JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

Combined 3D-QSAR, Molecular Docking, and Molecular Dynamics Study on Piperazinyl-Glutamate-Pyridines/Pyrimidines as Potent $P2Y_{12}$ Antagonists for Inhibition of Platelet Aggregation

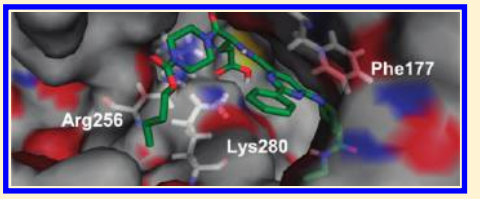
Ming Hao,[†] Yan Li,[*,†] Yonghua Wang,[‡] Yulian Yan,[†] and Shuwei Zhang[†]

[†]Department of Materials Science and Chemical Engineering, Dalian University of Technology, Dalian, Liaoning, 116023, China

[‡]College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, 712100, China

Supporting Information

ABSTRACT: An unusually large data set of 397 piperazinyl-glutamate-pyridines/pyrimidines as potent orally bioavailable $P2Y_{12}$ antagonists for inhibition of platelet aggregation was studied for the first time based on the combination of three-dimensional quantitative structure−activity relationship (3D-QSAR), molecular docking, and molecular dynamics (MD) methods. The comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) studies have been performed with a training set of 317 compounds, estimating three superimposition methods. The best CoMFA and CoMSIA models, derived from superimposition I, shows leave-one-out cross-validation correlation coefficients (Q^2) of 0.571 and 0.592 as well as the conventional correlation coefficients (R^2_{ncv}) of 0.814 and 0.834, respectively. In addition, the satisfactory results, based on the bootstrapping analysis and 10-fold cross-validation, further indicate the highly statistical significance of the optimal models. The external predictive abilities of these models were evaluated using a prediction set of 80 compounds, producing the predicted correlation coefficients (R^2_{pred}) of 0.664 and 0.668, respectively. The key amino acid residues were identified by molecular docking, and the stability and rationality of the derived molecular conformations were also validated by MD simulation. The good concordance between the docking results and CoMFA/CoMSIA contour maps provides helpful clues about the rational modification of molecules in order to design more potent $P2Y_{12}$ antagonists. We hope the developed models could provide some instructions for further synthesis of highly potent $P2Y_{12}$ antagonists.



1. INTRODUCTION

As the most frequent cause of death in the world today, cardiovascular and cerebrovascular diseases are often a consequence of an arteriosclerotic process with a complex background. Atherosclerosis is a diffuse process that starts early in childhood and progresses asymptomatically through adult life. Later in life, it is clinically manifested as coronary artery disease, stroke, transient ischemic attack, and peripheral arterial disease.[1,2] Approaches that retard or even reverse atherosclerotic lesion development in humans for prevention of plaque disruption and acute coronary events include better control of risk factors, such as reducing the plasma cholesterol levels. Nevertheless, even when the atherosclerotic plaque disruption cannot be prevented, a beneficial effect of antiplatelet and anticoagulant agents has been observed in the prevention of acute coronary events. Currently, the most widely used antiplatelet agents are aspirin, dipyridamole, glycoprotein IIb/IIIa antagonists, and thienopyridines.[3] However, these drugs also have several limitations. For example, in patients with vascular disease aspirin was less efficacious than clopidogrel. For the latter, despite its effectiveness, clopidogrel negatively affects its clinical efficacy. The active metabolite of clopidogrel irreversibly and selectively inhibits the adenosine diphosphate (ADP) receptor. Once the receptor is activated, the drug will irreversibly bind to platelets. Consequently, clopidogrel has a slow onset and slow offset of pharmacological action.[4,5] For example, in some acute settings, such as suffering from bleeding or a trauma and requiring emergency surgery, it is difficult for clopidogrel to deal with. In addition, for those people who are resistant to the effects of clopidogrel,[6] this drug will be also less helpful. Therefore, intensive efforts have been devoted toward the discovery of innovative antiplatelet agents that are more potent and safe, direct acting, and orally bioavailable, which would be paramount to meeting the demands of the urgently needed therapeutic weapons against the ADP receptor.

ADP is an important platelet agonist, where even minute concentrations can induce rigid platelet aggregate formation.[7] Its crucial roles in modulating the thrombosis and hemostasis were rapidly recognized,[8] but the molecular identity of its receptors remained for a long time elusive. $P2Y_1$ and $P2Y_{12}$, as two important P2Y receptors, moderate the effect of ADP on platelets. The G_q-coupled $P2Y_1$ receptor[9] is responsible for a transient rise in Ca^{2+}, while the G_i-coupled $P2Y_{12}$ receptor[10] refers to the inhibition of adenylate cyclase, and both of them are implicated in the transduction of the ADP signal. For normal platelet aggregation,[11] it requires simultaneous activation of the G_q and G_i pathways by ADP. Among them, activation of the G_q pathway through $P2Y_{12}$ causes the changes of the platelet shape and a

Received: June 20, 2011
Published: September 16, 2011

rapidly reversible wave of platelet aggregation. While activation of the G_i pathway mediated by $P2Y_{12}$ magnifies the G_q-mediated responses and alone leads to slowly progressive and lasting aggregation. In addition, the $P2Y_{12}$ receptor can not only inhibit the adenylate cyclase and the subsequent reduction in intracellular cAMP content but also activate the glycoprotein IIB/IIIa integrin through a phosphoinositide 3-kinase pathway and/or another unidentified G protein. Therefore, $P2Y_{12}$ receptor has an important role in the irreversible wave of platelet aggregation that occurs on exposure to ADP. Experimental studies have shown that selective blockade of either receptor is enough to inhibit the platelet activation. However, $P2Y_{12}$ is a platelet specific receptor, whereas $P2Y_1$ is ubiquitous expression, which makes $P2Y_{12}$ a more promising therapeutic target for selective modulation of ADP-induced platelet activation.

The $P2Y_{12}$ gene, which encodes the 342 amino acid receptor, was recently identified.[10] It will be a great advantage for discovery of novel antiplatelet drugs if we know the structure of $P2Y_{12}$. However, as a membrane protein, it is very difficult to achieve this aim with experimental approaches. As an alternative, the quick development of molecular modeling methods has provided us with unprecedented power to those investigations in both the fundamental and industrial researches.[12] In light of this, Costanzi and co-workers[13] constructed an architecture of P2Y nucleotide receptors by homology to bovine rhodopsin. Based on the built P2Y receptor, they also performed a docking research. More recently, Deflorian and Jacobson[14] have compared three homology models of $P2Y_{12}$ based on different G-protein-coupled receptor (GPCR) structural templates including bovine rhodopsin, human A_{2A} adenosine receptor and human C−X−C chemokine receptor type 4. Taking into account all available pharmacology and mutagenesis researches, it can be concluded that C−X−C chemokine receptor type 4 based structure of $P2Y_{12}$ is the most accurate one, which will be helpful for the further research.

Since $P2Y_{12}$ has been identified a promising target to antiplatelet drugs, recently many classes of ADP receptor antagonists including prasugrel,[15] ticlopidine,[16] ticagrelor,[17] cangrelor,[18] 2-alkylthio-substituted analogues,[19] 6-amino-2-mercapto-3H-pyrimidin-4-one derivatives,[20] piperazinyl-glutamate-pyridines/pyrimidines-based antagonists,[21−24] piperazinyl-glutamate-quinolines,[25−27] anthraquinones,[28] adenosine analogues,[29] piperazinyl-pyridine ureas,[30] etc. were synthesized, some of which achieved promising inhibition potency, and others were currently in late-stage clinical trials. In spite of these encouraging advances, the development of antiplatelet drugs still remains a difficult task for researchers. To facilitate the drug discovery process, in silico modeling approaches[31−35] as a productive and cost-effective technology in design of novel lead compounds should be used in combination with experimental practices.[36−39] In view of this, Fujita's group has carried out excellent work to study the P2Y receptor using both the in silico and experimental methods.[40−42] In addition, our group also used two-dimensional quantitative structure−activity relationship (2D-QSAR) to predict series of $P2Y_{12}$ antagonists using a novel genetic algorithm-support vector machine coupled approach.[43] As we know, in many cases, 2D-QSAR often focuses on the predictive models[44−48] in which gaining an intuitive interpretation of important features from this 2D-QSAR study is not always simple. And it is also difficult to present a comprehensive feature for the ligand−receptor interactions, such as the hydrophobic contact, polar interactions between the key amino residues and agents.

Thus in the present work, a total of 397 piperazinyl-glutamate-pyridines/pyrimidines-based antagonists of $P2Y_{12}$ were collected to build 3D-QSAR models using comparative molecular field analysis (CoMFA)[49] and comparative molecular similarity indices analysis (CoMSIA) methods.[50] The reliability and robustness of the developed best models were estimated with bootstrapping analysis and 10-fold cross-validation. And the predictive abilities of the obtained models were validated statistically with an external test set of compounds. In addition, a combined computational approach including the docking analysis and molecular dynamics (MD) simulation was also employed to elucidate the probable binding modes of these antagonists at the binding site of the $P2Y_{12}$ receptor. The good concordance between the 3D contour maps and the docking result provides our identification of several key features of the binding mechanism for these piperazinyl-glutamate-pyridines/pyrimidines-based antagonists. Until recently, to our best knowledge, this is the first 3D-QSAR report for this set of compounds, and we hope the developed models could provide some meaningful clues in the future synthesis of highly potent and orally bioavailable $P2Y_{12}$ antagonists.

