Journal of Thermal Analysis and Calorimetry, Vol. 72 (2003) 415–420

USING SAMPLE QUALIFICATION TECHNIQUES TO CLASSIFY DSC RECORDS

*J. Sempere, R. Nomen**, *E. Serra and D. de Armas*

Institut Químic de Sarrià, Universitat Ramon Llull, Via Augusta, 390, 08017 Barcelona, Spain

Abstract

The Similarity Match, Distance Match, Discriminant Analysis, Search Standards, and QC Compare Search techniques of statistical analysis are applied to DSC records in order to distinguish between two forms or classes of a new pharmaceutical active component. The classes are defined according to its therapeutic activity as 'positive' and 'negative'. Excellent results are obtained using the QC Compare Search Method.

Keywords: DSC, sample classification, statistical analysis

Introduction

During the development of the industrial process to produce a new pharmaceutical active component, our group detected that samples from different batches could be classified into two categories (A – good – and B), according to the level of their biological activity. However, the analytical techniques used (HPLC, IR, RX diffraction) do not detect chemical or physical significant differences between them. Optical microscopy (Fig. 1) shows different size and agglomeration between the crystals corresponding to the two categories, but not a clear different crystalline form.

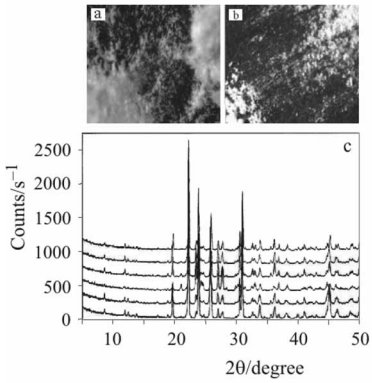


Fig. 1 Optical microscopy of crystal of classes a – A and b – B and c – X-ray diffraction patterns of sample corresponding to both classes

* Author for correspondence: E-mail: rnomen@iqs.edu

1388–6150/2003/ $ 20.00

Akadémiai Kiadó, Budapest
Kluwer Academic Publishers, Dordrecht

DSC records of different batches gave slightly different curves (Fig. 2), but classical evaluation methods don't allow to carry out a satisfactory classification of samples. The classification of such series of records is a challenge for the DSC users [1]. Promising results using simple methods of classification on MSExcel®. As a consequence, it was decided to export the data to the Turbo Quant Analyst® software, usually applied to process spectroscopic (near-IR) data.

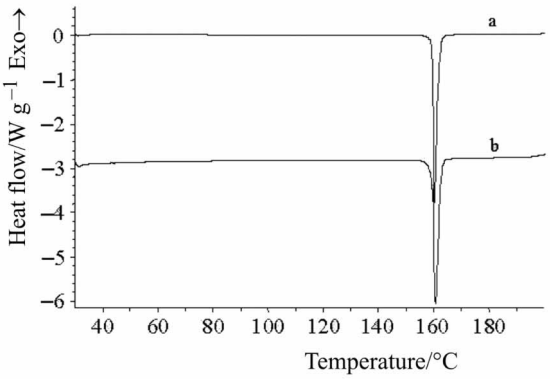


Fig. 2 Typical DSC curves for a – accepted and b – refused batches

The objectives of this work are applying to the DSC data of different batches using the mathematical and statistical techniques developed for the spectroscopic data, and establishing a procedure to classify the available samples.

Experimental

DSC records are obtained using a Mettler Toledo STAR^e system with a 821/700 cell. 2–4 mg of material are introduced in a standard aluminium pan (ref. ME-5111987) with a calibrated pierced lid (ref. ME-51140832). The scanning is from 30 to 200°C with a heating rate of 10 K min^{-1}, using a flow of 80 mL min^{-1} of dry nitrogen. A total of 23 pre-classified samples were recorded.

