

# Quantitative structure–polarization relationships (QSPR) study of BTEX tracers for the formation of antibody–BTEX–EDF complex

Taesung Moon,<sup>a</sup> Myung Whan Chi,<sup>b</sup> Myung Ja Choi<sup>b</sup>  
and Chang No Yoon<sup>b,\*</sup>

<sup>a</sup>*Nanormics, Inc., 10-57 Hawolgokdong, Sungbukku, Seoul 136-130, South Korea*

<sup>b</sup>*Bioanalysis and Biotransformation Research Center, Korea Institute of Science and Technology,  
PO Box 131 Cheongryang, Seoul 130-650, South Korea*

Received 14 May 2003; revised 19 April 2004; accepted 20 April 2004

**Abstract**—The multiple linear regression (MLR) analysis and back propagation neural networks (NN) were performed to examine the quantitative structure–polarization relationships (QSPR) for the formation of antibody–BTEX–EDF complex. Five descriptors out of 18 ones were selected for both MLR and NN, respectively, and the selected descriptors in MLR were the same as those in NN. These descriptors were the number of atoms, which can form hydrogen bonds (HA), connolly surface area (Area), the highest occupied molecular orbital energy (HOMO), partial charge of C<sub>3</sub> carbon atom (C<sub>3</sub>), and HOMO  $\pi$  coefficient of C<sub>2</sub> carbon atom (P<sub>2</sub>). The fact that the descriptors in MLR are identical to those in NN suggests that these descriptors have good linear relationships and play a significant role in the formation of antibody–tracer complex.

© 2004 Elsevier Ltd. All rights reserved.

## 1. Introduction

BTEX compounds (benzene, toluene, ethylbenzene, and xylene) are the most toxic aromatic constituents of petroleum products such as diesel fuel and gasoline. Benzene and ethylbenzene are used in the production of synthetic materials and consumer products, such as synthetic rubber, plastics, nylon, and paints. Toluene is used as a solvent for paints, coatings, gums, oils, and resins. Xylene is often used as a solvent in printing, rubber, and leather industries.<sup>1</sup> The main source of BTEX contamination is the leakage of gasoline from poorly maintained underground storage tanks. BTEX is soluble in water in the  $\mu\text{g/mL}$  range and at least benzene is considered highly toxic and carcinogenic.<sup>2</sup> The maximum contaminant level for benzene in drinking water is only 1–5  $\mu\text{g/mL}$ .<sup>3</sup> A lot of analytical methods have been used to determine BTEX. These include gas chroma-

tography,<sup>4,5</sup> sensor and absorption spectra,<sup>6–9</sup> and fluorescence polarization immunoassay (FPIA).<sup>10</sup> Fluorescence polarization immunoassay is a useful approach to detect BTEX with the advantage of simplicity and high precision. The fluorescence polarization is based on the principle that molecule rotates in liquid and the rate of rotation is related with the size of molecule. Unbound fluorescence-labeled BTEX (BTEX–EDF) tracer rotates more rapidly than the antibody–tracer complex. When sample contains BTEX, the BTEX–EDF tracer competes with BTEX at the antibody binding sites and the polarization signal decreases in the formation of antibody–BTEX–EDF complex.

The significance of fluorescence polarization immunoassay (FPIA) is the simplicity and high precision for detection of BTEX in compared with time-consuming and very expensive traditional methods. Since the experimental data can be obtained by using FPIA with ease, the quantitative structure–polarization relationships (QSPR) models can be achieved and used for the detection of BTEX. The QSPR studies were carried out to obtain further insight into the relationships between the chemical structure and fluorescence polarization of

**Keywords:** BTEX tracer; QSPR; Multiple linear regression; Neural network.

\* Corresponding author. Tel.: +82-2-958-5068; fax: +82-2-958-5059; e-mail: [cody@kist.re.kr](mailto:cody@kist.re.kr)

BTEX tracer. The multiple linear regression (MLR) and neural network (NN) were performed to find out the relationships. While there are several neural network methods, the back propagation algorithm was used in this study.

