A combination of support vector regression, structure-based pharmacophore and virtual screening for cytochrome P450 2C9 inhibitors from Chinese herbs

Fang Lu, Ganggang Luo, Ludi Jiang, Yilian Cai, Yanling Zhang*

Key Laboratory of TCM-information Engineer of State Administration of TCM, School of Chinese Material Medica, Beijing University of Chinese Medicine, Beijing, 100102, China

e-mail: collean_zhang@163.com (Yanling Zhang)

Keywords: CYP2C9 inhibitors; support vector regression; structure-based pharmacophore; virtual screening; drug-drug interaction; traditional Chinese medicine

Abstract. Cytochrome P450 2C9 (CYP2C9), one important isoform of the cytochrome P450 (CYPs), mediates the oxidation of some important drugs. The inhibitors of this target often affect the metabolic rate of the corresponding metabolites and then result in undesirable drug-drug interactions (DDIs) in clinical. In order to discover potential CYP2C9 inhibitors, a support vector regression (SVR) model with good predictive ability were constructed. The correlation coefficient (R^2) and mean square error (MSE) values of the optimal SVR model were 0.952 and 0.003. Meanwhile, a structure-based pharmacophore (SBP) model was generated based on the crystal complex of CYP2C9 (PDB ID: 4NZ2) to refine the results of SVR model and elucidate the mechanism of the inhibitors. The best SBP model consists of one hydrogen bond acceptor feature, three hydrophobes features and six exclusion volumes. Then, both the optimal models of SVR and SBP were utilized to predict compounds in Traditional Chinese Medicines Database (TCMD) to identify potential CYP2C9 inhibitors from Chinese herbs. Finally, 1514 compounds were reserved, whose predicted active values obtained from SVR model and Fitvalues obtained from SBP model were all higher than the corresponding values of the initial compound in 4NZ2. Among them, ID 14767, which has higher predicted values and better mapping results with the SBP model, might exhibit inhibition effect on CYP2C9. Both the SVR model and SBP model might be applied in discovering potential CYP2C9 inhibitors from Chinese herbs, and also provide reference for the rational application of drugs in clinical.

Introduction

In modern clinical medical, traditional Chinese medicine (TCM) have been widely used to promote health and treat illnesses [1]. Sometimes TCM is also combined with Western medicine. Since the complex of the chemical compositions in TCM, the irrational application of drug combination may cause undesirable drug-drug interactions (DDIs). DDI is generally considered a modification of pharmacological and/or side effects of one drug by another [2]. Most clinical studies report that the metabolism-mediated drug interactions were primarily. And in many cases, DDI can be put down to regulation of drug-metabolizing enzymes, namely cytochrome P450 (CYP) [3].

CYPs belong to the superfamily of proteins including a heme cofactor and, therefore, are hemoproteins. Human CYPs are primarily membrane-associated proteins [4] which metabolize thousands of endogenous substrates and xenobiotics. Since changes in CYPs enzyme activity, such as inhibiting the activity of the CYP, may affect the metabolism and then result in unwanted DDIs. Cytochrome P450 2C9 (CYP2C9) is among the most important drug metabolizing isoform [5]. 16% of oxidative metabolism of all therapeutics is controlled by CYP2C9 and it has adverse drug effects, such as enzyme inhibition [6]. Hence, to distinguish CYP2C9 inhibitors before treatment in clinical can reduce the occurrence of adverse reactions.

In recent years, computational methods, such as pharmacophore, molecular docking, molecular dynamics (MD) [6], support vector machine (SVM) [7] have been used to ease the problem of

© 2015. The authors - Published by Atlantis Press

time-consuming and high-cost in drug research, which caused by traditional discovery CYPs inhibitors experiments. In this paper, a quantitative model was constructed by SVR to predict and screen potential inhibitors of CYP2C9. Then, SBP model was built to refine the results of SVR model as a cross-linking method and elucidate the mechanism of inhibitors. The purpose of our study is to build virtual screening models of the inhibitors of CYP2C9 and discover CYP2C9 inhibitors from TCMD. And also provides reference for rational use of drugs in clinical.