Table 1 Processing conditions for the different methods of calculation

Method	Pathlength	Classes	Number of standards	
			Class A	Class B
Similarity Match	Constant	A	7	0
Distance Match	Peak ratio or Normalise, Area 386–397	A+B	5	7
Discriminant Analysis	Multiplicative signal correction	A+B	6	7
Search Standards	–	A	6	0
QC Compare Search	–	A+B	6	7

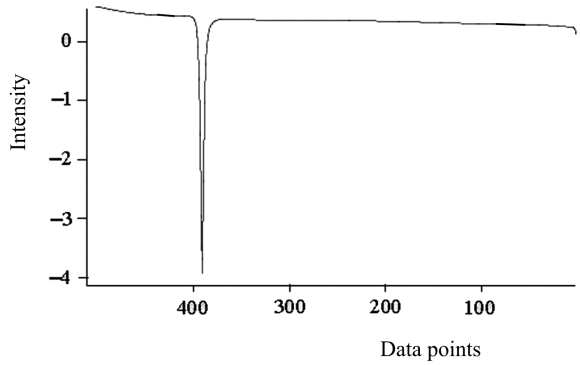


Fig. 3 Example of a DSC curve exported to Turbo Quant Analyst[R]. The *x*-axis is reversed by the program, which uses the DSC data as a spectrum

The DSC curves collected using the Star[e] software (Mettler Toledo), version 6.01, are exported to MSExcel® files and from this to the formate used by the Turbo Quant Analyst software included with the 6.1 NIR Antaris® Near IR instrument (Nicolet).

Finally, the DSC curves are treated using different methods supplied by Turbo Quant Analyst®. The set spectra is divided into two groups: a calibration set, and a validation set. Table 1 shows the conditions and sets of calibration for each method applied.

Results and discussion

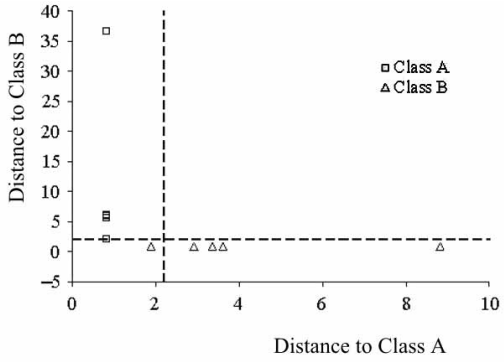
The statistical package being used includes five different methods to classify samples [2–5]. Each procedure gives results according different criteria [6, 7], but it is very difficult to predict which methodology will produce the best classification. In addition, it should be considered that DSC curves have completely different properties than spectra. As a consequence it is necessary to check all of them. Table 2 shows the results obtained using the different methods.

- *Similarity Match* looks for similarities and differences between a curve of an unknown sample and one reference class. It allows classifying in only one class, according to a similarity degree. It produces an indication of the residual similarities and differences.
- *Distance Match* consists in comparing point-by-point the curve of one unknown sample with each curve contained in a reference library, corresponding to known classes. The addition of the discrepancies is then used as an indication of the similarity between the problem and each class. It doesn't produce any percentage of similarity or confidence range for the results.
- *Discriminant Analysis* has been extensively used to the identification of raw materials in the chemical and pharmaceutical industries. It tries to obtain the class that could enclose the curve of the unknown sample. Essentially it consists in an analysis similar to the distance match method, including additional principal component selection and statistical considerations.

- *Search standards* searches for the difference between the curve of the unknown sample and all the spectra in the library. Finally, it looks after the minimum difference and classifies the sample in the same class as the nearest reference curve.
- *QC Compare Search* is an application of the known KNN (*k*-nearest neighbour) method with *k*=1. Each curve is treated as a vector and its distance to the references is computed. It tries to discriminate if a curve belongs to a one class between several classes. So, it requires more than one reference class to make the classification.

Table 2 Results of classification

	Fault/%	Results
Similarity Match	17	Asking for a similarity level of 99%. The method recognised all 'good' samples, but it includes some wrong results.
Distance Match	21	It classified as 'Class A' all 'A' samples, but also some 'B'.
Discriminant Analysis	25	Classified correctly all the 'A' samples, but makes some mistakes with 'B' sample.
Search Standards	3	Only one curve is out of its place.
QC Compare Search	0	Full classification.