## 2. Results and discussion

### 2.1. Multiple linear regression analysis

Table 1 shows the chemical structures and difference of fluorescence polarization (FP) between bound and unbound BTEX tracers<sup>10</sup> used in this study. Of 27 BTEX tracers, the 18 ones were selected as training set members by using D-optimal design (Table 2).<sup>11</sup> The four ones (2,3-dimethylbenzoic acid, 2,4-dimethylbenzoic acid, 2,5-dimethylbenzoic acid, and 3,5-dimethylbenzoic acid) in Table 1 were not included in analyses because the values of fluorescence polarization were not clear in experimental procedures. When they were included in MLR, the predictivity in test set was very poor whereas the regression in training set gave high  $r^2$  value of 0.867.

The high crossvalidated  $r^2$  value of 0.654 was achieved using five descriptors of the number of atoms, which can

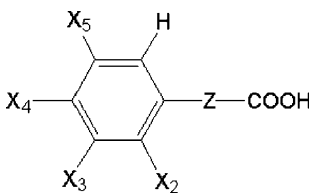
form hydrogen bonds (HA), connolly surface area (Area), the highest occupied molecular orbital energy (HOMO), partial charge of C<sub>3</sub> carbon atom of aromatic ring (C<sub>3</sub>), and HOMO  $\pi$  coefficient of C<sub>2</sub> carbon atom of aromatic ring (P<sub>2</sub>). The optimum regression model is as follows

$$\begin{aligned} \text{FP} = & -2.404[\text{HA}] + 0.230[\text{Area}] - 10.027[\text{HOMO}] \\ & (\pm 0.575) \quad (\pm 0.032) \quad (\pm 1.770) \\ & + 28.566[\text{C}_3] + 4.474[\text{P}_2] - 123.525 \\ & (\pm 17.188) \quad (\pm 1.520) \quad (\pm 20.592) \end{aligned} \quad (1)$$

$$N = 18; \quad r^2 = 0.853; \quad s = 1.679; \quad F = 13.919$$

This result suggests that the difference of fluorescence polarization (FP) between bound and unbound BTEX tracers is linearly dependent on these descriptors. However, it is not always possible to discuss about the effect of individual descriptor with the value of its coefficient since the coefficient results from the combination of all the descriptors. When these descriptors were used in MLR individually, the HA, Area, HOMO, C<sub>3</sub>, and P<sub>2</sub> gave  $r^2$  values of 0.021, 0.258, 0.003, 0.018, and 0.215, respectively. The difference of fluorescence polarization between bound and unbound BTEX tracers