Materials and methods

Data preparation and data set splitting. 30 inhibitors of CYP2C9 were collected from reference and were defined as data set for the construction of SVR model [8]. 1481 molecular descriptors were calculated for all compounds to reflect their molecular structural characteristics by using Dragon2.1. Then, Kennard-Stone (KS) algorithm was utilized to split the 30 inhibitors into two parts, including training set (24 compounds) and test set A (6 compounds) [9]. This data decomposition approach can ensure the compounds in the training set had good representative. Subsequently, BestFirst and CfsSubsetEval within Weka3.6.10 were performed to select crucial molecular descriptors based on the data of training set [10]. After this, the combination of descriptors, which has a higher correlation coefficient with the active values of the compounds in training set, was obtained.

Development of SVR. SVR is an important and efficient statistical regression procedure, which can achieve generalized performance. In this paper, the network sharing program LibSVM3.1 was used to run the SVR algorithm. During this process, the kernel function and corresponding parameters should be chosen suitably. In this study, Radial Basis Function Kernel (RBF Kernel), one of the commonly used kernel functions, was used to build SVR models. C and γ, which are two important parameters of RBF Kernel and can affect the prediction accuracy of SVR model, were determined by three optimization methods, including Grid Search (GS), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO).

Before constructing the SVR models, three data processing methods, including non-treatment, ScaleForSVR function in LibSVM and principal component analysis in SPSS, were performed. Thus, three data processing methods combined with three parameter optimization methods were employed to build SVR models. Finally, a total of nine prediction quantitative models of CYP2C9 were constructed.

Validation of the SVR model. Two evaluation indicators, named correlation coefficient (R^2) and mean square error (MSE), were used to validate the nine SVR models based on test set A. The closer R^2 is to 1, indicating a higher correlation between real active value and predicted active value of compounds; the smaller MSE value, the predicted active value is more accurate.

Preparation of SBP. Four crystal structures of CYP2C9 were derived from PDB (http://www.rcsb.org/pdb/home/home.do). Wherein, only one structure (PDB ID: 4NZ2) which contains CYP2A9 inhibitor, was applied to perform SBP study. 23 inhibitors and 69 non-inhibitors were selected from The Binding Database (http://www.bindingdb.org/bind/index.jsp) as test set B. Three-dimensional structures of all the compounds were constructed by Discovery Studio 4.0 (DS 4.0). Compounds conformations were generated by BEST method and the maximum number of conformations was set to 255.

Generation and validation of SBP. SBP model can be generated based on the structure of protein or a protein-ligand complex [11]. In this study, the SBP model was constructed by DS4.0, in which the From Current Selection tool was applied to identify active pockets, and one pocket around the initial ligand was chosen. With Interaction Generation protocol employed, features related to the possible interaction points in the active site were generated. All crucial features containing A (Hydrogen Bond Acceptor Feature), D (Hydrogen Bond Donor Feature), H (Hydrophobic Features) of the active site would be mapped. A, D and H, three classes of feature in receptor structure would be analyzed respectively. Furthermore, A, D and H features were clustered automatically and useless features which were inconsistent with the features of the initial compound were deleted manually by employing Cluster Current Features protocol. Exclusion volumes were

also added based on receptor structure.

Then, the test set B was used to evaluate the SBP models. And the evaluation indicators were shown as follows: A%, represents the ability to identify active compounds from the test database; N, represents the ability to identify active compounds and CAI is the comprehensive appraisal index [12].

Database search. Based on TCMD, which contains approximately 23033 natural compounds from 6735 medicinal plants, the optimal SVR model was applied to predict the active value of compounds. Meanwhile, the best SBP model was employed to refine the results of SVR model and elucidate the mechanism of the inhibitors. In addition, the initial compound [13] was also studied by the optimal SVR model and the best SBP model. The predicted active value and FitValue of the initial compound was set as a threshold value for virtual screening. Finally, the natural compounds were retained, whose predicted active values obtained from SVR model and Fitvalues obtained from SBP model were all higher than the corresponding values of the initial ligand.

