Quantitative Structure−Activity Relationship for Cyclic Imide Derivatives of Protoporphyrinogen Oxidase Inhibitors: A Study of Quantum Chemical Descriptors from Density Functional Theory

Jian Wan,[†] Li Zhang,[†] Guangfu Yang,*[,†] and Chang-Guo Zhan[‡]

Key Laboratory of Pesticide & Chemical Biology (CCNU), Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, P. R. China, and Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, Kentucky 40536

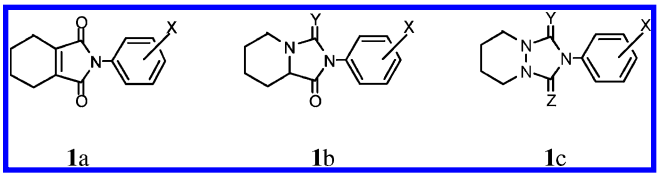
Received June 25, 2004

This study examined the applicability of various density functional theory (DFT)-based descriptors, such as energy gap (ΔE) between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), weighted nucleophilic atomic frontier electron density (WNAFED, F_i^N), mean molecular polarizability (α), and net atomic charge (Q_i), in quantitative structure−activity relationship (QSAR) studies on a class of important protoporphyrinogen oxidase (Protox) inhibitors including a series of cyclic imide derivatives with various heterocyclic rings and substituents. Our QSAR analysis using the quantum chemical descriptors calculated at the B3LYP/6-31G(d,p) level led to a useful explicit correlation relationship, i.e. $p\mathbf{I}_{50} = -5.7414 + 0.1424\alpha - 0.0003\alpha^2 - 0.4546F_{C\bullet}^N + 0.2974Q_{N\bullet\bullet}$ (n=26, R^2=0.87), showing that descriptors mean molecular polarizability, α, and WNAFED $F_{C\bullet}^N$ of a critical carbon atom and net atomic charge (Q_i) in the molecules are most likely responsible for the in vitro biological activity of cyclic imides. It has been shown that the use of the DFT-based quantum chemical descriptors indeed led to a better QSAR equation than that obtained from the use of the corresponding descriptors calculated at a semiempirical PM3 level. The present work demonstrates that the DFT-based quantum chemical descriptors are potentially useful in the future QSAR studies for quantitatively predicting biological activity, and, therefore, the DFT-based QSAR approach could be expected to help facilitate the design of additional substituted cyclic imide derivatives of Protox inhibitors with the potentially higher biological activity.

INTRODUCTION

Protoporphyrinogen oxidase (Protox, EC 1.3.3.4), the last common enzyme in chlorophylls biosynthesis,[1] has been identified as the target site of action for several photodynamically active porphyric herbicides of diphenyl ethers, phenyl heterocycles, and heterocyclic carboxamides.[2] These known Protox inhibitors apparently compete with protoporphyrinogen IX (Protogen) at or near the catalytic site on the enzyme, so Protogen accumulates in plastids and diffuses through the plastidial membranes into the cytosol, where it is oxidized to protoporphyrin IX (Proto) by a plasma membrane-bound Protox.[3] The Proto reacts with light to produce singlet oxygen, leading to lipid peroxidation and the destruction of cellular components.[4] The attributes of low application rates, good crop selectivity, low residue, and environmental safety exhibited by these inhibitors are important characteristics for modern agrochemicals, which have led to the rapid success of Protox inhibitors as herbicidal products and attracted a worldwide research commitment.

Cyclic imides are one of the most important kinds of Protox inhibitors. Many cyclic imide compounds, such as flumlpropyn, fluthiacet-methyl, and pentoxazone, have been developed as commercial herbicides for years.[5] Quantitative structure−activity relationship (QSAR) study is a helpful approach to quantitatively understand the relationships between molecular structures and biological activities. The first QSAR analysis of herbicidal N-aryl-3,4,5,6-tetrahydrophthalimides **1**a (see Figure 1) and related cyclic imides, such as 3-aryl-1,5-tetramethylenehydantoins **1**b (Y═O), 4-aryl-1,2-tetramethylene-triazolidine-3,5-diones **1**c (Y═Z═O) and the corresponding thiocarbonyl compounds (**1**b:Y═S; **1**c: one or both of Y and Z═S), were reported by Fujita and co-workers in 1980.[6] They found that the steric effects of aromatic substituents, as represented by Verloop's STERIMOL values, played an important role in determining the inhibitory potency of the compounds, which were confirmed by K. Wakabayashi[7] who carried out a QSAR analysis for forty N-aryl-3,4,5,6-tetrahydrophthalimides (**1**a) in 1988. However, a subsequent QSAR analysis for 11 cyclic imides (**1**a, X=4-substituents) reported by Nicolaus[8] showed



1a **1**b **1**c

Figure 1. The structures of some cyclic imide compounds involved in the QSAR analysis.

* Corresponding author: e-mail: gfyang@mail.ccnu.edu.cn.
[†] Central China Normal University.
[‡] University of Kentucky.

10.1021/ci049793p CCC: $27.50 © 2004 American Chemical Society
Published on Web 10/13/2004

that only hydrophobic property of the ligands can significantly affect the herbicidal activity of those compounds. These earlier QSAR studies were limited to the use of the physicochemical descriptors or the electronic structure descriptors based on semiempirical molecular orbital (MO) calculations.

