Fuzzy Adaptive Least Squares and Its Use in Quantitative Structure–Activity Relationships

Ikuo MORIGUCHI,* Shuichi HIRONO, Qian LIU, Yasuo MATSUSHITA, and Takako NAKAGAWA

School of Pharmaceutical Sciences, Kitasato University, 5–9–1 Shirokane, Minato-ku, Tokyo 108, Japan. *Received July 10, 1990*

Fuzzy adaptive least squares (FALS), a pattern recognition method designed to correlate molecular structure with activity rating, has been developed. A novel feature of FALS is that the degree to which each sample belongs to an activity class is given using a membership function. The algorithm involves an iterative modification of forcing factors to maximize the sum of the membership function values over all samples. This paper first describes the method and calculation procedure of FALS89 (1989 version of FALS), and then shows its application to the correlation of structure with a potency rating of anticarcinogenic mitomycin derivatives and arginine-vasopressin antagonists. FALS89 applied to these samples showed considerably high reliability in both recognition and leave-one-out prediction.

Keywords fuzzy set; adaptive least squares; activity rating discrimination; pattern recognition: QSAR; mitomycin; vasopressin antagonist

Introduction

It is well known that structurally similar compounds exhibit similar biological activities. Discriminating between these similarities is a problem of pattern recognition. There are two aspects involved in pattern discrimination for structure–activity studies. One is the discrimination of the type of action from molecular structure. For this purpose, methods for independent-category discrimination such as linear discriminant analysis,[1] SIMCA,[2] and linear learning machine (LLM)[3] are used. The other is the discrimination of activity ratings (−, +, ++, *etc.*), which are ordered categories. For this purpose, we developed adaptive least squares (ALS) in 1977.[4] ALS is a nonparametric pattern classifier, and is devised to formulate a quantitative structure–activity relationships (QSAR) in a single mathematical equation, irrespective of the number of activity ratings by an error-correcting feedback adaptation. Because the adaptation is done on the forcing factors described later, the ALS calculation is efficient, and is applicable to linearly inseparable samples, unlike LLM.

Ordered categories comprise not only statistical vagueness, such as inaccuracy of measurement, but also intrinsic vagueness, such as subjective criteria for classification. Such indefiniteness can be grasped by the concepts of fuzzy variance.[5] To ALS, therefore, we have introduced a membership function[5] which is assumed to be a fuzzy degree of membership in classes. The product is fuzzy adaptive least-squares (FALS).[6]

This paper first describes the method of the 1989 version of FALS (FALS89), and then shows its application to the correlation of structure with a potency rating of anticarcinogenic mitomycin derivatives and arginine-vasopressin antagonists.

FALS89 Like ALS, FALS makes decisions for ordered m-class ($m \geq 2$) discrimination by a single discriminant function as

$$Z = w_0 + w_1 x_1 + w_2 x_2 + \cdots w_p x_p \tag{1}$$

where $x_k = k$th descriptor ($k = 1, 2, \cdots, p$) for structure; w_k = weight coefficient; and Z = discriminant score. For a set of n compounds, Eq. 1 can be rewritten as Eq. 2.

$$\boldsymbol{Z} = \boldsymbol{XW} \tag{2}$$

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$$

In the matrix $\boldsymbol{X}$, x_{ki} ($k = 1, 2, \cdots, p$ and $i = 1, 2, \cdots, n$) is the kth descriptor for the ith compound.

Starting scores, a_j ($j = 1, 2, \cdots, m$), for the members of class j are assumed, and then class boundaries, b_j ($j = 1, 2, \cdots, m-1$), are fixed in advance. In fuzzy ALS, as well as ALS, a_j is assumed by Eq. 3 where a_j is dependent on the size of classes, or by Eq. 4 where a_j is only dependent on the number of classes. b_j is taken as the midpoint between a_j and a_{j+1} as Eq. 5.

$$a_j = 4\left(\sum_{g=1}^{j-1} n_g + n_j/2\right) \Big/ n - 2 \tag{3}$$

where n_g = size of group g and n_j = size of group j.

$$a_j = (4j-2)/m - 2 \tag{4}$$

$$b_j = (a_j + a_{j+1})/2 \tag{5}$$

A membership function, $M(Z)$, is assumed to give the grade of membership of classes for compounds. The value of $M(Z)$ (membership grade) ranges from 0 to 1, and is taken to be 0.5 at the class boundaries. Figure 1 shows the function used in FALS89. In Fig. 1, fuzzy level, Fl_j, is the parameter for fuzziness of the boundary between class j and class $j+1$. Two levels of slopes, 'steep' and 'gentle', are generally used. $M(Z)$ for class j can be written as Eq. 6.

$$M(Z) = \begin{cases} 1/[1 + \{(Z - b_{j-1})/Fl_{j-1} - 1\}^4] & Z \leq b_{j-1} + Fl_{j-1} \\ 1 & b_{j-1} + Fl_{j-1} < Z \leq b_j - Fl_j \\ 1/[1 + \{(b_j - Z)/Fl_j - 1\}^4] & b_j - Fl_j < Z \end{cases} \tag{6}$$

