Comprehensive Interactions of ACE Inhibitors With Their Receptor by a Support Vector Machine Model and Molecular Docking: ACE Inhibitors with SVM Model and Docking

Article *in* Journal of the Chinese Chemical Society · February 2017

DOI: 10.1002/jccs.201600803

CITATIONS
6

READS
192

5 authors, including:

Yong-Hong Zhang
Chongqing Medical University
37 PUBLICATIONS **567** CITATIONS

SEE PROFILE

Article

JOURNAL OF THE CHINESE CHEMICAL SOCIETY

Comprehensive Interactions of ACE Inhibitors With Their Receptor by a Support Vector Machine Model and Molecular Docking

Ya'nan Liang,[a] Dongya Qin,[a] Yonghong Zhang,[b] Wanqian Liu[a]* and Guizhao Liang[a]*

[a]*Key Laboratory of Biorheological Science and Technology, Ministry of Education, School of Bioengineering, Chongqing University, Chongqing 400044, P. R. China*

[b]*Medicine Engineering Research Center & School of Pharmacy, Chongqing Medical University, Chongqing 400016, P. R. China*

(Received: November 1, 2016; Accepted: December 16, 2016; Published Online: February 1, 2017; DOI: 10.1002/jccs.201600803)

In this work, we characterize the interaction of angiotensin-I-converting enzyme (ACE) inhibitors with their receptor derived from the Binding Database by combining ligand-based and structure-based methods. The ligand-based quantitative structure–activity relationship (QSAR) model by support vector machine (SVM) achieves an overall accuracy of 88.74%, Matthews correlation coefficient of 0.678, and area under the receiver operating characteristic curve of 0.914 with leave-one-out (LOO) cross-validation on 444 training samples. The predictive ability of the model obtained is further verified by predictions on two test sets including 110 and 114 compounds. We show that the SVM-based model, with 2D and 3D QSAR advantages, is simple, accurate, and robust and can be used to predict and identify new ACE inhibitors. The four descriptors, namely the capacity factor, volume, standard deviation, and hydrophilic–lipophilic features, in the QSAR model can well represent the SAR of these inhibitors. In parallel, the structure-based molecular docking studies reveal that hydrogen bond is an important force for the binding affinities of the ACE inhibitors with the receptor. This work is useful in understanding the interaction mechanisms of ACE inhibitors with their receptor, as well as designing of new ACE inhibitors.

Keywords: Angiotensin-I-converting enzyme; Quantitative structure–activity relationship; Support vector machine; Molecular docking.

INTRODUCTION

Hypertension is characterized by a kind of arterial pressure, which could supervene with heart, blood vessels, brain, kidney, and other organs with functional or organic change in systemic diseases.[1] The angiotensin-I-converting enzyme (ACE) can hydrolyze HIS-LEU from the inactive decapeptide angiotensin I (AngI) with the C-terminal, forming an active vasoconstrictor octapeptide called angiotensin II (AngII).[2] AngII can contract blood vessels and increase blood volume to stimulate the release of aldosterone, which is essential for the regulation of blood pressure and results in elevated blood pressure eventually. ACE inhibitors can prevent the generation of angiotensin II and promote the formation of bradykinin, which reduces blood pressure. At present, both in theoretical research and clinical treatment, ACE inhibitors have been extensively applied to the prevention and treatment of high blood pressure.[3–5] Both polypeptides and small molecules can be used as ACE inhibitors. However, polypeptide-based inhibitors have the shortcomings of the low oral availability and hydrolyzing inactivation *in vivo*. It is important, therefore, to explore small molecules as inhibitors against hypertension.[6,7]

As we know, it is not feasible to test the bioactivities for all ACE inhibitors by experimental means, so computational methods have been used as important tools for predicting and screening inhibitors with specific activities. Although many quantitative prediction models have been proposed for screening ACE inhibitors in previous studies,[8,9] qualitative models remain comparatively few in the literature.[10] Both ligand-based and receptor-based methods have been used to predict ACE inhibitors. Up to now, compared to molecular

*Corresponding author. Email: wqliu@cqu.edu.cn; gzliang@cqu.edu.cn

dynamics simulations and quantum chemical calculations, quantitative structure–activity relationship (QSAR) and docking methods need less computational time and cost for such predictions.[11]

QSAR is a ligand-based method that has been popularly applied for predicting biological activity.[12–14] Both 2D QSAR and 3D QSAR are widely used for modeling. Several 2D physicochemical descriptors, including VDW's volume, net charge index and hydrophobic parameter,[15] *Z*-score,[13] FASGAI,[16] VHSE,[17] G-scales,[18] etc., have been used for modeling and designing peptide-based ACE inhibitors. Generally, these models are based on linear (e.g., multiple linear regression (MLR), ordinary least squares (OLS), principal component regression (PCR), partial least squares (PLS), etc.) and nonlinear methods (e.g., artificial neural networks (ANNs)[19]). The two-dimensional QSAR method focuses on more physical information but ignores spatial structural characteristics related to the activity.[20] As a classic 3D QSAR method, comparative molecular field analysis (CoMFA)[21] has been widely applied in molecular modeling.[22,23] To design hypertensive agents, CoMFA has been employed to build a 3D QSAR model[24] for the ACE inhibitors.[25] However, it has been reported that the CoMFA method has several limitations. For instance, in the absence of a reliable 3D receptor-bound structure, a multitude of ligand alignment protocols with different parameter settings is required to yield different sets of lead compounds for increasing prediction uncertainties.[26] Other 3D QSAR models based on interaction energy descriptors[27] and pharmacophore[28] were also used for the development of new potent hits for ACE inhibitors.

