CLASSIFICATION OF CHILLI SAUCES: MULTIVARIATE PATTERN RECOGNITION USING SELECTED GCMS RETENTION TIME PEAKS OF CHILLI SAUCE SAMPLES

Low Kah Hin*[1], Sharifuddin M. Zain[1], Mohd. Radzi Abas[1], Mustafa Ali Mohd[2].

[1]*Department of Chemistry, Faculty of Science, University Malaya, 50603, Kuala Lumpur, Malaysia*
[2]*Shimadzu-UMMC Centre for Xenobiotics Studies, University of Malaya Medical Centre, University Malaya, 50603, Kuala Lumpur, Malaysia*

Abstract

As a preliminary work on the possibility of separating classes of chili sauces based on taste or customer preferences, organic compounds from different kinds of chili sauces of various brands were separated and analyzed by gas chromatography/mass spectrometry (GC/MS). It was found that these organic compounds do form a basis for separation of different types of sauces. The similarity and dissimilarity of chromatograms due to the organic composition of the chili sauces were explored by multivariate pattern recognition techniques based on cluster analysis (CA) and principal component analysis (PCA). Both CA and PCA results exhibit four linearly separable classes, namely general sauces, hot sauces, sauces with benzoic acid and sauces with garlic. It was concluded that by using chosen retention peaks in the chromatograms of various sauce samples as multivariate features, CA and PCA can be successfully used to reveal the natural clusters existing in chili sauces according to their organic composition.

Introduction

Originally, chilli sauces were produced at home. Today, this practice still preservere but most chilli sauces available nowadays are mass manufactured in factories. Today, much effort is spent to ensure that mass manufactured sauces resemble as close as possible home made ones. Some brands do this better than others. The ingredients used in chilli sauces are very similar to those used in tomato ketchup. However, the methods of preparation and the amount of the ingredients used may vary considerably. The primary ingredients in chili sauce are red tomatoes and chilies. The usual sweetener is sugar (sucrose). The proportions of the various spice ingredients are not standardized between manufacturers [1].

Humans are very good at perceiving similarities and differences between objects of different shapes. The goal of pattern recognition in analytical chemistry is finding similarities and differences between chemical samples based on measurements made on the sample [2]. Therefore, pattern recognition appears to be a useful tool to authenticate foodstuffs according to their quality/variety/brand, when a set of samples whose classification is known a priori is available.

This study initially involved data collection via instrumental analysis, basically GC/MS. By using only chromatographic techniques, we able to separate the chemical compounds in the sauces, which are mainly capsaicinoid compounds, the pungent principle of capsicum fruits, benzoic acid which is a preservative, and other compounds due to the ingredients of the sauce.

Capsaicinoid compounds are a group of pungent compounds found mostly in capsicum fruits - the main components are acid amides of vanillylamine and $C_9 - C_{11}$ branched-chain fatty acids. There are five naturally occurring capsaicinoids, which have been reported, namely capsaicin, nordihydrocapsaicin, dihydrocapsaicin, homocapsaicin and homodihydrocapsaicin. Of these, capsaicin and dihydrocapsaicin are the major compounds of most capsicum species. Therefore these compounds have been considered as major indicators of chili product quality [3].

Techniques using GC/MS provide accurate and efficient analysis of content and type of capsaicinoids present in a chili sauce sample. These results are probably the major factors in determining the qualities/variety/brand of those sauces, and are the most important parameters used in this statistical study.

In this work, different multivariate chemometric methods been employed in the exploration; this primarily involve the use of cluster analysis (CA) and principal component analysis (PCA). Cluster analysis, which involving the search for natural grouping among samples is a preliminary way to study data sets and to discover the structure residing in them. PCA is used to transform the original data matrix ($X_{n \times m}$) into a product of two matrices, once which contains information about the objects (Scores matrix, $S_{n \times m}$) and the other about the variables (Loadings matrix, $L_{n \times m}$) [4]. PCA enables one to study data structure in reduced dimensions.

