Faculty of Pharmaceutical Sciences, Kyushu University Fukuoka 812, Japan

SADAO IGUCHI
TSUYOSHI GOROMARU
ATSUKO NODA
KENJI MATSUYAMA
KEIZO SOGABE

Received July 30, 1977

[Chem. Pharm. Bull. 25(10)2800—2802(1977)]

UDC 615.277.3.011.5.015.11 : 541.69

Adaptive Least-Squares Classification applied to Structure-Activity Correlation of Antitumor Mitomycin Derivatives

An adaptive least-squares (ALS) classification which is capable of relating structure to activity rating of chemical compounds has been developed. The ALS method makes decisions for multicategory pattern classification by a single discriminant function. For the set of 16 mitomycin derivatives belonging to five activity classes used in this study, the present method is shown to be efficacious compared to other classification procedures such as the simple least-squares method, the Rao-type discriminant analysis, and the *K*-nearest neighbor method.

Keywords——structure-activity relationships; pattern recognition; nonparametric linear classifiers; mitomycin; cancer chemotherapy; drug design

The biological potency of drugs has been often represented in a form of activity rating. To such a discrete type of activity data, usual QSAR (quantitative structure-activity relationship) techniques such as the Hansch analysis[1] and the Free-Wilson approach[2] are not applicable. For the purpose of relating structure to activity class, a new classification method utilizing adaptive least-squares (ALS) technique was developed.

The ALS method makes decisions for multicategory pattern classification by a single discriminant function represented by a dot product as

$$L = w_0 + w_1x_1 + w_2x_2 + \cdots + w_mx_m = \boldsymbol{W}\cdot\boldsymbol{X} \qquad (1)$$

where $\boldsymbol{X}$ is the pattern vector $(1,x_1,x_2,\ldots,x_m)$ for m structural features of molecules, and $\boldsymbol{W}$ the weight vector $(w_0,w_1,w_2,\ldots,w_m)$. The discrimination lines between classes are fixed in advance as follows: If L_i $(=\boldsymbol{W}\cdot\boldsymbol{X}_i$; $\boldsymbol{X}_i$ is the pattern vector for the ith molecule$)\leqq 1$, then assign the ith molecule to class 1; if $1<L_i\leqq 2$, then assign to class 2; if $2<L_i\leqq 3$, then assign to class 3;....and if $L_i>k-1$, then assign to class k for classification into k classes.

The weight vector $\boldsymbol{W}$ is estimated by ALS calculation. A parameter $S^{(j)}$ described below is assumed instead of L as $S_i^{(j)}=\boldsymbol{W}\cdot\boldsymbol{X}_i$ $(i=1,2,\ldots,n)$ for the set of n compounds, and the least-squares estimate[3] of $\boldsymbol{W}$ is calculated to be $\boldsymbol{W}^{(j)}$. The superscript (j) denotes the number of iteration times. Then, $L_i^{(j)}$ $(=\boldsymbol{W}^{(j)}\cdot\boldsymbol{X}_i)$ is computed and used for classification.

$S^{(1)}$ is given by

$$S_i^{(1)} = N_i - 0.5 \qquad i = 1, 2, \cdots, n \qquad (2)$$

where N_i is the number in ascending order of activity class to which the ith molecule was observed to belong. And $S^{(j+1)}$ $(j\geqq 1)$ is adapted as

1) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
2) S.M. Free and J.W. Wilson, *J. Med. Chem.*, **7**, 395 (1964).
3) B.W. Bolch and C.J. Huang, "Multivariate Statistical Methods for Business and Economics, "Prentice-Hall Inc., Englewood Cliffs, 1974, Chapter 4.

$$S_i^{(j+1)} = \begin{cases} L_i^{(j)} & \text{if correctly classified} \\ L_i^{(j)} - C_i^{(j)} & \text{if misclassified} \end{cases} \qquad i = 1, 2, \cdots, n \qquad (3)$$

where $C_i^{(j)}$ is the correction term. For $C_i^{(j)}$ Eq. 4 was used in this work.

$$C_i^{(j)} = \pm 0.1 (L_i^{(j)} - N_i)^{\alpha} \qquad i = 1, 2, \cdots, n \qquad (4)$$

The sign for $C_i^{(j)}$ was chosen to correspond with that for $(L_i^{(j)} - N_i)$, and $\alpha=2$ or 3 for efficient adaptation.

