Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 1371–1379

Identification of phosphodiesterase-1 and 5 dual inhibitors by a ligand-based virtual screening optimized for lead evolution

Kazuto Yamazaki,* Naoto Kusunose, Katsuya Fujita, Hideshi Sato, Shigehiro Asano, Akihito Dan and Masaharu Kanaoka

Sumitomo Pharmaceuticals Co., Ltd, 1-98, Kasugade Naka 3-Chome, Konohana-ku, Osaka 554-0022, Japan

Received 11 August 2005; revised 31 October 2005; accepted 14 November 2005
Available online 6 December 2005

Abstract—We identified new lead candidates which showed potent dual inhibition against phosphodiesterase-1 and 5 by a ligand-based virtual screening optimized for lead evolution. This virtual screening method, consisting of classification and regression tree analysis using 168 2-center pharmacophore descriptors and 12 macroscopic descriptors, demonstrated a high predictive ability for bioactivity of new chemical compounds. The obtained lead candidates were structurally diverse, although only the structure–activity relationship data of hydroxamic acid derivatives were used to configure the prediction model for the virtual screening.
© 2005 Elsevier Ltd. All rights reserved.

cGMP is one of the second messengers of intra-cellular signal transduction and mediates various cell functions. The signal transduction is triggered by phosphorylation of some target proteins catalyzed by cGMP-dependent protein kinase. Phosphodiesterase types 1 (PDE-1) and 5 (PDE-5) are the major cGMP hydrolyzing enzymes in blood vessel[1] and regulate the level of the mediator in concert with guanylyl cyclase which catalyzes the synthesis of cGMP from GTP. PDE-1 and 5 have also been found in platelet.[2] Therefore, inhibitors of PDE-1 and 5 are expected as therapeutics for cardiovascular diseases, such as hypertension, angina, cardiac failure, and obstructive arteriosclerosis.

SCH51866 is known as a dual inhibitor of PDE-1 and 5 (Fig. 1A).[3] We also identified hydroxamic acid derivatives (Fig. 1B) as potent dual inhibitors. These dual inhibitors show cGMP-dependent vasodilatory effects in an isolated rat blood vessel.[4] Although the compounds had a promising therapeutic effect for cardiovascular symptoms that related to cGMP level, subsequent attempt of lead optimization was not successful, because bioavailability of these compounds was too low for an oral drug candidate. We started exploratory studies for potent dual inhibitors of PDE-1 and 5 that would be structurally different from the hydroxamic acid derivatives.

Keywords: Phosphodiesterase; PDE-1; PDE-5; Dual inhibitors; Virtual screening; CART; Pharmacophore.
* Corresponding author. E-mail: kazuto-yamazaki@ds-pharma.co.jp

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.11.046

For lead discovery, usually a core substructure or scaffold is identified and followingly optimum substituents. In general, success rate of finding the former is lower than the latter,[5] hence virtual screening is expected to increase the success rate through identification of multiple and diverse lead candidates. We deviced a virtual screening method optimized for lead evolution, that is, discovery of a new lead candidate with a novel scaffold.

In this paper, we report identification of new lead candidates with potent dual inhibitions against PDE-1 and 5. The virtual screening method used in this study demonstrated remarkable ability to distinguish structurally diverse active compounds from inactive ones.

Computational strategies. We used classification and regression tree (CART) analysis to configure a prediction model for virtual screening. An optimum set of structural descriptors were selected as explanatory variables for CART analysis.

CART is a method with the most active tree pruning among decision tree analyses.[6] Pruning is carried out to prevent overlearning for training data. In recent years, not only structure–activity relationship analysis using CART methods but also structure–property relationships,[7] such as CYP inhibition[8] and oral absorption,[9] were reported. Mazzatorta et al. studied structure–toxicity relationships of 235 pesticides with seven classification algorithms, containing CART method.[10] They concluded

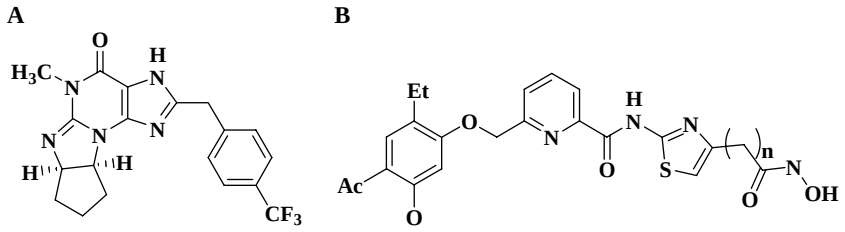


Figure 1. Dual PDE-1 and 5 inhibitors (A) SCH51866 and (B) representative compounds of the hydroxamic acid analogs.

that CART was the best among the algorithms they tested. For virtual screening, prediction of unknown data should be more emphasized than explaining training data. We expected that the classification ability of the CART method would give a good prediction in virtual screening.

