

Rational Prediction of the Herbicidal Activities of Novel Protoporphyrinogen Oxidase Inhibitors by Quantitative Structure–Activity Relationship Model Based on Docking-Guided Active Conformation

BEILEI LEI,[†] JIAZHONG LI,[†] JING LU,[†] JUAN DU,[†] HUANXIANG LIU,[‡] AND
 XIAOJUN YAO^{*,†,§}

[†]State Key Laboratory of Applied Organic Chemistry and Department of Chemistry, Lanzhou University, Lanzhou 730000, China, [‡]School of Pharmacy, Lanzhou University, Lanzhou 730000, China, and [§]Key Lab of Preclinical Study for New Drugs of Gansu Province, Lanzhou University, Lanzhou 730000, China

Molecular docking-guided active conformation selection was used in a quantitative structure–activity relationship (QSAR) study of a series of novel protoporphyrinogen oxidase (PPO) inhibitors with herbicidal activities. The developed model can be used for the rational and accurate prediction of herbicidal activities of these inhibitors from their molecular structures. Molecular docking study was carried out to dock the inhibitors into the PPO active site and to obtain the rational active conformations. Based on the conformations generated from molecular docking, satisfactory predictive results were obtained by a genetic algorithm-multiple linear regression (GA-MLR) model according to the internal and external validations. The model gave a correlation coefficient R^2 of 0.972 and 0.953 and an absolute average relative deviation AARD of 2.24% and 2.75% for the training set and test set, respectively. The results from this work demonstrate that the molecular docking-guided active conformation selection strategy is rational and useful in the QSAR study of these PPO inhibitors and for the quantitative prediction of their herbicidal activities. The results obtained could be helpful to the design of new derivatives with potential herbicidal activities.

KEYWORDS: Protoporphyrinogen oxidase (PPO); molecular docking; Quantitative Structure–Activity Relationship (QSAR); Genetic Algorithm-Multiple Linear Regression (GA-MLR)

INTRODUCTION

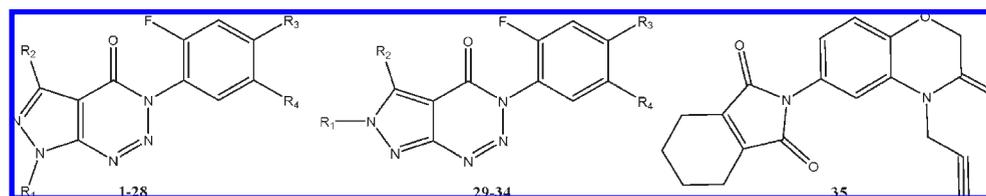
Inhibition of chlorophyll and heme biosynthesis has been an excellent approach in the development of herbicides. Protoporphyrinogen oxidase (PPO; EC 1.3.3.4) is an essential enzyme in chlorophyll biosynthesis. PPO, which oxidizes protoporphyrinogen IX (Protogen) to protoporphyrin IX (Proto) in the penultimate step of porphyrin biosynthesis, has been identified as the target of inhibitors with herbicidal activity. Inhibition of this highly regulated enzymatic conversion of Protogen to Proto leads to an unregulated extraplastidic accumulation of Proto (*1*). Accumulation of this photodynamic chlorophyll precursor is responsible for the light-dependent herbicidal action of PPO-inhibiting herbicides (*2*). All known PPO inhibitors apparently target at or near the catalytic site on the enzyme and compete with Protogen (*3–7*). On the basis of their chemical structures, these compounds can be classified into diphenyl ethers, benzoxazinones, phenyl imides, triazolinones, tetrazolinones, oxadiazolones, thiadiazolidines, isothiazolones, arylpyrroles, etc. (*8*).

Recently, Li et al. reported a series of novel pyrazolo-triazin-based PPO inhibitors (*9*). This class of inhibitors is different from all the ones mentioned above. These inhibitors exhibited good herbicidal activity. To further improve the herbicidal activity of these inhibitors, new compounds based on the experimentally active inhibitors should be designed. Then it will be quite necessary to analyze the quantitative structure–activity relationship (QSAR) of the known inhibitors to perform rational molecule design.

In recent years, several QSAR studies have been performed on different PPO inhibitors (*10–12*). All these works used the precise density functional theory (DFT)-based calculation to obtain the lowest-energy molecular conformation, and then built QSAR models with quantum chemical descriptors. The derived models have better performance than the semiempirical level based models and have potential predictive ability, because DFT is a more precise method, which can obtain the optimal lowest-energy conformation. However, the DFT calculation is time-consuming, and an approach with both good model performance and time-saving calculation will be of more practical use. Another important issue that needs to be paid attention to in these QSAR studies is that the conformation of the inhibitor obtained from DFT calculation may not be the active conformation.

*Corresponding author. Tel: +86-931-891-5756 Fax: +86-931-891-2582. E-mail: xjyao@lzu.edu.cn.

