Modeling the Toxicity of Aromatic Compounds to *Tetrahymena pyriformis*: The Response Surface Methodology with Nonlinear Methods

Shijin Ren*

Center for Environmental Biotechnology, 676 Dabney Hall, University of Tennessee, Knoxville, Tennessee 37996-1605, and Department of Statistics, 331 Stokely Management Center, University of Tennessee, Knoxville, Tennessee 37996-0532

Received March 7, 2003

Response surface models based on multiple linear regression had previously been developed for the toxicity of aromatic chemicals to *Tetrahymena pyriformis*. However, a nonlinear relationship between toxicity and one of the molecular descriptors in the response surface model was observed. In this study, response surface models were established using six nonlinear modeling methods to handle the nonlinearity exhibited in the aromatic chemicals data set. All models were validated using the method of cross-validation, and prediction accuracy was tested on an external data set. Results showed that response surface models based on locally weighted regression scatter plot smoothing (LOESS), multivariate adaptive regression splines (MARS), neural networks (NN), and projection pursuit regression (PPR) provided satisfactory power of model fitting and prediction and had similar applicabilities. The response surface models based on nonlinear methods were difficult to interpret and conservative in discriminating toxicity mechanisms.

INTRODUCTION

In ecotoxicology, quantitative structure−activity relationships (QSARs) have been developed and used for assessing the hazards of chemical pollutants in the environment. QSARs aim at relating toxicity or other biological activities of chemical compounds to their molecular structures and/or properties. Using a learning database containing compounds with measured toxicity potencies, QSARs can be developed by some statistical means and can be used to predict the toxicity of chemical compounds not in the learning database. To serve this purpose, several requirements must be met in developing QSARs, e.g., high learning data quality and effective model validation. In a recent study, Schultz and Cronin[1] described thoroughly the essential and desired characteristics of QSARs.

It is well-known that chemical compounds exhibit toxicity via different mechanisms of toxic action. For example, some compounds exhibit general toxicity and others exhibit specific toxicity. General, or narcotic, toxicity is caused by nonspecific disruption of the proper functioning of cell membrane, which is generally thought to be the site of toxic action. Specific, or reactive, toxicity is caused by interaction of toxicants with defined receptors at the site of toxic action.[2] The difference in mechanism of toxic action led to different QSAR models for narcotic and reactive compounds. As an example, the toxicity of narcotic compounds can be modeled solely by hydrophobicity, and the toxicity of electrophiles, some of the most reactive compounds, can be modeled by reaction alone.[3]

Compounds acting strictly by the narcotic or reactive mode represent the extreme cases. Between chemicals acting nonreactively and those acting reactively there is a continuum, as suggested by Ferguson[4] and McFarland.[5] Later Mekenyan and Veith[6] described a response surface which expressed toxicity as a function of hydrophobicity (quantified by the logarithm of the octanol/water partition coefficient logK_{ow}) and electrophilicity (quantified by the average acceptor superdelocalizability S_{av}^{N}). The interpretation of the response surface is straightforward considering the physical/chemical meanings of hydrophobicity and electrophilicity descriptors: toxicity is the result of uptake of a chemical compound and interaction between the compound and the site of action. Detailed descriptions and discussions of the response surface methodology can be found in, e.g., ref 7. Generally, the response surface methodology was found to be able to incorporate chemical compounds with different mechanisms of toxic action into a single QSAR.

A number of toxicity modeling studies was conducted based on the response surface methodology. For example, Wang et al.[8] studied the toxicity of selected phenols to *Cucumis sativus* and obtained the following model.

$$\log 1/\mathrm{IGC}_{50} = 0.77 \log K_{ow} - 0.39 E_{\mathrm{lumo}} + 2.36 \qquad (1)$$
$$n = 22,\ r^2 = 0.89,\ p < 0.0001$$

In the above equation, IGC_{50} is the 50% inhibition concentration, E_{lumo} is the lowest unoccupied molecular orbital energy which can be used to quantify electrophilicity, n is the number of observations, r^2 is the coefficient of determination, and p is the probability for the Fisher's test. Response surface models as represented by eq 1 were successful. However, success was usually achieved using structurally or mechanistically similar compounds. It was found that the performance of the response surface method was less satisfactory as structural diversity in a data set

* Corresponding author phone: (865)974-9004; fax: (865)974-7076; e-mail: sjren@utk.edu. Corresponding author address: Department of Chemical Engineering, 419 Dougherty Engineering Building, University of Tennessee, Knoxville, TN 37996-2200.

10.1021/ci034046y CCC: $25.00
Published on Web 07/11/2003

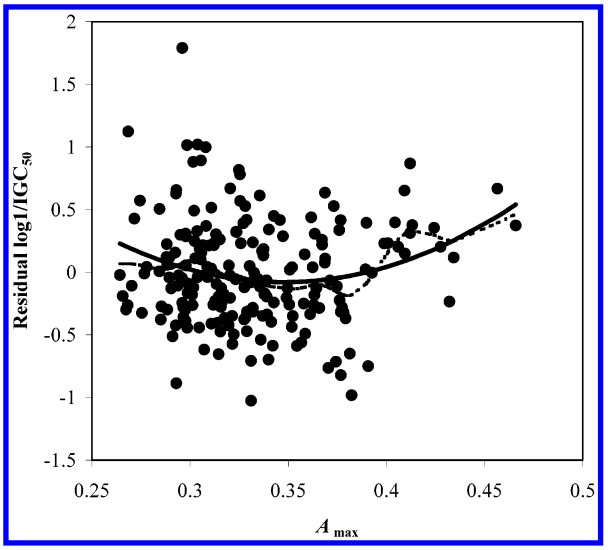

Figure 1. Residual log1/IGC_{50} versus A_{max}. The solid line represents a fitted quadratic model, and the dashed line represents a fitted spline with $\lambda = 0.1$.

increased.[7] As an example, Cronin et al.[9] obtained the following model for the toxicity of aromatic chemicals to *Tetrahymena pyriformis*. A large variety of chemical structures and several mechanisms of toxic action were involved in the data set

$$\log1/IGC_{50} = 0.40 \log K_{ow} - 0.94E_{lumo} - 1.27 \quad (2)$$
$$n = 203,\ r^2 = 0.60,\ s = 0.49,\ F = 151$$

where s is the root-mean-square error and F is the value of the Fisher statistic. A common approach to improve the model fitting in the response surface methodology is to identify statistical outliers and remove them from model fitting. These statistical outliers are usually associated with specific chemical structures and/or mechanisms of toxic action (e.g., refs 10 and 11). In the study of Cronin et al.,[9] several statistical outliers were identified but not removed. Instead, the authors used another electrophilicity descriptor A_{max} (the maximum acceptor superdelocalizability) and obtained an improved fit of the data.

