Revision of the Dual Scaling Method for Successive Categories Data and Its Application to Quantitative Structure–Activity Relationships (QSAR)

Tatsuya TAKAGI,* Koichi TANGE, Takahiro JIKIHARA, Norihisa ONOZAWA, Kazumi SAKASHITA, Hideaki FUJIWARA, and Yoshio SASAKI

Faculty of Pharmaceutical Sciences, Osaka University, 1-6 Yamada-oka, Suita, Osaka 565

(Received June 7, 1993)

Nishisato's Dual Scaling Method for successive categories data has been revised in order to apply it to quantitative structure–activity relationships. The revised version can give category values of successive categories, even though the original one can give only boundary values between two adjacent successive categories. The new version can be applied to an analysis of the antimicrobial activities of polymethylene bi(pyridineamine) derivatives, which are potential agents for controlling and/or preventing the formation of dental plaque. The results show that the proper molecular length of the agents leads to the maximum antimicrobial activity.

An effective method for the quantitative structure–activity relationships (QSAR) analysis of successive categories data has long been one of the most important subjects for QSAR analysts. As long as the techniques are concerned which analyze successive categories data having criterion variables, some effective ones have already been reported. These include the FALS[1] method, the logistic regression analysis method,[2] the ORMUCS[3,4] method, and regression discriminant analysis.[5] Nevertheless, very few techniques have been proposed and applied to analyze successive categories data without criterion variables. For example, when the antimicrobial activities of a set of compounds are available in the form of successive categories data, we can perform more precise QSAR analyses if the proper intervals are revealed between the categories. Hence, an effective method has been awaited for the analysis of successive categories data without any criterion variables.

In these situations, Nishisato's dual scaling method for successive categories data (DUS3)[6] is almost a unique technique for this purpose. However, the original DUS3 method is unfavorable for QSAR analyses; although this method does not afford "category values," which are scale values of successive categories on the same scale as the boundary values and the scale values of stimuli, it does afford boundary values between two adjacent successive categories (Fig. 1). The desired data for QSAR analyses are required to be expressed by an interval scale.

Here, the authors revised the original DUS3 method in order to reveal category values of successive categories (DUS3R), and applied it to a QSAR analysis of antimicrobial activities[7] of polymethylene bi(pyridineamine) derivatives, which are potential agents for controlling and/or preventing the formation of dental plaque.

Computation

A calculation with DUS3R, a principal factor analysis, and multiple-regression analyses were all performed using the MVA package program[8] developed and coded by the authors and co-workers, which uses a BASIC version of NICER[9] for the generalized eigenvalue-eigenvector solutions.

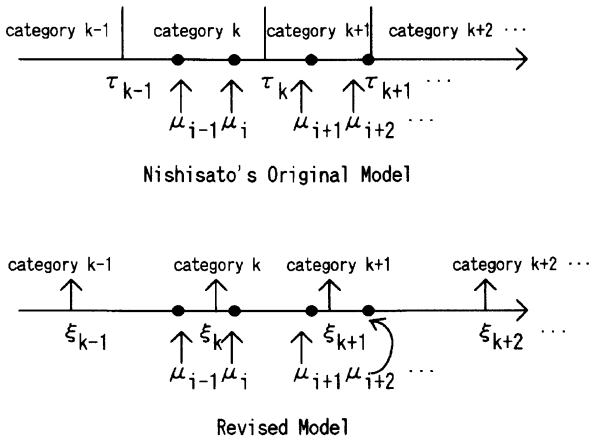


Fig. 1. Revised model and original model of the dual scaling method for successive categories data.

For the values of the communalities, we have adopted squared multiple-correlation coefficients (SMC) which indicate the lower bounds for the communalities. The factor scores were estimated by the least-squares method.

All of the computations were carried out on NEC PC9801/VX21 and EPSON PC286V personal computers.

Theory

Before describing the revision made by the authors, Nishisato's original dual scaling method for successive categories data needs be briefly outlined.

