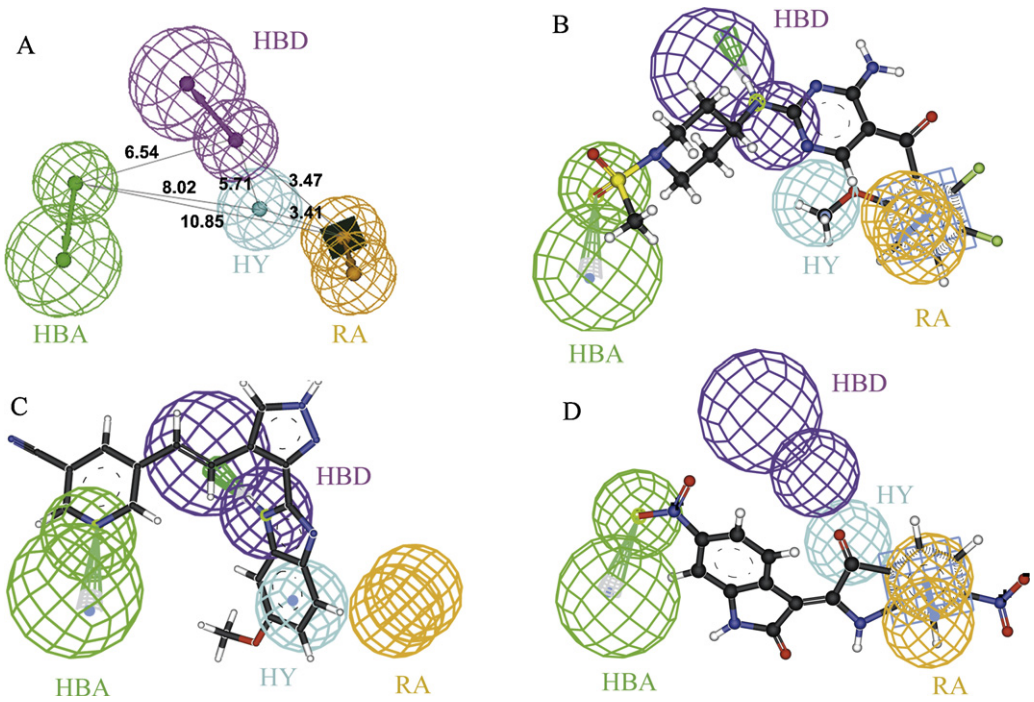

Fig. 2. Hypo1 and its mapping to representative compounds. (A) The topological features of Hypo1: the cyan contour represents hydrophobic (HY), green contour represents hydrogen bond acceptor (HBA), purple contour represents the hydrogen bond donor (HBD) and orange contour represents the ring aromatic (RA) features; (B) compound **1** mapped with Hypo1; (C) compound **13** mapped with Hypo1; (D) compound **22** mapped with Hypo1.

site search utility. For the generation of different ligand conformations, we used variable numbers of Monte Carlo simulations. All the calculations during the docking step were performed under the CFF 1.02 Force Field formalism. A short rigid body minimization was then performed and ten poses for each ligand were saved. Scoring was performed with a set of scoring functions implemented in LigandFit module, including LigScore1, LigScore2 [40], PLP1, PLP2 [41], and PMF [42]. CFF 1.02 force field was used for LigScore calculations.

2.6.2. Flexidock [43]

The binding pocket of CDK1 was defined to cover all residues within 4 Å of the ligand in the homology modeling structure of CDK1. All single bonds of residue side chains inside the CDK1 binding pocket were regarded as rotatable or flexible. The docked compound was allowed to rotate on all of single bonds and move flexibly within the tentative binding pocket. The obtained enzyme–ligand complex was further minimized with 500 iterations of steepest descent by using TRIPOS force field and calculated the interaction energy [44] between the CDK1 and ligand as below:

$$E_{\text{int}} = E(A, B) - (E(A) + E(B)),$$

where $E(A)$ and $E(B)$ are the energies of CDK1 and ligand respectively, and $E(A,B)$ the energy of docked complex.

3. Result and discussion

3.1. Characterization of pharmacophore model

HypoGen generated ten chemical-feature-based pharmacophore hypotheses (Table 3), among which Hypo1 was the best one characterized by the highest cost difference, lower errors and root mean square divergence (RMSD), and best correlation coefficient. As shown in Fig. 2A, it consists of one hydrogen-bond acceptor (HBA), one hydrogen-bond donor (HBD), one hydrophobic group (HY), and one ring aromatic (RA) feature. Compound **1**, the most potent one (actual activity: IC_{50} = 1 nM and estimated activity: IC_{50} = 2.1 nM), mapped well with all of the four features of Hypo1 (Fig. 2B). The twelve moderately active compounds (IC_{50}: 250–10,000 nM) either missed or partially overlapped with one feature, e.g. compound **13** (Fig. 2C). And the weakly active compounds (IC_{50} > 10,000 nM) missed or overlapped with some of the features, e.g. compound **22** (Fig. 2D). The above mapping results suggest that presence of four features are essential for the highly active compounds and three features are sufficient to account for the moderate activity. In other words, three-feature-hypothesis explained the moderately active compounds in the IC_{50} value of 250–10,000 nM, and the four-feature-hypothesis explained the remarkably improvement on biological activity of compound **1** (IC_{50} = 1 nM) over that of compound **13** (IC_{50} = 1000 nM). It is suggested that the established pharmacophore is capable of mapping structurally diverse compounds effectively.