In the CART analysis, a prediction model is configured with selected explanatory variables in the learning step. The selection of explanatory variables is based on discrimination indices of the training set. However, in many cases the number of active compounds is inadequate for the number of explanatory variables. In such a situation, an intrinsic variable for biological activity should be selected for prediction model or the model is pruned needlessly. Thus, it is best for CART analysis to use the smallest possible number of variables that include essential factors for biological activity and are independent from each other.

To select the optimum set of explanatory variables, we examined the general evaluation function for binding energy used in molecular modeling and simulation.[11] This function consists of potential energy term with solvation energy, and cratic and conformational entropy terms.

Potential energy term with solvation energy, a major constituent of binding energy, is approximated by summing up van der Waals potential term, electrostatic potential term, hydrogen bond energy term, and solvation free energy term. When a ligand binds to the target protein, they are complementary so that every emerging term value is within a favorable range. A pharmacophore represents this situation qualitatively and illustratively, providing what features are participating in the binding and the relative locations of the features. All or a part of pharmacophore is used as pharmacophore descriptor in chemo-informatics analyses.[12] So-called *n*-center (2-center, 3-center, etc.) pharmacophore descriptor consists of n number of binding features and distances between the features. We adopted 2 for *n*, the minimum number for *n*, so that the redundancy and the number of explanatory variables were minimized.

We used six types of binding features (Table 1) in a pharmacophore so as to represent potential energy function. Hydrogen bond donor and acceptor features correspond to hydrogen bond energy term of potential energy function. Positive ionizable and negative ionizable features correspond to electrostatic potential term. Hydrophobic and aromatic features correspond to solvation free energy term. van der Waals potential is included indirectly in the topology of these features, because it is a term reflecting the molecular shape.

Distance between two binding features is represented as a class variable for a pharmacophore descriptor. If only a few classes of distance between binding features were defined, it would be easy to find the pharmacophore descriptors that were common among active compounds, but would be difficult to find ones that

Table 1. Structural descriptors in the CART analysis

Category	Name		Representation
	(feature1_feature2_distance)		
Pharmacophore descriptors 168	Feature	D	Hydrogen bond donor
		A	Hydrogen bond acceptor
		P	Positive ionizable
		N	Negative ionizable
		H	Hydrophobic center
		R	Aromatic center
	Distance	0	2.00–4.00
		1	3.75–5.75
		2	5.50–7.50
		3	7.25–9.25
		4	9.00–11.00
		5	10.75–12.75
		6	12.50–14.50
		7	14.25–16.25
Macroscopic descriptors 12	pvdw		Polar van der Waals surface area
	nvdw		Non-polar van der Waals surface area
	tvdw		Total van der Waals surface area
	pasa		Polar solvent accessible surface area
	nasa		Non-polar solvent accessible surface area
	tasa		Total solvent accessible surface area
	mw		Molecular weight
	rbonds		Number of rotatable bonds
	hbd		Number of hydrogen bond donors
	hba		Number of hydrogen bond acceptors
	A log *P*		Water/*n*-octanol partition coefficient
	dg		Solvent free energy

differentiate active compounds from inactive ones. Too many distance classes would bring about the opposite. We considered that an index of diversity for a chemical library based on the pharmacophore descriptors with the optimum class definition would give the best score. Hence, we searched the best number of distance classes by studying the diversity indices using 10,000 diverse compounds selected randomly from a commercially available chemical library.[13] As a result, the eight classes shown in Table 1 were concluded as the optimum class definition of distance between binding features.

Six types of binding features and eight classes of distance between them give 168 2-center pharmacophore descriptors. Each pharmacophore descriptor is represented as a bit in a 168 bit string. A compound is represented as a bit string which is the result of logical OR operation over all the bit strings that represent energetically acceptable conformations for the compound.

Some macroscopic descriptors were also used for the CART analysis to compensate with quantitative parameters for qualitative nature of pharmacophore descriptors. Conceptually, binding of a ligand to the target protein can be regarded as aqueous/non-aqueous phase partition phenomenon, because the interior portion of a protein is usually non-aqueous environment. Therefore, water/*n*-octanol partition coefficient ($\log P$) can be regarded as quantitative index of binding energy. Actually, there are many reports in which $\log P$ is described to be correlated with binding activities for the target proteins.[14] In addition, the number of hydrogen bond donors and acceptors can be regarded as quantitative index of hydrogen bond energy term. Van der Waals surface area can be regarded as quantitative indices of van der Waals energy and solvent accessible surface area as quantitative indices of electrostatic and solvation free energies.[15] Cratic entropy is correlated with size and shape of the ligand, and conformational entropy with flexibility of the ligand.[16] We used molecular weight and the number of rotatable bonds as quantitative indices of these entropies. Overall, we used 168 pharmacophore and 12 macroscopic descriptors in the CART analysis (Table 1).