**Table 1.** Chemical structures and difference of fluorescence polarization of BTEX tracers<sup>10</sup>



No.	BTEX tracers	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	Z (liker)	FP <sup>a</sup>
1	Benzoic acid	H	H	H	H	—	2.5
2	3-Phenylpropionic acid	H	H	H	H	CH <sub>2</sub> CH <sub>2</sub>	8.7
3	4-Phenylbutyric acid	H	H	H	H	CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub>	14.5
4	6-Phenylhexanoic acid	H	H	H	H	CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub>	19.9
5	<i>o</i> -Toluic acid	CH <sub>3</sub>	H	H	H	—	4.0
6	<i>m</i> -Toluic acid	H	CH <sub>3</sub>	H	H	—	4.4
7	<i>p</i> -Toluic acid	H	H	CH <sub>3</sub>	H	—	4.7
8	<i>o</i> -Tolylacetic acid	CH <sub>3</sub>	H	H	H	CH <sub>2</sub>	2.8
9	<i>m</i> -Tolylacetic acid	H	CH <sub>3</sub>	H	H	CH <sub>2</sub>	4.4
10	<i>p</i> -Tolylacetic acid	H	H	CH <sub>3</sub>	H	CH <sub>2</sub>	4.4
11	<i>o</i> -Tolylpropionic acid	CH <sub>3</sub>	H	H	H	CH <sub>2</sub> CH <sub>2</sub>	4.4
12	<i>p</i> -Tolylpropionic acid	H	H	CH <sub>3</sub>	H	CH <sub>2</sub> CH <sub>2</sub>	3.7
13	2,3-Dimethylbenzoic acid	CH <sub>3</sub>	CH <sub>3</sub>	H	H	—	6.9
14	2,4-Dimethylbenzoic acid	CH <sub>3</sub>	H	CH <sub>3</sub>	H	—	7.0
15	2,5-Dimethylbenzoic acid	CH <sub>3</sub>	H	H	CH <sub>3</sub>	—	7.2
16	3,4-Dimethylbenzoic acid	H	CH <sub>3</sub>	CH <sub>3</sub>	H	—	7.1
17	3,5-Dimethylbenzoic acid	H	CH <sub>3</sub>	H	CH <sub>3</sub>	—	7.2
18	3,4-Dimethylphenoxyacetic acid	H	CH <sub>3</sub>	CH <sub>3</sub>	H	OCH <sub>2</sub>	5.8
19	3,4-Dimethylphenylthioacetic acid	H	CH <sub>3</sub>	CH <sub>3</sub>	H	SCH <sub>2</sub> CH <sub>2</sub>	6.3
20	3,4-Dimethylanilide succinic acid	H	CH <sub>3</sub>	CH <sub>3</sub>	H	NHCOCH <sub>2</sub> CH <sub>2</sub>	5.4
21	4-Tolylidide succinic acid	H	H	CH <sub>3</sub>	H	NHCOCH <sub>2</sub> CH <sub>2</sub>	4.4
22	3,4-Dichloranilide succinic acid	H	Cl	Cl	H	NHCOCH <sub>2</sub> CH <sub>2</sub>	5.9
23	3,4-Dichlorophenylacetic acid	H	Cl	Cl	H	CH <sub>2</sub>	10.5
24	3,4-Dichlorophenoxyacetic acid	H	Cl	Cl	H	OCH <sub>2</sub>	9.2
25	2,4-Dichlorophenoxyacetic acid	Cl	H	Cl	H	OCH <sub>2</sub>	8.3
26	2,4,5-Trichlorophenoxyacetic acid	Cl	H	Cl	Cl	OCH <sub>2</sub>	10.5
27	Ferroceneacetic acid	H	H	H	H	CH <sub>2</sub>	3.0

<sup>a</sup> FP = (fluorescence polarization of unbound BTEX tracers–fluorescence polarization of bound tracers)/10.

**Table 2.** Descriptors used in multiple linear regression (MLR) and neural network (NN)

BTEX tracers	HA <sup>a</sup>	Area <sup>b</sup>	HOMO <sup>c</sup>	C <sub>3</sub> <sup>d</sup>	P <sub>2</sub> <sup>e</sup>
[Training set]					
Benzoic acid	2.000	156.487	−10.062	−0.145	−0.510
4-Phenylbutyric acid	2.000	206.299	−9.588	−0.127	0.303
6-Phenylhexanoic acid	2.000	235.223	−9.516	−0.127	0.306
<i>o</i> -Toluic acid	2.000	163.132	−9.736	−0.150	0.529
<i>o</i> -Tolylacetic acid	2.000	177.053	−9.497	−0.132	−0.534
<i>m</i> -Tolylacetic acid	2.000	178.402	−9.527	−0.067	−0.186
<i>p</i> -Tolylacetic acid	2.000	174.442	−9.463	−0.129	0.257
<i>o</i> -Tolylpropionic acid	2.000	197.068	−9.471	−0.129	−0.529
<i>p</i> -Tolylpropionic acid	2.000	195.521	−9.415	−0.127	−0.266
3,4-Dimethylbenzoic acid	2.000	168.730	−9.648	−0.080	0.162
3,4-Dimethylphenylthioacetic acid	3.000	228.739	−8.306	−0.047	0.238
3,4-Dimethylanilide succinic acid	4.000	241.074	−8.704	−0.107	0.353
4-Tolylidide succinic acid	4.000	225.920	−8.770	−0.106	−0.303
3,4-Dichloranilide succinic acid	6.000	234.810	−9.216	−0.019	0.248
3,4-Dichlorophenoxyacetic acid	5.000	209.239	−9.532	−0.006	0.220
2,4-Dichlorophenoxyacetic acid	5.000	205.344	−9.512	−0.080	0.351
2,4,5-Trichlorophenoxyacetic acid	6.000	221.222	−9.657	−0.073	0.388
Ferroceneacetic acid	2.000	166.957	−9.721	−0.129	0.287
[Test set]					
3-Phenylpropionic acid	2.000	194.127	−9.672	−0.127	−0.299
<i>m</i> -Toluic acid	2.000	166.442	−9.767	−0.082	−0.378
<i>p</i> -Toluic acid	2.000	166.491	−9.844	−0.145	−0.236
3,4-Dimethylphenoxyacetic acid	3.000	201.059	−8.987	−0.029	−0.225
3,4-Dichlorophenylacetic acid	4.000	195.163	−9.857	−0.042	0.030

