Classification of HDAC8 Inhibitors and Non-Inhibitors Using Support Vector Machines

Guang Ping Cao, Sundarapandian Thangapandian, Shalini John and Keun Woo Lee*

Division of Applied Life Science (BK21 Program), Systems and Synthetic Agrobiotech Center (SSAC), Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Research Institute of Natural Science (RINS), Gyeongsang National University (GNU), Jinju, Korea

SYNOPSIS

Introduction: Histone deacetylases (HDAC) are a class of enzymes that remove acetyl groups from ε-N-acetyl lysine amino acids of histone proteins. Their action is opposite to that of histone acetyltransferase that adds acetyl groups to these lysines. Only few HDAC inhibitors are approved and used as anti-cancer therapeutics. Thus, discovery of new and potential HDAC inhibitors are necessary in the effective treatment of cancer.

Materials and Methods: This study proposed a method using support vector machine (SVM) to classify HDAC8 inhibitors and non-inhibitors in early-phase virtual compound filtering and screening. The 100 experimentally known HDAC8 inhibitors including 52 inhibitors and 48 non-inhibitors were used in this study. A set of molecular descriptors was calculated for all compounds in the dataset using ADRIANA. Code of *Molecular Networks*. Different kernel functions available from SVM Tools of free support vector machine software and training and test sets of varying size were used in model generation and validation.

Results and Conclusion: The best model obtained using kernel functions has shown 75% of accuracy on test set prediction. The other models have also displayed good prediction over the test set compounds. The results of this study can be used as simple and effective filters in the drug discovery process.

Collection of Dataset (HDAC8 inhibitors) → Calculation of Descriptors → Generation of SVM Models

From published literature / Using ADRIANA.Code / Using LIBSVM software
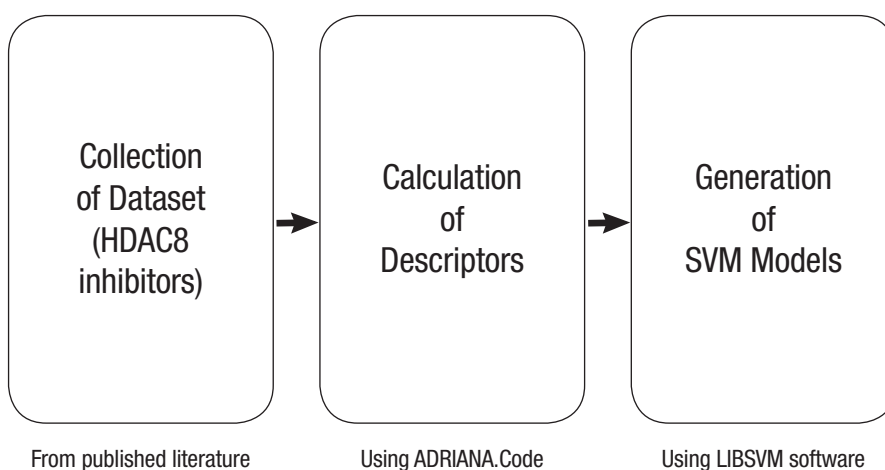
Key Words: histone deacetylase 8; support vector machine; ADRIANA.Code; libSVM tool; classification model; drug discovery process

