Discriminative Structural Analysis Using Pattern Recognition Techniques in the Structure–Taste Problem of Perillartines

YOSHIMASA TAKAHASHI ‡, YOSHIKATSU MIYASHITA *, YUICHIRO TANAKA *, HIROSHI HAYASAKA *, HIDETSUGU ABE ‡, and SHIN-ICHI SASAKI *x

Received July 19, 1982, from the *School of Materials Science and the ‡Research Center for Chemometrics, Toyohashi University of Technology, Tempaku-cho, Toyohashi 440, Japan. Accepted for publication April 15, 1983.

Abstract □ Pattern recognition techniques have been applied to the study of structure–taste correlations for perillartine derivatives. The structure of each compound was described by hydrophobicity (log P), logarithm of water solubility (log S), and topological descriptors relating to some positions which were assigned by superposing each compound on a "template" structure. The fragment molecular connectivities were calculated as the topological descriptors. The discriminant functions between the sweet and bitter taste classes were computed by the use of the simplex optimization technique, which correctly recognized most of the compounds under investigation. It was found that the hydrophobicity and one or two topological descriptors concerned with a specific part of the molecules contributed significantly to the discrimination. The discriminant function obtained correctly classified seven of nine compounds (which were not involved in the data set for developing the function) into the taste class to which they belonged.

Keyphrases □ Structure–taste relationships—discriminative structural analyses using pattern recognition techniques, perillartine derivatives □ Perillartine derivatives—taste, discriminative structural analyses using pattern recognition techniques □ Topology—structure–taste relationships, perillartine derivatives, pattern recognition techniques

Inspection of various classes of sweet-tasting substances shows a very diverse group of chemical compounds having common activity. To establish a structure–taste relationship, a large number of compounds have been tested, and several hypotheses for the sweet taste impression have been proposed by different groups (1). Shallenberger and Acree (2) postulated the first glucophore for sweet-tasting compounds. Their model is referred to as AH—B theory, where A and B are electronegative atoms and H is a hydrogen atom. This AH—B system was extended by Kier (3), who pointed out that the third feature was a nonpolar region, X, which was capable of interacting with a corresponding site on a receptor by dispersion forces. Although there are many exceptions, it seems that the extended theory, the AH—B—X system, is the most sophisticated model among the various theories. However, there remain many problems to be solved for developing new synthetic sweeteners, including that sweet-tasting compounds are often accompanied by bitter and/or other unpreferable tastes. This fact has puzzled chemists interested in searching for new sweeteners. Relating to this taste dichotomy, Belitz and his co-workers (4) suggested the presence of a sweet–bitter receptor, which could be given by formal representation as a hydrophobic pocket with a bipolar system.

Recently, Acton and Stone (5) synthesized 51 perillartine derivatives to develop new potent sweeteners. In their article, taste potencies and taste qualities (sweetness, bitterness, and other tastes) for the perillartines are represented by relative potencies to sucrose and percentages for total taste potencies, respectively. Using this data set, Iwamura (6) implemented the regression analysis with physicochemical and the STERIMOL parameters (7) of the compounds. From the results, he suggested that sweet and bitter receptors have common features. Furthermore, we (8) have shown that the sweet taste receptor might be more restricted (from the viewpoint of spatial shape) than the bitter receptor, by means of class discrimination using the STERIMOL parameters. On the other hand, Kier tried to discriminate between these sweet- and bitter-tasting compounds using discriminant analysis with molecular connectivities (9). He pointed out that unsaturation or heteroatoms are not critical to the structures that influence the sweet and bitter tastes.

In this article, we discuss the results of a discriminative structural analysis of perillartines using pattern recognition techniques; to describe the topological features of these molecules, molecular connectivity indices are also used. However, they are not calculated for whole molecules, but for the fragments related to some specific positions based on a template structure. The use of the indices may give some local topological information on a molecule.

In view of the class discrimination between the sweet and bitter tastes of perillartines, these indices are used in developing the best discriminant function together with two other descriptors. In such classification problems, statistical discriminant analysis (9–11) is frequently used. But the results of such statistical analysis is less reliable when the size of the data set available is not large enough. In contrast, nonparametric pattern recognition is applicable for such small data sets with proper reliability. Therefore, we employed the simplex pattern recognition method for the present study.

EXPERIMENTAL

Data Selection—In the present study, 22 compounds (11 sweet and 11 bitter) were selected as representative compounds from those previously reported by Acton (5). Each possesses >50% of the total taste potency with respect to only the most potent taste quality (Fig. 1). Nine compounds were used as a prediction set for discriminating the taste classes of unknowns by the use of the discriminant function, which was obtained by the former set of compounds. All of the compounds selected are aldoximes.