Classification techniques have been used to predict biological activities in pharmaceutical field.[29] Support vector machine (SVM) as a machine learning method has been widely applied in a variety of pattern recognition problems, and has revealed robust predictions in the classification and identification of peptides, proteins, genes, and so on.[30,31] Molecular docking is one of the most frequently used methods in structure-based drug design because of its ability not only to predict ligand-binding affinity by empirical or semiempirical scoring functions but also to offer information of the interaction between ligand and receptor in detail.[32] Therefore, methods based on QSAR modeling and molecular docking are often combined to clarify the relationship between the studied inhibitors and their enzyme targets.[33]

In this work, we explored the interaction mechanisms of ACE inhibitors with their receptor. First, the SVM-based QSAR model was established by representing 3D structural information of the compounds to predict a large set of ACE inhibitors derived from the Binding Database,[34] and two external test sets were used to verify the predictive ability of the model. The model was described by 3D structural information and established by SVM, thus it has the advantages of 2D and 3D QSAR models; i.e., it is simple, rapid, and robust. We then analyzed the SAR of the inhibitors with ACE using the SVM-based model. Second, we studied the interaction characteristics between these inhibitors and the receptor by molecular docking, which performed a consensus evaluation to enhance the prediction confidence of the SVM model. This work is helpful for understanding the mechanisms of ACE inhibitors with their receptor, which helps designing and virtually screening of new ACE inhibitors.

PRINCIPLES AND METHODS

Dataset

A large dataset including 1886 hits of pIC_{50} ($-\log IC_{50}$) data for polymerid was collected from the Binding Database.[34] There were 683 inhibitors of human ACE with experimental pIC_{50} values. After removing duplicate compounds with same structures but different pIC_{50} values, we then obtained 554 inhibitors as experimental samples. Of these, 422 samples with $pIC_{50} \geq 5$ ($IC_{50} \leq 10000$ nm) were regarded as inhibitors, and 132 samples with $pIC_{50} < 5$ ($IC_{50} > 10000$ nm) were regarded as non-inhibitors.

Two test sets were used to validate the external predictability of the trained model. These 554 samples were randomly divided into 4:1; as a result, 444 compounds (338 positive samples and 106 negative samples) as training set were used to establish the classification model, and 110 compounds as Test 1 (84 positive samples and 26 negative samples) were treated to verify the predictability of the model. In parallel, we collected 114 samples with pIC_{50} values as Test 2 (76 positive samples and 38 negative samples), but not included in the 554 data samples, from the published literature[35] to validate the extensive predictability and applicability of the model.

Structural description

Vsurf_descriptors are a category of variables including surface area, volume, shape, polarity, hydrophobicity, and the balance between them. It has been reported that these descriptors can characterize 3D physicochemical properties into a small number of quantitative numerical descriptors, thereby interpret the SAR of the molecules studied.[36] In this work, a total of 86 Vsurf_descriptors[37] involving surface area, volume, and shape descriptors calculated by the MOE software (http://www.chemcomp.com/) were used to construct the SVM model. To simplify the model, correlation analysis and Fselect[38] were applied to select the significant variables related to the inhibitory activity.

SVM modeling

SVM is an excellent method for the classification of two or more types of sample intervals, and uses an optimal hyperplane to classify samples with a maximum margin.[39] It has outstanding advantages for a small dataset, and uses nonlinear and high-dimensional pattern recognition to analyze data and build models for classification and regression. Here, the radial basis function (RBF) kernel was used to build the model. Two parameters, namely the regularization parameter C[40] and the kernel width parameter γ,[41] are the key elements of a successful SVM model. C is the maximum spacing parameter to adjust the balance of regulation and minimize the training error, and γ is the width of the Gaussian kernel parameter to influence the regulation and generalization for the SVM. A grid search method is applied to search criteria of (C, γ) to fit the highest accuracy and generate the final model. The SVM model was built by LIBSVM-3.20 (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) software.

Model validation

It is necessary to combine internal validation with external validation to verify a predictive model. A leave-one-out (LOO) cross-validation[42] was applied to investigate the predictability of the SVM model. It meant that one sample was extracted from the total N samples for test, and the remaining ($N - 1$) samples were regard as the training set. The cross-validation was repeated N times. The predictive performance was recalculated each time and then averaged 444 times. In the external validation, two test sets (Test 1 and Test 2) were used to validate the predictability and reliability of the predictor.

Model evaluation

The predictive results were evaluated using the statistical parameters[43] of accuracy (Acc), sensitivity (Sen), specificity (Spe), area under receiver operating characteristics curve (AUC),[44] and Matthews correlation coefficient (MCC),[45] which are shown in Eqs. (1)–(4):

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} \times 100 \quad (1)$$

$$\text{Sensitivity} = \frac{a}{a+c} \times 100 \quad (2)$$

$$\text{Specificity} = \frac{d}{b+d} \times 100 \quad (3)$$

$$\text{MCC} = \frac{ad-bc}{\sqrt{(a+b)(a+c)(c+d)(b+d)}} \quad (4)$$

where a is the number of inhibitors predicted as inhibitors, which is true positive; b means the number of non-inhibitors predicted as inhibitors, which stands for false positive; c means the number of inhibitors predicted as non-inhibitors, which is false negative; d means the number of non-inhibitors predicted as non-inhibitors, which is true negative; and Acc calculates the percentage of samples that are correctly identified. Sen and Spe are the percentage of the predicted inhibitors and non-inhibitors, respectively. However, Sen and Spe, as threshold-dependent measurements, fail to present all the information provided by one predictive method. Thus, MCC and AUC are used to account for unbalancing (both over and underprediction), and thereby to assess the modeling performance. Generally, the higher the parameters values, the better the prediction performance.[46]

Molecular docking

Molecular docking is one of the most popular methods for structure-based drug design. Here, we explored the binding poses by Surflex-dock in Sybyl 8.1.[47] The 3D coordinates of the X-ray crystal structure of human ACE[48] were downloaded from the RCSB Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do). Before molecular docking, the protein was prepared by the structure preparation tool module in Sybyl 8.1, and water molecules and the original

ligand were removed except for the cofactors Zn and Cl atoms. All 554 compounds were optimized by Tripos force field and Gasteiger–Huckel charges, and the termination gradient was set to 0.05 kcal/mol Å. We then added polar hydrogen atoms, repaired side chains of incomplete residues, and optimized the side chains for the protein. Missing residues were built using the Swiss-PdbViewer software.[49] During the docking process, the whole internal cavity in the pocket was used to generate the thresh of 0.5 and a bloat of 5 for the protomol. For each ligand, the docking poses were set as 20. Other docking parameters were set as default.