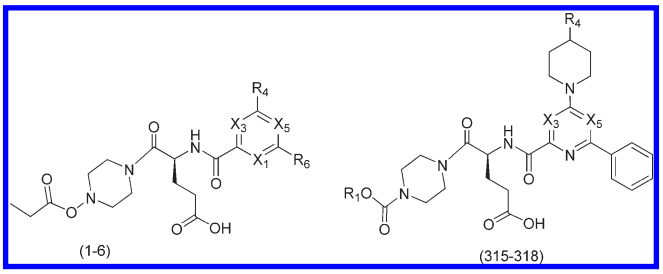
2. METHODS AND COMPUTATIONAL DETAILS

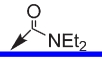
Data Set and Biological Activity. Eliminating those compounds with unspecified antagonistic activity and redundant structures, a total of 397 piperazinyl-glutamate-pyridines/pyrimidines as potent $P2Y_{12}$ antagonists for inhibition of platelet aggregation was collected from the continuous work of Parlow et al and used as the data set for molecular modeling in this study.[21−24] In this work, the in vitro biological activities of these compounds were converted into the corresponding pK_i ($-\log K_i$) values, which were used as dependent variables in the current QSAR analyses. The collected compounds were divided into a training set (including 317 compounds) and a test set (including 80 compounds) in an approximate ratio of 4:1. All structures and the K_i values of the data set against $P2Y_{12}$ receptor as well as their belongings to the training or test set are listed in Table S1 (Supporting Information). The principle for selection of the test set chemicals was to ensure that both their pK_i values are randomly but uniformly distributed in the range of the values for the whole set, and their structures cover as large a diversity as possible of the data set. Table 1 depicts several representative skeletons and K_i values (nM) of the data set.

Molecular Modeling. All molecular modeling and 3D-QSAR studies were performed using the SYBYL6.9 molecular modeling software package (Tripos Associates, St. Louis, MO). Partial atomic charges were calculated by the Gasteiger−Huckel method,[51] and energy minimization and conformational search were performed using Tripos molecular mechanics force field[52] by the Powell method with a convergence criterion of 0.05 kcal/mol·Å with the maximum iterations set to 2000.

Molecular Alignment. Molecular alignment of compounds is a crucial step in the development of 3D-QSAR models.[53] It was assumed that each molecule binds into the active site in a similar mode, since these compounds share a common scaffold. To derive the best possible 3D-QSAR statistical model, three different superimposition rules were investigated. Initially, the ligand-based superimposition (superimposition I) using database alignment technology was adopted. During the process, the most potent antagonist (compound 316) was chosen as a template to fit the remaining training and test sets of compounds by using the substructure-alignment function available in SYBYL. Figure 1A describes the common substructure for the alignment, which is marked in bold. Figure 1B shows the resulting superimposition

Table 1. Representative Structures and Binding Activities for a Set of 397 $P2Y_{12}$ Antagonists with Diverse Structures



Compd.	Substituent					K_i (nM)	Ref.
	R_4	R_6	X_1	X_3	X_5		
1	Ph	Ph	N	CH	CH	15	[21]
2*	H	Ph	N	CH	CH	209	[21]
3	Ph	Ph	CH	CH	CH	373	[21]
4	H	Ph	CH	CH	CH	3640	[21]
5	Ph	Ph	CH	CH	N	484	[21]
6*	H	Ph	N	CH	N	356	[21]
	X_3	X_5	R_1	R_4			
315	CH	N	Pent			2.5	[23]
316	CH	N	Bu			0.38	[23]
317	CH	N	Pent			0.60	[23]
318*	CH	N	Bu			0.79	[23]

*Denotes a test set.

I model. Superimposition II is the docking-based alignment which is shown in Figure 1C. The conformation with the highest total score for each ligand of data set was aligned automatically together inside the binding pocket of $P2Y_{12}$ and used directly for CoMFA and CoMSIA research. Superimposition III is a common scaffold-based alignment same as superimposition I but with all molecular conformations obtained from the superimposition II. And the final superimposition is shown in Figure 1D.

CoMFA and CoMSIA Field Calculation. The CoMFA and CoMSIA models were generated by using SYBYL 6.9 with the default parameters. Detailed algorithms of CoMFA and CoMSIA can be easily referred to many literatures,[54,55] thus we only introduce the modeling parameters in this article.

To derive the CoMFA and CoMSIA descriptor fields, the aligned training set molecules were placed in a 3D cubic lattice ($18 \times 15 \times 14 = 3780$ grid points) with grid spacing of 2 Å in *x*, *y*, and *z* directions such that the entire set was included in it. In CoMFA, the steric and electrostatic fields were calculated separately for each molecule using sp^3 carbon as the probe atom with a charge of +1.00 (default probe atom in SYBYL) and energy cutoff values of 30 kcal/mol for both the steric and electrostatic fields. The CoMFA fields generated automatically were scaled by the CoMFA-STD method in SYBYL. CoMSIA similarity index descriptors were derived according to Klebe et al.[50] with the same lattice boxes as used in CoMFA calculations. In CoMSIA, the steric, electrostatic, hydrophobic, and hydrogen bond (H-bond) donor and acceptor descriptors were calculated using a probe atom of radius 1.0 Å, charge +1.0, and hydrophobicity +1.0. A Gaussian function was used to evaluate the mutual distance between the probe atom and each molecule atom. Because of the different shape of the Gaussian function, CoMSIA similarity indices (A_F) for molecule j with atom *i* at grid point *q* were calculated by equation

$$A^q_{F,k}(j) = -\sum \omega_{probe,k}\omega_{ik} e^{-\alpha r^2_{iq}} \quad (1)$$

where *k* represents the steric, electrostatic, hydrophobic, or H-bond donor or acceptor descriptor; $\omega_{probe,k}$ is the probe atom with radius 1.0 Å, charge +1, hydrophobicity +1, H-bond donating +1, and H-bond accepting +1; ω_{ik} is the actual value

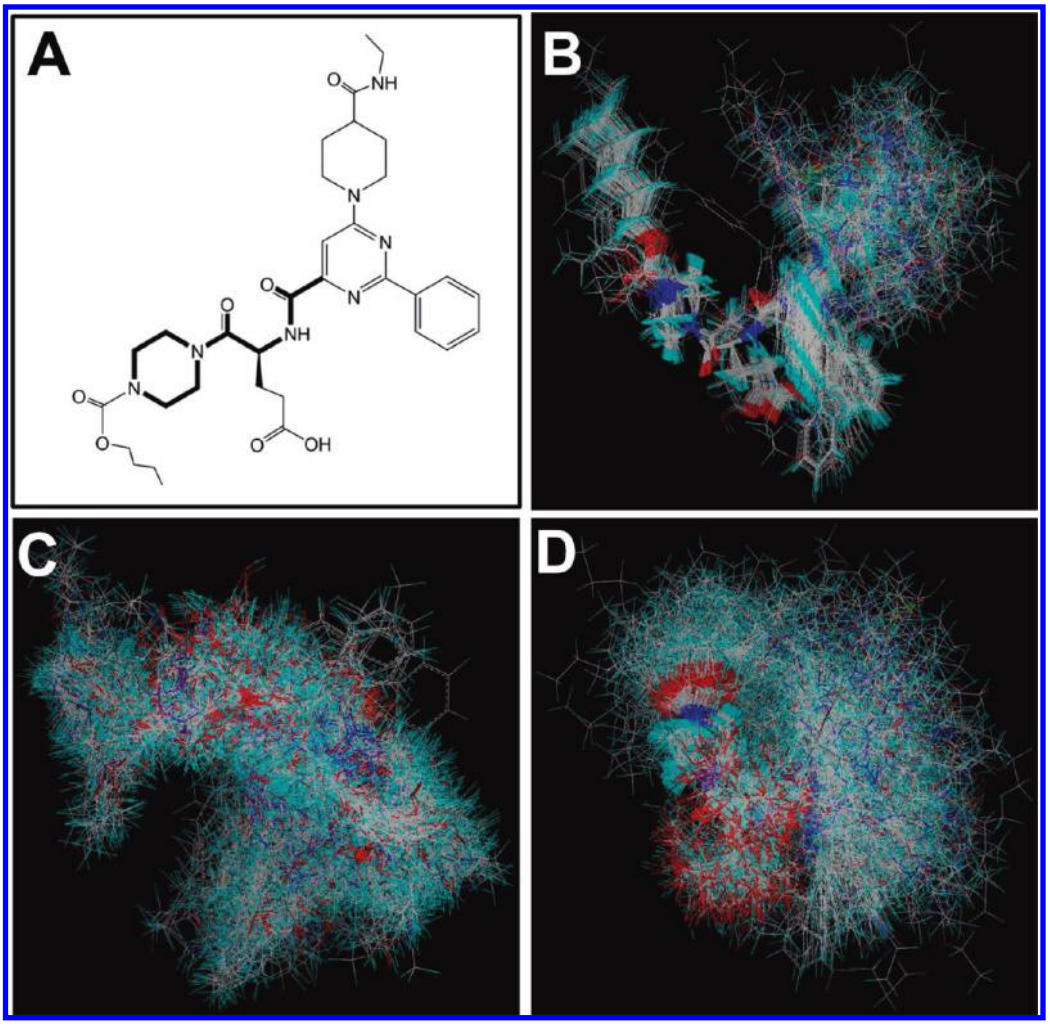


Figure 1. The alignment of all molecules in the data set. (A) Compound 316 is used as the template for ligand-based alignment. (B−D) present the alignments from the superimposition I, II, and III, respectively.

of the physicochemical property k of atom i; and r_{iq} is the mutual distance between the probe atom at grid point q and atom i of the test molecule. The default value of 0.3 was used as attenuation factor (α).

