A Chemometric Strategy for the Classification of Lecithins According to Process Performance

Peter Kaufmann*, **Urban Olsson** and **Bengt G. Herslof**
Department of Analytical Chemistry, University of Stockholm

Lecithins are widely used as multipurpose additives in the pharmaceutical and food industries. This implies the need for an analytical means for the assessment of process performance prior to full-scale processing. A general methodology was developed for the classification of lecithins with respect to this property. The strategy developed utilizes pattern recognition methods, fatty acid composition of the phospholipid classes (phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol) and lipid class analysis by high performance liquid chromatography to identify the lecithins with acceptable performance, e.g., emulsifying behavior.

KEY WORDS: Classification, gas chromatography, high performance liquid chromatography, lipids, multivariate, phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, phospholipids and thin-layer chromatography.

The term "lecithin" generally refers to a mixture obtained from crude vegetable oils in the degumming process (treatment with water). In this process the phospholipids, partial glycerides, sterols, glycosylglycerides and other substances are removed in an emulsion.

The most commonly marketed lecithin originates from soybean oil. It contains phospholipids (60%), triglycerides (35%) and other lipids (such as partial glycerides, glycosylglycerides, sterols) and nonlipid materials such as pigments and carbohydrates. Commercial lecithins are widely used as multifunctional additives in a variety of foods, but they also have found a significant and growing use in nonfood applications (1).

Due to the complexity and variability of the lecithin mixture it is difficult to correlate functional properties to discrete constituents in the mixture or its composition. The situation is obscured even further by the fact that the chemical composition is often analyzed by a combination of several different methods and techniques.

Multivariate analysis of chromatographic data has been used in studies to classify olive oils according to their geographical origin (2), polymer batches in quality control (3), honeybees according to subspecies (4), ants from different colonies (5), fungi and molds according to specie and strain (6–8), brain tissue as normal or tumoral (9), and diagnosis of liver disorders (10), among numerous other applications.

Many papers in the literature describe the principles and theory behind the use of different methods of pattern recognition such as clustering, linear discriminant analysis, linear learning machine, nonlinear mapping (11,12), principal components and partial least squares regression (12–18).

*To whom correspondence should be addressed at: Karlshamns LipidTeknik AB, P.O. Box 15200, S-104 65 Stockholm, Sweden.

The problem addressed in this study was one of screening different batches of commercially available lecithins, e.g., classifying the lecithins according to emulsifying performance in an industrial food process, in order to ensure a continuously high process quality.

Conventional quality-control data, such as fatty acids of the total material, etc., showed seemingly spurious between-batch variation. There were no clear trends that could account for the great differences in emulsifying performance observed. This implied that either the data measured were irrelevant to the problem or the variables chosen were highly interdependent. As the data was analyzed using traditional univariate analysis, no covariance could be identified.

The objectives were therefore to find a relevant set of variables and to interpret the results using multivariate statistical analysis. Variation in the fatty acid composition of the isolated phospholipid classes from lecithin, analyzed by a combination of preparative thin-layer chromatography and gas chromatography, would most probably contain relevant information. Furthermore, the data from analytical high performance liquid chromatography (HPLC) of the lipid classes could contain useful information if analyzed using multivariate methods.

MATERIALS AND METHODS

Soybean lecithins, of commercial grade and of various origin (Lucas Meyer, Federal Republic of Germany; and Stern Chemie, Federal Republic of Germany) were obtained from Karlshamns AB (Karlshamn, Sweden), These were categorized according to results obtained from a full scale industrial emulsification process: 1, good performance; 2, good at double dose; and 3, no good at any dose.

High performance liquid chromatography. The lipid classes (glycerides, phospholipids, etc.) were separated by straight-phase HPLC (19), Licrospher 100 Diol 5 μm 250 × 4 mm. A solvent system of hexane:isopropanol:water:acetic acid (80:10:10, w/w/w) was used, and detection was by a light-scattering detector (ACS 750/14, Mass Detector, England).