RESULTS AND DISCUSSION

Modeling results

A total of 86 3D descriptors were used to characterize the structures of the inhibitors studied. To simplify the model, we selected the variables by two methods as follows: First, correlation analysis was employed to eliminate highly correlated variables. After deleting the variables with a correlation coefficient larger than 0.85, we obtained the remaining 14 descriptors for SVM modeling. Second, using the Fselect tool module in Libsvm-3.20, we chose the first four descriptors with F-score values larger than 0.25 to build a robust predictor. The four kinds of descriptors are given in Supplementary Material (Tables S1 and S2, Supporting information). Generally, a large F-score expresses the more important variable for the QSAR model. It can be seen from Figure 1 that the four descriptor sets, i.e., capacity factor, hydrophilic–lipophilic balance, standard deviation, and volume descriptors, were significant in the SVM model.

The optimized SVM model ($C = 64$ and $\gamma = 0.016$) achieved Acc = 88.74%, Sen = 90.91%, Spe = 84.43%, and MCC = 0.68 (Table 1) on 444 training samples by the grid search (Figure 2). Based on the SVM model, two further external validations were operated on Test 1 and Test 2. The performance of the overall Acc, Sen, Spe, and MCC were 86.74%, 87.10%, 82.35%, and 0.60 for Test 1, and 71.93%, 94.00%, 54.69%, and 0.51 for Test 2, respectively (Table 1); moreover, the AUC on the 444 training samples, Test 1, and Test 2 were 0.91, 0.91, and 0.81, respectively (Figure 3), revealing the favorable prediction and generalization ability of the model obtained.

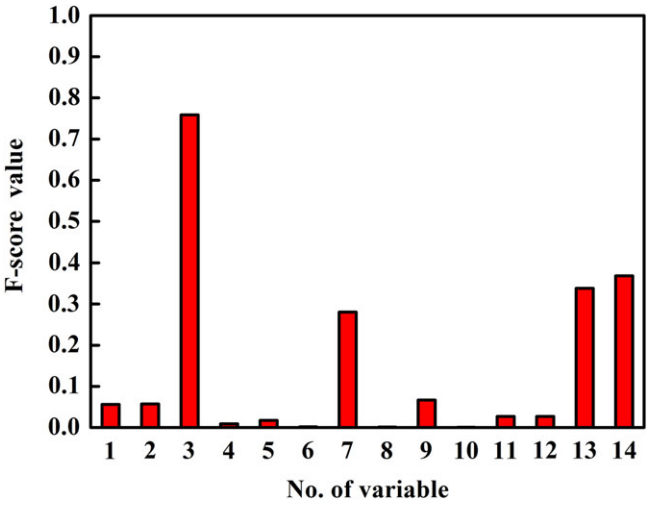

Fig. 1. *F*-scores of 14 variables in the SVM model (the larger the *F*-score, the more important the corresponding variable. The top four important descriptors are 3, 7, 13, and 14, representing capacity factor, hydrophilic–lipophilic, standard deviation, and volume, respectively.

To determine whether there was an accidental correlation between these descriptors and indicator variables, we randomly divided 554 samples into training set and test set in the ratio of 4: 1 (444: 110) to train the SVM model 10 times. We found the overall accuracy of training set was >85%, and AUC was close to 0.90, and the predictive performance of two test sets was also robust (Figure 4). The random validation results for 10 models demonstrated the reliability and stability of the SVM model.

In the used dataset, compounds with $pIC_{50} \geq 5.00$ were regarded as inhibitors, while those with $pIC_{50} < 5.00$ were regarded as non-inhibitors. Thus, there may be a "gray zone" where it may be inappropriate to regard a compound with a pIC_{50} value of 5.1 as an inhibitor and a compound with a pIC_{50} value of 4.9 as an inactive. To detect the "gray zone" in the dataset, we removed the data including $4.75 \leq pIC_{50} \leq 5.25$, of which there were 30 samples in proportion of 5% with 554 samples. The remaining 524 were randomly divided into 4:1 to build an SVM model. The prediction results of 10 times are shown in Figure S1. By comparing with the prediction results of the model based on all the samples, we found that the "gray zone" accounted only for a small part in the total dataset and had no significant

Table 1. Predicted results by LOO cross-validation by different methods

Correlation method	Samples	Acc (%)	Sen (%)	Spe (%)	MCC
Supporting vector machine	Train	88.74	90.91	84.43	0.68
	Test 1	86.37	87.10	82.35	0.60
	Test 2	71.93	94.00	54.69	0.51
K-nearest neighbor	Train	88.96	92.13	78.22	0.70
	Test 1	83.64	86.67	70.00	0.51
	Test 2	69.30	80.60	53.19	0.35
Random forest	Train	97.97	98.81	95.33	0.94
	Test 1	87.28	89.77	77.27	0.63
	Test 2	71.93	83.33	56.25	0.41
Naive Bayes	Train	88.29	92.31	75.47	0.68
	Test 1	88.18	89.01	84.21	0.65
	Test 2	73.68	87.10	57.69	0.47
Ensemble learning	Train	89.63	91.48	82.61	0.70
	Test 1	89.09	90.00	85.00	0.68
	Test 2	76.32	83.56	63.41	0.48
Discriminant analysis	Train	86.94	89.77	76.09	0.63
	Test 1	87.27	88.04	83.33	0.62
	Test 2	74.56	83.10	60.47	0.45

influence on the overall prediction ability of the model (Figure S2).

Comparison of different methods

To date, several methods have been used to develop predictors.[50] One of the reasons why our model had a robust predictive ability was the use of SVM modeling. Here, we compared the predictive results for the studied inhibitors using five different classification methods, namely K-nearest neighbor,[51] random forest,[52] naive Bayes,[53] ensemble learning,[54] and discriminant analysis,[55] to provide valuable information for classification of ACE inhibitors. As summarized in Table 1, these five classification methods presented relatively favorable predicted results as the SVM model; however, the SVM model is superior to any other classification model on the prediction of two test sets. Besides, this demonstrated that the studied ACE inhibitor dataset from the Binding Database[34] had strong modeling adaptability for different classification methods.