Template Structure and Fragment Molecular Connectivity—For generating the structural descriptors, a template structure was considered based on the chemical structure of perillartine and related compounds (Fig. 2). Ten positions of interest were considered. Assignments of the positions on the structure for the topological descriptors were made as shown in Fig. 2. For each compound to be analyzed, the positions a–j were assigned by superposing the structure in the manner of planar representation on the template structure. The α,β-unsaturated bonds of skeletons always correspond to the edge between a and b on the template structure. Three positions (h, i, and j) were designated in the descending order of the bulkiness or size of the side chains on each position. As topological descriptors, 20 fragment molecular connectivity (FMC) indices (${}^1\chi_a$–${}^1\chi_j$, ${}^1\chi^v_a$–${}^1\chi^v_j$) related to the 10 positions were adopted for use in this study. These indices are those for path 1 fragments with respect to the given position, which were calculated as the sum of the first-order terms for each edge joining the position of interest (12). Thus, for this study, the path 1 FMC value ${}^1\chi$ (or ${}^1\chi^v$) for the vertex k is given by:

$$\mathrm{FMC}_k = \sum_l (\delta_k \cdot \delta_l)^{-1/2} \qquad \text{(Eq. 1)}$$

0022-3549/84/0600-0737$01.00/0

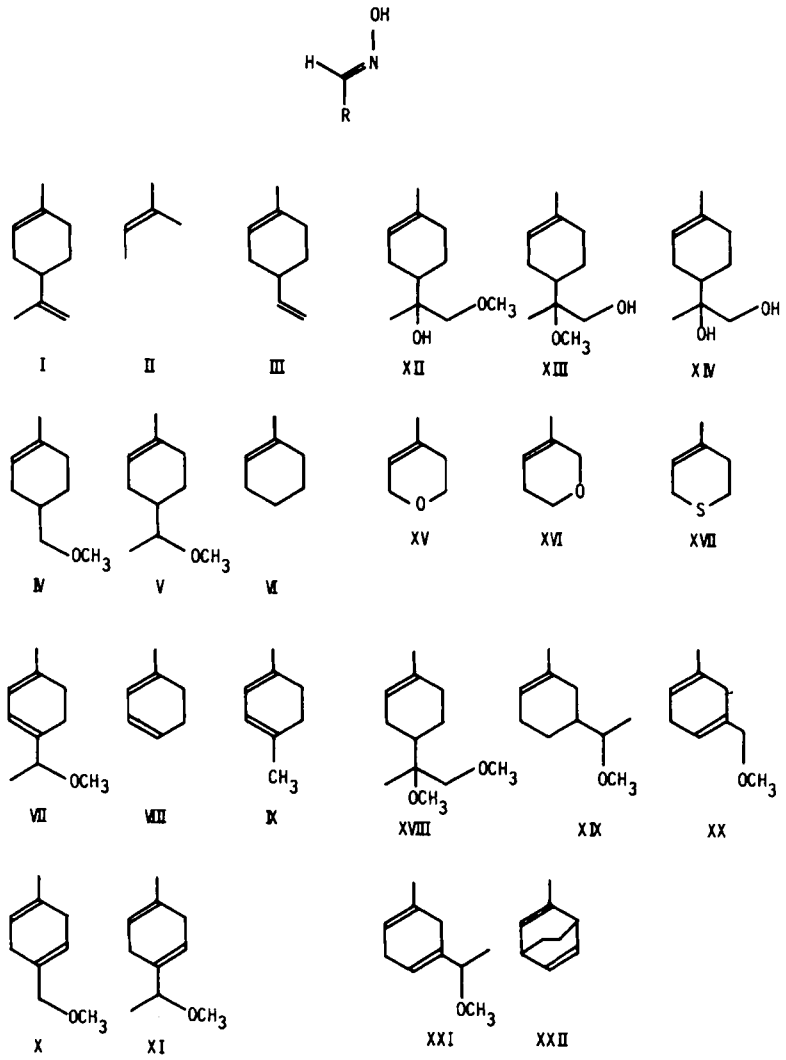

Figure 1—*Perillartine analogues.*

where δ is the valence of a vertex which corresponds to an atom and l is a vertex adjacent to k on the template structure. As a convention, positions were coded with zero for FMC if there were no substituents at the given positions. These structural descriptors were used to describe local topological features of the molecules.