Thin-layer chromatography. The phospholipid classes—phosphatidylcholine (PC), phosphatidylinositol (PI) and phosphatidylethanolamine (PE)—were isolated by preparative TLC. This was done on preparative plates, Silica 60, 0.5 mm 20 × 20 with concentration zone (Merck, Darmstradt, Federal Republic of Germany). The solvent system used was chloroform:methanol:isopropanol:0.25% aqueous KCl: ethylacetate, (30:9:25:6:18, v/v/v/v/v). Three hundred microliters of a 10% w/w solution of the lecithin, dissolved in chloroform, was applied as a band, the plates were developed twice (Fig. 1) and the bands were detected by spraying the plates with Zinzadze's reagent (20). The bands were scraped off the plates and extracted with chloroform:methanol, 1:1 (v/v). These

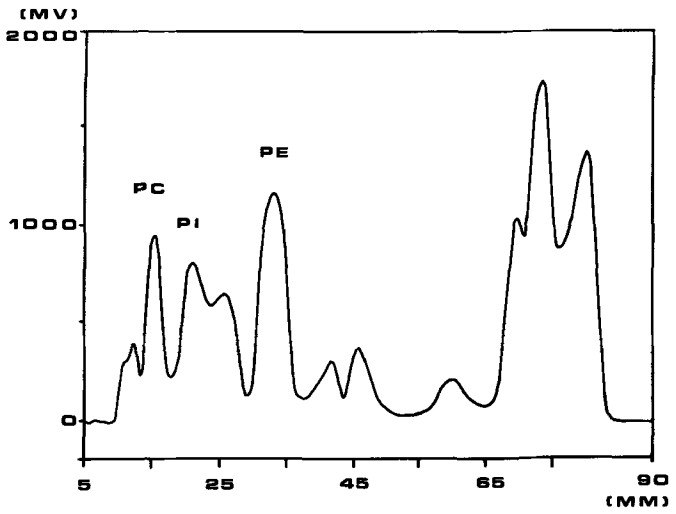

FIG. 1. Typical TLC-chromatogram of the lecithins, showing the main phospholipid classes. Conditions as in Materials and Methods.

TABLE 1

HPLC variables[a]	GC variables[a]
1. Tri-, diglycerides	19. Palmitic acid (PC)
2. Monoglycerides	20. Stearic acid (PC)
3. Phosphatidic acid	21. Oleic acid (PC)
4. Sterylglycosides	22. *cis*-Vaccenic acid (PC)
5. 4.6, Unidentified peak	23. Linoleic acid (PC)
6. Phosphatidylethanolamine	24. α-Linolenic acid (PC)
7. Phosphatidylcholine	25. Palmitic acid (PI)
8. Lysophosphatidylethanolamine	26. Stearic acid (PI)
9. Phosphatidylinositol	27. Oleic acid (PI)
10. Lysophosphatidylcholine	28. *cis*-Vaccenic acid (PI)
11. Sucrose	29. Linoleic acid (PI)
12. 11.8, Unidentified peak	30. α-Linolenic acid (PI)
13. 12.8, Unidentified peak	31. Palmitic acid (PE)
14. 14.8, Unidentified peak	32. Stearic acid (PE)
15. 15.5, Unidentified peak	33. Oleic acid (PE)
16. Raffinose	34. *cis*-Vaccenic acid (PE)
17. Stachyose	35. Linoleic acid (PE)
18. Sum of the above (%)	36. α-Linolenic acid (PE)

[a]Variables used in the multivariate analysis. Those underlined were chosen on the basis of their modeling power for the class model "good".

TABLE 2

Phospholipid class	Fatty acid	Mean content (area %)	Standard dev.
Phosphatidylcholine	16:0	11.85	1.68
	18:1(n-9)	6.58	1.58
	18:1(n-7)	0.18	0.29
	18:2(n-6)	71.85	2.18
	18:3(n-3)	5.77	1.04
Phosphatidylethanolamine	16:0	15.70	3.71
	18:0	1.47	0.38
	18:1(n-9)	7.01	3.48
	18:1(n-7)	0.72	0.24
	18:2(n-6)	68.81	6.49
	18:3(n-3)	5.70	1.03
Phosphatidylinositol	16:0	29.72	2.57
	18:0	4.22	0.53
	18:1(n-9)	5.22	1.69
	18:1(n-7)	0.91	0.09
	18:2(n-6)	53.14	3.11
	18:3(n-3)	5.79	0.92

fractions were subjected to analytical HPTLC using the same solvent system as above, and scanned by photodensitometry (Desaga CD60) utilizing iodine vapors to check for purity.