$$\log1/IGC_{50} = 0.37 \log K_{ow} + 13.1 A_{max} - 4.30 \quad (3)$$
$$n = 203,\ r^2 = 0.70,\ s = 0.42,\ F = 237$$

Equation 3 indicates that an improvement of model fitting (an increase of r^2 from 0.6 to 0.7) can be achieved by carefully choosing the model descriptor. Note that response surface models such as eqs 2 and 3 were obtained using multiple linear regression which assumes a linear relationship between toxicity (log1/IGC_{50}) and the descriptors. Model residuals of eq 3 were inspected for potential patterns of nonlinearity. No special pattern was observed in the plot of residuals versus logK_{ow}. However, the plot of residual log1/IGC_{50} versus A_{max}, shown in Figure 1, revealed a concave pattern. Using residual log1/IGC_{50} as the dependent variable and A_{max} as the independent variable, a quadratic model was fitted (not shown). The quadratic term in the model was highly significant with a p value of 0.0023. The intrinsic nonlinearity may be more complicated than a quadratic pattern, as demonstrated by the fitted smoothed curve (obtained with JMP 5.0, SAS Institute, Cary, North Carolina) in Figure 1. This smoothed curve was obtained with a λ value of 0.1 (smoothing methods are explained in most standard statistical textbooks, e.g., ref 12). The existence of a nonlinear relationship between toxicity and A_{max} in the aromatic chemicals data suggested that a statistical assumption for multiple linear regression was violated when eq 3 was developed.

Cronin and Schultz[13] pointed out that biology and the modeling of biology is a nonlinear phenomenon and that always expecting linear relationships in biological modeling is not realistic. Neural networks, as reflected in the recent review articles of Kaiser,[14,15] has been used in aquatic toxicity modeling as a useful tool when the relationship between toxicity and molecular descriptors is intrinsically nonlinear (e.g., ref 16). In addition to neural networks, a number of nonlinear modeling methods have been developed in the field of statistics to handle nonlinearity exhibited in a given data set. In this study, five nonlinear statistical modeling methods as well as neural networks were employed to analyze the aromatic chemicals data set mentioned previously using the response surface methodology. The performances of the response surface models obtained using the nonlinear methods were assessed, the applicabilities of the models were inspected, and model transparency and interpretability were discussed.

METHODS

Nonlinear Methods. Six nonlinear modeling methods were considered in the present study to caputer the intrinsic nonlinearity in the aromatic chemicals data set described later. These nonlinear methods included generalized additive model (GAM), least absolute shrinkage and selection operator version 2 (LASSO2), locally weighted regression scatter plot smoothing (LOESS), multivariate adaptive regression splines (MARS), neural networks (NN), and projection pursuit regression (PPR). Details of the computational algorithms of these nonlinear regression methods were not the focus of this study and therefore were not provided. However, a brief introduction together with a key reference for each of the six methods were given in Table 1.

Data Set and Data Preparation. The data set reported in the study of Cronin et al.[9] was used. Note this was also the data set based on which eqs 2 and 3 were developed. This data set involved 203 aromatic compounds containing a nitro or cyano group. The toxicity of these compounds was evaluated with the *Tetrahymena pyriformis* testing system,[23] and the toxicity data were expressed as the negative logarithm of the 50% growth inhibition concentrations (log1/IGC_{50}). The values of the hydrophobicity descriptor logK_{ow} and electrophilicity descriptors E_{lumo} and A_{max} were presented in the study of Cronin et al.[9] The calculation methods for these descriptors were described by the authors. The original data set was randomly split into two disjoint subsets: one training data set for model development (162 observations, approximately 80% of the original data) and one test data set for model validation and model prediction assessment (41 observations, approximately 20% of the original data). The compounds included in the validation set were presented in Table 2.

Model Development. Log1/IGC_{50} was used as the dependent variable, and logK_{ow} and A_{max} were used as the

Table 1. Brief Introduction to Nonlinear Regression Methods Used in This Study

method	brief description	ref
generalized additive model (GAM)	GAM extends the concept of linear regression by allowing linear functions of the predictors to be replaced by smooth functions of the predictors. The model assumes that the smooth predictor functions have an additive structure.	17
least absolute shrinkage and selection operator version2 (LASSO2)	LASSO2 is similar to the ridge regression method. Instead of taking penalties on the squares of the parameters, it makes constraints on the absolute value of the parameters.	18
locally weighted regression scatter plot smoothing (LOESS)	LOESS fits successive linear regression function in local neighborhoods and thus obtains a smothered curve.	19
multivariate adaptive regression splines (MARS)	MARS is an adaptive procedure for regression. It is a generalization of stepwise linear regression, but the regression is fitted using piecewise basis functions.	20
neural networks (NN)	NN is a nonlinear method generally consisting of input, hidden, and output layers. The nonlinear model fitting is implemented in the hidden layer with nonlinear transformations. Model fitting is repeated if there is more than one hidden layer.	21
projection pursuit regression (PPR)	These models can be simply described as additive functions of linear combinations of the attributes. Projection pursuit regression uses the backfitting algorithm to iteratively obtain the additive terms of the model.	22