The procedure is designed to maximize approximately

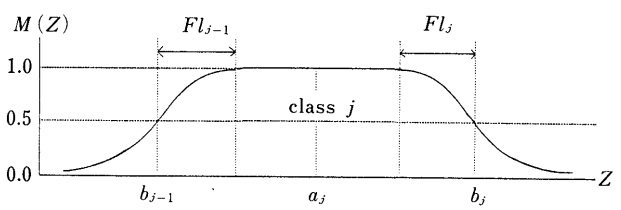


Fig. 1. Membership Function for Class j

the total sum of the membership grade over all compounds in the set. The calculation begins by setting the initial forcing factors $S_i^{(1)}$ ($i=1, 2, \cdots, n$), which are taken to be

$$S_i^{(1)}=a_j \tag{7}$$

where a_j=the starting score for class j to which the ith compound was observed to belong. Generally, classes are numbered in ascending order of biological potency. By use of $S_i^{(1)}$ in place of Z in Eq. 1 (or Eq. 2) as

$$\boldsymbol{S}^{(1)}=\boldsymbol{XW} \tag{8}$$

where $\boldsymbol{S}^{(1)}=(S_1^{(1)}, S_2^{(1)}, \cdots, S_n^{(1)})'$ (the prime denotes the transposition), the least-squares estimate of the initial weight, $\boldsymbol{W}^{(1)}$, is written as

$$\boldsymbol{W}^{(1)}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{S}^{(1)} \tag{9}$$

$\boldsymbol{W}^{(1)}$ is computed by ordinary least-squares. Then, $Z_i^{(1)}$ for each substance is calculated from Eq. 1 (or Eq. 2) using $\boldsymbol{W}^{(1)}$ as

$$\boldsymbol{Z}^{(1)}=\boldsymbol{XW}^{(1)} \tag{10}$$

The membership grade, MG_i, is calculated based on $Z_i^{(1)}$ for all the compounds.

At iteration 2 and thereafter, the forcing factor $S_i^{(t+1)}$ ($t\geq 1$) is adapted using the correction term $C_i^{(t)}$ when compound i actually belongs to class j, as

$$S_i^{(t+1)}=Z_i^{(t)}+C_i^{(t)} \tag{11}$$

$$C_i^{(t)}=\begin{cases}\alpha\sqrt{(1-MG_i)Fl_{j-1}} & Z_i^{(t)}\leq a_j \\ -\alpha\sqrt{(1-MG_i)Fl_j} & Z_i^{(t)}>a_j\end{cases} \tag{12}$$

In Eq. 12, α is the constant giving below.

Then, the least-squares estimate of $W_k^{(t+1)}$ is computed from Eq. 13, and $Z_i^{(t+1)}$ is calculated from Eq. 1 (or Eq. 2) using $\boldsymbol{W}^{(t+1)}$ for the evaluation of membership grade and classification.

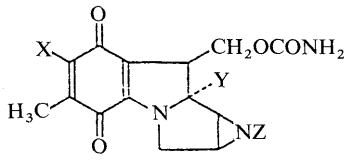
$$\boldsymbol{W}^{(t+1)}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{S}^{(t+1)} \tag{13}$$

The adaptive least-squares calculation is iteratively carried out so as to minimize $\Sigma(S_i-Z_i)^2$, or ΣC_i^2. Therefore, we can expect to obtain a discriminant function giving almost maximum MG_i for the set of compounds.

The iterative learning of the discriminant function is actually carried out in two steps. In step 1 ($1\leq t\leq 10$), a weight vector giving a good discrimination result is roughly searched using all combinations of fuzzy levels and a greater α value to avoid falling into a local optimum. Since there is some arbitrariness in the membership function, the product of mean membership grade (MMG) and Spearman's rank correlation coefficient (R_S) is used as the criterion for the best discrimination. Thus, the weight vector giving the greatest value of the product $MMG\cdot R_S$ in step 1 is selected for the starting vector of step 2. In step 2 ($t\geq 11$), the iterative calculation with the best combination of fuzzy levels chosen in step 1 and a smaller α value is performed until the discrimination is no longer improved within a maximum of 20 times of iteration.

The results of FALS89 are validated by leave-one-out prediction.[3] The discriminant function with a scientifically reasonable subset of descriptors giving the best leave-one-out prediction is finally adopted.