3.2. Pharmacophore model examination

As shown in Table 3, a promising cost value was observed for Hypo1 (the fixed cost, null cost and total cost values of the hypothesis were 108.453, 199.495 and 120.236, respectively). Also, a high correlation coefficient of 0.951 (Fig. 3A) was observed as well as the low RMS value and configuration cost of 0.875 and 16.12, respectively. A test set containing 70 structural diverse compounds was further analyzed by Hypo1 (Table 2), giving a good a correlation coefficient of 0.711 (Fig. 3B), confirming the validity of the established HypoGen hypothesis in predicting novel CDK1 inhibitors which are not in training set.

The result of Fisher's cross validation clearly showed that among the 19 randomization trials no one had a cost value lower or close to that of Hypo1 (Fig. 4) when the experimental activities of compounds in the training set were scrambled randomly. It is indicated that there is a strong correlation between chemical structures and biological activities for Hypo1 and it is not generated by chance.

Table 3
Information of statistical significance and features measured for top 10 hypotheses as a result of HypoGen generation process.

Hypo.[a] no.	Total cost	Error cost	RMS	Corr. (r)	Features[b]
1	120.236	101.165	0.875	0.951	HBA, HBD, HY, RA
2	125.756	108.069	1.130	0.917	HBA, HBD, HY, RA
3	126.527	108.723	1.152	0.914	HBA, HBD, HY, RA
4	128.797	110.957	1.221	0.903	HBA, HBA, HBD, HY
5	128.828	111.191	1.228	0.901	HBA, HBD, HBD, HY
6	129.671	110.888	1.219	0.904	HBA, HBD, HY, RA
7	129.831	111.496	1.238	0.900	HBA, HBD, HY, RA
8	130.161	112.525	1.268	0.895	HBA, HBA, HY, RA
9	130.59	112.953	1.281	0.892	HBA, HBA, HY, RA
10	131.076	113.397	1.293	0.890	HBA, HBD, HY, RA

[a] Null cost = 199.495; fixed cost = 108.453; configuration cost = 16.5109. All costs are in units of bits.
[b] HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; HY, hydrophobic; RA, ring aromatic.

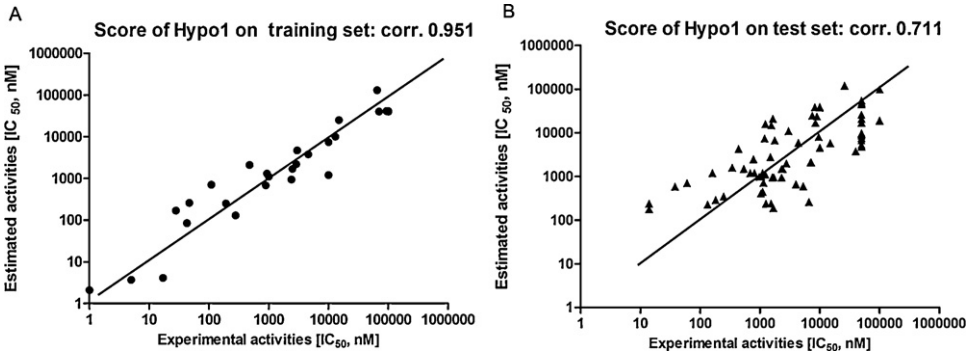

Fig. 3. The prediction ability of Hypo1. (A) The regression of experimental versus estimated activities by the Hypo1 for training set; (B) the regression of experimental versus estimated activities by the Hypo1 for test set inhibitors.

In ROC study, Hypo1 was further validated for picking up active molecules in the database, a spiked database including 5000 inactive compounds and 15 known inhibitors (10 highly active and 5 moderately active molecules). The AUC of ROC curve was 0.8928 against the randomly selection value of "0.5" (Fig. 5), demonstrating the reliability of Hypo1 which might be valuable in identifying diverse compounds with CDK1 inhibitory activity.

3.3. Homology modeling of CDK1

In the absence of a 3D structure for the CDK1, it is necessary to create a homology model from related proteins. The high degree of primary sequence identity between CDK2 and CDK1 (64.1%) indicated that CDK2 structures were suitable templates for CDK1. The sequence alignment result from CLUSTALW program is shown in Fig. 6A, which was consistent with the literature [45], in which the authors constructed reliable binary complexes between CDK1 and its inhibitors (roscovitine and flavopiridol). The total energy of 200 representative snapshots of the ATP-bound CDK1 complex recorded during a length of 1 ns of dynamic simulation. As shown in Fig. 6B, the total energy of ATP-bound CDK1 complex reached the relatively lowest arrived state when the dynamic simulation was performed from 0.91 to 0.96 ns (frame 182–192), and the lowest energy conformation (−3548.41 kcal/mol) of ATP-bound CDK1 complex, snapshooting at 0.93 ns during the simulation, was selected as preferable constructure for further study. As shown in Fig. 6C, the 3D structure of CDK1 is typical of the kinase family, and folded into the typical bilobal structure. It comprises an N-terminal domain of β-sheets and a larger C-terminal domain mainly constituted by α-helices.