Gas chromatography. The fatty acid composition of the isolated phospholipid classes was determined by transesterification to form methyl esters (FAME) and subsequent analysis by capillary gas chromatography (GC).

FAME were prepared by a method utilizing sodium methoxide—each extract from the preparative TLC runs was dissolved in 2 mL of isooctane:dimethylcarbonate (1:1), and 1 mL of a 0.5 mol/dm^3 solution of sodium methoxide was added. The solution was vortexed for 1 min at ambient temperature to ensure complete transesterification, 3 mL of water was added, the mixture centrifuged and the upper phase carefully removed and 0.5 μL injected on GC.

GC analysis was done on a capillary column, polar phase (DB-Wax, J&W Scientific, Folsom, CA) 30 m × 0.32 mm, He carrier gas, on-column injection, 130–220°C, 1°C/min. All solvents and reagents were of analytical grade. The reproducibility of the combined TLC and GC analysis was checked with respect to the fatty acid composition of the phospholipid classes in repetitive runs (n=4) on the same lecithin, and was found to differ not more than 3.8%.

The data were statistically analyzed using univariate Analysis of Variance (ANOVA), multivariate Principal Components Analysis (PCA) and Partial Least

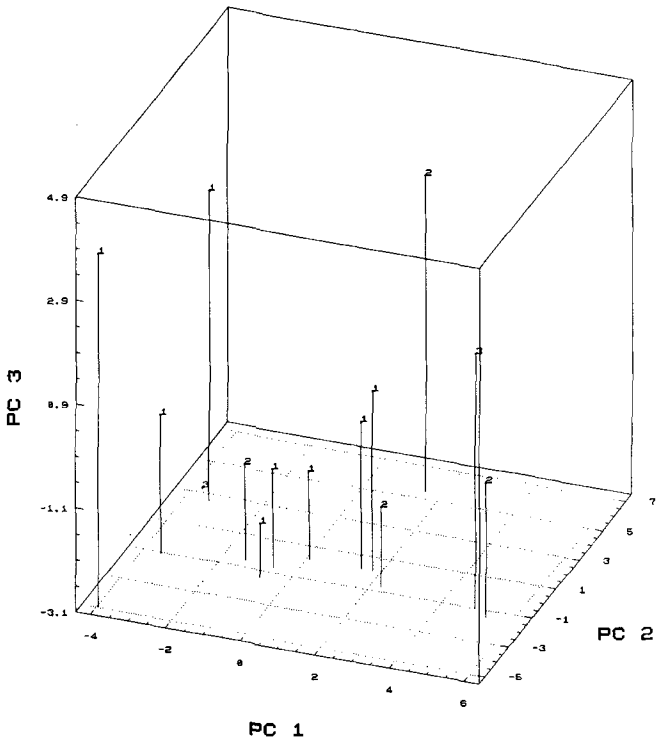

FIG. 2. Plot of the first three principal components. Axes PC1, PC2 and PC3 are linear combinations of all the variables.

Squares Regression (PLS). ANOVA was performed by comparing the variance between different lecithin samples (batches) with the mean of the replication variance (combined TLC and GC analysis), 14 different samples and two replicate runs on each batch.

PCA and PLS were done by utilizing the SIMCA program [Soft Independent Modeling of Class Analogy, Sepanova AB Stockholm Sweden (21)]. Projecting many dimensions (in this case 36) down onto a two-dimensional plane makes it possible to see natural groupings present in the data. These groupings can then be interpreted in terms of known behaviour, and hence form the basis for classification.

Using SIMCA to perform PCA is an interactive and an iterative process which is described as follows: Each individiual is described by a vector containing the different measured variables and is represented as a point in Measurement space. In order to "open a window" into M-space and see the structure of the data, the following procedure is employed (22): i) The data are autoscaled, to give a mean of zero and variance of 1 for the variables, thereby removing the influence of different units of measurement. ii) Two principal components of the whole data set utilizing all the variables are extracted. These are plotted (the window), revealing eventual groupings. iii) Individual class models are developed for each of the found groups, if possible. Outliers detected are removed for later inspection and variables found not to be relevant to the problem are discarded. iv) These disjoint principal components models can be used to classify unknown objects.