Nishisato's Original Dual Scaling Method for Successive Categories Data. Let A_1, A_2, ···, A_j, ···, A_n be n stimuli, μ_1, μ_2, ···, μ_j, ···, μ_n the corresponding scale values, x_{ij} (i=1, 2, ···, N; j=1, 2, ···, n) the observation values expressed in the form of successive categories data, and 1, 2, 3, ···, k, ···, m, $m+1$ a set of $(m+1)$ successive categories which divide the judgmental continuum into $(m+1)$ sections with m category boundaries, τ_1, τ_2, ···, τ_k, ···, τ_m. The response

variable (f) and the $N \times nm$ partitioned matrix ($\boldsymbol{F}$) can be defined as follows:

$$ {}_{i}f_{jk} = \begin{cases} 1, x_{ij} \leq k \\ -1, x_{ij} > k, \end{cases} \tag{1} $$

$$ \boldsymbol{F} = [\boldsymbol{F}_1, \boldsymbol{F}_2, \cdots, \boldsymbol{F}_j, \cdots, \boldsymbol{F}_n], \tag{2} $$

and

$$ \boldsymbol{F}_j = \begin{bmatrix} {}_{1}f_{j1}, & {}_{1}f_{j2}, & \cdots, & {}_{1}f_{jm} \\ {}_{2}f_{j1}, & {}_{2}f_{j2}, & \cdots, & {}_{2}f_{jm} \\ \vdots & \vdots & & \vdots \\ {}_{N}f_{j1}, & {}_{N}f_{j2}, & \cdots, & {}_{N}f_{jm} \end{bmatrix}. \tag{3} $$

The matrix $\boldsymbol{F}$ can be transformed to $\boldsymbol{Q}$, which is the desired matrix of incidences for parameters (τ_k, μ_j) by using

$$ \boldsymbol{Q} = \boldsymbol{F}\boldsymbol{A} \tag{4} $$

and

$$ \boldsymbol{A} = \begin{bmatrix} \boldsymbol{E}_m, & -\boldsymbol{e}_m, & \boldsymbol{0}, & \boldsymbol{0}, & \cdots, & \boldsymbol{0} \\ \boldsymbol{E}_m, & \boldsymbol{0}, & -\boldsymbol{e}_m, & \boldsymbol{0}, & \cdots, & \boldsymbol{0} \\ & & \vdots & & & \vdots \\ \boldsymbol{E}_m, & \boldsymbol{0}, & \boldsymbol{0}, & \boldsymbol{0}, & \cdots, & -\boldsymbol{e}_m \end{bmatrix}. \tag{5} $$

In Eq. 5, $\boldsymbol{A}$ is the designed matrix of order $nm \times (n+m)$, $\boldsymbol{E}_m$ is the unit matrix of order m, $\boldsymbol{e}_m$ is the unit column vector of order m, and 0 is the zero column vector of order m. It should be noted that the row marginals of $\boldsymbol{Q}$ are all zero.

In this original method, the solution vector of τ_k and μ_j ($\boldsymbol{\theta}$) and the weighted vector for the subjects ($\boldsymbol{\omega}$), which are defined as Eqs. 6 and 7, can be determined so as to maximize the between-column (between-parameter) sum of squares (SS_{b}) relative to the total sum of squares (SS_{t}).

$$ \boldsymbol{\theta}^{\mathrm{t}} = [\tau_1, \tau_2, \cdots, \tau_k, \cdots, \tau_m, \mu_1, \mu_2, \cdots, \mu_j, \cdots, \mu_n] \tag{6} $$

$$ \boldsymbol{\omega}^{\mathrm{t}} = [w_1, w_2, \cdots, w_i, \cdots, w_N] \tag{7} $$

The ratio $SS_{\mathrm{b}}/SS_{\mathrm{t}}$, which is called the squared correlation ratio, is indicated by η^2:

$$ \eta^2 = SS_{\mathrm{b}}/SS_{\mathrm{t}}, \tag{8} $$

where SS_{b} and SS_{t} are expressed as

$$ SS_{\mathrm{b}} = \boldsymbol{\omega}^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{\omega} \tag{9} $$

and

$$ SS_{\mathrm{t}} = \boldsymbol{\omega}^{\mathrm{t}}\boldsymbol{D}_{\mathrm{N}}\boldsymbol{\omega}. \tag{10} $$