In addition, two other descriptors were used in this study, hydrophobicity (log P) and the logarithm of water solubility (log S) of the compounds. These values were taken from the literature (5, 6). The values of ${}^{1}\chi_a - {}^{1}\chi_j$, log P, and log S for the compounds are shown in Table I.

In the present study, each value of the descriptors employed was normalized in such manner that the mean value was zero and the standard deviation was one for each of 22 descriptors. This process, which is sometimes called autoscaling, will not alter any of the classification results since no change of the relative positions of the pattern points occurs.

Feature Selection—Descriptors not relevant to the classification problem should be eliminated because they may disturb the correct classification or enlarge the computational work. An exhaustive search for the best subset of features is only possible for a small number of descriptors. In this study, the number of original features is too large, and it was necessary to reduce the dimensionality due to the reason mentioned above.

Feature (descriptor) selection was performed with the Fisher ratio (13). The Fisher ratio is a quantitative estimate of the significance of a given parameter for separating two classes. The Fisher ratio of the descriptor k (F_k)

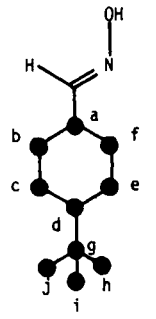

Figure 2—*Template structure for perillartines.*

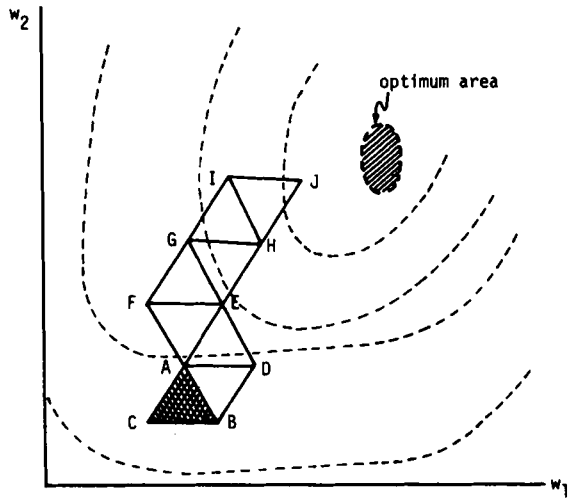

Figure 3—*Progress of the two-dimensional simplex toward the optimum area.*

Table I—Percentages of Taste Qualities and Descriptor Values

Compound	S/B[a]	log P[b]	log S[c]	$^1\chi_a$	$^1\chi_b$	$^1\chi_c$	$^1\chi_d$	$^1\chi_e$	$^1\chi_f$	$^1\chi_g$	$^1\chi_h$	$^1\chi_i$	$^1\chi_j$
I	60/25	2.58	−3.301	1.224	0.908	0.908	1.149	0.908	0.908	1.488	0.577	0.577	0.000
II	55/8	0.87	−0.824	1.393	1.115	0.707	0.000	0.000	0.577	0.000	0.000	0.000	0.000
III	50/10	2.28	−3.699	1.224	0.908	0.908	1.224	0.908	0.908	1.115	0.707	0.000	0.000
IV	53/22	1.10	−1.523	1.224	0.908	0.908	1.224	0.908	0.908	0.908	1.207	0.000	0.000
V	50/20	1.40	−2.523	1.224	0.908	0.908	1.129	0.908	0.908	1.318	1.115	0.577	0.000
VI	50/15	1.48	−1.745	1.224	0.908	1.000	1.000	1.000	0.908	0.000	0.000	0.000	0.000
VII	65/7	1.10	−2.398	1.224	0.908	0.908	1.149	0.908	0.908	1.318	1.115	0.577	0.000
VIII	70/3	1.48	−1.824	1.224	0.908	1.000	1.000	1.000	0.908	0.000	0.000	0.000	0.000
IX	78/3	0.78	−3.222	1.224	0.908	0.908	1.393	0.908	0.908	0.577	0.000	0.000	0.000
X	90/2	0.80	−1.699	1.224	0.908	0.908	1.224	0.908	0.908	0.908	1.204	0.000	0.000
XI	92/1	1.10	−2.398	1.224	0.908	0.908	1.149	0.908	0.908	1.318	1.115	0.577	0.000
XII	6/70	−0.10	−1.921	1.224	0.908	0.908	1.105	0.908	0.908	1.644	0.853	0.500	0.500
XIII	4/93	−0.10	−1.796	1.224	0.908	0.908	1.105	0.908	0.908	1.495	1.060	0.500	1.060
XIV	5/73	−0.92	0.301	1.224	0.908	0.908	1.105	0.908	0.908	1.642	1.060	0.500	0.500
XV	0/50	−0.72	−0.155	1.224	0.908	1.000	1.000	1.000	0.908	0.000	0.000	0.000	0.000
XVI	2/65	−0.72	−1.097	1.224	0.908	1.000	1.000	1.000	0.908	0.000	0.000	0.000	0.000
XVII	0/70	0.34	−1.260	1.224	0.908	1.000	1.000	1.000	0.908	0.000	0.000	0.000	0.000
XVIII	0/78	0.72	−2.046	1.224	0.908	0.908	1.105	0.908	0.908	1.495	0.853	0.500	1.060
XIX	2/52	1.40	−2.523	1.224	0.908	1.000	0.908	1.149	0.816	0.000	0.000	0.000	0.000
XX	0/75	0.80	−1.699	1.224	0.908	1.000	0.908	1.224	0.816	0.000	0.000	0.000	0.000
XXI	0/67	1.10	−2.398	1.224	0.908	1.000	0.908	1.149	0.816	0.000	0.000	0.000	0.000
XXII	0/50	1.90	−1.921	1.149	0.816	1.224	0.908	0.908	1.149	0.000	0.000	0.000	0.000