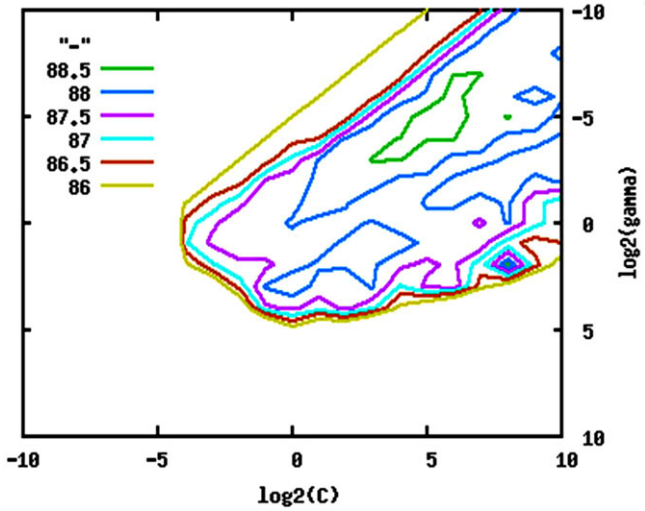

Fig. 2. Parameters by the grid search by LOO cross-validation on 444 training samples. The optimal SVM model with Acc = 88.74% was produced at C = 32 and γ = 0.03125.

Descriptor interpretation

As described above, the SVM model included four variables, i.e., capacity factor, volume, standard deviation, and hydrophilic–lipophilic balance descriptors. These variables generally describe the hydrophobic and hydrophilic properties mediated by surface properties such as shape, electrostatic interaction, hydrogen bonding, and hydrophobicity. The capacity factor[56] of molecules was calculated at the −0.2 kcal/mol energy level. It represents the ratio between the hydrophilic regions and the molecular surface, and the amount of hydrophilic regions per surface area unit.[57,58] The volume descriptors characterize the van der Waals volume

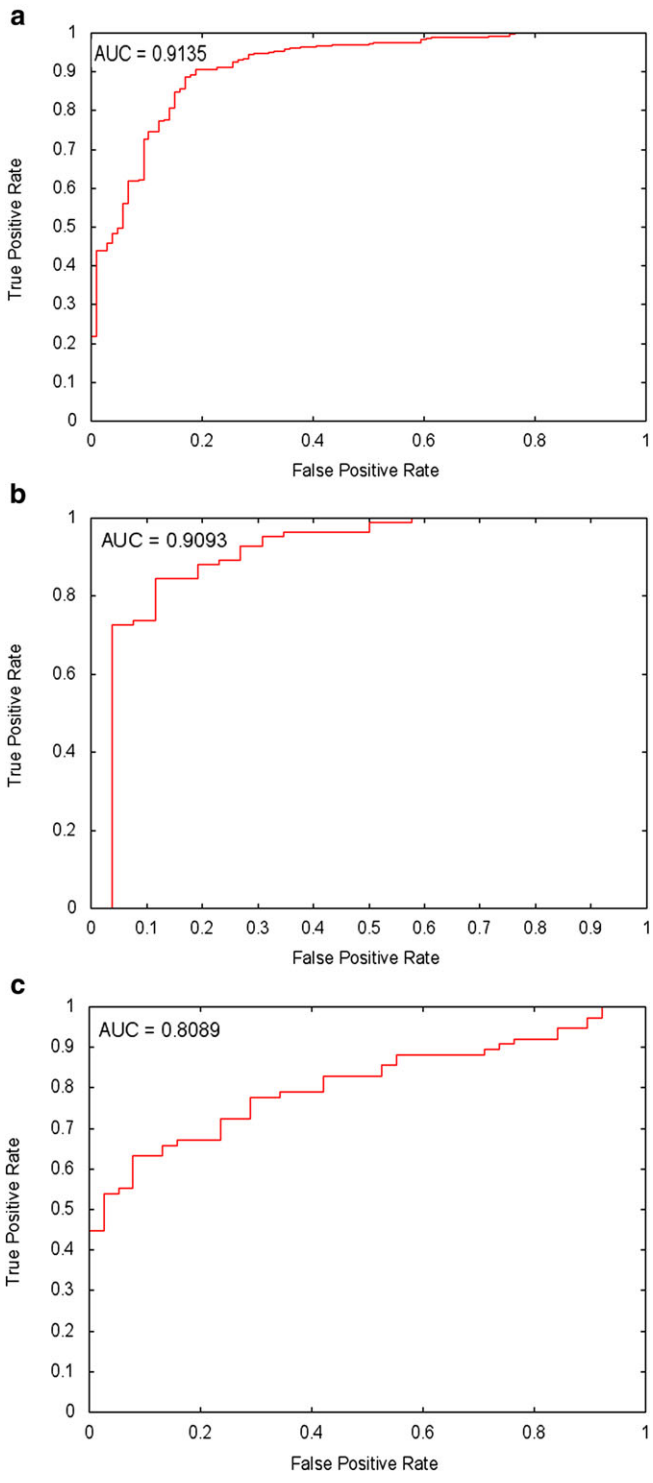

Fig. 3. Area under the receiver operating characteristic curve with LOO cross-validation on 444 training samples (a), Test 1 (110 samples) (b), and Test 2 (114 samples) (c).

