JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

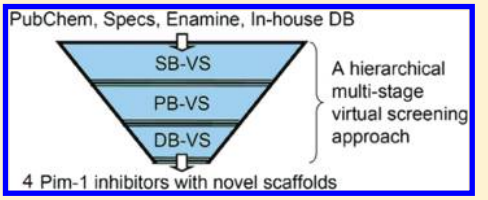
Discovery of Novel Pim-1 Kinase Inhibitors by a Hierarchical Multistage Virtual Screening Approach Based on SVM Model, Pharmacophore, and Molecular Docking

Ji-Xia Ren, Lin-Li Li, Ren-Lin Zheng, Huan-Zhang Xie, Zhi-Xing Cao, Shan Feng, You-Li Pan, Xin Chen, Yu-Quan Wei, and Sheng-Yong Yang*

State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, Chengdu, Sichuan 610041, China

Supporting Information

ABSTRACT: In this investigation, we describe the discovery of novel potent Pim-1 inhibitors by employing a proposed hierarchical multistage virtual screening (VS) approach, which is based on support vector machine-based (SVM-based VS or SB-VS), pharmacophore-based VS (PB-VS), and docking-based VS (DB-VS) methods. In this approach, the three VS methods are applied in an increasing order of complexity so that the first filter (SB-VS) is fast and simple, while successive ones (PB-VS and DB-VS) are more time-consuming but are applied only to a small subset of the entire database. Evaluation of this approach indicates that it can be used to screen a large chemical library rapidly with a high hit rate and a high enrichment factor. This approach was then applied to screen several large chemical libraries, including PubChem, Specs, and Enamine as well as an in-house database. From the final hits, 47 compounds were selected for further in vitro Pim-1 inhibitory assay, and 15 compounds show nanomolar level or low micromolar inhibition potency against Pim-1. In particular, four of them were found to have new scaffolds which have potential for the chemical development of Pim-1 inhibitors.



INTRODUCTION

The serine−threonine protein kinase proviral insertions in murine (Pim)-1 was first identified in studies of retroviral insertional mutagenesis of c-myc-induced murine lymphomas.[1] Afterward a lot of studies have indicated that Pim-1 plays important roles in the regulation of cell survival, proliferation, differentiation, and apoptosis.[2−4] Pim-1 has been found to be involved in several essential signaling pathways, typically, janus kinase/signal transducer and activator of transcription (JAK/STAT) and phosphatidylinositol 3-kinase/protein kinase B (or PKB)/mammalian target of rapamycin (PI3K/Akt/mTOR).[3,4] Pim-1 can promote the cell cycle progression, including G1/S transition by phosphorylating cell division cycle 25 homolog A (Cdc25A) and inactivating the cyclin kinase inhibitor (CKI) p21 Cip1[5,6] and G2/M transition by activating Cdc25C.[7] Pim-1 was found to be able to inactivate the pro-apoptotic protein Bad by phosphorylating it on the Ser112 gatekeeper site.[8] Elevated expression of Pim-1 is associated with the development and the progression of several hematopoietic malignancies and solid tumors, including acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic myeloid leukemia (CML), non-Hodgkin's lymphoma (NHL), and prostate cancer.[9,10] Recently it has been found that Pim-1 controls rapamycin-resistant T cell survival and activation.[11] Pim-1 has also been demonstrated to phosphorylate breast cancer resistance protein (BCRP), one of the multidrug resistant proteins and thereby promote its multimerization and drug-resistant activity in human prostate cancer cells.[12] Pim-1 was found to be a critical component of a survival pathway [STAT3/Pim-1/nuclear factor κB (NF-κB)] activated by docetaxel and to promote survival of docetaxel-treated prostate cancer cells.[13] Collectively, these findings established Pim-1 as a proto-oncogene and an important player in the process of malignant transformation and the development of resistance to chemotherapeutic agents. On the other hand, Pim-1 knockout mice show no obvious phenotype, implying small toxicity when targeting Pim-1.[3] Taken together, all of this evidence demonstrates that Pim-1 is a potential target for cancer intervention.

Due to the therapeutic values in cancer, the discovery of Pim-1 inhibitors has increasingly attracted much attention in recent years. Bregman and Meggers identified an organoruthenium half-sandwich complex as a potent Pim-1 inhibitor, which perfectly mimics the binding mode of the broad-spectrum kinase inhibitor staurosporine.[14] Imidazo[1,2-b]pyridazines were developed as Pim kinases inhibitors by the Knapp group and the SuperGen company independently.[15,16] The most potent compound of this series, namely SGI-1776 developed by SuperGen, is in the phase I clinical trials for treating docetaxel refractory prostate cancer and relapsed/refractory NHL (http://apps.who.int/trialsearch/trial.aspx?TrialID=NCT00848601; identifier: NCT00848601). Recently, the Vertex company reported triazolo[4,3-b]-pyridazines, benzo[c]isoxazole, indole, and quinolin-2(1H)-one derivatives as potent Pim-1 inhibitors.[17,18] During the period

Received: November 28, 2010
Published: May 27, 2011

Table 1. Molecular Descriptors Selected by the GA-CG Algorithm for the Generation of SVM Classification Model of Pim-1 Inhibitors and Noninhibitors

descriptor class	number of descriptors	descriptors
constitutional descriptors	16	Br_Count, C_Count, Cl_Count, F_Count, N_Count, O_Count, S_Count, Molecular_Weight, Num_Negative atoms, Num_Bridge bonds, Num_Rings, Num_Ring5, Num_Aromatic rings, Num_Fragments, Num_Ring4, Num_Ring assemblies
estate keys	20	ES_Sum_sCH3,x ES_Sum_dsCH, ES_Sum_sssCH, ES_Sum_tsC, ES_Sum_sNH2, ES_Sum_ssNH, ES_Sum_dsN, ES_Sum_aaN, ES_Sum_sssN, ES_Sum_sOH, ES_Sum_dO, ES_Sum_aaO, ES_Sum_ssS, ES_Sum_aaS, ES_Sum_ddssS ,ES_Count_dssC, ES_Count_aasC, ES_Count_sssN, ES_Count_ddsN, ES_Count_ddssS
lipophilicity descriptor	1	Apol
hydrogen-bonding descriptors	2	Num_H_Acceptors, Num_H_Donors
surface area and volume descriptor	1	Molecular_PolarSASA
topological descriptor	10	BIC, CIC, E_ADJ_equ, E_DIST_equ, E_DIST_mag, CHI_3_P, CHI_V_3_P, JY, Kappa_3, PHI

of preparation of this manuscript, a number of novel Pim-1 inhibitors have also been discovered by several groups.[19−21] In spite of the great progress in developing Pim-1 inhibitors, there is only one compound SGI-1776 in clinical studies now. Therefore, discovering more potent Pim-1 inhibitors, particularly with novel chemical scaffolds, is still needed and important, which could provide more lead candidates for anticancer drug development.

In silico VS has been thought as an economical and a rapid approach for lead discovery. Currently several VS methods have been well established, typically including molecular docking- and pharmacophore-based VS.[22,23] Recently, supporting vector machine (SVM), a machine learning method, has also been introduced into the virtual screening.[24−26] These methods have been broadly applied and are becoming a major source of lead compounds in drug discovery. However, these methods are individually far from perfect in many aspects, which will be discussed below. The docking-based VS (DB-VS) method, which is the most widely used VS method, involves docking of compounds from a virtual library into a protein target followed by applying a scoring function to estimate the likelihood that the ligand will bind to the protein with high affinity. Since DB-VS is based on the structural information of target, the hit compounds by DB-VS are supposed to be able to completely satisfy the requirements of the target active site in the molecular size, shape, and physical chemical properties. The DB-VS method also tends to select novel scaffolds. However, DB-VS still faces some difficulties, particularly, for example, lack of consideration of the receptor and/or ligand flexibilities and inaccurate scoring functions, which often lead to a low hit rate and a low enrichment factor as well as a high false positive rate. If these problems, even just part of them, are taken into account, then the computing time will dramatically increase, which limits its use in screening a large virtual chemical library. In the pharmacophore-based VS (PB-VS) approach, a pharmacophore hypothesis is taken as a template. The purpose of screening is actually to find such molecules (hits) which have chemical features similar to those of the template. Compared with DB-VS, PB-VS reduces the problems arising from inadequate consideration of protein flexibility and the use of insufficiently designed or optimized scoring functions. However, as we have indicated previously,[27] PB-VS results usually bear a higher false positive rate, which stems from the insufficient consideration of receptor information, such as the receptor flexibility and the steric restriction by the receptor. In addition, PB-VS can be still time consuming due to the sampling of the conformational space of small molecules, although it is not worse than DB-VS in terms of the screening speed. Different from DB-VS and PB-VS, the SVM-based VS (SB-VS) method identifies active compounds by using a SVM classifier derived from a set of known active (positive) and inactive (negative) compounds. The biggest advantage of SB-VS lies in the fast screening speed, which makes it possible to screen a vast chemical library rapidly. However, the SB-VS still suffers from a low hit rate and a high false positive rate, which might originate from the lack of consideration of the information regarding the macromolecule target and the pharmacophore features of small molecules.