In the above equations, $\boldsymbol{D}$ and $\boldsymbol{D}_{\mathrm{N}}$ are the diagonal matrices of order $n+m$ and N, respectively, and are defined as

$$ \boldsymbol{D} = \mathrm{diag}[\underbrace{nN, nN, \cdots, nN}_{m \text{ terms}}, \underbrace{mN, mN, \cdots, mN}_{n \text{ terms}}] \tag{11} $$

and

$$ \boldsymbol{D}_N = \mathrm{diag}[2nm, 2nm, \cdots, 2nm]. \tag{12} $$

If we define

$$ \boldsymbol{B} = \boldsymbol{D}_N{}^{-1/2}\boldsymbol{Q}\boldsymbol{D}^{-1/2}, \tag{13} $$

the maximization of η^2 with respect to $\boldsymbol{\omega}$ results in the eigenequation

$$ \boldsymbol{B}\boldsymbol{B}^{\mathrm{t}}\boldsymbol{V} = \eta^2\boldsymbol{E}\boldsymbol{V}. \tag{14} $$

We can then calculate $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$ as

$$ \boldsymbol{\omega} = \boldsymbol{D}_N{}^{-1/2}\boldsymbol{v} \tag{15} $$

and

$$ \boldsymbol{\theta} = (1/\eta)\boldsymbol{D}^{-1}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{\omega}, \tag{16} $$

where $\boldsymbol{v}$ indicates the eigenvector which is associated with the largest eigenvalue of Eq. 14, η^2.

Revised Dual Scaling Method for Successive Categories Data. Since the category values cannot be obtained using the original method, the authors converted the definition of ${}_{i}f_{jk}$ in Eq. 1 to that in Eq. 17 so as to obtain category values instead of boundary values,

$$ {}_{i}f_{jk} = \begin{cases} 1, x_{ij} < k \\ 0, x_{ij} = k \\ -1, x_{ij} > k. \end{cases} \tag{17} $$

Consequently, the definition of a diagonal matrix of order N, $\boldsymbol{D}_N$, should be converted as

$$ \boldsymbol{D}_N = \mathrm{diag}[3nm, 3nm, \cdots, 3nm]. \tag{18} $$

The order of $\boldsymbol{D}$ therefore changes to $(n+m+1)$, that is,

$$ \boldsymbol{D} = \mathrm{diag}[\underbrace{nN, nN, \cdots, nN}_{m+1 \text{ terms}}, \underbrace{mN, mN, \cdots, mN}_{n \text{ terms}}] \tag{19} $$

Eq. 17 has been derived in light of the dual scaling method of rank order data.

We can thus obtain the following solution vector ($\boldsymbol{\theta}_{\mathrm{r}}$):

$$ \boldsymbol{\theta}_{\mathrm{r}}{}^{\mathrm{t}} = [\xi_1, \xi_2, \cdots, \xi_k, \cdots, \xi_{m+1}, \mu_1, \mu_2, \cdots, \mu_k, \cdots, \mu_n], \tag{20} $$

where ξ_k denotes the category values.

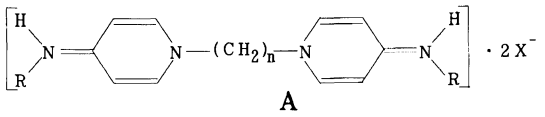
Application

In recent years, most adults have periodontal disease, in which dental plaque — a dense bacterial matrix that adheres avidly to tooth surface — plays an

important role. Hence, the potential antidental-plaque agents have been desired in order to prevent periodontal disease, especially alveolar pyorrhea. Regarding these situations, Bailey et al.[7] have reported on polymethylene bi(pyridineamine) derivatives (Table 1) which are a new class of topical antimicrobial agents as inhibitors of dental plaque that retain efficiency in the presence of saliva. They both synthesized these compounds and determined their antimicrobial minimum inhibitory concentrations (MICs) against a variety of bacteria. Nevertheless, it is difficult to apply a factor analysis or principal component analysis to their data matrix, since some of the MICs values were expressed with a sign of inequality, such as ">250".

Here, the authors tried to apply the above-mentioned DUS3R method to a QSAR analysis of the antimicrobial polymethylene bi(pyridineamine) derivatives (as mentioned below).