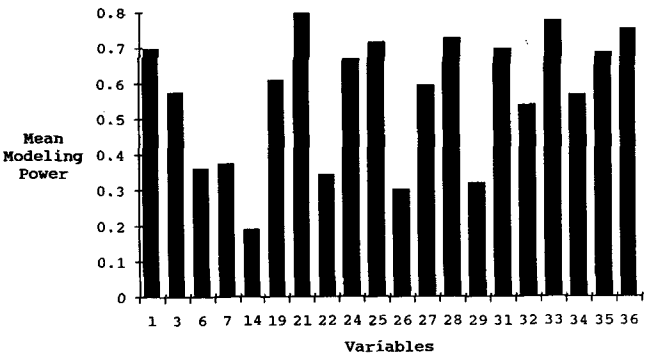

FIG. 3. Plot of the mean modeling power of the variables, taken over the four principal components of the model.

Multivariate regression using PLS is closely related to the above, except that the data are divided into an x-block and a y-block. The x-block variables are used to predict the y-block variable(s).

RESULTS AND DISCUSSION

The differences in the fatty acid content of the phospholipid classes (Table 1) were statistically significant ($p=<0.05$), with two exceptions (oleic and linoleic acid in the PI fraction). The experimental data (Table 2), i.e., the fatty acid content of the phospholipid classes and the lipid class analysis, were then used in the multivariate analysis.

Three principal components (linear combinations of the variables) utilizing all the variables (Table 1) and all of the objects (lecithin samples) were calculated and the results plotted in three dimensions (Fig. 2). In this first exploratory step prior knowledge of each lecithin sample's process performance was not used. This plot (scores) shows the projection of each lecithin sample onto a three-dimensional hyperplane in the 36-dimensional measurement space which is spanned by the analytical variables (Table 1).

As can be seen in this plot (Fig. 2), no straightforward groupings are present. However, this plot reveals a structure encountered in quality control, that of asymmetric classes (23). This is suggested by the fact that all the lecithins labeled 1 (good emulsifying performance) are embedded amongst all the other types of lecithins (2, good at double dose; or 3, no good at any dose), with "good" none the less constituting a class. It would seem that defining the class "good" is possible, a lecithin can show "good" emulsifying performance in only one way, while the contrary is not true, which is indicated by the scatter of the lecithins labeled "2" and "3". A lecithin can be a "bad"emulsifier in numerous ways and "bad" is therefore not a class-defining characteristic. Removing the irrelevant variables, according to the modeling power criterion (22), and recalculating the model gave a plot with essentially the same structure.

TABLE 3

Lecithin sample[a]	Process performance[b]	RSD[c]	PLS regression prediction
1. sc537	1	0.175	1.02
2. lm546	1	0.149	0.98
3. sc626	1	0.074	1.61
4. sc546	1	0.053	1.16
5. lm72053	1	0.248	0.81
6. sc710	1	0.295	1.16
7. scBras	1	2.335	1.10
8. lm531	2	0.78	2.05
9. lm539	2	0.772	1.59
10. lm76005	2	2.017	2.02
11. lm76001	2	2.646	2.13
12. lm533	3	2.902	2.97
13. lm71465	3	2.013	2.72
14. sc611	1	3.688	0.69

[a]lm, Lucas Meyer; and sc, Stern Chemie.
[b]These values are obtained from an industrial emulsification process; 1, good performance; 2, good at double dose; and 3, no good at any dose.
[c]Values in this column should be compared to the class "good" RSD of 0.38.

A separate class model was then developed for each of the classes, "good" and "bad". In this second step of the analysis, knowledge of each lecithin sample's process performance was used to choose samples belonging to the group "good", but this information was withheld from the subsequent calculations. The class "bad" did prove impossible to define, and no statistically significant principal components could be calculated, according to the crossvalidation criterion (23).

For the class "good", a model was developed with four significant principal components explaining 85% of the total variance in the data. Twenty relevant variables (Table 1, underlined) chosen on the basis of their modeling power, were used to define this class. These variables consisted mainly of the fatty acid composition of the different main phospholipid classes. Total concentration of PE and PC are also important in the classification, as was phosphatidic acid. By graphically representing the mean of the modeling power for each variable, the influence of each variable on the classification can be seen (Fig. 3).

Each lecithin sample was fitted to the model in order to check its validity. The degree of fit of each lecithin sample to the model is expressed by the residual standard deviation (RSD) (22); samples with values close to the class RSD are members of the class "good" (Table 3). Predictions of the PLS regression show good agreement with observed process performance (Table 3).