INTRODUCTION

Much has been written recently concerning the impact of drug/ nondrug (inhibitors/non-inhibitors) classification in the field of drug discovery. Early-phase virtual screening and compound library design often employs filtering routines which are based on binary classifiers and are meant to eliminate potentially unwanted molecules from a compound library[1-3]. The support vector machine (SVM) is the most often used classifier in these applications. The SVM is firstly proposed for classification by V. Vapnik in 1995[4]. It has been widely applied to various areas of research in drug discovery[5-8], and first application in molecular informatics and pharmaceutical research have been described[9-11]. The standard scenario for SVM classifier can be summarized in two stages: training and testing. In first stage, sample data are basically n-dimensional vectors which are calculated by descriptor algorithms with a class membership label attached. And the SVM generates a classifier for prediction of the class label of test data during the second stage.
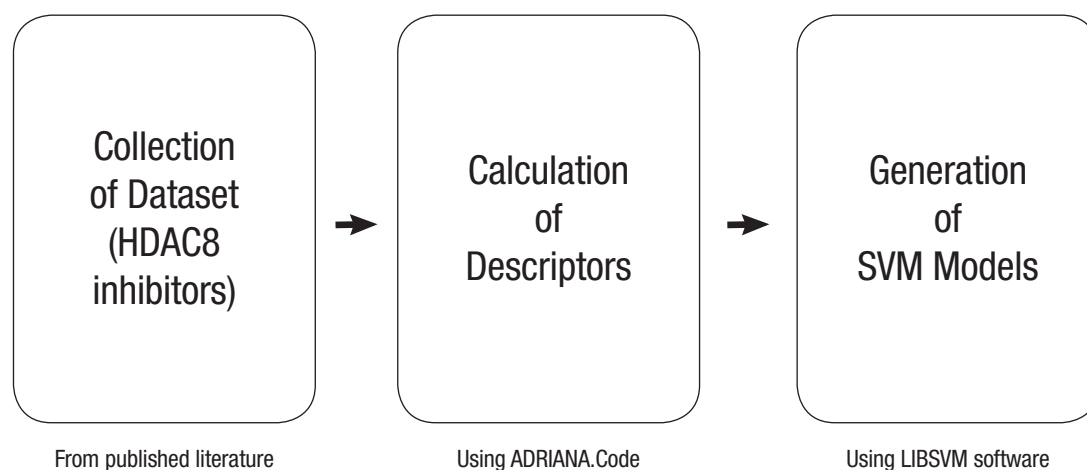
Histone deacetylases (HDACs) are a class of enzymes that remove acetyl groups from ε-N-acetyl lysine amino acids of histone proteins (Table 1). Their action is opposite to that of histone acetyltransferases (HATs) that add acetyl groups to the same lysine residues[12]. Acetylation is a post-translational modification that controls the biological function and stability of proteins in eukaryotic cells[13]. HDAC enzymes are classified into four different classes based on their phylogeny and domain organization[14].

Table 1. Different classes and members in the family of HDACs

Class	HDACs
Class I	HDAC1, HDAC2, HDAC3, HDAC8
Class II	HDAC4, HDAC5, HDAC6, HDAC7A, HDAC9, HDAC10
Class III	Sirtuins (SIRT1, SIRT2, SIRT3, SIRT4, SIRT5, SIRT6, SIRT7)
Class IV	HDAC11

The HDAC8 enzyme belongs to the class I enzymes which are found primarily in the nucleus[15]. Except HDAC8, functional HDACs are not found as single peptides but as multimeric complexes of higher molecular weight and also most of the purified HDAC enzymes are functionally inactive[13,16]. Along with this advantage, expression of HDAC8 notably correlates with the disease stage of neuroblastoma, a highly malignant childhood cancer derived from the sympathetic nervous system[17,18]. Moreover, an RNA interference study showed that HDAC8 is involved in the regulation of proliferation, clonogenic growth and neuronal differentiation of neuroblastoma cells. Inv1, an abnormal fusion protein formed during acute myeloid leukemia binds HDAC8, is also associated with aberrant, constitutive genetic repression[19]. Therefore, HDAC8 is considered to be the best model among other mammalian HDACs from a structural biology and drug discovery perspective. Only a few HDAC8 inhibitors are approved by FDA and being used as anti-cancer therapeutics. Thus, the discovery of new and potential HDAC8 inhibitors is necessary in the effective treatment of cancer. All HDAC inhibitors till date possess common structural features including metal binding and surface binding moieties along with a linker of four to six carbon chains long that connects metal and surface binding moieties[20]. This arrangement of chemical features is necessary to bind the tunnel like active site present in HDACs[21].

In this study, we used the SVM algorithm to classify HDAC8 inhibitors and non-inhibitors in early-phase virtual compound filtering and screening of drug discovery process. A set of molecular descriptors was calculated for all compounds in the dataset using ADRIANA.Code of Molecular Networks Inc. Different kernel functions available from *SVM Tools* were used in model generation and validation (Figure 1). The proposed method is effective and efficient in classifying known HDAC8 inhibitors with high correlation.