Obviously, each of these VS methods has its own advantages and disadvantages. Each one might not perform optimally when used alone in terms of the speed and effectiveness of VS. An alternative approach is a combination of these methods. Since inherent limitations of each of these screening techniques are not easily resolved, their combination in a hybrid protocol can help to mutually compensate for these limitations and capitalize on their mutual strengths. For example, PB-VS usually lacks the consideration of steric restriction due to the receptor, which could be made up by the succeeding use of DB-VS. In addition, DB-VS and PB-VS are often slow in screening a large chemical library, a preceding use of SB-VS that is much faster than DB-VS and PB-VS could significantly reduce the number of chemicals needed to screen. Of course, the higher false positive rate of SB-VS could be decreased by the subsequent use of PB-VS and DB-VS.

In this investigation, we propose a hierarchical multistage VS approach based on SB-VS, PB-VS, and DB-VS. In this approach, the three VS methods are applied in an increasing order of complexity so that the first filter (SB-VS) is fast and simple, while successive ones (PB-VS and DB-VS) are more time-consuming but are applied only to a small subset of the entire database. This approach will be used to screen several large chemical libraries, including PubChem (18 831 686 compounds), Specs (202 408 compounds), and Enamine (980 000 compounds) as well as an in-house database (445 compounds) to retrieve novel potent Pim-1 inhibitors. Obviously, the proposed hierarchical multistage approach allows us to screen large chemical libraries efficiently and effectively in a reasonable time scale. We also hope that the use of the hierarchical multistage VS approach can help us to retrieve some new potent Pim-1 inhibitors particularly with novel chemical scaffolds.

Table 2. Results of the SVM Model Validation by Five-Fold Cross Validation and an Independent Validation Set[a]

	number of compounds			positive			negative			
method	total	negative	positive	TP	FN	SE (%)	TN	FP	SP (%)	Q (%)
five-fold cross validation	37 557	37 057	500	463	37	92.6	37 011	46	99.88	99.78
independent validation set	143	47	96	84	12	87.5	45	2	95.74	90.21

[a] TP, true positive; TN, true negative; FP, false positive; FN, false negative; SE(%): sensitivity, SE = TP/(TP + FN); SP (%): specificity, SP = TN/(TN + FP); Q (%): overall accuracy, Q = (TP + TN)/(TP + TN + FP + FN.

RESULTS AND DISCUSSION

Establishment and Validation of the SVM Classification Model of Pim-1 Inhibitors and Noninhibitors. For the generation of SVM classification model of Pim-1 inhibitors and noninhibitors, we first constructed a training set to train the SVM model. The training set comprises 37 557 compounds, including 500 known Pim-1 inhibitors[15−18,28−34] and 37 057 putative noninhibitors (a detailed description about the generation of the training set, see Materials and Methods Section). Initially 252 molecular descriptors, which cover various molecular properties, including geometrical, topological, and electronic properties, were selected. The initial descriptors were preprocessed, whose purpose is to eliminate those obvious "bad" descriptors. Here, the following descriptors were removed: (1) descriptors with too many zero values; (2) descriptors which are highly correlated with others (correlation coefficients >95%); and (3) descriptors with very small standard deviation values (<0.5%). After the preprocessing, a total of 82 molecular descriptors remained. These descriptor values were scaled to a range of −1 to +1, which is necessary since the different ranges of descriptor values will influence the quality of SVM model generated. Then, the 82 scaled descriptors underwent a further feature selection process by using our previously proposed genetic algorithm-conjugate gradient (GA-CG) method.[35] And 50 molecular descriptors were finally chosen for building the SVM model; these descriptors can be roughly grouped into several categories: constitutional descriptors (16), estate keys (20), lipophilicity descriptor (1), hydrogen-bonding descriptors (2), surface area and volume descriptor (1), topological descriptor (10) (see Table 1).

The generated SVM model was first evaluated by five-fold cross-validation. The prediction results for the training set agents are given in Table 2. From Table 2, we can see that, of the 500 Pim-1 inhibitors, 463 were correctly predicted (TP, Table 2). The accuracy for the prediction of Pim-1 inhibitors (SE, Table 2) is 92.6%. In the 37 057 Pim-1 noninhibitors, 37 011 were properly predicted (TN, Table 2). The accuracy for the prediction of Pim-1 noninhibitors (SP, Table 2) is 99.88%. The overall prediction accuracy (Q) is 99.78%. This indicates that the generated SVM classification model is quite good for the prediction of training set agents.

Subsequently the established SVM model was further validated by an independent test set, with the purpose of assessing the prediction ability of SVM model to external compounds outside of the training set. The independent test set used here comprises 96 inhibitors and 47 noninhibitors.[28,29,33,34] Of the 96 inhibitors, 84 (TP, Table 2) were correctly predicted, indicating a prediction accuracy of 87.5% for the inhibitors (SE, Table 2). For the 47 noninhibitors, 45 (TP, Table 2) were properly predicted. The accuracy for the prediction of noninhibitors (SP, Table 2) is 95.74%. Of all the 143 agents (inhibitors and noninhibitors), 129 were correctly predicted, and 14 were wrongly predicted (see Table 2). The overall prediction accuracy (Q) is 90.21%, which is comparable with that for the training set.

Pharmacophore Modeling of Pim-1 Inhibitors and Model Validation. A pharmacophore model is an ensemble of essential chemical features necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger its biological response. The purpose of pharmacophore modeling is to extract these essential chemical features from a set of known ligands, called a training set. In this investigation, eight Pim-1 inhibitors (compounds **1**−**8**, see Figure 1),[15,18,28,30−32,34,36] were selected to form the training set; these compounds were chosen since they are all the most active compounds and bear structural diversity. The HipHop algorithm implemented in the Discovery Studio (DS) program package (Accelrys, San Diego, CA) was used for the pharmacophore modeling. In the HipHop run, four pharmacophore hypotheses were finally generated. The top-ranked hypothesis Hypo1, which will be used for the subsequent virtual screening, contains one hydrogen-bond acceptor, one hydrogen-bond donor, and one hydrophobic aromatic feature. The 3D space and distance constraints of Hypo1 are shown in Figure 2A. Figure 2B depicts the mapping of Hypo1 with one of the most potent compound, **1** (SGI-1776). Clearly, compound **1** is mapped very well with Hypo1. In a recent study, Redkar et al.[37] docked SGI-1776 into the active site of Pim-1 and found that the 3-phenyl of SGI-1776 forms hydrophobic interactions with Pro123, Pro125, and Leu120. Forming a hydrogen-bond interaction with Lys67, the 1-N of imidazo[1,2-b] pyridazine motif acts as a hydrogen acceptor. The 6-NH motif was found to be necessary to improve the inhibition potency of the analogous compounds; a hydrogen-bond interaction was supposed to be formed between the 6-NH motif and a residue of Pim-1. All of these demonstrate that Hypo1 is a good pharmacophore model and contains the essential chemical features necessary for potent Pim-1 inhibitors.

Then, a test set, called M-CMC that contains 588 known Pim-1 inhibitors and 6978 decoys from Comprehensive Medicinal Chemistry (CMC) library, was used to examine whether the pharmacophore model Hypo1 has the ability to differentiate Pim-1 inhibitors and noninhibitors (the selection method of the decoys see Materials and Methods Section). Each of the M-CMC compounds was mapped onto Hypo1. Of the 588 Pim-1 inhibitors, 489 were found to map very well with Hypo1 (fitness ≥ 1.8), indicating a yield of 83.16%, and 1586 of the 6978 (22.73%) decoys also mapped very well with Hypo1. These results reveal two facts: (1) Hypo1 is capable of discriminating to some extent the Pim-1 inhibitors and noninhibitors; and (2) VS by using Hypo1 alone may suffer a high false positive rate, which is one of the most important reasons why we adopted the combined multivirtual-screening strategy here.