DUS3R Analysis. The MICs for the alkyl-substituted polymethylene bi(pyridineamine) derivatives (**A**, R=alkyl and n=4—14) against five kinds of bacteria (*S. aureus* Smith, *E. coli* Vogel, *K. pneumoniae* 39645, *P. mirabilis* MGH-1, and *Ps. aeruginosa* MGH-2) were chosen for this analysis. The data of MICs against two



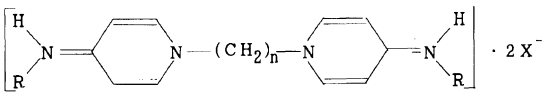
A

Chart 1.

kinds of bacteria, *E. coli* AB1932-1 and *E. coli* 100/B22, which are similar to *E. coli* Vogel, and those for a series of compounds of non-substituted polymethylene bi(pyridineamine) derivatives (**A**, R=H) were excluded from the data matrix (Chart 1). The MICs values were classified into 13 successive categories, as shown in Table 2. The data matrix of order (58×5), thus prepared, are shown in Table 3, which were analyzed using the above-mentioned DUS3R algorithm.

The resulting category values (ζ_c) are given in Table 2 along with the rank of the categories (ν_c). Although we had wanted to apply the normal multiple-regression analyses using predictor variables, such as the partition coefficients, which are frequently used for QSAR analyses, no such variable could be prepared. One of the reasons was that the series of compounds vary the length of the methylene chains as well as the substituents. The

Table 1. Alkyl-Substituted Polymethylene Bi(pyridineamine) Derivatives



Compd	R	n	X
17	n-C_6H_{13}	4	Br
18	n-C_6H_{13}	5	Br
19	n-C_6H_{13}	6	Br
20	n-C_6H_{13}	7	Br
21	n-C_6H_{13}	8	Br
22	n-C_6H_{13}	9	Br
23	n-C_6H_{13}	10	Br
24	n-C_6H_{13}	12	Cl
25	n-C_6H_{13}	14	Cl
26	n-C_7H_{15}	4	Br
27	n-C_7H_{15}	5	Br
28	n-C_7H_{15}	6	Cl
29	n-C_7H_{15}	7	Br
30	n-C_7H_{15}	8	Cl
31	n-C_7H_{15}	9	Cl
32	n-C_7H_{13}	10	Cl
33	n-C_7H_{15}	12	Cl
34	n-C_7H_{15}	14	Cl
35	C_8H_{17}[a]	4	Br
36	C_8H_{17}[a]	5	Br
37	C_8H_{17}[a]	6	Br
38	C_8H_{17}[a]	7	Br
39	C_8H_{17}[a]	8	Br
40	C_8H_{17}[a]	9	Br
41	C_8H_{17}[a]	10	Br
42	C_8H_{17}[a]	12	Br

Table 1. (Continued)

Compd	R	n	X
43	n-C_8H_{17}	6	Cl
44	n-C_8H_{17}	7	Br
45	n-C_8H_{17}	8	Cl
46	n-C_8H_{17}	9	Cl
47	n-C_8H_{17}	10	Cl
48	n-C_8H_{17}	12	Br
49	n-C_8H_{17}	14	Br
50	n-C_9H_{19}	6	Cl
51	n-C_9H_{19}	7	Br
52	n-C_9H_{19}	8	Br
53	n-C_9H_{19}	9	Br
54	n-C_9H_{19}	10	Br
55	n-C_9H_{19}	12	Br
56	n-$C_{10}H_{21}$	6	Br
57	n-$C_{10}H_{21}$	7	Br
58	n-$C_{10}H_{21}$	8	Br
59	n-$C_{10}H_{21}$	9	Br
60	n-$C_{10}H_{21}$	10	Br
61	n-$C_{10}H_{21}$	12	Br
62	n-$C_{12}H_{25}$	6	Br
63	n-$C_{12}H_{25}$	7	Br
64	n-$C_{12}H_{25}$	8	Br
65	n-$C_{12}H_{25}$	9	Br
66	n-$C_{12}H_{25}$	10	Br
67	n-$C_{12}H_{25}$	12	Br
68	C_6H_{11}[b]	4	Br
69	C_6H_{11}[b]	5	Br
70	C_6H_{11}[b]	6	Br
71	C_6H_{11}[b]	7	Br
72	C_6H_{11}[b]	8	Br
73	C_6H_{11}[b]	9	Br
74	C_6H_{11}[b]	10	Br

a) 2-Ethylhexyl. b) Cyclohexyl.

