Differentiation of Soy Sauce Produced from Whole Soybeans and Defatted Soybeans by Pattern Recognition Analysis of HPLC Profiles

Emiko Kinoshita,* Toshinobu Sugimoto, Yoshinori Ozawa, and Tetsuo Aishima

Research and Development Division, Kikkoman Corporation, 399 Noda, Noda, Chiba 278, Japan

Nonvolatile components in commercial fermented soy sauce produced from whole soybeans and defatted soybeans were analyzed by reversed phase HPLC (RP-HPLC). Differentiation of soy sauce produced from two types of soybeans was attempted by pattern recognition analysis of HPLC profiles. Two distinct groups were observed in both the factor score plot and the clustering dendrogram of HPLC profiles. Two soy sauce groups were clearly differentiated by soft independent modeling of class analogy (SIMCA), where ferulic acid was identified as one of the key components for the differentiation. A fractional factorial experimental design combined with multiple linear regression analysis was applied to determine components associated with types of soybeans. Daidzein and three isoflavone derivatives besides ferulic acid were identified as typical components that increased when soy sauce was produced from whole soybeans.

Keywords: *Soy sauce; pattern recognition; cluster analysis; SIMCA; fractional factorial design*

INTRODUCTION

Fermented soy sauce is the essential seasoning for Japanese cuisine, but consumption of deep-colored soy sauce was the largest among various types of soy sauces mostly produced from wheat and soybeans (Fukushima, 1985). Today, two types of deep-colored soy sauce are being distributed in Japan. The critical difference between them lies in the types of soybeans used as the starting materials. Originally, soy sauce had been produced from wheat and whole soybeans, referred to as "M-soy sauce" hereafter, since the 16th century in Japan. On the other hand, soy sauce production using wheat and defatted soybeans, referred to as "K-soy sauce" hereafter, was developed just after World War II to efficiently utilize defatted soybeans. K-Soy sauce has acquired popularity since then because of its stronger flavor. Until the 1980s, nearly 99.8% of deep-colored soy sauce distributed in Japan was K-soy sauce (Fukushima, 1985). However, recently, the production volume of M-soy sauce has been increasing again to meet consumers' demand for soy sauce with a milder flavor. Consequently, according to the statistics in 1996, the production ratio of M-soy sauce has reached ~10% in the total production of deep-colored soy sauce. In the market, the prices of M-soy sauce are generally higher than those of K-soy sauce because whole soybeans are more expensive than defatted ones. Thus, objective differentiation of K- and M-soy sauces is needed to authenticate products.

Manufacturing of deep-colored soy sauce is performed through several steps and usually takes 6−8 months. Therefore, the quality of soy sauce depends on conditions in each production step: raw materials; ratios of wheat to soybeans; cultivars of soybeans and/or wheat; defatted or whole soybeans; conditions for heat treatments of raw materials; microorganisms grown in koji and moromi stages (Ishihara et al.,1996); aeration for moromi; and temperature and time for pasteurization of raw soy sauce. Composition of amino acids, sugars, organic acids, and other minor components produced by enzymatic action of various microorganisms differs among products produced by different manufacturers. Furthermore, contents of other minor components, such as isoflavones, originally contained in soybeans (Wang et al., 1990; Coward et al., 1993; Kinoshita et al., 1993; Wang and Murphy, 1994) were also different from soy sauce to soy sauce. About 20% of crude fat is contained in whole soybeans, while the fat content in defatted soybeans is only a trace. Thus, lipophilic compounds including aroma volatiles might transfer to the soybean oil fraction from soy sauce during the production process, if a large amount of oil is to be present.

Multivariate pattern recognition techniques are useful to examine whether two sample sets can be classified on the basis of their instrumental data (Aishima and Nakai, 1991). Gas chromatographic data of volatile components in soy sauce were studied in detail by using multivariate techniques by Aishima (1983). Recently, new isoflavones were identified as key components to differentiate soy sauce produced by different manufacturers in pattern recognition analysis of HPLC profiles of nonvolatile components (Kinoshita et al., 1997).

In this research, pattern recognition techniques were applied to HPLC profiles of nonvolatile components of commercial K- and M-soy sauces to examine whether they could be differentiated on the basis of their components. Soy sauce production based on a fractional factorial experimental design was conducted to identify components associated with types of soybeans used. Furthermore, elucidation of chemical structures of peak components that played key roles in the differentiation was attempted to understand quality differences of soy sauce in terms of chemical components.

* Author to whom correspondence should be addressed (telephone +81 471 23 5590; fax +81 471 23 5550; e-mail QZS01157@niftyserve.or.jp).

S0021-8561(97)00461-5 CCC: $15.00 © 1998 American Chemical Society
Published on Web 02/07/1998

Table 1. Experimental Conditions Based on Fractional Factorial Design and Responses from Them

expt no.	factor					peak area y
	x_1	x_2	x_3	x_4	x_5	
1	−1	−1	−1	−1	1	549440[a]
2	1	−1	−1	−1	−1	548330
3	−1	1	−1	−1	−1	492570
4	1	1	−1	−1	1	527670
5	−1	−1	1	−1	−1	272270
6	1	−1	1	−1	1	273470
7	−1	1	1	−1	1	232600
8	1	1	1	−1	−1	260600
9	−1	−1	−1	1	−1	513410
10	1	−1	−1	1	1	420350
11	−1	1	−1	1	1	360490
12	1	1	−1	1	−1	234180
13	−1	−1	1	1	1	174460
14	1	−1	1	1	−1	237600
15	−1	1	1	1	−1	242340
16	1	1	1	1	1	152000
17	1	1	1	1	1	171410
18	1	1	1	1	1	191750
19	1	1	1	1	1	166810

1[b]	whole	5/week[c]	10^5 cells[d]	*A. sojae*	65/35
−1	defatted	1/week	0	*A. oryzae*	55/45

[a] Areas of peak 34. [b] Factor level. [c] Frequency of aeration. [d] Number of cells/mL.