Table 2. Predicted Toxicity for Compounds in the External Test Set

compound	observed	GAM	LASSO2	LOESS	MARS	PPR	NN
3-cyanopyridine	−0.74	−0.60	−0.55	−0.64	−0.48	−0.66	−0.68
4-fluorobenzonitrile	−0.26	0.14	0.13	0.19	0.16	0.15	0.16
3-chlorobenzonitrile	−0.06	0.63	0.55	0.63	0.55	0.59	0.57
2-methoxy-2-nitropyridine	−0.01	0.45	0.46	0.37	0.44	0.31	0.40
4-cyanobenzaldehyde	0.04	−0.16	−0.13	−0.16	−0.08	−0.12	−0.15
3-methoxybenzonitrile	0.05	−0.07	−0.03	0.02	0.03	−0.03	0.08
2-amino-5-nitropyridine	0.22	−0.11	−0.16	−0.23	−0.13	−0.04	−0.25
4-cyanoaniline	0.24	−0.42	−0.40	−0.51	−0.32	−0.41	−0.39
4-fluoro-2-nitrotoluene	0.25	0.77	0.72	0.77	0.70	0.59	0.72
4-methyl-2-nitroaniline	0.37	0.56	0.48	0.57	0.49	0.55	0.52
ethyl-4-cyanobenzoate	0.37	0.21	0.22	0.32	0.26	0.23	0.31
5-nitroquinoline	0.39	0.53	0.49	0.52	0.48	0.54	0.45
methyl-4-nitrobenzoate	0.40	0.64	0.70	0.62	0.67	0.80	0.63
2,3-dinitrophenol	0.46	1.19	1.28	1.24	1.16	1.14	1.28
4-amino-3,5-dinitrobenzamide	0.51	1.26	1.24	1.46	1.23	1.22	1.45
2,6-dinitrophenol	0.54	1.10	1.17	1.25	1.08	1.28	1.28
2-methyl-5-nitrophenol	0.66	0.63	0.55	0.63	0.55	0.60	0.58
2,6-dichloro-4-nitrophenol	0.66	1.11	1.27	1.18	1.20	1.38	1.18
2-nitrophenol	0.67	0.48	0.41	0.44	0.41	0.51	0.38
1-cyanonaphthalene	0.69	0.27	0.25	0.43	0.31	0.41	0.54
ethyl-4-nitrobenzoate	0.71	0.63	0.54	0.62	0.54	0.59	0.57
2,6-dinitroaniline	0.84	1.02	1.09	0.96	1.00	1.12	1.05
1,3-dinitrobenzene	0.89	0.68	0.83	0.70	0.76	0.62	0.79
3,5-dinitroaniline	0.94	0.66	0.79	0.65	0.74	0.90	0.72
4-amino-2-nitrophenol	0.98	−0.04	−0.09	−0.11	−0.06	−0.10	−0.15
2-cyanonitrobenzene	1.08	0.30	0.42	0.35	0.39	0.29	0.38
5-fluoro-2-nitrophenol	1.12	0.63	0.68	0.60	0.65	0.76	0.62
3,5-dichloronitrobenzene	1.13	1.12	1.21	1.18	1.15	0.98	1.16
2,6-dibromo-4-nitrophenol	1.36	1.46	1.61	1.51	1.62	1.37	1.51
2-nitro-1-naphthol	1.36	1.05	1.05	1.08	1.01	0.91	1.05
2,4,6-trichloronitrobenzene	1.43	1.75	1.81	1.65	1.84	1.74	1.67
2,3,5,6-tetrachloronitrobenzene	1.47	2.20	2.26	2.10	2.44	2.33	2.06
3-phenylnitrobenzene	1.57	1.00	0.97	1.51	1.16	1.25	1.25
2-methyl-4,6-dinitrophenol	1.73	1.32	1.40	1.32	1.28	1.06	1.35
4-*tert*-butyl-2,6-dinitrophenol	1.80	1.86	1.92	1.70	1.90	1.90	1.74
2,6-diiodo-4-nitrophenol	1.81	1.41	1.56	1.48	1.56	1.33	1.47
2,6-dichloronitropyrimidine	2.03	2.12	1.72	2.01	1.89	1.92	1.98
4-iodo-1,3-dinitrobenzene	2.12	2.29	1.94	1.98	2.05	1.93	1.96
3,5-dichloro-1,2-dinitrobenzene	2.42	2.38	2.06	2.00	2.14	1.95	1.99
2-chloro-3,5-dinitropyridine	2.64	1.30	1.96	3.19	2.53	2.39	3.01
4-chloro-3,5-dinitrobenzonitrile	2.66	1.37	2.30	3.35	2.93	2.68	3.27

independent variables. Response surface models were developed using the six before-mentioned nonlinear methods on the training data set. The statistical software R 1.4.1 (The R Foundation, Vienna, Austria, downloadable at http://www.r-project.org) was used for all statistical analyses. Key R codes for each of the nonlinear methods were provided in the Appendix. For GAM whose smooth terms are represented using penalized regression splines, the default value of 1 was used as the dimension of the basis for the smooth terms (i.e., one-dimensional smooth). Normal distribution was specified

and used as the link function in model fitting. For LASSO2 which fits a regression model with L1 constraint on the parameters (see reference in Table 1 for L1 constraint), the constraint value was set to 10 which was an absolute bound. For LOESS, the parameter that controls the degree of smoothing was 0.75, and quadratic polynomials (degree of polynomials = 2) were used in local regression. By default, normal distribution was implied and model fitting was performed by least squares. A control parameter was specified (surface="direct") so that the fitted surface was to be computed exactly. For MARS, the maximum interaction degree was 1, the cost per degree of freedom charge was 2 (which served as a "penalty" in fitting the splines), forward stepwise process was carried out, and forward stepwise stopping threshold was 0.001. By default, the fitted model was pruned in a backward stepwise fashion. For PPR which involves ridge function, the Friedman's super smoother (see reference in Table 1 for this smoother) was used for smoothing the ridge functions. The super smoother span control value was 0 which allowed automatic span selection by local cross-validation. The smoother bass tone control value was 0 and was used with automatic span selection. For NN, a model with a single hidden layer and three units in the hidden layer was developed. The value of the default error function in R was monitored during training. The maximum number of iterations was determined such that the average predicted mean square error in cross-validation (see later methods) was the smallest. This number turned out to be 50. Consequently, the final neural networks model was trained for 50 iterations.

Using the response surface models developed, the fitted log1/IGC_{50} values of compounds in the training set were compared with the observed values. Compounds with residuals larger than ±1, if any, were judged to be statistical outliers and removed. The model fitting process was then repeated using the remaining data. The correlation between fitted and observed toxicity was inspected with n, r^2, s, F, and p values noted. Note that due to the nonlinear and/or nonparametric nature of the modeling methods used in this study, mathematical expressions of the most of the resulting models are not available.

Model Validation. Each response surface model was validated since, as pointed out by Schultz and Cronin,[1] all ecotoxicity-related QSAR models require validation. The validation procedure recommended by Tropsha et al.[24] was used in this study, i.e., cross-validation (CV) followed by test on an external data set. The latter step is sometimes viewed as the true prediction accuracy assessment.