Table 2. Successive Categories Defined from the Metric Data of Antimicrobial Activities

Rank of category (ν_c)	MIC/g m^{-3} a)	Category value (ζ_c)
1	0.2	2.058
2	0.3	1.972
3	0.4	1.942
4	0.5	1.840
5	1.0	1.573
6	1.5	1.385
7	2.0	1.162
8	3.0—3.9	0.658
9	7.8	0.158
10	12.1—48.4	−0.314
11	62.5, >62.5	−0.801
12	125, >125, 250, >250	−1.509
13	500, >500	−2.060

a) From Ref. 7.

predictor variable (N_c), which is sum of the number of carbon atoms in the substituent R and in the methylene chain, was thus defined in order to perform normal multiple-regression analyses. It should be noted that the N_c values of 2-ethylhexyl and cyclohexyl define 6 and 4, respectively, so as to express the length of the compounds. This variable is thought to indicate the relative size and approximately the relative hydrophobicity of the series of compounds.

Multiple Regression Analyses. The results of the multiple-regression analyses are given below which utilize the category values (ζ_c) obtained by the DUS3R method together with those using the rank of category (ν_c) as a criterion variable.

S. aureus

$$\zeta_c = -6.696{N_C}^2 + 6.639N_C \quad (21)$$
$$(\pm 0.956) \quad (\pm 0.956)$$
$$n = 58, R = 0.884, s = 0.480, F = 98.63^{**}$$

$$\nu_c = 6.211{N_C}^2 - 6.046N_C \quad (22)$$
$$(\pm 1.149) \quad (\pm 1.149)$$
$$n = 58, R = 0.828, s = 0.576, F = 59.76^{**}$$

E. coli

$$\zeta_c = -6.264{N_C}^2 + 6.550N_C \quad (23)$$
$$(\pm 0.920) \quad (\pm 0.920)$$
$$n = 58, R = 0.893, s = 0.462, F = 108.47^{**}$$

$$\nu_c = 6.057{N_C}^2 - 6.291N_C \quad (24)$$
$$(\pm 1.077) \quad (\pm 1.077)$$
$$n = 58, R = 0.850, s = 0.541, F = 71.47^{**}$$

K. pneumoniae

$$\zeta_c = -5.594{N_C}^2 + 5.961N_C \quad (25)$$
$$(\pm 1.086) \quad (\pm 1.086)$$
$$n = 58, R = 0.848, s = 0.545, F = 70.16^{**}$$

$$\nu_c = 5.499{N_C}^2 - 5.811N_C \quad (26)$$
$$(\pm 1.198) \quad (\pm 1.198)$$
$$n = 58, R = 0.811, s = 0.601, F = 52.68^{**}$$

P. mirabilis

$$\zeta_c = -3.396{N_C}^2 + 4.014N_C \quad (27)$$
$$(\pm 1.260) \quad (\pm 1.260)$$
$$n = 58, R = 0.788, s = 0.632, F = 45.06^{**}$$

$$\nu_c = 3.499{N_C}^2 - 4.082N_C \quad (28)$$
$$(\pm 1.311) \quad (\pm 1.311)$$
$$n = 58, R = 0.768, s = 0.658, F = 39.47^{**}$$

Ps. aeruginosa

$$\zeta_c = -3.319{N_C}^2 + 3.956N_C \quad (29)$$
$$(\pm 1.297) \quad (\pm 1.297)$$
$$n = 58, R = 0.797, s = 0.621, F = 47.76^{**}$$

$$\nu_c = 3.403{N_C}^2 - 4.003N_C \quad (30)$$
$$(\pm 1.297) \quad (\pm 1.297)$$
$$n = 58, R = 0.774, s = 0.651, F = 41.00^{**}$$

For an adequate comparison among all of these regression equations, all of the variables were normalized. Hence, the constant term disappears in these regression equations. In the case of using the ζ_c values, Eqs. 21, 23, 25, 27, and 29, show better multiple-correlation coefficients than in the case of using the ν_c values. These results suggest that the DUS3R method can sufficiently convert the successive categories data to metric data.