MATERIALS AND METHODS

Materials. Two types of fermented deep-colored soy sauce produced by the same manufacturer were periodically purchased at a supermarket for about one year. M-Soy sauce was produced from only whole soybeans, whereas K-soy sauce was produced from only defatted soybeans according to the lists of their ingredients. The numbers of K- and M-soy sauces analyzed by HPLC were 14 and 15, respectively. Soybeans and wheat used in the soy sauce production based on a fractional factorial design were imported from the United States and Canada, respectively. Defatted soybeans were purchased from Nissin Flour Milling Co. Ltd. (Tokyo, Japan). Salt was purchased from Japan Tobacco Inc. (Tokyo, Japan). Authentic ferulic acid and daidzein were obtained from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan) and Funakoshi Co., Ltd. (Tokyo, Japan), respectively.

Fractional Factorial Design for Soy Sauce Production. A fractional factorial design (FFD; Morgan, 1991) was used to examine significant factors in the production process and interaction between them for generating nonvolatile components in soy sauce. A two-level FFD, 2^{5-1}, was used to obtain estimates of effects, and the design matrix is given in Table 1. Nineteen experiments including three replicates were designed by combining high, +1, and low, −1, levels of five factors considered as potentially important: x_1, type of soybeans; x_2, aeration; x_3, yeast; x_4, type of mold, and x_5, ratio of the amount of soybeans to that of wheat. No experiment at center point was designed because the type of mold is not a continuous variable but only a category variable. Replicate production was conducted so that all analytical results could be analyzed by analysis of variance. By using FFD and multiple linear regression (MLR) analysis, an estimate of all main effects and all possible two-variable interactions was attempted. We assumed that interactions among three or more factors were negligible and that the variation in peak areas, y, could be adequately described by a polymonial in the experimental factors including linear terms, $b_i x_i$, and interaction or cross-product terms, $b_{ij} x_i x_j$; e represents the lack of fit to the model, or error:

$$y = b_0 + b_i x_i + b_j x_j + b_{ij} x_i x_j + e$$

To make regression coefficients comparable among different peaks, regression coefficients were calculated after peak areas were transformed to Z scores (Norušis, 1993). Otherwise, the magnitude of regression coefficients simply depends on the absolute peak areas. In other words, if the area of a specific peak is large, then the regression coefficient for the peak would automatically become large regardless of its importance, and *vice versa*.

Soy sauce production was performed according to the authentic method generally conducted by Japanese manufacturers (Fukushima, 1985). That is, after wheat and soybeans were treated with heat, *Aspergillus koji* starter was inoculated to the mixture of wheat and soybeans and then incubated for 2 days under a controlled temperature to make dry mash or koji. Completed koji was mixed with saline water in a 2 L round plastic bottle (12 cm i.d. × 24 cm height) to make wet mash or moromi, in which the concentration of NaCl was 17% (w/v). Moromi fermentation was continued for 6 months at 28 °C, during which time *Saccharomyces* yeast was added, if necessary. Moromi was periodically aerated by stirring using a stainless rod. After the completed moromi was filtered with Celite, the raw soy sauce obtained was pasteurized at 80 °C for 30 min and at 58 °C for 72 h. Nineteen soy sauce samples thus prepared were analyzed by HPLC.

HPLC Analysis. HPLC analysis were carried out using a Shimadzu liquid chromatograph system with an SIL-10A autoinjector, LC-10AD pumps, and an SPD-10A detector (Shimadzu Corp., Kyoto, Japan). Ten microliters of each soy sauce sample was directly injected into a Wakosil-II 5C18 HG column (4.6 mm i.d. × 250 mm; Wako Pure Chemicals Industries, Tokyo, Japan) fitted with a precolumn (4.6 mm i.d. × 30 mm) packed with the same material. The flow rate and column temperature were 0.8 mL/min and 15 °C, respectively, and monitored at 280 nm. The linear gradient chromatography was carried out by using two solvents such as 0.05% trifluoroacetic acid (TFA) (I) and 90% acetonitrile containing 0.05% TFA (II). Solvent I was kept at 100% for 20 min, and then solvent II was increased from 0 to 25% for 270 min and then further increased from 25 to 50% for 50 min. After each analysis, the column was washed with tetrahydrofuran. Chromatographic data were accumulated by Labchart 180 (System Instruments Co., Ltd., Tokyo, Japan).