TABLE I. Structures and Descriptors of Mitomycin Derivatives



Compd.	Structure			Descriptors used in Eqs. 14 and 15				
	X	Y	Z	σ_m(X)	V_w(X)	σ^*(Y)	B_1(Z)	B_4(Z)
1[a]	NH_2	OMe	H	−0.16	0.177	1.81	1.00	1.00
2	NHEt	OMe	H	−0.24	0.493	1.81	1.00	1.00
3[b]	NH_2	OMe	Me	−0.16	0.177	1.81	1.52	2.04
4	NH_2	OMe	Et	−0.16	0.177	1.81	1.52	2.97
5	NH_2	OMe	Ac	−0.16	0.177	1.81	1.90	2.93
6	NH_2	OH	Me	−0.16	0.177	1.55	1.52	2.04
7	NMe_2	OMe	H	−0.15	0.441	1.81	1.00	1.00
8	NH_2	OMe	COPh-*o*-Cl	−0.16	0.177	1.81	2.36	5.98
9	NH_2	OMe	COPh-*p*-Cl	−0.16	0.177	1.81	2.36	5.98
10	NHPh	OMe	H	−0.12	0.892	1.81	1.00	1.00
11[c]	OMe	OMe	H	0.12	0.304	1.81	1.00	1.00
12	OMe	OMe	Me	0.12	0.304	1.81	1.52	2.04
13[d]	OMe	OH	Me	0.12	0.304	1.55	1.52	2.04
14	NH_2	H	Me	−0.16	0.177	0.49	1.52	2.04
15	NH_2	OMe	SO_2Me	−0.16	0.177	1.81	2.11	3.15
16	OMe	H	Me	0.12	0.304	0.49	1.52	2.04

a) Mitomycin C. *b*) Porfiromycin. *c*) Mitomycin A. *d*) Mitomycin B.

TABLE II. FALS89 Discriminant Functions and Their Reliability

Eq. No.		N[b]	Calcd			Predicted[a]		
			MMG[c]	N_{mis}[d]	R_s[e]	MMG[c]	N_{mis}[d]	R_s[e]
14[f]	Solid sarcoma 180 $Z=-5.654\,\sigma_m(X)-3.198\,V_w(X)+1.664\,\sigma^*(Y)$ [0.70][g] [0.59] [0.72] $-1.638\,B_1(Z)+0.579$ [0.74]	16	0.925	1 (0)	0.969	0.883	2 (0)	0.901
15[h]	Ascites Hirosaki sarcoma $Z=-1.911\,V_w(X)+1.729\,\sigma^*(Y)-0.551\,B_4(Z)$ [0.37][g] [0.79] [0.41] −0.876	14	0.859	2 (0)	0.876	0.857	2 (0)	0.876

a) Using leave-one-out technique. *b*) Number of compounds. *c*) Mean membership grade. *d*) Number misclassified. The figure in parentheses is the number misclassified by two ratings. *e*) Spearman rank correlation coefficient. All the values are highly significant at the level of $p<0.001$. *f*) Derived with fuzzy levels {0.5, 0.1}. *g*) Contribution index (=|coef|·SD of the descriptor), which is a measure of the contribution of the descriptor to discriminant score. *h*) Derived with fuzzy levels {0.1, 0.5}.

Methods

In FALS89 calculation, two kinds of fuzzy levels ($Fl=0.1$ for 'steep' and $Fl=0.5$ for 'gentle') were used for the membership function. The α value was taken to be 2.0 and 0.5 in step 1 and step 2, respectively. All calculations were carried out on a Sony NWS-830 computer and a Kobe Steel KTR-B08 transputer attached to an Epson PC-286VF microcomputer.

Results and Discussion

Structure Activity Correlation of Carcinogenic Mitomycin Derivatives Mitomycins are antitumor antibiotics discovered by HATA *et al.* of the Kitasato Institute.[7] The first example of the application of FALS89 is for mitomycin derivatives which have been well studied by ALS.[8] Their structures and descriptors are shown in Table I. Five descriptors for the structural features were the Hammett constant,[9] σ_m (X), and van der Waals volume,[10] V_w(X), for the substituent X, polar effect, σ^*(Y),[11] for Y, and Sterimol width parameters,[12] B_1(Z) and B_4(Z), for Z. The data set was the same as that used in the study by ALS.[8]

Anticarcinogenic activity in mice was observed against a solid sarcoma 180 with 16 compounds and against an ascites Hirosaki sarcoma with 14 compounds, and was allotted three ratings for both subsets (See Table IV).

FALS89 calculation was effectively done using Eq. 4 for the starting scores. The resultant discriminant functions and their reliability are listed in Table II. The four-descriptor equation (Eq. 14) and the three-descriptor equation (Eq. 15) were best for anticarcinogenic activity against the solid sarcoma 180 and ascites Hirosaki sarcoma, respectively. The squared correlation matrix for the descriptors included in Eqs. 14 and 15 appears in Table III. The combinations of descriptors were validated by the leave-one-out technique.