Determining Parameters and Scoring Functions for the Docking Study. Unlike the SB-VS and PB-VS, in which prediction models should be established in advance before performing

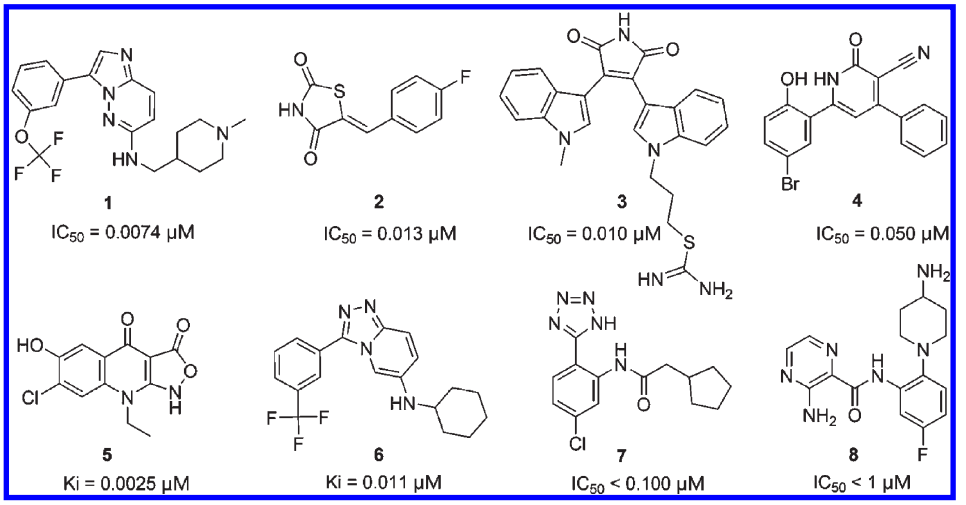


Figure 1. Chemical structures of Pim-1 kinase inhibitors in the training set together with their biological activity data (IC_{50} or K_i values) for the HipHop run.

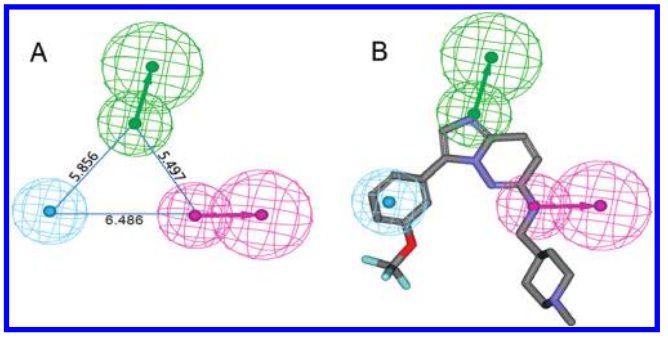


Figure 2. (A) The best pharmacophore model of Pim-1 inhibitors generated by HipHop. (B) The best HipHop model aligned with one of the most active compounds, **1** (IC_{50} = 0.0074 μM), in the training set. The features are color coded: green is hydrogen-bond acceptor; magenta is hydrogen-bond donor; and blue is hydrophobic aromatic feature.

VS, DB-VS seems straightforward if the receptor structure is known. However, docking parameters and scoring functions have been thought to have considerable effects on the final results of DB-VS. Thus an optimization of docking parameters and scoring functions is still necessary prior to executing the actual DB-VS.

In this investigation, GOLD 4.0[38] was adopted for the DB-VS; GOLD has been thought as one of the best docking programs.[39] The crystal structure (PDB entry 3BGQ) of the kinase domain of Pim-1 complexed with compound **6** (see Figure 1) was chosen as the reference structure of the receptor since it has the highest resolution (2.00 Å) among all the known Pim-1 crystal structures.[17,18] In order to determine the best docking parameters, seven active compounds that have been cocrystallized with Pim-1 (compounds **4**, **6**, and **9**−**13**, see Figure S1 in Supporting Information) were docked back to the active site of Pim-1.[16−18,20,28,40] We adjusted the docking parameters until the docked structures (both the poses and positions of heavy atoms) are as close as possible to their original crystallized structures in the binding site of Pim-1. The finally optimized docking parameters mainly include: the "number of dockings" was set to 10 without using early termination option; the "detect cavity" was turned on; the optimized positions of polar protein hydrogen atoms were saved; the GA parameter was set to "2 times speedup"; "Flip Planar R-NR1R2" was turn off; the top 10 scoring poses were saved for each compound; and the other parameters were kept as their default settings. The use of these parameters results in very small root-mean-square deviation (RMDS) values between the docked poses of the ligands and their corresponding bound conformations in the crystal structures. Table 3 presents the calculated RMDS values. Clearly, all the RMDS values are less than 2 Å except compound **13** (PDB entry 3JPV), whose RMDS value is 3.08 Å. These indicate that the GOLD[38] software is able to reproduce the correct poses of the ligands and is a reliable method for docking studies.

Table 3. RMSD Values of the Selected Compounds between Their Pim-1 Crystallized Conformations and Docked Conformations

compound	PDB entry	RMDS (Å)
4	2OBJ	0.91
6	3BGQ	0.66
9	3JXW	0.73
10	3F2A	1.32
11	3BGP	1.65
12	2C3I	1.07
13	3JPV	3.08

For the selection of scoring functions, we again chose a set of known Pim-1 inhibitors whose IC_{50} values span a range of three orders. These inhibitors were docked into the active site of Pim-1. Different scoring functions, including GoldScore, ChemScore, and a modified ChemScore that is an optimized scoring function for the kinase-related docking [hereafter called kinase scoring function (KCS)],[41] were calculated. Then the correlation coefficient between the experimentally measured IC_{50} values and the scoring function values was calculated for each individual scoring function. It was found that the scoring function KCS gave the best correlation coefficient. Therefore, the KCS scoring function will be adopted in the subsequent DB-VS study.

Evaluation of the Performance of the Hierarchical Multistage VS Approach in Virtual Screening. As indicated above,

dx.doi.org/10.1021/ci100464b |J. Chem. Inf. Model. 2011, 51, 1364–1375

the proposed hierarchical multistage VS approach is a combination of SB-VS, PB-VS, and DB-VS. These filters are used in a cascade fashion with the fastest filter SB-VS applied first, followed by the moderately fast filter PB-VS, and the slowest filter DB-VS last (see Figure 3, left). Before the actual application of the hierarchical multistage VS approach, an assessment of its performance in virtual screening will be made.

In order to carry out this evaluation, we constructed a test library, called M-MDDR, which comprises 203 known Pim-1 inhibitors (these compounds have never been used in previous modeling and model validations, see Materials and Methods Section) and 117 877 decoy compounds obtained from MDL Drug Data Report (MDDR) library (the selection method of the decoys see Materials and Methods Section). For the performance of VS, the yield (percentage of predicted compounds in known inhibitors), hit rate (percentage of known inhibitors in predicted compounds), and enrichment factor (ratio of hit rate to the percentage of known inhibitors in M-MDDR), which shows the magnitude of hit rate improvement over random selection, were evaluated.

First, SB-VS, PB-VS, and DB-VS are individually employed to screen the M-MDDR database. For SB-VS, the number of the predicted positive is 1530, and that of the total hits is 182 with a yield of 89.66% (182 known inhibitors out of 203, see Table 4). The hit rate and enrichment factor are 11.90% and 69.19, respectively. The time used for the screening of M-MDDR by SB-VS is about 1 min (0.017 h) on a PC desktop equipped with Intel E5420 (2500 MHz) processor if the molecular descriptors have been prepared (the time cost for the calculation of molecular descriptors of the M-MDDR compounds is about 5 h on the same computer). For PB-VS, the number of the predicted positive is 43 790, and that of the total hits is 173 with a yield of 85.22% (173 known inhibitors out of 203, see Table 4). The hit rate and enrichment factor are 0.40% and 2.30, respectively. The time cost for the screening of M-MDDR by PB-VS is about 6 h. For DB-VS, the number of the predicted positive is 24 152 (KCS value ≥ 35, since a test study indicates that most of the known Pim-1 inhibitors (more than 80%) afford a KCS value greater than 35, and that of the total hits is 163 with a yield of 80.30% (163 known inhibitors out of 203, see Table 4). The hit rate and enrichment factor are 0.67% and 3.89, respectively. The time cost is about 3118 h on the same computer as before, which is extremely slow compared with SB-VS and PB-VS.

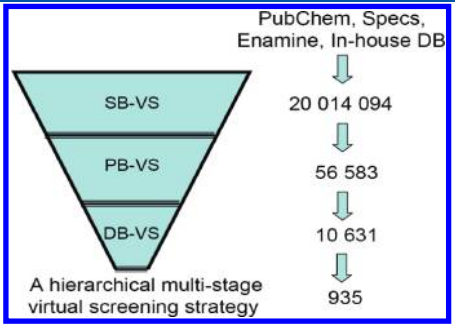


Figure 3. A schematic workflow of the proposed hierarchical multistage VS approach.

Further, the SB-VS and PB-VS were combined to screen the M-MDDR library with SB-VS first performed followed by PB-VS. The first step (SB-VS) identified 1530 positive compounds. These compounds were further filtered by PB-VS, and 989 (1530/989) compounds passed this filter. The number of total hits is 172 (182/172) with a yield of 84.73% (172 known inhibitors out of 203, see Table 4). The hit rate and enrichment factor are 17.39% and 101.16, respectively. Obviously, the combination of SB-VS and PB-VS considerably increases the hit rate and the enrichment factor compared with the sole use of SB-VS or PB-VS, and the time used is also much reduced compared with the sole use of PB-VS.