Isolation of Key Components. For elucidating chemical structures of components of peaks that showed significance for differentiating HPLC profiles in pattern recognition, peak components were isolated and purified through the following steps. Three liters of soy sauce was extracted with 1.5 L of ethyl acetate three times at room temperature for 3 min. The upper layer was collected and concentrated to ≈9 mL by a rotary evaporator, and then 330 μL of the concentrate was repeatedly injected into preparative HPLC equipped with a Wakosil-II 5C18 HG column (10 mm i.d. × 50 mm + 10 mm i.d. × 300 mm) to separate peak components. The collected fractions containing objective peak components were concentrated and applied to rechromatography with preparative HPLC or redissolved with 50 mM acetic acid in water for further purification with Bio-Gel P2 column (20 mm i.d. × 380 mm, Bio-Rad Laboratories, Cambridge, MA) chromatography eluted with 50 mM acetic acid to remove higher molecular contaminants. The UV spectra of components were used as indicators in the isolation and purification process by monitoring with an SPD-M10A photodiode array detector (Shimadzu Corp., Kyoto, Japan).

Instrumental Analysis. Melting points of purified components were measured by a Yanagimoto micromelting point apparatus (Yanagimoto Co., Kyoto, Japan). UV spectra of their water solution were measured by a 557 double-wavelength double-beam spectrometer (Hitachi Ltd., Tokyo, Japan). IR spectra of the KBr disks were obtained with an FT/IR-7300 (Jasco Corp., Tokyo, Japan). NMR spectra of purified components were obtained with a JNM-FX200 (JEOL Ltd., Tokyo, Japan). The mass spectrum was obtained by an FRIT-FAB JMS-AX2000 LC/MS system (JEOL Ltd.).

Pattern Recognition. The HPLC data set of commercial soy sauce was analyzed by pattern recognition techniques using areas of 42 major peaks as predictor variables. In unsupervised pattern recognition, cluster analysis and factor analysis were performed using SPSS for Windows ver. 6.1

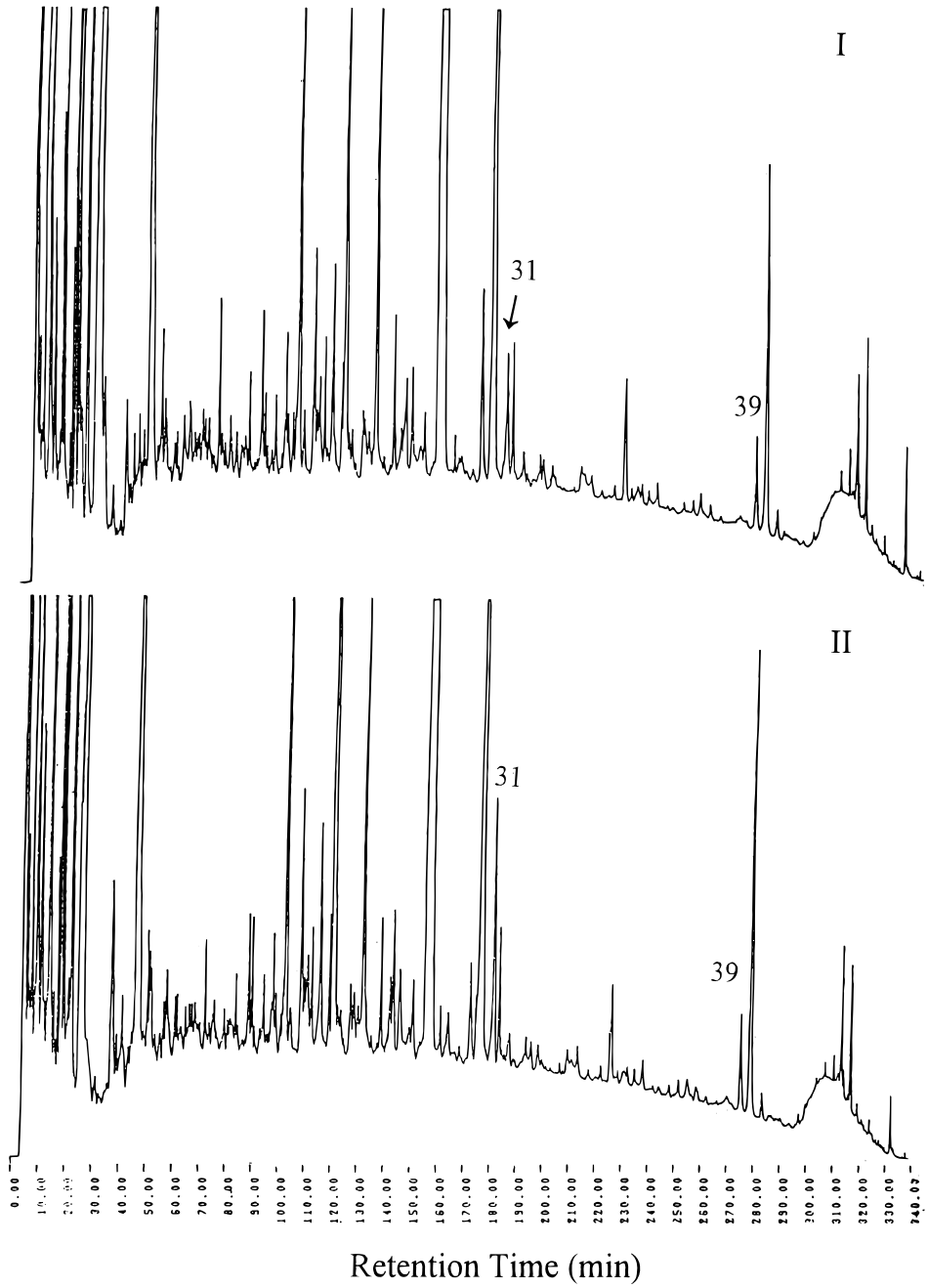


Figure 1. Typical RP-HPLC profiles of K-(I) and M-soy sauces (II).

