Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Bioorganic & Medicinal Chemistry 12 (2004) 701–713
Bioorganic & Medicinal Chemistry

# QSAR and kinetics of the inhibition of benzaldehyde derivatives against *Sacrophaga neobelliaria* phenoloxidase

Wei Li and Isao Kubo*

*Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720-3112, USA*

Received 26 September 2003; accepted 17 November 2003

**Abstract**—A group of 18 benzaldehyde derivatives was characterized as a family of mixed type inhibitors on the oxidation of L-3, 4-dihydroxyphenylalanine (L-DOPA) catalyzed by *Sacrophaga neobelliaria* phenoloxidase (PO), presumably by forming a Schiff base with a primary amino group of the enzyme. Inhibition constants and $IC_{50}$ were determined. Cuminaldehyde was the most active inhibitor with an $IC_{50}$ of 0.0067 mM. Vanillin was the least active inhibitor with an $IC_{50}$ of 38 mM. Four physicochemical descriptors were identified by stepwise multiple regressions as significant predictors on the inhibition activity, which were further rotated by principle component analysis yielding two significant principle properties. It was shown that hydrophobicity of the substituent at the para position of the aldehyde group played major role on inhibition activity: one unit increase in Hansch–Fujita $\pi$ value of the substituent led to about 4.5 [95% confidence interval is (7.9, 2.6)] fold increase on $IC_{50}$. Electron-donating effect of the substituent at the *para* position of the aldehyde group was less important than hydrophobicity. Hydroxyl group at the ortho position of the aldehyde group contributed to higher inhibition activity, presumably by forming a quasi-six-membered ring with the unshared pair of electrons on the nitrogen atom of the amino group through intramolecular hydrogen bonding.


## 1. Introduction

Proteins with phenoloxidase (PO) (tyrosinase or polyphenol oxidase) activity are widely distributed among animals, plants, fungi, and prokaryotes.[1] They incorporate oxygen into other molecules, converting monophenols to *o*-diphenols (monophenolase) and diphenols to the corresponding *o*-quinones (diphenolase).[2] Insect PO (Monophenol, dihydroxyphenylalanine: oxygen oxidoreductase; EC 1.14.18.1) is present in insect hemolymph or in cuticle in an inactive form called prophenoloxidase (ProPO).[3,4] After activation of ProPO, *o*-quinones are subsequently formed from tyrosine derivatives. Production of *o*-quinones by PO has been known to be an initial step in the biochemical cascade of B-sclerotization, quinone tanning and melanin biosynthesis, processes that occupy several major roles in insect development and immunity, such as cuticle sclerotization, wound healing, and defense against foreign pathogens.[5–8] Thus, studying the enzymatic behavior of PO could be used for finding ways of weakening the protection afforded by the cuticle and defense reactions of insect, and could provide theoretic basis for designing more effective and environmentally safe pest management agents.

In our previous study, we found that benzaldehyde and its derivatives were a family of potential inhibitors against mushroom tyrosinase, presumably by forming a Schiff base with a primary amino group of the enzyme.[9–11] In our continuous search for alternative pest insect control agents, a variety of naturally occurring aldehydes common in plants were screened, among which a number of promising pest insect control agents were identified.[12,13] In this paper, we explored the inhibition behavior of a group of 18 benzaldehyde derivatives (Table 1) against PO of *Sacrophaga neobelliaria*. From experimental data, we determined the inhibition types, computed inhibition constants and $IC_{50}$ values of the compounds. With the data obtained from the kinetics study, we tried to relate the physicochemical properties of the compounds to their inhibition activities to uncover the structure and activity relationship of these inhibitors against PO.

*Keywords:* QSAR; Enzyme inhibition; Enzyme kinetics; Schiff base; Benzaldehyde derivatives; Phenoloxidase; *Sacrophaga neobelliaria*.
* Correspondence author. Tel.: +1-510-643-6303; fax: +1-510-643-0215; e-mail: ikubo@uclink.berkeley.edu

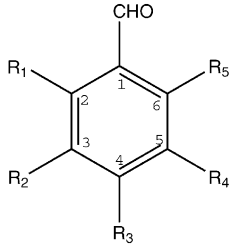
0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.11.014

## 2. Results

Activities and types of the inhibition of the group of 18 compounds (Table 1) on the oxidation of L-DOPA catalyzed by *S. neobelliaria* PO are summarized in Table 1. Type of inhibition is determined by comparing the goodness-of-fit of each inhibition model to the data, which is measured by AIC values. The model thus selected is mixed type of inhibition for all compounds. Figure 1 defines the mechanism of a mixed type inhibition: inhibitor (I) binds to an inhibitor site and brings about a conformational change in the active site of the enzyme (E), which prevents the enzyme from converting the bound substrate (EIS) to product (P).[14] Because a mixed inhibitor can bind to either the free form of enzyme (E) or the enzyme–substrate complex (ES), there are two inhibition constants: $K_{ic}$ measures EI affinity, and $K_{iu}$ measures EIS affinity (Table 1). $IC_{50}$ provides another measurement on inhibition activity, which tells the concentration of the inhibitor under which the enzyme shows only 50% of the original activity (Fig. 2). $IC_{50}$ depends on not only the properties of the inhibitor itself but also the substrate concentration. All measures of $IC_{50}$ as shown in Table 1 were obtained at 1 mM of L-DOPA. The binding of the inhibitor to the enzyme may either reduce enzyme-substrate affinity (increase $K_m$) or has no effect on substrate binding at all ($K_m$ unchanged). In the latter case, it is specifically called noncompetitive inhibitor. For both cases, $V_{max}$ decreases. Table 1 shows the effect of the inhibitor on $K_m$ and $V_{max}$ by two measurements: $K'_m/K_m$ and $V'_{max}/V_{max}$, where $K'_m$ and $V'_{max}$ are corresponding apparent values measured in the presence of inhibitors at the concentration of $IC_{50}$.

Physicochemical properties of an inhibitor determine its inhibitory effect on an enzyme. There are 18 compounds in our set of inhibitors, with the aldehyde group as their common functional group and any other groups on the benzene ring as substituents. Each substituent has its own electronical, hydrophobic, and steric properties, which were measured by Hammett's substituent constant (σ), Hansch–Fujita constant (π), and molar refractivity (MR), respectively. σ reflects the electron donating or accepting properties of a substituent.[15] π describes the contribution of a substituent to the lipophilicity of a compound.[16,17] MR is the molar volume corrected by the refractive index and measures the size and polarizability of the substituent.[18] For the set of inhibitors we tested, their physicochemical properties were summarized in Table 2, based on which we computed their dissimilarity matrix and grouped them into different clusters (Fig. 3). The results of a cluster analysis can be visualized by a dendrogram, from which the dissimilarity between two compounds can be read from the height at which they join a single group. Figure 3a shows the dendrogram ($D_1$) clustered on $V_1$ through $V_{13}$ as in Table 2. Figure 3b shows the dendrogram ($D_2$) clustered on $V_6$, $V_4$, $V_2$ and $V_{12}$, which were selected by the stepwise regression as a parsimonious subset of pre-

**Table 1.** Chemical structure of benzaldehyde derivatives and their inhibitory effects on the oxidation of L-DOPA catalyzed by *S. neobelliaria* phenoloxidase. Index numbers for 18 compounds are used throughout the paper




| Compd | Substituents | $IC_{50}$ (mM)[a] | $K_{iu}$ (mM) | $K_{ic}$ (mM) | $K'_m/K_m$[b] | $V'_{max}/V_{max}$[c] |
|---|---|---|---|---|---|---|
| **1** | $R_1 = R_3 = OMe$ | 4.5171 ± 0.2134 | 12.4670 | 2.5769 | 2.02 | 0.73 |
| **2** | $R_3 = NMe_2$ | 1.0492 ± 0.0372 | 1.0032 | 1.1081 | 0.95 | 0.49 |
| **3** | $R_1 = OH$ | 4.2442 ± 0.2464 | 13.8336 | 2.4222 | 2.11 | 0.77 |
| **4** | $R_1 = R_3 = OH$ | 2.0721 ± 0.2249 | 7.3917 | 1.0566 | 2.31 | 0.78 |
| **5** | $R_3$ = Butyl-t | 0.0190 ± 0.0007 | 0.0299 | 0.0131 | 1.50 | 0.61 |
| **6** | $R_3 = OMe$ | 0.6449 ± 0.0183 | 0.9358 | 0.4609 | 1.42 | 0.59 |
| **7** | $R_3$ = Propyl-i | 0.0067 ± 0.0001 | 0.0381 | 0.0035 | 2.47 | 0.85 |
| **8** | Benzaldehyde | 0.2499 ± 0.0211 | 0.4081 | 0.1673 | 1.55 | 0.62 |
| **9** | $R_1 = OH$, $R_3 = OMe$ | 2.2506 ± 0.1109 | 45.2806 | 0.9727 | 3.16 | 0.95 |
| **10** | $R_2 = R_3 = OMe$ | 2.3811 ± 0.0832 | 4.3858 | 1.5195 | 1.66 | 0.65 |
| **11** | $R_2 = OH$, $R_3 = OMe$ | 1.5415 ± 0.0689 | 11.3654 | 0.6903 | 2.85 | 0.88 |
| **12** | $R_1 = R_2 = OH$ | 2.5389 ± 0.1030 | 24.1689 | 1.2086 | 2.81 | 0.90 |
| **13** | $R_2 = R_3 = OH$ | 0.2522 ± 0.1571 | 0.3418 | 0.1838 | 1.36 | 0.58 |
| **14** | $R_1 = R_4 = OH$ | 1.1825 ± 0.0523 | 13.7334 | 0.5500 | 2.90 | 0.92 |
| **15** | $R_2 = OMe$, $R_3 = OH$ | 37.9076 ± 0.9874 | 145.3261 | 20.0758 | 2.29 | 0.79 |
| **16** | $R_1 = OH$, $R_2 = OMe$ | 10.8216 ± 0.3226 | 214.6025 | 4.7104 | 3.14 | 0.95 |
| **17** | $R_1 = R_5 = OH$ | 6.1195 ± 0.1809 | 21.4362 | 3.1786 | 2.28 | 0.78 |
| **18** | $R_2 = R_4 = OH$ | 1.8096 ± 0.1669 | 21.4486 | 0.8240 | 2.95 | 0.92 |