[a] Compounds with (sweet/total taste) ≥0.5 are taken as sweet and those with (bitter/total taste) ≥0.5 as bitter. [b] Taken from Ref. 6. [c] Taken from Ref. 5.

is calculated from Eq. 2; it is the ratio between the square of the difference in interclass means and the sum of squared intraclass standard deviations:

$$F_k = \frac{(\bar{x}_k^{(1)} - \bar{x}_k^{(2)})^2}{\sigma_k^{(1)2} + \sigma_k^{(2)2}} \qquad \text{(Eq. 2)}$$

where $\bar{x}_k^{(1)}$ and $\bar{x}_k^{(2)}$ are the mean values of the descriptor k in classes 1 and 2, respectively, and $\sigma_k^{(1)}$ and $\sigma_k^{(2)}$ are the standard deviations of those classes. The features are ranked according to their importance, and less important features are discarded.

Development of the Discriminant Function—The simplex pattern recognition method was applied to develop a near optimum discriminant function for discrimination between the sweet- and bitter-tasting compounds. The simplex method is an approach to optimize an objective function (*e.g.*, discriminant function). This optimization technique was previously applied to the optimization of linear discriminant functions for chemical applications by Ritter *et al.* (14). In the present work, the simplex pattern recognition program(SPLX), which was developed in our laboratory according to their algorithm with a small modification, was used.

In the case of two classes, if pattern points on one side of a discriminant surface always belong to class 1 and those on the other side belong to class 2, the data are linearly separable by a linear discriminant function of the following form:

$$g(\mathbf{X}) = w_1x_1 + w_2x_2 + \ldots + w_dx_d = |\mathbf{W}||\mathbf{X}|\cos\theta \qquad \text{(Eq. 3)}$$

where x_i is the ith component of a pattern $\mathbf{X}$ and w_i is the weight assigned to that component. Hence, the classification into two categories can be implemented with the sign of $g(\mathbf{X})$ given by the dot product of a pattern vector $\mathbf{X}$ and a weight vector $\mathbf{W}$. The optimization problem of that function, in this case, can be placed in that of the corresponding weight vector in the weight space.

A simplex is a geometric figure in d-dimensional space, which consists of $d + 1$ vertices. For example, a two-dimensional simplex is a triangle, a three-dimensional simplex is a tetrahedron, *etc.* Figure 3 shows a two-dimensional simplex superimposed on a contour map, which represents the various equiresponse lines of a function of two variables. In this case, the search variables are the components of the weight vector. On the other hand, two criterion functions are evaluated in terms of the response. These are given by:

$$R = N_c/N_t \qquad \text{(Eq. 4)}$$

$$P = \sum_{\mathbf{X}_i \in S_E} |\mathbf{W} \cdot \mathbf{X}_i| \qquad \text{(Eq. 5)}$$

In Eq. 4, N_t is the total number of compounds tested and N_c is that of compounds correctly classified, so R is a recognition rate. On the other hand, in Eq. 5, S_E is the set of all patterns misclassified by $\mathbf{W}$. Thus, P is the sum of the absolute values of the dot products of all misclassified patterns by the use of the weight vector $\mathbf{W}$. In general, the value of P will become small if few compounds are misclassified or if those misclassified lie close to the discriminant surface. This criterion is used only as a smoothing function, so that the ultimate solution will be given by the minimum P value for a weight vector giving maximum recognition rate.