Molecular Docking. Docking simulations of piperazinyl-glutamate-pyridines/pyrimidines into the $P2Y_{12}$ binding pocket were performed using the Surflex-dock module of SYBYL package in this study. Surflex-dock utilizes a so-called "whole" molecule alignment algorithm based on morphological similarity between the ligand and target.[56] The docking method aligns the ligand to a "protomol" or idealized ligand in the active site of the target. The detailed algorithm for Surflex-dock is presented in the literature.[57] For our studies, the coordinate of $P2Y_{12}$ was obtained from Jacobson's homology modeling.[14] In this article, the $P2Y_{12}$ homology model has been successfully established based on the human C−X−C chemokine receptor type 4.[14] Prior to the docking, the ligand and substructure were extracted from the crystal structure, and hydrogen atoms were added to the protein in standard geometry using the biopolymer modulators. In this study, automatic-based mode was adopted to generate the protomol in the Surflex-dock program, and two parameters that significantly affect the size and extend of the protomol generated are the threshold and the bloat values. In the present work, the threshold and bloat values were set to 0.5 and 0, respectively. During the present molecular docking process, the protein was considered to be rigid, and the ligand molecules were flexible. Other parameters were established by default values in the software. In the current work, 20 conformations were obtained though Surflex-dock for each ligand, and all conformations were extracted from the optimized antagonist−$P2Y_{12}$ complex. The conformations with the highest total scores for each ligand of the training set were aligned automatically together inside the binding pocket of $P2Y_{12}$ and used directly for CoMFA and CoMSIA analyses to explore the interactions between the ligands and $P2Y_{12}$.

MD Simulations. The MD simulations were performed with GROMACS software package[58] using the GROMOS96 force field.[59] The molecular topology file for the ligand in protein was generated by the program PRODRG 2.5.[60,61] The simulation cell was a cubic periodic box with a side length of 114.32 Å, and the minimum distance between the protein and box walls was set to larger than 10 Å. In order to neutralize the total charge, 22 chloridion ions were placed randomly in the box. The total number of the atoms of the simulation system was 149 736, including the protein complexes and waters. The remaining box volume was filled using the simple point charge (SPC) water.

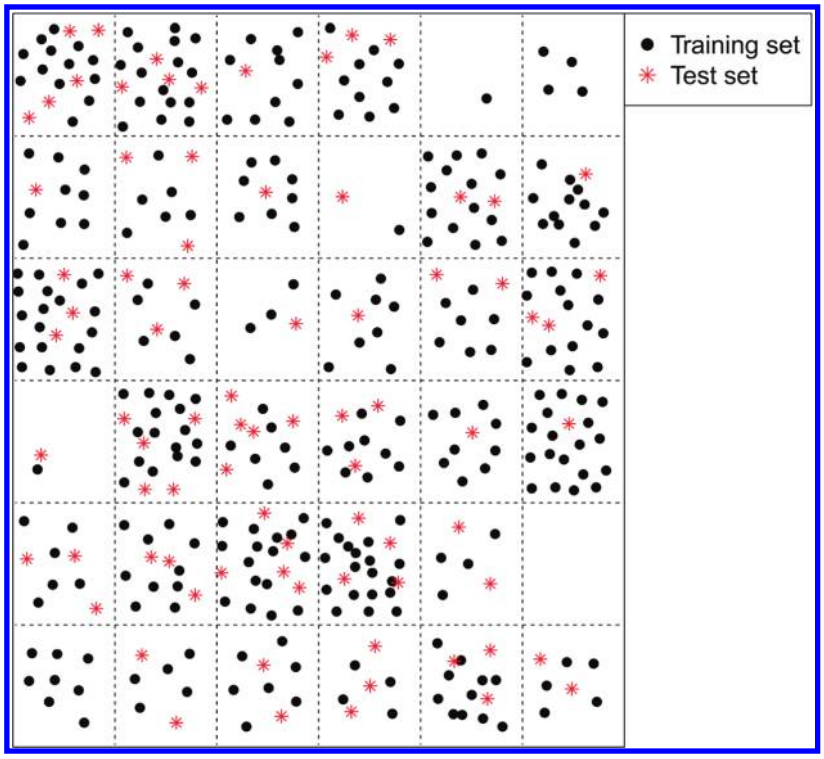


Figure 2. SOM top map indicating the distribution of the training and external prediction sets. The training set is labeled in black dot and the prediction set in red asterisk.

Prior to the simulation, an energy minimization was applied to the full system without constraints using the steepest descent integrator for 9896 steps, then the system was equilibrated via a 200 ps MD simulation at 300 K. Finally, a 5 ns simulation was performed with a time step of 2 fs. During MD simulation, the standard parameters and main calculation methods were set as follows: The model used NPT ensemble at 300 K with periodic boundary conditions, the temperature was kept constant by the Berendsen thermostat, the values of the isothermal compressibility were set to 4.5×10^{-5} bar^{-1} while the pressure was maintained at 1 bar using the Parrinello−Rahman scheme,[62] electrostatic interactions were calculated using the particle mesh Ewald method,[63,64] and cutoff distances for the calculation of Coulomb and van der Waals interactions were 1.0 and 1.4 nm, respectively. All the MD simulations lasted 5 ns to ensure that the whole systems were stable.

3D-QSAR Models Calculation and Statistical Validation. In order to generate statistically significant 3D-QSAR models, partial least-squares (PLS) regression was used to analyze the training set by correlating the variation in their pK_i values (the dependent variable) with variations in their CoMFA/CoMSIA interaction fields (the independent variables). PLS is a statistical approach that generalizes and combines features from principal component analysis and multiple regressions.[65,66] It is particularly useful to predict a set of dependent variables from a large set of independent variables and when the matrix of predictors has more variables than observations.

To evaluate the reliability of the models generated from the PLS analysis, cross-validation analysis was accomplished with the leave-one-out (LOO) methodology, wherein one compound was moved away from the data set and its activity was predicted by the model derived from the rest of the data set. A cross-validated correlation coefficient, Q^2 (also called R^2_{cv}), was subsequently obtained and provided as a statistical index of the predictive power. Then, a noncross-validation analysis was carried out, and the Pearson coefficient (R^2_{ncv}) and standard error of estimates (SEE) were calculated based on the obtained optimal principal components by cross-validation. Finally, the CoMFA/CoMSIA results were graphically represented by field contour maps, where the coefficients were generated using the field type "StDev*Coeff". The cross-validated coefficient, Q^2, was calculated using the following equation:

$$Q^2 = 1 - \frac{\sum_{i=1}^{train} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{train} (y_i - \overline{y}_{tr})^2} \quad (2)$$

where y_i, $\hat{y}_i$, and $\overline{y}_{tr}$ are the observed, predicted, and mean values of the target property (pK_i), respectively, for the training set.

In order to evaluate the real predictive ability of the best models generated by the CoMFA/CoMSIA analyses using the training set, the pK_i values of 80 compounds were treated as the external validation set. A predictive R^2_{pred} value was then obtained with the following formula:

$$R^2_{pred} = 1 - \frac{PRESS}{SD} \quad (3)$$

where PRESS is the predictive residual sum of squares for the test set and SD denotes the sum of squared deviation between the biological activities of the test set molecules and the mean activity of the training set molecules.

3. RESULTS AND DISCUSSION

Split the Training and Test Sets. Rational selection of training and test sets is one of the important and challenging steps for the development of validated QSAR models.[67,68] A basic rule to implement such selection is to guarantee that the points of the training set are distributed evenly within the whole area occupied by representative points and that the closeness condition of the test set points to the training ones is satisfied.[69] In the current work, to investigate the descriptor space, totally 929 2D descriptors were calculated using the Dragon software package (Dragon Professional, version 5.4) for each $P2Y_{12}$ antagonist. Then these original Dragon molecular descriptors were undertaken a preprocessing process as follows: (1) those descriptors containing larger than 85% zero values were removed; (2) zero and near-zero variance predictors were removed because such descriptors may cause the model to crash or the fit to be unstable; and (3) one of the two descriptors that has the absolute correlation above 0.95 was omitted. After these steps, the number of the original descriptors was reduced to 237 for further self-organizing map (SOM) research.

As a special kind of neural network that can be used for clustering, visualization, and abstraction tasks, SOM is especially suitable for data survey due to its prominent visualization properties, as illustrated by our previous work of data set splitting.[70,71] SOM creates a set of prototype vectors representing the data set and carries out a topology preserving projection of the prototypes from the d-dimensional input space onto a low-dimensional grid, which is a convenient visualization space for showing the cluster structure of the data. In the present work, the construction of the training and test sets was made based on the SOM visualization of the whole data.[72] A small Kohonen network with 6 * 6 = 36 neurons was employed, producing a map with 36 positions. All the compounds with 237 molecular descriptors were placed onto the 36 positions (neurons) of the Kohonen map. The distribution of the molecules is demonstrated in Figure 2, where the training set is labeled in black dot and the prediction set in red asterisk, respectively. The purpose we performed the SOM simulation on the data set is to guarantee that the representative points of the training set are, on one hand, distributed evenly within the whole area of the descriptor space occupied by the data set and, on the other hand, close to those of the test set, which ensures the reliability of the simulation results. In addition, it should be noted that the biological activity distribution is a crucial factor to the selection of training and test sets. For the present work, the training and test sets both present a uniform distribution with average activity of 8.022 and 8.026, respectively. All these results indicate that the split of the data set is rational.

CoMFA and CoMSIA Statistical Results. To judge whether a QSAR model is reliable fitting for prediction of unknown molecules, several statistical parameters including especially the cross-validated correlation coefficient (Q^2), noncross-validated correlation coefficient (R^2_{ncv}), and standard error of estimate (SEE), and F statistic values as well as the optimum number of components (OPN) should be evaluated. For the $P2Y_{12}$ 3D-QSAR studies, good correlations were observed in the obtained CoMFA and CoMSIA models demonstrated by the high values of Q^2 and other statistical results. Table 2 summarizes the statistical results of the CoMFA and CoMSIA analyses.

During the molecular modeling process of pK_i, 317 compounds out of the total 397 $P2Y_{12}$ receptor antagonists were used as the training set, and the remaining 80 compounds were used as test set to validate the developed models. As it is known, the CoMFA and CoMSIA models are alignment sensitive, with the quality and the predictive ability of the models directly dependent on the alignment rules,[73] and differences in the statistical values are observed with different superimpositions. As seen from Table 2, superimposition I, in both CoMFA and CoMSIA models, gives the better statistical results than those obtained from superimpositions II and III. Thus, our main analysis was restricted to the superimposition I model for the prediction of $P2Y_{12}$ antagonists, which presents Q^2 = 0.571 with six principal components, R^2_{ncv} = 0.814, SEE = 0.301, and F value of 225.527 for CoMFA model. It is noted that for this model, the steric feature is found to make slightly higher contribution (55.7%) to the $P2Y_{12}$ antagonist activity than that of the electrostatic feature (44.3%).