[a] Data are presented as estimate ± S.E; for details of the estimation procedure, see enzyme kinetics data analysis section.
[b] $K'_m$ and $V'_{max}$ are corresponding apparent values measured in the presence of inhibitors at the concentration of $IC_{50}$.
[c] $K'_m$ and $V'_{max}$ are corresponding apparent values measured in the presence of inhibitors at the concentration of $IC_{50}$.

dictors from $V_1$ through $V_{13}$. Figure 3c shows the dendrogram ($D_3$) clustered on $V_6$ and $V_4$. Figure 3d shows the dendrogram ($D_4$) clustered on the first two principle components ($PC_1$ and $PC_2$) obtained by principle component analysis (PCA) on $V_6$, $V_4$, $V_2$, and $V_{12}$, which therefore kept almost all clustering features in $D_2$. Clusters shown here provide a framework for comparing inhibition activity of the compounds. For example, compound **5** and **7** are uniquely grouped into a cluster; both have alkane substituents at the para position of the aldehyde group, and they are the two most active compounds of all against the enzyme.

To find the relationship between the inhibition activity and the physicochemical properties of the set of inhibitors, we performed regression analysis, for which $\log_e(IC_{50})$ is the response variable, and the physicochemical properties are the predictor variables. Stepwise regression was used to select predictors from $V_1$ to $V_{13}$ according to the improvement of goodness-of-fit. The selection procedure starts with an empty subset and at each step add the predictor that gives the largest reduction of the residual sum of squares. When each new variable is added to the subset, partial correlations are considered to see if any of the variables in the subset should now be dropped.[19] $V_6$, $V_4$, $V_2$ and $V_{12}$ were selected at the end (**Model 1**). Figure 4a–d shows the relationship between these four variables and the logarithm of $IC_{50}$. It appears that predictor $V_6$ and $V_4$ are highly correlated with $\log_e(IC_{50})$. Table 3 allow us to read off the estimated regression line as follows

$$\log_e(IC_{50}) = 0.18 - 1.24 \times V_6 + 0.79 \times V_4 + 0.53 \times V_2 - 0.43 \times V_{12}$$

(Model 1)

Table 3 shows that the effect of $V_6$ and $V_4$ on $\log_e(IC_{50})$ is significant, while the effect of $V_2$ and $V_{12}$ is not significant at 5% significance level. **Model 1** tells us that, one unit increase in π value of the substituent at the para position of the aldehyde group led to about 3.5 [95% confidence interval is (6.9, 1.8)] -fold increase in inhibition activity. One unit increase in MR value of the substituent at the meta position of the aldehyde group led to about 2.2 [95% confidence interval is (4.4, 1.1)] -fold decrease in inhibition activity, as far as $IC_{50}$ is

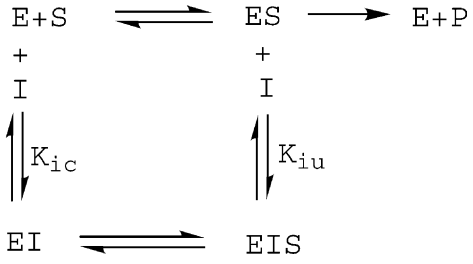



**Figure 1.** Mechanism of mixed type of inhibition. Inhibitor (I) binds to an inhibitor site and brings about a conformational change in the active site of the enzyme (E), which prevents the enzyme from converting the bound substrate (EIS) to product (P). Because a mixed inhibitor can bind to either the free form of enzyme (E) or the enzyme-substrate complex (ES), there are two inhibition constants: $K_{ic}$ measures EI affinity, and $K_{iu}$ measures EIS affinity. The binding of the inhibitor to the enzyme may either reduce enzyme-substrate affinity (increase $K_m$) or has no effect on substrate binding at all ($K_m$ unchanged). In the latter case, it is specifically called noncompetitive inhibitor. For both cases, $V_{max}$ decreases.

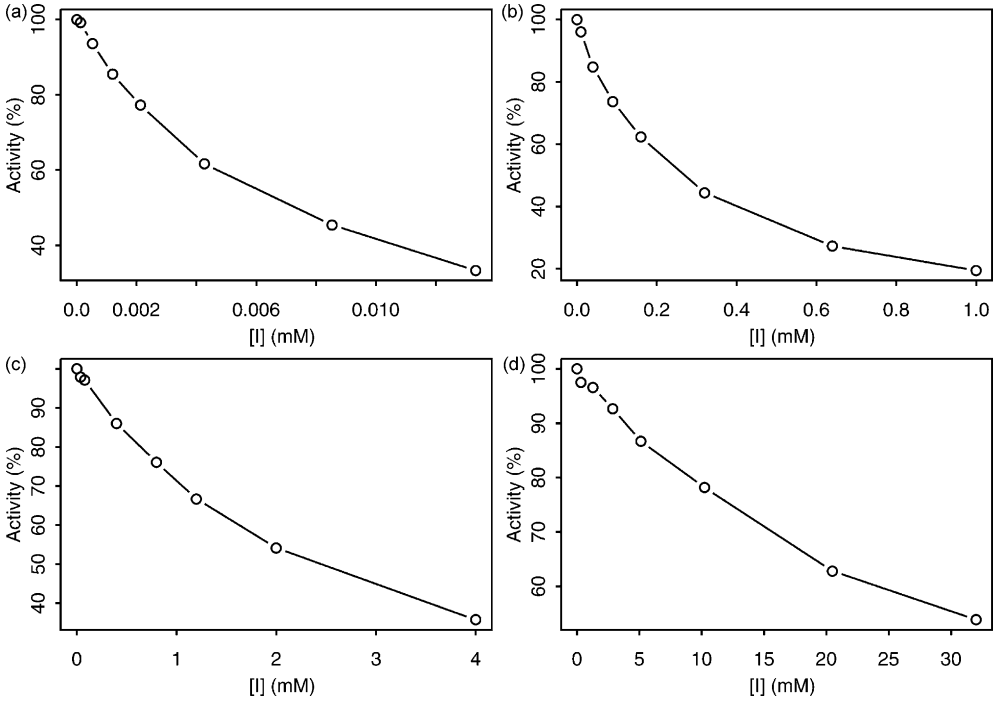



**Figure 2.** The inhibitory effect on the oxidation of L-DOPA catalyzed by *S. neobelliaria* phenoloxidase represented by four benzaldehyde derivatives. a through d are cuminaldehyde (compound **7**), 2-hydroxy-4-methoxy-benzaldehyde (compound **8**), benzaldehyde (compound **10**) and vanillin (compound **15**), respectively. For experimental conditions, see method of enzyme assay, with L-DOPA concentration fixed at 1 mM.

**Table 2.** Physicochemical properties of 18 benzaldehyde derivatives measured by individual substituents on the benzene ring