For CV, the method of 10-fold CV was used. As such, data in the training set, which was used to develop the models, were randomly divided into 10 CV subsets (note that the data set in Cronin et al.[9] was ordered according to log 1/IGC_{50}). Each CV subset was used as a left-out set. The toxicity of compounds in any of the 10 left-out sets was predicted using the response surface models developed on data in the remaining nine CV subsets. At the end of the 10-fold CV procedure, the toxicity of each of the compounds in the entire training set was predicted and compared with the corresponding observed value. The predicted mean square error (PMSE) for each subset was calculated. The average PMSE value calculated over the 10 CV subsets and the associated standard error for each of the response surface models were recorded. Additionally, the CV predicted toxicity was correlated with observed toxicity, and r^2_{CV} (coefficient of determination in 10-fold CV) value was calculated and recorded for each of the nonlinear methods.

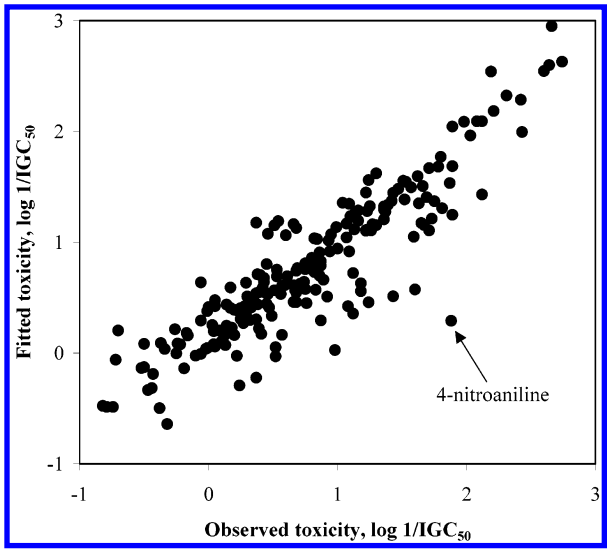

Figure 2. Fitted toxicity versus observed toxicity in the PPR response surface model before removing 4-nitroaniline.

Using the models developed on the training set, the toxicity of compounds in the external test set (Table 2) was predicted. The predicted toxicity was then compared with the corresponding observed toxicity. Correlations between predicted and observed toxicity was obtained for each of the nonlinear methods, and r^2_{pred} (coefficient of determination of model prediction), s, F, and p values were obtained and recorded for each of the models.

RESULTS

Model Development. The r^2 values in the initial model fitting process using all 162 data points in the training set ranged from 0.67 for LASSO2 to 0.77 for PPR response surface. A plot for the fitted toxicity versus observed toxicity for the PPR response surface model was shown in Figure 2. Plots for other nonlinear methods were similar but had (slightly) more scattering of the data. Examining the residuals revealed that 4-nitroaniline was a statistical outlier in the response surface models developed using all six nonlinear methods. This compound was indicated in Figure 2. 4-Nitroaniline was then removed from model fitting. Using the remaining 161 compounds, the final response surface models were obtained. A summary of these models was presented in Table 3. As Table 3 indicated, all the nonlinear methods used in this study provided comparable power of model fitting, with the response surface based on PPR being slightly better than the others. A plot of the fitted versus observed toxicity using the PPR model after removing 4-nitroaniline changed little from Figure 2 and was consequently not provided. The residual plot for the PPR model after removing 4-nitroaniline was shown in Figure 3. No special pattern was observed. Residual plots for the other five models (not shown) did not reveal any special pattern either.

Model Validation. The results of the 10-fold CV of the six response surface models were summarized in Table 4. Results in Table 4 indicated that all the models had r^2_{CV} close

Table 3. Summary of Response Surface Model Fitting and Prediction[a]

nonlinear method	r^2 (r^2_{pred})	s	F
GAM	0.75	0.32	464.66
	(0.61)	(0.46)	(60.07)
LASSO2	0.73	0.33	430.96
	(0.69)	(0.42)	(87.96)
LOESS	0.75	0.32	483.81
	(0.73)	(0.46)	(105.91)
MARS	0.73	0.33	426.92
	(0.74)	(0.42)	(112.71)
PPR	0.80	0.29	647.95
	(0.71)	(0.43)	(93.49)
NN	0.73	0.33	438.49
	(0.73)	(0.45)	(107.21)

[a] All p values for the F ratio are < 0.0001. Statistics in parentheses are those for model prediction.

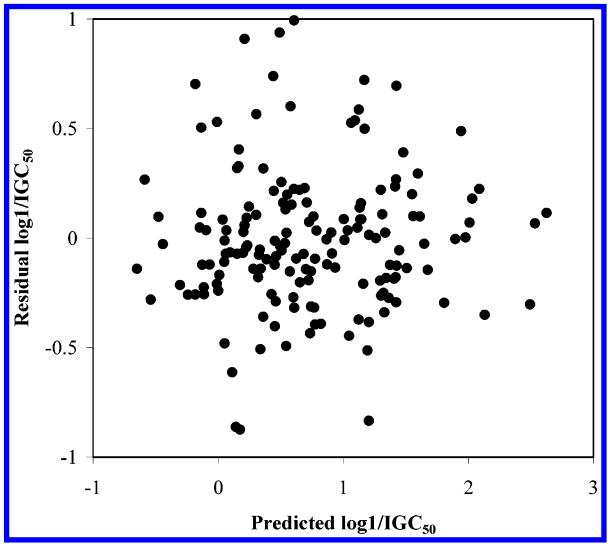

Figure 3. Residual plot for the PPR response surface model.

Table 4. Summary of Results of 10-Fold Cross-Validation of the Response Surface Models

nonlinear method	r^2_{CV}	average PMSE	standard error of average PMSE
GAM	0.70	0.16	0.05
LASSO2	0.70	0.16	0.06
LOESS	0.70	0.17	0.05
MARS	0.69	0.17	0.06
PPR	0.71	0.17	0.06
NN	0.70	0.16	0.05

to 0.70 and that all models had similar average PMSE values and associated standard errors. The statistics in Table 4 indicated that all six models were almost equally robust.