Collection of Dataset (HDAC8 inhibitors) → Calculation of Descriptors → Generation of SVM Models

From published literature   Using ADRIANA.Code   Using LIBSVM software

Figure 1. Architecture of the proposed study.

RESULTS AND DISCUSSION

Model generation

The main aim of this study is to find a valid SVM model to classify HDAC8 inhibitors and non-inhibitors, we tried many experiments with the biological responses and calculated molecular descriptors by SVM software and compared the performance based on different input patterns and different kernel functions. Four basic kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid kernel functions were used in this study (Table 2).

The effective and efficient model was obtained by all 4 kinds of kernel functions. The SVM models were generated with different number of descriptors from the total of 23 calculated descriptors using different kernel functions. The performances of polynomial kernel function and RBF kernel function were more stable than the other generated models. But the model of linear kernel function had a good performance in the case of using few descriptors.

SVM performance

In this study, we trained different models using two training sets and performances of the best models were compared with each other. The performances of the developed SVM models were tested on two test sets. The test set 1 contained 32 inhibitors and 28 non-inhibitors whereas the test set 2 contained 20 inhibitors and 20 non-inhibitors. On these two test sets, the best model obtained has shown 75% of accuracy, the other kernel

functions have also displayed good prediction over the test set compounds. But the model with 75% accuracy was developed with 18 descriptors, which is not acceptable from the medicinal chemistry point of view[22]. This study is focused on developing SVM models that can be used as simple filters in the early stage of drug discovery process. Thus the models classify the samples using high number of descriptors also were considered in this study. But high importance is given towards the models classifying the samples with less number of descriptors. Model 4 and 5 of training set 1 are developed with 4 and 3 descriptors, respectively (Table 3). The descriptors such as HDon, Hacc, XlogP, and NRotBond were used in the development of model 4 and 5 (Table 4). This explains that these descriptors are of great influence in classifying the HDAC8 inhibitors. The trend of decreasing prediction percentage with increased number of descriptors has shown the negative influence of other descriptors. The proposed algorithm did not adjust the classification boundary in order to include all inhibitors into non-inhibitors class since this would have resulted in an unacceptably low specificity. Model 5 that is generated using training set 1 (20 inhibitors and 20 non-inhibitors) was selected as best model. In the case of low numbers of sample data, the effect of model of linear kernel function which is composed of few descriptors prevails with 70% of accuracy, the model was made up of Hacc, Xlogp and

Table 2. Four basic kernel function used in SVM modeling

Kernel function	Expression
Linear kernel	$K(x_i, x_j) = x_i^T \cdot x_j$
Polynomial kernel	$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
Radial basis function (RBF)	$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
Sigmoid kernel function	$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r), \gamma > 0$

Table 3. Comparison of the results of different SVM models generated with training set 1

Model	Descriptors	Training set		Test set		Correctly predicted in (%)	Kernel function
		Inh	Non-inh	Inh	Total		
Model 1	All 23 descriptors	20	20	40	60	66.67	Polynomial
Model 2	18 descriptors	20	20	39	60	65	RBF
Model 3	10 descriptors	20	20	40	60	66.67	Polynomial
Model 4	4 descriptors	20	20	41	60	68.33	Linear
Model 5	3 descriptors	20	20	42	60	70	Linear

Inh, Inhibitor; Non-inh, Non-inhibitor.