| Compd[a] | Position[b] 2 | | Position 3 | | | Position 4 | | | Position 5 | | | Position 6 | | logP[f] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\pi^c$ $V_1$ | MR[d] $V_2$ | $\pi$ $V_3$ | MR $V_4$ | $\sigma^e$ $V_5$ | $\pi$ $V_6$ | MR $V_7$ | $\sigma$ $V_8$ | $\pi$ $V_9$ | MR $V_{10}$ | $\sigma$ $V_{11}$ | $\pi$ $V_{12}$ | MR $V_{13}$ | |
| **1** | −0.02 | 7.87 | 0 | 1.03 | 0 | −0.02 | 7.87 | −0.27 | 0 | 1.03 | 0 | 0 | 1.03 | 1.53 |
| **2** | 0 | 1.03 | 0 | 1.03 | 0 | 0.18 | 15.55 | −0.83 | 0 | 1.03 | 0 | 0 | 1.03 | 2.07 |
| **3** | −0.67 | 2.85 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | 0 | 1.03 | 1.39 |
| **4** | −0.67 | 2.85 | 0 | 1.03 | 0 | −0.67 | 2.85 | −0.37 | 0 | 1.03 | 0 | 0 | 1.03 | 1 |
| **5** | 0 | 1.03 | 0 | 1.03 | 0 | 1.98 | 19.62 | −0.2 | 0 | 1.03 | 0 | 0 | 1.03 | 3.43 |
| **6** | 0 | 1.03 | 0 | 1.03 | 0 | −0.02 | 7.87 | −0.27 | 0 | 1.03 | 0 | 0 | 1.03 | 1.65 |
| **7** | 0 | 1.03 | 0 | 1.03 | 0 | 1.53 | 14.96 | −0.15 | 0 | 1.03 | 0 | 0 | 1.03 | 3.02 |
| **8** | −0.67 | 2.85 | 0 | 1.03 | 0 | −0.02 | 7.87 | −0.27 | 0 | 1.03 | 0 | 0 | 1.03 | 1.27 |
| **9** | 0 | 1.03 | −0.02 | 7.87 | 0.12 | −0.02 | 7.87 | −0.27 | 0 | 1.03 | 0 | 0 | 1.03 | 1.53 |
| **10** | 0 | 1.03 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | 0 | 1.03 | 1.78 |
| **11** | 0 | 1.03 | −0.67 | 2.85 | 0.12 | −0.02 | 7.87 | −0.27 | 0 | 1.03 | 0 | 0 | 1.03 | 1.27 |
| **12** | −0.67 | 2.85 | −0.67 | 2.85 | 0.12 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | 0 | 1.03 | 1 |
| **13** | 0 | 1.03 | −0.67 | 2.85 | 0.12 | −0.67 | 2.85 | −0.37 | 0 | 1.03 | 0 | 0 | 1.03 | 1 |
| **14** | −0.67 | 2.85 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | −0.67 | 2.85 | 0.12 | 0 | 1.03 | 1 |
| **15** | 0 | 1.03 | −0.02 | 7.87 | 0.12 | −0.67 | 2.85 | −0.37 | 0 | 1.03 | 0 | 0 | 1.03 | 1.27 |
| **16** | −0.67 | 2.85 | −0.02 | 7.87 | 0.12 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | 0 | 1.03 | 1.27 |
| **17** | −0.67 | 2.85 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | 0 | 1.03 | 0 | −0.67 | 2.85 | 1 |
| **18** | 0 | 1.03 | −0.67 | 2.85 | 0.12 | 0 | 1.03 | 0 | −0.67 | 2.85 | 0.12 | 0 | 1.03 | 1 |

[a] See Table 1 for chemical structures.
[b] See Table 1 for chemical structures.
[c] Hansch–Fujita constant.
[d] Molar refractivity.
[e] Hammett's substituent constant.
[f] The logarithm of the partition coefficient for *n*-octanol/water are computed by Crippen's fragmentation.[28]

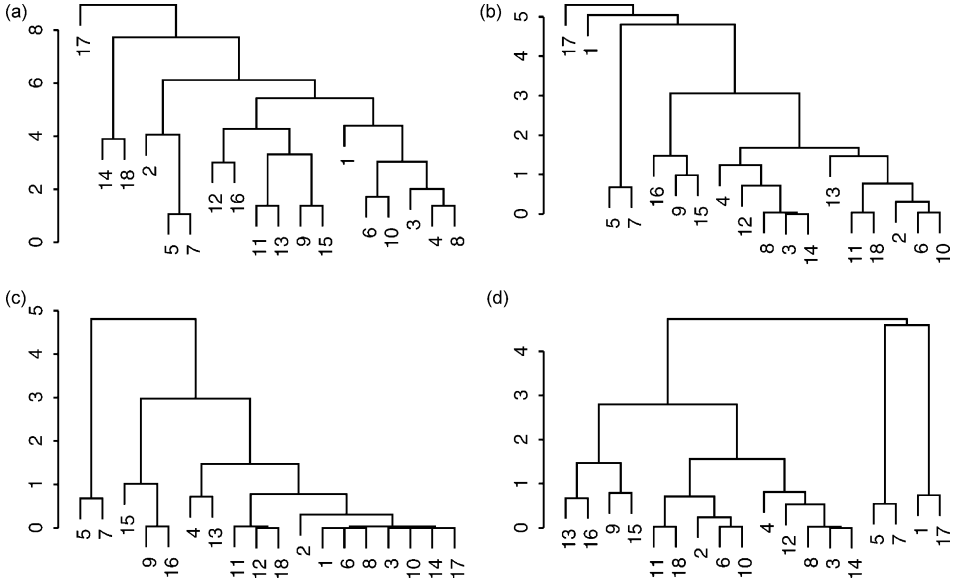



**Figure 3.** Dendrograms obtained by agglomerative hierarchical clustering on the physicochemical property data of 18 compounds. Distance between compounds was measured by eq 5. Distance between clusters was measured by eq 6. (a) Clustering on $V_1$ through $V_{13}$ as listed in Table 2; (b) clustering on $V_6$, $V_4$, $V_{12}$ and $V_2$; (c) Clustering on $V_6$ and $V_4$; (d) Clustering on $PC_1$ and $PC_2$. Dendrogram indicates not only which clusters are joined but also the distance at which they are joined. The distance between two cases can be read from the height at which they join a single group.

compared. How good is the fitted linear regression model? Figure 5a gives a plot of the response against the fitted values with the solid line being visually equivalent to the regression line, which appears to model the trend of the data reasonably well. Another useful diagnostic plot is the normal quantile plot of residuals (Fig. 6a), which provides a visual test of the assumption that the model's errors are normally distributed. Here we see that most of the ordered residuals cluster along the superimposed quantile–quantile line except compound **13**, which possibly is an outlier. Cook's distance plot (Fig. 6b) is a measure of the influence of individual observations on the regression coefficients. We see compound **13** and **7** are the most heavily influential observations on the regression coefficient. The multiple $R^2$ and F-statistic also strongly support the model (Table 3), which tells us that it explains about 74.5% of the variation in $\log_e(IC_{50})$.

Ideally, the predictors selected do not overlap in what they described, i.e., they should be orthogonal, having minimal covariance. However, for example, the corre-

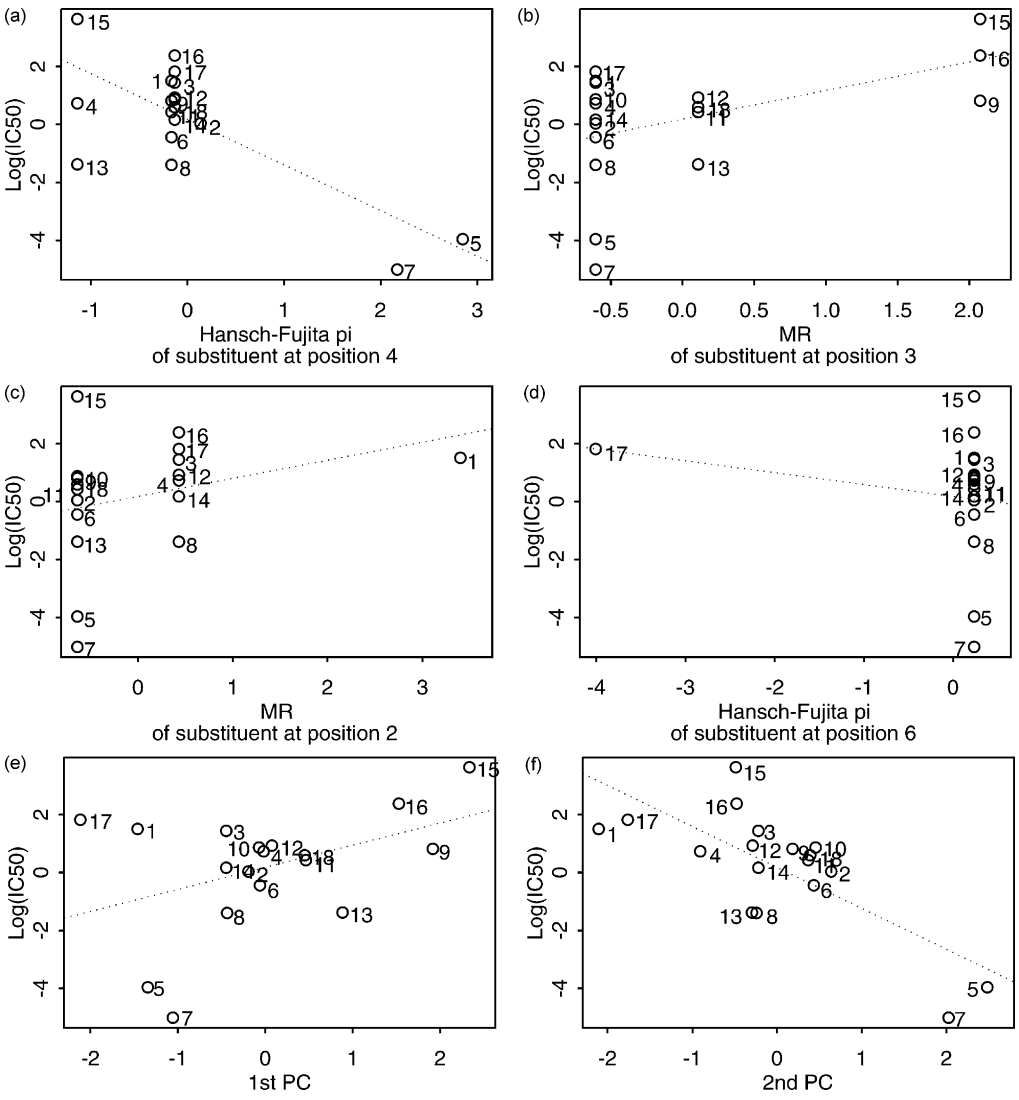



**Figure 4.** The relationship between physicochemical properties and inhibitory activities of the compounds against the oxidation of L-DOPA catalyzed by *S. neobelliaria* phenoloxidase. (a) $V_6$ versus $\log_e(IC_{50})$; (b) $V_4$ versus $\log_e(IC_{50})$; (c) $V_2$ versus $\log_e(IC_{50})$; d) $V_{12}$ versus $\log_e(IC_{50})$; (e) $PC_1$ versus $\log_e(IC_{50})$; (f) $PC_2$ versus $\log_e(IC_{50})$; Physicochemical properties showed here are identified by stepwise regression and PCA as a parsimonious subset of predictors to include in a least squares multiple regression on $\log_e(IC_{50})$. Dotted lines are regression lines obtained from simple linear regression showing the linear relationship between the predictor and $\log_e(IC_{50})$.