The results of this study show that by using a combination of several analytical techniques to chemically characterize the lecithins, a set of data relevant to the classification problem can be generated. The multivariate statistical analysis was able to separate the lecithin samples into two nested subgroups. One of the groups was definable as a statistically valid class, the "good" emulsifiers, and the other was a group of "all the rest", and therefore impossible to classify. The subsequent multivariate PLS regression was able to predict the observed process performance of the lecithins.

ACKNOWLEDGMENTS

This work was fully financially supported by the "Karlshamns Research Foundation", Karlshamn, Sweden.

REFERENCES

1. Forina, M., C. Armanino, S. Lanteri and E. Tisciornia, in *Food Research and Data Analysis*, edited by H. Martens and H. Russwurm, Jr., Applied Science Publishers, London, 1983.
2. Hagman, A., K. Karlsson and S. Jacobsson, *HRC CC, 11(1)*:46 (1988).
3. Mar, T., J. Brill, W. Bertsch, D.J.C. Fletcher and R. Crewe *J. Chromatogr. 399*:277 (1987).
4. Lavine, B.K., P.C. Jurs, D.R. Henry, R.K. Vander Meer, J.A. Pino and J.E. McMurry, *Chemometr. Intel. Lab. Systems 3*:79 (1988).
5. Blomquist, G., E. Johansson, B. Söderström and S. Wold, *J. Chromatogr. 173*:19 (1979).
6. Blomquist, G., E. Johansson, B. Söderström and S. Wold, *J. Anal. Appl. Pyrolysis 1*:53 (1979).
7. Blomquist, G., E. Johansson, B. Söderström and S. Wold, *J. Chromatogr. 173*:7 (1979).
8. Wold, S., E. Johansson, E. Jellum, I. Björnson and R. Nesbakken, *Anal Chim. Acta 133*:251 (1981).
9. Cohen, M.E., D.L. Hudson, L.T. Mann, J. Van Den Bogaerde and N. Gitlin, *J. Chromatogr. 384*:145 (1987).
10. Kowalski B.R., and C.F. Bender, *J. Amer. Chem. Soc. 94 (16)*:5632 (1972).
11. Kowalski, B.R., and C.F. Bender, *Ibid. 95 (3)*:686 (1973).
12. Wold, S., *Kemisk Tidskrift 7-8*:38 (1975).
13. Wold, S., *Ibid. 9*:36 (1975).
14. Wold, S., *Ibid. 10*:40 (1975).
15. Wold, S., *Pattern Recogn. 8*:127 (1976).
16. Wold, S., S. Albano, W.J. Dunn III, K. Esbensen, S. Hellberg, E. Johansson and M. Sjöström, in *Food Research and Data Analysis*, edited by H. Martens and H. Russwurm, Jr., Applied Science Publishers, London, 1983.
17. Massart, D.L., B.G.M. Vandeginste, S.N. Deming, Y. Michotte and L. Kaufman, in *Chemometrics: A Textbook Data Handling in Science and Technology, 2*, edited by B.G.M. Vandeginste and L. Kaufman Elsevier Science Publishers, Amsterdam, 1987.
18. Christie, W.W., in *Lipid Analysis Isolation, Separation, Identification and Structural Analysis of Lipids, 2nd ed.*, Pergamon Press, Oxford, 1982.
19. Herslöf, B., *Proceedings of the 5th International Colloquium on Lecithins: Biochemical, Pharmaceutical and Analytical Considerations*, in press.
20. Wold, S., *SIMCA 3B*, Sepanova Ab, Östrandsv. 14, S-122 43, Enskede, Sweden.
21. Albano, C., G. Blomqvist, D. Coomans, W.J. Dunn III, V. Edlund, B. Eliasson, S. Hellberg, E. Johansson, B. Nordén, D. Johnels, M. Sjöström B. Söderström, H. Wold and S. Wold in *Proceedings of Symposium on Applied Statistics*, edited by A. Höskuldsson and K. Esbensen, Danish Society of Theoretical Statistics, Neucc, Recau, Recku, Copenhagen, 1981.
22. Dunn, W.J. III, and S. Wold, *J. Med. Chem. 23*:595 (1980).
23. Wold, S., *Technometrics, 20(4)*:397 (1978).

[Received September 26, 1989; accepted April 10, 1990]