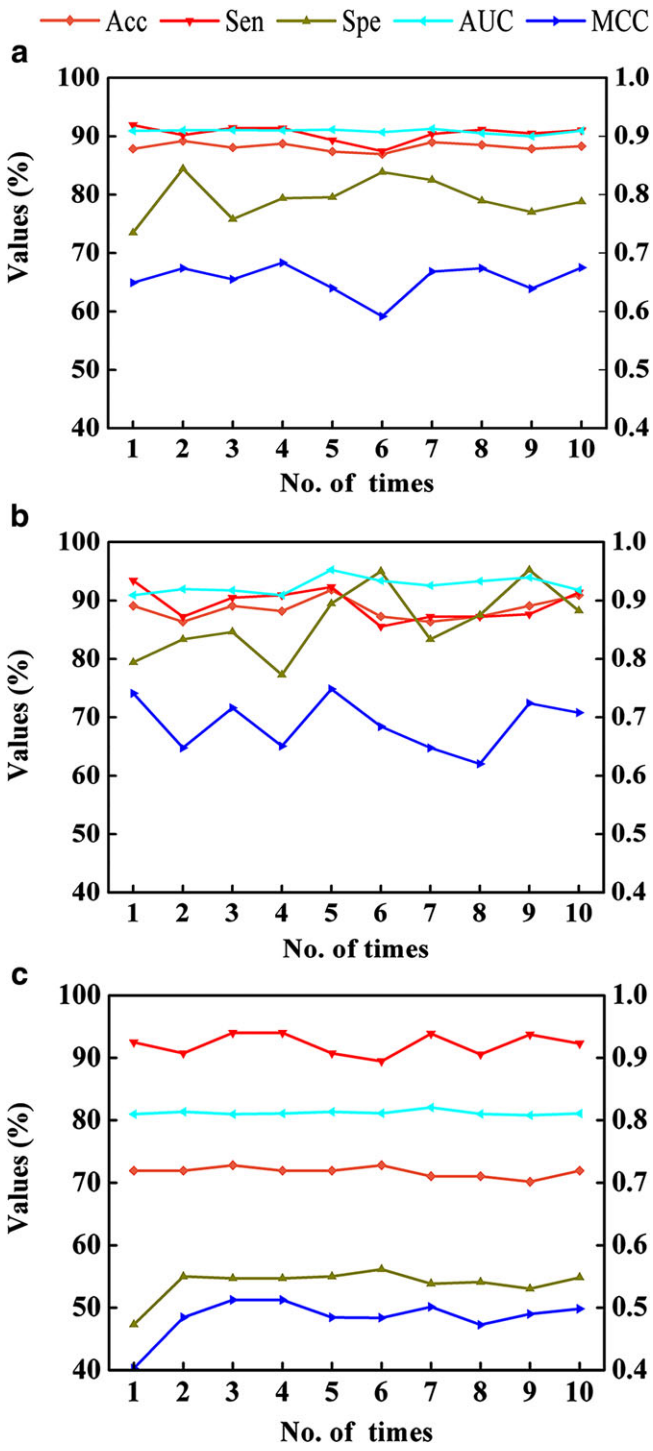

Fig. 4. Predicted results on 444 training samples (a), Test 1 (b), and Test 2 (c) for 10 times.

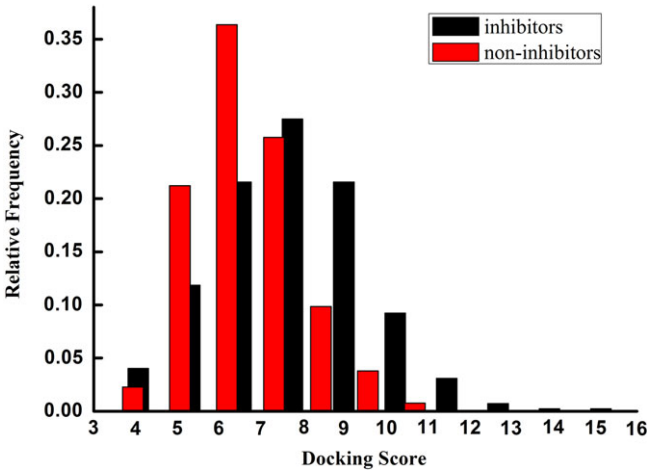

Fig. 5. Relative frequency of docking scores for inhibitors (black) and non-inhibitors (red).

calculated by a grid approximation (spacing 0.75 Å). The standard deviation descriptor is calculated by a standard dimension that describes the square root of the third largest eigenvalue with covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis. Hence, the descriptor characterizes the steric features of the studied molecules. The hydrophilic–lipophilic balance descriptor, which was calculated at the hydrophilic regions measured at −4 kcal/mol and the hydrophobic regions measured at −0.8 kcal/mol, represents the molecular polarity affected by volume, shape, and surface area. In conclusion, these four descriptors characterize the physicochemical and spatial properties of the inhibitors related to their inhibitory activities.

Binding poses of the inhibitors with ACE

As we know, the QSAR method characterizes the relationship between inhibitors and their enzyme targets only from aspects of ligand-based structures. Here, the four descriptors in the SVM model revealed that molecular features of ligands contributed to their activities. However, the structural information of the receptor may have been neglected; in other words, the interactive residues between ligands and receptor were not fully considered. Therefore, molecular docking would be applied to explore the spatial binding poses of ligands with their receptor in the following section.

Surflex-dock, as one of the most classical docking methods, has been employed to explore the interaction of inhibitors with enzymes. In this method, the docking conformation of each ligand is ranked by the total score, which consists of several functions, representing hydrophobic, entropic, polar, electrostatic, and crash terms, to make the consistent evaluation criterion and reveal the optimal conformation. The frequency distribution of the docking scores (Table S2) of the inhibitors and non-inhibitors are shown in Figure 5. Generally, the larger the docking score, the better the conformation. It can be seen that the non-inhibitors exist in the low-score region. In contrast, the inhibitors are preferentially located in the high score-region, especially where the scores are larger than 10.0; in fact, almost all inhibitors have high scores. This suggested that the inhibitors had stronger binding forces with ACE relative to the non-inhibitors.

The conformations with docking scores larger than 12.00 are shown in Figure 6. It is seen that the inhibitors are located in the ACE active site, which includes the zinc ion. We used the docking scores as independent variables to establish an SVM model by Libsvm-3.20. As a result, Acc was 79.05% and AUC was 0.72 for the 444 training samples; Acc was 79.09% and AUC was 0.67 for Test 1; and Acc was 70.18% and AUC was 0.56 for Test 2. This further demonstrated that there was a slight difference on the predictive ability between the model by four Vsurf_descriptors and the model by docking scores, which showed that docking scores cannot fully characterize the interaction between the ligand and the receptor.

To determine whether the combined model by five variables (four Vsurf_descriptors and docking scores) had better predictive ability than the independent model, we built a new SVM-based model with these five variables. The new model produced Acc of 88.29% and AUC of 0.90 on the training set; Acc of 90% and AUC of 0.95 on Test 1; and Acc of 71.93% and AUC of 0.81 on Test 2. Through comparison, we can see that the predictive ability of the model with five variables is comparable to that of the model with four Vsurf_descriptors. The SVM model-bonded docking scores involving ligands and receptor information may assist in enhancing the confidence of the SVM model based on four descriptors.