<sup>a</sup> HA: the number of atoms, which can form hydrogen bonds.<sup>b</sup> Area: connolly surface area.<sup>c</sup> HOMO: the highest occupied molecular orbital energy.<sup>d</sup> C<sub>3</sub>: partial charge of C<sub>3</sub> carbon atom.<sup>e</sup> P<sub>2</sub>: HOMO  $\pi$  coefficient of C<sub>2</sub> carbon atom.

means the degree of the formation of antibody–BTEX–EDF complex. The greater the difference, the better the complex formation. The descriptor, HA, suggests that the hydrogen bond between tracer and antibody plays an important role in the formation of antibody–tracer complex. The descriptor, Area, is related with the molecular shape, which is important factor in the binding of ligand to antibody. The substituent at 2 or 3 position appears to be involved in the interactions with antibody since the partial charge and the HOMO  $\pi$  coefficient of ring carbon mainly relies on its substituent. The HOMO has an effect on the formation of complex cooperatively with other terms whereas has little effects by itself ( $r^2 = 0.003$ ). On the other hand, the number of linker atoms (Table 1) also appears as a significant monoparametric parameter ( $r^2$  value of 0.341 and coefficient of 0.639), which means that the size of the linker between benzene ring and carboxyl group is important in the formation of antibody–BTEX–EDF complex.

## 2.2. Neural networks

The neural networks were carried out with descriptors selected by using pruning method.<sup>11–13</sup> The pruning was performed with using one input layer–one hidden layer–one output layer neural network architecture. The learning rate and momentum were set to 0.3 and 0.7, respectively. The network was not allowed to run more than 10,000 epochs since full network training requires

lots of computation time. In order to avoid the local minimum, the neuron sensitivity was obtained by averaging the sensitivity over 10 independent networks. The 5–5–1 (an input layer with 5 neurons, a hidden layer with 5 neurons, and an output layer with 1 neuron) network architecture was finally selected by starting from 18–18–1 architecture. The selected descriptors were the same as those of multiple linear regression analysis; the number of atoms, which can form hydrogen bonds (HA), connolly surface area (Area), the highest occupied molecular orbital energy (HOMO), partial charge of C<sub>3</sub> carbon atom (C<sub>3</sub>), and HOMO  $\pi$  coefficient of C<sub>2</sub> carbon atom (P<sub>2</sub>). The fact that the selected descriptors in NN are identical to those in MLR shows that these descriptors play an important role in the formation of antibody–tracer complex. The full back propagation neural network (NN) trainings were then performed with the 5–5–1 network architecture. The calculated output was obtained by averaging neural network prediction over 50 independent networks in order to avoid the local minimum. The experimental and calculated FP (difference of fluorescence polarization) values of training set in MLR and NN are shown in Table 3 in which the training set shows good correlations between experimental and calculated values in MLR and NN. The  $r^2$  values of MLR and NN in training set were 0.853 and 0.870, respectively (Table 4). In addition, the predictive  $r^2$  values of them in test set were 0.762 and 0.690, respectively. Little improvement in NN compared with MLR implies that the descriptors used in MLR have