Experimental

Sample preparation
About 1 mL of sauce was weighed and extracted with 5 mL dichloromethane in a vial for overnight. Then 1 mL of sauce extracts was drawn and filtered using a 0.4μm nylon filter membrane with a syringe filter into a small glass vial. The filtrate was dried under nitrogen gas, and reconstituted with 200 μL of chloroform and was later used for a GC/MS analysis with injection volume of 1 μL.

GC/MS analysis
Samples were assayed using a Gas Chromatograph (Shimadzu GC-17A) with a Mass Spectrometer (Shimadzu QP-5000). A 30 m × 0.247 mm × 0.25 μm capillary column, with a stationary phase of (5% - phenyl) – methyl polysiloxane (J & W Scientific, DB-5), was used. For analysis, the initial temperature was 60°C held for 1 min. The temperature was increased from 60 to 280°C at the rate of 10°C/min and held at 280°C for 22 min. The total run time was 38 min. The injection volume was 1.0 μL and the total flow of carrier gas was 39.5 mL/min helium.

Data processing
Data preprocessing is a very important part of any chemometric data analysis project. It consist of whatever mathematical manipulation of data prior to primary analysis. From the obtained GC profiles, 12 marked peaks were considered as the multivariate variables. They are the compounds eluting at retention times, 7.6 (1), 9.2 (2), 12.2 (3), 17.0 (4), 19.2 (5), 19.5 (6), 21.0 (7), 21.2 (8), 24.4 (9), 25.4 (10), 25.7 (11), and 36.9 (12) minutes. These peaks were examined for their mass spectra and identification of these peaks were attempted. The peaks' area per weight of sample were preprocessed, then arranged the in an X-matrix (201 x 12). The values of each row represent the measurement of different variable for each sample, and the values of each column represent the measurement of different samples for a particular variable. Each replicate was treated as an individual sample in the data matrix, which then undergo CA and PCA. All these were done by combination of several statistical software packages such a Microsoft® Excel, Teach/Me®-Data Analysis, JMP® and JMP IN®.
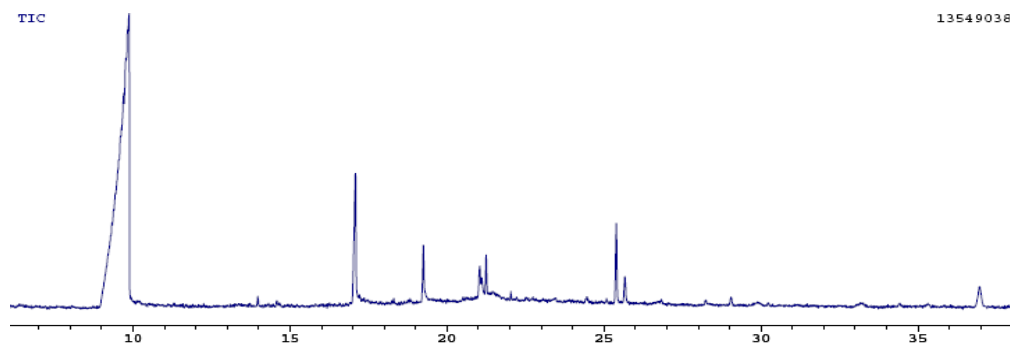

Figure 1: Typical GC/MS chromatogram of a chili sauce extract.

Table 1: Peak areas at particular retention times for a part of sauce samples

Retention Time	GCS0036 S01a	GCS0035 S01b	GCS0035S01c	GCS0035 S02a	GCS0035 S02b
7.5	782155	1387435	1613369	801396	968122
9.2	0	0	0	0	0
12.2	186056	597293	819216	386139	385629
17.0	28568683	23716034	37251905	18332483	14260443
19.2	12228225	6519874	11069449	4543486	3855149
19.5	0	0	0	0	0
21.0	5353903	4026651	6242748	923887	1528811
21.2	9774055	10841376	14612206	8586443	9105060
24.4	1426550	1902577	2076992	1914108	1396709
25.4	10714811	13662227	19084453	10339062	11473199
25.7	6485459	8002463	11712994	6540138	7115043
36.9	2731411	4064320	6083270	2461134	2353073

In this work, CA was applied to the autoscaled data, the sample were calculated on the basis of Euclidean distances, while Ward hierarchical agglomerative method was used to establish the clusters. Next, PCA was performed on the unscaled data, used to provide data structure in a reduced dimension, retaining the maximum amount of variability present in the data.