(SPSS Inc., Chicago, IL). Clustering of samples was carried out on the basis of Euclidean distances among samples and Ward's method (Norušis, 1994), respectively. Prior to cluster analysis, each peak area was transformed into *z*-value, i.e., $z_{ij} = (x_{ij} - x_j)/s$. Here, *x* and *s* stand for the mean and standard deviation of areas of the *j*th peak. For supervised pattern recognition, soft independent modeling of class analogy (SIMCA) was carried out by Unscrambler ver. 5.3 (CAMO AS, Trondheim, Norway). The HPLC data set obtained from soy sauce produced on the basis of FFD was analyzed by MLR using SPSS for Windows ver. 6.1.

RESULTS

HPLC Analysis and Pattern Recognition. Typical HPLC profiles of K- and M-soy sauces are comparatively shown in Figure 1. The coefficients of variations, CV = (standard deviation/mean) × 100, in the 42 major peaks calculated from eight consecutively repeated HPLC analyses for the same sample deviated from 1 to 27%, but those in most peaks were <10%. All peaks were commonly found in all samples of K- and M-soy sauces. All HPLC profiles of soy sauce were clearly classified into two groups by cluster analysis (Figure 2) and factor score plots of factors 1 and 2 (Figure 3). Clear classification thus obtained without using any criterion for the grouping indicated the existence of indigenous difference between the compositions of K- and M-soy sauces. Contribution proportions of factors 1 and 2 were 44.2 and 16.2%, respectively, and so 60.4% of total variance was explained by the first two factors; however, contribution of factor 1 to the separation was far more predominant, as clearly described in Figure 3. Factor loadings of factors 1 and 2 are shown in Figure 4. Peaks 2 and 31 were highly positively loaded on factor 1, whereas loadings of peaks 8, 13, and 33 were highly negative. Next, SIMCA was performed using three principal components (PC) separately calculated from

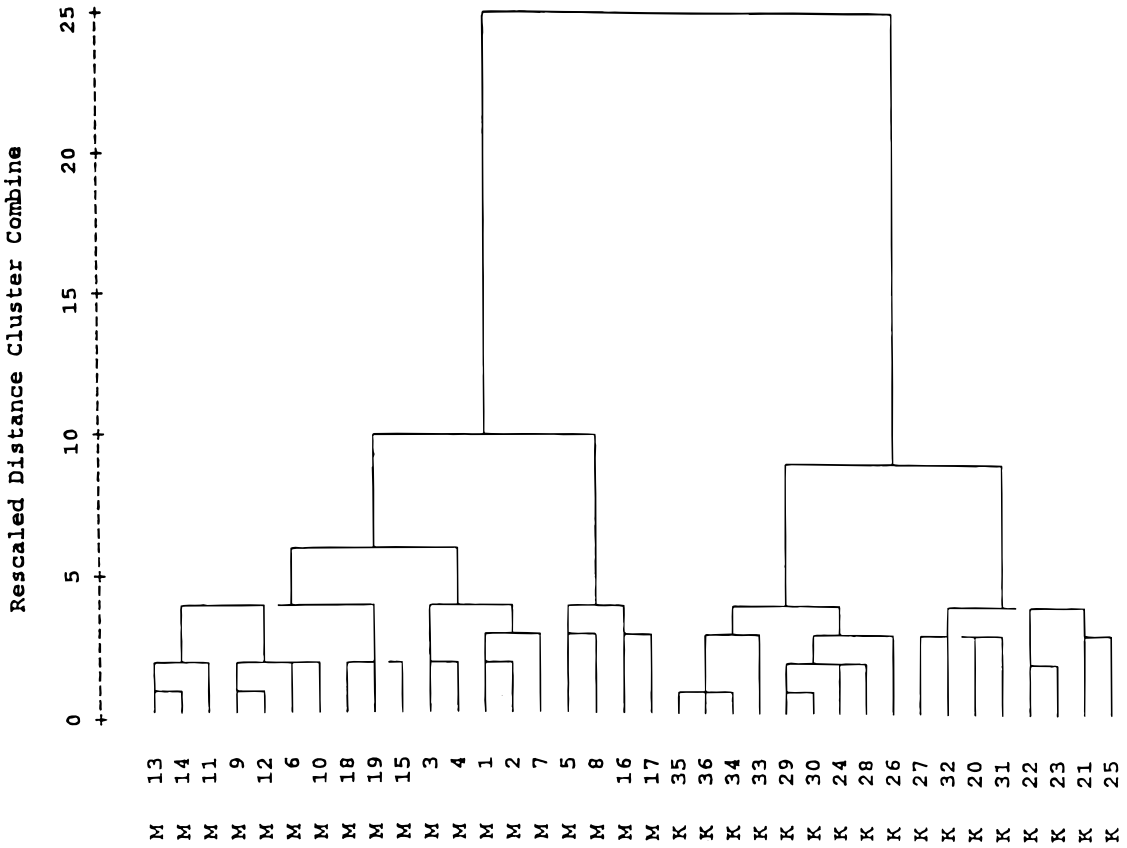


Figure 2. Clustering of soy sauce samples: K, K-soy sauce; M, M-soy sauce.