Table 1. Chemical Structures of All the PPO Inhibitors



no.	R ₁	R ₂	R ₃	R ₄	no.	R ₁	R ₂	R ₃	R ₄
1	CH ₃	H	Cl	OCH ₂ C≡CH	18 ^a	<i>n</i> -C ₃ H ₇	H		R ₃ R ₄ =OCH ₂ CONCH ₂ C≡CH
2 ^a	CH ₃	H	Cl	OCH ₂ CH=CH ₂	19	<i>n</i> -C ₃ H ₇	H		R ₃ R ₄ =OCH ₂ CONCH ₂ CH=CH ₂
3	CH ₃	H	Cl	OCO ₂ CH ₂ CH ₃	20	CH ₃	H	Cl	OH
4	CH ₃	CF ₃	Cl	OCH ₂ C≡CH	21 ^a	CH ₃	H	Cl	OCH ₃
5	CH ₃	CH ₃	Cl	OCH ₂ C≡CH	22	CH ₃	H	Cl	OC ₂ H ₅
6 ^a	CH ₃	CH ₃	Cl	OCH ₂ CH=CH ₂	23	CH ₃	H	Cl	O(<i>n</i> -C ₄ H ₉)
7	CH ₃	H		R ₃ R ₄ =OCH ₂ CONCH ₂ C≡CH	24	CH ₃	H	Cl	O(<i>n</i> -C ₅ H ₁₁)
8 ^a	CH ₃	H		R ₃ R ₄ =OCH ₂ CONCH ₂ CH=CH ₂	25	CH ₃	H	Cl	OCH ₂ CH=CHCl(E)
9	CH ₃	H		R ₃ R ₄ =OCH ₂ CON(<i>n</i> -C ₃ H ₇)	26 ^a	CH ₃	H	Cl	OCH ₂ C(Cl)=CH ₂
10	CH ₃	CF ₃		R ₃ R ₄ =OCH ₂ CONCH ₂ C≡CH	27	CH ₃	H	Cl	OCH ₂ CN
11 ^a	CH ₃	CF ₃		R ₃ R ₄ =OCH ₂ CONCH ₂ CH=CH ₂	28	CH ₃	H	Cl	OCH ₂ OC ₂ H ₄ OCH ₃
12	CH ₃	CH ₃		R ₃ R ₄ =OCH ₂ CONCH ₂ C≡CH	29	allyl	H	Cl	OCH ₂ C≡CH
13	CH ₃	CH ₃		R ₃ R ₄ =OCH ₂ CONCH ₂ CH=CH ₂	30 ^a	allyl	H	Cl	OCH ₂ CH=CH ₂
14	allyl	H	Cl	OCH ₂ C≡CH	31	allyl	H		R ₃ R ₄ =OCH ₂ CONCH ₂ C≡CH
15 ^a	allyl	H	Cl	OCH ₂ CH=CH ₂	32 ^a	allyl	H		R ₃ R ₄ =OCH ₂ CONCH ₂ CH=CH ₂
16	allyl	H		R ₃ R ₄ =OCH ₂ CONCH ₂ C≡CH	33	<i>n</i> -C ₃ H ₇	H		R ₃ R ₄ =OCH ₂ CONCH ₂ C≡CH
17 ^a	allyl	H		R ₃ R ₄ =OCH ₂ CONCH ₂ CH=CH ₂	34	<i>n</i> -C ₃ H ₇	H		R ₃ R ₄ =OCH ₂ CONCH ₂ CH=CH ₂

^a Test samples.

In QSAR studies, the molecular descriptors are important intermediates correlating the molecular structures with their target activities. The molecular descriptors are often calculated from three-dimensional molecular structures. Any in-depth representation of a molecule structure should take into account its 3D structure (13). So how to obtain the appropriate 3D molecular conformation is of great importance. Generally, the lowest-energy molecular conformations are used to calculate descriptors in QSAR study, but in most cases, the active conformations of inhibitors are different from the lowest-energy molecular conformations. Becker et al. have attempted to build a conformation space and created a QSAR type descriptor to quantify the effect of conformation constraints on bioactivity, which was shown to be in excellent correlation with the observed activity of the molecules (14), but the method is too complex to be used in QSAR study.

In this work, the docking-guided molecular conformation selection strategy docking method was introduced into a QSAR study of PPO inhibitors to obtain rational active conformations. The docking program AutoDock4.0 (15) was used to dock all inhibitors into the active site of PPO. Based on the generated active conformations, a variety of molecular descriptors were calculated and genetic algorithm (GA) was used to select the most relevant descriptors to build QSAR model. By explaining the physical-chemical meaning of the selected descriptors, we can investigate the most important structural factors influencing the activity.

MATERIALS AND METHODS

Data Set. The 35 3*H*-pyrazolo[3,4-*d*][1,2,3]triazin-4-one derivatives used in this study are novel protoporphyrinogen oxidase (PPO) inhibitors available in the literature (9). The herbicidal activities expressed as *pI*₅₀ values were used as dependent variable in the following analyses. The structures of all compounds are shown in Table 1, and the biological data are listed in Table 2.

Molecular Docking and Molecular Descriptor Calculation. All molecular structures were drawn with the molecular sketch program in the

SYBYL 6.9 molecular modeling package (16), and energy minimizations were performed using the MMFF94 force field (17) with a distance-dependent dielectric function and Powell method with a convergence criterion of 0.01 kcal/mol. Then to locate the appropriate binding orientations and conformations of these derivatives interacting with PPO, molecular docking program Autodock version 4.0 (15) was used for the automated molecular docking studies. The PPO crystal structure was obtained from the RCSB PDB database (18) (PDB: 1SEZ) (19), and the contained waters were deleted.