Warner et al.[10] reported on QSAR analyses for the antibacterial activities of carbamidates, bisguanides, and bisbiguanides, which are well known to be anti-dental plaque agents against *S. mutans*, and pointed out that the optimum regions of hydrophobicity lead to the optimum antibacterial activities. The above-mentioned results using the DUS3R method agreed well with those of Warner et al. The optimum number of carbon atoms (N_{opt}) can be calculated from the equation

$$N_{opt} = -\frac{b\sqrt{V_{N_c^2}}}{2a\sqrt{V_{N_c}}}, \quad (31)$$

where a and b are the regression coefficients in the regression equation using normalized variables, $y = aN_c{}^2 +$

Table 3. N_c Values and Rank of Antimicrobial Activities of Alkyl-Substituted Polymethylene Bi(pyridineamine) Derivatives

Compd[a]	Rank of antimicrobial activity										N_c
	S. aureus		*E. coli*		*K. pneumo*		*P. mirab*		*Ps. aerog*		
	ν_c	ζ_c	ν_c	ζ_c	ν_c	ζ_c	ν_c	ζ_c	ν_c	ζ_c	
17	9	0.158	12	−1.509	12	−1.509	12	−1.509	12	−1.509	16
18	7	1.162	12	−1.509	12	−1.509	12	−1.509	12	−1.509	17
19	4	1.840	10	−0.314	12	−1.509	12	−1.509	12	−1.509	18
20	3	1.942	10	−0.314	10	−0.314	12	−1.509	12	−1.509	19
21	1	2.058	8	0.658	9	0.158	12	−1.509	12	−1.509	20
22	3	1.942	7	1.162	8	0.658	12	−1.509	12	−1.509	21
23	1	2.058	7	1.162	8	0.658	10	−0.314	10	−0.314	22
24	1	2.058	5	1.573	5	1.573	9	0.158	10	−0.314	23
25	4	1.840	5	1.573	5	1.573	8	0.658	8	0.658	24
26	5	1.573	10	−0.314	12	−1.509	12	−1.509	12	−1.509	18
27	4	1.840	10	−0.314	10	−0.314	12	−1.509	12	−1.509	19
28	1	2.058	8	0.658	8	0.658	12	−1.509	11	−0.801	20
29	1	2.058	7	1.162	8	0.658	12	−1.509	12	−1.509	21
30	4	1.840	7	1.162	7	1.162	11	−0.801	10	−0.314	22
31	1	2.058	5	1.573	7	1.162	10	−0.314	10	−0.314	23
32	1	2.058	4	1.840	5	1.573	9	0.158	9	0.158	24
33	1	2.058	4	1.840	7	1.162	7	1.162	8	0.658	26
34	5	1.573	7	1.162	8	0.658	8	0.658	7	1.162	28
35	5	1.573	11	−0.801	12	−1.509	12	−1.509	12	−1.509	16
36	4	1.840	10	−0.314	12	−1.509	12	−1.509	12	−1.509	17
37	4	1.840	9	0.158	12	−1.509	12	−1.509	12	−1.509	18
38	4	1.840	8	0.658	9	0.158	12	−1.509	12	−1.509	19
39	2	1.972	8	0.658	9	0.158	12	−1.509	12	−1.509	20
40	2	1.972	5	1.573	9	0.158	12	−1.509	12	−1.509	21
41	1	2.058	6	1.385	8	0.658	10	−0.314	10	−0.314	22
42	4	1.840	5	1.573	7	1.162	8	0.658	8	0.658	24
43	1	2.058	5	1.573	7	1.162	10	−0.314	10	−0.314	22
44	4	1.840	7	1.162	5	1.573	9	0.158	10	−0.314	23
45	1	2.058	5	1.573	5	1.573	8	0.658	9	0.158	24
46	5	1.573	5	1.573	7	1.162	10	−0.314	8	0.658	25
47	5	1.573	5	1.573	7	1.162	7	1.162	8	0.658	26
48	4	1.840	7	1.162	5	1.573	7	1.162	7	1.162	28
49	5	1.573	9	0.158	9	0.158	8	0.658	8	0.658	30
50	5	1.573	7	1.162	8	0.658	9	0.158	10	−0.314	24
51	4	1.840	5	1.573	7	1.162	8	0.658	10	−0.314	25
52	5	1.573	8	0.658	7	1.162	8	0.658	9	0.158	26
53	7	1.162	7	1.162	8	0.658	8	0.658	8	0.658	27
54	7	1.162	7	1.162	8	0.658	8	0.658	8	0.658	28
55	7	1.162	8	0.658	8	0.658	8	0.658	8	0.658	30
56	5	1.573	7	1.162	8	0.658	7	1.162	9	0.158	26
57	5	1.573	8	0.658	9	0.158	8	0.658	9	0.158	27
58	5	1.573	7	1.162	8	0.658	9	0.158	9	0.158	28
59	5	1.573	7	1.162	10	−0.314	9	0.158	9	0.158	29
60	7	1.162	9	0.158	10	−0.314	9	0.158	10	−0.314	30
61	7	1.162	9	0.158	10	−0.314	10	−0.314	10	−0.314	32
62	9	0.158	10	−0.314	11	−0.801	11	−0.801	11	−0.801	30
63	8	0.658	11	−0.801	11	−0.801	11	−0.801	11	−0.801	31
64	9	0.158	10	−0.314	11	−0.801	11	−0.801	11	−0.801	32
65	9	0.158	11	−0.801	11	−0.801	11	−0.801	11	−0.801	33
66	10	−0.314	11	−0.801	11	−0.801	11	−0.801	11	−0.801	34
67	10	−0.314	11	−0.801	11	−0.801	11	−0.801	11	−0.801	36
68	12	−1.509	12	−1.509	13	−2.060	13	−2.060	13	−2.060	12
69	12	−1.509	13	−2.060	13	−2.060	13	−2.060	13	−2.060	13
70	10	−0.314	12	−1.509	12	−1.509	12	−1.509	12	−1.509	14
71	9	0.158	12	−1.509	12	−1.509	12	−1.509	12	−1.509	15
72	8	0.658	11	−0.801	12	−1.509	12	−1.509	12	−1.509	16
73	5	1.573	10	−0.314	12	−1.509	13	−2.060	13	−2.060	17
74	1	2.058	10	−0.314	12	−1.509	12	−1.509	12	−1.509	18