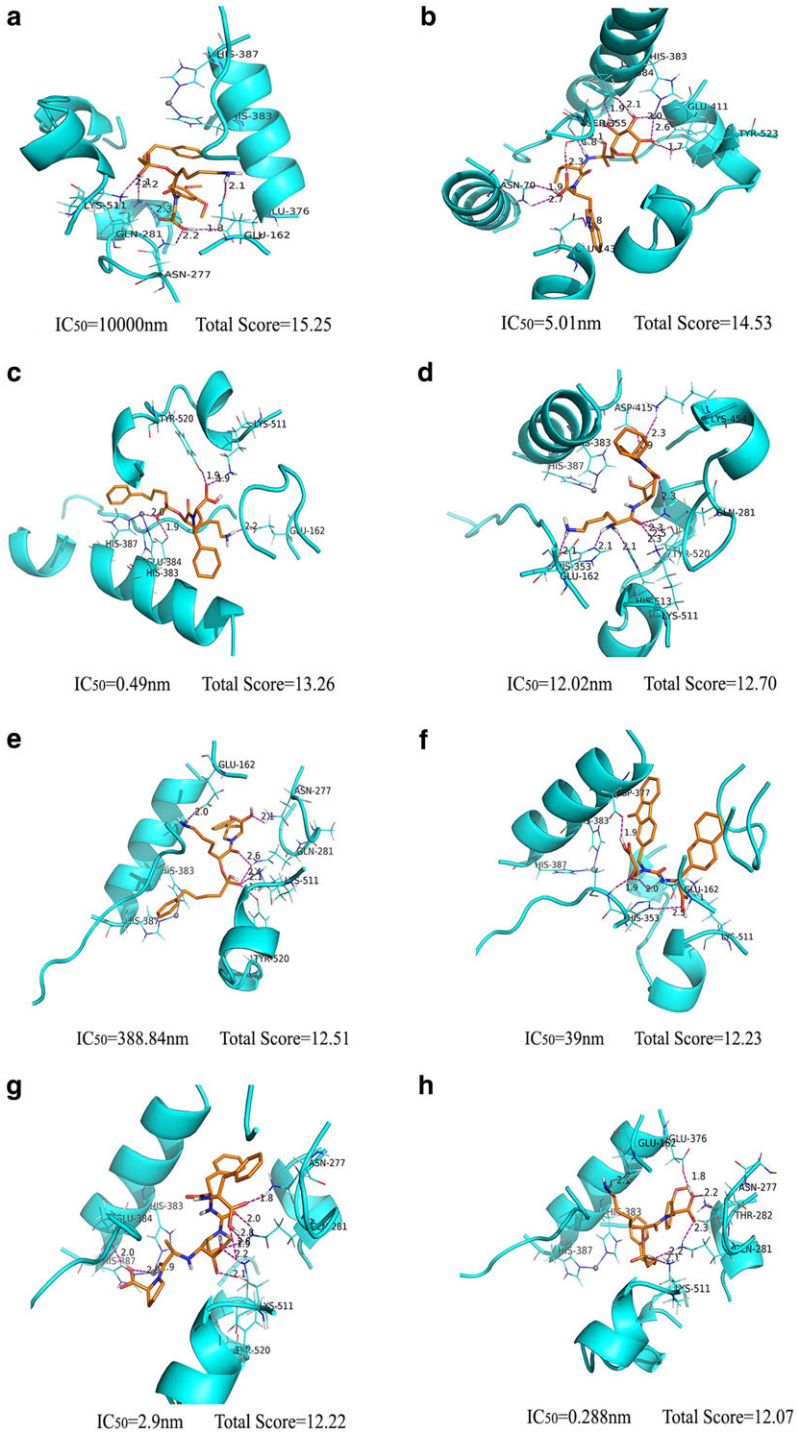

Fig. 6. Interaction between the inhibitors and ACE with docking scores larger than 12.00. Residues are rendered as line model and cartoon model in cyan; ligand is rendered as a stick model (orange: carbons; red: oxygen; blue: nitrogen); and the zinc ion is rendered as sphere in gray. The dotted lines represent the hydrogen bonds. (a) BDBM 50302107 (The Binding Database ID, the same below), (b) BDBM 50006118, (c) BDBM 50406394, (d) BDBM 50406924, (e) BDBM 50406388, (f) BDBM 50373667, (g) BDBM 50405271, (h) BDBM 50406375.

Table 2. Occurrence frequency of the residues and hydrogen bonds within 3 Å distance to ligands in the optimal 100 conformations for ACE inhibitors

Amino acid	Frequency	No. of H-bonds	Amino acid	Frequency	No. of H-bonds
TYR 523	66	11	HIS 353	26	7
TYR 520	57	34	HIS 513	24	9
HIS 383	55	5	GLU 162	19	13
GLN 281	47	19	ALA 356	17	11
HIS 387	42	9	ASP 415	16	10
GLU 384	37	26	PHE 457	11	5
LYS 511	34	16			

We observed the distribution of hydrophilic amino acids (HIS, GLU, GLN, LYS, ASP, SER, ASN, THR, ARG, etc.) and hydrophobic amino acids (TYR, ALA, PHE, VAL, TRP, PRO,LEU, etc.) within inhibitors and non-inhibitors. As displayed in Figure 7, the occurrence frequency for hydrophilic and hydrophobic amino acids within 100 inhibitors was 359 and 199, respectively. It is obvious that hydrophobic residues are fewer than hydrophilic amino acids, also within 100 non-inhibitors (290 for hydrophilic residues and 199 for hydrophobic residues). Generally, a hydrophobic region is formed in the active site of the target protein. These results suggest that there is a hydrophilic region in the ligand-binding pocket to maintain the hydrophilic–hydrophobic balance in the active site, which could well interpret the physicochemical meaning of the hydrophilic–lipophilic descriptor in the SVM model.

To further determine which residue made more contribution to the binding of the inhibitors with ACE, we computed the occurrence frequency larger than 10 for amino acids, which affected the inhibitory activity in the binding pocket (within 3 Å). The results of 100 inhibitors and 100 non-inhibitors are listed in Tables 2 and in 3, respectively. It can be seen that there are several key residues, including TYR 523, TYR 520, HIS 383, GLN 281, HIS 387, GLU 384, LYS 511, HIS 353, and HIS 513, around the active site for binding with different ligands. These molecular docking results are in agreement with previous experimental data.[59] We found that the occurrence frequencies of TYR 523 and TYR 520 were the maximum, which could be explained by the following reasoning: TYR is an aromatic hydrophobic amino acid, and the polar hydroxyl group of its side chain prefers to form hydrogen bonds with the inhibitors.