All inhibitors were docked into the PPO crystal structure to obtain the active conformations. AutoDockTools 1.4.5 (ADT) (20) was used to add polar hydrogens and to assign Kollman-all-atom charges for PPO and Gasteiger–Marsili charges for the inhibitors. AutoGrid 4.0 (15) was used to create affinity grids centered on the active site based on the location of the cocrystallized ligand. Each grid enclosed an area of 60 Å × 60 Å × 60 Å with 0.375 Å spacing. All bond rotations and torsions for the ligand were automatically set in the ADT. The number of individuals in each population was set at 50. The rest of the parameters were taken as default. Finally, the docking was performed by AutoDock 4.0 (15) employing the Lamarckian genetic algorithm (LGA) (15) method for conformational search and docking. The resulted docking conformations were transferred into DRAGON 5.4 (21) to calculate the molecular descriptors.

Data Splitting Based on Principal Component Analysis (PCA).

To build and validate a QSAR model, a representative training set is needed to develop the model, and a test set is needed to validate the external predictive ability of the derived model. In this study, the principal component analysis (PCA) method was used to obtain such a training set and test set (22). PCA is a computational tool that reduces the dimensionality of molecular descriptor space while retaining an accurate representation of the intermolecular distances. PCA based on the generated descriptors of the whole data set was performed and the descriptor space was explored using the obtained principal components. In this space the representative subsets were selected.

Descriptor Selection and Model Construction. The next step was to search the feature space and select pertinent descriptors correlated with the herbicidal activity. Here, genetic algorithm (GA) (23) was used due to its good performance in feature selection (24–26). In general, the process of GA proceeds as follows. First of all, GA generates a set of solutions randomly which is called an initial population. Each solution is called

Table 2. The Experimental and Predicted Activities by the GA-MLR Model and the Relevant Descriptors in the QSAR Model

no.	exp p_{50}	pred p_{50} MLR	absolute error	SCBO ^a	GGI9 ^a	R2u+ ^a	C- 033 ^a
1	7.12	6.96	-0.16	34.5	0.131	0.084	0
2 ^b	6.33	6.65	0.32	33.5	0.131	0.087	0
3	6.31	6.32	0.01	35.5	0.182	0.090	0
4	5.71	5.71	0.00	38.5	0.298	0.080	0
5	6.49	6.77	0.28	35.5	0.151	0.089	0
6 ^b	6.10	6.49	0.39	34.5	0.151	0.091	0
7	7.85	7.59	-0.26	39.5	0.182	0.077	0
8 ^b	7.34	7.22	-0.12	38.5	0.182	0.082	0
9	7.37	7.35	-0.02	37.5	0.182	0.070	0
10	5.74	5.54	-0.20	43.5	0.388	0.082	0
11 ^b	5.41	5.14	-0.27	42.5	0.388	0.088	0
12	6.63	6.69	0.06	40.5	0.241	0.088	0
13	6.28	6.41	0.13	39.5	0.241	0.090	0
14	7.94	7.81	-0.13	37.5	0.161	0.064	0
15 ^b	7.19	6.94	-0.25	36.5	0.161	0.086	0
16	8.02	7.94	-0.08	42.5	0.212	0.074	0
17 ^b	7.78	7.75	-0.03	41.5	0.212	0.073	0
18 ^b	7.87	7.84	-0.03	41.5	0.212	0.070	0
19	7.75	7.79	0.04	40.5	0.212	0.064	0
20	6.23	6.11	-0.12	29.5	0.060	0.108	0
21 ^b	6.12	5.98	-0.14	30.5	0.105	0.099	0
22	6.53	6.22	-0.31	31.5	0.121	0.091	0
23	6.61	6.56	-0.05	33.5	0.151	0.081	0
24	6.32	6.54	0.22	34.5	0.186	0.073	0
25	6.55	6.72	0.17	34.5	0.151	0.083	0
26 ^b	6.58	6.67	0.09	34.5	0.161	0.080	0
27	6.64	6.85	0.21	34.5	0.131	0.088	0
28	6.09	6.32	0.23	36.5	0.196	0.091	0
29	4.51	4.46	-0.05	37.5	0.161	0.092	1
30 ^b	<4.50	4.12	-	36.5	0.161	0.096	1
31	5.27	5.38	0.11	42.5	0.212	0.075	1
32 ^b	4.83	4.75	-0.08	41.5	0.212	0.089	1
33	5.00	5.22	0.22	41.5	0.212	0.073	1
34	5.04	4.76	-0.28	40.5	0.212	0.081	1
35	8.49	8.46	-0.03	37.0	0.101	0.066	0

^aThe detailed definitions can be found in the section Molecular Descriptor Explanation. ^bTest set compounds.

a chromosome, and it is usually represented in the form of a binary string. After the generation of the initial population, a new population is formed to consist of the fittest chromosomes as well as offspring of these chromosomes based on the notion of survival of the fittest. The value of the fitness for each chromosome is calculated based on leave-one-out (LOO) cross validation (Q_{loo}^2). In general, offspring are generated by applying genetic operators. Among various genetic operators, selection, crossover and mutation are the most fundamental and popular operators. These steps of evolution continue until the stopping conditions are satisfied (27, 28).