In previous studies on the molecular similarity of Protox inhibitors,[9,10] the cyclic imides and related herbicides were similar to each other not only sterically but also electrostatically. So Uraguchi et al.[11] obtained the correlation between the Protox inhibiting activity ($p\mathbf{I}_{50}$) and the molecular shape similarity index (S), i.e. $p\mathbf{I}_{50} = 20.3(\pm 3.31)S - 8.10$, ($n$=26, R=0.93, s=0.46, F=160.5). Moreover, Uraguchi and coworkers[11,12] also studied the steric similarity between cyclic imides and Protogen. By examining the superposition patterns, cyclic imides were found to match with the C- and D-rings moiety of Protogen, which suggested that the 3D molecular-shape similarity of inhibitors to substructural Protogen are highly important in determining the inhibitory potency of a compound. However, they still used the semiempirical molecular orbital methods, MNDO and PM3, to optimize the geometries for a series of cyclic imides and Protogen. As the potential limitations of semiempirical MO methods are well-known, it is interesting to explore and determine better molecular descriptors based on electronic structure calculations at a higher level of theory for the development of a QSAR model with higher predictability.

As well-known, QSAR studies could also help to understand the molecular mechanism of biological activities of environmentally important molecules including medicines and pesticides. It is essential for this purpose to obtain QSAR equations with high-quality descriptors, because the success of a QSAR model is highly dependent upon the choice of descriptors. So exploring the use of reliable descriptors, especially descriptors based on reasonably high level electronic structure calculations, could lead to significantly better QSAR results. Semiempirical molecular orbital (MO) calculations have been used to obtain electronic descriptors for many years. However, the latest development of the computer technology and software of electronic structure theory allows calculating quantum chemical descriptors at first-principles levels, such as DFT, with higher accuracy including some effective consideration of electron correlation effects. We are particularly interested in testing quantum chemical descriptors based on density functional theory (DFT), because DFT has played an important role in many research areas of modern computational chemistry.

There have been dramatic changes in computational chemistry over the last two decades. Electronic structure theory has become an effective and powerful tool for use in predicting the properties of a wide range of molecules including geometries, energetics, reactivity, and spectroscopic properties. One of the major reasons for the acceleration of the use of electronic structure theory in predicting molecular properties for larger molecules has been the development of density functional theory (DFT). The combination of relatively low computational cost with reasonable accuracy has led to the successful application of the DFT method to the prediction of a broad range of properties of molecules. So, DFT has emerged as a practical and versatile tool to obtain accurate information on molecular systems of chemical interest.[13−22] This approach, which includes the dynamic correlation effects, represents a valid alternative to the Hartree−Fock (HF) theory, or to more sophistical post-HF methods, such as those of the Møller−Plesset theory, coupled-cluster theory, or configuration interaction approach, that are highly demanding in terms of CPU time. One of the most popularly used hybrid density functionals is Becke's three-parameter hybrid exchange functional and the Lee−Yang−Parr correlation functional (B3LYP).[16,17] DFT has become a powerful computational approach to investigate the precise electronic characteristic of molecular structure, which is the key factor to determine the interactions between the receptors and the ligands. However, DFT-based quantum chemical descriptors were seldom used to carry out quantitative structure−activity relationships study.[23]

As a part of our research interest on the structural modification and molecular design of novel Protox inhibitors, we carried out a DFT-based quantitative structure−activity relationship (DFT/QSAR) study for structurally diverse cyclic imide compounds **2** as shown in Table 1. The equilibrium geometries and electronic structures of a series of structural diverse cyclic imides were examined by the DFT method using the B3LYP functional. The quantum chemical descriptors,[24] such as the energy gap (ΔE) between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), weighted nucleophilic frontier electron densities, net atomic charge, and molecular polarizability, were derived from the DFT calculations. These precise quantum chemical descriptors were then used to carry out the QSAR study. To our knowledge, this is the first report for a DFT/QSAR study on the structurally diverse cyclic imides.

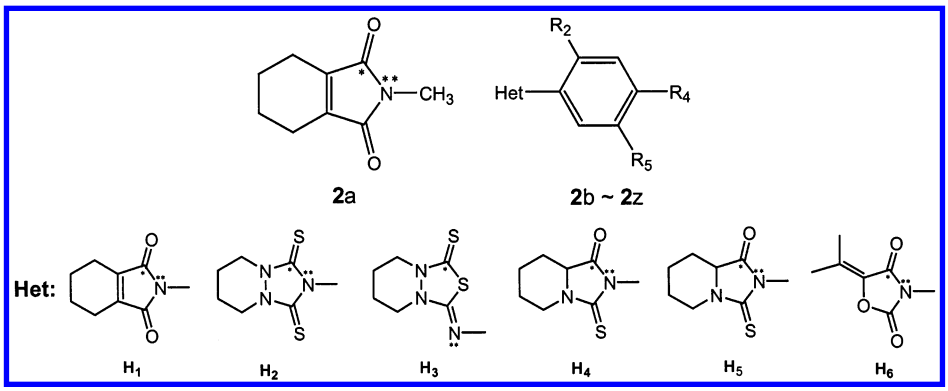
METHODS

Test Compounds. All cyclic imide compounds and their in vitro biological activity, i.e., $p\mathbf{I}_{50}$ values, used for the QSAR analysis were taken from the literature[11] and listed in Table 1.

Calculation Methods. The geometries of all cyclic imide molecules involved in this study were fully optimized by using the DFT[13−22] method with the B3LYP[17,18] functional. A standard valence double-ζ basis set, 6-31G, augmented with polarization functions, i.e., 6-31G(d,p),[25] was used for all kinds of atoms involved in cyclic imides. Vibrational frequencies were computed at the same B3LYP/6-31G(d,p) level to characterize the stationary points on the corresponding potential energy surfaces. The equilibrium geometry of Protogen was also optimized at the same B3LYP/6−31(d,p) level. A useful kind of net atomic charges, called electrostatic potential (ESP)-fitting charges, were derived from the DFT-calculated molecular electrostatic potential distribution using the Merz−Singh−Kollman (MK) method.[24,25] All of DFT calculations were performed using the Gaussian 03 program[26] on an 8-processors SGI Origin 300 server.