Results and discussion

Mass spectra analysis
From the total ion chromatograms of chili sauces extracts, compounds eluting at particular retention times were examined for their mass spectra. Compounds 1, 3, 6 and 8 were not resolved. The m/z values obtained from mass spectra of the molecular ions (M^+) and fragment ions of selected compounds are shown in Table 2.

Table 2: m/z values of selected compounds.

	Retention time / Compound		Mass/charge, m/z					Molecular Ion, M^+
			Fragment ions					
1	7.5	unresolved						
2	9.2	Benzoic acid	51	65	77	94	105	122
3	12.2	unresolved						
4	17.0	Myristic acid	60	73	85	129	185	228
5	19.2	Palmitic acid	60	73	85	129	213	256
6	19.5	unresolved						
7	21.0	Linoleic acid	67	81	95	109	123	280
8	21.2	unresolved						
9	24.4	Nordihydrocapsaicin	43	122	137	152	195	293
10	25.4	Capsaicin	43	122	137	152	195	305
11	25.7	Dihydrocapsaicin	43	122	137	152	195	307
12	36.9	Vitamin E	43	57	71	165	205	430

As mentioned before, the pungent capsaicinoids have been considered as major indicators of chilli product quality. Typical mass spectra of dihydrocapsaicin is shown in *Figure 3*. The molecular ions (M^+) of the capsaicinoid compounds were confirmed from their mass spectra. The base peak of each compound was found to be common at m/z = 137. Other fragment ions had m/z = 43, 122, 152 and 195.
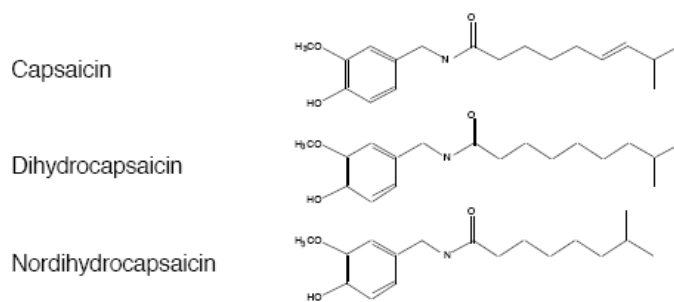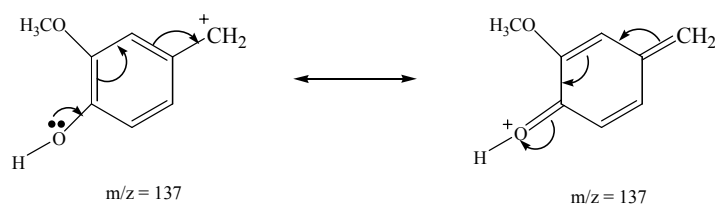
Figure 2: The capsaicinoids commonly found in chilies.

A possible empirical formula for fragment ion m/z = 137 from isotopic abundance calculation is $C_8H_9O_2$. An achievable fragment ion is shown below as:-