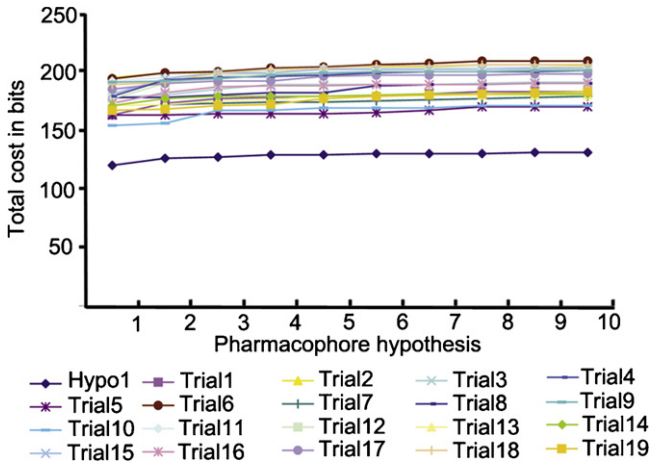

Fig. 4. The difference in costs between the HypoGen runs (Hypo1) and the scrambled runs (Trial1–Trial19).

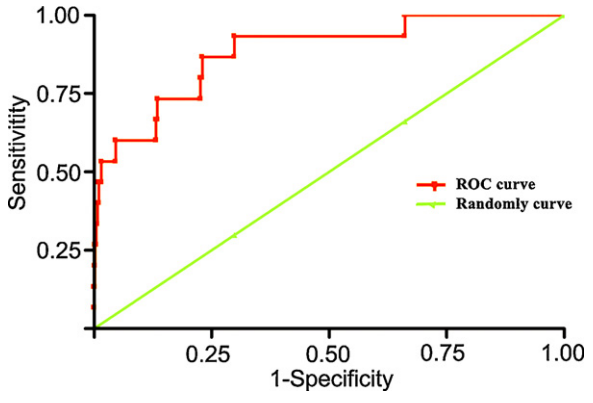

Fig. 5. ROC curve obtained by Hypo1 against randomly curve in the database screening.

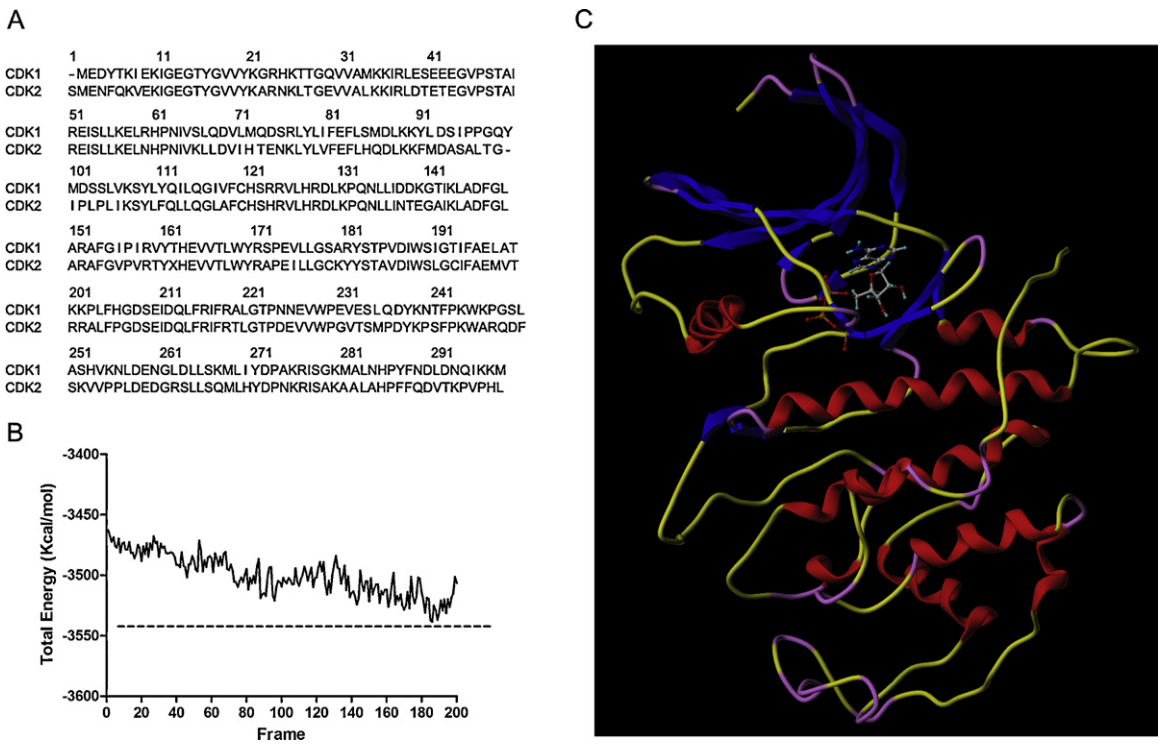

Fig. 6. The homology modeling of CDK1–ATP complex. (A) Pair-wise alignment between CDK1 and CDK2 sequences. Only two very small gap insertions were required to optimize the matrix score, as the two sequences are highly conserved. (B) Total energy of the trajectory of CDK1–ATP complex during the MM/MD simulation. (C) Final 3D structure of CDK1–ATP complex.