When building the CoMSIA models, the data sets are the same as those used in establishing the CoMFA ones. Since the five descriptors (steric, electrostatic, hydrophobic, and H-bond donor and acceptor) are not completely independent of each other and since such dependency among individual fields may reduce the model significance and generalization,[65,74] all 31 possible descriptors' combinations were calculated with their respective Q^2 value using the optimum number of components, for estimating

Table 2. Optimal CoMFA and CoMSIA Results Based on Different Superimposition Methods[a]

	Superimposition Methods					
	I		II		III	
PLS analysis	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
Q^2	0.571	0.592	0.161	0.330	0.144	0.281
OPN	6	8	4	5	5	5
R^2_{ncv}	0.814	0.834	0.644	0.539	0.740	0.635
SEE	0.301	0.284	0.414	0.472	0.354	0.420
F value	225.527	193.930	141.111	72.581	177.444	108.121
R^2_{bs}	0.850	0.880	0.770	0.620	0.841	0.733
SEE_{bs}	0.269	0.240	0.331	0.430	0.277	0.359
$R^2_{cv(mean)}$	0.570	0.589	0.149	0.320	0.123	0.259
R^2_{pred}	0.664	0.668	0.172	0.282	0.179	0.358
	Relative Contribution (%)					
S	55.7	22.1	39.1	100	53.8	57.7
E	44.3	28.6	60.9	–	46.2	–
H	–	30.0	–	–	–	–
D	–	–	–	–	–	–
A	–	19.3	–	–	–	42.3

[a] Q^2, cross-validated correlation coefficient after the LOO procedure. OPN, optimal number of principal components. R^2_{ncv}, noncross-validated correlation coefficient. SEE, standard error of estimate. F, the value of F statistic. R^2_{bs}, the average R^2 value from a bootstrapping analysis for 100 runs. SEE_{bs}, the average SEE value from a bootstrapping analysis for 100 runs. $R^2_{cv(mean)}$, the average R^2_{cv} from 10 times 10-fold cross-validation. R^2_{pred}, predicted correlation coefficient for the test set of compounds. Superimposition method: I, from the database alignment; II, from docking alignment; and III, from database alignment based on the docking conformations.

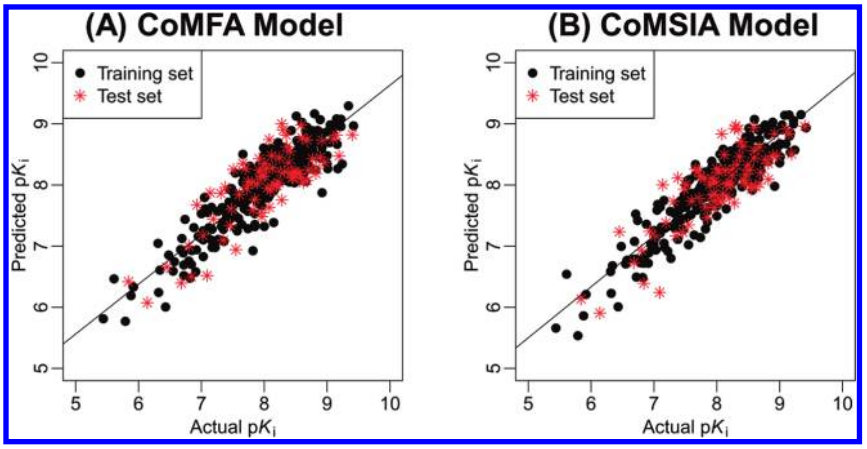


Figure 3. Predicted versus actual pK_i for the optimal (A) CoMFA and (B) CoMSIA models based on the superimposition I. The training set is marked in black dot and test set marked in red asterisk.

which one of the five CoMSIA fields is actually needed for the generation of a predictive model. Finally, the model with steric, electrostatic, hydrophobic, and H-bond acceptor fields appears to be superior among all the models derived. For the optimal CoMSIA model based on the superimposition I, it gives the statistical results of $Q^2 = 0.592$, $R^2_{ncv} = 0.834$, SEE = 0.284, and $F = 193.930$ based on the use of eight optimum components. In this case, the steric, electrostatic, hydrophobic, and H-bond acceptor fields contribute 22.1, 28.6, 30, and 19.3%, respectively.

In addition to the LOO cross-validation, the cross-validation in groups using 10 groups repeating the procedure 10 times was also carried out. The mean of 10 readings was given as $R^2_{cv(mean)}$. To further assess the robustness and statistical confidence of the derived models, a bootstrapping analysis for 100 runs was performed. Here, we just report the average R^2 and SEE values of these 100 analyses (namely, R^2_{bs} and SEE_{bs}, respectively). Bootstrapping involves the generation of many new data sets from the original data set and is obtained by randomly choosing of samples from the original data set. The statistical calculation is performed on each of these bootstrapping samplings. The difference between the parameters calculated from the original data set and the average of the parameters calculated from the many bootstrapping samplings is a measure of the bias of the original calculations. Table 2 illustrates all corresponding results, where the R^2_{bs} and SEE_{bs} values are 0.850 and 0.269 for the best CoMFA model as well as 0.880 and 0.240 for the best CoMSIA model based on the superimposition I, respectively, proving the robustness of the present models. Besides, the $R^2_{cv(mean)}$ values for both the CoMFA and CoMSIA models also present 0.570 and 0.589, respectively, indicating highly statistical significance.

As a further test of robustness of the CoMFA and CoMSIA models, we also randomized the target values for 50 times. As a result, none of those models had significant Q^2. The Q^2 obtained were in the range from −0.204 to 0.01 for CoMFA and from −0.214 to 0.014 for CoMSIA, respectively. This indicates that the Q^2 in both the optimal CoMFA and CoMSIA models with original data is not due to chance correlations.

Validation implies a quantitative assessment of the robustness and predictive power of a QSAR model. This predictive power can be defined as the model's capability to predict accurately the modeled property of new compounds. In general, a $Q^2 > 0.5$ is considered proof of the acceptable internal predictive ability. It is also reported that the high value of Q^2 appears to be the necessary but not the sufficient condition for the model to have a high predictive power. Thus, the external validation becomes a necessary way to establish a reliable QSAR model.[75] For the present work, an external test set including 80 compounds was used to validate the model, which exhibits the R^2_{pred} of 0.664 and 0.668 for the CoMFA and CoMSIA models based on the superimposition I, respectively.

Outliers from a QSAR are compounds that do not fit the model or are poorly predicted by it.[76] Many reasons may account for the presence of outliers from a QSAR. Typically, however, such compounds have been recognized as acting by a different mechanism of action from other compounds which may be well modeled by the QSAR. When performed correctly, removal of significant outliers will allow for the development of stronger and more significant models. Thus, one should perform an outlier check test in the built models. However, in the present work, after establishment of the models, no outliers are found in both the training and test sets based on the optimal superimposition I models, despite that they are built based on an unusually large of 397 $P2Y_{12}$ antagonists. Figure 3 depicts the correlations between the experimental and the predicted activities for both the training and test sets for the optimal CoMFA and CoMSIA models, respectively. Clearly, a good agreement between the predicted activities and experimental data was observed, and all the points are rather uniformly distributed around the regression line, suggesting the satisfactory predictive capability of the models. The predicted values by the two optimal CoMFA and CoMSIA models are shown in Table S2 (Supporting Information).

Contour Maps. The optimum CoMFA and CoMSIA models from superimposition I are selected to construct the StDev*Coeff contour maps to view the field effects on the target features. To aid in visualization, compound 316, as the most active $P2Y_{12}$ antagonist in the series, is shown superimposed with the CoMFA and CoMSIA contour maps as depicted in Figures 4 and 5.

CoMFA Contour Maps. The CoMFA result is usually represented as 3D coefficient contour. It shows regions where variations of the steric and electrostatic nature in the structural features of the different molecules contained in the training set cause the increase or decrease in the activity. For steric fields, Figure 4A shows the contour map of sterically favored (green) and disfavored (yellow) regions. A large green-colored contour is mapped near the distal position of butyl carbamate on the piperazine nitrogen, suggesting that bulkier groups are favored

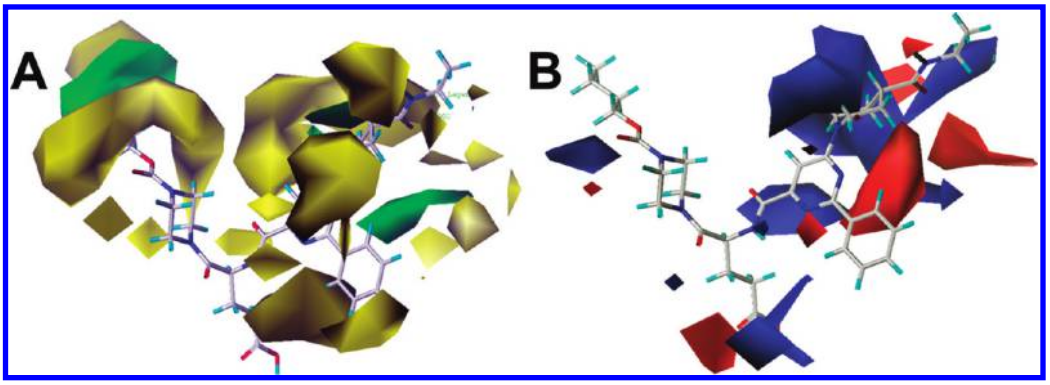


Figure 4. CoMFA StDev*Coeff contour plots with the combination of compound 316. (A) Steric contour map. Green contours indicate regions where bulky groups increase activity (favored level 80%), and yellow contours indicate regions where bulky groups decrease activity (disfavored level 20%). (B) Electrostatic contour map. Red contours indicate regions where negative charges increase activity (disfavored level 20%), and blue contours indicate regions where positive charges increase activity (favored level 80%).