**Table 3.** Experimental versus calculated FP values in multiple linear regression (MLR) and neural network (NN)

BTEX tracers	FP <sup>a</sup> (exp)	MLR		NN	
		FP (calcd)	Residual	FP (calcd)	Residual
[Training set]					
Benzoic acid	2.5	2.1	−0.4	2.2	−0.3
4-Phenylbutyric acid	14.5	12.9	−1.6	13.8	−0.7
6-Phenylhexanoic acid	19.9	18.8	−1.1	17.9	−2.0
<i>o</i> -Toluic acid	4.0	4.8	0.8	5.0	1.1
<i>o</i> -Tolyacetic acid	2.8	1.4	−1.4	1.8	−1.0
<i>m</i> -Tolyacetic acid	4.4	5.4	1.0	5.1	0.7
<i>p</i> -Tolyacetic acid	4.4	4.1	−0.3	4.0	−0.4
<i>o</i> -Tolypropionic acid	4.4	5.8	1.4	4.7	0.3
<i>p</i> -Tolypropionic acid	3.7	6.1	2.4	5.3	1.6
3,4-Dimethylbenzoic acid	7.1	5.6	−1.5	5.7	−1.4
3,4-Dimethylphenylthioacetic acid	6.3	4.8	−1.5	4.3	−2.0
3,4-Dimethylanilide succinic acid	5.4	8.0	2.6	7.3	1.9
4-Tolyidide succinic acid	4.4	2.3	−2.1	2.0	−2.4
3,4-Dichloranilide succinic acid	5.9	8.9	3.0	8.7	2.8
3,4-Dichlorophenoxyacetic acid	9.2	8.9	−0.3	9.2	−0.0
2,4-Dichlorophenoxyacetic acid	8.3	6.3	−2.0	5.8	−2.5
2,4,5-Trichlorophenoxyacetic acid	10.5	9.3	−1.2	9.2	−1.3
Ferroceneacetic acid	3.0	5.1	2.0	5.0	2.1
[Test set]					
3-Phenylpropionic acid	8.7	8.3	−0.4	7.6	−1.1
<i>m</i> -Toluic acid	4.4	3.8	−0.6	3.5	−0.9
<i>p</i> -Toluic acid	4.7	5.5	0.8	5.4	0.7
3,4-Dimethylpheoxyacetic acid	5.8	3.7	−2.1	3.4	−2.4
3,4-Dichlorophenylacetic acid	10.5	9.4	−1.1	9.8	−0.7

<sup>a</sup> FP = (fluorescence polarization of unbound BTEX tracers–fluorescence polarization of bound tracers)/10.

**Table 4.** Statistics of multiple linear regression (MLR) and neural network (NN)

	Training set		Test set	
	$r^{2a}$	$s^b$	$r^2$	$s$
MLR	0.853	1.679	0.762	1.163
NN	0.870	0.580	0.690	1.328

<sup>a</sup>  $r^2$ :  $r$  square.

<sup>b</sup>  $s$ : standard error.

good linear relationships. When the linearity of MLR descriptors is good, the improvement in NN is small. It is interesting that the selected descriptors in NN are identical to those in MLR, which suggests that these descriptors play a significant role in the formation of antibody–tracer complex. The neural network models using the smaller number of adjustable weights showed the loss of predictivity, for example, the 5–4–1 network architecture gave the  $r^2$  value of 0.843 in training set and predictive  $r^2$  value of 0.335 in test set.