The results of the assessment of the prediction accuracy of the six models on the external test set were presented in Table 3. Plots showing predicted versus observed toxicity for compounds in the test set were shown in Figure 4. Of particular note is Figure 4(a),(b) that corresponds to the models based on GAM and LASSO2, respectively. As seen in Figure 4(a), the toxicity of two compounds with high observed toxicity, i.e., 2-chloro-3,5-dinitropyridine and 4-chloro-3,5-dinitrobenzonitrile (the last two compounds in Table 2), was significantly underpredicted by the GAM model. In Figure 4(b), although a significant correlation was obtained between toxicity predicted by the LASSO2 response surface and the observed toxicity, it appeared that toxicity might be systematically underestimated for several compounds with high observed toxicity. For these reasons, caution should be exercised should the GAM and LASSO2 response surface models be used.

Table 5. Outliers to the Response Surface Model Based on Multiple Linear Regression in Ref 9 and Their A_{Max} Values

compound	A_{max}
4-nitroaniline	0.2961
2-hydroxy-4-methyl-3-nitropyridine	0.3311
3,4-dinitrophenol	0.3823
3-methyl-4-nitrophenol	0.3039
4-nitrophenol	0.3080
4-amino-2-nitrophenol	0.2985
4,6-dichloro-5-nitropyrimidine	0.4121
3-cyano-4,6-dimethyl-2-hydroxypyridine	0.2930
4-nitrodiphenylamine	0.3016
4-nitro-1-naphthalamine	0.3056

In addition to the above observations, the response surface models based on GAM and LASSO2 had relatively low prediction capacity as demonstrated by the r^2_{pred} value in Table 3 and the degree of scattering in Figure 4(a),(b). The performances of the other four models (LOESS, MARS, PPR, and NN) were similar for the aromatic chemicals data set. Identifying the "best" model could be subjective and was not the aim of the present study.

DISCUSSION

The molecular descriptors used in the developing the response surface models in this study are well-established and toxicologically relevant descriptors. The hydrophobicity descriptor logK_{ow} captures the uptake, transport, and distribution of a toxicant, and the electrophilicity descriptor A_{max} captures the interaction between a toxicant and the site of toxic action. Together with molecular descriptors such as E_{lumo}, these descriptors are known to be "global" molecular descriptors and are frequently used in the response surface methodology (e.g., ref 25). In this study, however, logK_{ow} was used with A_{max} instead of E_{lumo} because Cronin et al.[9] demonstrated that A_{max} is superior to E_{lumo} for modeling the acute toxicity of aromatic compounds to *T. pyriformis*, as demonstrated by eqs 2 and 3.

The response surface methodology is considered mechanistic-based because it involves preselected global descriptors that characterize toxicokinetic and toxicodynamic effects.[8] A response surface model essentially expresses toxicity as the result of the uptake of the toxicant molecule into the biophase and the interaction between the toxicant molecule and the site of action, a notion proposed by McFarland.[5] Variations of existing response surface models are caused primarily by the choices of molecular descriptors in the model, e.g., logK_{ow} versus log$D$ (where D is the ionization-corrected octanol/water partition coefficient) and E_{lumo} versus A_{max}. Very occasionally, indicator variables describing the presence or absence of certain characteristic groups (e.g., $-NH_2$, $-NO_2$) are included in response surface models. In this circumstance, the models are referred to as modified response surface models.[26] Introducing additional descriptors to a response surface model is possible. However, this may lead to deviations from the notion proposed by McFarland[5] and compromises the straightforward interpretation of the model. For this reason, although nonlinear relationships

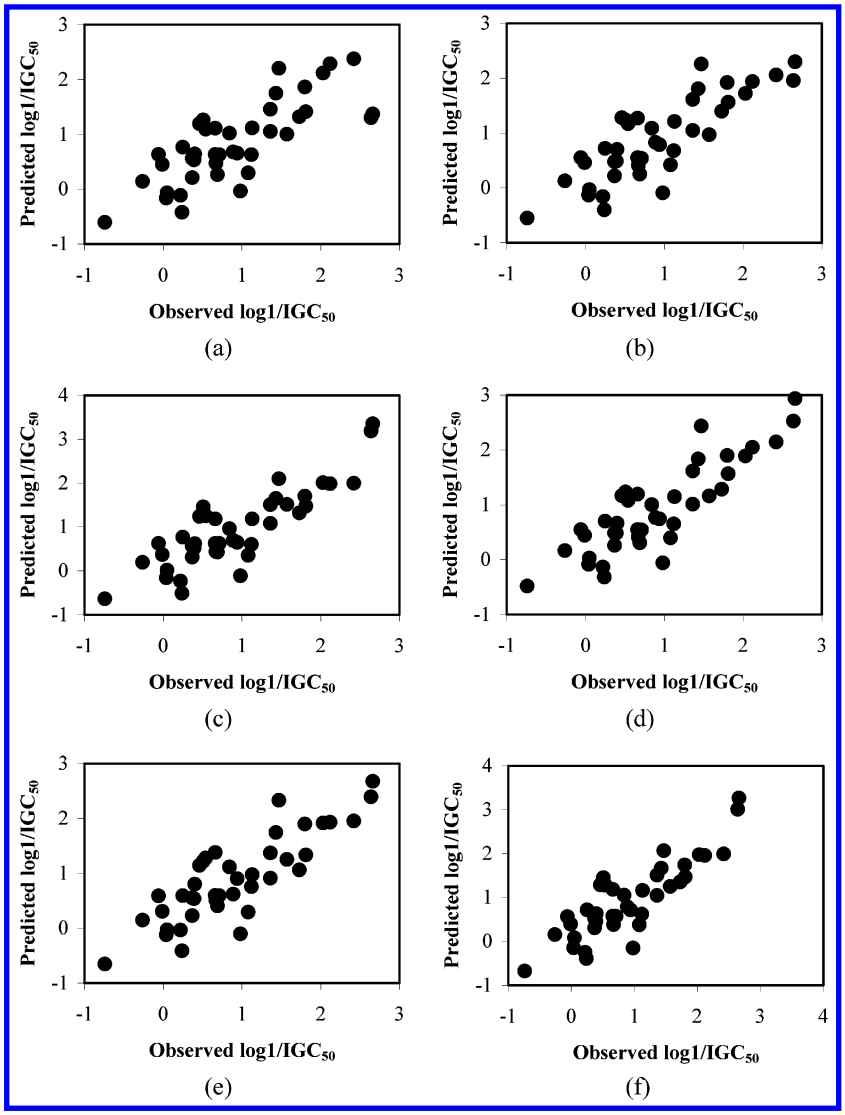

Figure 4. Predicted versus observed toxicity for compounds in the test set: (a) GAM, (b) LASSO2, (c) LOESS, (d) MARS, (e) PPR, and (f) NN. Note change of scales in (f).

between toxicity and other molecular descriptors may exist, the scope of the present study is limited to the response surface methodology involving logK_{ow} and A_{max} only.