Finally the hierarchical multistage VS approach (SB-VS/PB-VS/DB-VS) was employed to screen M-MDDR library with SB-VS performed first, followed by PB-VS, and finally DB-VS. As mentioned above, the combined SB-VS/PB-VS method predicted 989 (1530/989) positive compounds. These compounds were further screened by DB-VS. The final number of positive compounds predicted by SB-VS/PB-VS/DB-VS is 377 (1530/989/377), and that of the total hits is 158 (182/172/158) with a yield of 77.83% (158 known inhibitors out of 203, see Table 4). The hit rate and enrichment factor are 41.91% and 243.78, respectively, which are significantly higher than the corresponding values of the individual VS methods and the combined SB-VS and PB-VS. The time totally used is 23 h which is significantly shortened compared with the sole use of DB-VS. From here, we can conclude that, in terms of the overall performance, the hierarchical multistage VS approach outperforms the sole SB-VS, PB-VS, and DB-VS as well as the combined SB-VS and PB-VS. This superiority of the hierarchical multistage VS approach is expected to be more significant in the screening of extremely large databases like PubChem.

Virtual Screening by Using the Hierarchical Multistage VS Approach for Retrieving Novel Potent Pim-1 Inhibitors. The proposed hierarchical multistage VS method was applied to screen several large chemical libraries including PubChem (18 831 686 compounds), Specs (202 408 compounds), and Enamine (980 000 compounds) as well as an in-house database (445 compounds) to retrieve new potent Pim-1 inhibitors. The first filter SB-VS had 56 583 compounds pass through. These compounds then underwent the second filtering process by PB-VS, and 10 631 compounds remained. Finally, the 10 631 compounds were subjected to docking study by using GOLD 4.0,[38] and 935 passed through the DB-VS filter.

Table 4. Evaluation Results of the Performance of Various VS Methods by Screening a Chemical Database that Comprises 203 Known Pim-1 Inhibitors and 117 877 Decoys from MDDR Library

method	predicted inhibitors	hits	yield (%)	hit rate (%)	enrichment factor	time cost (hours)
SB-VS	1530	182	89.66	11.90	69.19	0.017
PB-VS	43 790	173	85.22	0.40	2.30	6
DB-VS	24 152	163	80.30	0.67	3.89	2240.5
SB/PB	989 (1530/989)	172 (182/172)	84.73	17.39	101.16	0.100
SB/PB/DB	377 (1530/989/377)	158 (182/172/158)	77.83	41.91	243.78	19

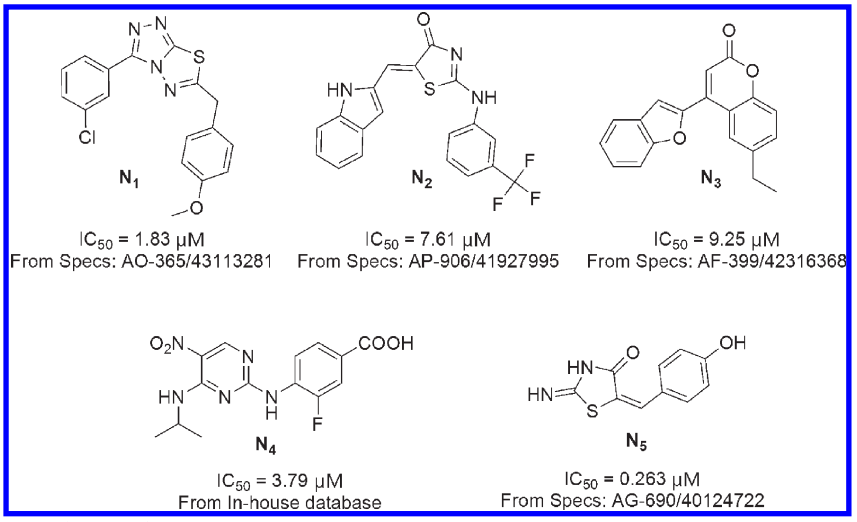


Figure 4. Four Pim-1 inhibitors (**N1**−**N4**) with new scaffolds and the most active compound **N5** obtained by using the hierarchical multistage VS approach together with their measured bioactivities.

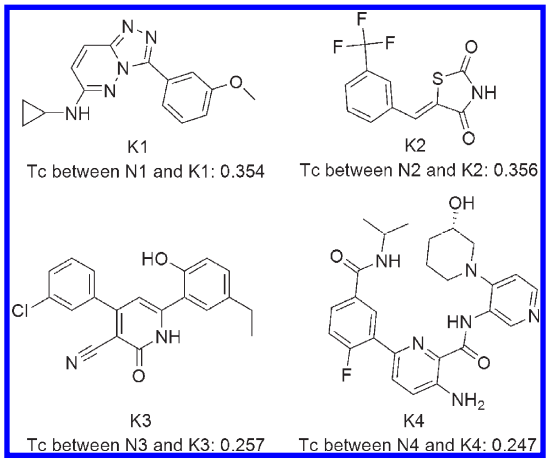


Figure 5. The known Pim-1 inhibitors (**K1**−**K4**) that have the biggest Tanimoto coefficients with **N1**−**N4**, respectively.

The final hit compounds obtained by the hierarchical multistage VS approach were further visually inspected to check whether some important interactions with the active site of Pim-1 kinase could be kept, for example, interactions with residues in the hinge region and other residues in the active site of Pim-1 kinase, including Lys67, Glu89, Asp186, and Asn172. A total of 47 compounds were finally selected for the subsequent in vitro kinase assay.

In Vitro Pim-1 Kinase Inhibitory Assay. An in vitro Pim-1 kinase inhibitory assay was carried out for the selected 47 compounds. In the first round, the inhibition potency was tested with the compound at a fixed concentration of 10 μM. In the second round, we selected those compounds that afford a higher inhibition rate at 10 μM and that have a new scaffold to perform IC_{50} profiling. Fifteen compounds show nanomolar level or low micromolar inhibitory potency or an inhibition rate greater than 50% at 10 μM (see Table S4 in Supporting Information). Of the 15 compounds, 4 compounds were found to have new scaffolds (**N1**−**N4**, see Figure 4); compounds with these scaffolds have never been reported as Pim-1 inhibitors. These scaffolds are [1,2,4]triazolo[3,4-b] [1,3,4]thiadiazole for **N1**, thiazol-4(5H)-one for **N2**, 2H-chromen-2-one for **N3**, and *N*-phenylpyrimidin-2-amine for **N4**. The IC_{50} values are 1.83, 7.61, 9.25, and 3.79 μM for **N1**−**N4**, respectively (see Figure 4). In order to further check the novelty of **N1**−**N4**, the Tanimoto coefficients between the four compounds and all the known Pim-1 inhibitors were calculated through a similarity analysis based on the extended connectivity fingerprints (ECFPs).[42] Figure 5 presents the known Pim-1 inhibitors (**K1**−**K4**) who have the largest Tanimoto coefficients with **N1**−**N4**, respectively; the Tanimoto coefficients of **N1**−**K1**, **N2**−**K2**, **N3**−**K3**, and **N4**−**K4** are all less than 0.4. The differences of chemical structures between **Ni** and K_i (i = 1, 2, 3, 4) are indeed large (see Figure 4 and 5). The most active compound against Pim-1 obtained in this screening is also given in Figure 4 (compound **N5**), whose IC_{50} is 0.263 μM. **N5** is very similar with one of the training set compounds (compound **2**, see Figure 1). The dose−response curves of compounds **N1**−**N5** are presented in Figure 6.

Figure 7A and B, as examples, illustrates the possible interaction modes of **N1** and **N2** with the kinase domain of Pim-1, respectively. For comparison, compound **6** that has been complexed with Pim-1 (PDB entry 3BGQ) is also shown. An overview is that the binding poses and interactions of **N1** and **N2** with the kinase domain of Pim-1 are very similar with that of compound **6**. From Figure 7A, we can see that the triazole ring of compound **N1** forms a hydrogen (H)-bond with Lys67. The anisole group is directed toward the solvent accessible region. And the chlorophenyl moiety forms hydrophobic interactions with Pro123, Val 126, and Leu174. From Figure 7B, we can see that the thiazol-4(5H)-one motif of **N2** forms two hyrogen bonds with Lys67. Again the *p*-trifluoromethylphenyl group is directed toward the solvent accessible region. The indole motif forms an hydrogen bond with Glu121. The good binding poses and interactions of **N1** and **N2** with the active site of Pim-1 provide a solid basis for their bioactivity against Pim-1.