From the sign of the coefficients with these descriptors, it was concluded that electron-donating and less bulky substituents for X, electron-withdrawing groups for Y, and thin substituents for Z are favorable for higher potency against both solid sarcoma and ascites sarcoma. The findings seem to support the mechanism of action of mitomycin C postulated by Andrews *et al.*[13] The mechanism involves A) enzymatic one-electron reduction of the quinone moiety to a semiquinone radical followed by B) elimination of the methoxy group to form a reactive planner intermediate, C) opening of the aziridine ring to cross-link deoxyribonucleic acid (DNA), and D) oxidation of the semiquinone radical to cause DNA strand cleavage. The descriptor V_w(X) may be related to step A), σ^*(Y) to B), B_1(Z) and B_4(Z) to C), and σ_m(X) to D).

The calculated and predicted membership grades and ratings in both activities for each compound using Eqs. 14 and 15 are listed in Table IV, along with observed ratings. The results from ALS81 analysis[8] are also listed in Table IV. FALS89 and ALS81 gave the same calculated ratings for both activities. However, FALS89 was found to have better predictive ability of activity against solid sarcoma 180, despite the use of descriptor subsets selected by ALS81.

Structure Activity Correlation of Vasopressin Antagonists The second example of the application of FALS89 studied 33 antagonists of antidiuretic and vasopressor responses to arginine-vasopressin. Vasopressin acts on the membranes of the distal convoluted tubules and collecting ducts of the kidney, causing them to become water permeable. This permits reabsorption of water by osmosis. Vasopressin also acts as a vasoconstrictor of vascular

TABLE III. Squared Cross-Correlation Matrix of Descriptors

	σ_m(X)	V_w(X)	σ^*(Y)	B_1(Z)	B_4(Z)
σ_m(X)	1.000				
V_w(X)	0.001	1.000			
σ^*(Y)	0.056	0.029	1.000		
B_1(Z)	0.000	0.330	0.028	1.000	
B_4(Z)	0.004	0.346	0.011	0.898[a]	1.000

a) B_1(Z) and B_4(Z) were not used simultaneously in Eqs. 14 and 15.

TABLE IV. Observed and Calculated Activities of Mitomycin Derivatives

Compd.	Solid sarcoma 180							Ascites Hirosaki sarcoma						
		FALS89 (Eq. 14)				ALS81[a]			FALS89 (Eq. 15)				ALS81[a]	
	Obsd[a]	Calcd		Pred		Calcd	Pred	Obsd[a]	Calcd		Pred		Calcd	Pred
	RT[b]	*MG*	RT	*MG*	RT	RT	RT	RT	*MG*	RT	*MG*	RT	RT	RT
1	3	1.000	3	1.000	3	3	3	3	1.000	3	1.000	3	3	3
2	3	1.000	3	1.000	3	3	3	3	1.000	3	0.987	3	3	3
3	3	1.000	3	1.000	3	3	3	3	1.000	3	0.994	3	3	3
4	3	1.000	3	1.000	3	3	3	2	1.000	2	1.000	2	2	2
5	3	0.809	3	0.356	2	3	2	2	1.000	2	1.000	2	2	2
6	3	0.990	3	0.923	3	3	3	2	1.000	2	1.000	2	2	2
7	3	1.000	3	1.000	3	3	3	3	1.000	3	1.000	3	3	3
8	2	1.000	2	1.000	2	2	2							
9	2	1.000	2	1.000	2	2	2							
10	2	1.000	2	1.000	2	2	2	2	1.000	2	0.996	2	2	2
11	2	0.994	2	0.856	2	2	2	3	1.000	3	1.000	3	3	3
12	2	1.000	2	1.000	2	2	2	2	1.000	2	1.000	2	2	2
13	1	1.000	1	1.000	1	1	2	1	0.024	2	0.015	2	2	2
14	1	1.000	1	0.997	1	1	1	1	1.000	1	1.000	1	1	1
15	1	0.000	2	0.000	2	2	2	3	0.001	2	0.000	2	2	2
16	1	1.000	1	1.000	1	1	2	1	1.000	1	1.000	1	1	1

a) Ref. 8. b) Rating.

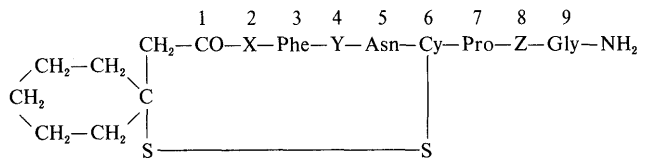
TABLE V. Structures and Descriptors of Arg-Vasopressin Antagonists