**Table 3.** Results obtained from regression analysis. Response variable is $\log_e(IC_{50})$. **Model 1** uses $V_6$, $V_4$, $V_2$ and $V_{12}$ as predictors. **Model 2** uses $PC_1$ and $PC_2$ as predictors. **Model 4** and **5** are the same as **Model 1** and **2**, respectively, except removing compound **13** from the data set

| Model | Term | Coef. | S.E. | Pr(>\|t\|) | $R^2$ | Pr(>F) |
|---|---|---|---|---|---|---|
| **1** | $V_6$ | −1.24 | 0.32 | 0.0017 | 0.7449 | 0 |
| | $V_4$ | 0.79 | 0.32 | 0.0276 | | |
| | $V_2$ | 0.53 | 0.31 | 0.1083 | | |
| | $V_{12}$ | −0.43 | 0.30 | 0.1725 | | |
| **2** | $PC_1$ | 0.76 | 0.24 | 0.0064 | 0.7317 | 0 |
| | $PC_2$ | −1.41 | 0.25 | 0.0001 | | |
| **4** | $V_6$ | −1.50 | 0.26 | 0.0001 | 0.8539 | 0 |
| | $V_4$ | 0.71 | 0.26 | 0.0174 | | |
| | $V_2$ | 0.35 | 0.25 | 0.1817 | | |
| | $V_{12}$ | −0.39 | 0.24 | 0.1261 | | |
| **5** | $PC_1$ | −1.03 | 0.21 | 0.0002 | 0.8231 | 0 |
| | $PC_2$ | −1.34 | 0.21 | 0.0000 | | |

lation between $V_6$ and $V_4$ is −0.29, which indicates that they are correlated negatively. PCA is a data reduction method by eigendecomposition of the matrix of the original predictors to construct a small set of new orthogonal variables, which are called principle components (PC). Results of a PCA on $V_6$, $V_4$, $V_2$ and $V_{12}$ were summarized in Figure 7. Figure 7A shows that about 64% of the total variance is accounted by the first two PCs ($PC_1$ and $PC_2$). Figure 7B shows that $V_4$ has the biggest loading on $PC_1$. Figure 7C shows that $V_6$ and $V_2$ have the biggest loading on $PC_2$. Figure 7D shows both the observations and the original variables in a two dimensional space defined by $PC_1$ and $PC_2$. A multiple linear regression with these PCs as predictors was performed, which showed that only $PC_1$ and $PC_2$ were significant. After dropping the insignificant terms, we obtained the following regression line

$$\log_e(IC_{50}) = 0.18 + 0.71 \times PC_1 - 1.41 \times PC_2$$

(Model 2)

Both the effects of $PC_1$ and $PC_2$ are very significant (Table 3). **Model 2** tells us that, one unit increase in $PC_1$value led to about 2.1 [95% confidence interval is (3.6, 1.3)] -fold decrease in inhibition activity. One unit increase in $PC_2$ value led to about 4.1 [95% confidence interval is (7.1, 2.4)] -fold increase in inhibition activity, as far as $IC_{50}$ is compared. **Model 2** fits the trend of the data reasonably well (Fig. 5b), and the model explains about 73.2% of the variation in $\log_e(IC_{50})$ (Table 3). The model errors are approximately normal except for compound **13** being a possible outlier (Fig. 6c). Compound **7**, **13**, **15** and **17** are the most heavily influential observations on the regression coefficient (Fig. 6d).

Is **Model 2** sufficient for the data? We could see some nonlinear trend from Figure 4e and f. An additive model extends the notion of a linear model by allowing linear functions of the predictors to be replaced by arbitrary smooth functions of the predictors. Figure 8a and b shows the transformation of $PC_1$ and $PC_2$ by cubic B-spline smoothing. Using the transformed $PC_1$ and $PC_2$ as predictors, we obtained **Model 3**. Figure 5c shows that **Model 3** fits the data a little better than **Model 2**. Now we have three regression models: **Model 2** has only two predictors and is the simplest model; **Model 1** has four predictors; **Model 3** includes non-linear terms as predictors and is the most complex model. To test which is the best model statistically, we did an ANOVA test (Table 4) to compare **Model 1** and **3** with the simplest **Model 2**. There is no evidence that either **Model 1** or **3** is better than **Model 2**.

Both Figure 6a and c indicated that compound **13** was a possible outlier. We eliminated this compound from the set of inhibitors and repeated the above analysis. By stepwise multiple regression, we obtained the following regression line,

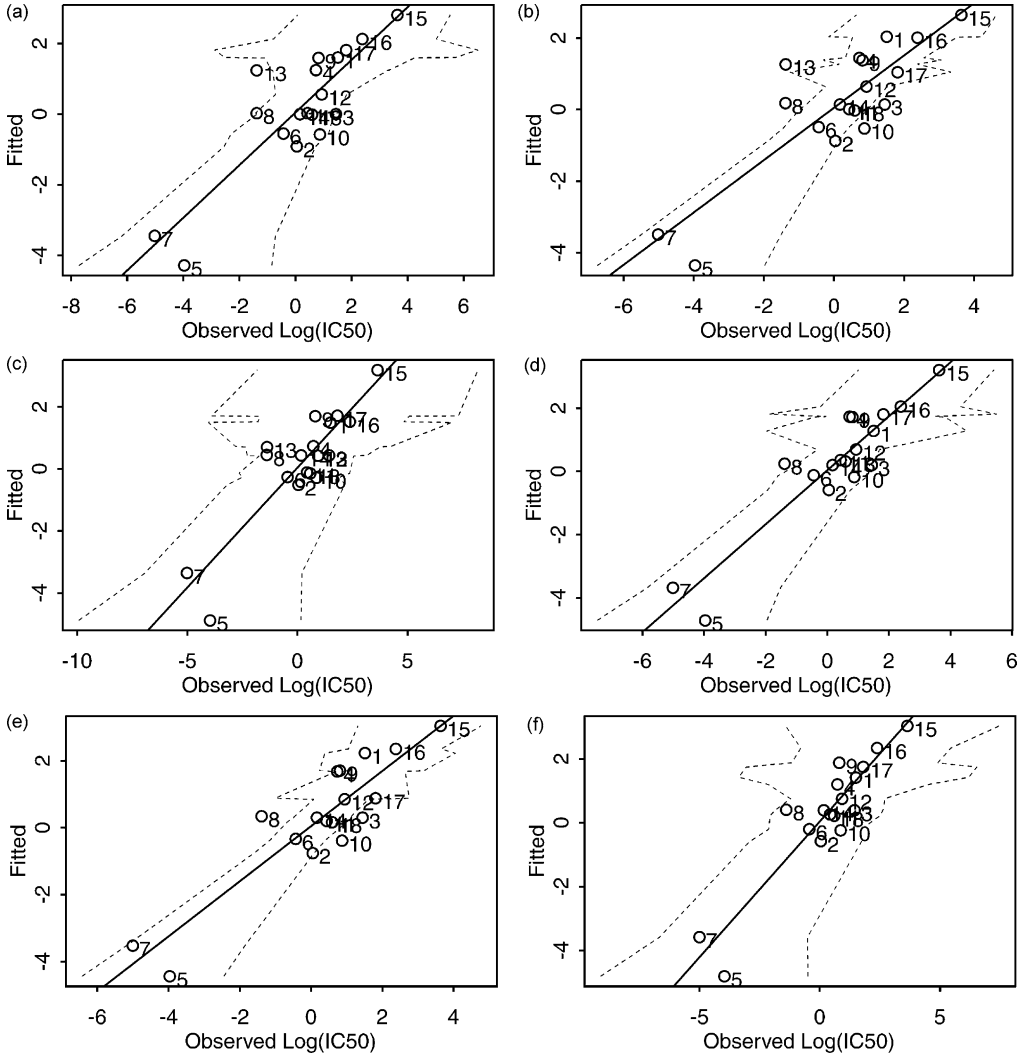



**Figure 5.** Observed versus fitted values of logarithm of $IC_{50}$. Fitted values were obtained from: (a) multiple linear regression with $V_6$, $V_4$, $V_2$ and $V_{12}$ as predictors; (b) multiple linear regression with $PC_1$ and $PC_2$ as predictors; and (c) additive regression models using $PC_1$ and $PC_2$ transformed by cubic B-spline as predictors. d, e and f are the same as a, b and c, respectively, except that compound **13** was not in the data set. The solid lines are visually equivalent to the regression lines, which evaluate how well the models capture the trend of the data. Dotted lines show simultaneous 95% confidence intervals of the fitted values.

