Communications to the Editor

[Chem. Pharm. Bull. 33(9)4081—4084(1985)]

AQUATIC TOXICITY: A PREDICTIVE STRUCTURE-ACTIVITY MODEL

Katsuichiro Komatsu, Shuichi Hirono, and Ikuo Moriguchi*
School of Pharmaceutical Sciences, Kitasato University,
Shirokane, Minato-ku, Tokyo 108, Japan

A quantitative structure-toxicity model has been developed for the estimation of aquatic toxicity ratings for a wide variety of organic chemicals using the adaptive least-squares (ALS 81) method. The discriminant function is based on substructural fragments obtained for 280 compounds. The data base used for the derivation of the function was obtained from the Registry of Toxic Effects of Chemical Substances of U.S. Department of Health and Human Services. The accuracy of classifying the chemicals into three toxicity ratings was 86% in the recognition and 82% in the leave-one-out prediction. All misclassified compounds were assigned to the neighboring ratings in both recognition and prediction.

KEYWORDS———structure-toxicity model; aquatic toxicity; organic chemical; adaptive least-squares; quantitative structure-activity relationship; pattern recognition; substructural descriptor; discriminant function; activity rating

The aquatic toxicity of chemical substances is extremely important information contributing to a better understanding of potential occupational hazards and helping to bring about a more healthful environment. Furthermore, aquatic toxicity is considered to be a rough measure of general toxicity for higher animals and humans.

Major efforts are now being devoted to structure-toxicity correlations to examine the possibility of predicting toxicity on the basis of the chemical constitution of substances.[1] However, one of the most serious difficulties confronting toxicologists is the quantification of the biological response,[2] and toxic activity has often been described in crude expressions such as toxicity ratings. This is due in part to the fact that toxic measurements involve the results of effects on multiple biological receptors. In addition, any test endpoint as complex as mortality may involve various combinations of toxic manifestations, and different aquatic species usually respond differentially to the same toxicant.

To correlate activity ratings with structural features of chemicals, we have developed the adaptive least-squares (ALS) method[3] which is a nonparametric pattern classifier. The present version of ALS is ALS 81.[4] In the present communication, a quantitative structure-toxicity model has been developed for the estimation of aquatic toxicity for a wide variety of organic chemicals by the use of ALS 81.

The data base used for the derivation of the discriminant function was obtained from the Registry of Toxic Effects of Chemical Substances of U.S. Department of

Table I. Toxicity Discriminant Function[a)]

No.	Descriptor	Coef.	CI[b)]
1	Number of benzene rings	0.639	0.268
2	Number of Cl atoms	0.222	0.239
3	IV[c)] for -CN (except nitrile)	2.009	0.238
4	IV for $(-O-)_2P(=X)-X-$ (X:O,S)	1.858	0.221
5	IV for naphthalene ring	1.437	0.208
6	IV for PhX_2 (X:halogen)	1.145	0.179
7	IV for $CH_2=CHX$ (X:CHO,CH_2OH,COOH,CN)	1.452	0.172
8	IV for -CON≺	1.283	0.170
9	IV for $HOCH_2CH_2OCH_2CH_2OR$ (R:alkyl)	1.491	0.154
10	IV for m-X-PhOH (X:electron-attracting group)	0.799	0.106
11	IV for $CH_3COOC_nH_{2n+1}$ ($n \geqq 4$)	0.677	0.090
12	IV for Cl_3C-R (R:electron-attracting group)	0.654	0.078
13	IV for 1,3-butadiene moiety	0.626	0.074
14	IV for ethereal -O-	-0.607	0.176
15	IV for alkane or alkene	-0.572	0.172
16	IV for $X_1CH=CX_2X_3$ ($X_1 \sim X_3$:H,Cl)	-1.207	0.160
17	IV for XCH_2CH_2X' (X,X':H,Cl,OH)	-1.062	0.141
18	IV for ≻N-C(=X)-N≺ (X:O,S)	-1.184	0.140
19	IV for R-CHO (R:alkyl having even number of carbon atoms in the main chain)	-0.965	0.128
20	IV for $X_1X_2CX_3X_4$ ($X_1 \sim X_4$:H,halogen)	-0.729	0.121
21	IV for -COO- (ester)	-0.359	0.119
22	IV for polycarboxylate	-0.906	0.108
23	IV for $\succ NCH_2CH_2\overset{\prime}{N}CH_2CH_2N\prec$	-0.882	0.105
24	IV for nitrile	-0.597	0.099
25	IV for -COOH	-0.331	0.070
26	IV for aliphatic OH	-0.135	0.053
27	IV for ≻C=O	-0.117	0.026
Constant		-0.013	

a) $L=\Sigma$(coef. × descriptor) + constant; class I, $L < -1.225$; class II, $-1.225 \leqq L < 0.121$; and class III, $L \geqq 0.121$. These cutting points are determined on the basis of the size of classes used for the derivation of the discriminant function.[3b)] b) Contribution index (=|coef.| × S.D. of descriptor), which is a measure of the contribution of the descriptor to the discriminant score (L).[3b)] c) Indicator variable which is assigned the value of 1 for the presence of the given substructure and 0 for its absence.