Table 4. Descriptors used in each model

Model	Descriptors
Model 1	Weight, HDon, HAcc, XlogP, TPSA, Polariz, Dipole, LogS, NRotBond, NViolationsRo5, NViolationsExtRo5, NAtoms, NStereo, Complexity, RComplexity, Diameter, InertiaX, InertiaY, InertiaZ, Span, Rgyr, Eccentric, Aspheric
Model 2	HDon, HAcc, XlogP, TPSA, Polariz, Dipole, LogS, NRotBond, NViolationsRo5, NViolationsExtRo5, NAtoms, NStereo, Complexity, RComplexity, Diameter, Rgyr, Eccentric, Aspheric
Model 3	HDon, HAcc, XlogP, TPSA, Polariz, Dipole, LogS, NRotBond, NViolationsRo5, NAtoms
Model 4	HDon, HAcc, XlogP, NRotBond
Model 5	Hacc, Xlogp, NRotBond
Model 6	Weight, HDon, HAcc, XlogP, TPSA, Polariz, Dipole, LogS, NRotBond, NViolationsRo5, NViolationsExtRo5, NAtoms, NStereo, Complexity, RComplexity, Diameter, InertiaX, InertiaY, InertiaZ, Span, Rgyr, Eccentric, Aspheric
Model 7	HDon, HAcc, XlogP, TPSA, Polariz, Dipole, LogS, NRotBond, NViolationsRo5, NViolationsExtRo5, NAtoms, NStereo, Complexity, RComplexity, Diameter, Rgyr, Eccentric, Aspheric
Model 8	HDon, HAcc, XlogP, TPSA, Polariz, Dipole, LogS, NRotBond, NViolationsRo5, NAtoms
Model 9	HDon, HAcc, XlogP, NRotBond
Model 10	HAcc, Xlogp, NRotBond

NRotBo descriptors. The prediction results are 66.67% with polynomial kernel function for model1, 65% with RBF kernel function for model 2, 66.67% with polynomial function for model 3, both model 4 and model 5 used linear kernel function, the results are 68.33% and 70%, respectively (Table 3).

In terms of the second set of models developed using training set 2, we observed 75% correct prediction over the test set using 18 descriptors with polynomial kernel function. All prediction results were over 70% when polynomial kernel function is used (Table 5). But the predictions percentages were only 67.5% when the RBF and linear kernel functions are used in the development model 9 and 10, respectively (Table 5). Models 9 and 10 were developed with 4 and 3 descriptors, respectively, with 67.5% predictive ability are of high significance in terms of identifying the key properties classifying the HDAC8 inhibitors. Interestingly, these models also were developed with same set of descriptors, which contains HDon, Hacc, XlogP, and NRot-Bond.

Comparison of the models developed using two different training and test sets with four different kernel functions shown that the classification accuracy improved with the increasing number of training samples. The global molecular descriptors are more relevant than the size and shape descriptors in the prediction results of the developed models.

CONCLUSION AND PROSPECTS

The focus of this study was to evaluate the performance of SVMs in classification problems. The SVM has been widely applied to various fields, especially drug discovery. The developed models in this work can be used as simple filters in the HDAC 8 inhibitors discovery process. Selecting relevant descriptors is both important and difficult for any machine learning method. In this study, a set of molecular descriptors was calculated for all compounds in the dataset using ADRIANA.Code of Molecular Networks Inc. And the results showed that these descriptors can effectively be used in drug design. The developed models of high accuracy such as model 5, 6, and 7 using support vector machine with different kernel functions are suitable to classify

HDAC8 inhibitors from non-inhibitors. Three global molecular descriptors such as HDon, HAcc, and NRotBond along with XlogP were the influencing factors in the classification using SVM models. This explains both the hydrophilic and hydrophobic nature of the HDAC8 inhibitors detailing the polar nature of the hydroxamic acid moiety and the hydrophobic nature of the tunnel binding and surface binding moieties. Thus the developed models in this study could effectively classify HDAC8 inhibitors and non inhibitors with high correlation.

MATERIALS AND METHODS

Collection of data set
A total of more than 500 compounds were collected from various literature resources including patents. Among these, 100 compounds with HDAC8 inhibitory activity values predicted under same biological assay conditions were selected to be used in SVM model development and validation. This data set has included 52 inhibitors and 48 non-inhibitor compounds, which was done based on IC_{50} values of the compounds. Based on their IC_{50} values, this data set was finally divided into two different training and test sets. Training set 1 is comprised of 40 compounds including 20 inhibitors and 20 non-inhibitors. Training set 2 was made of 60 compounds including 32 inhibitors and 28 non-inhibitors. Both the training sets were diverse in terms of chemical diversity and biological activity ranging from 0.008 to 35 μM.