We shed light on the roles of hydrogen bonds between the inhibitors and ACE. The occurrence frequency of the hydrogen bonds within 3 Å distance to the docking conformation of 100 inhibitors and 100 non-inhibitors are shown in Figure 7. It is clearly seen that more hydrogen bonds are formed in inhibitors than in non-inhibitors. Besides, by comparison with 100 inhibitors (Table 2), we find that the frequency of the hydrogen bonds significantly decreases in the 100 non-inhibitors (Table 3) with the key residues. Therefore, we conclude that the hydrogen bonds play an important role in the binding affinities of these inhibitors with ACE.

Table 3. Occurrence frequency of the residues and hydrogen bonds within 3 Å distance to ligands in the optimal 100 conformations for non-ACE inhibitors

Amino acid	Frequency	No. of H-bonds	Amino acid	Frequency	No. of H-bonds
TYR 520	49	9	LYS 511	30	2
TYR 523	44	6	HIS 513	26	10
HIS 383	43	4	HIS 353	22	6
GLU 384	37	16	ALA 356	14	1
HIS 387	37	6	ASP 453	11	7
GLN 281	30	3	PHE 527	11	3

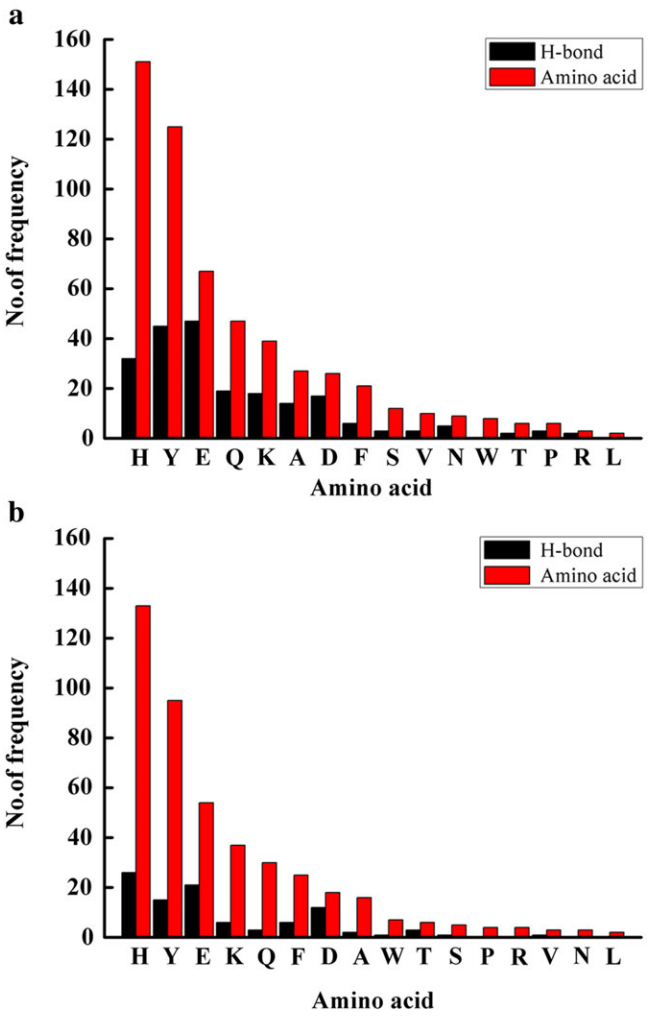

Fig. 7. Occurrence frequency of the amino acids and hydrogen bonds within 3 Å distance to the ligands. (a) 100 inhibitors, (b) 100 non-inhibitors.

CONCLUSIONS

We explored the interaction of ACE inhibitors with their receptors by combining an SVM model with molecular docking. We showed that the SVM model could achieve favorable prediction on the ACE inhibitors. This model with four descriptors is simple, accurate, and robust. Also, this model could well represent the SAR of ACE inhibitors.

Structure-based molecular docking results indicated that hydrogen bond is an important force affecting the binding affinity of the binding sites (TYR 523, TYR 520, HIS 383, GLN 281, HIS 387, GLU 384, LYS 511, HIS 353, and HIS 513) of ACE with the inhibitors. The structure–function relationships of ACE inhibitors represented by the SVM model and molecular docking are useful for screening ACE inhibitors. Our future work will focus on exploring further mechanisms of inhibitors with ACE by high-precision calculation such as molecular dynamics simulation and quantum chemistry, along with biochemical and medical experiments, which will be helpful in explore the pathogenesis of hypertensive disorders to reduce the risk of disease.

ACKNOWLEDGMENTS

We gratefully acknowledge support of this research by the National Natural Science Foundation of China (31571782 and 10901169), the Fundamental Research Funds for the Central Universities (CQDXWL-2014-Z009 and 106112016CDJZR235517), and Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ1500205).

Supporting information

Additional supporting information is available in the online version of this article.

REFERENCES

1. C. Harrison, K. R. Acharya, *J. Cell Commun. Signal.* **2014**, *8*, 195.
2. E. D. Sturrock, R. Natesh, V. R. Jm, K. R. Acharya, *Cell. Mol. Life Sci.* **2004**, *61*, 2677.
3. C. S. Anthony, G. Masuyer, E. D. Sturrock, K. R. Acharya, *Curr. Med. Chem.* **2012**, *19*, 845.
4. M. A. Zaman, S. Oparil, D. A. Calhoun, *Nat. Rev. Drug Discov.* **2002**, *1*, 621.
5. N. J. Brown, D. E. Vaughan, *Circulation* **1998**, *97*, 1411.
6. P. Vlieghe, V. Lisowski, J. Martinez, M. Khrestchatisky, *Drug Discov. Today* **2010**, *15*, 40.
7. D. J. Craik, D. P. Fairlie, S. Liras, D. A. Price, *Chem. Biol. Drug Des.* **2013**, *81*, 136.
8. M. Foltz, L. Van Buren, W. Klaffke, G. S. M. J. E. Duchateau, *J. Food Sci.* **2009**, *74*, H243.
9. J. Wu, R. E. Aluko, *J. Pept. Sci.* **2007**, *13*, 63.
10. P. Zhou, X. M. Chen, Y. Wu, Z. Shang, *Amino Acids* **2009**, *38*, 199.
11. N. F. Bras, P. A. Fernandes, M. J. Ramos, *ACS Catal.* **2014**, *4*, 2587.
12. J. Wu, R. E. Aluko, S. Nakai, *J. Agric. Food Chem.* **2006**, *54*, 732.
13. J. Wu, R. E. Aluko, S. Nakai, *QSAR Comb. Sci.* **2006**, *25*, 873.