The *T. pyriformis* data set was chosen for response surface analyses in this study because of its high data quality. Unlike some data compiled in the literature that came from various sources, the *T. pyriformis* data set used in this study were obtained in the same laboratory and the toxicity measured by a single protocol, some of the desired characteristics of the biological data for developing QSARs.[1] Since the aim of the present study was to develop response surface models considering curvatures in the surface (caused by nonlinearity), it was crucial that such curvatures were not the results of experimental errors, i.e., the intrinsic (experimental) errors contained in the data should be as small as possible. In a recent study in which a similar *T. pyriformis* data set for aromatic compounds[10] was analyzed, Mekapati and Hansch[27] demonstrated the excellent experimental work for data acquisition of the *T. pyriformis* testing system.

In the above-mentioned study, Mekapati and Hansch[27] also showed that by achieving "uniform activity", i.e., compounds acting by the same mechanism, an r^2 value of 0.9 could be obtained in modeling the *T. pyriformis* data. The advantage of doing so was that a high quality QSAR could be established which could be used to predict the toxicity of compounds with the same mechanism (i.e., mechanism-based QSAR). While this is desirable, in another study of Cronin et al.,[26] the authors obtained a response surface model (after deleting some statistical outliers thought to be compounds associated with specific mechanisms of toxic action) for *T. pyriformis* data with an r^2 value of 0.81 and an *s* value of 0.34. Without further deleting statistical outliers, the authors stated that with these r^2 and *s* values, little further improvement could be made to the model. This statement was valid because, as mentioned in the Introduction, the response surface methodology was intended to encompass compounds with more than one toxicity mechanism, and excessive deletion of compounds (acting by different mechanisms) from the response surface model was contradictory to this purpose, though better statistical fit could be obtained.

Direct comparisons between the response surface models developed in the present study with that developed by Cronin et al.[26] mentioned above would not be appropriate due to the different data set and model descriptors involved. However, the criteria of an r^2 value being approximately 0.8 and an *s* value being approximately 0.3, though arbitrary,

can be used as informal indicators of satisfactory model fit of the response surface methodology for the *T. pyriformis* data. All six nonlinear models developed in this study had *s* values close to 0.3 and r^2 values slightly lower than 0.8. However, examining the predictive capacity of the models revealed that only models based on LOESS, MARS, PPR, and NN offered satisfactory predictive power (and subsequent discussions will be focused on these four models). The potential inadequacy of the models based on GAM and LASSO2 has been described previously. An additional note for the model based on GAM is that its prediction accuracy for the external validation set was relatively low (r^2_{pred} = 0.61, see Table 3). According to Tropsha et al.,[24] an r^2_{cv} substantially lower than r^2 indicates that the model is unstable. This was not the case for the model based on GAM because the r^2_{cv} value (0.70, Table 4) for this model was only slightly lower than the r^2 value (0.75, Table 3). Cronin et al.[26] pointed out that the process of validation is highly dependent on the selection of the test data set. Thus, the low r^2_{pred} value for the GAM model may be attributable to some peculiar pattern in the external validation set that challenged the GAM model only. In addition to detecting potential inadequacies of the GAM and LASSO2 models, results of toxicity prediction for the external test set also provided valuable information on the applicability of the other four models. The LOESS, MARS, PPR, and NN models provided toxicity predictions for 28, 32, 31, and 31 of the 41 compounds in the external validation set that were within 0.5 log units of the corresponding observed values, respectively. All these four models overpredicted the toxicity of several compounds with dinitro groups in the 2,3-, 3,5-, and 3,6- positions. No common structural element was identifiable for the compounds whose toxicity was underpredicted. Therefore, the applicabilities of the four models are similar.

The response surface methodology has been used not only for toxicity modeling and prediction but also for the determination of the mechanisms of toxic action of compounds. As such, a response surface can be developed, and those compounds whose toxicity cannot be adequately described by the response surface are identified. These compounds are then further studied, in particular their chemical structures, and mechanistic explanations are provided to account for their departures from the response surface. While several mechanisms of toxic action are well established (e.g., ref 28) and mechanism-based QSARs are developed (e.g., ref 29), the use of QSARs without discriminating between the toxicity mechanisms of chemical compounds is also desirable.[9,11] Cronin et al.[9] identified 10 compounds (shown in Table 5) as outliers to the response surface represented by eq 3, the response surface model based on linear regression. Examination of these compounds by the authors revealed that most of them were 4-substituted nitrophenols and nitroanilines that might be metabolized or differed from other compounds in the data set in other ways. Further inspection of these compounds revealed that they have A_{max} values (also shown in Table 5) approximately between 0.30 and 0.42. This A_{max} range corresponds to a region in which the curvilinear effect of A_{max} on toxicity is most severe (Figure 1). All the above-mentioned 10 compounds except 4-amino-2-nitrophenol were included in the training set in the present study. Results in this study showed that, among the compounds used in the training set, only 4-nitroaniline was identified as an outlier in this study, suggesting that more structural characteristics and potentially more mechanisms can be captured in the response surface models based on nonlinear methods. For 4-amino-2-nitrophenol, the difference between predicted and observed toxicity (Table 2) may suggest that this compound could be considered another outlier to most of the response surface models developed.

Results of this study also seemed to suggest that the need for mechanism-based QSARs is created at least in part by the modeling approaches that are commonly used today. This should not be taken to mean that mechanism-based QSARs are not important, but rather it suggests that a more "versatile" QSAR can be developed using more powerful statistical models (results of this study) and/or statistically more efficient model descriptors (results of ref 9). A good analogy in this regard is the use of $\log K_{DMPC}$ (logarithm of the dimyristoyl phophatidyl-choline/water partition coefficient) to replace $\log K_{ow}$ in modeling the toxicity of narcotic compounds. It is well-known that polar narcotic compounds have toxicity higher than that of nonpolar narcotic compounds and that separate $\log K_{ow}$-based QSARs need to be developed for each type of narcotic compounds. However, several authors (e.g., refs 30 and 31) have demonstrated that using $\log K_{DMPC}$ one single QSAR was needed to replace separate $\log K_{ow}$-dependent QSARs for narcotic compounds and that the difference between polar and nonpolar narcotic compounds disappeared. This exemplifies how a change in the modeling method (in this particular case, a change in the model descriptor) can mitigate the effect of knowledge about toxicity mechanisms on developing QSARs.