Health and Human Services.[5)] The substances used for the ALS calculation were 280 organic chemicals including 45 aliphatic hydrocarbons and halides, 139 other aliphatic compounds, 16 alicyclic compounds, 63 benzene and naphthalene derivatives, 6 heterocyclic compounds, and 11 other chemicals.

The aquatic toxicity ratings[5)] are defined on the basis of TLm 96 which is the

concentration that will kill 50% of the exposed organisms (early stages of fish and amphibians) within 96 hours. In this study, a three-class discrimination with the following classes was performed: class I, 143 not or weakly toxic compounds (TLm 96 > 100 ppm); class II, 97 moderately toxic compounds (10 ppm < TLm 96 ≦ 100 ppm); and class III, 40 strongly toxic compounds (TLm 96 ≦ 10 ppm).

In all, 60 structural features were investigated to describe the molecules. They include four physicochemical descriptors (log P,[6] $(\log P)^2$, V_w,[7] and V_w^2), seven numerical descriptors (numbers of F, Cl, aromatic Cl, Br, phenyl, Ph-NH_2, and double bonds), and 49 indicator variables for the presence of various substructural fragments and moieties. Forty out of the 60 descriptors were preliminarily chosen by means of backward stepwise elimination, and then the several subsets of descriptors were selected using every-possible-combination technique. The final subset of descriptors was determined on the basis of the results of leave-one-out prediction.[8]

In Table I, we show the discriminant function thus obtained. The function includes two numerical descriptors and 25 indicator variables for substructural fragments and moieties. Note that physicochemical descriptors disappeared from the final subset. The factors of hydrophobicity and steric effects may be implied in the 27 substructural descriptors. A descriptor with a positive coefficient is considered to contribute in a positive sense to an estimate of toxicity, while a descriptor with a negative coefficient contributes in a negative way. However, these coefficients cannot be used to make inferences about the contribution of each fragment. They are valid only when used in the context of this multidimensional model. The primary purpose of the function is to provide for a method of classification of toxicity ratings, not to elucidate cause-effect relationships.

The result of discrimination of the toxicity ratings for the compounds used in the design of the model is fairly satisfactory, as shown in Table II. The accuracy of classification into three classes was 86% in the recognition, and 82% in the leave-one-out prediction.[8] All misclassified compounds were assigned to the neighboring ratings; no compounds were strongly misclassified in both recognition and prediction, in spite of the diversity of the molecular structure of organic chemicals investigated in this study.

Table II. Result of Toxicity Discrimination (n=280, 3 classes)

Recognition

N_{mis}[a]=39(0), 86.1% success, R_S[b]=0.839

	Calcd.		
Obsd.	I	II	III
I	127	16	0
II	14	81	2
III	0	7	33

Leave-one-out prediction

N_{mis}=51(0), 81.8% success, R_S=0.803

	Calcd.		
Obsd.	I	II	III
I	126	17	0
II	17	77	3
III	0	14	26

a) Number of misclassified compounds. The figure in parentheses is the number of compounds misclassified by two grades. b) Spearman rank correlation coefficient.

It is evident from these results that a reasonably accurate discriminant func-

tion can be designed for the estimation of aquatic toxicity ratings. It will be desirable to test the discriminant function against independently evaluated compounds. As new data become available, it will be important to develop this model further in order to reflect the latest state of the art.

REFERENCES AND NOTES

1) L. Golberg (ed.), "Structure-Activity Correlation as a Predictive Tool in Toxicology," Hemisphere Publ. Corp., Washington, 1983.
2) R. Dagani, Chem. Eng. News, 59, No. 10, 26 (1981).
3) a) I. Moriguchi and K. Komatsu, Chem. Pharm. Bull., 25, 2800, 3440 (errata) (1977); b) I. Moriguchi, K. Komatsu, and Y. Matsushita, J. Med. Chem., 23, 20 (1980).
4) I. Moriguchi and K. Komatsu, Abstract of Papers, 8th Symposium on Structure-Activity Relationships, Tokyo, Oct. 1981, p.55; K. Akahane, D. Momose, K. Iizuka, T. Miyamoto, M. Hayashi, K. Iwase, and I. Moriguchi, Eur. J. Med. Chem., 19, 85 (1984).
5) R.L. Tatken and R.J. Lewis, Sr. (eds.), "Registry of Toxic Effects of Chemical Substances 1981-82 Edition," U.S. Department of Health and Human Services, National Institute for Occupational Safety and Health, Cincinnati, 1983.
6) Hydrophobic constant calculated by the method of Hansch and Leo; C. Hansch and A. Leo, "Substituent Constants for Correlation Analysis in Chemistry and Biology," John Wiley and Sons, New York, 1979.
7) van der Waals volume calculated by the method of Moriguchi; I. Moriguchi, Y. Kanada, and K. Komatsu, Chem. Pharm. Bull., 24, 1799 (1976); I. Moriguchi and Y. Kanada, Chem. Pharm. Bull., 25, 926 (1977).
8) A.J. Stuper and P.C. Jurs, J. Am. Chem. Soc., 97, 182 (1975).

(Received June 27, 1985)