Quantum Chemical Descriptors. Quantum chemical descriptors (listed in Table 2) obtained from the DFT calculations were used to analyze variations in the herbicidal activity of cyclic imide compounds **2**a−**2**z by using the Hansch-Fujita method. The lowest-energy conformations were considered as the bioactive conformations, in this study, and were used to calculate electronic descriptors such as ΔE

Table 1. Molecular Structure Formula and Herbicidal Activities of Cyclic Imide Compounds, Together with the DFT-based QSAR Predicted Values[a]



	substituents				pI_{50}				
no.	Het	R_2	R_4	R_5	obsd	eq 1	δ_1[b]	eq 2	δ_2[b]
2a					4.41	4.06	0.35	4.09	0.32
2b	H_1	H	H	H	5.80	7.10	−1.30	6.99	−1.19
2c	H_1	H	Cl	H	7.60	7.65	−0.05	7.54	0.06
2d	H_1	H	Cl	OCH_2CCH	9.40	8.47	0.93	8.39	1.01
2e	H_1	H	Cl	$CO_2C_3H_7$-i	9.00	8.74	0.26	8.60	0.40
2f	H_1	H	Cl	$CO_2C_4H_9$-n	8.04	8.72	−0.68	8.59	−0.55
2g	H_1	H	Cl	$CO_2CH_2CO_2CH_3$	8.80	8.80	0.00	8.64	0.16
2h	H_1	H	Cl	$CO_2C_5H_{11}$-n	8.12	8.66	−0.54	8.55	−0.43
2i	H_1	H	Cl	$CO_2C_8H_{17}$-n	7.60	7.78	−0.18	7.79	−0.19
2j	H_1	F	Cl	OCH_2CCH	8.74	8.63	0.11	8.50	0.23
2k	H_1	F	Cl	$CO_2C_3H_7$-i	8.86	8.96	−0.10	8.73	0.13
2l	H_2	H	Br	H	8.10	8.26	−0.16	8.50	−0.40
2m	H_2	H	Cl	H	8.17	8.09	0.08	8.33	−0.16
2n	H_2	F	Cl	H	8.14	8.15	−0.01	8.42	−0.28
2o	H_2	F	Cl	OC_3H_7-i	8.92	8.41	0.51	8.73	0.19
2p	H_2	F	Cl	OCH_2CCH	9.05	8.32	0.73	8.64	0.41
2q	H_3	H	Br	H	6.13	6.24	−0.11	6.09	0.04
2r	H_3	H	Cl	H	5.96	6.17	−0.21	6.02	−0.06
2s	H_3	F	Cl	H	5.96	6.38	−0.42	6.31	−0.35
2t	H_3	F	Cl	OC_3H_7-i	6.21	6.30	−0.09	6.29	−0.08
2u	H_3	F	Cl	OCH_2CCH	6.96	6.54	0.42	6.58	0.38
2v	H_4	H	Br	H	7.77	8.18	−0.41	8.08	−0.31
2w	H_5	H	Br	H	8.29	8.44	−0.15	8.34	−0.05
2x	H_6	H	Cl	H	7.66	6.93	0.73	7.03	0.63
2y	H_6	F	Cl	O-cyclo-C_5H_9	8.28	8.23	0.05	8.35	−0.07
2z	H_6	F	Cl	$OCH(CH_3)CCH$	8.57	8.33	0.24	8.39	0.18

[a] *Carbon atom used in the approximate nucleophilic superdelocalizability descriptor. **Nitrogen atom used in the net charge descriptor. [b] Difference between observed and predicted values.

Table 2. Quantum Chemistry Descriptors Calculated at DFT-B3LYP/6-31G(d,p) Level for the Cyclic Imides of Interest

	descriptors				descriptors		
no.	α	$F^N_{C\bullet}$	$Q_{N\bullet\bullet}$	no.	α	$F^N_{C\bullet}$	$Q_{N\bullet\bullet}$
2a	100.51[a]/81.81[b]	2.61/1.31	−0.17/−0.15	**2**n	193.81/174.36	3.98/3.56	0.90/0.18
2b	151.76/125.42	3.21/1.40	−0.52/−0.09	**2**o	232.48/201.91	3.98/3.50	1.05/0.17
2c	166.17/137.19	3.36/1.48	−0.52/−0.09	**2**p	223.89/198.82	4.24/3.51	0.92/0.17
2d	202.79/168.47	3.63/1.53	−0.45/−0.09	**2**q	215.85/188.95	8.58/6.69	−0.54/−0.06
2e	219.27/176.32	3.34/1.49	−0.52/−0.09	**2**r	207.06/187.32	8.59/6.77	−0.51/−0.05
2f	231.59/183.49	3.34/1.49	−0.54/−0.09	**2**s	207.33/189.77	8.15/6.66	−0.52/−0.04
2g	225.27/184.57	3.23/1.52	−0.53/−0.09	**2**t	246.09/211.43	8.14/6.65	−0.48/−0.04
2h	242.96/190.27	3.23/1.49	−0.49/−0.09	**2**u	238.92/212.69	7.81/6.41	−0.47/−0.05
2i	276.62/212.32	3.35/1.49	−0.54/−0.09	**2**v	175.85/144.57	2.96/0.93	0.06/0.03
2j	202.33/170.63	3.27/1.44	−0.34/−0.09	**2**w	187.04/153.33	3.07/0.99	0.01/0.03
2k	219.60/178.09	2.88/1.40	−0.43/−0.10	**2**x	159.06/135.16	4.27/2.44	−0.30/−0.06
2l	202.36/174.94	4.06/3.53	0.76/0.17	**2**y	216.88/178.35	4.41/2.34	−0.10/−0.05
2m	194.23/172.80	4.12/3.57	0.68/0.17	**2**z	204.92/174.49	3.98/1.37	−0.06/−0.06

[a] Quantum chemical descriptors based on B3LYP/6-31G(d,p) level. [b] Quantum chemical descriptors based on PM3 method.