In the present work, genetic algorithm and multiple linear regression (GA-MLR) were performed in MobyDigs (29) using GA-VSS (genetic algorithm-variable subset selection) and the ordinary least-squares regression (LS) method. The corresponding parameters used were as follows: population size, 100; maximum allowed variables in a model, 3; and reproduction/mutation trade-off (T), 0.5. The crossover and mutation values are all based on T and calculated automatically in the software.

Furthermore, particular attention was paid on the collinearity of the selected molecular descriptors by applying the QUIK rule (Q under influence of K) (30). Only models with a global correlation of the $[X + Y]$ block (K_{XY}) greater than the global correlation of the X block (K_X) variable can be accepted, where X is the descriptor matrix and Y is the dependent variable. The detailed theory of the QUIK rule can be found in ref 30.

QSAR Model Validation. Before a QSAR model is used to predict the activities for new compounds, it should be validated both internally and externally to ensure that the built model is robust, reliable, stable and

predictive. In the current work, several statistic terms such as correlation coefficient (R^2), leave-one-out (LOO) cross-validated Q_{loo}^2 , root-mean-square error (RMSE), and the absolute average relative deviation (AARD) were used to assess the internal predictive ability of the proposed model. Besides, the Y scrambling technique was also employed to exclude the possibility of chance correlation and to check for reliability and robustness by permutation testing. In Y scrambling, new models were recalculated for randomly reordered responses (Y scrambling). The resulting models obtained with randomized response should have significantly lower Q_{loo}^2 values than the proposed one because the relationship between the structures and activities is broken (31). Y scrambling was performed by response scrambling with maximum iterations of 500, and then the mean Q_{loo}^2 of the Y scrambling was reported. Furthermore, the built model was also validated externally using the test set compounds due to the fact that the best way to evaluate the predictive ability of a QSAR model is its validation using compounds not included in the training set with known activities (31, 32). The corresponding statistical parameters were defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_{ie} - y_{ip})^2}{n}}$$

$$\text{AARD} = \frac{100}{n} \sum_{i=1}^n \frac{|y_{ie} - y_{ip}|}{y_{ie}}$$

where i represents the i th molecule, y_{ie} is the desired output (experimental property), y_{ip} is the actual output, and n is the number of compounds in the training or test set.

Applicability Domain of the QSAR Model. The applicability domain (AD) of the model is evaluated using the hat value, which is defined as follows:

$$h_i = x_i(X^T X)^{-1} x_i^T \quad (i = 1, \dots, m)$$

where x_i is the descriptor row-vector of the query compound i , m is the number of query compounds, and X is the $n \times k$ matrix of the training set (k is the number of model descriptors and n is the number of training set samples). To visualize the AD of a QSAR model, the plot of cross-validated standardized errors versus hat values (the Williams plot) achieves an immediate, and simple, graphical detection of both Y outliers and X outliers in a model. In the plot the horizontal and vertical dashed lines indicate the limits of normal values: the first for the Y outliers (i.e., compounds with cross-validated standardized errors greater than 3.0 standard deviation units, $\pm 3.0\sigma$) and the second for X outliers (i.e., compounds with hat values greater than h^*). The warning hat h^* is fixed at $3k'/n$ generally, where n is the number of training compounds and k' is the number of model parameters plus 1.

RESULTS AND DISCUSSION

Docking Analysis. All inhibitors were docked into the active site of PPO to obtain the active conformations for the following QSAR analysis. The automated molecular docking produced 50 binding conformations. The obtained 50 conformations of each inhibitor were clustered according to the rmsd (root mean squared deviation) values relative to the initial conformation, and the most possible binding conformations were selected according to both the interacting energy and the number of conformations in a cluster. In this way, the most likely binding conformations for all compounds were obtained.

The binding mode of the most active compound **16** in the PPO active site (the main residues in interactions are displayed) is displayed in **Figure 1(a)**, and corresponding 2D interaction mode is shown in **Figure 1(b)**. These two figures illustrate how the carbonyl oxygen atom of the triazine ring of **16** forms two H-bond interactions (a strong one with bond length 2.59 Å and a weak one of 3.19 Å) with the residue Arg98. In addition, this inhibitor is located in a hydrophobic pocket composed by residues Leu334,