Results and analysis

Molecular descriptors selection. 1481 molecular descriptors were computed by Dragon 2.1. Then eighteen molecular descriptors were selected by using BestFirst and CfsSubsetEval algorithms. Thus, nine SVR models were built based on the eighteen molecular descriptors. The names of molecular descriptors were list in Table 1.

Table 1 The names of molecular descriptors

No.	Symbol	Definition	Class
1	PJI2	2D Petitjean shape index	topological descriptors
2	SIC2	structural information content (neighborhood symmetry of 2-order)	topological descriptors
3	MWC10	molecular walk count of order 10	molecular walk counts
4	JGI5	mean topological charge index of order5	Galvez topol. charge indices
5	GATS2m	Geary autocorrelation - lag 2 / weighted by atomic masses	2D autocorrelations
6	GATS5m	Geary autocorrelation - lag 5 / weighted by atomic masses	2D autocorrelations
7	Mor21m	3D-MoRSE - signal 21 / weighted by atomic masses	3D-MoRSE descriptors
8	E1v	1st component accessibility directional WHIM index / weighted by atomic van der Waals volumes	WHIM descriptors
9	R1u+	R maximal autocorrelation of lag 1 / unweighted	GETAWAY descriptors
10	R6u+	R maximal autocorrelation of lag 6 / unweighted	GETAWAY descriptors
11	R1e+	R maximal autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors
12	R4p+	R maximal autocorrelation of lag 4 / weighted by atomic polarizabilities	GETAWAY descriptors
13	R5p+	R maximal autocorrelation of lag 5 / weighted by atomic polarizabilities	GETAWAY descriptors
14	nCaH	number of unsubstituted aromatic C(sp2)	functional groups
15	C-027	R--CH--X	atom-centred fragments
16	C-035	X--CX..X	atom-centred fragments
17	C-040	R-C(=X)-X / R-C#X / X-=C=X	atom-centred fragments
18	N-075	R--N--R / R--N--X	atom-centred fragments

Construction and validation of SVR model. Based on the training set and the 18 selected molecular descriptors, nine SVR models were constructed. After validating the candidate models by using test set A, model 2 was chosen as the optimal model for further study, which obtained higher evaluation indicators both in training set and test set. The data processing method of this model was non-treatment and the corresponding parameter method was PSO, which was used to select the optimal (C, γ), and the optimal (C, γ) was (4.9636, 0.01). Meanwhile, the prediction R^2 and MSE of model 2 was 0.952 and 0.003 respectively, which indicated that the trend of prediction and the true value may be roughly consisted. The results of the parameters optimization and the trend of prediction were shown as Fig. 1. Then, the R^2 and MSE of the test set A were 0.954 and 0.010 respectively, which suggested that the quantitative model of CYP2C9 had good ability to predict the compounds in TCMD.

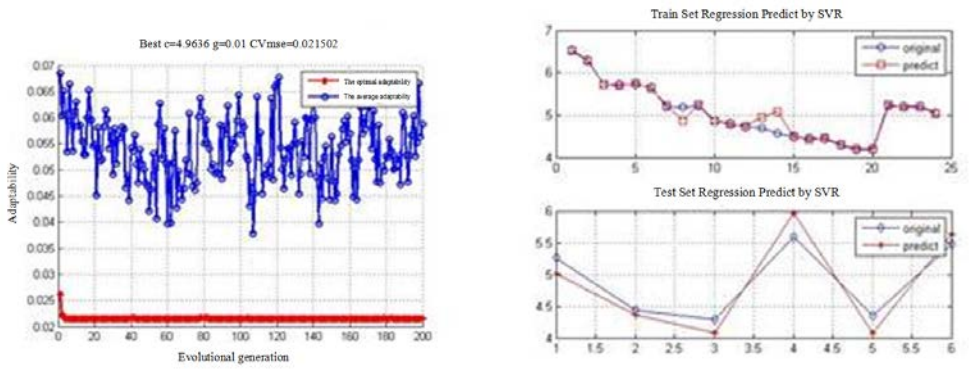

Fig. 1 The results of the parameters optimization and the trend of prediction

Generation and validation of SBP model. One active pocket which was around the initial ligand was chosen to generate pharmacophore models. The relevant evaluation parameters which were obtained by testing set B were showed in Table 2. The tenth pharmacophore model with the highest CAI value, 2.09, was chosen as the best SBP pharmacophore model, which included one A feature, three H features and six Ev (as Fig. 2 showed).