An Analysis of Necessity for the Use of the Hierarchical Multistage VS Approach. Here one may argue the necessity of using such an elaborate hierarchical multistage VS approach in this study. As we have mentioned before, each of the three VS methods, namely SB-VS, PB-VS, and DB-VS, has its own advantages and disadvantages. The most obvious difference among them is the speed of virtual screening; SB-VS is the

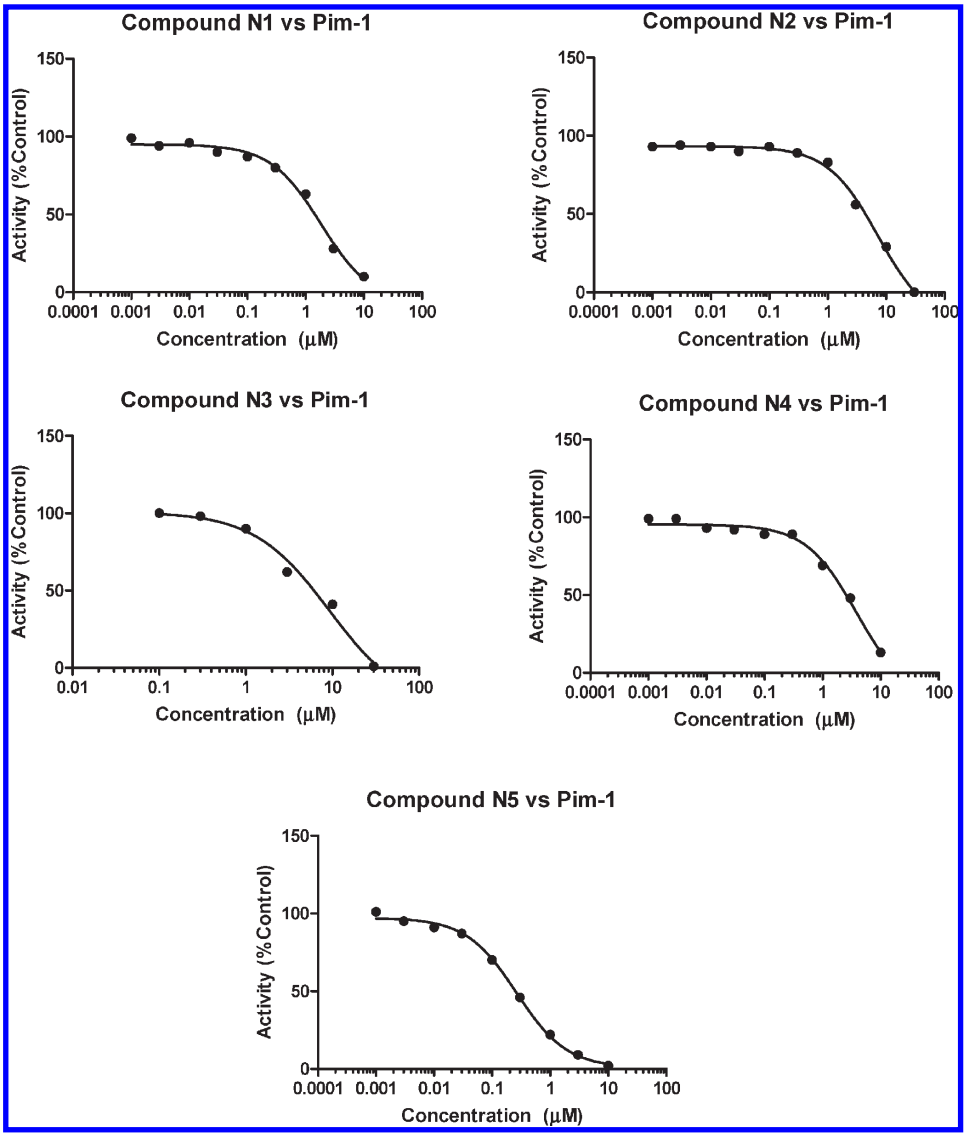


Figure 6. Dose−response curves of **N1**−**N5**.

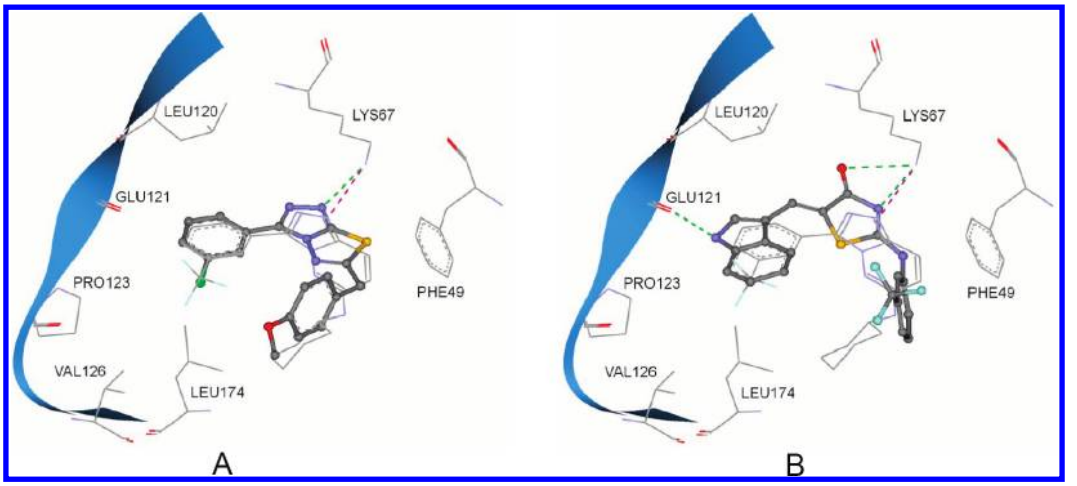


Figure 7. Binding modes of compound **N1** (A) and **N2** (B) in the active site of Pim-1. Compound **6** complexed with Pim-1 (PDB entry 3BGQ) is also shown for comparison (in line form). Dotted lines represent the hydrogen-bonding interactions between ligands and Pim-1.

dx.doi.org/10.1021/ci100464b |*J. Chem. Inf. Model.* 2011, 51, 1364–1375

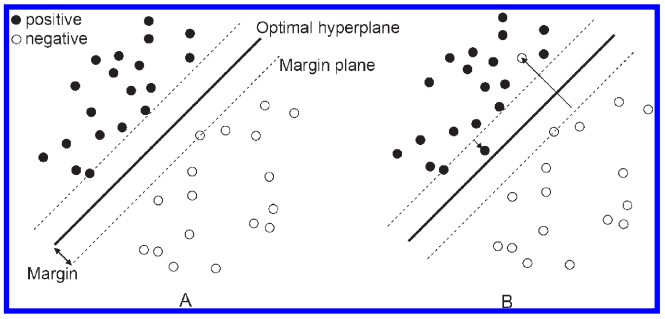


Figure 8. A schematic hyperplane of SVM separating objects represented as filled (positive) and unfilled (negative) circles into two different classes with a maximum margin. (A) The linearly separable case in which the two classes of objects can be perfectly separated by an optimal hyperplane. (B) The linearly nonseparable case in which no any hyperplane to be able to perfectly separate two classes of objects.

fastest one, followed by PB-VS, and DB-VS is the slowest one. In the hierarchical multistage VS approach, they are arranged in an increasing order so that the fastest filter is carried out first, and the slowest one is executed in the end. This arrangement allows an efficient screening of large chemical databases, like PubChem that contains 18 831 686 compounds, in a short time scale. According to our current computational power, screening a large chemical library, like PubChem, by DB-VS is impractical. On the other hand, the combination of SB-VS, PB-VS, and DB-VS allows them to mutually compensate for their limitations and hence increasing the hit rate and the enrichment factor (as mentioned above). The test results of the hierarchical multistage VS approach through screening the M-MDDR library showed that the hit rate and the enrichment factor reached 41.91% and 243.78, respectively, which are much larger than those of individual VS methods. In our practical screening of Pim-1 inhibitors against PubChem, Specs, Enamine, and an in-house database, a higher hit rate (31.91%, 15 actives out of 47 selected compounds) was also achieved. Finally, we have to mention that a simple similarity search is not comparable to the hierarchical multistage VS approach in at least the following aspects: (1) A simple similarity search is often difficult to find compounds with novel scaffolds, for example, in our case the compounds **N1**−**N4** that have new scaffolds cannot be located by a simple similarity search; (2) the hit rate and enrichment factor of a simple similarity search is usually lower, for example, and several screening tests against PubChem, Specs, Enamine and an in-house database by using the similarity search based on ECFPs, which were made by us, gave hit rates less than 5% and enrichment factors less than 10.

MATERIALS AND METHODS

SVM Modeling. *SVM.* SVM is a supervised machine learning method based on the structural risk minimization principle. Many studies have demonstrated that SVM is one of the best methods for classification modeling.[43,44] In this investigation, an optimal SVM modeling method, namely SVM method combined with GA for feature selection and CG for parameter optimization (GA-CG-SVM), which was proposed by our group recently,[35] was used. Detailed description of the proposed GA-CG-SVM method can be found in refs.[35,45] Here we just make a short summary to the basic idea of SVM and GA-CG-SVM.