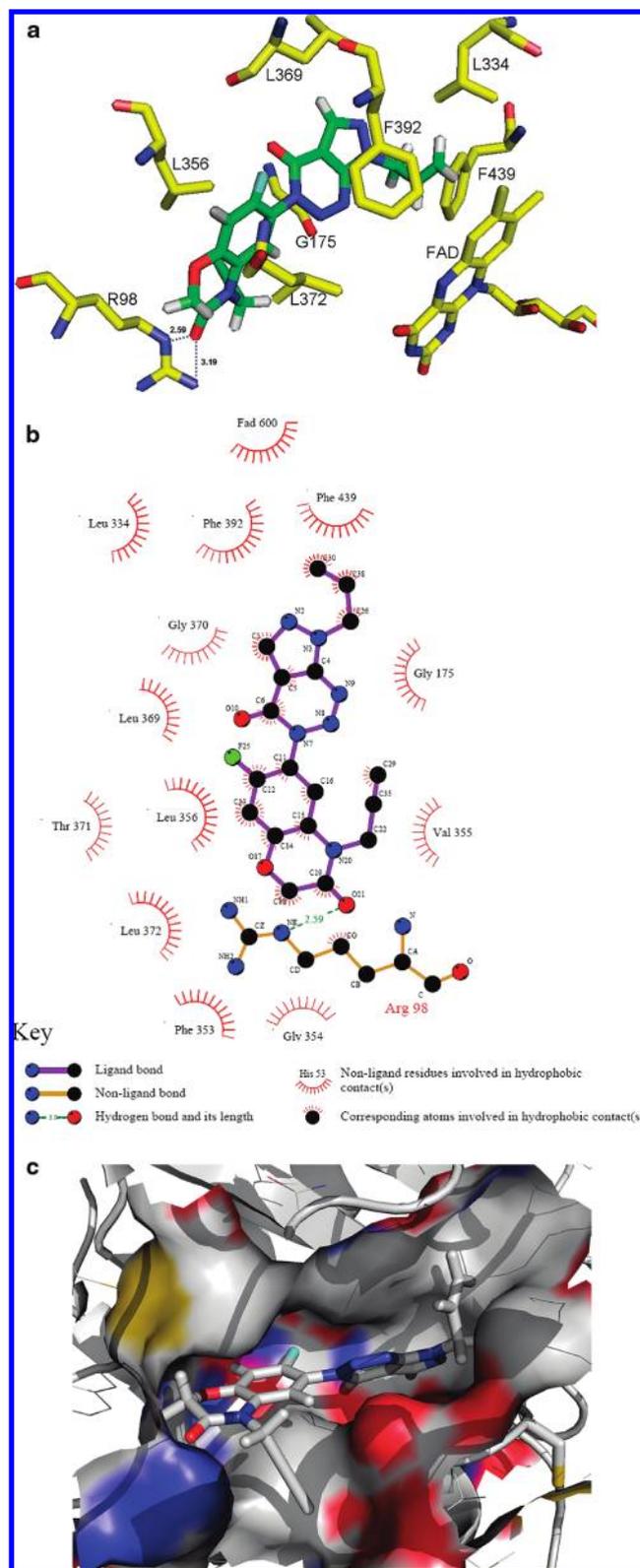


Figure 1. The binding mode of the most active compound **16**. (a) The binding interactions of **16** in the active site of PPO by using the PyMOL program (40). The yellow sticks represent the amino acid residuals in the active site, and the green sticks describe **16**; the blue dashed lines and the numbers represent H-bond interactions and the corresponding bond length between the heavy atoms. (b) Schematic representation of interactions between **16** and PPO produced using the Ligplot program developed by Wallace et al. (41). (c) **16** in the binding pocket of PPO by the PyMOL program (40).

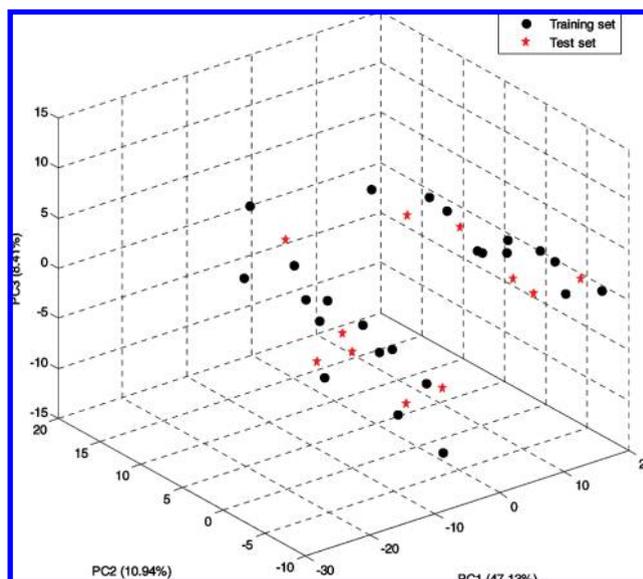


Figure 2. PCA plot of the compounds in the training and test sets.

Phe392, Phe439, Gly370, Gly175, Leu369, Thr371, Leu356, Val355, Leu372, Phe353 and Gly354. **Figure 1(c)** indicates that PPO interacts with the inhibitor mainly through hydrophobic interactions. **16** is located in the hydrophobic pocket like a “key” in the “lock”. All other inhibitors bound to the active site of PPO in a similar way with **16**. Based on the reasonable binding conformations, the following analyses were performed.

PCA Analysis of the Training and Test Sets. The PCA method was used to visualize the descriptor space of the data set and show the distance between each pair of compounds. The PCA plot is shown in **Figure 2**. The PC1, PC2, and PC3 made 47.13%, 10.94%, and 8.41% contributions to the total PCs, respectively. In all, these three PCs account for 66.48% of the descriptors in the data set. **Figure 2** illustrates the data set clustered into two classes, which agrees with the two framework structures of the data set. That is to say, the PCA plot represents the molecular diversity of these compounds appropriately. Therefore, compounds in the training and test sets were selected based on the 3D plot considering that the training set was representative of the whole data set. Eventually, 24 compounds were selected to the training set and 11 compounds were in the test set as listed in **Table 1**.