Table 2 The validation results of SBP pharmaphore models

Feature	A	D	Ha	Ht	A%	N	CAI
AADHHEv5	23	92	5	38	21.74%	0.53	0.12
ADHHEv2	23	92	8	30	34.78%	1.07	0.37
ADHHEv7	23	92	4	18	17.39%	0.89	0.15
AAHHEv2	23	92	9	28	39.13%	1.29	0.5
AAHHEv7	23	92	9	14	39.13%	2.57	1.01
AHHHEv3	23	92	14	30	60.87%	1.87	1.13
AHHHEv8	23	92	14	20	60.87%	2.8	1.7
AHHHEv8	23	92	11	17	47.82%	2.59	1.24
AHHHEv8	23	92	8	10	34.78%	3.2	1.11
AHHHEv6	23	92	13	14	56.52%	3.71	2.09

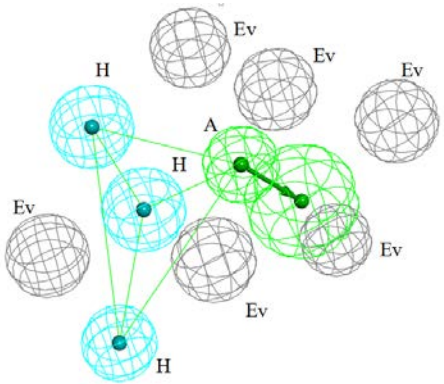

Fig. 2 The best SBP pharmacophore model of CYP2C9

Database search. Firstly, by using the optimal SVR and SBP model, the active value and FitValue of the initial compound in the crystal structure of CYP2C9 enzyme was 5.27, 1.63 respectively, which was regarded as the threshold values in identifying potential CYP2C9 inhibitors. Finally, a hit list of 1514 compounds was obtained which the active values and the FitValues were all higher than the corresponding values of the initial compound. Among these compounds, ID 14767, which derives from *Leontopodium leontopodioides*, has the effect of anti-inflammatory and obtain better prediction results [14]. The mapping results of ID14767 were similar to the initial compound (shown as Fig. 3).

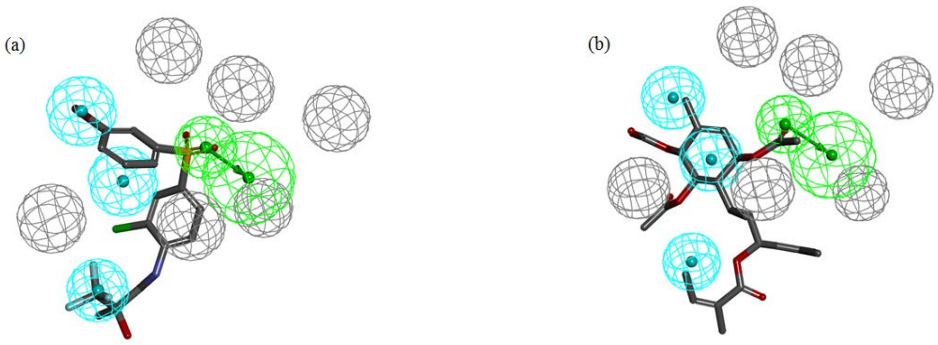

Fig. 3 The best SBP model mapped with the initial ligand (a) and ID14767 (b)

Conclusion

In this paper, a rational strategy for discovering CYP2C9 inhibitors was carried out by two computational methods, SVR and SBP. Both the optimal model of SVR and SBP were utilized to predict and discover potential inhibitors of CYP2C9 enzyme. Finally, 1514 compounds were retained, whose predicted active values and Fitvalues obtained from SVR and SBP model respectively were all higher than the corresponding values of the initial compound in 4NZ2. Among them, ID 14767, which has higher predicted values and better mapping results with the SBP model, might have promising inhibition effect on CYP2C9.