Mechanistic studies are valuable in developing a science to understand the interaction between chemical compounds and living organisms. In addition to several other methods (e.g., refs 32−34), the response surface methodology has been used for mechanism identification, as mentioned previously. While providing higher modeling capacity, the response surface models developed using nonlinear method have relatively limited use for mechanistic studies. As mentioned previously, Cronin et al.[9] identified several compounds as outliers to the response surface represented by eq 3. In the present study, however, the number of statistical outlier was narrowed down, and 4-nitroaniline and 4-amino-2-nitrophenol were considered outliers. Cronin et al.[9] suggested that 4-amino-2-nitrophnol may have the potential for metabolism to redox cycler but provided no reasons for the outlying phenomenon of 4-nitroaniline. What caused 4-nitroaniline to be different from other 4-substituted nitrophenols or nitroanilines (that were no longer statistical outliers in the present study) was not investigated in this study (this assumes that there was not unduly experimental error and the molecular descriptors were correctly measured/calculated). However, this suggested that the response surface models based on nonlinear methods were "conservative" in identifying compounds with different mechanisms and were not well suited for mechanistic studies. On the other hand, compared to a response surface based on linear regression, a response surface model based on nonlinear methods would provide a stronger signal that a different mechanism is involved if the toxicity of a compound cannot be adequately described by it.

In addition to relatively limited use in mechanistic studies, the response surface models based on most nonlinear methods (such as LOESS, MARS, PPR, and NN) have the disadvantage of not being able to be explicitly expressed. Although disputed by some researchers,[14,35] NN is generally conceived as a "black-box" methodology. The relationship between the dependent and independent variables based on NN cannot be explicitly expressed. LOESS, MARS, and PPR are similar to NN in this sense because there are no regression coefficients and consequently the models cannot be described by mathematical equations. The traditional response surface methodology based on linear regression is superior in this regard since the importance of QSAR model transparency and interpretability has been discussed in details by Cronin and Schultz[13] and Schultz and Cronin.[1] Although the response surface methodology with nonlinear modeling methods does not deviate from the basic response surface concept (toxicity is the combined result of uptake and interaction) because of the model predictors used ($\log K_{ow}$ and A_{max}), the response surface methodology with most nonlinear regression methods does not allow direct rationalization of the toxic effects on living organisms because interpretation of the resulting models is difficult or not possible at all.

Among the four nonlinear methods, LOESS, MARS, and PPR are within the realm of statistical methods (they are considered nonlinear/nonparametric regression methods). NN, on the other hand, is sometimes not considered a statistical method, as seen in the title of the work of Basak et al.[16] The underlying computation algorithms of the nonlinear regression methods and NN are very different. LOESS, MARS, and PPR make use of "smoothers", whereas NN employs nonlinear transformations of the input variables. Nonetheless, the resultant models based on all these methods allow for curvatures, an advantage over traditional linear regressions. Due to its wider use than LOESS, MARS, and MARS, NN has received greater attention in the literature. Given the similar applicability and predictive capacity of the LOESS, MARS, PPR, and NN for the aromatic chemicals data set as demonstrated previously, the choice of modeling method may simply be the personal preference of the model developer. However, note that the data set used in this study can only be considered a small, or at most, a medium sized set. For large data sets with intrinsic nonlinearity, the advantage of NN over nonlinear regression methods may be more evident.

CONCLUSIONS

A data set with the toxicity of aromatic compounds to *Tetrahymena pyriformis* was analyzed using the response surface methodology with nonlinear methods. The nonlinear regression methods considered were GAM, LASSO2, LOESS, MARS, and PPR. All models were cross-validated, and their predictive powers were tested on an external data set. Results showed that response surface models with satisfactory modeling and prediction accuracy could be developed using LOESS, MARS, PPR, and NN, with 4-nitroaniline and 4-amino-2-nitrophenol considered outliers to the response surfaces. All these four models had similar applicabilities. Compared to response surface models based on linear regression, the response surface models based on nonlinear methods could be used to describe the toxicity of compounds with greater structural variability and a larger variety of mechanisms of toxic action. However, they were conservative in identifying compounds with different mechanisms and were difficult to interpret.

ACKNOWLEDGMENT

The author wishes to thank Helen M. Delgado for motivation and support. The author is grateful to the reviewers for their highly constructive comments on the manuscript.

APPENDIX

Key R Codes Used in Nonlinear Modeling. (X: data matrix with all variables, x: matrix of independent variables, y: dependent variable, "Toxicity" represented $\log 1/IGC_{50}$); GAM: obj <- gam(Toxicity~s(logKow)+s(Amax),data=X, family=Gaussian()); LASSO2: obj <- l1ce(Toxicity~logKow +Amax,data=X,bound=10,absolute.t=TRUE); LOESS: obj <- loess(y~logKow+Amax,X,control=loess.control(surface ="direct")); MARS: obj <- mars(x,y); PPR: obj <- ppr-(x,y,nterms=2); NN: obj <- nnet(x,y,size=3,decay=1e-3, linout=T,skip=T,maxit=50).