Results of the Final QSAR Model. To select most relevant descriptors to the pI_{50} of the compounds, GA was performed to do the feature selection based on the training samples only. The optimum number of descriptors (D_n) was determined when adding new descriptors did not improve the performance of the model significantly. In this work, the optimum D_n was four. The best 4-parameter model and corresponding validating parameters were as described below:

$$pI_{50} = 0.2234 \text{ SCBO} - 13.58 \text{ GGI9} \\ - 29.32 \text{ R2u} + - 2.532(\text{C-033}) + 3.498$$

$$N_{\text{tr}} = 24, Q^2_{\text{loo}} = 0.954, R^2_{\text{tr}} = 0.972, K_X = 37.34, K_{XY} \\ = 48.80, Q^2_{\text{Y-scrambling}} = 0.087, \text{RMSE}_{\text{tr}} \\ = 0.168, \text{AARD}_{\text{tr}} = 2.24\%$$

RMSE represents the root-mean-square error, and AARD means the absolute average relative deviation. From the statistical parameters, it can be seen that the MLR model is stable and

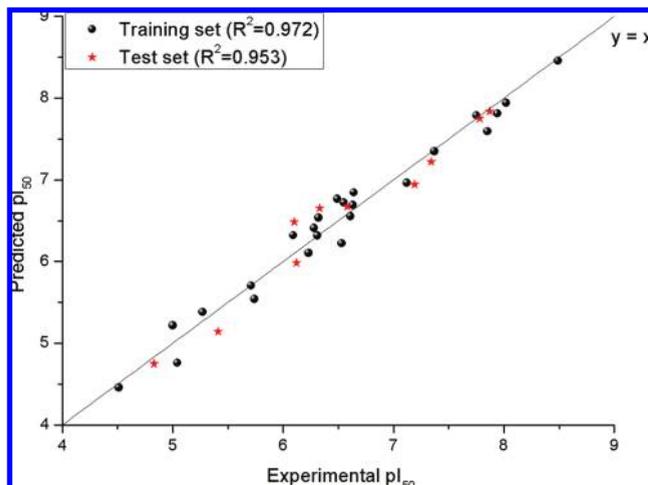


Figure 3. The experimental pI_{50} versus the predicted pI_{50} for the compounds in the training and test sets.

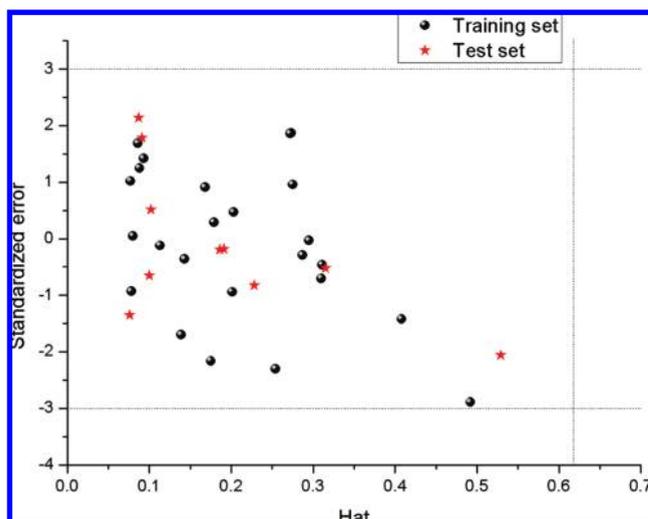


Figure 4. Williams plot for the original data set and external test set. The dashed lines are, respectively, the 3σ limit and the warning value of hat ($h^* = 0.62$).

has good internal predictive ability. The model can predict the 10 external test set compounds very well with R^2_{test} value of 0.953, AARD_{ts} 2.75% and RMSE_{ts} 0.208. The prediction for **30** in the test set is 4.12, which is also consistent with its experimental activity less than 4.50. The predicted pI_{50} values are listed in **Table 2**. **Figure 3** shows the regression plot of predicted pI_{50} vs experimental values. In addition, from the Williams plot (**Figure 4**) it can be seen that there is no Y outlier and there is no X outlier. All these results indicate that the built model is robust, reliable, stable and predictive.

Molecular Descriptor Explanation. By interpreting the descriptors in the regression model, it is possible to gain some insight into factors that are likely to govern the activities of these compounds. The values of descriptors for each compound are listed in **Table 2**. The relative importance of the descriptors was determined by their standardized regression coefficients (Std Coeff). The Std Coeff values of the four descriptors are SCBO (0.783), GGI9 (−0.876), R2u+ (−0.299), and C-033 (−0.937), respectively. It can be seen that the most important descriptor is the C-033, which is an atom centered fragment descriptor defined for each ring atom with three neighbors. It represents the number of the