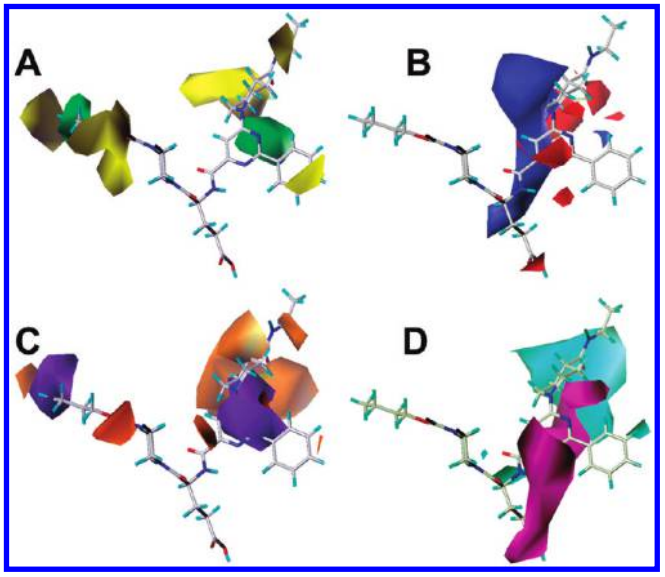


Figure 5. CoMSIA StDev*Coeff contour plots with the combination of compound 316. (A) Steric contour map. Green contours indicate regions where bulky groups increase activity (favored level 0.02), and yellow contours indicate regions where bulky groups decrease activity (disfavored level −0.002). (B) Electrostatic contour map. Red contours indicate regions where negative charges increase activity (disfavored level 20%), and blue contours indicate regions where positive charges increase activity (favored level 80%). (C) Hydrophobic contour map. Purple contours indicate regions where hydrophobic substituents enhance activity (favored level 80%), and orange contours indicate regions where hydrophobic substituents decrease activity (disfavored level 20%). (D) H-bond acceptor contour map. Magenta contours indicate regions where H-bond acceptor substituents increase activity (favored level 80%), and cyan contours indicate the disfavor regions for H-bond acceptor groups (disfavored level 20%).

at the position. However, it can be noted that at the back of the green contour there is also a yellow-colored map. Therefore for this position, a careful selection of the carbon chain length is required. The fact that compound 316 (having −Bu at this position) is more potent in activity than compound 317 (with −Pent at the same location) is a good example. And it is the same case for the comparisons between compounds 318 and 319 and compounds 297 and 298 as well as compounds 308 and 309. Around the carbonyl close to the piperazine ring, it is noted that there is a semicircle yellow-colored map indicating a large group in this position is disadvantageous to the inhibitory activity. This can be supported by the fact that compound 21 with an −*i*Pr substituent shows about 10-fold potency less than its counterpart 14 with a −Pr group. Another example is that compound 11 with a larger substituent (*m*-tolyl) toward this spatial distribution is less active than compound 3 with a relatively small substituent. Thus, it can be concluded that in this position, an aliphatic straight chain of four carbon length linked to the carbamate is beneficial to the binding activity, while when a branching with the bulk steric is introduced into the carbamate side chain, the binding activity is disfavored. In addition, a green contour map is also observed around the benzene ring linked to the pyrimidine, indicating that its preference for a large substitute here. This may explain the phenomenon why that almost all $P2Y_{12}$ antagonists in the data set contain the large benzene ring in this position. In fact, this conclusion is fully consistent with observations obtained from ref 21, who systematically compared the effect of changing the substitutes in this position by experiments and found that if the −Ph was replaced with the H atom the binding potency will be sharply reduced. It should be noted that, some yellow contours are located at the distal part of amide group, showing the substituent with the bulky steric is detrimental to the binding potency. By comparing compounds 316 and 318 and compounds 326 and 327 as well as compounds 328 and 329, it is found that the presence of a larger group is harmful to the activity.

As shown in Figure 4B, the electrostatic contour map with the combination of compound 316 from the CoMFA analysis is depicted. The large red polyhedron surrounding the terminal carboxylic acid of the glutamate depicts their favor for the electronegative substituent. The fact that compounds 8 and 9 with the low binding potency contain no electro-riched substituent, such as $-(CH_2)_2COOH$, at this position proves this finding. Moreover, at o-position of the pyrimidine, a small red contour is located, indicating that the negatively charged groups are favored at this position. Compound 1 having an N atom instead of the C at this position exhibits higher activity than that of its counterpart compound 3. Additionally, a large red contour is found near the N atom at p-position on the pyrimidine ring suggesting that electronegative substituents are favorable for the binding activity. This is why compounds 264, 266, 268, 270, 272, 274, and 275 (with the electronegative N atom instead of the

corresponding C) present larger potency than their counterparts 355, 369, 372, 373, 366, 368, and 359, respectively. Since there are several red-cord maps surrounding the piperidine ring, the introduction of electron-rich groups in these positions is beneficial to improve the inhibitory potency. A blue-colored map is also noted near the carbonyl group on the piperazine ring, which denotes that electropositive groups are desirable at this position for the increase of $P2Y_{12}$ antagonist inhibitory activity.

CoMSIA Contour Maps. Since similar results are obtained for the steric and electrostatic contours of CoMSIA as those of the CoMFA model, only the characteristics of the hydrophobic and H-bonding interaction fields are described here.

For the CoMSIA hydrophobic field contour map as depicted in Figure 5C, the purple contours are the regions where hydrophobic substituents are favorable for the $P2Y_{12}$ inhibition, whereas the orange contours are unfavorable. The appearance of purple contours at the terminal butyl carbamate indicates that substitution by hydrophobic groups at this location is extended to result in a higher $P2Y_{12}$ antagonist activity. Moreover, at the back of the purple-colored contour, there is an orange map, which indicates too long an alkyl chain entering this position would be unfavorable for the activity. For example, most of highly active compounds 314, 316, and 318 possess hydrophobic −Bu groups which create a hydrophilic environment around this contour, while their counterparts 315, 317, and 319 having larger −Pent groups instead of −Bu exhibit relatively low potency. Another large purple contour is located around the benzene ring, suggesting the occupancy of this contour by hydrophobic group would be favorable for the activity. The presence of a large orange contour map around the carbonyl group linked to the piperazine suggests that its occupancy by hydrophilic groups would favor the inhibitory activity. Compounds 11 and 12 (having hydrophobic *m*-tolyl groups at the position) present low activities, while other molecules with polar carbonyl group toward this contour possess larger potency. As seen in Figure 5C, the orange contour falls at the terminal amide group linked to the piperidine ring, suggesting the occupancy of this contour by hydrophilic group would be favorable for the activity as exemplified by the most potent compound 316. On the contrary, several chemicals such as 277, 298, 304, and 305 with nonpolar substitutes exhibit decreased activity.

The H-bond acceptor contour map of the CoMSIA model in the presence of the most potent compound 316 is depicted in Figure 5D. A magenta contour around the carbonyl group of the glutamate represents the increased activity of compounds if they have a H-bond acceptor group at this position, which on one hand is in accordance with the observations that most active compounds in the data set contain a H-bond acceptor group, such as −COOH, on the other hand indicates that a H-bond acceptor group at this location may be a necessity for strong binding activities. However, in the poorly active compounds 8 and 9, the substituent is a hydrogen atom which cannot form a H-bond with the $P2Y_{12}$ receptor close to the magenta contour. As seen from Figure 5D, the large magenta contour extends from the carbonyl group to the pyrimidine ring, indicating that H-bond acceptor groups in these positions are favorable for the $P2Y_{12}$ inhibitory potency. For example, compounds 290, 295, 312, 314, 319, 321, and 323 with H-bond acceptor at the p-position on the pyrimidine depict a larger potency than their counterparts 333, 337, 340, 346, 350, 352, and 354 without such a substituent in this position. Moreover, there is one large magenta-colored contour overlapping the O atom of amide linked to

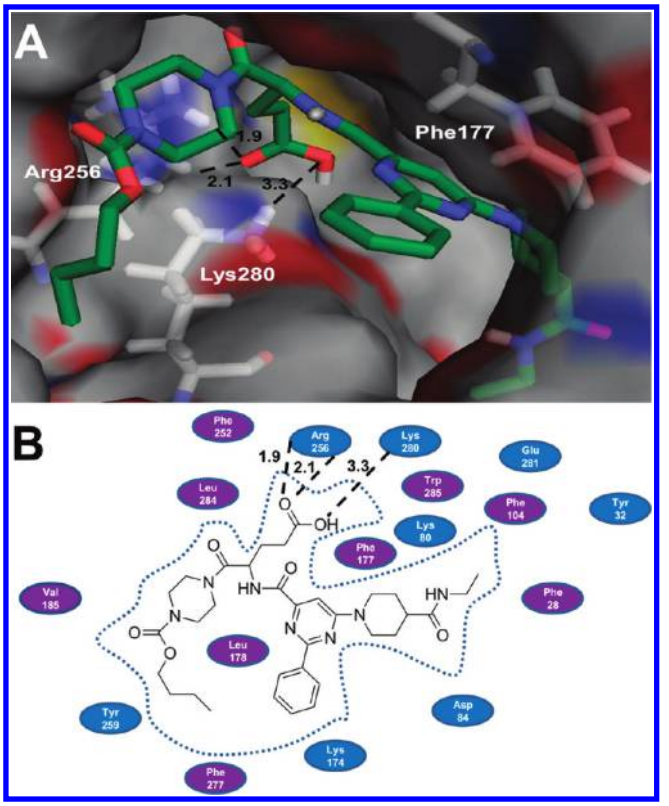


Figure 6. Molecular docking model of compound 316 in the binding site of $P2Y_{12}$. (A) A surface rendering to illustrate the interactions between compound 316 with the representative amino acids. The dashed lines show the formation and distance of the H-bonds. Compound 316 is depicted as carbon chain in bright-green. (B) The active site residues are represented as follows: polar residues in blue and hydrophobic residues in purple, respectively.

the pyrimidine of compound 316, which means that this group is favored in this region and will lead to an increase of the inhibitory potency. This may be the reason why most compounds in this database possess this structure which seems to play a key role in their predominantly positive effect on $P2Y_{12}$ inhibitory activity. In Figure 5D, the large cyan contour around −NH group of amide linked to the piperidine indicates that presence of strong H-bond donor is beneficial to the biological activity. This is in agreement with the experimental results that compound 326 with a H-bond donor group of −NHEt gives a higher activity than compound 327 without the H-bond donor substituent ($-NEt_2$). It is the same case for the compounds 317 and 319 as well as for compounds 316 and 318.