The optimization procedure begins by locating initial simplex (initial weight vectors) on the d-dimensional weight vector space. These were generated as follows:

$$\begin{aligned}
\mathbf{W}_1 &= (w_1, w_2, w_3, \ldots, w_d, w_{d+1}) \\
\mathbf{W}_2 &= (w_1 + \delta_1, w_2, w_3, \ldots, w_d, w_{d+1}) \\
\mathbf{W}_3 &= (w_1, w_2 + \delta_2, w_3, \ldots, w_d, w_{d+1}) \\
&\vdots \\
\mathbf{W}_{d+1} &= (w_1, w_2, w_3, \ldots, w_d + \delta_d, w_{d+1})
\end{aligned} \qquad \text{(Eq. 6)}$$

where $w_{d+1} = -1$ and δ_j was defined as:

$$\delta_j = \alpha w_j \qquad \text{(Eq. 7)}$$

The constraint of $w_{d+1} = -1$ describes a unique weight vector for a given decision surface (14). In Eq. 7, α is an arbitrary constant parameter concerned

Table II—Fisher Ratios of Descriptors

Descriptor	Fisher Ratio	Rank
log P	7.84×10^{-2}	1
$^1\chi_c$	4.21×10^{-2}	2
$^1\chi_j$	3.93×10^{-2}	3
log S	3.63×10^{-2}	4
$^1\chi_e$	2.61×10^{-2}	5
$^1\chi_h$	1.45×10^{-2}	6
$^1\chi_a$	1.44×10^{-2}	7
$^1\chi_b$	1.44×10^{-2}	8
$^1\chi_g$	5.58×10^{-3}	9
$^1\chi_f$	3.63×10^{-3}	10
$^1\chi_d$	1.97×10^{-3}	11
$^1\chi_i$	4.85×10^{-4}	12

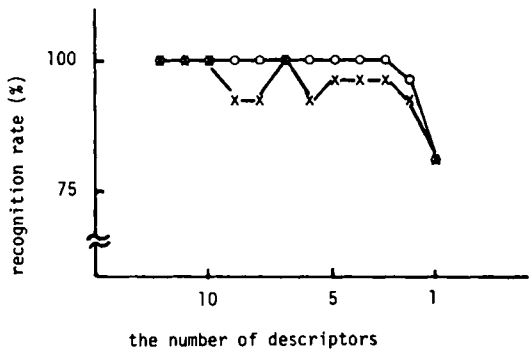

Figure 4—*Recognition rates obtained by the use of various descriptor set contained with $^1\chi$ FMC indices. Key: (○) simplex method; (X) discriminant analysis.*

Table III—Cross-Correlation Table of Five Descriptors

	log P	log S	$^1\chi_c$	$^1\chi_e$	$^1\chi_j$
log P	1.00				
log S	−0.76	1.00			
$^1\chi_c$	0.10	0.00	1.00		
$^1\chi_e$	−0.01	−0.19	0.65	1.00	
$^1\chi_j$	−0.39	0.16	−0.19	−0.03	1.00

with a kind of size on the initial simplex. In the present study, α was selected to be 0.5.

A prototype weight vector (initial weight vector) $\mathbf{W}_1$, in generating the initial simplex, was determined by the use of the bisecting approach in which the linear discriminant surface might be approximate as the hyperplane bisecting two class centers (14). Once the initial simplex was located in the objective space, optimization of the simplex (or weight vectors) by a certain moving algorithm was attempted. Moving the simplex in the optimization process was performed according to the algorithm suggested by Nelder and Mead (15), which is a logical algorithm consisting of reflection, expansion, and contraction rules. Since this procedure has been described in detail elsewhere (14–16), it will not be repeated here.

In addition, another approach was also employed for the development of discriminant functions. Ordinary statistical discriminant analysis was implemented to the same data set for comparison with the above method.

RESULTS AND DISCUSSION

The correlations between the descriptors under investigation were examined. The results of the correlation analysis showed that the values of descriptors $^1\chi$ and $^1\chi^v$ were highly correlated (correlation coefficients = 0.78–1.00) with each other on the same position of the molecule. In such classification problems, the use of highly correlated pairs among descriptors may lead to the reduction of reliability of discriminant functions, decreasing the intrinsic dimensionality in the pattern space and increasing the probability of chance classification. Thus, in the present work $^1\chi$ values were taken as the topological variables relating to each position on the template structure. Furthermore, for the purpose of reducing the number of descriptors, preselection of the descriptors was implemented by the use of the Fisher ratio, which is a measure of the significance of a given parameter for differentiating between the taste classes. Each descriptor was ranked in decreasing order of the Fisher ratio (Table II). According to the ranking, the descriptors were eliminated one at a time and then a linear discriminant function was developed with the descriptor set using the simplex method or statistical discriminant analysis. The plots of recognition rate R *versus* the number of descriptors are illustrated in Fig. 4.