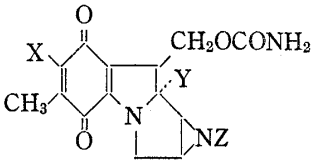
Similar results were obtained by the negative-feedback adaptation of $S^{(j+1)}$ using Eq. 3′ instead of Eq. 3.

$$S_i^{(j+1)} = \begin{cases} S_i^{(j)} & \text{if correctly classified} \\ S_i^{(j)} - C_i^{(j)} & \text{if misclassified} \end{cases} \qquad i = 1, 2, \cdots, n \qquad (3')$$

These adaptive calculations are repeated until the number of compounds misclassified becomes zero or smallest within a given iteration times.

The ALS method was successfully applied to structure-activity correlation of 16 mitomycin derivatives.[4] Five descriptors for the molecular features were used: field constant,[5] F_X, and van der Waals volume,[6] V_{W-X}, for substituent X, steric substituent constant,[7] E_{S-Z}, for substituent Z, and two descriptors Y_{OMe} and Y_{OH}, accounting for substituent Y=OMe and Y=OH, respectively. These descriptors are listed along with observed and calculated activity ratings in Table I.

Table I. Structural Features and Antitumor Activity of Mitomycin Derivatives



No.	X	Y	Z	F_X[a]	V_{W-X}[b]	Y_{OMe}[c]	Y_{OH}[d]	E_{S-Z}[e]	Activity[f] Obsd[g]	Activity[f] Calcd[h]
1[i]	NH_2	OMe	H	0.02	0.177	1	0	1.24	3+	3+
2	NHEt	OMe	H	−0.11	0.493	1	0	1.24	3+	3+
3[j]	NH_2	OMe	Me	0.02	0.177	1	0	0	2+	2+
4	NH_2	OMe	Et	0.02	0.177	1	0	−0.07	2+	2+
5	NH_2	OMe	Ac	0.02	0.177	1	0	−0.47[m]	2+	2+
6	NH_2	OH	Me	0.02	0.177	0	1	0	2+	2+
7	NMe_2	OMe	H	0.10	0.441	1	0	1.24	2+	2+
8	NH_2	OMe	COPh-*o*-Cl	0.02	0.177	1	0	−1.19[n]	+	+
9	NH_2	OMe	COPh-*p*-Cl	0.02	0.177	1	0	−1.19[n]	+	+
10	NHPh	OMe	H	−0.02	0.892	1	0	1.24	+	+
11[k]	OMe	OMe	H	0.26	0.304	1	0	1.24	+	+
12	OMe	OMe	Me	0.26	0.304	1	0	0	+	+
13[l]	OMe	OH	Me	0.26	0.304	0	1	0	±	±
14	NH_2	H	Me	0.02	0.177	0	0	0	−	−
15	NH_2	OMe	SO_2Me	0.02	0.177	1	0	−1.54[o]	−	+
16	OMe	H	Me	0.26	0.304	0	0	0	−	−

a) Reference 5. *b*) Reference 6. *c*) $Y_{OMe}=1$ for Y=OMe. *d*) $Y_{OH}=1$ for Y=OH. *e*) Reference 7. *f*) Against solid sarcoma 180. *g*) Reference 4. *h*) Using Eq. 5. *i*) Mitomycin C. *j*) Porfiromycin. *k*) Mitomycin A. *l*) Mitomycin B. *m*) The value for iso-Pr was used. *n*) The value for Ph(Me)CH was used. *o*) The value for *t*-Bu was used.