CART analysis for the structure activity relationship of hydroxamic acid analogs. A hundred and two hydroxamic acid analogs, which we reported previously, showed inhibitory activities against PDE-5 ($IC_{50} < 10$ μM).[4] Among them nine compounds showed potent inhibitory activities ($IC_{50} < 100$ nM). Inhibitory activities against PDE-1 of 60 compounds were also reported.[4] In order to study the correlation between PDE-1 and 5 activities of the analogs, we made cross-tabulation of the PDE-1 and 5 activities, which were classified into three classes according to the IC_{50} values. As a result, we found a positive correlation between PDE-1 and 5 activities (Fig. 2). On the whole the inhibitory activity against PDE-1 is weaker than that against PDE-5. Therefore, we used all the active compounds against PDE-5 as active set of training data in the CART analysis. The obtained CART model, we expected, would not only detect the active compounds against PDE-5 but also PDE-1.

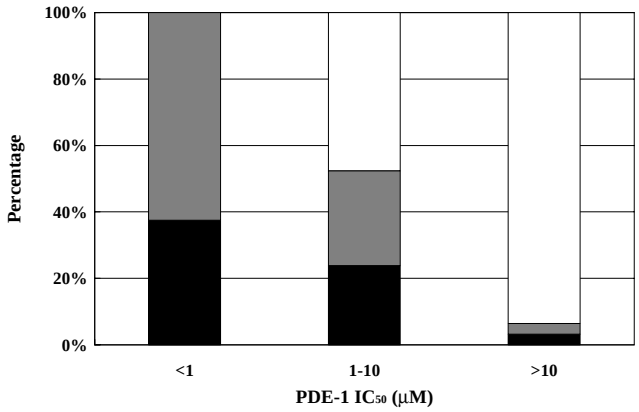


Figure 2. Correlation between PDE-1 and 5 activities of the hydroxamic acid analogs. PDE-5 $IC_{50} < 0.1$ μM (black), PDE-5 $IC_{50} = 0.1$–1 μM (gray), and PDE-5 $IC_{50} > 1$ μM (white).

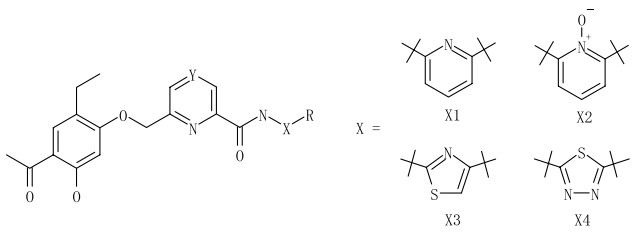
We presumed that the more training data a virtual screening based on the more accurate the prediction would be, and the more diverse candidate of active compounds it would produce. To realize the best possible predictability, we used not only the reported 102 hydroxamic acid analogs, but also additional 28 analogs (Table 2). The additional compounds were obtained by the same synthetic scheme as reported[4] in this laboratory. They showed inhibitory activities against PDE-5 ($IC_{50} = 0.13$–9.78 μM, in-house data).

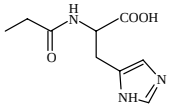
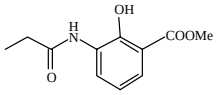
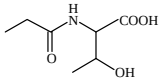
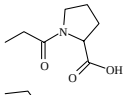
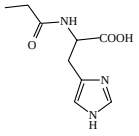
As active set of training data, 130 compounds mentioned above were used. As inactive set, 10,000 compounds, which were selected based on structural diversity from about 88,000 commercially available compounds supplied by SPECS Inc.[17], were used.

Pharmacophore descriptors were computed through three steps, namely conformational profiling, identification of binding features, and making bit strings based on the defined classes of distance between two features. Conformational profile and binding features were computed[18] with the Catalyst ver.4.0.[19] The procedure for converting the coordinates of binding features to bit strings of pharmacophore descriptors was performed with an in-house C program. Calculations of polar, apolar, and total van der Waals surface areas were carried out with another in-house Fortran program.[20] This program was also used to calculate solvent accessible surface area.[21] Other descriptors were computed with the Cerius2 ver.4.0.[19] The parameters by Eisenberg and McLachlan[15] were used for the calculation of solvent free energy of ligand. We used the Answer Tree ver.2.1J[22] for the CART analysis.[23]