In order to validate the accuracy of proposed CDK1 structure, Flexidock study was performed to check the binding mode between CDK1 and compound **1**. The simulated model of compound 1-PKB/Akt1 complex is depicted in Fig. 7A, the interaction mode of which is consistent with previous reported mode between CDK1 and its inhibitors (i.e. roscovitine and flavopiridol) [45]. Three key hydrogen-bond interactions are formed between the diaminopyrimidine moiety of compound **1** and Glu81, Leu83 in the hinge region of CDK1. The sulfonamide moiety of compound **1** formed hydrogen bonds to Lys89 of CDK1. The distance of the mentioned hydrogen bonds is listed in Table 4. Besides, the 2,6-disubstitution pattern on the phenyl ring of compound **1** forced it out of the plane of the pyrimidine core allowing the methoxyphenyl ring to pack into the hydrophobic pocket consisting of Leu55, Val64, Ala145 and Phe80 of CDK1 and to form π–π stacking interaction of Phe80 of CDK1.

3.4. Comparing the features between Hypo1 and CDK1 binding site

As shown in Fig. 7B, the resultant interaction mapping from docking analysis was compared with that of Hypo1, in order to assess the consistency between the results from Hypo1 mapping and molecular docking. Highly identical interaction was observed, such as (i) the HBA and HBD features corresponded to the key hydrogen bond interaction of CDK1 with the sulfonamide and diaminopyrimidine core, (ii) the RA feature was well mapped with the aromatic ring of compound **1**, indicating that the equally important role of the π–π stacking interactions which further enhances the interaction, (iii) the HY feature of pharmacophore surrounded the methoxyl group of compound **1** and fit well into the hydrophobic pocket of CDK1, although a slightly different orientation of the methoxyl group of compound **1** proposed by Hypo1 and molecular docking was observed.

Table 4
The distance of hydrogen bonds between compound **1** and CDK1: proposed by Flexidock.

Compd. **1**	CDK1	Distance (Å)
NH_2(a)	GLU81 (C=O)	1.82
NH_2(b)	GLU81 (C=O)	2.33
NH	LEU83 (C=O)	2.29
N	LEU83 (NH)	2.59
SO_2(a)	LYS89 (NH_2)	2.39
SO_2(b)	LYS89 (NH_2)	3.72

3.5. Database screening

Hypo1 was used as a 3D structural query to retrieve potent molecules from NCI and MayBridge database. All queries were performed using the Fast Flexible search databases/Spreadsheet method in Catalyst. A total of 6000 lead compounds were obtained after preliminary screening which led to 55 molecules by restricting the minimum estimated activity to 200 nM. Further Lipinski's filtration yielded 32 drug-like molecules whose docking analyses were carried out to reduce the rate of false positive. Since that pharmacophore mapping and molecular docking can be synergistically integrated to improve the drug design and development process, all the hits obtained by Hypo1 were further evaluated by docking studies to check their shape fits in the binding site of CDK1 using LigandFit, the performance of which has been evaluated through analysis of its docking accuracy on several enzymes [46]. At last, a total of nine compounds (Fig. 8) were selected were identified as preferable hits in both Hypo1-based screening and molecular docking studies. The estimated activities (IC_{50}) of them from Hypo1 prediction as well as docking scores from various functions are

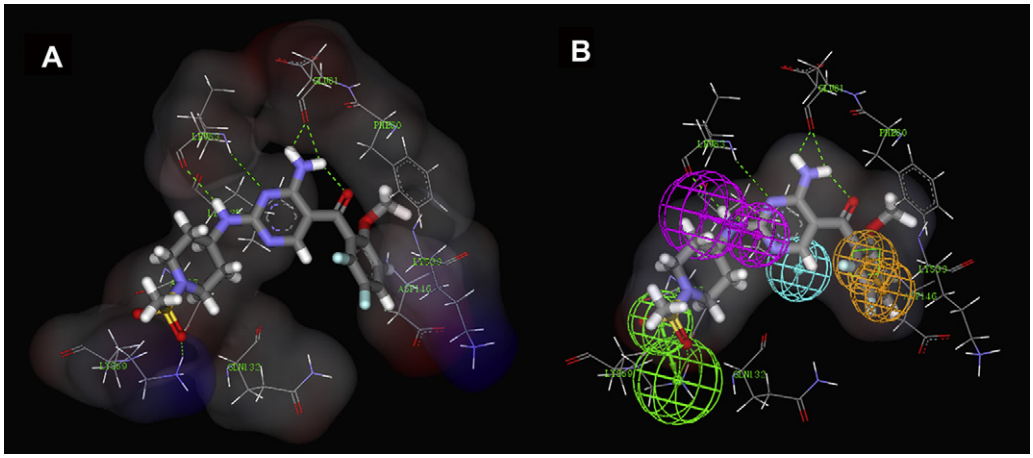

Fig. 7. The binding mode and consistency analysis between the results from Hypo1 and molecular docking. (A) The binding mode between CDK1 and compound **1**; (B) Hypo1 mapped to the complex of CDK1 with compound **1**. The hydrogen bonds interaction between ligand and enzyme were highlighted using dash line.