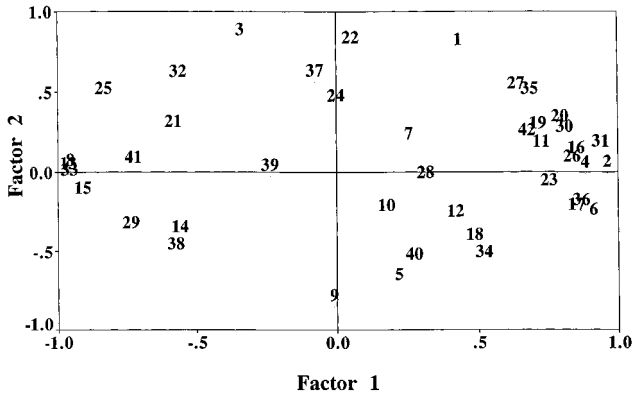


Figure 3. Factor score plot of soy sauce samples: K, K-soy sauce; M, M-soy sauce.

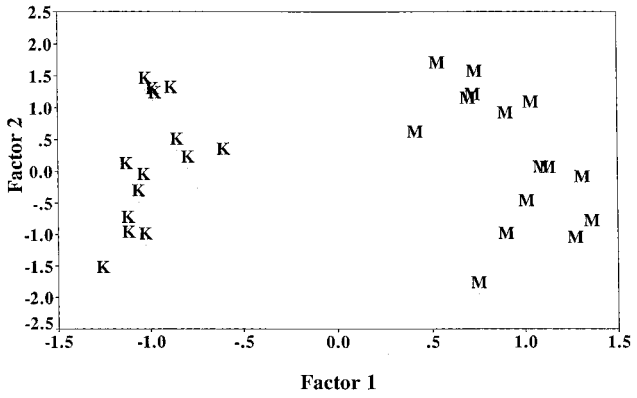


Figure 4. Factor loadings of factors 1 and 2.

each of the K- and M-soy sauce data sets, where 92.2 and 93.8% of variance contained in each K- and M-soy sauce were accumulated in the first three PCs. Figure 5 shows a Cooman's plot of samples; the abscissa represents the distance from the class model of K-soy sauce to each sample, while the ordinate represents that from the class model of M-soy sauce. From the Cooman's plot, it became clear that all samples were assigned to their original soy sauce groups with significance at the 5% level. Discrimination power indicated contribution of each peak to the unambiguous classification as described in Figure 6 (Sharaf et al., 1986). Among them, peaks 2, 8, 13, 31, and 33 showed considerably high discrimination power; however, peak 2 was a mixture of several components, and peak 33 was only a small one.

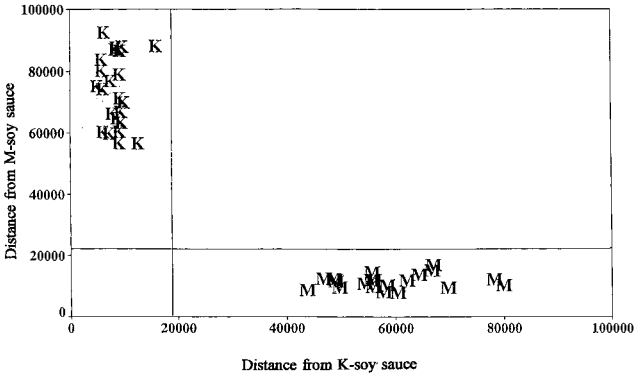


Figure 5. Cooman's plot of soy sauce samples in SIMCA: K, K-soy sauce; M, M-soy sauce.

Peaks Significant in FFD. Although MLR analysis was applied to 42 peaks to examine the relationship between factor levels and peak areas observed in 19 HPLC profiles, for simplicity, areas of peak 34 are given in Table 1 as an example to show the relationship between factor levels and peak areas. Regarding peak