In SVM, each object is described by a vector x_i of N real numbers (features or descriptors), which corresponds to a point in an N-dimensional space. The objects in the first class (positive) are each assigned a value of $y_i = +1$, the ones in the second class are $y_i = -1$. In linearly separable cases (see Figure 8, left), the objects can be correctly classified by

$$w \cdot x_i + b \geq +1, \qquad \text{for } y_i = +1 (\text{class1}) \tag{1}$$

$$w \cdot x_i + b \geq -1, \qquad \text{for } y_i = -1 (\text{class2}) \tag{2}$$

where w is a vector normal to the hyperplane, b is a scalar quantity.

The SVM attempts to find an optimal separating hyperplane with the maximum margin by solving the following optimization problem:

$$\max_{w,b} \frac{2}{||w||} \qquad \text{subject to } y_i(w \cdot x_i + b) - 1 \geq 0 \tag{3}$$

where $2/(||w||)$ is the margin.

Equation 3 can be solved by Lagrange multipliers method. Finally a classifying determination function is obtained as follows:

$$f(x) = \text{sgn}(w \cdot x + b) = \text{sgn}[\sum_{i=1}^{n} \alpha_i y_i (x \cdot x_i) + b] \tag{4}$$

The above concepts can also be extended to the linearly nonseparable cases (see Figure 8, right), in which no hyperplane can be used to perfectly separate two sets of points. In this case, we can introduce non-negative slack variables $\xi_i \geq 0$, $i = 1, ..., m$. Such that

$$w \cdot x_i + b \geq +1 - \xi_i, \qquad \text{for } y_i = +1 \tag{5}$$

$$w \cdot x_i + b \geq -1 + \xi_i, \qquad \text{for } y_i = -1 \tag{6}$$

The purpose here is to find a hyperplane that provides the minimum number of training errors, which means to keep the constraint violation as small as possible. The equation to be solved becomes:

$$\max_{w,b} \frac{2}{||w||} + C \sum_{i=1}^{m} \xi_i \qquad \text{subject to } y_i(w \cdot x_i + b) - 1 + \xi_i \geq 0 \tag{7}$$

where C is the penalty parameter, which should be predetermined by the user. The parameter C has important impact on the accuracy of the SVM classifier, thus it should be chosen carefully. Similar to the linearly separable cases, the eq 7 can also be solved by Lagrangian multipliers method.

The nonlinear (non)separable cases could be easily transferred to linear cases through projecting the input variable into a new high-dimensional feature space by using a kernel function $K(x_i,x_j)$. Several kernel functions including polynomial, radial basis function (RBF), and sigmoid kernel have been suggested. However, the RBF is the most widely used kernel function, and it performed very well in most cases. The RBF kernel function will also be used in this study:

$$k(x_i, x_j) = \exp(-\gamma \ || \ x_i - x_j \ || \ 2) \tag{8}$$

where γ is a parameter which should be specified by the user in advance.

Obviously, optimal parameters C and γ have significant influence on the quality of SVM model. Further, it has been shown that the feature selection and the parameter setting influence each other in SVM modeling, which means that they should be dealt with simultaneously. Thus, an integrated scheme for the simultaneous treatment of both feature selection and parameter optimization was proposed by us. In this scheme, a GA-based method is used for the feature selection and a conjugate gradient (CG) method for the parameter optimization. It is expected that the use of the integrated scheme can help to develop optimal prediction models of specific properties with fewer descriptors but higher accuracy.

Data Sets. A total of 846 compounds including 799 positives (Pim-1 inhibitors, $IC_{50} \leq 10$ μM) and 47 negatives (Pim-1 noninhibitors, $IC_{50} > 10$ μM) were collected from different literature resources.[15−18,28−34] The 799 positive compounds were first grouped into different categories based on their scaffolds. Then we chose compounds from each category randomly but approximately in proportion to construct the training, test, and independent validation sets, which could help to guarantee the three data sets having a relatively equal distribution in the chemical structural space. The formed training set for the SVM modeling contains 500 positives. The test set for the validation of SVM model comprises 96 positives and the collected 47 negatives. The remaining 203 positive compounds together with the 117 877 decoys obtained from MDDR (the selection method of decoys see below) were used for the evaluation of performance of the proposed hierarchical multistage VS approach (detailed information is available in the Supporting Information, see Table S1−S3).

In the SVM modeling, a sufficient number of negative compounds in the training set have been thought to be very important for the quality of SVM model. However, negative compounds that were definitely reported are very limited. Thus we adopted a method suggested by Chen et al to generate putative negatives;[46] this method can create putative negatives without requiring the knowledge of actual negatives. For the putative negatives generation, 18.83 M compounds from the PubChem were clustered into 8913 compound families by using K-means clustering based on their molecular descriptors, which were calculated by Discovery Studio 2.55 software. And 37 057 putative negatives were randomly selected from families that do not contain any of the known Pim-1 inhibitors. These compounds together with 500 known positives constitute the training set for the SVM model training.

In the assessment of VS methods, two test sets, namely M-CMC and M-MDDR, were constructed. The M-CMC contains 588 known Pim-1 inhibitors and 6978 decoys from CMC library. The M-MDDR comprises 203 known Pim-1 inhibitors and 117 877 decoys from MDDR. The decoys were selected such that they have similar physical properties with the annotated Pim-1 inhibitors, so that enrichment is not simply a separation of trivial physical features. But they are chemically distinct from the annotated Pim-1 inhibitors so that they might not bind to Pim-1 active site.[47,48] Here a very similar strategy as that used by Shoichet and Irwin[49] was adopted to construct the decoy sets from CMC and MDDR. First, the known 799 Pim-1 inhibitors were seeded among 8720 CMC or 164 365 MDDR compounds. Compounds with Tanimoto coefficient less than 0.6 to any annotated Pim-1 inhibitors were selected; the Tanimoto coefficient was calculated through a similarity analysis based on the ECFPs.[42] Second, four physical properties, including molecular weight, number of hydrogen-bond donors and acceptors, number of rotatable bonds, and log P, were calculated for the annotated Pim-1 inhibitors and the selected compounds from the previous step. Compounds were chosen if their calculated physical properties were similar to those of any of the annotated Pim-1 inhibitors, otherwise they were ignored. Finally 6978 compounds from CMC library and 117 877 from MDDR library were selected as the decoys to construct the test sets M-CMC and M-MDDR, respectively.

Details of SVM Modeling. All the SVM modeling calculations were carried out by using GA-CG-SVM program.[35] The initial descriptors used in this study were calculated by Accelrys Discovery Studio 2.55 program package. The termination criterion is that the generation number reaches 200 or that the fitness value does not improve during the last 10 generations. The crossover rate was set to 0.5, and the mutation rate was 0.1. The starting values of the parameters (C, γ) were set to (8192, 0.38) in the program.

Pharmacophore Modeling. The HipHop algorithm implemented in Accelrys Discovery Studio 2.55 program package was employed for the pharmacophore modeling. Eight Pim-1 inhibitors (Compound **1**−**8**, see Figure 1) were chosen to form the training set;[15,18,28,30−32,34,36] these compounds were chosen since they are all the most active compounds and bear structural diversity. Compounds **1** (SGI-1776, IC_{50} = 7.4 nM) and **6** (K_i = 11 nM) (see Figure 1) were selected as 'reference compounds' specifying a 'principal' value of 2 and a 'MaxOmitFeat' value of 0. The 'principal' and 'MaxOmitFeat' values were set to 1 for the remaining 6 compounds. The 'minimum Interfeature Distance' value was set to 1.0 Å. All the rest HipHop parameters were kept at their default values.

Molecular Docking Study. All the molecular docking studies were carried out by Genetic Optimization of Ligand Docking (GOLD) 4.0.[38] GOLD adopts the genetic algorithm to dock flexible ligands into protein binding sites. The crystal structure (PDB entry 3BGQ) of the kinase domain of Pim-1 bound to the inhibitor compound **6** was used for the docking studies. Hydrogen atoms were added to the protein by using Accelrys Discovery Studio 2.55. The Charmm force field was assigned. The binding site was defined as a sphere containing the residues that stay within 10 Å from the ligand, which is large enough to cover the ATP binding region at the active site.

General Chemistry. Unless otherwise noted, reagents and solvents were obtained from commercial suppliers and were used without further purification. According to the information provided by the suppliers, all the purchased compounds were purified by flash chromatography and/or crystallization; the purity of these compounds is >98%. The compounds from our in-house database were also purified by flash chromatography and/or crystallization, with purity larger than 98%.