REFERENCES AND NOTES

(1) Schultz, T. W.; Cronin, M. T. D. Essential and desirable characteristics of ecotoxicity quantitative structure−activity relationships. *Environ. Toxicol. Chem.* **2003**, *22*, 599−607.
(2) Gunatilleka, A. D.; Poole, C. F. Models for estimating the non-specific aquatic toxicity of organic compounds. *Anal. Commun.* **1999**, *36*, 235−242.
(3) Bearden, A. P.; Schultz, T. W. Structure−activity relationships for *Pimephales* and *Tetrahymena*: A mechanism of action approach. *Environ. Toxicol. Chem.* **1997**, *16*, 1311−1317.
(4) Ferguson, J. The use of chemical potentials as indices of toxicity. *Proc. R. Soc. London B* **1939**, *127*, 387−403.
(5) McFarland, J. W. On the parabolic relationship between drug potency and hydrophobicity. *J. Med. Chem.* **1970**, *13*, 1092−1196.
(6) Mekenyan, O. G., and Veith, G. D. Relationships between descriptors for hydrophobicity and soft electrophilicity in predicting toxicity. *SAR QSAR Environ. Res.* **1993**, *1*, 335−344.
(7) Schultz, T. W.; Cronin, M. T. D.; Netzeva, T. I. The present status of QSAR in toxicology. *J. Mol. Struct.* (*THEOCHEM*) **2003**, *622*, 23−38.
(8) Wang, X.; Sun, C.; Wang, Y.; Wang, L. Quantitative structure−activity relationships for the inhibition toxicity to root elongation of *Cucumis sativus* of selected phenols and interspecies correlation with *Tetrahymena pyriformis*. *Chemosphere* **2002**, *46*, 153−161.
(9) Cronin, M. T. D.; Manga, N.; Seward, J. R.; Sinks, G. D.; Schultz, T. W. Parametrization of electrophilicity for the prediction of the toxicity of aromatic compounds. *Chem. Res. Toxicol.* **2001**, *14*, 1498−1505.
(10) Cronin, M. T. D.; Schultz, T. W. Development of quantitative structure−activity relationships for the toxicity of aromatic compounds to *Tetrahymena pyriformis*: Comparative assessment of the methodologies. *Chem. Res. Toxicol.* **2001**, *14*, 1284−1295.
(11) Schultz, T. W.; Cronin, M. T. D. Response-surface analyses for toxicity to *Tetrahymena pyriformis*: Reactive carbonyl-containing aliphatic chemicals. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 304−309.
(12) Neter, J.; Kuter, M. H; Nachtscheim, C. J.; Wasserman, W. *Applied Linear Regression Models*; Irwin: Chicago, IL, 1996.
(13) Cronin, M. T. D.; Schultz, T. W. Pitfalls in QSAR. *J. Mol. Struct.* (*THEOCHEM*) **2003**, *622*, 39−51.
(14) Kaiser, K. L. E. The use of neural networks in QSARs for acute aquatic toxicological endpoints. *J. Mol. Struct.* (*THEOCHEM*) **2003**, *622*, 85−95.
(15) Kaiser, K. L. E. Neural networks for effect prediction in environmental and health issues using large datasets. *QSAR Comb. Sci.* **2003**, *22*, 185−190.
(16) Basak, S. C.; Grunwald, G. D.; Gute, B. D.; Balasubramanian, K.; Opitz, D. Use of statistical and neural net approaches in predicting toxicity of chemicals. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 885−890.
(17) Hastie, T. J.; Tibshirani, R. J. *Generalized Additive Models*; Chapman and Hall: New York, 1990.

(18) Osborne, M. R.; Presnell, B.; Turlach, B. A. On the LASSO and its dual. *J. Comput. Graphical Stat.* **2000**, *9*, 319−337.
(19) Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *J. Ame. Stat. Asso.* **1979**, *46*, 175−185.
(20) Friedman, J. H. Multivariate additive regression splines. *Annals. Stat.* **1991**, *19*, 1−141.
(21) Müller, B. *Neural Networks*: *An Introduction*; Springer-Verlag: New York, 1990.
(22) Friedman, J. H.; Stuetzle, W. Projection pursuit regression. *J. Ame. Stat. Asso.* **1981**, *76*, 817−823.
(23) Schultz, T. W. TETRATOX: *Tetrahymena* population growth impairment endpoint − A surrogate for fish lethality. *Toxicol. Methods* **1997**, *7*, 289−309.
(24) Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69−77.
(25) Dimitrov, S. D.; Mekenyan, O. G.; Sinks, G. D.; Schultz, T. W. Global modeling of narcotic chemicals: Ciliate and fish toxicity. *J. Mol. Struct.* (*THEOCHEM*) **2003**, *622*, 63−70.
(26) Cronin, M. T. D.; Aynur, O. A.; Duffy, J. C.; Netzeva, T. I.; Rowe, P. H.; Valkova, I. V.; Schultz, T. W. Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to *Tetrahymena pyriformis. Chemosphere* **2002**, *49*, 1201−1221.
(27) Mekapati, S. B.; Hansch, C. On the parametrization of the toxicity of organic chemicals to *Tetrahymena pyriformis.* The problem of establishing a uniform activity. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 956−961.
(28) Schultz, T. W.; Sinks, G. D.; Bearden, A. P. QSAR in aquatic toxicology: A mechanism of action approach comparing toxic potency to *Pimephales promelas*, *Tetrahymena pyriformis*, and *Vibrio fischeri*. In *Comparative QSAR*; Devillers, J., Ed.; Taylor & Francis: New York, 1998; pp 51−109.
(29) Wang, X.; Dong, Y.; Wang, L.; Han, S. Acute toxicity of substituted phenols to *Rana japonica* tadpoles and mechanism-based quantitative structure−activity relationship (QSAR) study. *Chemosphere* **2001**, *44*, 447−455.
(30) Schultz, T. W.; Seward, J. R. Dimyristoyl phosphatidylcholine/water partitioning-dependent modeling of narcotic toxicity to *Tetrahymena pyriformis*. *Quant. Struct.-Act. Relat.* **2000**, *19*, 339−344.
(31) Vaes, W. H. J.; Urrestarazu Ramos, E.; Verhaar, H. J. M.; Hermens, J. L. M. Acute toxicity of nonpolar versus polar narcosis: Is there a difference? *Environ. Toxicol. Chem.* **1998**, *17*, 1380−1384.
(32) Ren, S. Determining the mechanism of toxic action of phenols to *Tetrahymena pyriformis*. *Environ. Toxicol.* **2002**, *17*, 119−127.
(33) Ren, S. Predicting three narcosis mechanisms of aquatic toxic action. *Tox. Lett.* **2002**, *133*, 127−139.
(34) Russom, C. L.; Bradbury, S. P.; Broderium, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environ. Toxicol. Chem.* **1997**, *16*, 948−967.
(35) Niculescu S. P. Artificial neural networks and genetic algorithms in QSAR. *J. Mol. Struct.* (*THEOCHEM*) **2003**, *622*, 71−83.

CI034046Y