Compd.	Structure			Descriptors used in Eqs. 16—21 (Table VI)							
	X	Y	Z	*D*(X)	*Phe*(X)	*L*(X)	*HB*(Y)	*H*(Y)	*E*(Y)	*T*(Y)	*OMH*(Y)	*D*(Z)
1	D-Phe	Val	L-Arg	1	1	2.06	0	0	0.68	−0.62	0.91	0
2	D-Phe	Ile	L-Arg	1	1	2.06	0	0	0.67	−0.70	1.25	0
3	D-Phe	Thr	L-Arg	1	1	2.06	1	0	0.13	0.06	−0.28	0
4	D-Phe	Ala	L-Arg	1	1	2.06	0	1	−0.23	−0.47	−0.40	0
5	D-Phe	Gln	L-Arg	1	1	2.06	1	1	0.12	−0.20	−0.91	0
6	D-Phe	Lys	L-Arg	1	1	2.06	1	1	−0.33	0.28	−0.67	0
7	D-Phe	Phe	L-Arg	1	1	2.06	0	1	0.26	−0.37	1.92	0
8	D-phe	Leu	L-Arg	1	1	2.06	0	1	0.23	−0.62	1.22	0
9	D-Phe	Gly	L-Arg	1	1	2.06	0	1	−0.42	0.57	−0.67	0
10	D-Phe	Tyr	L-Arg	1	1	2.06	1	1	0.40	0.38	1.67	0
11	D-Phe	Ser	L-Arg	1	1	2.06	1	0	−0.17	0.21	−0.55	0
12	L-Tyr (Me)[a]	Val	D-Arg	0	0	3.98	0	0	0.68	−0.62	0.91	1
13	L-Tyr (Et)	Val	D-Arg	0	0	4.92	0	0	0.68	−0.62	0.91	1
14	L-Tyr (iso-Pr)	Val	D-Arg	0	0	4.92	0	0	0.68	−0.62	0.91	1
15	L-Tyr (Pr)	Val	D-Arg	0	0	5.03	0	0	0.68	−0.62	0.91	1
16	L-Tyr (Me)	Val	L-Arg	0	0	3.98	0	0	0.68	−0.62	0.91	0
17	L-Tyr (Et)	Val	L-Arg	0	0	4.92	0	0	0.68	−0.62	0.91	0
18	L-Tyr (iso-Pr)	Val	L-Arg	0	0	4.92	0	0	0.68	−0.62	0.91	0
19	L-Tyr (Pr)	Val	L-Arg	0	0	5.03	0	0	0.68	−0.62	0.91	0
20	D-Tyr (Me)	Val	D-Arg	1	0	3.98	0	0	0.68	−0.62	0.91	1
21	D-Tyr (Et)	Val	D-Arg	1	0	4.92	0	0	0.68	−0.62	0.91	1
22	D-Tyr (iso-Pr)	Val	D-Arg	1	0	4.92	0	0	0.68	−0.62	0.91	1
23	D-Tyr (Pr)	Val	D-Arg	1	0	5.03	0	0	0.68	−0.62	0.91	1
24	D-Tyr (Me)	Val	L-Arg	1	0	3.98	0	0	0.68	−0.62	0.91	0
25	D-Tyr (Et)	Val	L-Arg	1	0	4.92	0	0	0.68	−0.62	0.91	0
26	D-Tyr (iso-Pr)	Val	L-Arg	1	0	4.92	0	0	0.68	−0.62	0.91	0
27	D-Tyr (Pr)	Val	L-Arg	1	0	5.03	0	0	0.68	−0.62	0.91	0
28	D-Tyr	Val	D-Arg	1	0	2.74	0	0	0.68	−0.62	0.91	1
29	D-Tyr	Val	L-Arg	1	0	2.74	0	0	0.68	−0.62	0.91	0
30	D-Phe	Abu[b]	L-Arg	1	1	2.06	0	1	—	—	—	0
31	D-Phe	Cha[c]	L-Arg	1	1	2.06	0	1	—	—	—	0
32	D-Phe	Nle[d]	L-Arg	1	1	2.06	0	1	—	—	—	0
33	D-Phe	Nva[e]	L-Arg	1	1	2.06	0	1	—	—	—	0

a) Tyrosine alkyl ether. *b*) α-Aminobutyric acid. *c*) Cyclohexyl alanine. *d*) Norleucine. *e*) Norvaline.

smooth muscle. Further, recent work has revealed that vasopressin acts as a neurohormone in the central nervous system controlling the cardiovascular, renal, and thermoregulatory systems.[14] Vasopressin antagonists are considered valuable pharmacological tools for investigating physiological and behavioral functions of the posterior pituitary hormone.

Arginine-vasopressin antagonists with the following general structure were synthesized, and their anti-antidiuretic and anti-vasopressor activities were measured by Manning *et al.*[15-18] The detailed structures are listed in Table V.



The biological activities were observed *in vivo* as an effective dose (ED) in rats, and all were not measured at the same time. Therefore, the potencies were allotted into three ratings (See Table IX) and were subjected to FALS89 analysis.