Molecular Docking. The docking results could illustrate us the interaction modes between antagonists and the adenosine diphosphate receptor ($P2Y_{12}$). In the present work, the most potent compound 316 is taken as the representative to elucidate this point in detail. As illustrated in Figure 6, compound 316 locates at the active site of the receptor and binds primarily through the hydrophobic and H-bond interactions. A narrow, long hydrophobic cavity is formed among Val185, Leu178, and Phe277 which accommodates a long chain hydrophobic substituent, and this conclusion is also inferred from both the steric green and hydrophobic purple maps of the CoMSIA model. It should also be noted that at the back of terminal butyl carbamate on the piperazine nitrogen, there exists the side chain of Phe277.

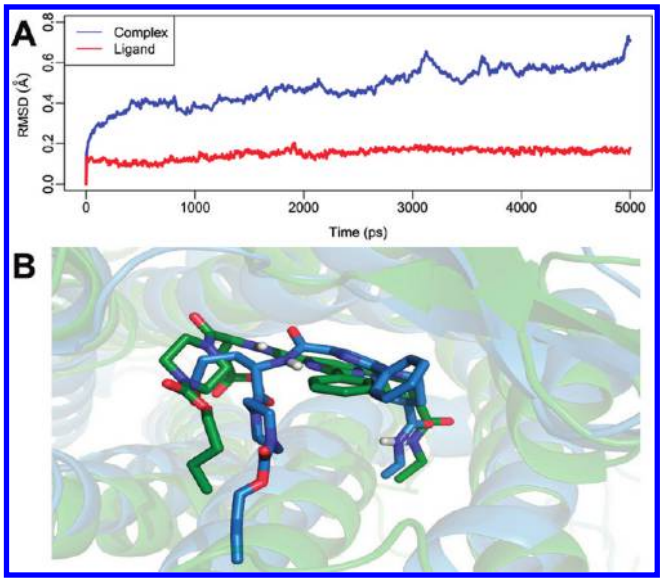
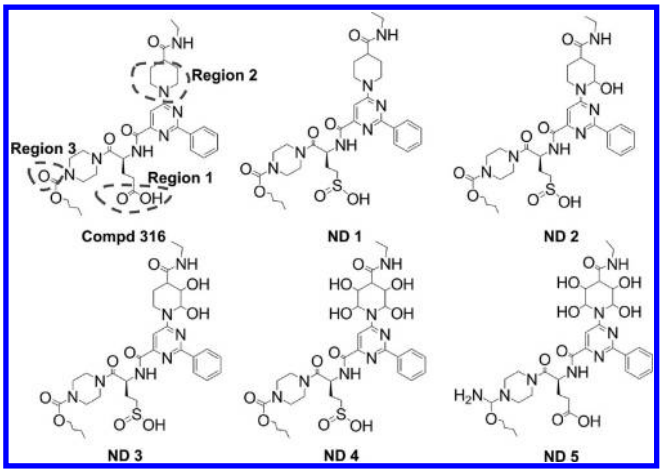


Figure 7. MD simulation results. (A) Plot of the rmsd of the docked complex versus the MD simulation time in the MD-simulated structure. (B) View of the superimposed backbone atoms of the average structure of the MD simulation (light-blue) and the initial structure (bright-green) for compound 316-P2Y$_{12}$ complex. Compound 316 is represented as carbon chain in bright-green and light-blue for the initial and average complexes, respectively.

Therefore, substituents with too long straight chain at this position are unfavorable to the activity probably due to the crashing with this residue. Thus, it can be concluded that those substituents containing a linear alkyl chain (with a four to five carbon length) can result in an increase of the activity. This is in agreement with our CoMFA and CoMSIA steric maps. Another hydrophobic center is formed between Leu178, Phe177, and Lys174, thus a large substituent, such as the benzene ring, in this position will increase the activity. This observation is fully supported by our CoMFA and CoMSIA steric maps results. Additionally, the aromatic ring of Phe177 anchors the aromatic moieties of the pyridine and phenyl rings of the ligand through $\pi-\pi$ stacking. The third hydrophobic interaction is observed at the distal of the piperidine ring of compound 316 including Glu281, Phe104, Trp285, Tyr32, Phe28, Lys80, and Asp84. In this hydrophobic pocket, one can notice that most of residues include aromatic rings indicating a hydrophobic substituent linked to the tail of piperidine could increase the inhibition potency. In addition, at the terminal ethyl amide, there are three residues of Phe28 and Leu284 blocking the bulky substituents in this position, which is consistent with our 3D-QSAR results also. Compounds 321, 322, and 315 with ring and branching structures show less activity than that of 316 with a relatively small −Et group, which is just a good illustration.

Besides the important role of hydrophobic interactions, several H-bond interactions formed between the ligand and P2Y$_{12}$ also attract our attention. For example, the carbonyl O atom as a H-bond acceptor forms two H-bonds with two N atoms of Arg256 with the bond lengths of 1.9 and 2.1 Å, respectively. Another H-bond interaction can also be formed between the residue Lys280 and −OH of the ligand with the bond length of 3.3 Å. Thus, the antagonist is stabilized in the binding pocket by H-bond interactions between the carboxyl group of the compound and the residues Arg256 and Lys280. It can also be noted

Table 3. Structures and CoMFA and CoMSIA Predicted Activity of New Designed P2Y$_{12}$ Antagonists



	predicted pK_i	
no.	CoMFA	CoMSIA
compd. 316	8.967	8.934
ND 1	9.865	9.123
ND 2	10.058	9.484
ND 3	10.117	9.800
ND 4	10.422	10.265
ND 5	9.556	10.185

that Tyr259 is in close proximity of the esther group attached to the piperazine ring. All these findings can be supported by the previous report.[14] In this docking simulation, two amino acid residues (Arg256 and Lys280) have been identified to play an important role in the binding interactions with the ligands, which can also be supported by previous published site-directed mutagenesis study.[77]

Molecular Dynamic Simulation. Presently, a 5 ns simulation of the docked complex structure of P2Y$_{12}$ with antagonist 316 was performed to obtain a dynamic picture of the conformational changes that occur in an aqueous solution, with main emphasis to explore the conformational change that takes place in the antagonist 316 and P2Y$_{12}$ receptor.

Figure 7A shows the root-mean-square deviation (rmsd) of the trajectory for the complex with respect to the initial structure (in blue line), and the graph presents that the rmsd reaches about 0.6 Å, which suggests that a relatively stable conformation of the protein is achieved through the MD simulation. Figure 7A also gives the rmsd of the ligand 316 (in red line) in the binding site of P2Y$_{12}$. It can be clearly noted that the rmsd for the ligand reaches about 0.1 Å from the beginning of MD simulation and retains this value throughout the simulation, suggesting that the changes of complex are mainly caused by the protein. In fact, it can be noticed that the P2Y$_{12}$ used in the present work has a long tail,[14] which may be one probable reason why the rmsd values present a little change at the end the MD simulation. In order to compare the structures from MD simulations and docking, a superimposition of both the structures in the last 2 ns is shown in Figure 7B, where the bright-green ribbon represents the initial structure for the docked complex, the light-blue ribbon represents the average structure, with compound 316 represented as carbon chain in

bright-green for the initial complex and carbon chain in light-blue for the average complex, respectively. It can be noticed that from this figure, basically, there is no significant difference between the average structure extracted from MD simulations and the docked model of the complex, especially at the right part of the pyrimidine ring. One can also notice that the butyl carbamate part (light-blue) presents a slight movement out of the docking one (bright-green). Although the complex has undergone several movements during MD simulation, both the binding pocket and the conformation of the ligand are still stable, suggesting rationality and validity of the docking model.

Design of New $P2Y_{12}$ Antagonists. According to the detailed contour analyses of CoMFA and CoMSIA models, some useful information on the structural requirements for the observed inhibitory activity is obtained. Based on this information, five new pyrimidines analogues showing enhanced predicted inhibitory activity compared with the most potent compound 316, which is used as a reference molecule, were designed by us. At present, the modified parts are focused on three regions (i.e., region 1, region 2, and region 3) in the reference molecule. As shown in Table 3, all the designed molecules show better pK_i than the reference molecule 316. Since the present work focuses on the computational investigation of the SAR of piperazinyl-glutamate-pyridines/pyrimidines derivatives as $P2Y_{12}$ antagonists and the exploration of the interaction mechanism between $P2Y_{12}$ and its antagonists by modeling results, the validation of the inhibitory activity of the novel-designed five compounds needs further experimental evaluations.