m/z = 137   m/z = 137

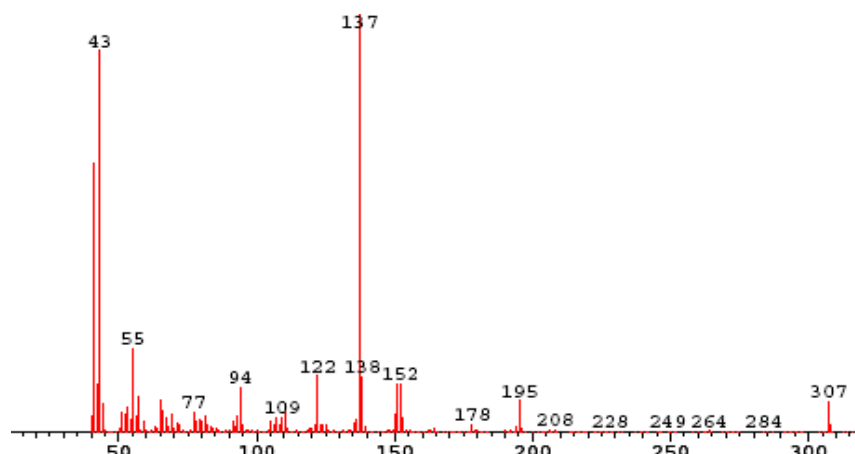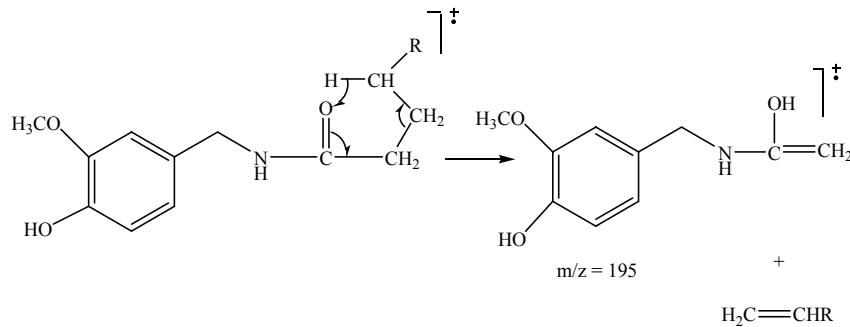The resonance stabilization of this fragment ion suggested its occurrence as the base peak [5].

Figure 3: Mass spectrum of dihydrocapsaicin.

The isotopic abundance calculation of the fragment ion m/z = 195 suggested an empirical formula of $C_{10}H_{13}O_3N$. A possible fragment ion is shown in the decomposition reaction below [6]:-

m/z = 195

$H_2C \!=\! CHR$

Fragment ions with m/z = 153 and 152 were perhaps due to the ion decomposition reaction below:-
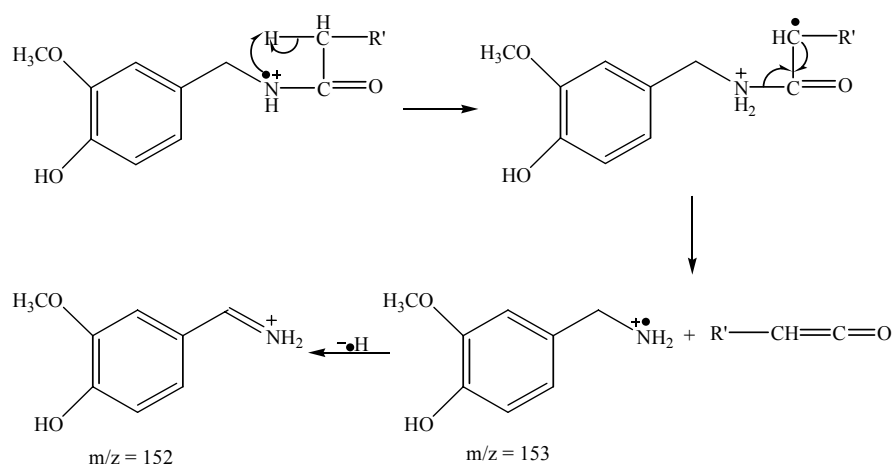
Multivariate correlation analysis

From the correlation matrix for GC peaks, peak no. 1 and no. 3 correlate strongly ($r = 0.9861$); followed by inter-correlation between peaks no. 9, nordihydrocapsaicin, no. 10, capsaicin and no. 11, dihydrocapsaicin. Others correlate moderately or weaker. The strong correlation between nordihydrocapsaicin, capsaicin and dihydrocapsaicin is expected because these compounds are the major pungent capsaicinoids in capsicum fruits. In other word, we can easily conclude that capsaicinoids present in chili sauces or natural chili are in a constant ratio. On the other hand, we suspect that the unknown peaks no.1 and no.3 also show similar characteristics as the capsaicinoids, but this needs further investigation.