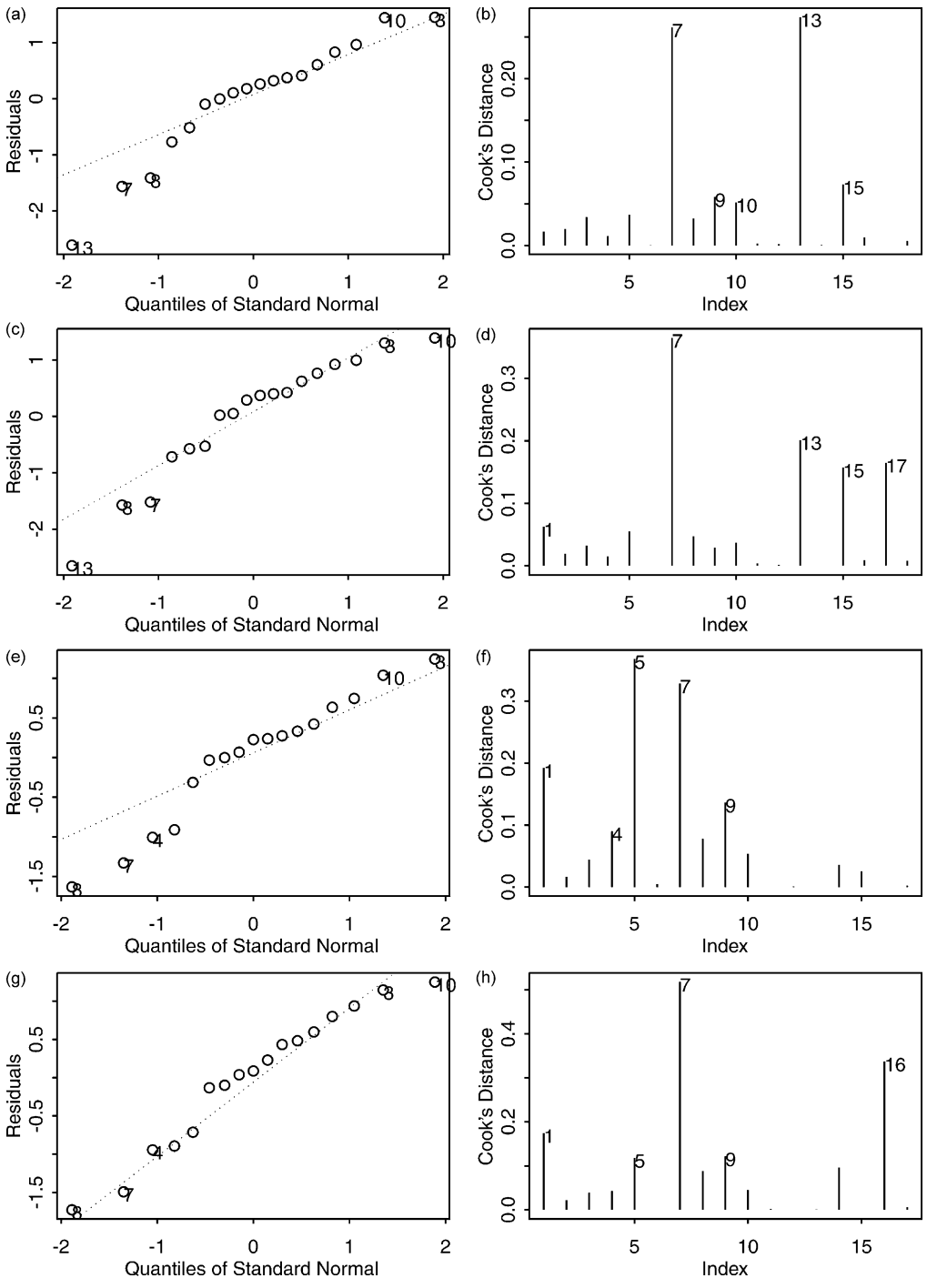



**Figure 6.** Diagnostic plots for multiple linear regressions. (a) Normal quantile plot of residuals obtained from the regression on logarithm of $IC_{50}$ with $V_6$, $V_4$, $V_2$ and $V_{12}$ as predictors; (b) Cook's distance plot obtained from the same regression as above; (c) Normal quantile plot of residuals obtained from the regression on logarithm of $IC_{50}$ with $PC_1$ and $PC_2$ as predictors; (b) Cook's distance plot obtained from the same regression as above. e, f, g and h are the same as a, b, c and d, respectively, except that compound **13** was not in the data set.

$$\log_e(IC_{50}) = 0.27 - 1.50 \times V_6 + 0.71 \times V_4 + 0.35 \times V_2 - 0.39 \times V_{12}$$

(Model4)

which explains about 85.4% of the variation in $\log_e(IC_{50})$ (Table 3). $V_6$ and $V_4$ are significant predictors. One unit increase in π value of the substituent at the para position of the aldehyde group led to about 4.5 [95% confidence interval is (7.9, 2.6)] -fold increase in inhibition activity. One unit increase in MR value of the substituent at the meta position of the aldehyde group led to about 2.0 [95% confidence interval is (3.6, 1.2)] fold decrease in inhibition activity, as far as $IC_{50}$ is compared. By using the first two PCs as predictors, which was obtained by PCA on $V_6$, $V_4$, $V_2$ and $V_{12}$, we have the following regression line,

$$\log_e(IC_{50}) = 0.27 - 1.03 \times PC_1 - 1.34 \times PC_2$$

(Model5)

which explains about 82.3% of the variation in $\log_e(IC_{50})$ (Table 3). One unit increase in $PC_1$value led to about 2.8 [95% confidence interval is (4.4, 1.8)] -fold increase in inhibition activity. One unit increase in $PC_2$ value led to about 3.8 [95% confidence interval is (6.1, 2.4)] -fold increase in inhibition activity, as far as $IC_{50}$ is compared. Results of the PCA on $V_6$, $V_4$, $V_2$ and $V_{12}$ were summarized in Figure 7, which shows that about 65% of the total variance is accounted by $PC_1$ and $PC_2$ (Fig. 7E); $V_4$ has the biggest loading on $PC_1$ (Fig. 7F);

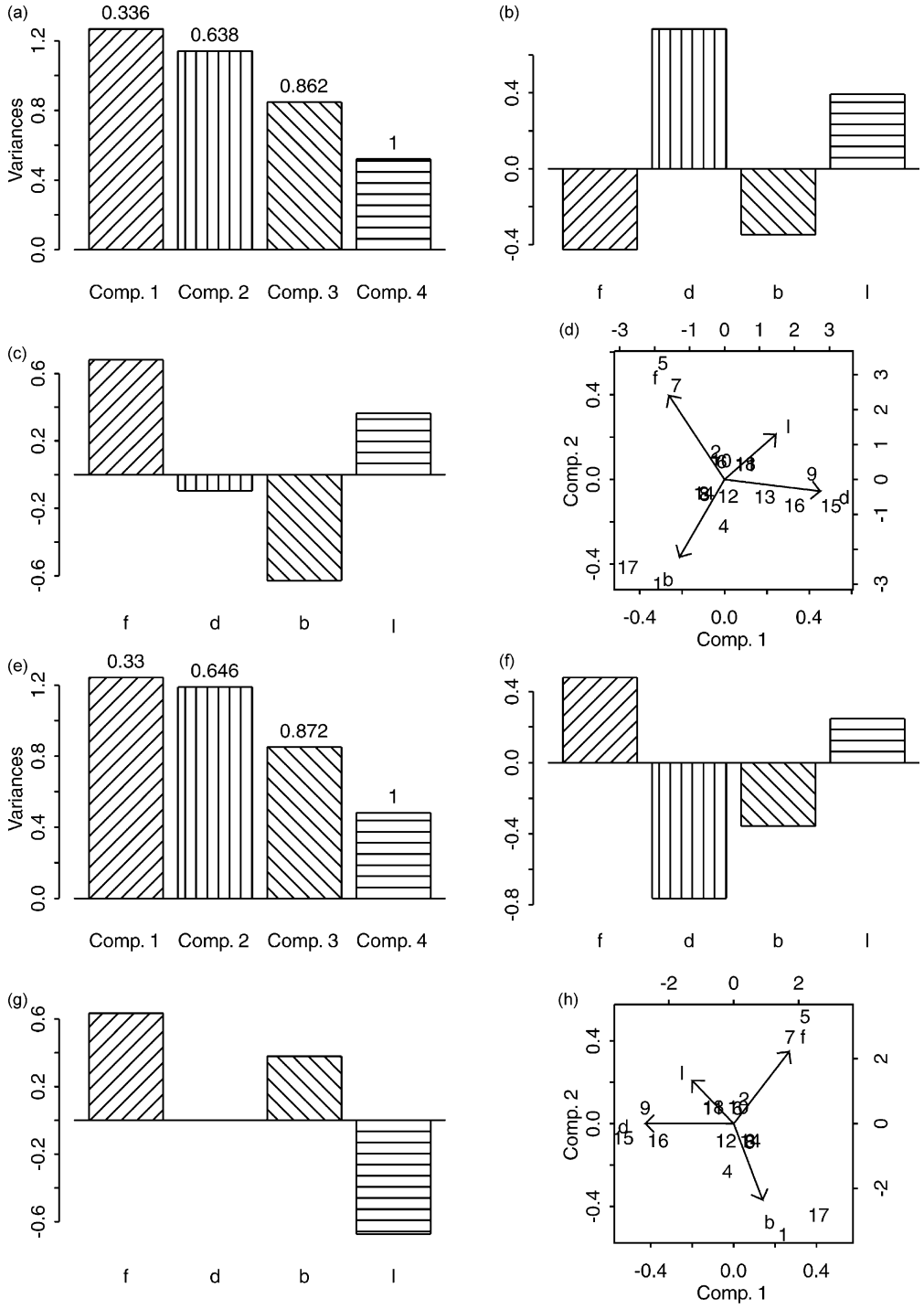



**Figure 7.** Results from PCA on $V_6$, $V_4$, $V_2$ and $V_{12}$. (A) Scree plot shows the total variance associated with each PC; (B) loadings of the original variables on $PC_1$; (C) Loadings of the original variables on $PC_2$; (D) Biplot represents both the observations (indexed by numbers from 1 to 18) and the variables (indexed by b, d, f and l, representing $V_2$, $V_4$, $V_6$ and $V_{12}$, respectively) in a two dimensional space defined by the first two PCs, with the length of the arrows proportional to the loadings of the original variables. E, F, G and H are the same as A, B, C and D, respectively, except that compound **13** was not in the data set.