As shown in Fig. 4, in most cases the results for discriminant analysis were not as good as those for the simplex method. This would be anticipated because the $^1\chi$ FMC indices show noticeably non-normal distributions (Table I). It seemed that the simplex pattern recognition was more suitable for the discriminative analysis for the present data set, rather than the other method.

It was shown, in this case, that only two or three descriptors were enough to discriminate between the two taste classes from the result of the simplex method: log P, $^1\chi_c$, and $^1\chi_j$. During the analysis described above, only the Fisher ratios were used for the feature selection. However, because the Fisher ratio is based on the global class separation information of the data structure, it is not always true that the obtained result is the best. Thus, further examinations would be required for the descriptors. Five of the best descriptors were selected on the basis of the Fisher ratio, log P, log S, $^1\chi_c$, $^1\chi_h$, and $^1\chi_j$. The cross-correlation table among the descriptors is shown in Table III. These descriptors were used in the following studies.

To develop a discriminant function with the best set of variables among the combinations of the above descriptors, implementations were performed. The best linear discriminant function with three or two descriptors is shown as follows:

$$g_3(\mathbf{X}) = 0.595 \cdot \log P - 0.773 \cdot {}^1\chi_c - 0.220 \cdot {}^1\chi_j - 0.882 \times 10^{-5} \quad \text{(Eq. 8)}$$

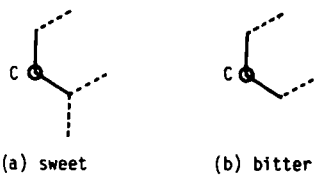

Figure 5—*Characteristic topological fragments for the sweet or bitter compounds.*

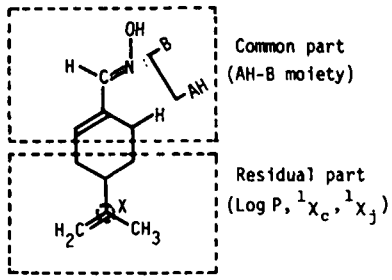

Figure 6—*AH—B—X postulation for perillaldehyde oxime (3).*

$$g_2(\mathbf{X}) = 0.696 \cdot \log P - 0.718 \cdot {}^1\chi_c - 0.220 \times 10^{-5} \quad \text{(Eq. 9)}$$

$$g(\mathbf{X}) \begin{cases} \geq 0 \text{ implies sweet class} \\ < 0 \text{ implies bitter class} \end{cases}$$

Equation 8 correctly classified all compounds into the observed classes. On the other hand, Equation 9 correctly discriminated all of the 11 sweet-tasting compounds and 10 of the 11 bitter-tasting compounds. The average recognition rate was 95%; only XVIII was misclassified.

Equations 8 and 9 suggest that the log P and fragment molecular connectivity index, $^1\chi_c$, are significant characteristics that influence whether a compound is sweet or bitter. In the two equations, as the hydrophobicity of a compound increases, the $g(\mathbf{X})$ values are larger due to the positive value of that coefficient. The compound then tends to be classified as sweet.

The negative coefficient for the $^1\chi_c$ term suggests that a smaller value of $^1\chi_c$ enhances the sweet-taste impression. For the data set employed in the present work, however, the values of $^1\chi_c$ do not vary much. In this case, it is helpful to consider the contribution of $^1\chi_c$ in terms of topological features. The $^1\chi_c$ index describes the branching information or the topological form concerned with the path 1 fragment at position c. Although each value does not always correspond to a unique fragment, it is meaningful to examine the topological fragment. The topological fragments concerned with those positions are illustrated in Fig. 5. Most compounds in the sweet class involve a fragment such as a in Fig. 5; the compounds in another class involve fragment b, excepting some compounds. It should be noted that fragment b is embedded in another fragment.

Furthermore, the third term (which was weighted by a negative coefficient in Eq. 8) suggests that bulky side chains on the j position may prevent the impression of sweetness. Alternatively, because $^1\chi_j$ is zero except for four compounds in the bitter class, this variable could be replaced by an indicator variable related to the presence of a quaternary carbon at the g position. Thus, it is postulated that the occupancy by a substituent at the position j may influence the taste quality of these compounds.