14. C. Polanco, T. Buhse, V. N. Uversky, *Acta Biochim. Pol.* **2016**, *63*, 229.
15. H. Long, Y. Wang, Y. Lin, Z. Lin, *J. Chin. Chem. Soc.* **2010**, *57*, 417.
16. G. Z. Liang, L. Yang, L. Kang, H. C. V. Der Mei, Z. Li, *Amino Acids* **2008**, *37*, 583.
17. H. Mei, Z. H. Liao, Y. Zhou, S. Z. Li, *Biopolymers* **2005**, *80*, 775.
18. X. Wang, J. Wang, Y. Lin, Y. Ding, Y. Wang, X. Cheng, Z. Lin, *J. Mol. Model.* **2010**, *17*, 1599.
19. R. Jahangiri, S. Soltani, A. Barzegar, *Pharmaceutical sciences* **2014**, *20*, 122.
20. M. Li, D. Wei, H. Zhao, Y. Du, *Chemosphere* **2014**, *95*, 220.
21. R. D. Cramer, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.* **1988**, *110*, 5959.
22. I. A. Doytchinova, D. R. Flower, *J. Med. Chem.* **2001**, *44*, 3572.
23. J. Caballero, M. Saavedra, M. Fernandez, F. D. Gonzalez-Nilo, *J. Agric. Food Chem.* **2007**, *55*, 8101.
24. S. Wu, W. Qi, R. Su, T. Li, D. Lu, Z. He, *Eur. J. Med. Chem.* **2014**, *84*, 100.
25. A. A. S. Juan, S. Cho, *Bull. Kor. Chem. Soc.* **2005**, *26*, 952.
26. R. D. Cramer, *J. Med. Chem.* **2003**, *46*, 374.
27. S. Kim, M. Chi, C. Yoon, H. Sung, *J. Biochem. Mol. Biol.* **1998**, *31*, 459.
28. M. Arooj, S. Thangapandian, S. John, S. Hwang, J. K. Park, K. W. Lee, *Int. J. Mol. Sci.* **2011**, *12*, 9236.
29. W. M. Czarnecki, S. Podlewska, A. J. Bojarski, *J. Cheminform.* **2015**, *7*, 38.
30. S. K. Kushwaha, P. Chauhan, K. Hedlund, D. Ahren, *Bioinformatics* **2016**, *32*, 1223.
31. Z. Zhou, C. Cheng, Y. Li, *SAR QSAR Environ. Res.* **2015**, *26*, 1001.
32. D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat. Rev. Drug Discov.* **2004**, *3*, 935.
33. J. Ren, L. Li, R. Zheng, H. Xie, Z. Cao, S. Feng, Y. Pan, X. Chen, Y. Wei, S. Yang, *J. Chem. Inf. Model.* **2011**, *51*, 1364.
34. T. Liu, Y. Lin, X. Wen, R. N. Jorissen, M. K. Gilson, *Nucleic Acids Res.* **2007**, *35*, D198.
35. S. A. Depriest, D. Mayer, C. B. Naylor, G. R. Marshall, *J. Am. Chem. Soc.* **1993**, *115*, 5372.
36. G. Cruciani, P. Crivori, P. Carrupt, B. Testa, *J. Mol. Struc.-Theochem* **2000**, *503*, 17.
37. P. Labute, *J. Mol. Graph. Model.* **2000**, *18*, 464.
38. C. Chang, C. Lin, *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27.
39. V. Vapnik, *IEEE Trans. Neural Netw.* **1999**, *10*, 988.
40. M. Pardo, G. Sberveglieri, *Sensor Actuat. B Chem.* **2005**, *107*, 730.
41. G. H. Fu, D. S. Cao, Q. S. Xu, H. D. Li, Y. Z. Liang, *J. Chemometr.* **2011**, *25*, 92.
42. J. Zhang, S. Wang, *Neural Comput. & Applic.* **2015**, *27*, 1717.
43. H. Kaur, G. P. S. Raghava, *Bioinformatics* **2004**, *20*, 2751.
44. J. M. Deleo, F. Pucino, K. A. Calis, K. W. Crawford, T. E. Dorworth, J. F. Gallelli, *Am. J. Health Syst. Pharm.* **1993**, *50*, 2348.
45. B. W. Matthews, *Biochim. Biophys. Acta* **1975**, *405*, 442.
46. C. Ruizsamblas, S. Medinarodriguez, V. Quirosrodriguez, A. M. Jimenezcarvelo, L. Valverdesom, A. Gonzalezcasado, L. Cuadrosrodriguez, *Anal. Methods* **2015**, *7*, 4192.
47. Y. Ai, S. Wang, P. Sun, F. Song, *Int. J. Mol. Sci.* **2011**, *12*, 1605.
48. R. Natesh, S. L. U. Schwager, E. D. Sturrock, K. R. Acharya, *Nature* **2003**, *421*, 551.
49. N. Guex, M. C. Peitsch, *Electrophoresis* **1997**, *18*, 2714.
50. S. Pirhadi, F. Shiri, J. B. Ghasemi, *RSC Adv.* **2015**, *5*, 104635.
51. D. Coomans, D. L. Massart, *Anal. Chim. Acta* **1982**, *136*, 15.
52. L. Breiman, *Mach. Learn.* **2001**, *45*, 5.
53. R. D. King, C. Feng, A. D. Sutherland, *Appl. Artif. Intell.* **1995**, *9*, 289.
54. L. Rokach, *Artif. Intell. Rev.* **2010**, *33*, 1.
55. W. J. Krzanowski, *J. Classif.* **1993**, *10*, 128.
56. H. Z. Si, K. J. Zhang, Z. D. Hu, B. T. Fan, *QSAR Comb. Sci.* **2007**, *26*, 41.
57. N. S. H. N. Moorthy, M. J. Ramos, P. A. Fernandes, *RSC Adv.* **2011**, *1*, 1126.
58. N. S. H. N. Moorthy, M. J. Ramos, P. A. Fernandes, *SAR QSAR Environ. Res.* **2012**, *23*, 521.
59. A. T. Girgih, R. He, R. E. Aluko, *J. Agric. Food Chem.* **2014**, *62*, 4135.