Virtual library and screening. A library of commercially available chemical compounds supplied by SPECS Inc. were screened for PDE-5 inhibitory activity by a virtual screening. The library was considered to be suitable to find diverse hit compounds which differed structurally from the hydroxamic acid analogs. We used only 50,520 compounds that were not listed on the catalog

Table 2. Additional hydroxamic acid analogs used as training data for the CART analysis



Compound	X	Y	R
1	1	C	–$CH_2CONHCH_2COOH$
2	1	C	–$CH_2CONHCH(Et)COOH$
3	1	C	–$CH_2CONHCH(i\text{-}Pr)COOH$
4	1	C	–$CH_2CONHCH(CH_2OH)COOH$
5	1	C	–$CH_2CONHCH(CH_2COOH)COOH$
6	1	C	–$CH_2CONHCH(CH_2CH_2COOH)COOH$
7	1	C	–$CH_2CONHCH(CH_2CH_2SCH_3)COOH$
8	1	C	–$CH_2CONHCH_2CH_2COOH$
9	1	C	–$CH_2CONHCH_2CONHCH_2COOH$
10	1	C	–$CH_2CONHCH(CH_2COOEt)COOEt$
11	1	C	–$CH_2CON(Et)CH_2CH_2OH$
12	1	C	–$CH_2CON(CH_2CH_2OH)_2$
13	1	C	
14	1	C	
15	2	N	–$CH_2CONHCH(Me)COOH$
16	2	N	–$CH_2CONHCH(Et)COOH$
17	2	N	–$CH_2CONHCH(i\text{-}Pr)COOH$
18	2	N	–$CH_2CONHCH(CH_2COOH)COOH$
19	2	C	–$CH_2CONHCH(Me)COOMe$
20	2	C	–$CH_2CONHCH(CH_2COOH)COOH$
21	3	N	–$CH_2CONHCH(CH_2COOH)COOH$
22	3	C	–$CH_2CONHCH(CH_2OH)COOH$
23	3	C	–$CH_2CONHCH(CH_2COOH)COOH$
24	3	C	–$CH_2CONHCH(CH_2CH_2COOH)COOH$
25	4	C	–$CH_2CONHCH(CH_2COOH)COOH$
26	2	N	
27	2	N	
28	3	N	

of SPECS Inc. in September 1998 but listed in October 1999, so that there would be a higher possibility that these compounds were actually in stock.

In the first step of the virtual screening, the compounds with unfavorable physicochemical and pharmacokinetic properties were filtered out based on a modified Lipinsky's rule of 5;[24] that is, molecular weight below 500, number of hydrogen bond acceptors below 10, number of hydrogen bond donors below 5, A log *P* below 6, and number of rotatable bonds below 12. In the next step, the compounds with favorable biological activities were selected with the CART model for PDE-5 inhibition built as described above. SPSS ver.10.0[22] was used in both the steps of the virtual screening.

Acquisition of hit compounds of the virtual screening and biological assay. We selected a structurally diverse subset of 100 compounds among the hits in the virtual screening. Nineteen drug-like compounds among them were selected by medicinal chemists' inspection and obtained from SPECS Inc. (Table 2). They were assayed for in vitro inhibitions against PDE-1 and 5.[4]

Prediction model by CART analysis. We analyzed the training data of 130 hydroxamic acid analogs as PDE-5 inhibition active set and 10,000 diverse compounds as inactive set by the CART method. Fourfold cross-validation demonstrated the high predictive ability and robustness of 96.6% ± 1.0%. The CART model, configured using the training data divided in a ratio of 3:1 randomly, was highly predictive and could predict correctly 100% of the active set and 96.3% of the inactive set (Fig. 3). This decision tree model had four pharmacophore descriptors but no macroscopic descriptors as branching factors and three active nodes.

Virtual screening for PDE-5 inhibitory activity. In the first step of virtual screening of a commercially available chemical library, 43,365 compounds were selected by a modified Lipinsky's rule of 5 (85.8%). Among them, 1821 compounds were finally selected as PDE-5 inhibitor candidates (3.6%).

Biological activities. Nineteen diverse and drug-like compounds **29**–**47** selected from the virtual screening hits were assayed for the inhibitions of PDE-1 and 5. Among them seven compounds, **29**–**35** (37%), showed inhibitory activities against both PDE-1 and 5 at below 10 μM as IC_{50} (Table 3). Compound **29**, one of the seven dual inhibitors, showed the most potent inhibitory activity for PDE-5 at below 1 μM as IC_{50}, and compound **35** was the most potent against PDE-1, below 1 μM as IC_{50}.

