Thermochimica Acta, 226 (1993) 265–274
Elsevier Science Publishers B.V., Amsterdam

Characterization of peanuts by DSC for country of origin [1]

Susan M. Dyszel
US Customs Service, 1301 Constitution Ave, NW, Rm 7113, Washington, DC 20229 (USA)
(Received 8 August 1992; accepted 13 May 1993)

Abstract

The Research Divison of the Office of Laboratories and Scientific Services, US Customs Service, recently characterized peanuts for country of origin. This study used differential scanning calorimetry (DSC) to profile oils extracted from peanuts grown in the United States, Argentina and China. The melting behavior of triglycerides and other components in the oil matrix was obtained by controlled heating from the sub-ambient. A series of variables for each oil was assigned to the observed DSC thermal curve for the temperature region between 240 and 340 K. These variables were tabulated and analyzed by multivariate analysis. A graphical presentation of the canonical discriminant functions scores grouped the samples into three areas by country of origin. Distinguishing samples by peanut variety was also explored. DSC results are compared with those obtained by high performance liquid chromatography (HPLC) and trace metal analysis.

INTRODUCTION

The Research Division of the Office of Laboratories and Scientific Services, US Customs Service, recently examined peanuts from a major importation, at the request of the U.S. Department of Agriculture. This importation was composed of individual bags of peanuts labelled as products of Argentina and China. The shipment entered the United States after the lifting of the absolute peanut quota and the origin of the peanuts had potential impact on specific country quotas, health and possible fraud issues. This importation also generated significant political pressure from domestic peanut growing interests.

The use of differential scanning calorimetry (DSC) for determining the country of origin of nuts, specifically pistachio and macadamia nuts, has been documented in the literature [1, 2]. The current study utilized multivariate data analysis of the DSC melting profiles. Multivariate data analysis has been previously applied to the characterization of tropical oils

[1] Presented at the 21st Annual NATAS Conference, Atlanta, GA, 13–16 September 1992.

by DSC [3]. These results were compared to those obtained by high performance liquid chromatography (HPLC) analysis of the extracted peanut oil triglycerides and by trace element analysis of the whole peanuts, conducted by other members of the Research Division [4].

EXPERIMENTAL

Seventy-nine peanut samples were analyzed in the DSC profiling. Of these, eleven were standards from the Department of Agriculture (USDA) Agricultural Research Service in New Orleans, sixty-five were from the import shipment, one was a commercially available peanut oil, and one, run as two separate samples, was supplied by the Peanut Council of America. Samples were prepared in two steps: (1) 25 g of raw peanuts were finely ground, and (2) the oil was extracted from a 5 g portion of the ground peanuts with petroleum ether. The method of extraction for the commercial peanut oil was unknown. DSC and HPLC utilized the same extracted oil; trace element analysis used the ground peanuts from the first step.

The DSC (DSC-2, Perkin-Elmer, Norwalk, CT) was programmed from 230 to 380 K at 5 K min^{-1}. Each encapsulated oil sample (5–10 mg) was introduced into the sample cell at the lower temperature under flowing nitrogen; the DSC signal baseline was allowed to stabilize and the heating was begun. The thermal curve was recorded and stored using the thermal analysis data station (TADS, Perkin-Elmer, Norwalk, CT). The sample was then cooled at 25 K min^{-1} back down to 230 K and allowed to equilibrate. A second heating run under the same conditions was then initiated. The results of the second run were used for data analysis, the first heating and cooling cycle having imposed a common thermal history on all samples. Reproducibility of the DSC curves was tested by running a third heating cycle. The resulting thermal curve could be superimposed on that from the second heating run. Similarly, the DSC curves from duplicate aliquots from the same oil were compared. Reproducible results were also obtained in these comparisons.