Preparation of compounds and descriptor calculation
Various set of molecular descriptors are available currently. The molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into an useful number or the result of some standardized experiment[23]. In this study, 23 molecular descriptors including 2D and 3D descriptors were calculated using ADRIANA.Code program available from Molecular Networks Inc. These 23 descriptors from ADRIANA.Code included various physicochemical properties such as global descriptors, shape and size-related descriptors (Table 6).

Support Vector Machine (SVM)
The SVM is a set of related supervised learning methods that analyze data and recognize patterns. To understand how a SVM classifier works, first think of the task of separating two classes of points in space. The SVM classification task can be separated into two kinds of cases, namely, linearly separable cases and non-separable cases. If two classes are linearly separable, the classifier can define optimal separating hyperplanes for their separation easily (Figure 2A). If the sample vectors are overlap,

Table 5. Comparison of the results of different SVM models generated with training set 2

Model	Descriptors	Training set		Test set		Correctly predicted in (%)	Kernel function
		Inh	Non-inh	Inh	Total		
Model 6	All 23 descriptors	30	30	28	40	70	Polynomial
Model 7	18 descriptors	30	30	30	40	75	Polynomial
Model 8	10 descriptors	30	30	23	40	57.5	RBF
Model 9	4 descriptors	30	30	27	40	67.5	Linear
Model 10	3 descriptors	30	30	27	40	67.5	Linear

Inh, Inhibitor; Non-inh, Non-inhibitor.