presented in Table 5. In case of AW00808, the most potent one predicted by Hypo1 (estimated IC_{50}: 27 nM), the hydroxyl group and 2-nitrogen of pyrimidine moiety of AW00808 are mapped well on HBD 1 and HBA of Hypo1, respectively, while the phenyl ring and 2-methyl on phenyl ring of AW00808 are mapping well on the RA and HY features, respectively (Fig. 9A). Beside, consensus scoring method which means the combination of multiple scoring functions and shows good efficiency in picking active compounds [47], was carried out to evaluate binding ability of tested compounds to CDK1. As shown in Table 5, all of the selected compounds shows good consensus results, with good performance in more than three scoring functions. As exampled by AW00808, the scores from LigScore1, LigScore2, PLP1, PLP2, and PMF of AW00808 (LigScore1: 4.03; LigScore2: 5.01; -PLP1: 70.92; -PLP2: 70.31; -PMF: 109.66) are promising and comparable to positive control that of compound **1** (LigScore1: 5.16; LigScore2: 5.29; -PLP1: 112.36; -PLP2: 95.75; -PMF: 99.14).

Moreover, Flexidock, an advanced molecular docking program, was applied to investigate more accurate binding mode and affinity between CDK1 and these nine compounds. The interaction energies of these compounds were calculated and presented in Table 5. Among them, AW00808 (−50.91 kcal/mol), DSHS00866

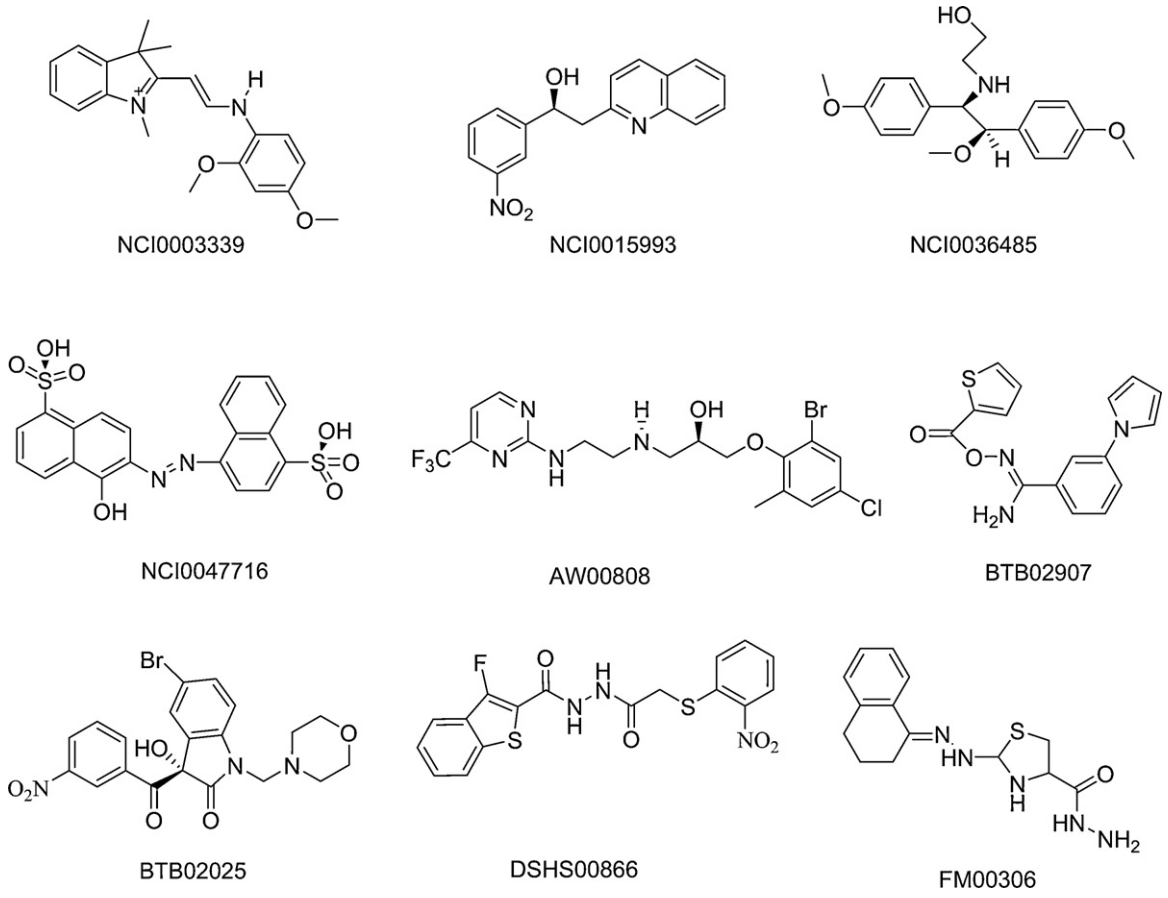

Fig. 8. Chemical structures of top ranked nine screening hits selected from NCI and MayBridge database.

Table 5
The estimated activity, interaction energy and LigandFit scoring results of top ranked nine compounds obtained from the combination of Hypo1-based virtual screening and molecular docking studies.