Data from the DSC were processed using the TADS Partial Area software Perkin-Elmer, Norwalk, CT), Quattro Pro spreadsheet (Borland International, Scotts Valley, CA) and the SAS/STAT multivariate analysis program, version 6 for personal computers (SAS Institute Inc., Cary, NC). The resulting DSC curves between 240 and 340 K were divided into partial areas based on the peak and valley position temperatures. The Partial Areas program allows for manual cursor selection of specific points along the DSC curve. The position temperatures were assigned manually with a valley in between when possible, as in Fig. 1 (p7 and p9 are peaks, p6 and p8 are valleys). In some cases, the above-baseline height difference between the peaks and valley was minimal, as seen in Fig. 2 and (p6 and p7). In such cases, the selection of the position was made by inspection to maximize the

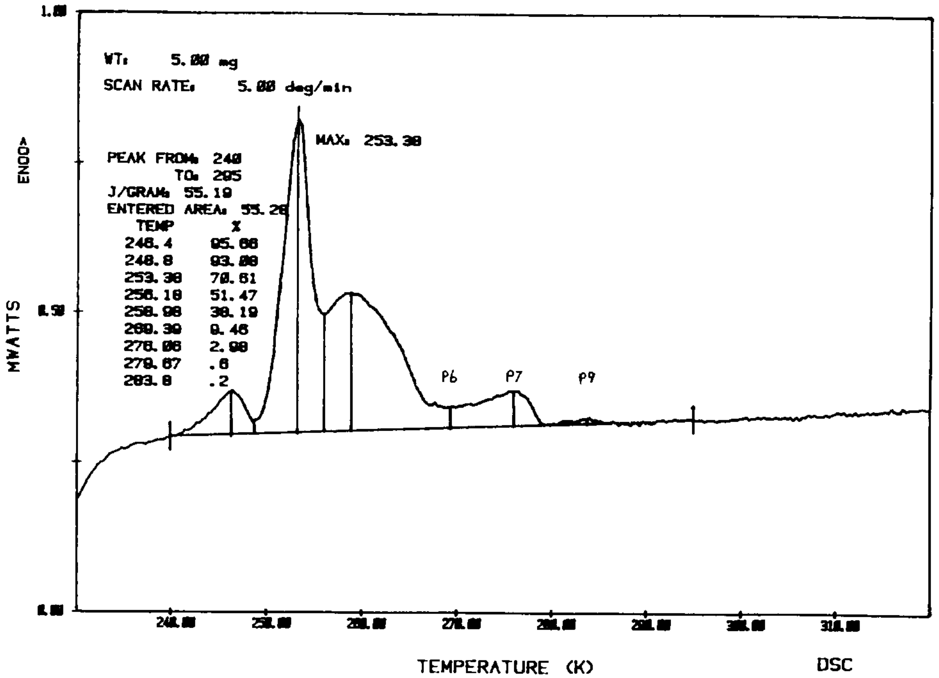

Fig. 1. DSC thermal curve of an Argentine-grown peanut oil; p6, p7 and p9 peak height measurement locations are indicated.

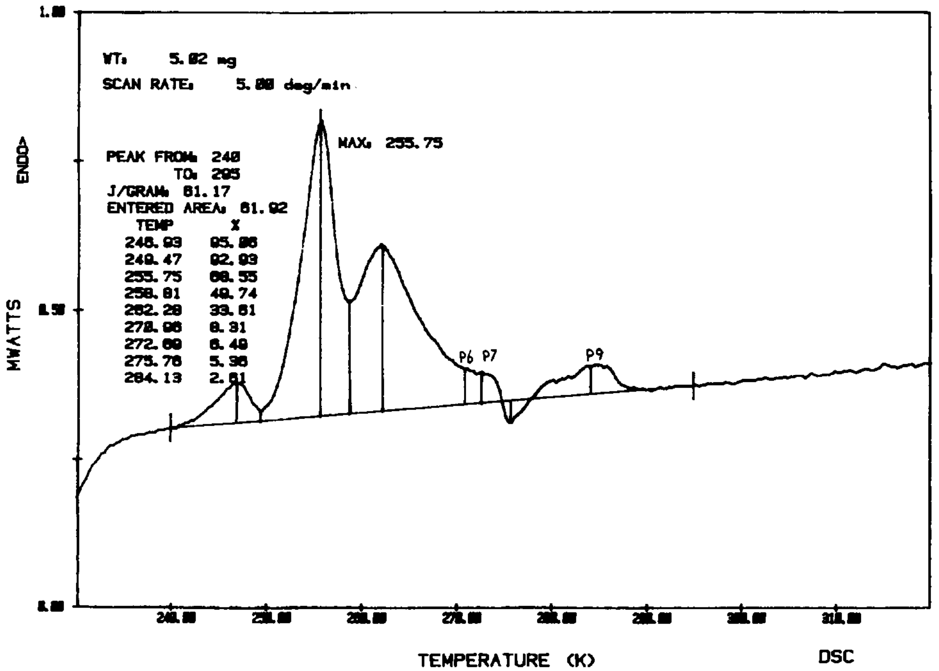

Fig. 2. DSC thermal curve of a Chinese-grown peanut oil.