Table 6. The molecular descriptors used in generating SVM models

Descriptor Name	Description	Abbreviation	Type of descriptors
Lecular weight	Molecular weight in [u] or [Da] derived from the gross formula	Weight	Global molecular descriptors
Number of hydrogen bonding acceptors	Number of hydrogen bonding acceptors derived from the sum of nitrogen and oxygen atoms in the molecule	HAcc	Global molecular descriptors
Number of hydrogen bonding donors	Number of hydrogen bonding donors derived from the sum of N-H and O-H groups in the molecule	HDon	Global molecular descriptors
Octanol/water partition coefficient (logP)	Octanol/water partition coefficient in [log units] of the molecule following the XlogP approach	XlogP	Global molecular descriptors
Topological polar surface area	Topological polar surface area in [Å^2] of the molecule derived from polar 2D fragments	TPSA	Global molecular descriptors
Mean molecular polarizability	Mean molecular polarizability in [Å^3] of the molecule	Polariz	Global molecular descriptors
Molecular dipole moment	Dipole moment in [Debye] of the molecule	Dipole	Global molecular descriptors
Aqueous solubility (logS)	Solubility of the molecule in water in [log units]	LogS	Global molecular descriptors
Number of rotatable bonds	Number of open-chain, single rotatable bonds	NRotBond	Global molecular descriptors
Number of Ro5 violations	Number of violations of the Lipinski's rule of 5 (Weight > 500, XlogP > 5, HDon > 5, HAcc > 10)	NViolationsRo5	Global molecular descriptors
Number of extended Ro5 violations	Number of violations of the extended Lipinski's rule of 5 (additional rule: number of rotatable bonds > 10)	NViolationsEx-tRo5	Global molecular descriptors
Number of atoms	Number of all atoms in the molecule (including hydrogen atoms)	NAtoms	Global molecular descriptors
Number of tetrahedral stereocenters	Number of tetrahedral chiral centers in the molecule	NStereo	Global molecular descriptors
Molecular complexity	Molecular complexity according to the approach by J. Hendrickson	Complexity	Global molecular descriptors
Ring complexity	Ring complexity according to the approach by J. Gasteiger and C. Jochum	RComplexity	Global molecular descriptors
Molecular diameter	Maximum distance between two atoms in the molecule in [Å]	Diameter	Size and shape descriptors
Principal moment of inertia of 1st principal axis	Principal component of the inertia tensor in xdirection in [Da·Å^2]	InertiaX	Size and shape descriptors
Principal moment of inertia of 2nd principal axis	Principal component of the inertia tensor in ydirection in [Da·Å^2]	InertiaY	Size and shape descriptors
Principal moment of inertia of 3rd principal axis	Principal component of the inertia tensor in zdirection in [Da·Å^2]	InertiaZ	Size and shape descriptors
Molecular span	Radius of the smallest sphere centered at the center of mass which completely encloses all atoms in the molecule in [Å]	Span	Size and shape descriptors
Molecular radius of gyration	Radius of gyration in [Å]	Rgyr	Size and shape descriptors
Molecular eccentricity	Molecular eccentricity	Eccentric	Size and shape descriptors
Molecular asphericity	Molecular asphericity	Aspheric	Size and shape descriptors
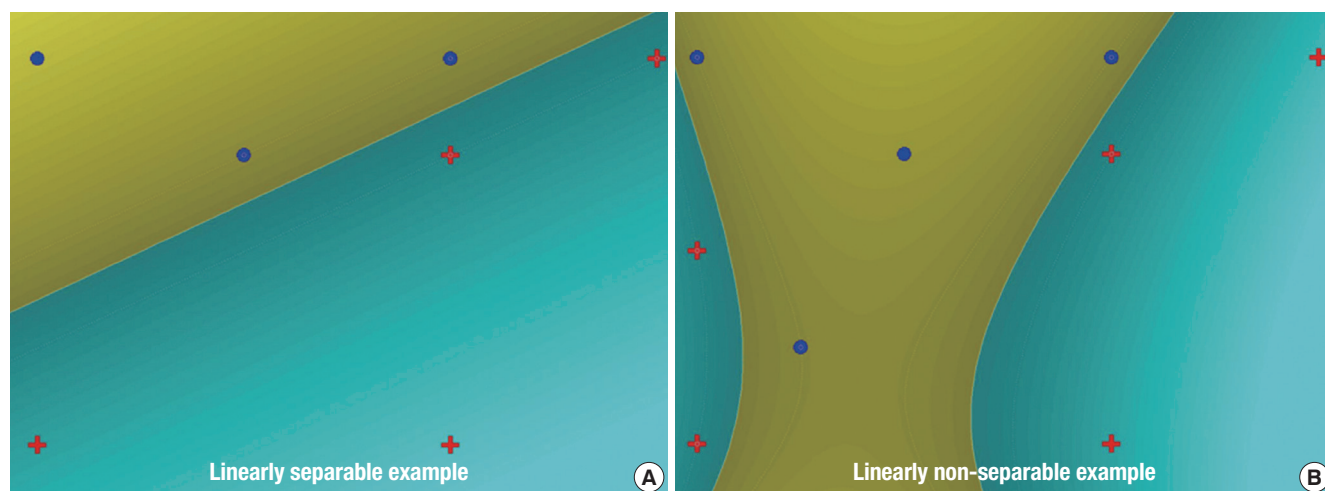
Figure 2. Two cases of SVM classification. (A) A linear separable case, the sample data are separated by linear kernel function. (B) A non-linear separable case, the sample data separated by polynomial kernel function, degree = 3.

classifier needs to generate nonlinear boundaries to separate them (Figure 2B). In short, the whole process of SVM classification can be summarized as a two-step procedure: First, the sample data vectors (descriptors of compounds) are mapped to a very high-dimensional feature space by kernel function. The dimension of this space is significantly larger than dimension of the original data space (Figure 3). It is not practical to use directly the feature function in computing the classification hyperplane. Instead, the nonlinear mapping induced by the feature functions is computed with special nonlinear functions
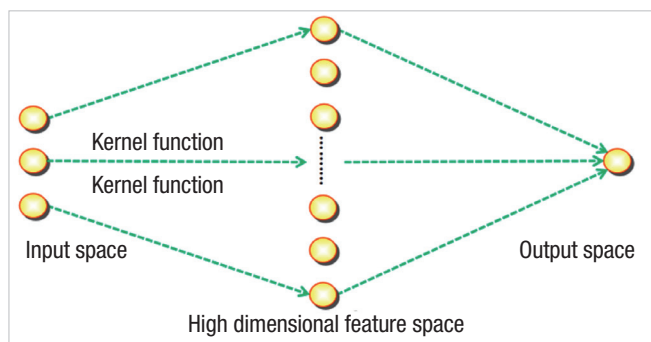
Figure 3. Mapping the input into a high dimensional feature space.