As to the structural descriptors in FALS89 analysis, those for X, Y, and Z in the general structure were considered. As for X, which included L-Tyr, D-Tyr, their alkyl ethers, and D-Phe, two indicator variables *D*(X) for X = D-amino acid residue and *Phe*(X) for X = Phe as well as *L*(X) for the length of the side chain were investigated. As for Y, which included 11 natural amino acid residues and 4 artificial amino acid residues, descriptors for hydrogen bonding, hydrophobicity,[19,20] and length of the side chain as well as several conformation parameters[21-23] were studied. Among them, descriptors selected in the QSAR equations in Table VI are two indicator variables *HB*(Y) for hydrogen bonding ability of the side chain and *H*(Y) for helix making properties, the optimal matching hydrophobicity *OMH*(Y),[20] and information measures for extended sheet conformation *E*(Y)[23] and for turns defined as for middle residues *T*(Y).[23] *H*(Y) is taken to be zero for amide-N having no hydrogen atom as Pro, branching at the β-position as Val and Ile, and a hydroxyl group attached to the β-carbon atom as Thr and Ser; and otherwise taken to be 1.[22] As for Z, which included D-Arg and L-Arg, an indicator variable *D*(Z) for D-Arg was used in FALS analysis.

The descriptor values used in the discriminant functions (Table VI) were listed in Table V. The values of *E*(Y), *T*(Y), and *OMH*(Y) were not available for four compounds having artificial amino acid residues for Y. Therefore, FALS89 calculation was carried out with two sets of data, one for 29 compounds excluding the four compounds, and the other for the whole 33 compounds.

TABLE VI. FALS89 Discriminant Functions and Their Reliability

Eq. No.		N	Calcd			Predicted		
			MMG	N_{mis}	R_s[a]	*MMG*	N_{mis}	R_s[a]
	Anti-antidiuretic activity							
16	$Z=$ 1.087 *D*(X) + 2.052 *Phe*(X) + 0.824 *L*(X) [0.49][b] [1.00] [1.09] −2.167 *H*(Y) − 1.569 *D*(Z) − 3.458 [0.93] [0.73]	29	0.925	2 (0)	0.955	0.877	3 (0)	0.922
17	$Z=$ 1.329 *D*(X) + 0.397 *L*(X) [0.59][b] [0.52] −1.637 *H*(Y) − 2.126 *D*(Z) − 1.293 [0.70] [0.98]	29	0.877	3 (0)	0.935	0.841	4 (0)	0.904
18	$Z=$ 1.056 *D*(X) + 1.965 *Phe*(X) + 0.802 *L*(X) [0.45][b] [0.98] [1.07] −2.105 *H*(Y) − 1.541 *D*(Z) − 3.279 [0.99] [0.69]	33	0.903	3 (1)	0.884	0.857	4 (1)	0.849
	Anti-vasopressor activity							
19	$Z=$ −0.136 *D*(X) − 0.311 *Phe*(X) + 3.596 *E*(Y) [0.06][b] [0.15] [1.26] −1.258 *OMH*(Y) − 0.160 *D*(Z) − 0.584 [0.87] [0.07]	29	0.852	4 (0)	0.902	0.785	5 (0)	0.860
20	$Z=$ −0.184 *D*(X) − 1.106 *Phe*(X) − 0.826 *T*(Y) [0.08][b] [0.54] [0.29] −0.514 *H*(Y) − 0.150 *D*(Z) + 0.240 [0.22] [0.07]	29	0.857	4 (1)	0.818	0.845	4 (1)	0.818
21	$Z=$ −0.137 *D*(X) − 1.094 *Phe*(X) − 0.520 *HB*(Y) [0.06][b] [0.54] [0.19] −0.642 *H*(Y) − 0.140 *D*(Z) + 0.885 [0.30] [0.06]	33	0.870	4 (1)	0.862	0.865	4 (1)	0.862

a) Highly significant at the level of $p<0.001$ for all the equations. *b*) Contribution index.

TABLE VII. Squared Cross-Correlation Matrix of Descriptors Appearing in Eqs. 16, 17, 19, and 20

	D(X)	*Phe*(X)	*L*(X)	*H*(Y)	*OMH*(Y)	*E*(Y)	*T*(Y)	*D*(Z)
D(X)	1.000							
Phe(X)	0.233	1.000						
L(X)	0.287	0.804	1.000					
H(Y)	0.121	0.521	0.419	1.000				
OMH(Y)	0.040	0.172	0.139	0.094	1.000			
E(Y)	0.139	0.598	0.482	0.559	0.564	1.000		
T(Y)	0.102	0.439	0.354	0.353	0.330	0.656	1.000	
D(Z)	0.064	0.275	0.221	0.143	0.047	0.165	0.121	1.000

TABLE VIII. Squared Cross-Correlation Matrix of Descriptors Appearing in Eqs. 18 and 21

	D(X)	*Phe*(X)	*L*(X)	*HB*(Y)	*H*(Y)	*D*(Z)
D(X)	1.000					
Phe(X)	0.266	1.000				
L(X)	0.318	0.832	1.000			
HB(Y)	0.057	0.214	0.178	1.000		
H(Y)	0.160	0.600	0.499	0.057	1.000	
D(Z)	0.083	0.313	0.260	0.067	0.187	1.000

The resulting discriminant functions are expressed as Eqs. 16—21 in Table VI. The best combination of fuzzy levels $\{Fl_1, Fl_2\}$ were {0.5, 0.1} in these equations except Eq. 19 {0.1, 0.1}, indicating that the boundary between higher potency classes was clearer. For anti-antidiuretic activity of 29 antagonists, five-descriptor equation Eq. 16 was the best. A high correlation ($r^2=0.80$) was found between *Phe*(X) and *L*(X) as shown in Table VII. However, there seemed to be no fear of chance correlation because Eq. 17, which was lacking in *Phe*(X), still gave fairly good results. For the data set of the whole 33 compounds, Eq. 18 with the same five descriptors as those in Eq. 16 was selected as the best equation.