difference in measured heights. Heights above the baseline at the 6, 7 and 9 positions (designated as p6, p7 and p9 on the thermal curve, respectively) were measured manually on the plotted, weight-normalized DSC curves. The heat of melting, designated "JGRAM", was calculated from the area under each DSC curve between the designated partial area limits, corrected for sample mass. In some samples, this temperature region included an exothermic event, recorded as a valley between the peaks, which dropped below the baseline constructed for the partial areas calculation. "EAREA", the entered area, another variable calculated by the partial area program, took the total area of the thermal events, both above and below the baseline, into account. EAREA was greater than JGRAM when the thermal curve dropped below the partial areas baseline.

All twenty-seven variables including partial areas, position temperatures with their corresponding height measurements, the heat of melting (JGRAM) and the total area under the curve (EAREA), were transferred to a spreadsheet for tabulation. The tabulated variables were input into the multivariate data analysis package for manipulation and analysis. Discriminant analysis and canonical discriminant analysis were used for both country of origin and variety determinations.

DISCUSSION

DSC observations

Figures 1 and 2 display DSC heating profiles for typical samples of Argentine and Chinese peanuts. In the region 270–285 K, the Chinese peanut oils showed evidence of polymorphism of the oil triglycerides, as demonstrated by a characteristic exothermic "valley" followed by an endothermic peak. The peak height measurements at p6, p7, p9 and their respective ratios helped to quantify this observed pattern. The polymorphic behavior of this Chinese peanut oil has been seen in the melting of other oils and has been previously discussed in the literature [5].

Other than the observed polymorphism in the Chinese peanut oils, there were insufficient observable differences between the DSC curves which could, by themselves, be used as the sole basis for determining the country of origin of the samples. Therefore, a numerical approach to the data analysis was undertaken using multivariate analysis.

Theory of multivariate analysis

Multivariate analysis may be defined as those statistical methods which analyze multiple measurements on each sample under investigation. Some of the more well-known techniques included in multivariate analysis are multiple regression, discriminant analysis and principal component analysis. For the particular problem of determining country of origin based on values derived from DSC melting behavior, discriminant analysis is an appropriate technique to use, because this technique relates a single dependent non-numeric vairable (country of origin) to a set of independent numeric variables (DSC values).

Two variations of this statistical technique were used in this study: discriminant analysis and canonical discriminant analysis. Discriminant analysis uses a sample set with known group membership called a calibration set to form a mathematical model or set of discriminant functions. These discriminant functions can then be used in resubstitution analysis to predict group membership of unknown samples. In addition to assigning an unknown sample to a particular group, it is also possible to calculate the probability of group membership using discriminant analysis resubstitution. These calculations involve the mathematical distance from the sample to each group and the variances, both inter- and intra-group, for the calibration set. In this manner, the degree of confidence in a group assignment can be assessed. Posterior probabilities of membership of less than 1.0000 indicate a greater degree of uncertainty in the classification.

Canonical discriminant analysis is a dimension reduction technique which is used to visualize group separations in two-dimensional plots. This technique uses linear combinations of the original values to form canonical functions which summarize the variance in the data. The number of canonical functions is significantly lower than the number of original variables, being one less than the number of classes of the dependent variable, i.e. countries of origin. Each sample can be represented by its set of scores on these functions, and because the most variance or information content is concentrated in the first two or three functions, this information can be displayed in two-dimensional plots. In these plots, samples having the same origin should cluster together, while samples with different origins should be separated, if a relationship exists between the country of origin and the DSC data.

Statistical analysis of DSC data for country of origin

In this study, the tabulated DSC data from the USDA peanut standards were used to form the discriminant analysis calibration set or model to predict country of origin of the imported, commercial oil and Peanut Council peanut samples. All variables were included on the data analysis with equal weighting. The resulting scores on the canonical discriminant functions, also based on all variables, were plotted to visualize the country of origin separations.

Using discriminant analysis resubstitution to predict the probability of group membership for all samples, all but one of the samples with an attributed origin were identified as belonging to that origin group. Based on this data treatment, as shown in Table 1, it appears that one of the Argentine samples exhibited thermal characteristics which more closely resembled the USA peanuts than those from Argentina. Three other samples are listed in Table 1. These were the commercial oil and the Peanut Council samples which had no country of origin attributed to them.