In Vitro Pim-1 Kinase Inhibitory Assays. Pim-1 kinase inhibitory assays were performed using labeled γ-^{33}P-ATP, and the incorporation of labeled phosphate onto the peptide KKRNRTLTV substrate was monitored. The Pim-1 kinase buffer used contains: 20 mM MOPS, 1 mM EDTA, 0.01% Brij-35, 5% Glycerol, 0.1% β-mercaptoethanol, and 1 mg/mL BSA. All assays were carried out robotically at room temperature (21 °C). In a final reaction volume of 25 μL, Pim-1 (h) (5−10 mU) is incubated with 8 mM MOPS pH 7.0, 0.2 mM EDTA, 100 μM KKRNRTLTV, 10 mM MgAcetate and γ-^{33}P-ATP (specific activity approximately 500 cpm/pmol, concentration as required). The reaction is initiated by the addition of the MgATP

mix. After incubation for 40 min at room temperature, the reaction is stopped by the addition of 5 μL of a 3% phosphoric acid solution. 10 μL of the reaction is then spotted onto a P30 filtermat and washed three times for 5 min in 75 mM phosphoric acid and once in methanol prior to drying and scintillation counting.

Prior to the Pim-1 kinase profiling, ATP competition assays were carried out with an initial 30 min preincubation of test compounds and protein. It was found that ATP could attenuate inhibition in a concentration dependent manner, suggesting reversible ATP competitive inhibitors. Due to the ATP concentration dependence of inhibitory potency, kinase profiling was performed at a kinase-specific ATP concentration, namely ATP Km (90 μM). Staurosporine was used as the positive control. The positive control wells contain all components of the reaction with the positive control (Staurosporine) replacing the test compound. DMSO is included in all wells to control for solvent effects. The blank wells contain all components of the reaction but not any test compound. Following subtraction of the blank well counts, the values of kinase activity remaining are expressed as a percentage of the positive control (i.e., % kinase activity remaining). The inhibition profile of the test compounds was expressed as the percentage of the residual kinase activity for an inhibitor concentration of 10 μM. The IC_{50} values of inhibitors were determined from dose−response curves after carrying out assays at 10 different concentrations for each compound. All the assays were replicated twice, and the means of the replicates were calculated.

CONCLUSIONS

In this account, the proposed hierarchical multistage VS approach, which is based on SB-VS, PB-VS, and DB-VS, was employed for the discovery of novel potent Pim-1 inhibitors. In this approach, the three VS methods are applied in an increasing order of complexity so that the first filter (SB-VS) is fast and simple, while successive ones (PB-VS and DB-VS) are more time-consuming but are applied only to a small subset of the entire database. The models used for the SB-VS and PB-VS were first established and comprehensively validated. And docking parameters and scoring functions for the DB-VS were also optimized in advance. Evaluation of this hierarchical multistage VS approach indicates that it drastically outperforms the solely used SB-VS, PB-VS, and DB-VS as well as the combined SB-VS and PB-VS in terms of the hit rate and the enrichment factor. This approach was then applied to screen several large chemical libraries including PubChem (18 831 686 compounds), Specs (202 408 compounds), and Enamine (980 000 compounds) as well as an in-house database (445 compounds). From the final hit compounds, 47 compounds were selected for further in vitro Pim-1 inhibitory assay, and 15 compounds show nanomolar level or low micromolar inhibition potency against Pim-1. Remarkably, 4 of the 15 compounds were found to have new scaffolds which have potentials for the later chemical development.

To sum up, this investigation is the first report that combines the SVM model-based VS with more classical pharmacophore-based VS and molecular docking; this approach allows accurate and efficient screening of large chemical databases in a short time scale. The high hit rate and enrichment factor and high speed as well as the discovery of new Pim-1 inhibitors particularly with new scaffolds demonstrate the validity of the hierarchical multistage VS approach. It is expected that this type of hierarchical multistage VS strategy could play more important roles in drug discovery especially at the current situation that each single VS method is far from perfect.

ASSOCIATED CONTENT

Supporting Information. The structures of the 799 Pim-1 inhibitors and 47 noninhibitors with the corresponding activities against Pim-1 (Tables S1−S3), the 15 hits together with their bioactivities (Table S4) as well as 7 compounds complexed with Pim-1 kinase (Figure S1) that were used for the docking parameter optimization. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: yangsy@scu.edu.cn. Telephone: 86-28-85164063.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (20872100) and SRFDP (20100181110025).

REFERENCES

(1) Cuypers, H. T.; Selten, G.; Quint, W.; Zijlstra, M.; Maandag, E. R.; Boelens, W.; van Wezenbeek, P.; Melief, C.; Berns, A. Murine leukemia virus-induced T-cell lymphomagenesis: integration of proviruses in a distinct chromosomal region. *Cell* **1984**, *37*, 141–150.

(2) Wang, Z.; Bhattacharya, N.; Weaver, M.; Petersen, K.; Meyer, M.; Gapter, L.; Magnuson, N. S. Pim-1: a serine/threonine kinase with a role in cell survival, proliferation, differentiation and tumorigenesis. *J. Vet. Sci.* **2001**, *2*, 167–179.

(3) Bachmann, M.; Möröy, T. The serine/threonine kinase Pim-1. *Int. J. Biochem. Cell Biol.* **2005**, *37*, 726–730.

(4) Brault, L.; Gasser, C.; Bracher, F.; Huber, K.; Knapp, S.; Schwaller, J. PIM serine/threonine kinases in the pathogenesis and therapy of hematologic malignancies and solid cancers. *Haematologica* **2010**, *95*, 1004–1015.

(5) Mochizuki, T.; Kitanaka, C.; Noguchi, K.; Muramatsu, T.; Asai, A.; Kuchino, Y. Physical and functional interactions between Pim-1 kinase and Cdc25A phosphatase - Implications for the Pim-1-mediated activation of the c-Myc signaling pathway. *J. Biol. Chem.* **1999**, *274*, 18659–18666.

(6) Wang, Z.; Bhattacharya, N.; Mixter, P. F.; Wei, W.; Sedivy, J.; Magnuson, N. S. Phosphorylation of the cell cycle inhibitor p21Cip1/WAF1 by Pim-1 kinase. *BBA-Molecular Cell Research* **2002**, *1593*, 45–55.

(7) Bachmann, M.; Kosan, C.; Xing, P. X.; Montenarh, M.; Hoffmann, I.; Moroy, T. The oncogenic serine/threonine kinase Pim-1 directly phosphorylates and activates the G2/M specific phosphatase Cdc25C. *Int. J. Biochem. Cell Biol.* **2006**, *38*, 430–443.

(8) Aho, T. L. T.; Sandholm, J.; Peltola, K. J.; Mankonen, H. P.; Lilly, M.; Koskinen, P. J. Pim-1 kinase promotes inactivation of the pro-apoptotic Bad protein by phosphorylating it on the Ser112 gatekeeper site. *FEBS Lett.* **2004**, *571*, 43–49.

(9) Amson, R.; Sigaux, F.; Przedborski, S.; Flandrin, G.; Givol, D.; Telerman, A. The human protooncogene product p33pim is expressed during fetal hematopoiesis and in diverse leukemias. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8857–8861.

(10) Valdman, A.; Fang, X. L.; Pang, S. T.; Ekman, P.; Egevad, L. In *PIM-1 expression in prostatic intraepithelial neoplasia and human prostate cancer*. Proceedings of the 99th Annual Meeting of the American Urological Association, San Francisco, CA, May 08−13, 2004; p 110.

(11) Fox, C. J.; Hammerman, P. S.; Thompson, C. B. The Pim kinases control rapamycin-resistant T cell survival and activation. *J. Exp. Med.* **2005**, *201*, 259–266.

(12) Xie, Y.; Xu, K.; Linn, D. E.; Yang, X.; Guo, Z.; Shimelis, H.; Nakanishi, T.; Ross, D. D.; Chen, H.; Fazli, L.; Gleave, M. E.; Qiu, Y. The 44-kDa pim-1 kinase phosphorylates BCRP/ABCG2 and thereby promotes its multimerization and drug-resistant activity in human prostate cancer cells. *J. Biol. Chem.* **2008**, *283*, 3349–3356.

(13) Zemskova, M.; Sahakian, E.; Bashkirova, S.; Lilly, M. The PIM1 kinase is a critical component of a survival pathway activated by docetaxel and promotes survival of docetaxel-treated prostate cancer cells. *J. Biol. Chem.* **2008**, *283*, 20635–20644.

(14) Bregman, H.; Meggers, E. Ruthenium half-sandwich complexes as protein kinase inhibitors: an N-succinimidyl ester for rapid derivatizations of the cyclopentadienyl moiety. *Org. Lett.* **2006**, *8*, 5465–5468.

(15) Bearss, D.; Liu, X.; Vankayalapati, H.; Xu, Y. Imidazo [1, 2-beta] pyridazine and pyrazolo [1,5-α] pyrimidine derivatives and their use as protein kinase inhibitors. U.S. Patent 2008/0261988 A1, 2008.