For anti-vasopressor activity of 29 antagonists, five-descriptor equations Eqs. 19 and 20 were selected. Although Eq. 19 is better in terms of R_S than Eq. 20 both in recognition and in leave-one-out prediction, the predictive *MMG* is somewhat inferior. Eq. 21 for the set of 33 compounds gave the best predictive results both in *MMG* and R_S. The squared correlation matrix for the descriptors appearing in Eqs. 18 and 21 for 33 compounds is listed in Table VIII.

The resulting recognition and leave-one-out prediction using these equations were reasonably good in terms of *MMG* as well as R_S, all of which indicated the significance level of $p<0.001$ (Table VI). The recognized and predicted membership grades and ratings in both activities for each compound using Eqs. 18 and 21 were listed in Table IX, along with observed ratings.

Native arginine-vasopressin contains L-Cys at position 1, L-Tyr for X, L-Gln for Y, and L-Arg for Z. The equations in Table VI indicate a clear contrast for X between the two kinds of activities; anti-vasopressor activity prefers L-Tyr the same as the position 2 residue of native vasopressin, whereas D-Phe, D-Tyr, and Tyr alkyl ether instead of L-Tyr are favorable for anti-antidiuretic activity. As for Y, residues having a tendency to not form helics, pleated sheets, nor turns are preferable for both activities. The dynamics and conformational energetics of lysine-vasopressin were theoretically studied by Hagler *et al.*,[24] and the predomi-

TABLE IX. Observed and Calculated Activities of Arg-Vasopressin Antagonists

Compd.	Anti-antidiuretic activity					Anti-vasopressor activity				
	Obsd Rating[a]	Calcd by Eq. 18		Pred		Obsd Rating[b]	Calcd by Eq. 21		Pred	
		MG	Rating	*MG*	Rating		*MG*	Rating	*MG*	Rating
1	3	1.000	3	1.000	3	2	0.950	2	0.984	2
2	3	1.000	3	1.000	3	2	0.950	2	0.984	2
3	3	1.000	3	1.000	3	1	0.950	1	0.766	1
4	2	0.123	1	0.091	1	1	0.997	1	0.996	1
5	1	0.982	1	0.973	1	3	0.000	1	0.000	1
6	1	0.982	1	0.973	1	1	1.000	1	1.000	1
7	1	0.982	1	0.973	1	1	0.997	1	0.996	1
8	1	0.982	1	0.973	1	1	0.997	1	0.996	1
9	1	0.982	1	0.973	1	1	0.997	1	0.996	1
10	1	0.982	1	0.973	1	1	1.000	1	1.000	1
11	3	1.000	3	1.000	3	1	0.950	1	0.766	1
12	1	1.000	1	1.000	1	3	0.936	3	0.991	3
13	1	1.000	1	1.000	1	3	0.936	3	0.991	3
14	1	1.000	1	1.000	1	3	0.936	3	0.991	3
15	1	0.998	1	0.989	1	2	0.170	3	0.035	3
16	2	0.978	2	0.529	2	3	0.946	3	1.000	3
17	2	1.000	2	1.000	2	2	0.014	3	0.000	3
18	2	1.000	2	1.000	2	3	1.000	3	1.000	3
19	2	1.000	2	0.897	2	3	1.000	3	1.000	3
20	1	0.854	1	0.061	2	2	1.000	2	1.000	2
21	2	1.000	2	1.000	2	2	1.000	2	1.000	2
22	2	1.000	2	1.000	2	2	1.000	2	1.000	2
23	3	0.000	2	0.000	2	2	1.000	2	1.000	2
24	3	1.000	3	1.000	3	3	0.946	3	1.000	3
25	3	1.000	3	1.000	3	3	0.946	3	1.000	3
26	3	1.000	3	1.000	3	3	0.946	3	1.000	3
27	3	1.000	3	1.000	3	2	0.161	3	0.082	3
28	1	1.000	1	1.000	1	2	1.000	2	1.000	2
29	2	0.995	2	0.956	2	3	1.000	3	1.000	3
30	3	0.000	1	0.000	1	1	0.997	1	0.996	1
31	1	0.982	1	0.973	1	1	0.997	1	0.996	1
32	1	0.982	1	0.973	1	1	0.997	1	0.996	1
33	1	0.982	1	0.973	1	1	0.997	1	0.996	1

a) 1, $ED > 4.0$ mmol/kg; 2, $1.7 < ED \leq 4.0$; 3, $ED \leq 1.7$. *b*) 1, $ED > 1.2$ mmol/kg; 2, $0.45 < ED \leq 1.2$; 3, $ED \leq 0.45$.

nant role of Phe at position 3 in the dynamic flexibility and multiple conformational state of the cyclic hexapeptide ring was revealed. Importance of the conformational property of Y located at position 4 seems to be understandable. As for Z, L-Arg, the same as the position 8 residue of native vasopressin, is favorable for both activities.