$V_6$ and $V_{12}$ have the biggest loading on $PC_2$ (Fig. 7G). Figure 7H shows both the observations and the original variables in a two dimensional space defined by $PC_1$ and $PC_2$. It appears that **Model 4** and **5** capture the trend of the data better than **Model 1** and **Model 2** (Fig. 5d and e). The errors of **Model 4** and **5** are approximately normally distributed (Fig. 6e and 6g). Compound **5** and **7** are the most heavily influential observations on the regression coefficients of **Model 4** (Fig. 6f). Compound **7** and **16** are the most heavily influential observations on the regression coefficients of **Model 5** (Fig. 6 h). To capture the possible nonlinearity contained in the data, we also tried additive regression models. Figure 8c and d shows the transformation of $PC_1$ and $PC_2$ by cubic B-spline smoothing. Using the transformed $PC_1$ and $PC_2$ as predictors, we obtained **Model 6**, which was shown no better than **Model 4** and **5** (Fig. 5f). There is no evidence that either **Model 4** or **6** are better than the simpler **Model 5** (ANOVA test in Table 4).

## 3. Discussion

We have previously characterized several benzaldehyde derivatives as mushroom tyrosinase inhibitors from various plants.[9–11] These compounds were found to inhibit the oxidation of L-DOPA by mushroom tyrosinase, presumably by forming a Schiff base between the aldehyde group of the inhibitors and a primary amino group of the enzyme (Fig. 9). The aldehyde group is generally known to react with biologically important nucleophilic groups such as sulfhydryl, amino, or hydroxyl groups. Formation of a Schiff base with a primary amino group of the enzyme is more likely, because the aromatic nucleus is known to stabilize it by conjugation.[20] Presumably, the primary amino group is not at the active site of the enzyme. However, the formation of the Schiff base brings about a conformational change in the active site of the enzyme, which prevents the enzyme from converting the bound substrate to product and shows up mixed type inhibition. Therefore, any substituent group of an inhibitor that can stabilize the formed Schiff base will increase its inhibition activity. Based on our results, we found the following physicochemical properties were crucial for a potent inhibitor.

First, hydrophobicity of the substituent at the para position of the aldehyde group plays key role in inhibition activity. In most cases, important functional sites of an enzyme are buried in the hydrophobic internal part of the enzyme. Therefore, hydrophobic compounds are easier to fit to the hydrophobic protein pockets than the hydrophilic ones. Results from the stepwise multiple regression analysis show that $V_6$ is the most important predictor among all physicochemical descriptors. $V_6$ is the Hansch-Fujita $\pi$ constant of the substituent at the para position of the aldehyde group. The Hansch-Fujita $\pi$ constant describes the contribution of a substituent to the lipophilicity of a compound.[17] A positive $\pi$ value of a substituent indicates that it is hydrophobic, whereas a negative value indicates that it is hydrophilic. **Model 4** tells us that, by fixing the other 3 predictors, one unit increase in $\pi$ value of the substituent at the para position of the aldehyde group results in about 4.5 (95% confidence interval is (7.9, 2.6)) fold increase in inhibition activity as far as $IC_{50}$ is compared. Figure 9 shows a hypothetic picture of the formation of the Schiff base between a primary amino group of the enzyme and

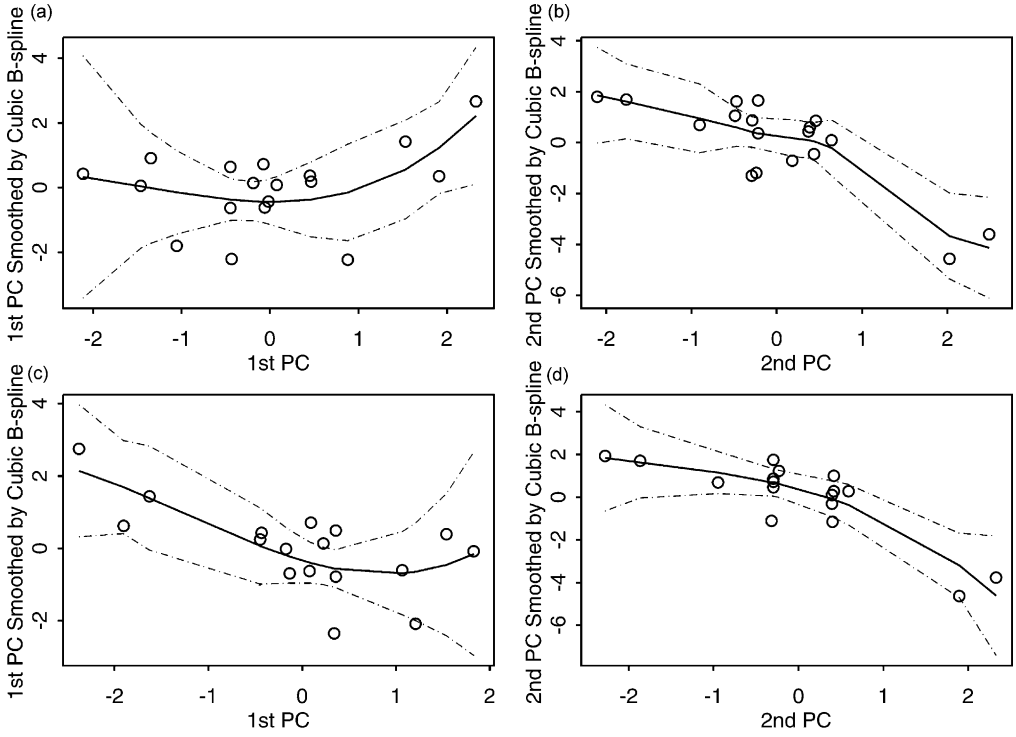



**Figure 8.** Transformation of the first two PCs by cubic B-spline smoothing. PCs in a and b were obtained from data containing all 18 compounds. PCs in c and d were obtained from data without compound **13**. a and c show $PC_1$ smoothed by cubic B-Spline; b and d show $PC_2$ smoothed by cubic B-Spline. Points are deviance residuals. Dotted lines are upper and lower pointwise twice-standard-error curves. Solid lines are fitted lines.

the aldehyde group of the inhibitor with its para group inserted in a hydrophobic protein pocket. Compound **5** and **7** are the most active inhibitors with $IC_{50}$ of 0.019 and 0.0067, mM respectively. Compound **5** has a *tert*-butyl group at the para position of the aldehyde group, whereas compound **7** has an iso-propyl group at the para position of the aldehyde group. *Tert*-butyl and isopropyl are two hydrophobic electron donor groups, which are able to not only stabilize the Schiff base through their inductive effect but also fit the formed Schiff base to the hydrophobic protein pocket (Fig. 9). The π value of the tert-butyl group in compound **5** and the iso-propyl in group 7 are 1.98 and 1.53 respectively, which are the most hydrophobic groups among all substituents in our data. These two compounds also have the largest logarithm of partition coefficients (n-Octanol/Water) (logP in Table 2).[21] However, prediction based on logP is less reliable than on the π value of the substituent at the para position of the aldehyde group, because logP measures the hydrophobicity of the whole compound and represents contributions from all substituents. For example, vanillin is the least active compound of all, whereas its hydrophobicity is not so bad as far as logP is compared with the other compounds. However, as far as its substituent at the para position of the aldehyde group is concerned, hydroxyl group has a π value of −0.67, which is the most hydrophilic group of all. It is worth to note that the effect of hydrophobicity on inhibition activity is not strictly linear and without limitation. For example, compound **5** has larger π value than **7** does; however, its $IC_{50}$ is 2.8 -fold larger than that of compound **5** (Table 1). It may be because the smaller isopropyl group of compound **7** fits the hydrophobic pocket better than the bigger *tert*-butyl group of compound **5**, which implies steric effects also play a role.