A subset of 17 peanut samples from the import shipment was analyzed by all three techniques: trace element analysis, HPLC and DSC. The results from all three techniques agree on the origin of these peanuts. Based on the results of DSC, the commercial peanut oil was classified as the USA origin. The commercial oil sample was run only by DSC. The duplicate samples of the Peanut Council unknown were assigned by discriminant analysis resubstitution to the Chinese origin group based on their DSC data runs. The Peanut Council peanuts, however, had been identified as of USA origin by trace element analysis and were not run by HPLC.

Turning to the graphical display of the results, the plot of the canonical scores for the DSC peanut samples related to origin is shown in Fig. 3. The canonical plot indicates three areas, enclosed by lines, which are associated with the three countries of origin. Each cluster is identified by a letter, "A", "C" or "U", designating Argentina, China or USA, respectively. The commercial oil unknown is shown in the USA cluster, identified by an "X". One of the Peanut Council duplicate sample runs, "P", is plotted in the area associated with the peanuts from China, while the other is in an area intermediate between the USA and Chinese clusters. The single Argentine peanut sample with thermal characteristics resembling those of the USA grown peanuts is shown as the "A" in the USA cluster.

TABLE 1

Posterior probability of membership by origin: discriminant analysis resubstitution

Run ID	Origin Attrib.	Origin Assign.	Probability		
			ARG	CHI	USA
PN160	ARG	USA	0.0020		0.9980
PN142 [a]	?	USA	0.0001		0.9999
PN182 [b]	?	CHI	0.0001	0.9726	0.0274
PN184 [b]	?	CHI		0.9745	0.0265

Probabilities less than 1.0000 indicate greater degrees of uncertainty in classification.
Key: ARG, Argentina; CHI, China; USA, United States; ?, no origin attributed to sample.
[a] Commercial peanut oil. [b] Peanut Council peanuts.

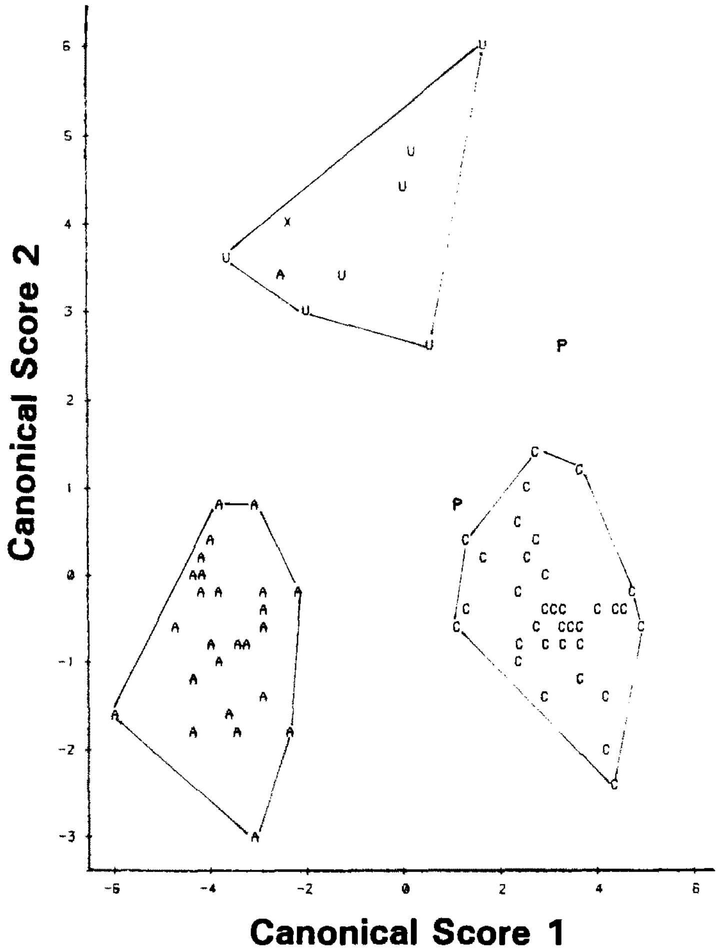

Fig. 3. Canonical value plotted for country of origin. Letter indicates Country: A, Argentina; C, China; U, USA; X, commercial oil; P, Peanut Council. Country of origin clusters are shown within the connecting lines.