CYP2C9 enzyme catalyzes the biotransformation of several important drugs in clinical, containing ibuprofen, naproxen, flurbiprofen [15] and so on. Wherein, the three drugs all play important roles on resisting inflammatory, which was consistent with the anti-inflammatory effect of ID14767. Taking DDIs into account, these medicines and ID14767 were not suggested to be taken at the same time in clinical. Therefore, it is indicated that it cannot take CYPs' inhibitors simultaneously when taking the drugs which can be metabolized by CYPs.

In summary, both the SVR model and SBP model built in this paper might be applied in discovering potential CYP2C9 inhibitors from Chinese herbs, and also provide reference for the rational application of drugs in clinical. Besides, computational methods should combine with biological experiments in the following study, for the clinical safety use of TCM.

Acknowledgement

The authors gratefully acknowledge the support of this work by the National Natural Science Foundation of China (No. 81173522) and Joint Construction Project of Beijing Municipal Commission of Education.

References

[1] Loudon, I. Western Medicine: An Illustrated History, Oxford University Express, Oxford, UK, 1997.

[2] Dhananjay Pal, Ashim K. Mitra. MDR- and CYP3A4-Mediated Drug-Drug Interactions, J. J Neuroimmune Pharmacol, 2006, 1: 323-339.

[3] Chen J, Liu D, Zheng X, et al. Relative contributions of the major human CYP450 to the metabolism of icotinib and its implication in prediction of drug-drug interaction between icotinib and CYP3A4 inhibitors/inducers using physiologically based pharmacokinetic modeling, J. Expert Opinion on Drug Metabolism & Toxicology, 2015, 11.

[4] Berka K, Hendrychová T, Anzenbacher P, et al. Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site, J. J.phys.chem.a, 2011, 115(41): 11248–11255.

[5] Lee, C. R.; Goldstein, J. A.; Pieper, J. A. Cytochrome P450 2C9 polymorphisms: a comprehensive review of the *in-vitro* and human data, J. *Pharmacogenetics*, 2002, *12*:251-263.

[6] Wang J F, Yan J Y, Wei D Q, et al. Binding of CYP2C9 with Diverse Drugs and its Implications for Metabolic Mechanism, J. Medicinal Chemistry, 2009, volume 5(3):263-270(8).

[7] Yap C W, Chen Y Z. Prediction of cytochrome P450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines, J. J.chem.inf.model, 2005, 45(4):982--992.

[8] Lather V, Fernandes M X. Comparative QSAR analyses of competitive CYP2C9 inhibitors using three-dimensional molecular descriptors, J. Chem Biol Drug Des, 2011, 78(1):112–123.

[9] Saptoro, A., Tadé, M. O., & Vuthaluru, H. A modified Kennard-Stone Algorithm for optimal division of data for developing artificial neural network models, J. Chem. Prod. Process. Model. 2012, 7(1).

[10] Azuaje, F. Review of" Data Mining: Practical Machine Learning Tools and Techniques" by Witten and Frank, J. Biomed. Eng. Online. 2006: 51.

[11] Yang S Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances, J.DRUG DISCOV TODAY, 2010, 15(11): 444-450.

[12] He Y, Jiang L, Yang Z, et al. A combination of pharmacophore modeling, molecular docking, and virtual screening for P2Y 12 receptor antagonists from Chinese herbs, J. Canadian Journal of Chemistry, 2015, 93(1):311-316(6).

[13] Brändén G, Sjögren T, Schnecke V, et al. Structure-based ligand design to overcome CYP inhibition in drug discovery projects, J. Drug Discov Today, 2014, 19(7):905–911.

[14] Li L, Ye J,Yin H, et al. Effect of hoontopodium leontopodioides (Willd.) Beauv. On Inflammation Induced by Animal Reversed Passive Arthus (RPA), J. China Journal of Chinese Materia Medica,1994.

[15] Evgeny Byvatov, Karl-Heinz Baringhaus, Gisbert Schneider, Hans Matter. A virtual Screening Filter for Identification of Cytochrome P450 2C9 (CYP2C9) Inhibitors, J.QSAR & Combinatorial Science, 2007,No.5,618-628.