a) See Table 1.

bN_c; here, $V_{N_c^2}$ and V_{N_c} are the variances of the original variables, N_c^2 and N_c, respectively. The optimum number of carbon atoms was calculated to be in the range of 22 to 26.

Factor Analysis Using the DUS3R Results. These days we must take into account two directions in designing antibacterial agents. One is to strengthen antibacterial activities against a wide range of bacteria. The other is to strengthen them against particular kinds of bacteria in order to prevent some side-effects,

such as candidiasis. Since a normal multiple-regression analysis might be useless for these purposes, the principal-component analysis (PCA) and the factor analysis (FA) have been widely used. Nevertheless, PCA is not applicable to sucessive categories data, and FA of successive categories data can hardly obtain factor scores. In these situations, the category values using the DUS3R method are thought to be very useful for designing practical antibacterial agents. The authors have thus applied principal factor analysis (PFA) to a QSAR analysis of the antimicrobial polymethylene bi(pyridineamine) derivatives using the category values obtained by the DUS3R method.

The two factors were extracted in order to account for the data satisfactorily (Table 4). Even for *S. aureus*, which exhibits the minimum communalities, about 80% of the variations in the antimicrobial activities data are explicable. Considering the errors in the data of antimicrobial activities, the variations in the results are sufficiently explicable by the two factors. We also attempted to extract more abstract factors. However, no biological or chemical features were revealed for these higher order factors. We therefore concluded that these two factors are those required, and are sufficient to interpret the data treated here.

The resulting loading matrix is shown in Table 4. The first factor is thought to be a sort of size factor: PFA has a tendency to extract a factor in common with the variables. Here, this size factor reflects the global antimicrobial activities for the five kinds of bacteria. The second factor reflects the diversity of antimicrobial activities among the species of bacteria; the loading values for the Gram-positive bacteria show positive values, and the ones for the Gram-negative bacteria show either negative values or a small absolute value.