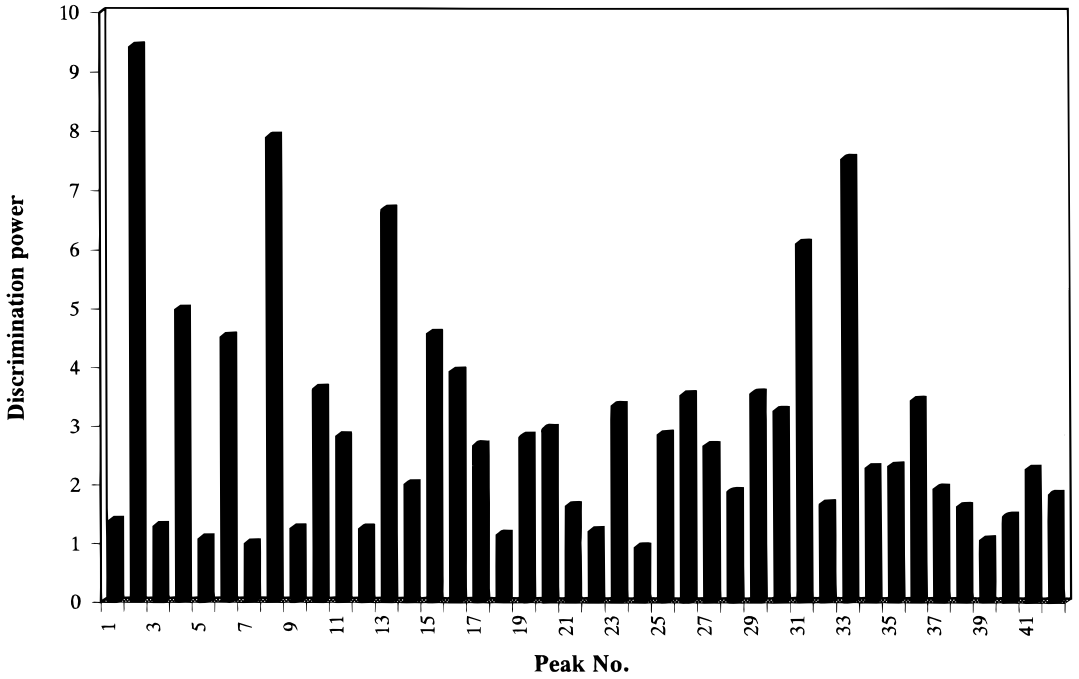


Figure 6. Discrimination power of 42 peaks.

34 shown in Table 1, regression coefficients (b_i, b_j, and b_{ij}) estimated by MLR with their statistical significance tested on the basis of their *t*-values are shown below. If a regression coefficient of a factor is large and statistically significant, then that factor or factors may be closely associated with the peak.

average area

$$b_0 = 344391 \ (t = 85.975, \ p = 0.0000)$$

main effects

$$b_1 = 0.784 \ (t = 27.81, \ p = 0.0001)$$

$$b_2 = -0.072 \ (t = -2.57, \ p = 0.0823)$$

$$b_3 = -0.206 \ (t = -7.31, \ p = 0.0053)$$

$$b_4 = 0.354 \ (t = 12.539, \ p = 0.0011)$$

$$b_5 = -0.041 \ (t = -1.441, \ p = 0.2452)$$

interaction effects

$$b_{12} = -0.091 \ (t = -3.212, \ p = 0.0489)$$

$$b_{13} = -0.161 \ (t = -5.693, \ p = 0.0107)$$

$$b_{14} = 0.165 \ (t = 5.859, \ p = 0.0099)$$

$$b_{15} = 0.102 \ (t = 3.609, \ p = 0.0366)$$

$$b_{23} = -0.046 \ (t = -1.642, \ p = 0.1991)$$

$$b_{24} = 0.128 \ (t = 4.545, \ p = 0.0200)$$

$$b_{25} = 0.139 \ (t = 4.913, \ p = 0.0161)$$

$$b_{34} = 0.092 \ (t = 3.249, \ p = 0.0475)$$

$$b_{35} = 0.095 \ (t = 3.362, \ p = 0.0437)$$

$$b_{45} = 0.049 \ (t = 1.734, \ p = 0.1813)$$

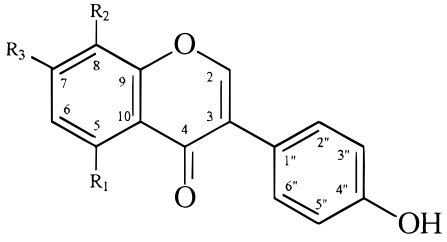
Adjusted square of multiple correlation (R^2) of this model was 0.987 and *F* value was 92.3, significant at 0.0016. Thus, FFD and MLR suggested that generation of peak 34, conjugated ethers of tartaric acid with daidzein (Kinoshita et al., 1997), was closely associated with whole soybeans. Besides peak 34, peaks 31 and 39 showed significantly positive association with whole soybeans (Table 2). Peak 31 was also suggested as one of the key components for differentiating K-and M-soy sauces in pattern recognition analysis. Fourteen peaks showed positive association with whole soybeans at 1% significance level, but seven peaks showed negative association. Among them, components of peaks 34, 37, and 38 have already been identified as new isoflavones conjugated with tartaric acid with ether linkage (Kinoshita et al., 1997). Therefore, we attempted to identify peaks 31 and 39.

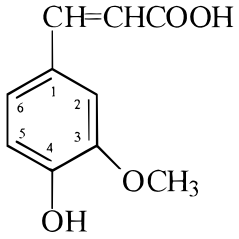
Identification of Peaks 31 and 39. Peak 31, selected as one of the most efficient for differentiating K- and M-soy sauces, was isolated, and 8.9 mg of purified component was obtained. The melting point of the pale yellow amorphous powder was 174 °C (dec) and UV_{max} in water was observed at 288 and 312 nm. ^{1}H NMR: δ (ppm) 3.89 (3H, s, CH_3-O-), 6.30 (1H, d, J = 15.9 Hz, $=CH-$), 6.80 (1H, d, J = 8.1 Hz, H-5), 7.05 (1H, dd, J = 2.0, 8.3 Hz, H-6), 7.15 (1H, d, J = 2.0 Hz, H-2), 7.57 (1H, d, J = 16.1 Hz, $=CH-$). ^{13}C NMR: δ