called kernels. Second, the classifier finds a hyperplane with the largest margin in this high-dimensional feature space with the largest margin separating classes of data. Sometimes it is not possible to find the hyperlane in high-dimensional feature space, so a tradeoff is introduced between the size of the separating margin and penalties for every vector which is within the margin[3,4]. The basic theory of SVM is briefly reviewed below:

The training data is defined as:
$$D = \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\}, x \in R^n, y \in \{-1, 1\}$$

The separating hyperplane is defined as:
$$D(x) = (w \cdot x) + b$$

Here x is a sample vector mapped to a high dimensional space, y is the class label of x, and w and b are parameter of the hyperplane that SVM classifier will estimate. A separating hyperplane in canonical form must satisfy the following constraints:
$$y^i [(w, x^i) + b] \geq |w| \tau, i = 1, 2, \ldots, n$$

The distance d(w, b, x) of a point x from the hyperplane (w, b) is:
$$d(w, b, x) = |w, x^i + b| / \|w\|$$

The margin can be expressed as a minimal τ, without loss of generality it applies a constraint $|w| \tau = 1$ to w, SVM training is becoming the problem finding the minimum of a function with the following constraints:
$$\eta(w) = 1/2 \|w\|^2$$
$$\text{subject to constraints } y^i [(w, x^i) + b] \geq 1$$

This problem is solved by Lagrange multipliers and minimization of the function:
$$\Phi(w, b, \alpha) = 1/2 \|w\|^2 - \sum_{i=1}^n \alpha_i \{y^i [(w, x^i) + b] - 1\}$$

Here α_i are Lagrange multipliers. Differentiating over w and x_i and substituting:

$$\max \Phi(\alpha) = \sum_{i=1}^n \alpha_i - 1/2 \sum_{i=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$
$$\text{subject to constraints } \sum_{i=1}^n y_i \alpha_i = 0, \alpha_i \geq 0, i = 1, \ldots, n$$

SVM introduced slack variable to solve the case that cannot be separate perfectly.
$$\text{minimize } \eta(w) = 1/2 \|w\|^2 + C \sum_i \xi_i$$
$$\text{subject to constraints } y_i [(w \cdot x_i) + b] \geq 1 - \xi_i$$

The ξ_i are slack variables, C is the error tradeoff parameter. The Lagrange multipliers can be obtained finally:
$$\max \Phi(\alpha) = \sum_{i=1}^n \alpha_i - 1/2 \sum_{i=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$
$$\text{subject to constraints } \sum_{i=1}^n y_i \alpha_i = 0, C \geq \alpha_i \geq 0, i = 1, \ldots, n$$

The freely available SVM software named LIBSVM (available at http://www.csie.ntu.edu.tw/~cjlin/libsvm) was used in generating SVM models[24]. Numbers of SVM models using different kernel functions were generated to accurately classify the training set compounds. The test compounds were used to validate the generated SVM models for their ability in classifying the external compounds that are not used in model generation.

ACKNOWLEDGEMENTS

REFERENCES

1. Clark, D.E., and Pickett, S.D. (2000). Computational methods for the prediction of 'drug-likeness'. *Drug Discov Today 5,* 49-58.
2. Schneider, G., and Böhm, H.J. (2002). Virtual screening and fast automated docking methods. *Drug Discov Today 7,* 64-70.
3. Byvatov, E., Fechner, U., Sadowski, J., and Schneider G. (2003). Comparison of Support Vector and Artificial Neural Network System for Drug/Nondrug Classification. *J Chem Inf Comput Sci 43,* 1882-1889
4. Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach Learn 20,* 273-297.
5. Han, L.Y., Zheng, C. J., Xie, B., Jia, J., Ma, X.H., Zhu, F., Lin, H.H., Chen, X., and Chen, Y.Z. (2007). Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. *Drug Discov Today 12,* 7-8.
6. Yap, C.W. and Chen, Y.Z. (2004). Prediction of cytochrome P450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines. *J Chem Inf Model 45,* 982-992.
7. Yang, Z.R., and Chou, K.C. (2004). Bio-support Vector Machines for Computational Proteomics. *Bioinformatics 20,* 735-741.