Analysis of weight values gives good insight for considering the contribution of input descriptors. In this work, the weight values of NN were compared with the regression coefficients of MLR. For direct comparison of weight values with regression coefficients, the MLR was carried out by using the descriptor values, which were rescaled to have values between ca. 0.1 and 1.0. The regression coefficients of HA, Area, HOMO, C<sub>3</sub> and P<sub>2</sub> were −0.495, 0.977, −0.935, 0.262, and 0.350, respectively. The total weights of descriptors were calculated by the following equation

$$W_{\text{tot}} = \sum_i W_{ij} \cdot W_{jk} \quad (2)$$

where  $W_{\text{tot}}$  is the total weight of descriptor  $i$ ,  $W_{ij}$  is the weight between input and hidden units, and  $W_{jk}$  is the weight between hidden and output units. The  $W_{\text{tot}}$ s of HA, Area, HOMO, C<sub>3</sub> and P<sub>2</sub> were −11.176, 20.053, −20.245, 6.915, and 8.772, respectively. The signs of  $W_{\text{tot}}$ s and regression coefficients were identical. The  $W_{\text{tot}}$ s and regression coefficients of HA and HOMO were negative and those of Area, C<sub>3</sub>, and P<sub>2</sub> were positive, which shows good agreement between  $W_{\text{tot}}$ s and regression coefficients.

### 3. Conclusion

In this study multiple regression (MLR) analysis and back propagation neural networks (NN) were carried out in order to obtain the information about the formation of antibody–BTEX–EDF complex. The MLR and NN gave good QSAR results ( $r^2$  value of 0.853 and 0.870, respectively). It is interesting that the selected descriptors in NN are identical to those in MLR, which suggests that these descriptors have good linear relationships and play a significant role in the formation of antibody–tracer complex. Although the four BTEX tracers (2,3-dimethylbenzoic acid, 2,4-dimethylbenzoic acid, 2,5-dimethylbenzoic acid, and 3,5-dimethylbenzoic acid) in Table 1 were not included in analyses because of experimental errors, the errors seem to be included in other data. It is thought that the errors included in the experimental data prevented the development of a better

model even if QSPR studies were based on a series of structurally similar compounds.

## 4. Experimental

### 4.1. Molecular modeling

The data on 27 fluorescence polarizations of BTEX tracers<sup>10</sup> were used for this study (Table 1). All BTEX derivative structures in this work were constructed by using InsightII package (Accelrys Inc.). All the rotatable bonds were searched from 0° to 360° in 60° increments in order to obtain low energy structures. The lowest energy conformer for a given molecule was minimized using va09a minimizer until maximum energy derivatives are less than 0.001 kcal/(mol Å). The minimized structures were then fully geometry optimized using AM1 Hamiltonian in MOPAC software package. The semiempirically derived low energy structures were used to compute the values of QSPR descriptors. The following set of descriptors was used in multiple linear regression; (1) number of atoms, (2) molecular weight, (3) volume, (4) connolly surface area, (5) the largest principal moment of inertia, (6) dipole moment, (7) the highest occupied molecular orbital energy, HOMO, (8) the lowest unoccupied molecular orbital energy, LUMO, (9) number of atoms, which can form hydrogen bonds, (10) partial charges of C<sub>2</sub>, C<sub>3</sub>, C<sub>4</sub>, and C<sub>5</sub> carbon atoms of aromatic ring, (11) HOMO  $\pi$  coefficients of C<sub>2</sub>, C<sub>3</sub>, C<sub>4</sub>, and C<sub>5</sub> carbon atoms of aromatic ring, and (12) number of linker atoms. The descriptors used in this study are conventional ones, which can be enough to explain the molecular properties of BTEX tracers. Descriptors (1)–(5) are the properties, which are dependent on the size and shape of compounds and descriptors (6)–(8) are electronic properties. Descriptor (9) means hydrogen-bonding capacity of compounds. In addition, descriptors (10)–(12) were used since the substituents of aromatic ring and the linker atoms might have an effect on the fluorescence polarization of BTEX tracers.