As mentioned in the previous section, inhibitory activity for PDE-5 was used as the measure in this virtual screening although the goal was to identify dual PDE-1 and 5 inhibitors. Therefore, the predictive ability of this virtual screening should be evaluated by the PDE-5 inhibition. Eleven compounds, **29**–**39**, showed inhibition of PDE-5 over 50% at 10 μM. This constitutes 58% of the 19 tested compounds. These compounds were structurally diverse, since they were a subset of the 100 diverse compounds extracted from the 1821 virtual hit compounds by a diversity analysis. In fact, the molecular weight range of the 11 active compounds was 318–483, the range of A log *P* 1.3–5.5, and the range of the number of rotatable bonds 4–9.

Prospect of lead generation and optimization of PDE-1 and 5 dual inhibitor. We obtained a variety of new lead

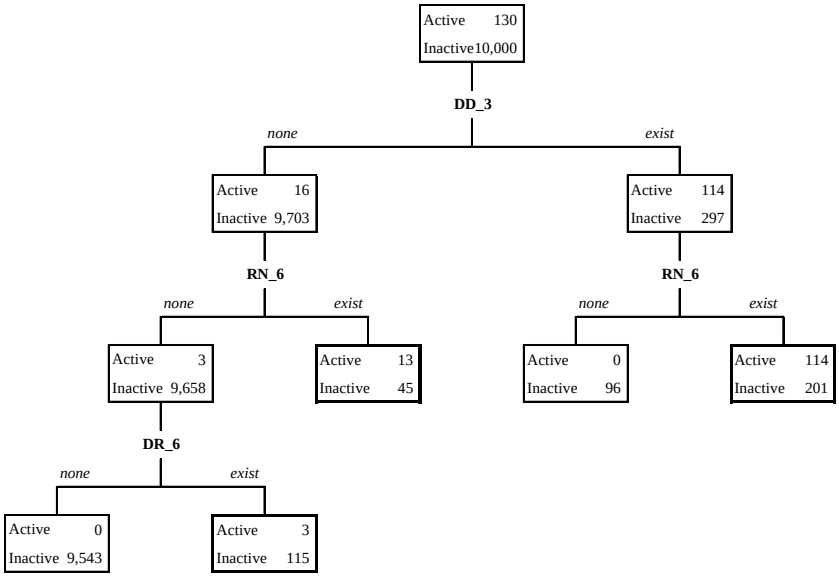


Figure 3. The prediction model for PDE-5 inhibition by the CART method. The terminal nodes for active classes are shown with bold lined boxes.

candidates which showed potent dual inhibition of PDE-1 and 5. We searched for seven compounds, **29**–**35**, among them in SciFinder®[25] (CAS registry numbers of them are listed in Table 3), and confirmed that none of the seven compounds had been reported of PDE inhibitory activity. On the other hand, these candidates lacked structural novelty, since they were derived from commercially available chemical library. However, none of the PDE inhibitors registered in MDDR (MDL Drug Data Report)[26] are structurally similar to the candidates. Therefore, by hit-to-lead derivatization, structurally novel PDE inhibitors will be identified without too much difficulty.

Generally, a lead compound has to have not only a potent in vitro activity but also acceptable physicochemical and pharmacokinetic properties, and consequently to show the expected therapeutic effects in the in vivo test. The inhibitors identified in this study are structurally diverse, therefore stand a good chance that some of them or derivatives of them show the expected in vivo effects.

Moreover, we have established a method of virtual screening for PDE-5 inhibitory activity. This method can be refined by using more information of structure–activity relationships. It can also be adapted for PDE-1 inhibitory activity. Hence, application of the method to derivatives of the lead candidates will also accelerate the lead discovery and optimization.

Advantage of the CART method for virtual screening. In this paper, we described a virtual screening method using carefully chosen structural descriptors and CART analysis, a method with the most active tree pruning among decision tree analyses. The virtual screening of a commercially available chemical library, which contained structurally diverse compounds, gave a good hit ratio for PDE-5 inhibitory activity; 58% of the compounds assayed showed the inhibitory activity of PDE-5 over 50% at 10 μM and 89% showed over 20% at 10 μM. The latter percentage was almost equal to the hit ratio of the training data. These results demonstrated that the method provided a highly predictive model for PDE-5 inhibitory activity and suggest that it can be the most appropriate among the extensively-used 'recursive partitioning' methods for ligand-based virtual screening.