Statistical data analysis of DSC data for peanut variety

In an attempt to understand why the DSC and the trace element analysis results for the country of origin of the Peanut Council sample differed, the DSC dataset for all samples was evaluated for the effect of peanut variety. In this case, there were six attributed varieties. Of these six attributed varieties, only the Runner variety with 27 samples and the Virginia variety with 34 samples had sample populations greater than one. The varieties represented by a single sample each were Valencia, Spanish, North Carolina and Windrow. There were also 12 samples with no attributed variety. These unattributed or unknown variety samples included some of the USDA standards, the commercial peanut oil and the Peanut Council samples.

TABLE 2

Posterior probability of membership by variety discriminant analysis resubstitution

Run ID	Variety attrib.	Variety assign.	Probability			
			NC	RUN	VA	WIN
PN030	RUN	NC	0.5979	0.4021		
PN012 [a]	?	VA			1.0000	
PN010 [a]	?	VA			1.0000	
PN138 [a]	?	VA			1.0000	
PN017 [b]	?	RUN		0.9972		0.0028
PN006 [b]	?	RUN	0.0167	0.9833		
PN140 [b]	?	WIN		0.0428		0.9572
PN015 [c]	?	NC	0.8690	0.0796		0.0514
PN002 [c]	?	WIN				1.0000
PN004 [c]	?	VA			1.0000	
PN142 [d]	?	NC	0.5248			0.4752
PN182 [e]	?	WIN			0.3478	0.6512
PN184 [e]	?	VA			1.0000	

Probabilities less than 1.0000 indicate greater degrees of uncertainty in classification.
Key: NC, North Carolina; RUN, Runner; VA, Virginia; WIN, Windrow; ?, No variety attributed to sample.
[a] USDA standard—China. [b] USDA standard—Argentina. [c] USDA standard—USA. [d] Commercial peanut oil. [e] Peanut Council peanuts.

With all six attributed varieties entered, the variety membership assignments from the discriminant analysis resubstitution matched those attributed to the samples for the majority of the samples for which a variety had been identified. Table 2 details the probability of membership by variety for those samples which were either grouped in a variety other than attributed or had no variety attributed. The canonical plot for variety is divided into two primary clusters, designated by r and v, which represent the Runner and Virginia varieties, respectively. There are more than just the three unknown samples indicated in Fig. 4, as the USDA samples had no listed variety. However, there are fewer than 12 "Xs" on the graph because some of the unattributed variety sample canonical scores fall coincident with those of other varieties and are thus hidden. For four of the samples which did not have attributed variety, the graphical plot of their canonical values in Fig. 4 showed that these samples appear very close to the location of one of two single samples (identified by n for North Carolina or W for Windrow). These four samples had been classified by the discriminant resubstitution analysis as North Carolina and Windrow varieties.

There is an added measure of uncertainty associated with the Windrow and North Carolina Variety Assignments, however. As these are geographical designations, the actual variety may well be "Runner" as indicated by the clustering of the North Carolina and Windrow samples