4. CONCLUSIONS

Presently, for the first time a large data set of 397 piperazinyl glutamate pyridines/pyrimidines as potent orally bioavailable $P2Y_{12}$ antagonists for the inhibition of platelet aggregation has been estimated for the purpose of developing 3D-QSAR models based on both the ligand- and receptor-based superimpositions. Statistically significant models have been derived with two 3D-QSAR methods of CoMFA and CoMSIA on the basis of the database alignment method. These two approaches produce equally good models expressed in terms of several rigorous evaluation criteria, such as Q^2 and R^2_{pred}, for both the internal and external data sets. Graphical interpretation of the optimal results, provided by the CoMFA and CoMSIA analyses, brings to light important structural features that could be responsible for the low- or high-bonding activity $P2Y_{12}$ antagonism: (i) Substituents with a proper length and size of carbamate (straight chain alkyl group with four-carbon) on the piperazine nitrogen are beneficial for the increase of potency; (ii) a bulky group at the distal of the amide linked to the piperidine ring is likely to decrease the $P2Y_{12}$ bonding; (iii) substituents with electro-riched groups at the o- and p-position on the pyrimidine are favored, and the carbonyl group of glutamate should be required for the ligand binding; (iv) the introduction of hydrophobic substituents at the terminal of the butyl carbamate on the piperazine nitrogen would enhance the binding; (v) H-bond acceptor groups near the carbonyl oxygen atom of glutamate and nitrogen atoms at p-position on the pyrimidine could result in larger binding affinity; (vi) furthermore, the key amino residues to the ligand−$P2Y_{12}$ interaction have been found, i.e., Arg256 and Lys280, which form three important H-bonding interactions between the ligand and the target.

In addition, a good consistency between the CoMFA and CoMSIA contour maps, molecular docking, and molecular dynamics simulations proves the reliability and the robustness of the developed models.

Overall, in this report, several reliable computation models between piperazinyl glutamate pyridines/pyrimidines and $P2Y_{12}$ have been built, which not only exhibit satisfied statistics but also provide several possible mechanism interpretations from a molecular level. We hope the models may provide an insight into some instructions for further synthesis of highly potent $P2Y_{12}$ antagonists.

■ ASSOCIATED CONTENT

Supporting Information. Table S1 containing the chemical structures and binding activity for a set of 397 $P2Y_{12}$ antagonists. Table S2 listing the actual and predicted values based on the optimal CoMFA and CoMSIA models from superimposition I, II, and III. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: yanli@dlut.edu.cn. Telephone: +86-411-84986062.

■ ACKNOWLEDGMENT

This work is financially supported by the National Natural Science Foundation of China (Grant no. 10801025). The authors thank Dr. Kenneth A. Jacobson (Laboratory of Bioorganic Chemistry & Molecular Recognition Section, NIDDK, U.S.A.) for his kindness in offering us the PDB file with the model $P2Y_{12}$−CXC in complex with Pfizer compound 47s as in their published paper (*J. Comput.-Aided Mol. Des.* **2011**, *25*, 329). The authors are also very grateful to Prof. Ling Yang for access to SYBYL software.

■ REFERENCES

(1) Viles-Gonzalez, J. F.; Fuster, V.; Badimon, J. J. Atherothrombosis: a widespread disease with unpredictable and life-threatening consequences. *Eur. Heart J.* **2004**, *25*, 1197–1207.

(2) Fuster, V.; Badimon, J. J.; Chesebro, J. H. Atherothrombosis: mechanisms and clinical therapeutic approaches. *Vasc. Med.* **1998**, *3*, 231–239.

(3) Horiuchi, H. Recent advance in antiplatelet therapy: The mechanisms, evidence and approach to the problems. *Ann. Med.* **2006**, *38*, 162–172.

(4) Meadows, T.; Bhatt, D. Clinical aspects of platelet inhibitors and thrombus formation. *Circ. Res.* **2007**, *100*, 1261–1275.

(5) Savi, P.; Pereillo, J.; Uzabiaga, M.; Combalbert, J.; Picard, C.; Maffrand, J.; Pascal, M.; Herbert, J. Identification and biological activity of the active metabolite of clopidogrel. *Thromb. Haemostasis* **2000**, *84*, 891–896.

(6) Gurbel, P.; Bliden, K.; Hiatt, B.; O'Connor, C. Clopidogrel for coronary stenting: response variability, drug resistance, and the effect of pretreatment platelet reactivity. *Circulation* **2003**, *107*, 2908–2913.

(7) Hall, M. W.; Goodman, P. D.; Solen, K. A.; Mohammad, S. F. Formation of occlusive platelet aggregates in whole blood caused by low concentrations of ADP. *ASAIO J.* **2000**, *46*, 693–695.

(8) Maffrand, J.; Bernat, A.; Delebassee, D.; Defreyn, G.; Cazenave, J.; Gordon, J. ADP plays a key role in thrombogenesis in rats. *Thromb. Haemostasis* **1988**, *59*, 225–230.

(9) Jin, J.; Daniel, J. L.; Kunapuli, S. P. Molecular basis for ADP-induced platelet activation. II: the P2Y1 receptor mediates ADP-induced intracellular calcium mobilization and shape change in platelets. *J. Biol. Chem.* **1998**, *273*, 2030–2034.

(10) Hollopeter, G.; Jantzen, H.; Vincent, D.; Li, G.; England, L.; Ramakrishnan, V.; Yang, R.; Nurden, P.; Nurden, A.; Julius, D. Identification of the platelet ADP receptor targeted by antithrombotic drugs. *Nature* **2001**, *409*, 202–207.

(11) Jin, J.; Kunapuli, S. P. Coactivation of two different G protein-coupled receptors is essential for ADP-induced platelet aggregation. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8070–8074.

(12) Chou, K. C. Structural bioinformatics and its impact to biomedical science. *Curr. Med. Chem.* **2004**, *11*, 2105–2134.

(13) Costanzi, S.; Mamedova, L.; Gao, Z. G.; Jacobson, K. A. Architecture of P2Y nucleotide receptors: structural comparison based on sequence analysis, mutagenesis, and homology modeling. *J. Med. Chem.* **2004**, *47*, 5393–5404.

(14) Deflorian, F.; Jacobson, K. Comparison of three GPCR structural templates for modeling of the $P2Y_{12}$ nucleotide receptor. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 329–338.

(15) Jakubowski, J. A.; Winters, K. J.; Naganuma, H.; Wallentin, L. Prasugrel: a novel thienopyridine antiplatelet agent. A review of pre-clinical and clinical studies and the mechanistic basis for its distinct antiplatelet profile. *Cardiovasc. Drug Rev.* **2007**, *25*, 357–374.

(16) Clement, D. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet* **1996**, *348*, 1329–1338.

(17) Springthorpe, B.; Bailey, A.; Barton, P.; Birkinshaw, T.; Bonnert, R.; Brown, R.; Chapman, D.; Dixon, J.; Guile, S.; Humphries, R. From ATP to AZD6140: the discovery of an orally active reversible $P2Y_{12}$ receptor antagonist for the prevention of thrombosis. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6013–6018.

(18) Ferreiro, J. L.; Ueno, M.; Angiolillo, D. J. Cangrelor: a review on its mechanism of action and clinical development. *Expert Rev. Cardiovasc. Ther.* **2009**, *7*, 1195–1201.

(19) Vasiljev, K. S.; Uri, A.; Laitinen, J. T. 2-Alkylthio-substituted platelet $P2Y_{12}$ receptor antagonists reveal pharmacological identity between the rat brain Gi-linked ADP receptors and P2Y12. *Neuropharmacology* **2003**, *45*, 145–154.

(20) Crepaldi, P.; Cacciari, B.; Bonache, M.-C.; Spalluto, G.; Varani, K.; Borea, P. A.; Kügelgen, I. v.; Hoffmann, K.; Pugliano, M.; Razzari, C.; Cattaneo, M. 6-Amino-2-mercapto-3*H*-pyrimidin-4-one derivatives as new candidates for the antagonism at the $P2Y_{12}$ receptors. *Bioorg. Med. Chem.* **2009**, *17*, 4612–4621.

(21) Parlow, J. J.; Burney, M. W.; Case, B. L.; Girard, T. J.; Hall, K. A.; Harris, P. K.; Hiebsch, R. R.; Huff, R. M.; Lachance, R. M.; Mischke, D. A.; Rapp, S. R.; Woerndle, R. S.; Ennis, M. D. Piperazinyl glutamate pyridines as potent orally bioavailable $P2Y_{12}$ antagonists for inhibition of platelet aggregation. *J. Med. Chem.* **2010**, *53*, 2010–2037.

(22) Parlow, J. J.; Burney, M. W.; Case, B. L.; Girard, T. J.; Hall, K. A.; Hiebsch, R. R.; Huff, R. M.; Lachance, R. M.; Mischke, D. A.; Rapp, S. R.; Woerndle, R. S.; Ennis, M. D. Piperazinyl-glutamate-pyridines as potent orally bioavailable $P2Y_{12}$ antagonists for inhibition of platelet aggregation. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4657–4663.

(23) Parlow, J. J.; Burney, M. W.; Case, B. L.; Girard, T. J.; Hall, K. A.; Hiebsch, R. R.; Huff, R. M.; Lachance, R. M.; Mischke, D. A.; Rapp, S. R.; Woerndle, R. S.; Ennis, M. D. Piperazinyl-glutamate-pyrimidines as potent $P2Y_{12}$ antagonists for inhibition of platelet aggregation. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6148–6156.

(24) Parlow, J. J.; Burney, M. W.; Case, B. L.; Girard, T. J.; Hall, K. A.; Harris, P. K.; Hiebsch, R. R.; Huff, R. M.; Lachance, R. M.; Mischke, D. A.; Rapp, S. R.; Woerndle, R. S.; Ennis, M. D. Part II: Piperazinyl-glutamate-pyridines as potent orally bioavailable $P2Y_{12}$ antagonists for inhibition of platelet aggregation. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1388–1394.

(25) Bryant, J.; Post, J. M.; Alexander, S.; Wang, Y.-X.; Kent, L.; Schirm, S.; Tseng, J.-L.; Subramanyam, B.; Buckman, B.; Islam, I.; Yuan, S.; Sullivan, M. E.; Snider, M.; Morser, J. Novel $P2Y_{12}$ adenosine diphosphate receptor antagonists for inhibition of platelet aggregation (I): In vitro effects on platelets. *Thromb. Res.* **2008**, *122*, 523–532.