Name	Est. (nM)	Interaction energy (kcal/mol)	LigScore1	LigScore2	-PLP1	-PLP2	-PMF
Comp.**1**	2.1	−64.59	5.16	5.29	112.36	95.75	99.14
AW00808	27	−50.91	4.03	5.01	70.92	70.31	109.66
DSHS00866	43	−51.23	3.85	5.47	72.23	63.12	88.13
BTB02025	48	−63.48	4.53	5.08	81.96	69.06	116.07
NCI0003339	66	−53.75	2.72	5.64	77.94	72.6	107.76
BTB02907	150	−43.96	3.52	5.74	66.85	58.65	87.01
FM00306	160	−47.02	3.5	5.99	55.88	39.46	89.53
NCI0015993	180	−50.38	4.47	5.13	62.89	62.73	75.03
NCI0036485	190	−49.22	3.51	3.72	76.21	75.39	71.04
NCI0047716	190	−19.65	5.03	3.85	75.48	75.95	107.84

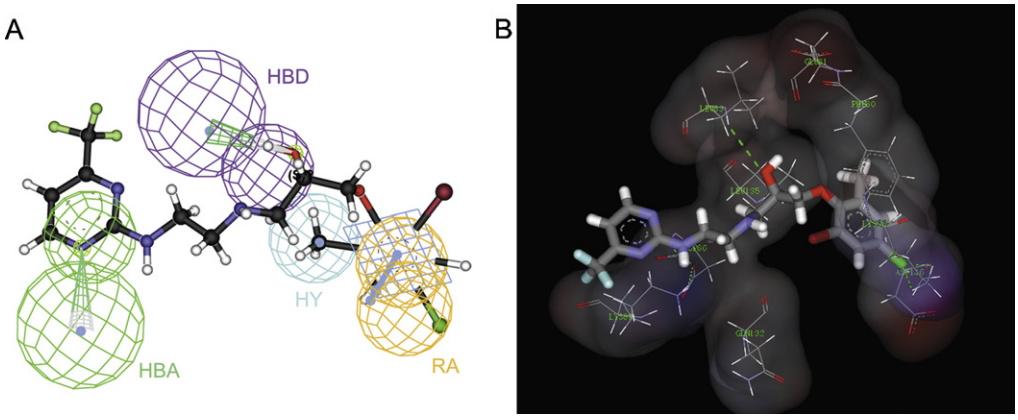

Fig. 9. The virtual evaluated results of preferable hit AW00808 by Hypo1 and molecular docking studies. (A) AW00808 mapped with Hypo1; (B) binding mode between AW00808 and CDK1, and the hydrogen bonds interaction between ligand and enzyme were highlighted using dash line.

(−51.23 kcal/mol) and BTB002025 (−63.48 kcal/mol) showed comparable interaction energy to compound **1** (−64.59 kcal/mol). For example, the exactly binding mode between CDK1 and AW00808, which is predicted as the most potent one by Hypo1, is displayed in Fig. 9B. Both of the results from LigandFit and Flexidock studies reveals that Leu83 in the hinge region made key hydrogen-bond interactions with the hydroxyl group of AW00808 and Asp86 form another hydrogen bond to the amine group on pyrimidine moiety of AW00808 (N—H···Asp86 2.19 Å; C—O···Leu83 3.13 Å). The 2,6-disubstitution phenyl ring played the same role as that of compound **1**, allowing the methyl group to pack into the hydrophobic pocket and forming a π–π stacking interaction with Phe80 of CDK1. Such an interactions mapping is consistent with previous discussion about Hypo1, in which there are four chemical features plays critical roles in the interaction between CDK1 and ligand.

4. Conclusions

In this study, HypoGen was used to generate pharmacophore models based on a series of known CDK1 inhibitors. The best pharmacophore model (Hypo1) shows good statistical parameters in validation process. The topological features of the established model were further confirmed by the consistent binding modes from the docking study using CDK1 Homology model. Then the model was used as a 3D query to screen the database including NCI and MayBridge. The hit compounds were subsequently subjected to Lipinski's filtration and docking study to refine the retrieved hits, and nine preferable compounds were selected for further investigation. These identified compounds maybe of interest for further structural optimization and biological evaluation of these compounds. We hope that the present data will be valuable to researchers for development of novel CDK1 inhibitors.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2012.04.003.