(the difference between LUMO and HOMO orbital energy), α (mean molecular polarizability), $F^N_{C\bullet}$ (weighted nucleophilic frontier electron density of the carbon atom marked with asterisk symbol on the heterocyclic ring in Table 1), and $Q_{N\bullet\bullet}$ (the net atomic charge of the nitrogen atom marked with two asterisk symbols on the heterocyclic ring).

Table 3. QSAR Development of Eqs 1 and 2

intercept	α	α^2	$F^N_{C\bullet}$	$Q_{N\bullet\bullet}$	s	r	F-test
4.7356	0.0145				1.1735	0.4061	4.74
−4.9468	0.1203	−0.0003			1.0409	0.6086	6.76
−6.3319	0.1493	−0.0003	−0.4733		0.5049	0.9264	44.41
−6.4056	0.1383	−0.0003	−0.0398	0.2076	0.4692	0.9398	39.69

Table 4. Correlation Matrix for Variables Used To Drive Eqs 1 and 2

	α	α^2	$F^N_{C\bullet}$	$Q_{N\bullet\bullet}$
α^2	0.9772	1.0000		
$F^N_{C\bullet}$	0.0660	0.0576	1.0000	
$Q_{N\bullet\bullet}$	0.0048	0.0099	0.0649	1.0000

The weighted electrophilic (F^E_i) and nucleophilic (F^N_i) atomic frontier electron density is defined as follows

$$F^{\mathrm{E}}_i = \frac{\sum (C_i^{\mathrm{HOMO}})^2}{\Delta E} \times 100$$

$$F^{\mathrm{N}}_i = \frac{\sum (C_i^{\mathrm{LUMO}})^2}{\Delta E} \times 100$$

where $C_i^{\mathrm{HOMO/LUMO}}$ are the coefficients of the atomic orbitals of atom i in the HOMO and LUMO and ΔE is the HOMO−LUMO energy gap.

Mean polarizability of the molecule[24] is defined as

$$\alpha = \frac{(\alpha_{xx} = \alpha_{yy} = \alpha_{zz})}{3}$$

where α_{xx}, α_{yy}, and α_{zz} are the diagonal elements of molecular polarizability tensor.

RESULTS AND DISCUSSION

QSAR Analysis and Model Validation. Variations in the herbicidal activity of cyclic imide compounds (**2**a−**2**z, Table 1) were analyzed using the quantum chemical descriptors obtained from the DFT calculations, and the QSAR equation with the largest correlation coefficient was obtained as follows:

$$p\mathbf{I}_{50} = \underset{(0.0990)}{-6.3319} + \underset{(0.0192)}{0.1493}\alpha - \underset{(0.0000)}{0.0003}\alpha^2 - \underset{(0.0544)}{0.4733}F^N_{C\bullet}$$

$$n = 26,\ Q^2 = 0.73(Q = 0.853),\ s_{press} = 0.700,$$
$$R^2 = 0.86(R = 0.926),\ s = 0.5049,\ F = 44.41 \quad (1)$$

The development of this equation and the intercorrelation of variables are shown in Tables 3 and 4, respectively. In this and other equations to be discussed below, n is the number of compounds, Q^2 is the cross-validated correlation coefficient using the Leave-One-Out (LOO) method, s_{press} is the root mean predictive error sum of squares, R^2 is the correlation coefficient, s is the standard error, and F is F-test value. The figures in parentheses under each coefficient are the 95% confidence intervals of the regression coefficient. α is the mean molecular polarizability, $F^N_{C\bullet}$ is the approximate nucleophilic superdelocalizability of the carbon atom marked with the asterisk symbol on the heterocyclic ring in Table 1, and $Q_{N\bullet\bullet}$ is the net atomic charge of the nitrogen atom marked with two asterisk symbols on the heterocyclic ring. The plot of the observed versus predicted $p\mathbf{I}_{50}$ values in eq 1 of different cyclic imide compounds is depicted in Figure 3 (a).

A stepwise regression analysis for the development of eq 1 in Table 3 indicates that α and α^2 terms account for 37% of the biological variance. Introducing the $F^N_{C\bullet}$ term improves the explained variance to 86%. These data suggest that α and $F^N_{C\bullet}$ terms are the most important descriptors in determining the biological activity of these compounds. In eq 1, the negative coefficient of the α^2 term indicates that variations in the activity are parabolically related to the mean polarizability of the molecule. The theoretically optimum α value in eq 1 is 219.67. The $F^N_{C\bullet}$ term represents the reactivity and the relative electronic effect of the carbon atom marked with the asterisk symbol on the heterocyclic ring in Table 1, suggesting that the ability of the carbonyl (or thiocarbonyl) group to accept electrons from an electron-rich receptor significantly affects the in vitro activity of the target compounds.