R-CH \cdots X fragment in a molecule with the meaning that a central carbon atom (C) on an aromatic ring has a carbon neighbor (R), a heteroatom neighbor (X) and the third hydrogen (H) neighbor outside the ring. “--” and “ \cdots ” stand for aromatic and aromatic single bonds, respectively (33). For these PPO inhibitors, the C-033 fragment indeed plays an important role in the binding process and can influence the herbicidal activity tremendously. There are two different frameworks of the inhibitors except **35**, and the difference lies in the C-033 fragment, i.e. the position of the R_1 substituent group as shown in **Table 1**. It can be seen from **Figure 1(c)** that the active pocket has not enough space to accommodate the R_1 group of **29–34** for steric block, which can explain their lower activities. The second descriptor is GGI9. GGI9 belongs to the Gálvez topological charge indices, which were proposed to evaluate the charge transfer between pairs of atoms, and therefore the global charge transfer in the molecule based on the corrected adjacency matrix. It is worthwhile to note that the molecular conformation can affect the distribution of charges, consequently change the GGI9 value. Therefore, reasonable conformation is very important to get correct descriptor values. GGI9 represents the ninth eigenvalue of the matrix of a molecule, and the lower values of these descriptors are required to improve the activity (34, 35). The third descriptor, SCBO, is a constitutional descriptor, which means the sum of conventional bond orders (H-depleted) (36). R2u+ is an R-GETAWAY (GEometry, Topology, and Atom-Weights Assembly) molecular descriptor, which is derived by employing the molecular influence matrix and geometric interatomic distances in the molecule (37, 38). It has been successfully used to predict activities of A_1 adenosine receptor agonists (39). The matrix is calculated from the spatial coordinates of the molecule atoms in a given conformation, which also emphasizes the importance of the reasonable conformation.

In conclusion, a quantitative structure–activity relationship study was performed on a series of PPO inhibitors based on docking-guided molecular conformation selection strategy. All inhibitors were docked into PPO active site to obtain the active conformations by using the software AutoDock4.0. PCA was used to split the data set into a training set and a test set. GA was used to select the most important descriptors based on the training set only. The obtained model is stable and has good predictive ability according to the internal and external validations. The four molecular descriptors contained in the model have specific physical-chemical meaning. Among them, the fragment C-033 plays the essential role in the correlation between inhibitors and PPO, which can also explain the lower activities of **29–34**. Moreover, a topological charge index descriptor and an R-GETAWAY descriptor indicate that the molecular conformation is greatly important in building a quantitative relationship between structure and activity. The present work demonstrates that the docking-guided molecular conformation selection strategy is very useful in QSAR studies for quantitative prediction of biological activity, and, therefore, could be expected to help facilitate the design of new derivatives with potential activity.

ACKNOWLEDGMENT

The authors wish to express their gratitude to the anonymous reviewers for their constructive suggestion in improving the paper.

LITERATURE CITED

- (1) Matringe, M.; Scalla, R. Effects of acifluorfen-methyl on cucumber cotyledons: Porphyrin accumulation. *Pestic. Biochem. Physiol.* **1988**, *32*, 164–172.