Sample	Actual Class	Calculated Class	Distance	Next Class	Next Distance	Distance to Class A	Distance to Class B
2B	A	A	0.8000	B	6.1154	0.8000	6.1154
11B	A	A	0.8000	B	36.6488	0.8000	36.6488
12B	A	A	0.8000	B	5.6723	0.8000	5.6723
10B	A	A	0.7998	B	6.0509	0.7998	6.0509
6M	B	B	0.7999	A	3.3553	3.3553	0.7999
9M	B	B	0.8002	A	1.8911	1.8911	0.8002
8M	B	B	0.7999	A	8.8186	8.8186	0.7999
14M	A	A	0.8002	B	2.1331	0.8002	2.1331
7M	B	B	0.8000	A	3.6144	3.6144	0.8000
2M	B	B	0.8000	A	2.9141	2.9141	0.8000

Fig. 4 Example of result of calibration

The software produces a table and a plot of results of calibration (Fig. 4), and – for some of the methods – a plot including the principal components, class averages and standard deviations (Fig. 5).

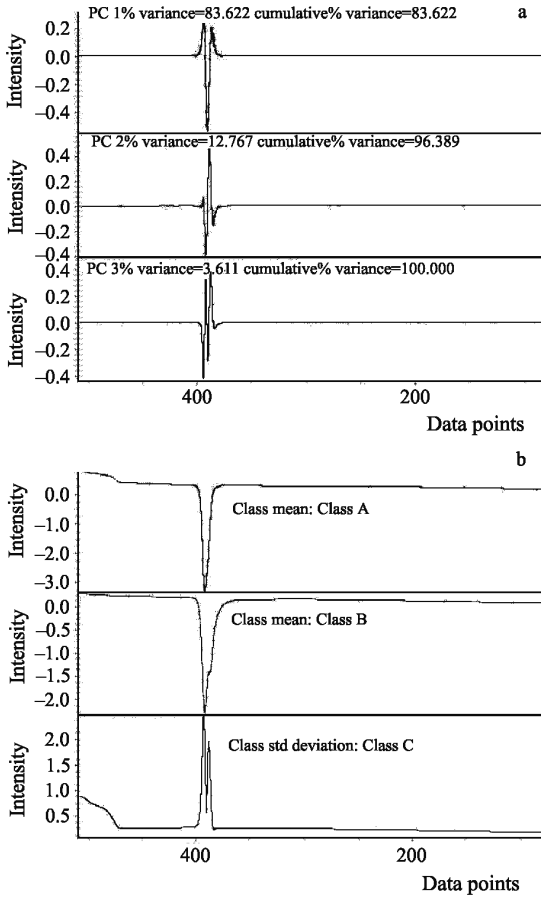


Fig. 5 Principal components and class averages

Table 2 shows the results obtained when the calibrated methods are applied to the validation set. The QC Compare Search method produces the complete classification of the set of validation samples.

Conclusions

DSC curves can be successfully analysed (classified) using mathematical and statistical methods initially developed for spectroscopy, expanding the possibilities of thermal analysis to cases in which only small – very difficult to evaluate – differences between different records are observed.

In the studied case the QC Compare Search method produces the best classification. However larger sets both for calibration and validation should be used to corroborate this result.

The authors would like to acknowledge the collaboration of IZASA, S.A., which made possible the use of the system Nicolete Antaris Near IR and its software for this work.

One of the authors D. A. would acknowledge the grant received from the MAE (Ministerio de Asuntos Exteriores de España, Becas unilaterales de la AECI para estudiantes extranjeros, BOE22/06/94), which made possible her participation in this work.

References

1 M. Sh. Lvova, E. I. Kozlov and I. A. Wishnevetzkaya, J. Thermal Anal., 40 (1993) 405.
2 T. Clermont and L. Luchetta, Nicolet Industrial Solution, Nicolet Instrument Corporation, Madison, USA WI, 1999.
3 S. Lowry, B. McCarthy and J. Hayatt, Nicolet Industrial Solution, Nicolet Instrument Corporation, Madison, USA WI, 2000.
4 W. J. McCarthy and S. Lowry, Nicolet Industrial Solution, Nicolet Instrument Corporation, Madison, USA WI, 2001.
5 M. S. Kemper and L. M. Luchetta, Nicolet Industrial Solution, Nicolet Instrument Corporation, Madison, USA WI, 2001.
6 S. R. Lowry, J. Hyatt and W. J. McCarthy, Appl. Spectrosc., 54 (2000) 450.
7 E. W. Ciurczak, NIR Analysis of pharmaceuticals, 'Handbook of Near Infrared Analysis' D. A. Burns and E. W. Ciurszak, Ed. Marcel Dekker, Inc., N.Y. pp. 550–551.