To investigate the electronic effects of the nitrogen atom marked with two asterisk symbols on the heterocyclic ring, the descriptor of its net charge is added into eq 1 for the QSAR study, and the new equation is obtained as follows:

$$p\mathbf{I}_{50} = \underset{(0.0961)}{-5.7414} + \underset{(0.0192)}{0.1424}\alpha - \underset{(0.0000)}{0.0003}\alpha^2 - \underset{(0.0542)}{0.4546}F^N_{C\bullet} + \underset{(0.1932)}{0.2974}Q_{N\bullet\bullet}$$

$$n = 26,\ Q^2 = 0.76(Q = 0.872),\ s_{press} = 0.669,$$
$$R^2 = 0.87(R{=}0.934),\ s = 0.4899,\ F = 35.98 \quad (2)$$

Comparing eqs 1 and 2, one can see that the result of QSAR in eq 2 is a bit better than that of eq 1 with one more descriptor. This indicates the net atomic charge derived from the molecular electrostatic potential distribution may be also an important descriptor in determining the biological activity. The plot of the observed versus predicted $p\mathbf{I}_{50}$ values in eq 2 of different cyclic imide compounds is depicted in Figure 3 (b).

Since electronic polarizability is not only a measure of molecular volume but also that of hydrophobicity, the above correlation equation does not exclude the possibility of steric and hydrophobic effects on the activity.[29] CMR is a kind of descriptor which may be used to represent the steric effect in most cases. So we compared the correlation between electronic polarizability and molar refractivity (CMR) and between electronic polarizability and hydrophobic descriptor (c **log** P)), leading to the following explicit correlation relationships:

$$\mathrm{CMR} = \underset{(0.0693)}{0.5081} + \underset{(0.0020)}{0.0384}\alpha$$

$$n = 26,\ R = 0.9688,\ s = 0.3536,\ F = 366.56 \quad (3)$$

$$c\ \mathbf{log}\ P = \underset{(0.1244)}{-2.0058} + \underset{(0.0036)}{0.0313}\alpha$$

$$n = 26,\ R = 0.8715,\ s = 0.6342,\ F = 75.80 \quad (4)$$

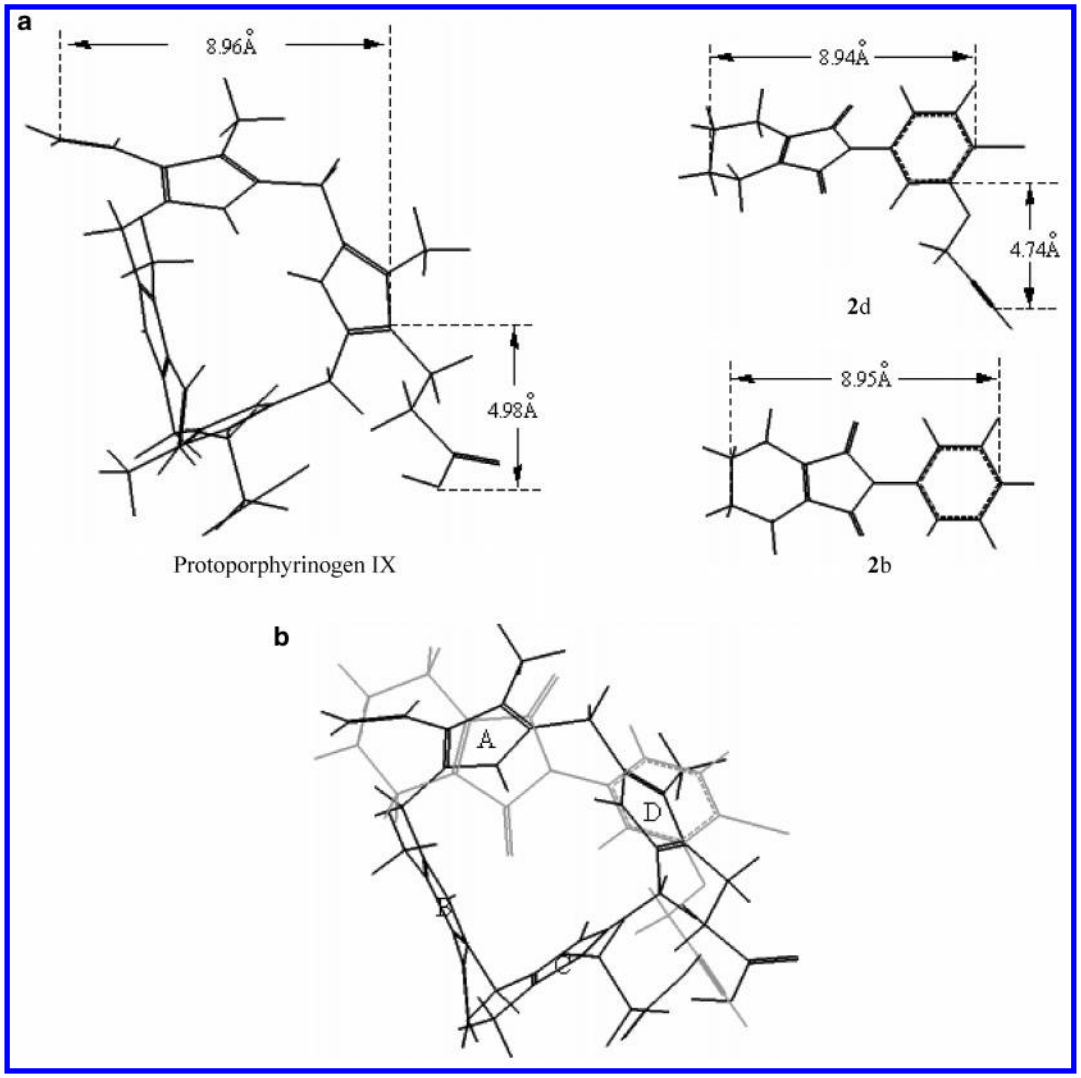


Figure 2. (a) Molecular size of compounds **2**d/**2**b and protoporphyrinogen IX. (b) Matching of compound **2**d with protoporphyrinogen IX.