Cluster analysis

CA is a well known technique of data analysis, commonly applied before other mutivariate procedures owing to its unsupervised character, that reveals the natural clusters existing in a data set on the basis of the information provided for the measured variables. The results obtained for the case at hand, using the Ward's method and Euclidean distance are presented as a dendrogram in Figure 4. At distance 700, 3 clusters that can be identified as 'similar' were found: from the top, the first cluster is composed of general sauces including a sub-cluster of hot sauces. The second cluster is a group made up of sauces containing benzoic acid as preservative. The last cluster is those sauces with the addition of garlic.
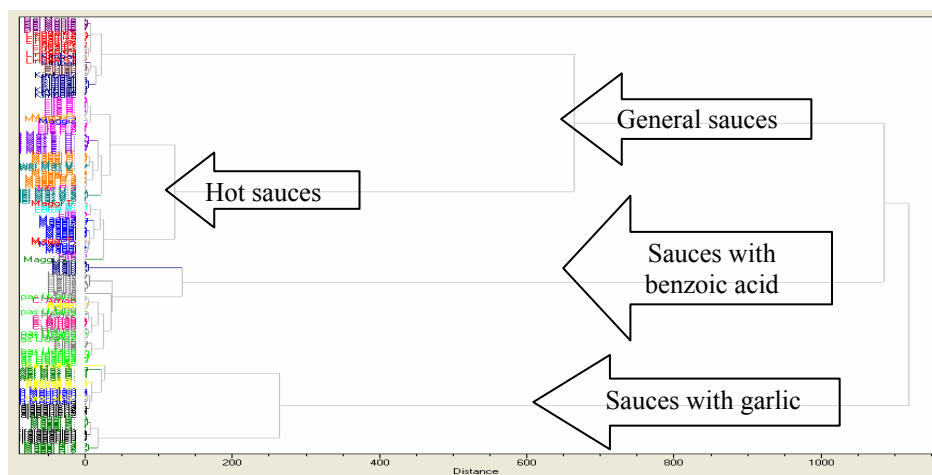

Figure 4: Dendrogram of Cluster analysis.

Principal component analysis

By performing PCA on the dataset, which was initially preprocessed from GC/MS data of the chili sauces extracts, a series of principal components (PCs) could obtained. Each PC is associated with an eigenvalue. PC1 has the largest eigenvalue and carries the largest variance of the original data, and subsequent PCs carry variance in a decreasing order.
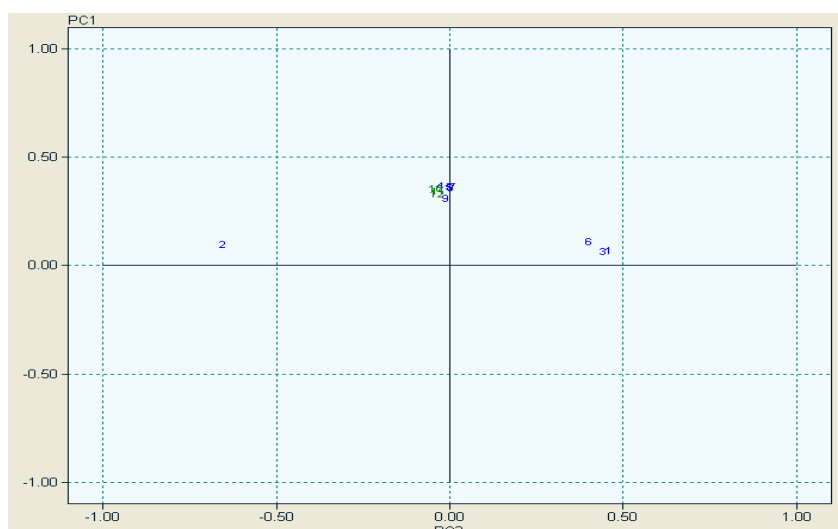
Figure 5: Loading Plot: PC1 vs. PC2.