- (2) Becerril, J. M.; Duke, S. O. Protoporphyrin IX content correlates with activity of photobleaching herbicides. *Plant Physiol.* **1989**, *90*, 1175–1181.
- (3) Dayan, F. E.; Meazza, G.; Bettarini, F.; Signorini, E.; Piccardi, P.; Romagnì, J. G.; Duke, S. O. Synthesis, herbicidal activity, and mode of action of IR 5790. *J. Agric. Food Chem.* **2001**, *49*, 2302–2307.
- (4) Dayan, F. E.; Duke, S. O.; Weete, J. D.; Hancock, H. G. Selectivity and mode of action of carfentrazone-ethyl, a novel phenyl triazolone herbicide. *Pestic. Sci.* **1997**, *51*, 65–73.
- (5) Dayan, F. E.; Armstrong, B. M.; Weete, J. D. Inhibitory activity of sulfentrazone and its metabolic derivatives on soybean (*Glycine max*) protoporphyrinogen oxidase. *J. Agric. Food Chem.* **1998**, *46*, 2024–2029.
- (6) Dayan, F. E.; Duke, S. O.; Reddy, K. N.; Hamper, B. C.; Leschinsky, K. L. Effects of isoxazole herbicides on protoporphyrinogen oxidase and porphyrin physiology. *J. Agric. Food Chem.* **1997**, *45*, 967–975.
- (7) Matringe, M.; Scalla, R. Characterization of [3H]acifluorfen binding to purified pea etioplasts, and evidence that protoporphyrinogen oxidase specifically binds acifluorfen. *Eur. J. Biochem.* **1992**, *209*, 861–868.
- (8) Huang, M. Z.; Huang, K. L.; Ren, Y. G.; Lei, M. X.; Huang, L.; Hou, Z. K.; Liu, A. P.; Ou, X. M. Synthesis and herbicidal activity of 2-(7-fluoro-3-oxo-3,4-dihydro-2H-benzo[b][1,4]oxazin-6-yl)isindoline-1,3-diones. *J. Agric. Food Chem.* **2005**, *53*, 7908–7914.
- (9) Li, H. B.; Zhu, Y. Q.; Song, X. W.; Hu, F. Z.; Liu, B.; Li, Y. H.; Niu, Z. X.; Liu, P.; Wang, Z. H.; Song, H. B.; Zou, X. M.; Yang, H. Z. Novel protoporphyrinogen oxidase inhibitors: 3H-pyrazolo[3,4-d]-[1,2,3]triazin-4-one derivatives. *J. Agric. Food Chem.* **2008**, *56*, 9535–9542.
- (10) Wan, J.; Zhang, L.; Yang, G. F.; Zhan, C. G. Quantitative structure-activity relationship for cyclic imide derivatives of protoporphyrinogen oxidase inhibitors: a Study of quantum chemical descriptors from density functional theory. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2099–2105.
- (11) Zhang, L.; Wan, J.; Yang, G. F. A DFT-based QSARs study of protoporphyrinogen oxidase inhibitors: phenyl triazolones. *Bioorg. Med. Chem. Lett.* **2004**, *12*, 6183–6191.
- (12) Wan, J.; Zhang, L.; Yang, G. F. Quantitative structure-activity relationships for phenyl triazolones of protoporphyrinogen oxidase inhibitors: a density functional theory study. *J. Comput. Chem.* **2004**, *25*, 1827–1832.
- (13) Gasteiger, J. Of molecules and humans. *J. Med. Chem.* **2006**, *49*, 6429–6434.
- (14) Becker, O. M.; Levy, Y.; Ravitz, O. Flexibility, conformation spaces, and bioactivity. *J. Phys. Chem. B* **2000**, *104*, 2123–2135.
- (15) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (16) *Sybyl version 6.9*; Tripos Associates: St. Louis, MO, 1999.
- (17) Halgren, T. A. Maximally diagonal force constants in dependent angle-bending coordinates. II. Implications for the design of empirical force fields. *J. Am. Chem. Soc.* **1990**, *112*, 4710–4723.
- (18) <http://www.rcsb.org/pdb/>.
- (19) Koch, M.; Breithaupt, C.; Kiefersauer, R.; Freigang, J.; Huber, R.; Messerschmidt, A. Crystal structure of protoporphyrinogen IX oxidase: a key enzyme in haem and chlorophyll biosynthesis. *EMBO J.* **2004**, *23*, 1720–1728.
- (20) Sanner, M. F. Python: A programming language for software integration and development. *J. Mol. Graphics Modell.* **1999**, *17*, 57–61.
- (21) Talete srl, DRAGON for windows (software for molecular descriptor calculations). Version 5.4, 2006. <http://www.talete.mi.it/>.
- (22) Ren, Y. Y.; Liu, H. X.; Yao, X. J.; Liu, M. C. Three-dimensional topographic index applied to the prediction of acyclic C5-C8 alkenes Kovás retention indices on polydimethylsiloxane and squalane columns. *J. Chromatogr. A* **2007**, *1155*, 105–111.
- (23) Goldberg, D. E. *Genetic algorithms in search, optimization and machine learning*; Addison-Wesley: Reading, MA, 1989.
- (24) Li, J. Z.; Lei, B. L.; Liu, H. X.; Li, S. Y.; Yao, X. J.; Liu, M. C.; Gramatica, P. QSAR study of malonyl-CoA decarboxylase inhibitors using GA-MLR and a new strategy of consensus modeling. *J. Comput. Chem.* **2008**, *29*, 2636–2647.
- (25) Liu, H. X.; Gramatica, P. QSAR study of selective ligands for the thyroid hormone receptor β . *Bioorg. Med. Chem.* **2007**, *15*, 5251–5261.
- (26) Lei, B. L.; Li, S. Y.; Xi, L. L.; Li, J. Z.; Liu, H. X.; Yao, X. J. Novel approaches for retention time prediction of oligonucleotides in ion-pair reversed-phase high-performance liquid chromatography. *J. Chromatogr. A* **2009**, *1216*, 4434–4439.
- (27) Han, J.; Kamber, M., *Datamining: concepts and techniques*; Morgan Kaufmann Publishers: San Francisco, CA, 2001.
- (28) Choi, J. H.; Ahn, H.; Han, I. Utility-based double auction mechanism using genetic algorithms. *Expert Syst. Appl.* **2008**, *34*, 150–158.
- (29) Todeschini, R.; Consonni, V.; Pavan, M., Software for multilinear regression analysis and variable subset selection by genetic algorithm. MOBY DIGS, Version 1.2 for Windows, Talete Srl, Milan, Italy, 2002.
- (30) Todeschini, R.; Consonni, V.; Maiocchi, A. The K correlation index: theory development and its applications in chemometrics. *Chemom. Intell. Lab. Syst.* **1999**, *46*, 13–29.
- (31) Gramatica, P.; Giani, E.; Papa, E. Statistical external validation and consensus modeling: A QSPR case study for Koc prediction. *J. Mol. Graphics Modell.* **2007**, *25*, 755–766.
- (32) Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694–701.
- (33) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
- (34) Galvez, J.; Garcia, R.; Salabert, M. T.; Soler, R. Charge indexes. New topological descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525.
- (35) Galvez, J.; Garcia-Domenech, R.; Julian-Ortiz, J. V. d.; Soler, R. Topological approach to drug design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272–284.
- (36) Todeschini, R.; Consonni, V. *Handbook of molecular descriptors: Methods and principles in medicinal chemistry*; WILEY-VCH Verlag GmbH: D-69469 Weinheim, 2000.
- (37) Consonni, V.; Todeschini, R.; Pavan, M. Structure/Response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682–692.
- (38) Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/Response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693–705.
- (39) González, M. P.; Terán, C.; Teijeira, M.; Besadac, P. Geometry, topology, and atom-weights assembly descriptors to predicting A1 adenosine receptors agonists. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2641–2645.
- (40) DeLano, W. L. *The PyMOL molecular graphics system*; DeLano Scientific: Palo Alto, CA, 2002.
- (41) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127–134.

Received June 12, 2009. Revised manuscript received September 4, 2009. Accepted September 8, 2009. This work was supported by the Program for New Century Excellent Talents in University (Grant No. NCET-07-0399) and the Project Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.