**Table 4.** ANOVA comparing regression models. Response variable is $\log_e(IC_{50})$. Model 1, 2, 4 and 5 are the same as those in Table 3. Model 3 uses $PC_1$ and $PC_2$ as predictors for additive regression model. Model 6 is the same as Model 3 except removing compound 13 from the data set

| | RSS | DF | F-value | Pr(>F) |
|---|---|---|---|---|
| Model 2 **versus** 1 | 0.98 | 2 | 0.32 | 0.73 |
| **Model 1** versus **3** | 2.05 | 2 | 0.67 | 0.53 |
| **Model 2** versus **3** | 3.04 | 4 | 0.50 | 0.74 |
| **Model 5** versus **4** | 2.21 | 2 | 1.27 | 0.32 |
| **Model 4** versus **6** | 0.32 | 2 | 0.15 | 0.86 |
| **Model 5** versus **6** | 1.89 | 4 | 0.44 | 0.78 |

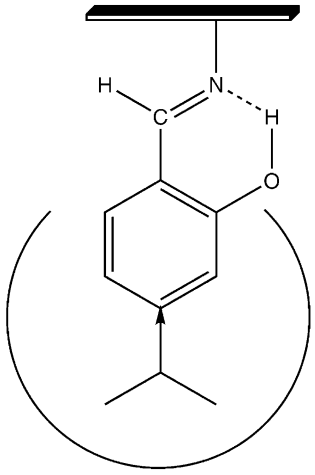



**Figure 9.** Benzaldehyde derivatives inhibit the oxidation of L-DOPA catalyzed by *S. neobelliaria* phenoloxidase, presumably by forming a Schiff base with a primary amino group of the enzyme. Alkane groups at the para position in benzaldehyde derivatives lead to higher inhibitory activity not only by fitting the formed Schiff base to the hydrophobic protein pocket (∪) but also by stabilizing the Schiff base through the inductive effect (→). The Schiff base was further stabilized by forming a quasi-six-membered ring between the hydroxyl group at the ortho position of the benzaldehyde and the unshared pair of electrons on the nitrogen atom of the amino group of the enzyme through intramolecular hydrogen bonding.

Second, compared to hydrophobicity, electron-donating effect of a substituent at the para position of the aldehyde group is less important on the inhibition activity. Hammett's substituent constant (σ) measures the electronic properties of a substituent. A positive σ value of a substituent indicates that it is electron withdrawing, whereas a negative value indicates that it is electron donating. Electronic effects are heavily influenced by the position of the substituent. For example, a methoxy group at the para position has a negative σ value of −0.27, whereas at the meta position it has a positive σ value of 0.12 (Table 2). Because of the proximity (steric) effects, electronic effect of ortho substituents is hard to quantify. Compound **5**, **7**, **2** and **6** are structurally similar to each other considering that para group is the only substituent. The σ values of compound **5** and **7** are −0.20 and −0.15, respectively, which are not the most electron donating groups among all substituents in our data. Actually, as far as σ is compared, both the dimethylamino group of compound **2** (σ = −0.83) and the methoxy group of compound **6** (σ = −0.27) are more electron donating than that of compound **5** and **7**. However, $IC_{50}$ of compound **2** is 1.0492 mM and $IC_{50}$ of compound **6** is 0.6449 mM; both of them are less effective than compound **5** and **7**. This suggests that hydrophobicity of a para group plays a more important role than electronic property on inhibition activity, because the dimethylamino group (π = 0.18) and the methoxy group (π = −0.02) are less hydrophobic than the tert-butyl and isopropyl group as far as the π value is compared.

Third, a hydroxyl group at the ortho position of the aldehyde group contributes to higher inhibition activity. The only difference between compound **6** and **8** is that compound **8** has an extra hydroxyl group at the ortho position, which makes it 2.6-fold more potent than compound **6**. The hydroxyl group at the ortho position is supposed to form a quasi-six-membered ring with the unshared pair of electrons on the nitrogen atom through intramolecular hydrogen bonding and produces a stable chelate structure, as illustrated in Figure 9. Ley and Bertram[22] reported a group of benzaldoximes as inhibitors on the oxidation of L-DOPA catalyzed by mushroom tyrosinase, which showed that the para rather than ortho position is best occupied by a

hydroxyl group. Crystal structure of catechol oxidase of sweet potato complexed with the inhibitor phenylthiourea (PTU) is recently published,[23] which shows that PTU binds to the binuclear copper center as a copper chelator. Based on the structural similarity between benzaldoximes and PTU, we propose that, rather than forming a Schiff base with a primary amino group of the enzyme, benzaldoximes would most likely bind to the binuclear copper center as a copper chelator. However, enzyme kinetics from vanillinoxime suggests that it acts as a noncompetitive inhibitor.[22] The detailed mechanism of the inhibitory effect of benzaldoximes on tyrosinase need further studies. Inhibitory effects of compound **13** and **15** on mushroom tyrosinase were also reported by Ley and Bertram,[22] which are about the same level as results obtained from our study as far as $IC_{50}$ is compared.

How could we explain the high inhibition activity of compound **13**? As we reasoned previously, the hydrophilic hydroxyl group at the para position of the aldehyde group of vanillin makes it the least active compound of all. Compound **13** also has a hydroxyl group at the para position. The normal quantile plot (Fig. 6a) in multiple regression analysis also identified compound **13** as the most possible outlier. How could this abnormal phenomenon be accounted? Compound **13** has two hydroxyl groups: one at the para position, another at the meta position. Therefore, it is structurally very similar to the diphenol substrate of PO, which gives it the ability to bind directly to the coupled binuclear copper active center as a substrate analogue.[2,24] It cannot serve as a substrate, because the aldehyde group deactivates its aromatic ring to prevent electrophilic attach by oxygen.[25] Vanillin has a hydroxyl group at the para position either, but it also has a methoxy group at the meta position. On one hand, it prevents the binding of the para hydroxyl group to the binuclear copper active center by steric hindrance. On the other hand, it destabilizes the Schiff base by its electron withdrawing effect at the meta position. Given further the fact that the hydroxyl group at the para position is hydrophilic, these properties make vanillin the least active inhibitor in the set.

Because PO plays several major roles in insect development and immunity, our results on PO inhibitors might be used for finding ways of weakening the protection afforded by the cuticle and defense reactions of insect. Most of the compounds in our study are listed as food flavor ingredients in Fenaroli's Handbook of Flavor Ingredients,[26] which could legitimate us for designing more effective and environmentally safe pest management agents.

## 4. Materials and methods

### 4.1. Animals

Pupae of *S. neobelliaria* (blowfly) were obtained from Carolina Biological Supplies Co., NC and maintained at 4 degree at cold room.

### 4.2. Chemicals

L-DOPA was purchased from Aldrich Chemical Co. (Milwaukee, WI). Cetylpyridinium chloride (CPC), dimethyl sulfoxide (DMSO), phenylmethanesulfonyl fluoride (PMSF), benzaldehyde and its derivatives (see Table 1 for chemical names) were obtained from Sigma Chemical Co. (St. Louis, MO).

### 4.3. Enzyme purification

All operations were carried out at cold room at 0–5 °C unless stated otherwise. 40 g of pupa were homogenized in 250 mL cold 0.01 M phosphate buffer, pH 7.0 containing 1 mM PMSF (buffer A). Homogeneous mixture was centrifuged at 13,000 *g* for 2 h. Supernatant was subjected to 35% ammonium sulfate saturation for 3 h. Proteins precipitated were discarded after centrifugation at 13,000 *g* for 30 min. The supernatant was brought to 70% ammonium sulfate saturation for 12 h and the proteins precipitated were collected by centrifugation at 10,000 *g* for 20 min. The pellet was dissolved in 25 mL buffer A and was chromatographed on a DEAE-Cellulose column (2.4×20 cm) pre-equilibrated with the same buffer. The column was washed extensively with this buffer and bound proteins were eluted by a linear gradient of water and 1 M sodium chloride. Fractions that contained ProPO were pooled, dialyzed and used as pure enzyme.

### 4.4. Enzyme assay

Enzyme used for the bioassay was purified as described above. PO functions both as a monophenolase and as an *o*-diphenolase. In the current experiment, L-DOPA was used as a substrate; therefore, inhibition of *o*-diphenolase is studied. Another substrate of this enzymatic reaction is oxygen, but the assay was carried out in air-saturated solutions. To obtain the active PO while eliminate the effects of residual CPC left in the enzyme solution, the following assay system is established. First, 0.3 mL of the purified enzyme was mixed with 0.1 mL of 0.2 M phosphate buffer pH 7.0 (buffer B) and 0.2 mL of 0.05% CPC in this order. Shake the mixture immediately for 15 s, incubate for another 60 seconds, and then take 0.5 mL, and add to 24.5 mL buffer B to produce mixture A, which was used as the activated enzyme solution. The final concentration of CPC in mixture A is 3.3 μg/mL, which is negligible in our preliminary test. In the 2 mL enzyme assay system, 0.5 mL inhibitor solution with different concentrations of inhibitor, and 0.5 mL substrate solution with different concentrations of L-DOPA was mixed with 1 mL mixture A, which was immediately monitored for the formation of dopachrome by measuring the linear increase in optical density at 475 nm up to the appropriate time (not exceeding 5 min, unless otherwise specified). All the samples were first dissolved in DMSO and used for the experiment at 30 times dilution. Each enzyme assay was repeated for three replications.

### 4.5. Data analysis

**4.5.1. Enzyme kinetics analysis.** The enzyme kinetics data were fit to various inhibition models by non-linear

regressions. The parameters were estimated by Marquardt-Levenberg algorithm.[14] Akaike's Information Criterion (AIC)[27] was used to determine the best model among all candidate models. For a regression model with $n$ observations, $p$ parameters and normally-distributed errors, AIC is defined by

$$\mathrm{AIC} = n\log(\mathrm{RSS}/n) + 2p + \mathrm{const} \tag{1}$$

where RSS is the residual sum of squares of the model. AIC is actually a penalized maximum likelihood value. The number of parameters included in the model is used to penalize the maximum likelihood value that is designed to balance the model complexity and fitting power.