### 4.2. Multiple linear regression analysis

Multiple linear regression analyses were carried out by using our laboratory-made program.<sup>11</sup> In order to ensure the reliability of our model, data set was divided into training and test sets. Several criteria have been used for selection of the best training set. One of the most popular criteria is the maximization of  $|X'X|$ . In this work, the stepwise addition method was applied to D-optimal design of Mitchell<sup>11,14</sup> for selection of training set. Prior to the D-optimal design, the cross-validation was performed in order to find the optimum number of descriptors. The all possible combinations of the optimum number of descriptors were considered in order to achieve the best regression model. The multicollinearity among variables was identified using variance inflation factor (VIF).<sup>15</sup> The VIF for the  $i$ th regression coefficient is expressed as

$$\text{VIF} = \frac{1}{1 - r_i^2} \quad (3)$$

where  $r_i$  is the coefficient produced by regressing the descriptor  $x_i$  against the other descriptors,  $x_j$  ( $j \neq i$ ). The models of which VIF is greater than 10 were not considered.

### 4.3. Neural networks

Back propagation neural networks were carried out by using our laboratory-made program.<sup>11</sup> The descriptors for neural networks were selected by pruning method as follows.<sup>11–13</sup> The importance of neurons in hidden or input layers was estimated according to sensitivity.

$$S_i = \Sigma \left( \frac{w_{ji}}{\max_a |w_{ja}|} \right)^2 \cdot S_j \quad (4)$$

where  $\max_a$  is the maximum weight of all weights ending a neuron  $j$  and  $S_j$  is a sensitivity in the upper layer. The neuron having the greatest value  $S_i$  gives the most important influence on all other neurons in the next layer. So the sensitivities of neurons in input and hidden layers were calculated and the neurons with low sensitivities were deleted. The pruning method was carried out iteratively to achieve a given number of neurons, which have small mean square error (MSE) according to the following procedure: (1) choose a large size network and determine the number of neurons to be pruned per step, (2) train the network, (3) compute neuron sensitivity for every N-epochs, (4) delete the neurons with low sensitivity, and (5) if stopping criterion is not met, go to step (2). The sensitivities of output layer neurons are set to 1. All the sensitivities in a layer are normalized to a maximum value of 1.

## Acknowledgements

This work was supported by National Research Laboratory program of the Ministry of Science & Technology, Korea.

## References and notes

- Kim, S. Y.; Lee, N. T.; Choi, M. J. *Bull. Korean Chem. Soc.* **2001**, *22*, 953.
- Snyder, R.; Witz, G.; Goldstein, B. D. *Environ. Health Perspect.* **1993**, *100*, 293.
- Odermatt, J. R. *Org. Geochem.* **1994**, *21*, 1141.
- Wittkamp, B. L.; Tilotta, D. C. *Anal. Chem.* **1995**, *67*, 600.
- Wang, Z.; Fingas, M. J. *Chromatogr. A* **1995**, *712*, 321.
- Shinar, R.; Liu, G.; Porter, M. D. *Anal. Chem.* **2000**, *72*, 5981.
- Fang, M.; Vetelino, K.; Rothery, M.; Hines, J.; Frye, G. C. *Sens. Actuators, B* **1999**, *56*, 155.
- Ueno, Y.; Horiuchi, T.; Norimoto, T.; Niwa, O. *Anal. Chem.* **2001**, *73*, 4688.
- Ueno, Y.; Horiuchi, T.; Niwa, O. *Anal. Chem.* **2002**, *74*, 1712.

10. Hong, J. Y.; Eremin, S. A.; Knopp, D.; Niessner, R.; Park, S. J.; Choi, M. J. *Anal. Bioanal. Chem.*, in press.
11. Moon, T.; Chi, M. H.; Kim, D. H.; Yoon, C. N.; Choi, Y. S. *Quant. Struct.-Act. Relat.* **2000**, *19*, 257.
12. Tetko, I. V.; Livingstone, D. J.; Luik, A. I. J. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 826.
13. Tetko, I. V.; Villa, A. E. P.; Livingstone, D. J. J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 794.
14. Mitchell, T. J. *Technometrics* **1974**, *16*, 203.
15. Myers, R. H. *Classical and Modern Regression with Applications*; PWS/KENT: Boston, 1990.