References

[1] M. Knockaert, P. Greengard, L. Meijer, Pharmacological inhibitors of cyclin-dependent kinases, Trends in Pharmacological Sciences 23 (2002) 417–425.
[2] M. Mihara, S. Shintani, A. Kiyota, T. Matsumura, D.T. Wong, Cyclin-dependent kinase inhibitor (roscovitine) suppresses growth and induces apoptosis by regulating Bcl-x in head and neck squamous cell carcinoma cells, International Journal of Oncology 21 (2002) 95–101.
[3] I. Collins, M.D. Garrett, Targeting the cell division cycle in cancer: CDK and cell cycle checkpoint kinase inhibitors, Current Opinion in Pharmacology 5 (2005) 366–373.
[4] M. Malumbres, S.L. Hunt, R. Sotillo, J. Martin, J. Odajima, A. Martin, P. Dubus, S. Ortega, M. Barbacid, Driving the cell cycle to cancer, Advances in Experimental Medicine and Biology 532 (2003) 1–11.
[5] O. Tetsu, F. McCormick, Proliferation of cancer cells despite CDK2 inhibition, Cancer Cell 3 (2003) 233–245.
[6] S. Ortega, I. Prieto, J. Odajima, A. Martin, P. Dubus, R. Sotillo, J.L. Barbero, M. Malumbres, M. Barbacid, Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice, Nature Genetics 35 (2003) 25–31.
[7] C. Berthet, E. Aleem, V. Coppola, L. Tessarollo, P. Kaldis, Cdk2 knockout mice are viable, Current Biology 13 (2003) 1775–1785.
[8] C.J. Sherr, J.M. Roberts, Living with or without cyclins and cyclin-dependent kinases, Genes and Development 18 (2004) 2699–2711.
[9] D. Santamaria, C. Barriere, A. Cerqueira, S. Hunt, C. Tardy, K. Newton, J.F. Caceres, P. Dubus, M. Malumbres, M. Barbacid, Cdk1 is sufficient to drive the mammalian cell cycle, Nature 448 (2007) 811–815.
[10] M. Leslie, The Journal of Cell Biology 178 (2007) 892.
[11] M. Malumbres, P. Pevarello, M. Barbacid, J.R. Bischoff, CDK inhibitors in cancer therapy: what is next? Trends in Pharmacological Sciences 29 (2008) 16–21.
[12] P.S. Sharma, R. Sharma, R. Tyagi, Inhibitors of cyclin dependent kinases: useful targets for cancer treatment, Current Cancer Drug Targets 8 (2008) 53–75.
[13] J. Vesely, L. Havlicek, M. Strnad, J.J. Blow, A. Donella-Deana, L. Pinna, D.S. Letham, J. Kato, L. Detivaud, S. Leclerc, et al., Inhibition of cyclin-dependent kinases by purine analogues, European Journal of Biochemistry 224 (1994) 771–786.