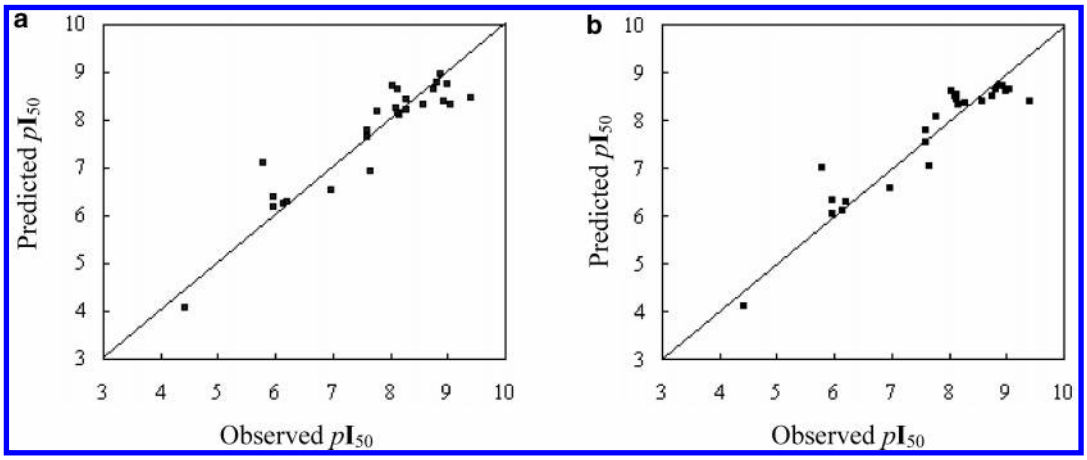


Figure 3. (a) Plot of observed (from eq 1) versus predicted activities pI_{50} and (b) plot of observed (from eq 2) versus predicted activities pI_{50} for cyclic imides.

From eqs 3 and 4, we can conclude that the descriptor α in eqs 1 and 2 is correlated with not only the steric effect but also the hydrophobicity of the target compounds. The negative coefficient of α^2 in eqs 1 and 2 also indicated that the target compounds should have suitable bulk and hydrophobic effects to match the receptor.

To compare the difference between the DFT and semiempirical methods, the QSAR analysis of compounds **2**a−**2**z was also carried out by using the descriptor values calculated by using MOPAC program[30] (version 6.0) with PM3 parametrizations. Based on the descriptor values calculated at the PM3 level, we obtained the follow-

ing explicit correlation relationship:

$$p\mathbf{I}_{50} = \underset{(0.1251)}{-4.9185} + \underset{(0.0316)}{0.1484\alpha} - \underset{(0.0001)}{0.0004\alpha^2} - \underset{(0.0751)}{0.4273F_{C\bullet}^{N}}$$

$$n = 26,\ Q^2 = 0.55(Q = 0.742),\ s_{press} = 0.900,$$
$$R^2 = 0.77(R{=}0.880),\ s = 0.6378,\ F = 25.10 \quad (5)$$

$$p\mathbf{I}_{50} = \underset{(0.1098)}{-3.0459} + \underset{(0.0288)}{0.1274\alpha} - \underset{(0.0000)}{0.0003\alpha^2} - \underset{(0.0671)}{0.4623F_{C\bullet}^{N}} + \underset{(1.1677)}{3.2057Q_{N\bullet\bullet}}$$

$$n = 26,\ Q^2 = 0.65(Q = 0.806),\ s_{press} = 0.818,$$
$$R^2 = 0.83(R = 0.913),\ s = 0.5600,\ F = 26.31 \quad (6)$$

A comparison of the DFT-based QSAR result (eqs 1 and 2) with the PM3-based QSAR result (eqs 5 and 6) clearly demonstrates that DFT-based quantum chemical descriptors led to the better correlation relationship than that the PM3-based, although the PM3-based QSAR results are also acceptable. As well-known, PM3 calculations cost less time than that of DFT calculations; however, the development of the computer technology fastens the CPU speed, which enables the DFT calculations to be routine to date.

Validation is a crucial aspect of any QSAR modeling. Most of the QSAR modeling methods implement the leave-one-out (or leave-some-out) cross-validation procedure.[31] The outcome from the cross-validation procedure is cross-validated Q^2, which is used as a criterion of both robustness and predictive ability of the model. It has been shown that a cross-validated Q^2 of 0.3 corresponds to a probability of chance correlation with an activity of less than 0.05, hence a Q^2 value of 0.3 or more is considered to be statistically significant.[32−34] Many authors consider high Q^2 (e.g. $Q^2 > 0.5$) as an indicator or even as the ultimate proof that the model is highly predictive.[35] The models of eqs 1, 2, 5, and 6 have the following statistic results $Q^2 = 0.73$, $s_{press} = 0.700$, $Q^2 = 0.76$, $s_{press} = 0.669$, $Q^2 = 0.55$, $s_{press} = 0.900$, $Q^2 = 0.65$, $s_{press} = 0.818$, respectively. These data clearly suggest that the model associated with eqs 1 and 2 have higher predictability and should be more reliable than that with eqs 5 and 6.