In the present data set, since the aldoxime moiety is a common part of all molecules, the variation of log P is due to the residual part. Thus, the observations are consistent with Kier's postulation of perillartine in his AH—B—X

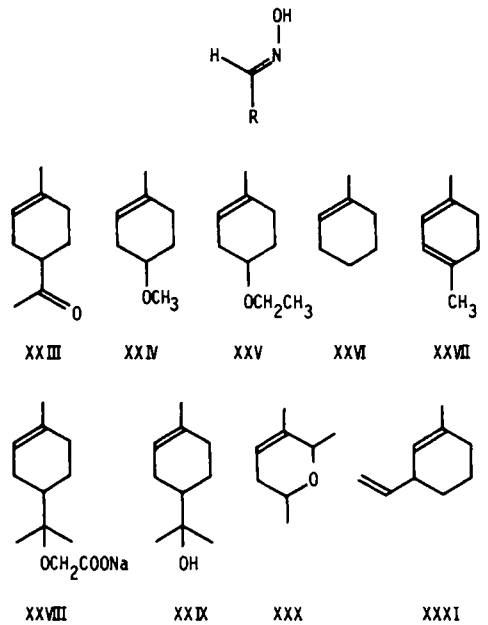

Figure 7—*A set of compounds for the taste predictions.*

Table IV—Descriptor Values and Percentages of Taste Qualities in Prediction Set

Compound	S/B	log P	log S	$^1\chi_a$	$^1\chi_b$	$^1\chi_c$	$^1\chi_d$	$^1\chi_e$	$^1\chi_f$	$^1\chi_g$	$^1\chi_h$	$^1\chi_i$	$^1\chi_j$
XXIII	23/14	0.32	−1.96	1.224	0.908	0.908	1.149	0.908	0.908	1.488	0.577	0.577	0.000
XXIV	42/26	0.60	−1.30	1.224	0.908	0.908	1.224	0.908	0.908	1.115	0.707	0.000	0.000
XXV	45/26	1.10	−1.96	1.224	0.908	0.908	1.224	0.908	0.908	0.908	1.207	0.000	0.000
XXVI	40/16	1.78	−1.77	1.224	0.908	1.000	1.000	1.000	0.908	0.000	0.000	0.000	0.000
XXVII	22/13	0.78	−3.22	1.224	0.908	0.908	1.393	0.908	0.908	0.577	0.000	0.000	0.000
XXVIII	0/38	−4.08	−2.30	1.224	0.908	0.908	1.105	0.908	0.908	1.642	0.500	0.500	0.853
XXIX	4/18	0.88	−1.57	1.224	0.908	0.908	1.105	0.908	0.908	0.788	0.500	0.500	0.500
XXX	0/36	−0.16	−1.64	1.149	0.908	0.908	1.393	0.908	1.318	0.577	0.000	0.000	0.000
XXXI	14/34	2.28	−3.70	1.224	0.816	1.224	0.908	1.000	0.908	0.000	0.000	0.000	0.000

theory (3). The AH—B—X system of perillartine is illustrated in Fig. 6. Furthermore, the aforementioned results suggest that the shape of the moiety concerned with the third site (X) plays an important role in the taste impression (especially, sweetness). This insight would be clear in the following predictive analysis.

The discriminant ability of the linear discriminant function obtained in the present study can be tested by predicting the taste classes of compounds not used in the development of the function. For this test, as shown in Fig. 7, nine compounds were selected from a previous study (5). These compounds are not dominant in one taste class, as seen by the total taste potency (Table IV). However, on the basis of the sweet-bitter taste ratios, each compound could be assigned to either the sweet or bitter taste class. The first five compounds (XXIII-XXVII) are categorized as sweet, and the other compounds (XXVIII-XXXI) are categorized as bitter. The prediction results of their taste classes are shown in Table V. Both discriminant functions, $g_2(X)$ and $g_3(X)$, correctly predicted the class of seven of the tested compounds.

Compounds XXIII and XXIX were misclassified (Fig. 7). They are the smallest in terms of the difference between the sweet and bitter taste ratios (Table IV). Considering that the compounds tested were not so typically sweet or bitter, this result may be more favorable.