Loadings graph (Figure 5) is used to determine which variables are important for describing the variation in the original data set, and there are no variables with loading close to ±1, revealing that no PC is closely aligned with any peaks. For PC1, peaks no. 1, 2, 3 and 6 do not contribute much to the variance described by PC1. For PC2, it is the other way round; peaks no. 4, 5, 7, 8, 9, 10, 11 and 12 have near zero loading. In addition, peak no. 2 show high negative loading value, which indicates benzoic acid is a meaningful parameter for PC2.

The final result of PCA, Figure 6, shows a plane spanned by first two PCs, representing 95.87 of the variation in the data. There are crowds of points discriminated by PC2 while PC1 primarily determines the spread of scores. Four major separable classes namely general sauces, hot sauces, sauces with benzoic acid and sauces with garlic can be observed from the plot.
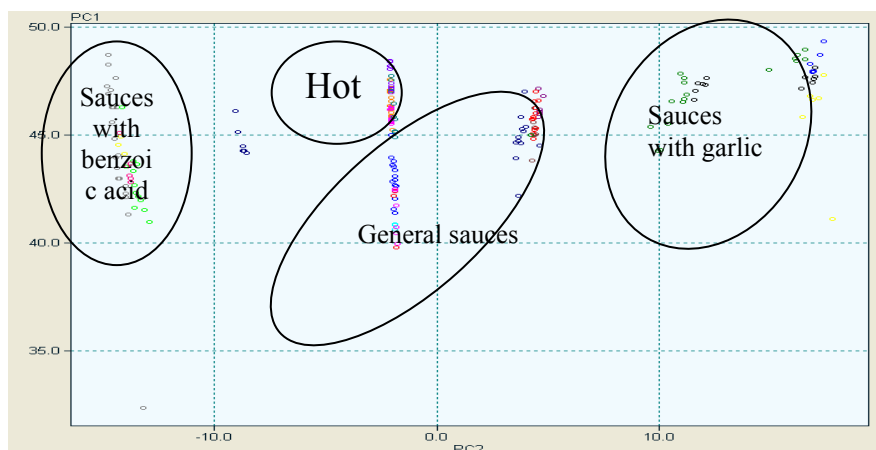

Figure 6: Score Plot of PC1 vs. PC2.

As had been discussed, peaks no. 1, 2, 3 and 6 are the dominant features in PC2. From scores plot, it is demonstrated that those sauces with benzoic acid clusters on the negative area of axis PC2 corresponding to its' negative loading, while those sauces with garlic are on the positive side. These indicates that the unknown peaks no. 1 and 3 should correspond to the garlic content of the sauces.

Conclusion

In this study, the feasibility of using just GC profiles to distiguish between chili sauces variety was investigated with multivariate chemometric methods. CA and PCA methods have demonstrated similar results; that using certain peaks in the GC profiles as multivariate parameters we are able to reveal the natural clusters existing in the chili sauce samples studied.

References
1. Grading Manual for Chili Sauce, United State, Department of Agriculture, Agricultural Marketing Service, Food and Vegetable Division, Processed Product Branch, 1954.
2. Beebe, K. R., Pell, R. J. and Seasholtz, M. B. 1998. Chemometrics: A Practical Guide, John Wiley & Sons, Inc., New York.
3. Rachaneewan Karnka, Mongkon Rayanakorn, Surasak Watanesk and Yuthsak Vaneesorn. 2002. Optimization of High Performance Liquid Chromatographic Parameters for Determination of Capsaicinoid Compounds Using the Simplex Method. Analytical Sciences. 18, 661-665.
4. Padin, P. M., Pena, R. M., Garcia, S., Iglesias, R., Barro, S. and Herrero. 2001. Characterization of Galician (N.W. Spain) quality brand potatoes: a comparison study of several pattern recognition technique. Analyst, 126, 97-103.
5. Ahmad Ab. Wahab. 1984. A Combined Gas Chromatography-Mass Spectrometry (GCMS) study on Pungent Principles in Chillies (Capsaicum Species), MARDI Res. Bull. 12,3, 290-297.
6. McLaferty, F.W. 1980. Interpretation of mass spectra. 3rd Ed. Mill Valley, California: University Science Books.