Generally macroscopic descriptors are considered to be important factors for structure–activity relationships between derivatives with a common scaffold. But, as shown in Figure 3, only pharmacophore descriptors were selected for the prediction model by the CART analysis, but not macroscopic descriptors. This result suggested that the inhibitory activity against PDE-5 of the hydroxamic acid derivatives was determined mainly by the pharmacophore factors. To compare the prediction model (global model) obtained, as described above, using the training data of 10,000 diverse compounds with what we call local model, a CART model configured using 152 hydroxamic acid analogs as inactive set ($IC_{50} > 10$ μM, data not shown). Fourfold cross-validation showed a moderate predictive ability and a high robustness of 70.9% ± 2.7%. The model obtained by the divided training data could predict correctly 84.6% of the active set and 73.7% of the inactive set (Fig. 4). The predictive ability of the local model was lower than that of the global model, but still was at a useful level. The hierarchical structure of the local model was deeper than that of the global model. There were four active nodes, and one of them was reachable only by the macroscopic descriptors and the others by both the pharmacophore and macroscopic descriptors. We configured another local model using the same data set but without the macroscopic descriptors. Fourfold cross-validation showed a predictive ability and robustness of 66.0% ± 2.8%, lower than that of the local model with the macroscopic descriptors. These results suggest that the macroscopic descriptors are actually important to discriminate the actives from the inactives among the hydroxamic acid derivatives, that is, among structurally

Table 3. PDE-1 and 5 inhibitory activities of the hit compounds in the virtual screening of the commercially available compounds

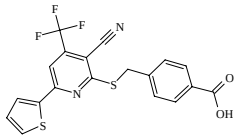
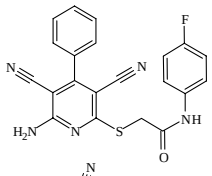
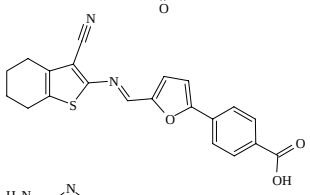
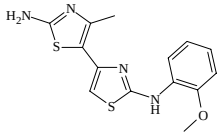
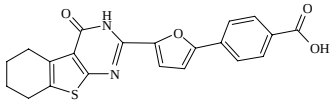
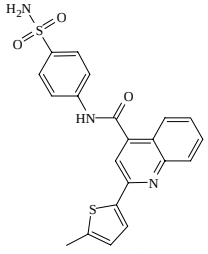
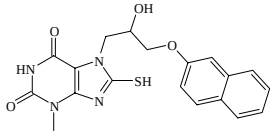
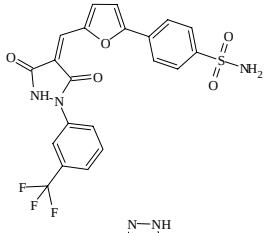
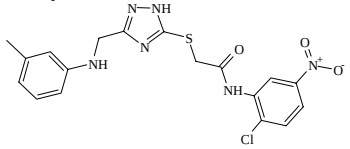
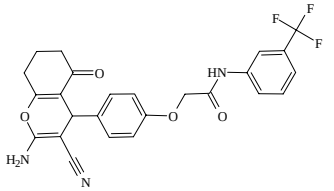
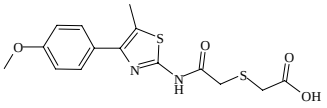
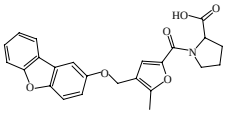
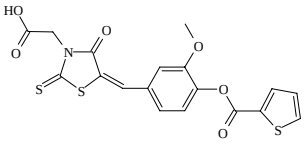
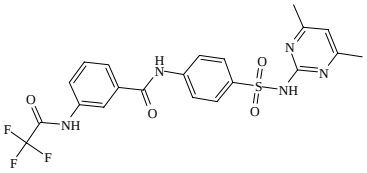
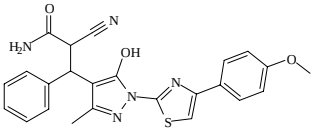
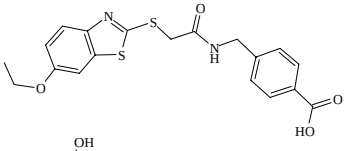
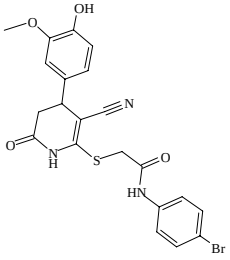
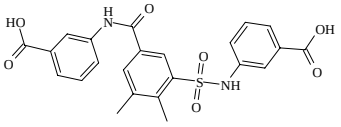
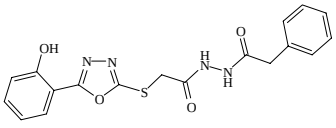
Compound	Structure	CAS registry no.	Catalog no.	IC_{50}/inhibition at 10 μM	
				PDE-1	PDE-5
29		336179-89-6	AK-777/11655009	IC_{50} = 1.9 μM	IC_{50} = 0.7 μM
30		354552-90-2	AM-807/12426108	IC_{50} = 5.7 μM	IC_{50} = 1.7 μM
31		357190-49-9	AE-641/09779061	IC_{50} = 2.5 μM	IC_{50} = 2.1 μM
32		315704-94-0	AG-690/12243209	IC_{50} = 2.5 μM	IC_{50} = 3.0 μM
33		357619-70-6	AM-900/12992016	IC_{50} = 3.7 μM	IC_{50} = 3.1 μM
34		353786-40-0	AK-968/12687248	IC_{50} = 2.5 μM	IC_{50} = 6.3 μM
35		303970-83-4	AG-690/12243056	IC_{50} = 0.3 μM	IC_{50} = 8.7 μM
36		—	AN-465/12849048	45%	66%
37		—	AG-690/13702382	22%	53%
38		—	AG-690/12870191	9%	52%