$IC_{50}$ values were obtained by fitting regression line: $1/v = \alpha + \beta[I]$, where [I] is the inhibitor concentration, $v$ is the corresponding enzyme activity, α and β are intercept and slope of the regression line. It can be shown that $IC_{50} = \alpha/\beta$. Standard error of the estimate of $IC_{50}$ was then estimated from the standard error of the intercept and the slope of the regression line.

**4.5.2. Physicochemical properties of the inhibitors.** Let $i$, ranging from 2 to 6, be the position of the substituent on the benzene ring of compound $k$. Let $j$ be one of the three types of physicochemical properties: Hammett's substituent constant (σ), Hansch-Fujita constant (π), and molar refractivity (MR). Then $x_{ijk}$ is the $j^{\mathrm{th}}$ type of physicochemical property of a substituent at position $i$ of compound $k$. Table 2 lists these values for each compound, which forms the data matrix used in cluster, principle component (PCA), and regression analysis. For example, $V_1$ measures the Hansch-Fujita π value of the substituent at position 2 on the benzene ring of a benzaldehyde compound in the set.

When variables are measured on different scales, as for different physicochemical properties, the choice of measurement units strongly affects subsequent data analysis. One way of circumventing this problem is to express all variables in standardized form; that is, transform all variables to have a mean of 0 and a standard deviation of 1. The standardization is achieved by

$$z_{ijk} = \frac{x_{ijk} - \bar{x}_{ij}}{s_{ij}} \tag{2}$$

where

$$\bar{x}_{ij} = \frac{1}{18}\sum_{k=1}^{18} x_{ijk} \tag{3}$$

and,

$$s_{ij} = \frac{1}{18}\sum_{k=1}^{18} \left| x_{ijk} - \bar{x}_{ij} \right| \tag{4}$$

which is the mean absolute deviation.

The logarithm of the partition coefficient for $n$-octanol/water of the inhibitors are computed by Crippen's fragmentation.[28]

**4.5.3. Cluster analysis of chemical structures.** Dissimilarity between compound $m$ and $n$ is a measure of how far apart they are, based on their physicochemical properties. Because of the continuous property of these physicochemical measurements, we used Euclidean distance to measure the dissimilarity between $m$ and $n$, which is defined by

$$d(m,n) = \sqrt{\sum_{i=2}^{6}\sum_{j=1}^{3}\left(z_{ijm} - z_{ijn}\right)^2} \tag{5}$$

According to group average method,[29] the dissimilarity between cluster $R$ and $Q$ is defined by

$$d(R,Q) = \frac{1}{|R|\,|Q|}\sum_{m\in R, n\in Q} d(m,n) \tag{6}$$

Agglomerative hierarchical clustering method was used to group the compounds,[29] which at the end yields a clustering tree in which the leaves represent compounds. The vertical coordinate of the place where two branches join equals the dissimilarity between the corresponding clusters.

**4.5.4. Principle component analysis.** Let $\mathbf{X}$ be the data matrix as listed in Table 2. Principle component analysis (PCA) seeks linear combinations (principle components or PCs) of the columns of $\mathbf{X}$ with maximal variance. Let $\mathbf{x}$ be a row vector of $\mathbf{X}$ with sample covariance matrix $\boldsymbol{\Sigma}$ . Let $\mathbf{xa}$ be a standardized linear combination of $\mathbf{x}$ with $\|\mathbf{a}\|^2 = \mathbf{a}^{\mathbf{T}}\mathbf{a} = \mathbf{1}$. The variance is maximized by taking $\mathbf{a}$ to be the column eigenvector corresponding to the largest eigenvalues of $\boldsymbol{\Sigma}$. Taking subsequent eigenvectors gives combinations with as large as possible variance that is uncorrelated with those that have been taken earlier. The $i$th principle component is then the $i$th linear combination picked by this procedure.

**4.5.5. Stepwise multiple regression.** Linear regression models the relationship between a response variable $Y$ and a set of predictor variables $X_1,\ldots,X_p$ by

$$Y = \alpha + \sum_{j=1}^{p}\beta_j X_j + \varepsilon \tag{7}$$

where α is the intercept of the model; β are coefficients of the predictor variables; and ε are random errors that can not be accounted by the predictors. The response variable in our analysis is the natural logarithm of $IC_{50}$ of 18 compounds. Each column ($V_1$ to $V_{13}$) in Table 2 represents a predictor variable measuring one specific physicochemical property of the compound. We used Efroymson's stepwise procedures to find a parsimonious subset of predictors to include in a least squares multiple regression.[30] The coefficient of each predictor was estimated. The significance of the coefficients and the overall fit of the model to the data were then tested statistically. The error ε is assumed to have a Gaussian

(normal) distribution with constant variance and to be independent of the predictors.

**4.5.6. Additive regression models.** Additive regression models are used to capture the non-linear patterns contained in the data. Compared to eq 7, additive models replace the linear function $\beta_j X_j$ by a non-linear function to get

$$Y = \alpha + \sum_{j=1}^{p} f_j(X_j) + \varepsilon \quad (8)$$

Function $f_j(.)$ is used to account for non-linearity of the data.[31] To find a smooth term $f_j(.)$, we use cubic B-spline functions. All data analysis is implemented with the standard statistical package Splus.

## References and notes


1. Lerch, K.; Germann, U. A. *Prog Clin Biol Res*. **1988**, *274*, 331.
2. Lerch, K. *Prog Clin Biol Res*. **1988**, *256*, 85.
3. Ashida, M. *Res Immunol*. **1990**, *141* (9), 908.
4. Hall, M.; Scott, T.; Sugumaran, M.; Soderhall, K.; Law, J. H. *Proc. Natl. Acad. Sci. U.S.A*. **1995**, *92* (17), 7764.
5. Soderhall, K.; Cerenius, L.; Johansson, M. W. *Ann. N. Y. Acad Sci*. **1994**, *712*, 155.
6. Chase, M. R.; Raina, K.; Bruno, J.; Sugumaran, M. *Insect Biochem. Mol. Biol*. **2000**, *30* (10), 953.
7. Cui, L.; Luckhart, S.; Rosenberg, R. *Insect Mol. Biol*. **2000**, *9* (2), 127.
8. Kwon, T. H.; Lee, S. Y.; Lee, J. H.; Choi, J. S.; Kawabata, S.; Iwanaga, S.; Lee, B. L. *Mol. Cells* **1997**, *7* (1), 90.
9. Kubo, I.; Kinst-Hori, I. *Planta Med*. **1999**, *65* (1), 19.
10. Lim, J. Y.; Ishiguro, K.; Kubo, I. *Phytother. Res*. **1999**, *13* (5), 371.
11. Kubo, I.; Kinst-Hori, I. *J. Agric. Food Chem*. **1999**, *47* (11), 4574.
12. Hammond, D. G.; Rangel, S.; Kubo, I. *J. Agric. Food Chem*. **2000**, *48*, 4410.
13. Rangel, S.; Wang, S.; Fujita, K; Kubo, I. In PostHarvest pesticide activity of naturally occuring aldehydes, 219th ACS National Meeting, San Francisco, CA, March 26–30, 2000.
14. Cornish-Bowden, A. *Analysis of Enzyme Kinetic Data*; Oxford University Press, 1995.
15. Hammett, L. P. *Physical Organic Chemistry: Reaction Rates, Equilibria, and Mechanisms*; McGraw-Hill Book Co: New York, 1940.
16. Fujita, T.; Iwasa, J.; Hansch, C. *J. Am. Chem. Soc*. **1964**, *86*, 5175.
17. Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society, 1995.
18. Wolff, M. E. *Principles and Practice, Fifth Edition,* Vol. 1; Wiley, 1995.
19. Miller, A. J. *Subset Selection in Regression,* Vol. 40; Chapman and Hall: London, 1990; p. 229.
20. Schauenstein, E. E. H.; Zollner, H. *Aldehydes in Biological Systems*; Pion: London, 1977; p. 205.
21. Leo, A.; Jow, P. Y.; Silipo, C.; Hansch, C. *J. Med. Chem*. **1975**, *18* (9), 865.
22. Ley, J. P.; Bertram, H. J. *Bioorg. Med. Chem*. **2001**, *9* (7), 1879.
23. Klabunde, T.; Eicken, C.; Sacchettini, J. C.; Krebs, B. *Nat. Struct. Biol*. **1998**, *5* (12), 1084.
24. Solomon, E. I.; Chen, P.; Metz, M.; Lee, S. K.; Palmer, A. E. *Angew. Chem. Int. Ed. Engl*. **2001**, *40* (24), 4570.
25. Conrad, J. S.; Dawso, S. R.; Hubbard, E. R.; Meyers, T. E.; Strothkamp, K. G. *Biochemistry* **1994**, *33* (19), 5739.
26. Burdock, G. A. *Fenaroli's Handbook of Flavor Ingredients, Third Edition*; CRC Press: Boca Raton, FL, 1995; p. 133.
27. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. *Akaike Information Criterion Statistics,* Vol. 1; D. Reidel Publishing Co: Dordrecht, 1986; p. 290.
28. Ghose, A. K.; Crippen, G. M. *J. Chem. Inf. Comput. Sci*. **1987**, *27* (1), 21.
29. Kaufman, L.; Rousseeuw, P. J. *Finding Groups in Data*; John Wiley & Sons Inc: New York, 1990; p. 342.
30. Efroymson, M. A. *Mathematical Methods for Digital Computers* **1960**, 191.
31. Hastie, T. J.; Tibshirani, R. J. *Generalized Additive Models,* Vol. 43; Chapman and Hall Ltd: London, 1990; p. 335.