Geometries of 2b, 2d, and Protogen IX. Through the comparison of the difference between observed and predicted values (Table 1) from QSAR eqs 1 and 2, we can find that compounds **2**b and **2**d are outlier. So we specially compared the geometries of compounds **2**d and **2**b with that of Protogen (Figure 2). From Figure 2 (a), we can see that the distances between the carbon atom at position 4 and the carbon atom at position 19 of compounds **2**d and **2**b are 8.94 Å and 8.95 Å, respectively, which are almost identical to that (8.96 Å) between the carbon atom at position 78 and the carbon atom at position 20 of Protogen. Moreover, the distance (4.74 Å) between the carbon atom at position 20 and the carbon atom at position 31 in the 3-propargyloxy group of compound **2**d corresponds to that (4.98 Å) between the carbon atom at position 20 and the oxygen atom at position 67 of the 2-carboxyethyl group of Protogen. Compound **2**b has no substituents other than hydrogen atoms on the phenyl group, which might be the reason compound **2**b is outlier. Uragu-chi[11] examined a number of superimposition patterns of compound **2**d with a substructural component of Protogen and found the phenyl ring and the imide moiety of compound **2**d were shown to closely superimpose the C and D rings of Protogen, respectively. However, the cyclohexene group binding to the imide moiety of compound **2**d did not match well with the methyl group on the D ring of Protogen, and the steric effect of the 2-carboxyethyl group on the D ring was not taken into consideration in their superimposition pattern. In our superimposition model, shown in Figure 2 (b), the steric dimensions of the cyclohexene group binding to the imide moiety of compound **2**d and allyl group on the A ring of Protogen are very similar. These results suggested that compound **2**d exhibited a great similarity to the A and D rings of Protogen, which can reasonably explain why compound **2**d has the highest activity.

CONCLUSION

The present study was performed to examine the applicability of DFT-based quantum chemical descriptors in QSAR analysis for studying the biological activity of a series of Protox inhibitors. The DFT-based quantum chemical descriptors were obtained at the B3LYP/6-31G(d,p) level. It has been shown that the use of the DFT-based quantum chemical descriptors indeed led to a better QSAR equation than that obtained from the use of the corresponding descriptors evaluated at a semiempirical PM3 level.

The obtained QSAR results based on the DFT-based descriptors demonstrate that the biological activity was correlated parabolically with the mean molecular polarizability α. The carbon atom marked with the asterisk symbol on the heterocyclic ring in Table 1 was an electron-accepting site binding to the electronegative region of receptor. The net charges of the nitrogen atom marked with two asterisk symbols on the heterocyclic ring are also important to the biological activity. Generally speaking, molecular polarization can be represented in terms of the *n*th order susceptibility tensors of the molecular bulk. The first-order term is referred to as the polarizability; the second-order term is called the first hyperpolarizability, etc. Thus, the most significant property of the molecular polarizability is the relation to the molecular bulk or molar volume. Polarizability values have been shown to be correlated with hydrophobicity and thus to other biological activities. The molecular polarizability term in eqs 1 and 2 can reflect both the steric influence and the hydrophobicity effects. Specifically, the comparison of geometries of compound **2**d (which has the highest activity) and Protogen suggests that 3D-structure similarity between the inhibitors and a part of the natural substrate of the enzyme is important for the molecular recognition at the active site of the Protox. The DFT-based QSAR could be expected to help facilitate the future design of additional substituted cyclic imide derivatives of Protox inhibitors with potentially higher biological activity.

ACKNOWLEDGMENT

The present work was supported by the National Key Project for Basic Research (2003CB114400, 2002CCA00500), National Natural Science Foundation of China (No. 20172017 and 20203009), Program for New Century Excellent Talents in University of China, Natural Science Foundation of Hubei

Province (No. 2002AB056), and the China Scholarship Council.

Supporting Information Available: The equilibrium geometries optimized at the B3LYP/6-31G(d,p) level for all cyclic imides and protoporphyrinogen IX (Table 1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Beale, S. I.; Weinstein, J. D. In *Biosynthesis of Heme and Chlorophylls*; Dailey, H. A., Ed.; McGraw-Hill: New York, 1990; pp 287−391.

(2) Anderson, R. J.; Norris, A. E.; Hess, F. D. In *Porphyric pesticides: Chemistry, Toxicology and Pharmaceutical Applications*; Duke, S. O., Rebeiz, C. A., Eds.; ACS Symposium Series 559, American Chemical Society: Washington, DC, 1994; pp 18−33.

(3) Dayan, F. E.; Duke, S. O. Phytotoxicity of protoporphyrinogen oxidase inhibitors: phenomenology, mode of action and mechanisms of resistance. In *Herbicide Activity: Toxicology, Biochemistry and Molecular Biology*; Roe, R. M., Burton, J. D., Kuhr, R. J., Eds.; I.O.S. Press: Amsterdam, The Netherlands, 1997; pp 11−35.

(4) Gupta, I.; Tripathy, B. C. Oxidative stress in cucumber (*Cucumis Sativus* L.) seedlings treated with acifluorfen. *Indian J. Biochem. Biophys.* **2000**, *37*, 498−505.

(5) Wakabayashi, K.; Böger, P. Peroxidiazing herbicides (II): Structure−activity relationship and molecular design. *Z. Naturforsch. C* **1995**, *50*, 591−601.

(6) Ohta, H.; Jikihara, T.; Wakabayashi, K.; Fujita, T. Quantitative structure−activity study of herbicidal N-aryl -3,4,5,6- tetrahydrophthalimides and related cyclic imides. *Pestic. Biochem. Physiol.* **1980**, *4*, 153−160.

(7) Wakabayashi, K. Molecular design of cyclic imide herbicides using biorational approaches. *J. Pestic. Sci*. **1988**, *13* (2), 337−361.

(8) Nicolaus, B.; Sandmann, G.; Böger, P. Molecular aspects of herbicide action on proporphyrinogen oxidase. *Z. Naturforsch. C* **1993**, *48*(3−4), 326−333.

(9) Wakabayashi, K.; Nakayama, A. Molecular similarity of peroxidizing herbicides: bioisosterism in Δ^2-1,2,4-thiadiazolines and related heterocyclic compounds. *J. Pestic. Sci*. **1994**, *19*, 111−117.