(16) Pogacic, V.; Bullock, A. N.; Fedorov, O.; Filippakopoulos, P.; Gasser, C.; Biondi, A.; Meyer-Monard, S.; Knapp, S.; Schwaller, J. Structural analysis identifies imidazo[1,2-b]Pyridazines as PIM kinase inhibitors with In vitro antileukemic activity. *Cancer Res.* **2007**, *67*, 6916–6924.

(17) Pierce, A.; Jacobs, M.; Stuver-Moody, C. Docking study yields four novel inhibitors of the protooncogene Pim-1 kinase. *J. Med. Chem.* **2008**, *51*, 1972–1975.

(18) Grey, R.; Pierce, A. C.; Bemis, G. W.; Jacobs, M. D.; Moody, C. S.; Jajoo, R.; Mohal, N.; Green, J. Structure-based design of 3-aryl-6-amino-triazolo[4,3-b] pyridazine inhibitors of Pim-1 kinase. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3019–3022.

(19) Akué-Gédu, R.; Rossignol, E.; Azzaro, S. p.; Knapp, S.; Filippakopoulos, P.; Bullock, A. N.; Bain, J.; Cohen, P.; Prudhomme, M.; Anizon, F.; Moreau, P. Synthesis, kinase inhibitory potencies, and in vitro antiproliferative evaluation of new Pim kinase inhibitors. *J. Med. Chem.* **2009**, *52*, 6369–6381.

(20) Tao, Z.-F.; Hasvold, L. A.; Leverson, J. D.; Han, E. K.; Guan, R.; Johnson, E. F.; Stoll, V. S.; Stewart, K. D.; Stamper, G.; Soni, N.; Bouska, J. J.; Luo, Y.; Sowin, T. J.; Lin, N.-H.; Giranda, V. S.; Rosenberg, S. H.; Penning, T. D. Discovery of 3H-Benzo[4,5]thieno-[3,2-d]pyrimidin-4-ones as potent, highly selective, and orally bioavailable inhibitors of the human protooncogene proviral insertion site in moloney murine leukemia virus (PIM) kinases. *J. Med. Chem.* **2009**, *52*, 6621–6636.

(21) Sliman, F.; Blairvacq, M.; Durieu, E.; Meijer, L.; Rodrigo, J.; Desmaele, D. Identification and structure-activity relationship of 8-hydroxy-quinoline-7 -carboxylic acid derivatives as inhibitors of Pim-1 kinase. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2801–2805.

(22) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.

(23) Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* **2010**, *53*, 539–558.

(24) Jorissen, R. N.; Gilson, M. K. Virtual screening of molecular databases using a support vector machine. *J. Chem. Inf. Model.* **2005**, *45*, 549–561.

(25) Liew, C.; Ma, X.; Liu, X.; Yap, C. SVM model for virtual screening of Lck inhibitors. *J. Chem. Inf. Model.* **2009**, *49*, 877–885.

(26) Liu, X.; Ma, X.; Tan, C.; Jiang, Y.; Go, M.; Low, B.; Chen, Y. Virtual screening of Abl inhibitors from large compound libraries by support vector machines. *J. Chem. Inf. Model.* **2009**, *49*, 2101–2110.

(27) Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today* **2010**, *15*, 444–450.

(28) Cheney, I. W.; Yan, S.; Appleby, T.; Walker, H.; Vo, T.; Yao, N.; Hamatake, R.; Hong, Z.; Wu, J. Z. Identification and structure-activity relationships of substituted pyridones as inhibitors of Pim-1 kinase. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1679–1683.

(29) Holder, S.; Zemskova, M.; Zhang, C.; Tabrizizad, M.; Bremer, R.; Neidigh, J. W.; Lilly, M. B. Characterization of a potent and selective small-molecule inhibitor of the PIM1 kinase. *Mol. Cancer Ther.* **2007**, *6*, 163–172.

(30) Kearney, P.; Brown, S.; Koltun, E. Aminopyrimidine, Aminopyridine and Aniline Derivatives Inhibitors of Pim-1 and/or Pim-3. WO Patent WO/2007/044,724, 2007.

(31) Burger, M.; Lindvall, M.; Han, W.; Lan, J.; Nishiguchi, G.; Shafer, C.; Bellamacina, C.; Huh, K.; Atallah, G.; Mcbride, C.; Pim kinase inhibitors and methods of their use. In WO Patent WO/2008/106,692: 2008.

(32) Tong, Y.; Stewart, K. D.; Thomas, S.; Przytulinska, M.; Johnson, E. F.; Klinghofer, V.; Leverson, J.; McCall, O.; Soni, N. B.; Luo, Y.; Lin, N.-h.; Sowin, T. J.; Giranda, V. L.; Penning, T. D. Isoxazolo [3,4-b]quinoline-3,4(1H,9H)-diones as unique, potent and selective inhibitors for Pim-1 and Pim-2 kinases: Chemistry, biological activities, and molecular modeling. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5206–5208.

(33) Olla, S.; Manetti, F.; Crespan, E.; Maga, G.; Angelucci, A.; Schenone, S.; Bologna, M.; Botta, M. Indolyl-pyrrolone as a new scaffold for Pim1 inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 1512–1516.

(34) Xia, Z. P.; Knaak, C.; Ma, J.; Beharry, Z. M.; McInnes, C.; Wang, W. X.; Kraft, A. S.; Smith, C. D. Synthesis and evaluation of novel inhibitors of Pim-1 and Pim-2 protein kinases. *J. Med. Chem.* **2009**, *52*, 74–86.

(35) Yang, S. Y.; Huang, Q.; Li, L. L.; Ma, C. Y.; Zhang, H.; Bai, R.; Teng, Q. Z.; Xiang, M. L.; Wei, Y. Q. An integrated scheme for feature selection and parameter setting in the support vector machine modeling and its application to the prediction of pharmacokinetic properties of drugs. *Artif. Intell. Med.* **2009**, *46*, 155–163.

(36) Bullock, A. N.; Debreczeni, J. E.; Fedorov, O. Y.; Nelson, A.; Marsden, B. D.; Knapp, S. Structural Basis of Inhibitor Specificity of the Human Protooncogene Proviral Insertion Site in Moloney Murine Leukemia Virus (PIM-1) Kinase. *J. Med. Chem.* **2005**, *48*, 7604–7614.

(37) Redkar, S.; Liu, X. H.; Gourley, E.; Lamb, J.; Grand, G.; Lloyd, M.; Wolfe, B.; Bearss, D.; Vankayalapati, H. In *Discovery of SGI-1776, a potent and selective PIM-1 kinase inhibitor*. Proceedings of the 100th Annual Meeting of the American Association for Cancer Research, San Diego, CA, April 18−22, 2009; Abstract no. 2013.

(38) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

(39) Li, X.; Li, Y.; Cheng, T.; Liu, Z.; Wang, R. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *J. Comput. Chem.* **2010**, *31*, 2109–2125.

(40) Qian, K.; Wang, L.; Cywin, C. L.; Farmer, B. T.; Hickey, E.; Homon, C.; Jakes, S.; Kashem, M. A.; Lee, G.; Leonard, S.; Li, J.; Magboo, R.; Mao, W.; Pack, E.; Peng, C.; Prokopowicz, A.; Welzel, M.; Wolak, J.; Morwick, T. Hit to lead account of the discovery of a new class of inhibitors of Pim kinases and crystallographic studies revealing an unusual kinase binding mode. *J. Med. Chem.* **2009**, *52*, 1814–1827.

(41) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793–806.

(42) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.

(43) Byvatov, E.; Schneider, G. Support vector machine applications in bioinformatics. *Appl. Bioinformatics* **2003**, *2*, 67–77.

(44) Yang, Z. R. Biological applications of support vector machines. *Briefings Bioinf.* **2004**, *5*, 328–338.

(45) Zhang, H.; Xiang, M. L.; Ma, C. Y.; Huang, Q.; Li, W.; Xie, Y.; Wei, Y. Q.; Yang, S. Y. Three-class classification models of logS and logP derived by using GA-CG-SVM approach. *Mol. Diversity* **2009**, *13*, 261–268.

(46) Han, L. Y.; Ma, X. H.; Lin, H. H.; Jia, J.; Zhu, F.; Xue, Y.; Li, Z. R.; Cao, Z. W.; Ji, Z. L.; Chen, Y. Z. A support vector machines approach for virtual screening of active compounds of single and

multiple mechanisms from large libraries at an improved hit-rate and enrichment factor. *J. Mol. Graphics Modell.* **2008**, *26*, 1276–1286.

(47) Irwin, J. J. Community benchmarks for virtual screening. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 193–199.

(48) Pham, T. A.; Jain, A. N. Parameter Estimation for Scoring Protein-Ligand Interactions Using Negative Training Data. *J. Med. Chem.* **2006**, *49*, 5856–5868.

(49) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.