Table 3 (*continued*)

Compound	Structure	CAS registry no.	Catalog no.	IC_{50}/inhibition at 10 μM	
				PDE-1	PDE-5
39		—	AI-020/37280112	21%	51%
40		—	AH-487/37146001	42%	48%
41		—	AM-879/37266002	59%	46%
42		—	AG-205/10457031	15%	44%
43		—	AK-968/12383011	9%	42%
44		—	AK-968/12265299	12%	39%
45		—	AE-848/11828113	25%	23%
46		—	AK-918/37228001	49%	19%
47		—	AG-690/13416257	22%	1%

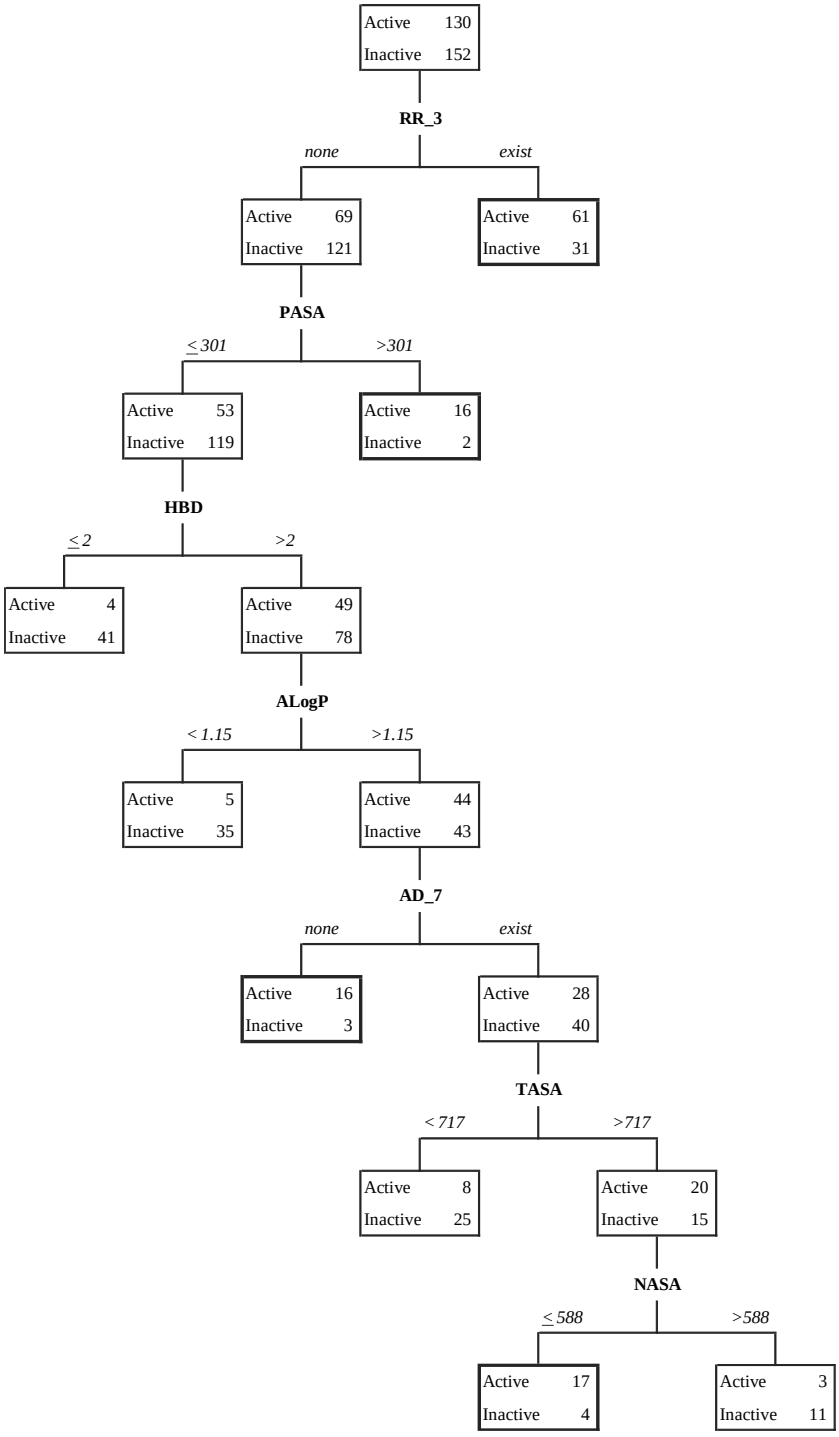


Figure 4. Local model for PDE-5 inhibition by the CART method. The terminal nodes for active classes are shown with bold lined boxes.