Table 2. Estimated Effects of Factors on Some Important Peaks

	peak[a]								
term	18	27	30	31	34	37	38	39	41
b_1[b]	0.310***	−0.504***	−0.403***	0.735***	0.784***	0.330***	0.659***	0.949***	0.896***
b_2	0.197**	−0.048	−0.036	0.077	−0.072	0.078*	−0.083	−0.016	0.042
b_3	0.248**	0.266**	0.142	−0.081*	−0.206**	−0.081*	−0.253*	−0.011	0.171**
b_4	0.810***	−0.430***	−0.431***	0.159	0.354**	0.869***	−0.262*	0.013	0.217**
b_5	−0.070*	−0.411**	−0.522***	0.006	−0.041	−0.094*	−0.082	−0.179*	−0.021
b_{12}[c]	0.073*	0.036	0.037	−0.012	−0.091*	0.037	−0.122	0.015	0.120*
b_{13}	0.050	−0.266**	−0.310**	0.000	−0.161*	0.004	−0.246*	−0.055	0.077*
b_{14}	0.007	−0.518***	−0.500***	−0.330*	0.165**	0.121**	0.135	0.331**	0.266**
b_{15}	−0.015	−0.043	−0.041	−0.171	0.102*	0.006	0.399**	−0.065	0.046
b_{23}	−0.098*	−0.020	−0.022	−0.020	−0.046	−0.071*	−0.168	−0.047	−0.075*
b_{24}	0.024	0.003	0.006	0.325	0.128*	0.106*	0.160	−0.157*	−0.130*
b_{25}	0.102*	−0.013	−0.067	−0.009	0.139*	−0.002	0.218	−0.033	−0.155**
b_{34}	−0.091*	−0.218**	−0.223**	0.198	0.092*	−0.058	0.190	0.034	−0.085*
b_{35}	−0.195**	0.085	0.089	0.068	0.095*	0.068*	0.137	0.025	−0.045
b_{45}	0.035	0.013	0.003	−0.114	0.049*	−0.092*	0.081	0.006	−0.117*

[a] * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. [b] Main effect of each factor. [c] Interaction effect of two factors: i = factors; 1, type of soybeans; 2, aeration; 3, yeast; 4, type of mold; 5, ratio of the amount of soybeans to that of wheat. See Materials and Methods for FFA.





Peak 31 (Ferulic Acid)

	R_1	R_2	R_3
Peak 34	H	H	-O-CH(COOH)-CH(OH)COOH
Peak 37	OH	H	-O-CH(COOH)-CH(OH)COOH
Peak 38	OH	OH	-O-CH(COOH)-CH(OH)COOH
Peak 39(daidzein)	H	H	H

Figure 7. Structures of peak components contributing to the differentiation of K- and M-soy sauce.

(ppm) 54.7 (CH_3−O−), 110.1 (CH), 114.5 (CH), 114.6 (CH), 121.9 (CH), 126.1 (C), 144.6 (CH), 147.5 (C), 148.5 (C), 169.3 (C═O). High-resolution FAB-MS: found, *m/z* 195.0658 $[M + H]^+$; calcd for $C_{10}H_{10}O_4$. IR_{max} (cm^{-1}): 3440, 2930, 1690, 1670, 1620, 1520, 1470, 1430, 1330, 1280, 1210, 1180, 1120, 1040, 950, 850, 800, 750. Comparison of the physical constants and ^1H NMR, ^{13}C NMR, IR, MS, and UV spectra of purified peak 31 with those for authentic ferulic acid strongly suggested the identity of the two compounds. Finally, an aliquot of purified compound with authentic ferulic acid was co-injected into HPLC and their elution at the same retention time was confirmed.

The component of peak 39 closely associated with whole soybeans in FFD was isolated, and 6.7 mg of purified component was obtained. The UV_{max} in water of the white amorphous powder was observed at 246 and 300 nm. ^1H NMR (200 MHz, DMSO-d_6): δ (ppm) 6.78 (2H, d, J = 8.79 Hz, H-3′ and -5′), 6.80 (1H, d, J = 2.2 Hz, H-6), 6.89 (1H, dd, J = 2.2, 8.79 Hz, H-6), 7.37 (2H,d, J = 8.79 Hz, H-2′ and -6′), 7.93 (1H, d, J = 8.79 Hz, H-5), 8.2 (1H, s, H-2), 8.48 (1H, br, OH). ^{13}C NMR (50 MHz, DMSO-d_6): δ (ppm) 101.8 (CH), 114.7 (CH), 115.6 (CH), 118.2 (C), 122.2 (C), 123.6 (C), 126.9 (CH), 129.7 (CH), 152.7 (CH), 156.9 (C), 157.0 (C), 163.9 (C), 172.8 (C═O). IR_{max} (cm^{-1}): 3230, 1630, 1600, 1560, 1520, 1460, 1390, 1310, 1280, 1240, 1200, 1100, 1050, 1020, 960, 900, 890, 850, 820, 790, 780, 550, 490. By comparing ^1H NMR, ^{13}C NMR, IR, and UV spectra of the purified component of peak 39 with those of authentic compounds, it was identified as daidzein. Finally, an aliquot of the purified component with authentic daidzein was co-injected into HPLC and their coelution was confirmed.

DISCUSSION

Chemometric pattern recognition applied to HPLC profiles showed effectiveness for classification of samples according to soy sauce produced from two types of soybeans. The difference in the HPLC profiles was very distinctive between K- and M-soy sauces in this research and may be because the soy sauce samples were produced by the same manufacturer. For soy sauce produced through the same process conditions at the same manufacturer, the effects of other factors can be negelected and so differences ascribable to types of soybeans would clearly appear in the composition of products. However, differentiation of K- and M-soy sauces produced by various manufaturers should be the ultimate objective in this research.