The results of this prediction encouraged us to apply the methodology of this study to the structure-taste problem for the series of perillartines. In such a discrimination, the influence of hetero atoms or unsaturated bonds on a structure was examined in the same manner by the use of FMC indices ($^1\chi^v_{a-j}$) with the weighted valences. The values of the weighted valences were obtained from a previous paper (12). But, the result was not as favorable, as shown in Fig. 8. Furthermore, an additional analysis in which $^1\chi$ and $^1\chi^v$ FMC indices were used together in developing the best discriminant function had the same results as using only $^1\chi$ FMC indices. These examinations suggest that the hetero atoms and unsaturated bonds are not so critical in the aforementioned residual part.

CONCLUSION

The structural discriminative analysis of the sweet and bitter perillartines using a template structure was successfully performed by pattern recognition techniques. It was shown that the fragment molecular connectivity indices are useful in describing the local features of molecules in terms of topological structure.

The results for the series of perillartines offer some explanations of the structure-taste correlation. The hydrophobicity of the residual part except for the AH—B moiety in the analogues influenced the taste quality, and furthermore the topological shape of the residual part is also predominantly concerned with the sweet or the bitter taste impression. The information obtained from the best discriminant functions was used to examine additional molecules, and it improved predictions of their taste classes. This study confirms Kier's original postulation (3) of the sweet pharmacophore for perillartine.

Table V—Prediction Results from Simplex Pattern Recognition

Compound	(S/B)	Prediction $g_2(X)$	Prediction $g_3(X)$
XXIII	(S)	X	X
XXIV	(S)	O	O
XXV	(S)	O	O
XXVI	(S)	O	O
XXVII	(S)	O	O
XXVIII	(B)	O	O
XXIX	(B)	X	X
XXX	(B)	O	O
XXXI	(B)	O	O

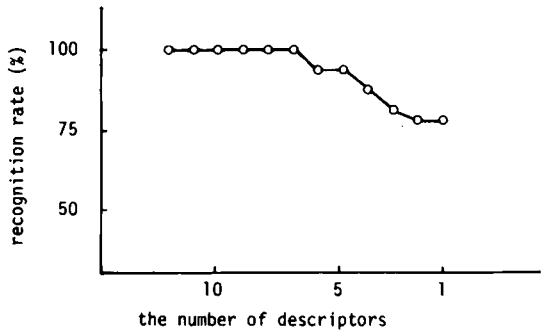

Figure 8—*Recognition rates obtained by the use of various descriptor sets with $^1\chi^v$ FMC indices using the simplex method.*

REFERENCES

(1) J. C. Boudreau (Ed.), "Food Taste Chemistry," ACS Symposium Series, No. 115, American Chemical Society, Washington, D.C., 1979.

(2) R. S. Shallenberger and T. E. Acree, *Nature (London)*, **216**, 480 (1967).

(3) L. B. Kier, *J. Pharm. Sci.*, **61**, 1394 (1972).

(4) H. D. Belitz, W. Chen, H. Jugei, R. Treleanco, H. Wieser, J. Gasteiger, and M. Marsili, in "Food Taste Chemistry," J. C. Boudreau, Ed., p. 93.

(5) E. M. Acton and H. Stone, *Science*, **193**, 584 (1976).

(6) H. Iwamura, *J. Med. Chem.*, **23**, 308 (1980).

(7) A. Verloop, W. Hoogenstraaten, and J. Tipker, in "Drug Design," vol. 7, E. J. Ariens, Ed., Academic, New York, N.Y., 1976, p. 165.

(8) Y. Takahashi, Y. Miyashita, Y. Tanaka, H. Abe, and S. Sasaki, *J. Med. Chem.*, **25**, 1245 (1982).

(9) L. B. Kier, *J. Pharm. Sci.*, **69**, 416 (1980).

(10) Y. C. Martin, J. B. Holland, C. H. Jarboe, and N. Plotnikoff, *J. Med. Chem.*, **17**, 409 (1974).

(11) D. R. Henry and J. H. Block, *J. Med. Chem.*, **22**, 465 (1979).

(12) L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research," Academic, New York, N.Y., 1976.

(13) B. R. Kowalski, "Chemometrics, Theory and Application," ACS Symposium Series, No. 52, American Chemical Society, Washington, D.C., 1977.

(14) G. L. Ritter, S. R. Lowry, C. L. Wilkins, and T. L. Isenhour, *Anal. Chem.*, **47**, 1951 (1975).

(15) J. A. Nelder and G. Mead, *Comput. J.*, **7**, 308 (1965).

(16) S. N. Deming and S. L. Morgan, *Anal. Chem.*, **45**, 278A (1973).

ACKNOWLEDGMENTS

The authors thank the Computer Center, Institute for Molecular Science for affording facilities for computation. The authors also thank Dr. C. Takayama for his invaluable discussion.