(10) Nakayama, A.; Hagiwara, K.; Hashimoto, S.; Hosaka, H. Quantitative structure−activity and molecular modeling studies of novel fungicides and herbicides having 1,2,4-thiadiazoline structures. *Classical and Three-Dimensional QSAR in Agrochemistry*; Hansch, C., Fujita, T., Eds.; ACS Symposium Series 606, American Chemical Society: Washington, DC, 1995; pp 213−228.

(11) Uraguchi, R.; Sato, Y.; Nakayama, A.; Sukekawa, M.; Iwataki, I.; Böger, P.; Wakabayashi, K. Molecular shape similarity of cyclic imides and protoporphyrinogen IX. *J. Pestic. Sci*. **1997**, *22*, 314−320.

(12) Sato, Y.; Nnakayama, A.; Sukekawa, M.; Iwataki, I.; Böger, P.; Wakabayashi, K. Molecular shape similarity between cyclic imides and protoporphyrinogen IX. *Pestic. Sci*. **1999**, *55*, 345−347.

(13) Par, R. G.; Yang, W. *Density-functional theory of atoms and molecules*; Oxford University Press: Oxford, 1989.

(14) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.

(15) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(16) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.

(17) (a) Becke, A. D. A new mixing of Hartree−Fock, local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377. (b) Becke, A. D. Density-functional thermochemistry III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(18) (a) Hirata, S.; Zhan, C.-G.; Aprà, E.; Windus, T. L.; Dixon, D. A. A new, self-contained asymptotic correction scheme to exchange-correlation potentials for time-dependent density functional theory. *J. Phys. Chem. A* **2003**, 107, 10154−10158. (b) Zhan, C.-G.; Dixon, D. A. A density functional theory approach to the development of *Q*−*e* parameters for the prediction of reactivity in free-radical copolymerizations. *J. Phys. Chem. B* **2002**, *106*, 10311−10325. (c) Zhan, C.-G.; Dixon, D. A.; Matsuzawa, N. N.; Ishitani, A.; Uda, T. Time-dependent density functional theory calculations of the photoabsorption of fluorinated alkanes. *J. Fluorine Chem.* **2003**, *122*, 27−35. (d) Zhan, C.-G.; Nichols, J. A.; Dixon, D. A. Ionization potential, electron affinity, electronegativity, hardness, and electron excitation energy: molecular properties from density functional theory orbital energies. *J. Phys. Chem. A* **2003**, *107*, 4184−4195. (e) Zhan, C.-G.; Dixon, D. A. Electronic excitations in pyrrole: A test case for determination of chromophores in the chromogenic effects of neurotoxic hydrocarbons by time-dependent density functional theory and single-excitation configuration interaction methods. *J. Mol. Spectrosc.* **2002**, *216*, 81−89.

(19) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **1992**, *45*, 13244−13249.

(20) Foresman, J. B.; Frisch, A. *Exploring chemistry with electronic structure methods*; Gaussian Inc.: Pittsburgh, PA, 1996.

(21) Sulpizi, M.; Folkers, G.; Rothlisberger, U.; Carloni, P.; Scapozza, L. Applications of Density Functional Theory-Based Methods in Medicinal Chemistry. *Quant. Struct.-Act. Relat.* **2002**, *21*, 173−181.

(22) Bernardi, F.; Bottoni, A.; Garavelli, M. Exploring Organic Chemistry with DFT: Radical, Organo-metallic, and Bio-organic Applications. *Quant. Struct.-Act. Relat.* **2002**, *21*, 128−148.

(23) Arulmozhiraja, S.; Morita, M. Structure−Activity Relationships for the Toxicity of Polychlorinated Dibenzofurans: Approach through Density Functional Theory-Based Descriptors. *Chem. Res. Toxicol.* **2004**, *17*, 348−356.

(24) Karelson, M.; Lobanov, V. S. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027−1043.

(25) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row atoms. *J. Chem. Phys.* **1988**, *89*, 2193−2218.

(26) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem*. **1990**, *11*, 431−439.

(27) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.

(28) Gaussian 03, Revision B.03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T., Jr.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 2003.

(29) Cammarata, A. An apparent correlation between the in vitro activity of chloramphenicol analogues and electronic polarizability. *J. Med. Chem.* **1967**, *10*, 525−527.

(30) J. J. P. Stewart: MOPAC version 6.0.

(31) Clark, R. D.; Sprous, D. G.; Leonard, J. M. Validating models based on large data sets. In *Rational Approaches to Drug Design*; Höltje, H.-D., Sippl, W., Eds.; Prous Science SA: Barcelona, 2001; pp 475−485.

(32) Agarwal, A.; Pearson, P. P.; Taylor, E. W.; Li, H. B.; Dahlgren, T.; Herslof, M.; Yang, Y.; Lambert, G.; Nelson, D. L.; Regan, J. W.; Martin, A. R. Three-Dimensional Quantitative Structure−Activity Relationships of S−HT Receptor Binding Data for Tetrahydropyridinylindole Derivatives: A Comparison of the Hansch and CoMFA Methods. *J. Med. Chem*. **1993**, *36*, 4006−4014.

(33) Clark, M.; Cramer, R. D. I.; Jonea, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative molecular field analysis (CoMFA). 2. Towards its use with 3-D structural databases. *Tetrahedron Comput. Methods*. **1990**, *3*, 47−59.

(34) Thomas, B. F.; ComDton, D. R.; Martin, B. R.; Seamus, S. F. Modeling the cannabinoid receptor: a three-dimensional quantitative structure−activity analysis. *Mol. Pharmacol*. **1991**, *40*, 656−665.

(35) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graph. Model.* **2002**, *20*, 269−276.

CI049793P