(26) Post, J. M.; Alexander, S.; Wang, Y.-X.; Vincelette, J.; Vergona, R.; Kent, L.; Bryant, J.; Sullivan, M. E.; Dole, W. P.; Morser, J.; Subramanyam, B. Novel $P2Y_{12}$ adenosine diphosphate receptor antagonists for inhibition of platelet aggregation (II): Pharmacodynamic and pharmacokinetic characterization. *Thromb. Res.* **2008**, *122*, 533–540.

(27) Wang, Y.; Vincelette, J.; Da Cunha, V.; Martin-McNulty, B.; Mallari, C.; Fitch, R.; Alexander, S.; Islam, I.; Buckman, B.; Yuan, S. A novel $P2Y_{12}$ adenosine diphosphate receptor antagonist that inhibits platelet aggregation and thrombus formation in rat and dog models. *Thromb. Haemostasis* **2007**, *97*, 847–855.

(28) Baqi, Y.; Atzler, K.; Köse, M.; Glänzel, M.; Müller, C. E. High-affinity, non-nucleotide-derived competitive antagonists of platelet $P2Y_{12}$ receptors. *J. Med. Chem.* **2009**, *52*, 3784–3793.

(29) Douglass, J. G.; deCamp, J. B.; Fulcher, E. H.; Jones, W.; Mahanty, S.; Morgan, A.; Smirnov, D.; Boyer, J. L.; Watson, P. S. Adenosine analogues as inhibitors of $P2Y_{12}$ receptor mediated platelet aggregation. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2167–2171.

(30) Bach, P.; Boström, J.; Brickmann, K.; van Giezen, J. J. J.; Hovland, R.; Petersson, A. U.; Ray, A.; Zetterberg, F. A novel series of piperazinyl-pyridine ureas as antagonists of the purinergic P2Y12 receptor. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 2877–2881.

(31) Wang, Y.; Li, Y.; Yang, S.; Yang, L. Classification of substrates and inhibitors of P-glycoprotein using unsupervised machine learning approach. *J. Chem. Inf. Model.* **2005**, *45*, 750–757.

(32) Wang, G.; Li, Y.; Liu, X.; Wang, Y. Understanding the aquatic toxicity of pesticide: structure-activity relationship and molecular descriptors to distinguish the ratings of toxicity. *QSAR Comb. Sci.* **2009**, *28*, 1418–1431.

(33) Xu, X.; Yang, W.; Wang, X.; Li, Y.; Wang, Y.; Ai, C. Dynamic communication between androgen and coactivator: Mutually induced conformational perturbations in androgen receptor ligand-binding domain. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1154–1171.

(34) Wang, Y.; Li, Y.; Ma, Z.; Yang, W.; Ai, C. Mechanism of MicroRNA-Target Interaction: Molecular Dynamics Simulations and Thermodynamics Analysis. *PLoS Comput. Biol.* **2010**, *6*, e1000866.

(35) Wang, X.; Yang, W.; Xu, X.; Zhang, H.; Li, Y.; Wang, Y. Studies of benzothiadiazine derivatives as hepatitis C virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics. *Curr. Med. Chem.* **2010**, *17*, 2788–2803.

(36) Wang, Y. H.; Li, Y.; Yang, S. L.; Yang, L. An *in silico* approach for screening flavonoids as P-glycoprotein inhibitors based on a Bayesian-regularized neural network. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 137–147.

(37) Sun, H. A universal molecular descriptor system for prediction of logP, logS, logBB, and absorption. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 748–757.

(38) Sun, H. Prediction of Chemical Carcinogenicity from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1506–1514.

(39) Sun, H. Pharmacophore-based virtual screening. *Curr. Med. Chem.* **2008**, *15*, 1018–1024.

(40) Nonaka, Y.; Hiramoto, T.; Fujita, N. Identification of endogenous surrogate ligands for human P2Y12 receptors by in silico and in vitro methods. *Biochem. Biophys. Res. Commun.* **2005**, *337*, 281–288.

(41) Hiramoto, T.; Nonaka, Y.; Inoue, K.; Yamamoto, T.; Omatsu-Kanbe, M.; Matsuura, H.; Gohda, K.; Fujita, N. Identification of endogenous surrogate ligands for human P2Y receptors through an in silico search. *J. Pharmacol. Sci.* **2004**, *95*, 81–93.

(42) Hiramoto, T.; Nemoto, W.; Kikuchi, T.; Fujita, N. Construction of hypothetical three-dimensional structure of $P2Y_1$ receptor based on fourier transform analysis. *J. Protein Chem.* **2002**, *21*, 537–545.

(43) Hao, M.; Li, Y.; Wang, Y.; Zhang, S. Prediction of $P2Y_{12}$ antagonists using a novel genetic algorithm-support vector machine coupled approach. *Anal. Chim. Acta* **2011**, *690*, 53–63.

(44) Hemmateenejad, B.; Elyasi, M. A segmented principal component analysis-regression approach to quantitative structure-activity relationship modeling. *Anal. Chim. Acta* **2009**, *646*, 30–38.

(45) Hemmateenejad, B.; Mehdipour, A. R.; Miri, R.; Shamsipur, M. Comparative QSAR studies on toxicity of phenol derivatives using quantum topological molecular similarity indices. *Chem. Biol. Drug Des.* **2010**, *75*, 521–531.

(46) Hemmateenejad, B.; Miri, R.; Jafarpour, M.; Tabarzad, M.; Shamsipur, M. Exploring QSAR for the inhibitory activity of a large set of aromatic/heterocyclic sulfonamides toward four different isoenzymes of carbonic anhydrase. *QSAR Comb. Sci.* **2007**, *26*, 1065–1075.

(47) Atabati, M.; Zarei, K.; Mohsennia, M. Prediction of λ_{max} of 1, 4-naphthoquinone derivatives using ant colony optimization. *Anal. Chim. Acta* **2010**, *663*, 7–10.

(48) Abreu, R.; Ferreira, I.; Queiroz, M. QSAR model for predicting radical scavenging activity of di(hetero)arylamines derivatives of benzo [*b*] thiophenes. *Eur. J. Med. Chem.* **2009**, *44*, 1952–1958.

(49) Richard, D.; David, E.; Jeffrey, D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

(50) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.

(51) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.

(52) Clark, M.; Cramer, R., III; Van Opdenbosch, N. Validation of the general purpose tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982–1012.

(53) AbdulHameed, M. D. M.; Hamza, A.; Liu, J. J.; Zhan, C. G. Combined 3D-QSAR modeling and molecular docking study on indolinone derivatives as inhibitors of 3-phosphoinositide-dependent protein kinase-1. *J. Chem. Inf. Model.* **2008**, *48*, 1760–1772.

(54) Murumkar, P. R.; Zambre, V. P.; Yadav, M. R. Development of predictive pharmacophore model for in silico screening, and 3D QSAR CoMFA and CoMSIA studies for lead optimization, for designing of potent tumor necrosis factor alpha converting enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 143–156.

(55) Srivastava, V.; Gupta, S. P.; Siddiqi, M. I.; Mishra, B. N. 3D-QSAR studies on quinazoline antifolate thymidylate synthase inhibitors by CoMFA and CoMSIA models. *Eur. J. Med. Chem.* **2010**, *45*, 1560–1571.

(56) Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499–511.

(57) Jain, A. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 281–306.

(58) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(59) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.

(60) Aalten, D. M. F.; Bywater, R.; Findlay, J. B. C.; Hendlich, M.; Hooft, R. W. W.; Vriend, G. PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 255–262.

(61) Schuttelkopf, A. W.; van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.* **2004**, *D60*, 1355–1363.

(62) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(63) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(64) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632–5633.

(65) Böhm, M.; Stürzebecher, J.; Klebe, G. Three-dimensional quantitative structure−activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **1999**, *42*, 458–477.

(66) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W., III The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743.

(67) Golbraikh, A.; Shen, M.; Xiao, Z. Y.; Xiao, Y. D.; Lee, K. H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241–253.

(68) Hemmateenejad, B.; Javadnia, K.; Elyasi, M. Quantitative structure-retention relationship for the Kovats retention indices of a large set of terpenes: A combined data splitting-feature selection strategy. *Anal. Chim. Acta* **2007**, *592*, 72–81.

(69) Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 357–369.

(70) Li, Y.; Wang, Y.; Ding, J.; Wang, Y.; Chang, Y.; Zhang, S. In silico prediction of androgenic and nonandrogenic compounds using random forest. *QSAR Comb. Sci.* **2009**, *28*, 396–405.

(71) Sun, X.; Li, Y.; Liu, X.; Ding, J.; Wang, Y.; Shen, H.; Chang, Y. Classification of bioaccumulative and non-bioaccumulative chemicals using statistical learning approaches. *Mol. Divers.* **2008**, *12*, 157–169.

(72) Zupan, J.; Novič, M.; Ruisánchez, I. Kohonen and counterpropagation artificial neural networks in analytical chemistry. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 1–23.

(73) Zhang, Z. Y.; An, L. Y.; Hu, W. X.; Xiang, Y. H. 3D-QSAR study of hallucinogenic phenylalkylamines by using CoMFA approach. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 145–153.

(74) Bringmann, G.; Rummey, C. 3D QSAR investigations on antimalarial naphthylisoquinoline alkaloids by comparative molecular similarity indices analysis (CoMSIA), based on different alignment approaches. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 304–316.

(75) Golbraikh, A.; Tropsha, A. Beware of q^2!. *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.

(76) Egan, W. J.; Morgan, S. L. Outlier detection in multivariate analytical chemical data. *Anal. Chem.* **1998**, *70*, 2372–2379.

(77) Hoffmann, K.; Sixel, U.; Di Pasquale, F.; von Kügelgen, I. Involvement of basic amino acid residues in transmembrane regions 6 and 7 in agonist and antagonist recognition of the human platelet $P2Y_{12}$-receptor. *Biochem. Pharmacol.* **2008**, *76*, 1201–1213.