Thus, FALS analysis successfully generated significant QSAR models which characterized structural features favorable to anti-antidiuretic and anti-vasopressor activities. The resemblance and difference between the characteristics of two kinds of receptors for Arg-vasopressin were interestingly suggested by the FALS calculation results.

The success achieved in these examples using FALS89 as an aid in the empirical rule making process leads us to believe that the new computerized pattern classifier, FALS, could become a useful tool in the development of structure–activity relations and drug design.

Acknowledgment This work was supported by a Grant-in-Aid for Scientific Research (No. 01571183) from the Ministry of Education, Science and Culture, Japan.

References

1) Y. C. Martin, J. B. Holland, C. H. Jarboe, and N. Protnikoff, *J. Med. Chem.*, **17**, 409 (1974).
2) W. J. Dunn and S. Wold, *J. Med. Chem.*, **21**, 922 (1978).
3) A. J. Stuper and P. C. Jurs, *J. Pharm. Sci.*, **67**, 745 (1978).
4) I. Moriguchi and K. Komatsu, *Chem. Pharm. Bull.*, **25**, 2800 (1977); I. Moriguchi, K. Komatsu, and Y. Matsushita, *J. Med. Chem.*, **23**, 20 (1980).
5) D. Dubois and H. Prade, "Fuzzy Sets and Systems: Theory and Application," Academic Press, New York, 1980.
6) I. Moriguchi, S. Hirono, and Q. Liu, Abstracts of Papers, 16th Symposium on Structure–Activity Relationships, Kyoto, Oct. 1988, p. 300; Q. Liu, S. Hirono, Y. Matsushita, T. Nakagawa, and I. Moriguchi, Abstracts of Papers, 17th Symposium on Structure–Activity Relationships, Osaka, Nov. 1989, p. 224.
7) T. Hata, Y. Sano, R. Sugawara, A. Matsuma, K. Kanamori, T. Shima, and T. Hoshi, *J. Antibiot.*, **9**, 141 (1956).
8) I. Moriguchi and K. Komatsu, "Structure–Activity Relationships-Quantitative Approaches," Vol. 2, ed. by T. Fujita, Nankodo, Tokyo, 1982, p. 303.
9) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, **16**, 1207 (1973).
10) I. Moriguchi, Y. Kanada, and K. Komatsu, *Chem. Pharm. Bull.*, **24**, 1799 (1976).
11) Y. C. Martin, "Quantitative Drug Design," Marcel Dekker, New York, 1978.
12) A. Verloop, W. Hoogenstraaten, and J. Tipker, "Drug Design," Vol. VII, ed. by E. J. Ariëns, Academic Press, New York, 1976, p. 165.
13) P. A. Andrews, S. Pan, and N. R. Bachur, *J. Am. Chem. Soc.*, **108**, 4158 (1986).
14) C. L. Riphagen and Q. J. Pitman, *Fed. Proc.*, **45**, 2318 (1986).
15) M. Manning, B. Lammek, and A. M. Kolodziejczyk, *J. Med. Chem.*, **24**, 701 (1981).

16) M. Manning, A. Olma, W. A. Klis, and A. M. Kolodziejczyk, *J. Med. Chem.*, **25**, 45 (1982).
17) M. Manning, A. Olma, W. A. Klis, J. Sato, and W. H. Sawyer, *J. Med. Chem.*, **26**, 1607 (1983).
18) M. Manning, E. Nawrocka, A. Misicka, A. Olma, W. A. Klis, J. Sato, and W. H. Sawyer, *J. Med. Chem.*, **27**, 423 (1984).
19) J. R. Fauchère and V. Pliška, *Eur. J. Med. Chem. Chim. Ther.*, **18**, 369 (1983).
20) R. M. Sweet and D. Eisenberg, *J. Mol. Biol.*, **171**, 479 (1983).
21) P. Y. Chou and C. D. Fasman, *Annu. Rev. Biochem.*, **47**, 251 (1978).
22) R. L. Kisliuk, "Principles of Medicinal Chemistry," ed. by W. O. Foye, Lea and Febiger, Philadelphia, 1981, Chapter 24.
23) B. Robson and J. Garnier, "Introduction to Proteins and Protein Engineering," Elsevier, Amsterdam, 1986, p. 373.
24) A. T. Hagler, D. J. Osguthorpe, P. D. Osguthorpe, and J. C. Hempel, *Science*, **227**, 1309 (1985).