The factor scores on the two extracted factors were analyzed using normal multiple-regression analysis; the results are as follows:

$$\begin{gathered} Fac1 = -5.558{N_{\mathrm{C}}}^2 + 6.019N_{\mathrm{C}} \\ (\pm 0.908) \quad (\pm 0.908) \\ n = 58, R = 0.895, F = 111.2^{**} \end{gathered} \tag{32}$$

Table 4. Factor Loadings Calculated by Principal Factor Analysis

	Factor 1	Factor 2
CCR[a)]	0.827	0.998
S. aureus	0.605	0.530
E. coli	0.918	0.309
K. pneumo	0.923	0.128
P. mirab	0.865	−0.417
Ps. aerug	0.871	−0.414

a) Cumulative contribution ratio.

$$\begin{gathered} Fac2 = -2.976{N_{\mathrm{CH2}}}^2 + 2.736N_{\mathrm{CH2}} \\ (\pm 1.018) \quad (\pm 1.022) \\ -0.807N_{\mathrm{R}} - 0.584I_{\mathrm{cyclo}} \\ (\pm 0.209) \quad (\pm 0.205) \\ n = 58, R = 0.796, F = 22.8^{**} \end{gathered} \tag{33}$$

Here, the predictor variables (N_{CH2} and N_{R}) denote the number of carbon atoms in the methylene chain and in the substituent R, respectively. Their detailed definitions are similar to those of the N_{c} variable. The dummy variable (I_{cyclo}) equals unity if the substituent R is a cyclic hydrocarbon and zero if R is not cyclic. The multiple correlation coefficients of Eqs. 21, 22, 23, 24, 25, 26, 27, 28, 29, and 30 and Eq. 32 were not significantly improved when variables, N_{CH2}, N_{R}, and I_{cyclo} were used.

The results of Eqs. 32 and 33 suggest that the global antimicrobial activity is mainly controlled by the length of the polymethylene bi(pyridineamine) derivatives, whereas the dependence of the antimicrobial activities on the bacterial species is due to three different factors: N_{CH2}, N_{R}, and I_{cyclo}.

Of course, these QSAR analyses in this study are at their initial stage. However, the authors consider that they are sufficient to show the efficiencies of the DUS3R method.

Conclusion

1) The dual-scaling method for successive categories data has been revised in order to obtain category values expressed in metric data.
2) This revised method was applied to a QSAR analysis of anti-dental plaque agents and was ascertained to be efficient for QSAR analyses.
3) The QSAR analysis of antimicrobial polymethylene bi(pyridineamine) derivatives revealed that the optimum number of carbon atoms is in the 22-to-26 range
4) The global antimicrobial activity is mainly due to the length of the polymethylene bi(pyridineamine) derivatives, whereas the dependence of the antimicrobial activities on bacterial species is due to three different factors: N_{CH2}, N_{R}, and I_{cyclo}.

References

1) I. Moriguchi, S. Hirono, Q. Liu, Y. Matsushita, and T. Nakagawa, *Chem. Pharm. Bull.*, **38**, 3373 (1990).
2) J. A. Anderson and P. R. Philipps, *Appl. Statist.*, **30**, 22 (1981).
3) Y. Takahashi, Y. Miyashita, H. Abe, and S. Sasaki, *Bunsekikagaku*, **33**, E487 (1984).
4) Y. Takahashi, S. Sasaki, M. Tamaru, I. Shimazaki, S. Ito, S. Kawada, and Y. Suda, *Quant. Struct.-Act. Relat.*, **6**, 17 (1987).
5) T. Takagi, N. Iwata, M. Manabe, H. Izaki, H. Fujiwara, and Y. Sasaki, "Proceedings of the 16th Symposium on Structure–Activity Relationships," 1988, p.304.

6) S. Nishisato, *Jpn. Psychol. Res.*, **22**, 134 (1980).
7) M. Baily, C. G. DeGrazia, S. Hoff, P. Schulenberg, J. R. O'Connor, D. A. Paris, and S. A. Mckenzie, *J. Med. Chem.*, **27**, 1457 (1984).
8) T. Takagi, K. Tange, N. Iwata, Y. Shindo, A. Iwata, T. Katayama, H. Izaki, S. Fujii, and Y. Sasaki, "Proceedings of the 4th Software Conference," 1988, p. 285.
9) Y. Beppu and I. Ninomiya, *QCPE* #409 (1981).
10) V. D. Warner, D. M. Lynch, K. H. Kim, and G. L. Grunewald, *J. Med. Chem.*, **22**, 359 (1979).