[14] M. Barvian, D.H. Boschelli, J. Cossrow, E. Dobrusin, A. Fattaey, A. Fritsch, D. Fry, P. Harvey, P. Keller, M. Garrett, F. La, W. Leopold, D. McNamara, M. Quin, S. Trumpp-Kallmeyer, P. Toogood, Z. Wu, E. Zhang, Pyrido[2,3-d]pyrimidin-7-one inhibitors of cyclin-dependent kinases, Journal of Medicinal Chemistry 43 (2000) 4606–4616.
[15] A. Wirger, F.G. Perabo, S. Burgemeister, L. Haase, D.H. Schmidt, C. Doehn, S.C. Mueller, D. Jocham, Flavopiridol, an inhibitor of cyclin-dependent kinases, induces growth inhibition and apoptosis in bladder cancer cells in vitro and in vivo, Anticancer Research 25 (2005) 4341–4347.
[16] G. Zhu, S.E. Conner, X. Zhou, C. Shih, T. Li, B.D. Anderson, H.B. Brooks, R.M. Campbell, E. Considine, J.A. Dempsey, M.M. Faul, C. Ogg, B. Patel, R.M. Schultz, C.D. Spencer, B. Teicher, S.A. Watkins, Synthesis, structure-activity relationship, and biological studies of indolocarbazoles as potent cyclin D1-CDK4 inhibitors, Journal of Medicinal Chemistry 46 (2003) 2027–2030.
[17] K.S. Joshi, M.J. Rathos, R.D. Joshi, M. Sivakumar, M. Mascarenhas, S. Kamble, B. Lal, S. Sharma, In vitro antitumor properties of a novel cyclin-dependent kinase inhibitor, P276-00, Molecular Cancer Therapeutics 6 (2007) 918–925.
[18] J.C. Byrd, T.S. Lin, J.T. Dalton, D. Wu, M.A. Phelps, B. Fischer, M. Moran, K.A. Blum, B. Rovin, M. Brooker-McEldowney, S. Broering, L.J. Schaaf, A.J. Johnson, D.M. Lucas, N.A. Heerema, G. Lozanski, D.C. Young, J.R. Suarez, A.D. Colevas, M.R. Grever, Flavopiridol administered using a pharmacologically derived schedule is associated with marked clinical efficacy in refractory, genetically high-risk chronic lymphocytic leukemia, Blood 109 (2007) 399–404.
[19] H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design—Recent Advances, vol. 3, Kluwer Escom, 1998, pp. 317–338.
[20] O. Dror, A. Shulman-Peleg, R. Nussinov, H.J. Wolfson, Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design, Current Medicinal Chemistry 11 (2004) 71–90.
[21] S. Vadivelan, B.N. Sinha, G. Rambabu, K. Boppana, S.A. Jagarlapudi, Journal of Molecular Graphics & Modelling 26 (2008) 935–946.
[22] A. Rinaldi, More biotechnology cosmetic products that claim drug-like properties reach the market, EMBO Reports 9 (2008) 1073–1077.
[23] L. Di, E.H. Kerns, Application of pharmaceutical profiling assays for optimization of drug-like properties, Current Opinion in Drug Discovery & Development 8 (2005) 495–504.
[24] S.L. Dixon, A.M. Smondyrev, S.N. Rao, Chemical Biology & Drug Design 67 (2006) 370–372.
[25] P. Ducrot, M. Legraverend, D.S. Grierson, 3D-QSAR CoMFA on cyclin-dependent kinase inhibitors, Journal of Medicinal Chemistry 43 (2000) 4098–4108.
[26] C. Kunick, K. Lauenroth, K. Wieking, X. Xie, C. Schultz, R. Gussio, D. Zaharevitz, M. Leost, L. Meijer, A. Weber, F.S. Jorgensen, T. Lemcke, Evaluation and comparison of 3D-QSAR CoMSIA models for CDK1 CDK5, and GSK-3 inhibition by paullones, Journal of Medicinal Chemistry 47 (2004) 22–36.
[27] Z.L. Wu, P. Aryal, O. Lozach, L. Meijer, F.P. Guengerich, Biosynthesis of new indigoid inhibitors of protein kinases using recombinant cytochrome P450 2A6, Chemistry & Biodiversity 2 (2005) 51–65.
[28] X.J. Chu, W. DePinto, D. Bartkovitz, S.S. So, B.T. Vu, K. Packman, C. Lukacs, Q. Ding, N. Jiang, K. Wang, P. Goelzer, X. Yin, M.A. Smith, B.X. Higgins, Y. Chen, Q. Xiang, J. Moliterni, G. Kaplan, B. Graves, A. Lovey, N. Fotouhi, Discovery of [4-amino-2-(1-methanesulfonylpiperidin-4-ylamino)pyrimidin-5-yl](2,3-difluoro-6-methoxyphenyl)methanone (R547), a potent and selective cyclin-dependent kinase inhibitor with significant in vivo antitumor activity, Journal of Medicinal Chemistry 49 (2006) 6549–6560.
[29] L. Vandromme, M. Legraverend, S. Kreimerman, O. Lozach, L. Meijer, D.S. Grierson, A Pd(0) based cross-coupling approach to the synthesis of 2-amidopurines and their evaluation as CDK inhibitors, Bioorganic and Medicinal Chemistry 15 (2007) 130–141.
[30] R. Lin, G. Chiu, Y. Yu, P.J. Connolly, S. Li, Y. Lu, M. Adams, A.R. Fuentes-Pesquera, S.L. Emanuel, L.M. Greenberger, Design, synthesis, and evaluation of 3,4-disubstituted pyrazole analogues as anti-tumor CDK inhibitors, Bioorganic and Medicinal Chemistry Letters 17 (2007) 4557–4561.
[31] B.R. Brooks, R.E. Brucolleri, B.D. Olafson, D.J.S.S. States, M. Karplus, CHARMM: a program for macromolecular energy minimization, and dynamics calculations, Journal of Computational Chemistry 4 (1983) 187–214.
[32] A. Smellie, S.D. Kahn, S.L. Teig, Analysis of conformational coverage. 1. Validation and estimation of coverage, Journal of Chemical Information and Computer Sciences 35 (1995) 285–294.
[33] W.M. Kazmierski, W. Andrews, E. Furfine, A. Spaltenstein, L. Wright, Discovery of potent pyrrolidone-based HIV-1 protease inhibitors with enhanced drug-like properties, Bioorganic and Medicinal Chemistry Letters 14 (2004) 5689–5692.
[34] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters 27 (2006) 861–874.
[35] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Research 22 (1994) 4673–4680.
[36] M.J. Sutcliffe, F.R. Hayes, T.L. Blundell, Knowledge based modelling of homologous proteins, part II: rules for the conformations of substituted sidechains, Protein Engineering 1 (1987) 385–392.
[37] K.Y. Cheng, M.E. Noble, V. Skamnaki, N.R. Brown, E.D. Lowe, L. Kontogiannis, K. Shen, P.A. Cole, G. Siligardi, L.N. Johnson, The role of the phospho-CDK2/cyclin A recruitment site in substrate recognition, Journal of Biological Chemistry 281 (2006) 23167–23179.
[38] X.Q. Xie, J.Z. Chen, E.M. Billings, 3D structural model of the G-protein-coupled cannabinoid CB2 receptor, Proteins 53 (2003) 307–319.
[39] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites, Journal of Molecular Graphics and Modelling 21 (2003) 289–307.
[40] A. Krammer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman, LigScore: a novel scoring function for predicting binding affinities, Journal of Molecular Graphics and Modelling 23 (2005) 95–407.
[41] D.K. Gehlhaar, G.M. Rejto, P.A. Verkhivker, C.J. Sherman, D.B. Fogel, L.J. Fogel, S.T. Freer, Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming, Chemistry and Biology 2 (1995) 317–324.
[42] I.Y. Muegge, C. Martin, A general and fast scoring function for protein-ligand interactions: a simplified potential approach, Journal of Medicinal Chemistry 42 (1999) 791–804.
[43] Sybyl6.9, Molecular Modelling System, Tripos Associates, St. Louis, MO, USA, 2003.
[44] R. Vijayalakshmi, V. Subramanian, B.R. Nair, A study of the interaction of Cr(III) complexes and their selective binding with B-DNA: a molecular modeling approach, Journal of Biomolecular Structure & Dynamics 19 (2002) 1063–1071.
[45] F. Canduri, H.B. Uchoa, W.F. de Azevedo Jr., Molecular models of cyclin-dependent kinase 1 complexed with inhibitors, Biochemical and Biophysical Research Communications 324 (2004) 661–666.
[46] E.M. Krovat, T. Langer, Impact of scoring functions on enrichment in docking-based virtual screening: an application study on rennin inhibitors, Journal of Chemical Information and Computer Science 44 (2004) 1123–1129.
[47] M. Montes, M.A. Miteva, B.O. Villoutreix, Structure-based virtual ligand screening with LigandFit: pose prediction and enrichment of compound collections, Proteins 68 (2007) 712–725.