8. Mahé, P., Ueda, N., Akutsu, T., Perret, J.L., and Vert, J.P. (2005). Graph Kernels for Molecular Structure-Activity Relationship Analysis with Support Vector Machines. *J Chem Inf Model 45,* 939-951.

9. Burbidge, R., Trotter, M., Buxton, B., and Holden, S. (2001). Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput Chem 26,* 5-14.

10. Warmuth, M.K., Liao, J., Ratsch, G., Mathieson, M., Putta, S., and Lemmen, C. (2003). Active learning with Support Vector Machines in the drug discovery process. *J Chem Inf Comput Sci 43,* 667-673.

11. Wilton, D., Willett, P., Lawson, K., and Mullier, G. (2003). Comparison of ranking methods for virtual screening in lead-discovery programs. *J Chem Inf Comput Sci 43,* 469-474.

12. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science 325*(5942)*,* 834-40.

13. Vannini, A., Volpari, C., Gallinari, P., Jones, P., Mattu, M., Carfí, A., De Francesco R., Steinkühler, C., and Di Marco, S. (2007). Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8-substrate complex. *EMBO Reports 8,* 879-884.

14. Dokmanovic, M., Clarke, C., and Marks, P.A. (2007). Histone deacetylase inhibitors: overview and perspectives. *Mol Cancer Res 5*(10), 981-9.

15. Valenzuela-Fernández, A., Cabrero J.R., Serrador J.M., and Sánchez-Madrid, F. (2008). HDAC6: a key regulator of cytoskeleton, cell migration and cell-cell interactions error. *Trends Cell Biol 18*(6), 291-297.

16. Bolden, J.E., Peart, M.J., and Johnstone, R.W. (2006). Anticancer activities of histone deacetylase inhibitors. *Nat Rev Drug Discov 5,* 769-784.

17. Brodeur, G.M. (2003). Neuroblastoma: Biological insights into a clinical enigma. *Nat Rev Cancer 3,* 203-216.

18. Oehme, I., Deubzer, H.E., Wegener, D., Pickert, D., Linke, J.P., Hero, B., Kopp-Schneider, A., Westermann, F., Ulrich, S.M., von Deimling, A., Fischer, M., and Witt, O. (2009). Histone deacetylase 8 in neuroblastoma tumorigenesis. *Clin Cancer Res 15,* 91-99.

19. Durst, K.L., Lutterbach, B., Kummalue, T., Friedman, A.D., and Hiebert, S.W. (2003). The inv(16) fusion protein associates with corepressors via a smooth muscle myosin heavy-chain domain. *Mol Cell Biol 23,* 607-619.

20. Thangapandian, S., John, S., Sakkiah, S., and Lee, K.W. (2010). Docking-enabled pharmacophore model for histone deacetylase 8 inhibitors and its application in anti-cancer drug discovery. *J Mol Graph Model 29,* 382-395.

21. Thangapandian S., John, S., Sakkiah, S., and Lee, K.W. (2010). Ligand and structure based pharmacophore modeling to facilitate novel histone deacetylase 8 inhibitor design. *Eur J Med Chem 45,* 4409-4417.

22. Abu-Awwad F. M. (2009). A computational Study of Histamine H1-Receptor Agonist Activity Using QSPR and Molecular Surface Electrostatic Potential. *Imt J ChemTech Res* 1(3), 742-750.

23. Todeschini, R., and Consonni, V. (2000). Handbook of Molecular Descriptors, Weinheim: Wiley-VCH.

24. Chang, C. C., and Lin, C. J. (2011). LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology doi.acm.org/10.1145/ 1961189.1961-199. Software available at http://www.csie.ntu.edu.tw/~cjlin/lib svm