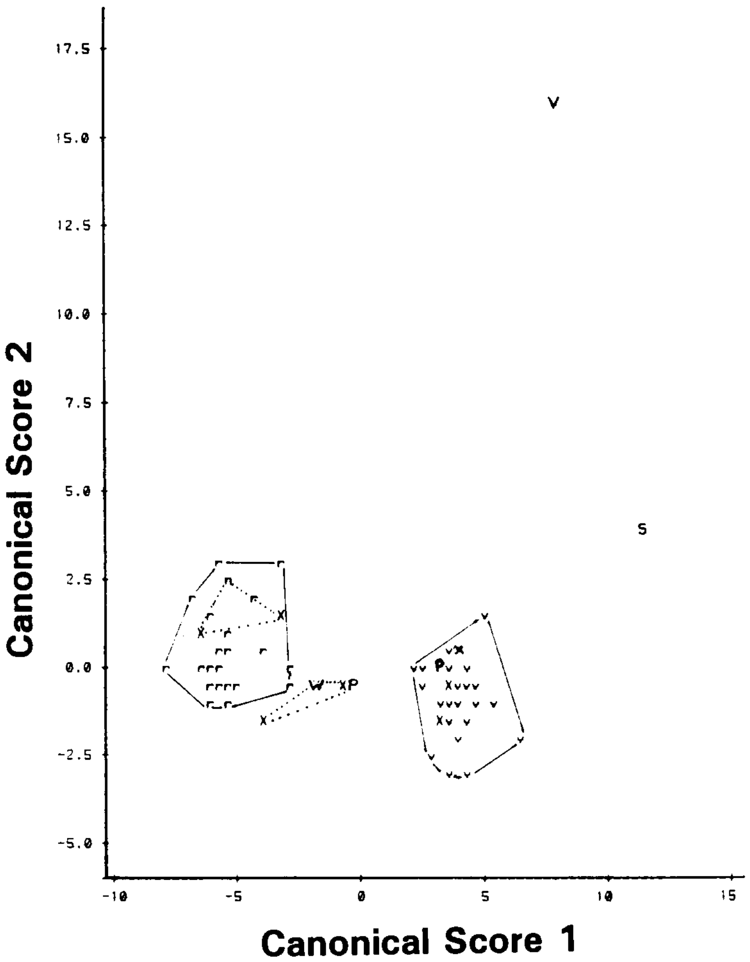

Fig. 4. Canonical values plotted for variety. Letter indicates variety: r, Runner; v, Virginia; S, Spanish; V, Valencia; n, North Carolina; X, unattributed variety; P, Peanut Council; W, Windrow. Varietal groups are shown by connecting lines. The North Carolina group appears to be within the Runner Group and the Windrow intermediate between the Runner and the Virginia groups.

within and near, respectively, the "Runner" data cluster on the canonical plot for variety. Removal of all four single varietal samples (Valencia, Spanish, North Carolina and Windrow) would have left only 2 varieties for the model, which would have been reduced by the canonical discriminant analysis to only one canonical score per sample, rendering the data incapable of being plotted on a two-dimensional plot.

By matching the variety grouping with the samples' attributed origin identifications, there was a correlation between samples attributing a Virginia variety and Chinese origin; likewise, the attributed Runner variety corresponded to Argentine origin. The peanuts of USA origin were also found to be in the Runner cluster. One of the peanut Council samples

(PN184) was identified as being of the Virginia variety by both the discriminant resubstitution analysis and canonical scores plot, which, for this universe of samples, has been associated with peanuts of Chinese origin. The other sample (PN182) was graphically located on the edge of the Windrow cluster on the canonical scores plot. This sample was also classified as belonging to the Windrow variety by the discriminant resubstitution analysis. However, the probability of belonging to this group was at the 65% level with a 35% probability that the sample might have been of the Virginia Variety. Accordingly, the possibility that Chinese origin might be associated with this sample also exists. However, it is apparent that the canonical boundaries of variety are not clearly defined between the Runner, Windrow and Virginia varieties based on this data universe. Variety alone cannot be used to determine country of origin.

CONCLUSION

DSC and multivariate analysis have been used to determine the country of origin of peanuts, differentiating those grown in China, Argentina and the United States. Based on a limited population of samples, it has been shown that the DSC results generally correlate with those obtained from HPLC and trace element analysis, and overall, the DSC data strongly suggest the significant roles that variety, along with origin, play in determining the thermal profile of peanuts. The availability of representative and authentic standards of known origin also has a direct bearing on the predictive certainty in country of origin determinations.

ACKNOWLEDGMENTS

I acknowledge the contributions to this study by my colleagues at the Research Division, Office of Laboratories and Scientific Services. Robert S. Schwartz, Ph.D. and Le Tiet Hecking, for their work on the trace element analysis, and the late Bruce C. Pettitt, Ph.D., for his HPLC analysis of the peanuts.

REFERENCES

1 S.M. Dyszel and B.C. Pettitt, Determination of the country of origin of pistachio nuts by DSC and HPLC, J. Am. Oil Chem. Soc., 67(12) (1990) 947.
2 S.M. Dyszel, Characterization of macadamia nuts by differential scanning calorimetry for country of origin, Thermochim. Acta, 166 (1990) 291.
3 S.M. Dyszel and S.K. Baish, Characterization of tropical oils by DSC, Proc. 20th NATAS Conf. (1991) 451.
4 B.C. Pettit, R.S. Schwartz and L.T. Hecking, Country of origin – peanuts: China vs. Argentian, Customs Laboratory Bulletin, 4(3) (1992) 89.
5 E. Kaiserberger, DSC Investigations of the thermal characterization of edible fats and oils, Thermochim. Acta, 151 (1989) 83.