4) S. Kinoshita, K. Uzu, K. Nakano, M. Shimizu, T. Takahashi, and M. Matsui, *J. Med. Chem.*, **14**, 103 (1971).
5) C. Hansch, A. Leo, S.H. Unger, K.H. Kim, D. Nikaitani, and E.J. Lien, *J. Med. Chem.*, **16**, 1207 (1973).
6) I. Moriguchi, Y. Kanada, and K. Komatsu, *Chem. Pharm. Bull.* (Tokyo), **24**, 1799 (1976).
7) R.W. Taft, Jr., "Steric Effects in Organic Chemistry," ed. by M.S. Newman, John Wiley and Sons, New York, 1956, p. 598.

The best discriminant function formulated within 100 times of iterative calculation was Eq. 5 which was obtained after 10 times of adaptation using Eq. 3 and 4 with $\alpha=2$.

$$L = -4.33 F_X - 2.46 V_{W-X} + 0.77 E_{S-Z} + 2.48 Y_{OMe} + 2.28 Y_{OH} + 1.42$$
$$n = 16, \quad \% \text{ correct} = 93.8 \qquad (5)$$

Goodness of fit for this classification was highly significant at the 0.001 level by the χ^2 test.[8] Equation 5 indicates that electron-donating substituent X, and OMe or OH (OMe being a slightly better) for substituent Y enhance the activity against solid sarcoma 180, whereas bulky substituents X and Z causing large steric hindrance reduce the effectiveness.

Besides the ALS method, three classification procedures were tested with the same structure-activity data for the purpose of comparison: the simple least-squares method[9] (5), the Rao-type discriminant analysis[10] (4), and the K-nearest neighbor method[11] with $K=1$ (7). The figures in the parentheses after the methods were the number of molecules misclassified. The best results were obtained by the ALS method: Only one compound was uncorrectly assigned.

In conclusion, the results of this study show that the ALS classification is effective for relating structure to activity rating of compounds. It is suggested that the ALS method can aid significantly in rational design of more potent drugs for cancer chemotherapy and other categories.

School of Pharmaceutical Sciences,
Kitasato University
Shirokane, Minato-ku, Tokyo, 108, Japan

Ikuo Moriguchi
Katsuichiro Komatsu

Received August 9, 1977

8) R.C. Campbell, "Statistics for Biologists," Cambridge University Press, Cambridge, 1974, Chapter 4.
9) B.R. Kowalski, P.C. Jurs, and T.L. Isenhour, *Anal. Chem.*, **41**, 695 (1969).
10) Y.C. Martin, J.B. Holland, C.H. Jarboe, and N. Plotnikoff, *J. Med. Chem.*, **17**, 409 (1974).
11) B.R. Kowalski and C.F. Bender, *Anal. Chem.*, **44**, 1405 (1972).

[Chem. Pharm. Bull. 25(10)2802—2805(1977)]

UDC 547.913.02 : 581.192

Structure of Stephanthraniline A

A new polyoxypregnane derivative, stephanthraniline A, was isolated from the stem of *Stephanotis japonica* Makino (Asclepiadaceae), and its structure was determined.

Keywords——*Stephanotis japonica* Makino; Asclepiadaceae; polyoxypregnane; stephanthraniline A; sarcostin; ester-linkage; ^{13}C-NMR

In an earlier investigation of constituents of *Stephanotis japonica* Makino (Asclepiadaceae), C/D-*cis*-polyoxypregnane, sarcostin, lineolon, deacylmetaplexigenin, and stephanol were isolated.[1] In this communication, we wish to describe the isolation and structure of a new compound from the same source. The aglycone mixture, obtained after a mild acid hydrolysis of the crude glycoside, was separated by silica gel column chromatography and preparative

1) M. Fukuoka and H. Mitsuhashi, *Chem. Pharm. Bull.* (Tokyo), **16**, 553 (1968); M. Fukuoka and H. Mitsuhashi, *ibid.*, **17**, 2248 (1968).