similar compounds. Virtual screening using such a local model should be useful during the hit-to-lead or lead optimization processes for prioritization among the derivatives of each lead candidate identified using the global model.

In conclusion, we identified new lead candidates of PDE-1 and 5 dual inhibitors through a ligand-based virtual screening optimized for lead evolution. This virtual screening method, consisting of CART analysis using 168 2-center pharmacophore descriptors and 12 macroscopic descriptors, showed a high predictive ability for bioactivity of new chemical compounds. With this virtual screening method, we expect that some lead compounds and, subsequently, a drug candidate will be found efficiently.

Acknowledgment

We thank Ms. Masako Sumitomo for her technical assistance.

References and notes

1. Rybalkin, S. D.; Yan, C.; Bornfeldt, K. E.; Beavo, J. A. *Circ. Res.* **2003**, *93*, 280.
2. Wallis, R. M.; Corbin, J. D.; Francis, S. H.; Ellis, P. *Am. J. Cardiol.* **1999**, *83*, 3C.
3. Vemulapalli, S.; Watkins, R. W.; Chintala, M.; Davis, H.; Ahn, H. S.; Fawzi, A.; Tulshian, D.; Chiu, P.; Chatterjee, M.; Lin, C. C.; Sybertz, E. J. *J. Cardiovasc. Pharmacol.* **1996**, *28*, 862.
4. Dan, A.; Shiyama, T.; Yamazaki, K.; Kusunose, N.; Fujita, K.; Sato, H.; Matsui, K.; Kitano, M. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4085.
5. Jenkins, J. L.; Glick, M.; Davies, J. W. *J. Med. Chem.* **2004**, *47*, 6144.
6. Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Wadsworth: Pacific Grove, CA, 1984.
7. Jeffrey, J. S.; Donald, F. W. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1028.
8. Kamaldeep, K. C.; Stuart, W. P.; Jaina, M.; Patrick, B.; Andrew, M. D. *J. Med. Chem.* **2005**, *48*, 5154.
9. Jane, P. F. B.; Andrey, U.; Gordon, C.; Han-Dan, H.; Volker, F.; Robert, T.; He-Qun, Y.; Cheng-Pang, H.; Lan, J.; Kin-Kai, H. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2061.
10. Mazzatorta, P.; Benfenati, E.; Lovenzini, P.; Vighi, M. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 105.
11. Gasteiger, J. *Handbook of Chemoinformatics: From Data to Knowledge*; Wiley-VCH GmbH and Co. KGaA: Weinheim, 2003.
12. Chen, X.; Rusinko, A., III.; Young, S. S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1054.
13. ISIS keys and Euclid distances were used for calculating the diversity indices.
14. Kubinyi, H. *Drug Discovery Today* **1997**, *2*, 538.
15. Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.
16. Jason, S. B.; Irene, L.; Ernesto, F. *Biochemistry* **1997**, *36*, 6588.
17. http://www.specs.net/.
18. The parameters were set to FAST as the option for the conformational profiling and 255 as the maximum number of conformers.
19. http://www.accelrys.com/.
20. Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
21. These surface areas were calculated for all conformations with acceptable potential energy and the Boltzman weighted average was assigned to each molecule. We defined oxygen, nitrogen, and hydrogen bonded to either oxygen or nitrogen as polar atoms for the calculations of the surface areas. The radius of the probe molecule for solvent accessible surface area was set to 1.4 A.
22. http://www.spss.com.
23. The training data set was divided into two sets at a ratio of 3 to 1 randomly, and the former was used to grow the decision tree and the latter to prune it. GINI index was used as the index of classification. We used four parameters for termination of tree growing. The minimum variation of impurity was set to 0.0001, the maximum tree size was 10, and the minimum number of records of parent and child nodes was 5 and 1, respectively. We assumed that the prior probability distribution was equal in both the active and inactive classes, because the number of data in the active set was far fewer than in the inactive set.
24. Lipinsky, C.; Lombardo, F.; Dominy, B.; Feeney, P. *Adv. Drug Delivery Rev.* **1997**, *23*, 3.
25. http://www.cas.org/.
26. http://www.mdli.com/.