The methodology using FFD did not aim only to identify peaks produced under the strong influence of soybean types but to clarify relationships between major nonvolatile components and all potential factors involved in the production process. Actually, several peaks were suggested to be produced under the strong interactive effects of soybeans and other factors as indicated in Table 1; we selected only peaks showing large main effects of soybean types for structure elucidation in this study because other factors may differ from manufacturer to manufacturer. Furthermore, the

contribution of other factors to the generation of peak components is another subject in soy sauce study, and so we focused on only the main effect of soybean types on components.

By using instrumental analyses, purified components of some peaks selected as considerably important for the differentiation of two types of soy sauce by pattern recognition and FFD were identified as ferulic acid, daidzein, and three isoflavone derivatives (Figure 7). Among them, ferulic acid was derived from lignin glycoside contained in wheat bran during the koji stage (Yokotsuka, 1953), and then liberated feruic acid was further transformed into 4-ethylguaiacol by Torulopsis yeast during the moromi stage (Asao et al., 1958). On the other hand, daidzein is liberated from its isoflavone glycoside originally contained in soybeans during the koji stage (Coward et al., 1993; Wang and Murphy, 1996). The other isoflavones identified as derivatives of daidzein and genistein were also produced during the soy sauce manufacturing process (Kinoshita et al., 1997). The difference in composition between the two types of soy sauce might derive from relatively poor growth of microorganisms and transfer of hydrophobic compounds to rich soybean oil in the M-soy sauce moromi. However, components of some peaks that showed large discrmimination power in SIMCA have not been identified yet. Identification of these components, which will reveal the implication of ferulic acid and isoflavones in differentiating two types of soy sauce, is now underway.

ABBREVIATIONS USED

FFD, fractional factorial design; MLR, multiple linear regression; SIMCA, soft independent modeling of class analogy; RP-HPLC, reversed phase HPLC.

ACKNOWLEDGMENT

We sincerely thank Ms. A. Nasu for NMR spectral measurements and Mr. K. Matsuura at JEOL Ltd. for mass spectral measurements.

LITERATURE CITED

Aishima, T. Relation between gas chromatographic profiles of soy sauce volatiles and organoleptic characteristics based on multivariate analysis. In *Instrumental Analysis of Foods*; Chalarambous, G., Inglett, G., Eds.; Academic Press: New York, 1983; Vol. 1, pp 37−56.

Aishima, T.; Nakai, S. Chemometrics in flavor research. *Food Rev. Int.* **1991**, *7*, 33−101.

Asao, Y.; Yokotsuka, T. Studies in flavorous substances in soy sauce. Part XVII. Guaiacyl components in soy sauce. (1) *Nippon Nōgeikagakku Kaishi* **1958**, *27*, 622−628.

Coward, L.; Barnes, N. C.; Setchell, K. D. R.; Branes, S. Geistein, daidzein, and their β-glycoside conjugates: Antitumor isoflavones in soybean foods from American and Asian diets. *J. Agric. Food Chem.* **1993**, *41*, 1961−1967.

Fukushima, D. Fermented vegetable protein and related foods of Japan and China. *Food Rev. Int.* **1985**, *1*, 149−209.

Ishihara, K.; Honma, N.; Matsumoto, I.; Imai, S.; Nakazawa, S.; Iwafuchi, H. Comparison of volatile components in soy sauce (*koikuchi shoyu*) produced using *Aspergillus sojae* and *Aspergillus oryzae*. *Nippon Shokuhin Kogaku Kaishi* **1996**, *43*, 1063−1074.

Kinoshita, E.; Yamakoshi, J.; Kikuchi, M. Purification and identification of an angiotensin I-converting enzyme inhibitor from soy sauce. *Biosci., Biotechnol. Biochem.* **1993**, *57*, 1107−1110.

Kinoshita, E.; Ozawa, Y.; Aishima, T. Novel isoflavone derivatives connecting with tartaric acid and playing key roles for differentiating Japanese soy sauce. *J. Agric. Food Chem.* **1997**, *45*, 3753−3759.

Morgan, E. D. *Chemometrics: Experimental Design*; Wiley: Chichester, U.K., 1991; pp 151−181.

Norušis, M. J. *SPSS for Windows, Base System User's Guide*, release 6.0; SPSS: Chicago, IL, 1993; p 342.

Norušis, M. J. Cluster analysis. In *SPSS Professional Stasistics 6.1*; SPSS: Chicago, IL, 1994; pp 83−109.

Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*; Wiley: New York, 1986; p 245.

Wang, G.; Kuan, S. S.; Francis, O. J.; Ware, G. M.; Carman, A. S. A simplified HPLC method for the determination of photoestrogens in soybean and its processed products. *J. Agric. Food Chem.* **1990**, *38*, 185−190.

Wang, H.-J.; Murphy, P. A. Isoflavone content in commercial soybean foods. *J. Agric. Food Chem.* **1994**, *42*, 1666−1673.

Wang, H.-J.; Murphy, P. A. Mass balance study of isoflavones during soybean processing. *J. Agric. Food Chem.* **1996**, *44*, 2377−2383.

Yokotsuka, T. Studies on flavous substances in soy. Part 12. Flavorous substances in row soy. *Nihon Nōgeikaga Kukaishi* **1953**, *27*, 549−553.

Received for review June 5, 1997. Revised manuscript received November 13, 1997. Accepted December 7, 1997.

JF970461C