Available online at www.sciencedirect.com
ScienceDirect
Journal of Pharmaceutical and Biomedical Analysis 47 (2008) 320–327
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Near-infrared reflectance spectroscopy as a process analytical technology tool in *Ginkgo biloba* extract qualification

Sílvia S. Rosa [a,b,*], Pedro A. Barata [a], José M. Martins [a], José C. Menezes [b]

[a] *Laboratórios Atral SA, Rua da Estação, 42, Vala do Carregado, P-2600-726 Castanheira do Ribatejo, Portugal*
[b] *Center for Biological and Chemical Engineering, IBB—Institute for Biotechnology and Bioengineering, IST, Technical University of Lisbon; Av. Rovisco Pais, P-1049-001 Lisbon, Portugal*

Received 23 October 2007; received in revised form 15 January 2008; accepted 17 January 2008
Available online 29 January 2008

Abstract

Here, we describe the use of near-infrared diffuse reflectance spectroscopy for qualification of *Ginkgo biloba* extract as raw material for use in pharmaceutical products. *G. biloba* extract shows unpredicted and uncontrolled variability in some of its quality specifications, intrinsic to its natural origin, which have influence on the manufacturing process of solid dosage forms (viz. granulation and compression). Some of these properties could not be determined by conventional quality control tests, so we investigated the use of NIR to qualify the batches of Ginkgo extract accordingly to its different features and establish a relationship with some of the manufacturing steps behaviour based on their qualification. Several approaches were evaluated, and the NIR method developed demonstrated to be sensitive to changes in important quality specifications and therefore adequate to qualify incoming batches of *G. biloba* extract. This could be considered a process analytical technology (PAT) application since it: (1) establishes the source of variability in a qualitative way, (2) explains its propagation to the final product quality attributes and (3) lays the basis for a control strategy to be applied in the manufacturing process.
© 2008 Elsevier B.V. All rights reserved.

Keywords: Near-infrared spectroscopy; Qualification; *Ginkgo biloba*; Pharmaceutical products; Process analytical technology (PAT)

1. Introduction

Near infrared (NIR) covers the wavelength range between the mid infrared (MIR) and the visible region: 780–2500 nm or 12,800–4000 cm^{-1} [1]. It is a fast and non-destructive technique that gives information about chemical and physical properties in almost all kinds of samples [2]. The growth and application of NIR techniques has been directly related to the development of equipment and software, and in the last 25 years near-infrared spectroscopy (NIRS) has become widely used for a range of analyses in various industries [3]. Nowadays, NIRS has found increased use in the pharmaceutical industry. The most common application of NIR in the pharmaceutical industry is the identification of raw materials [4,5], but in the last decade several applications have been described: the quantification of active pharmaceutical ingredients, the determination of water content, blend homogeneity and hardness in solid dosage forms formulations, among others [6].

Ginkgo biloba extract is obtained from the leaves of *G. biloba* tree, one of the most studied and used botanicals in the world, which is widely used in medicine due to its antioxidant, anti-ischemic and neuroprotective properties [7].

Most of the available ginkgo extracts on the marked are standardized for their content in terpene lactones and flavonol glycosides, which are the two most important pharmacologically groups of compounds present. A draft monograph of the European Pharmacopeia [9] for a purified Ginkgo dry extract, mentions standardized specifications for those parameters. Terpene lactones content is usually around 6.0%, and flavone glycosides around 24.0%. During the production of *G. biloba* extract, manufacturers must control the presence of these compounds as well as others (e.g. ginkgolic acids, responsible for the allergenic and cytotoxic effects) [8]. Even thought the most important and usual parameters of *G. biloba* extract are those mentioned above, there is a large number of other parameters

* Corresponding author. Tel.: +351 962023417.
E-mail address: silviarosa@ist.utl.pt (S.S. Rosa).

0731-7085/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2008.01.031

that should be verified: proanthocyanidin content, organic acid content, solubility, total residual organic solvents, sulphated ash, microbiological contamination, presence of phosphorous and chlorinated solvents, investigation on specific functional groups, pH value and particle size [8].

As said different batches of *G. biloba* extract reveal different behaviour in a pharmaceutical manufacturing process. This could be related to its natural variability (e.g. different growing cultivar conditions, different geographic origin), and slight differences on some of its quality attributes not usually determined in conventional quality control tests. In our work, batch results of conventional wet chemical tests were not sufficient to classify Ginkgo extracts and explain the behaviour observed during production, so we made a retrospective study on previous batches and tried to use NIR as a pattern-recognition technique for *G. biloba* extracts as described in previous studies [10,11]. We tried then to investigate the relationship between *G. biloba* extract quality parameters and intermediate (viz. granulates and cores) and final product (viz. coated tablets) analytical results.

This could be considered a process analytical technology (PAT) application since it is an efficient and innovative approach for the control of critical quality attributes of a raw material to a pharmaceutical manufacturing process [12]. Although this application is used in the first step of the process (raw material analysis), it will enable process control, thus reducing the vulnerability of the process and manufacturing time cycles, to use in the incoming raw-material with a positive impact in productivity and final product quality assurance.

2. Experimental

2.1. Materials

Thirty batches of *G. biloba* were used in this study (Table 1). All batches are from the same supplier, cover a period of about 4 years, and were selected to include the three main processability behaviours observed. A subset of 23 batches was used for method development. The remaining seven batches were used to challenge the method, and establish its predictive capabilities. Table 1 summarizes the Ginkgo batches used for method development and validation, the chemical properties and production processability behaviour for each one.

2.2. Near-infrared reflectance spectra

Spectra were recorded on a MPA spectrophotometer from Bruker Optics®, equipped with a reflectance diffuse fiber optic probe also from Bruker Optics®.

Spectra were recorded on the wavenumber range of 4000–12,000 cm^{-1}, from an average of 32 scans and with 8 cm^{-1} resolution. Spectra were obtained inserting directly the probe into the powdered samples. As a hand-held probe this operation needs special attention in order to ensure that the probe does not move during the spectra acquisition. In between samples, the probe was cleaned with water humidified paper and then paper dried.

2.3. Data analysis

Identification and cluster methods were developed using the Ident package of OPUS software, Version 4.2, from Bruker Optics®, while for classification principal component analysis (PCA) was used as implemented in the PLS Toolbox version 3.5 for Matlab version 6.5, both from The MathWorks Inc., 2002.

3. Results and discussion

G. biloba extract batches were studied according to the following quality control attributes and production processability parameters: (1) water dispersion, which is the test performed by the dispersion of about 100 mg of *G. biloba* extract in 2 ml of distilled water. The test result is positive if the sample disperses completely in the water without the lump formation, (2) hidroalcoholic solubility, performed in about 1000 mg of *G. biloba* extract, to which is added a mixture of 14.57 ml of ethanol 95° and 10 ml of water. The test result is positive if the result is a limpid solution; (3) water content, performed by the Karl Fischer titration method, which result should be less than 3%; (4) granulation and compression problems, both happening during manufacturing. Granulation problems due to the lump formation, and compression problems due to the hardness reduction; (5) compression problems, as the only problem happening during the manufacturing step, most related with hardness reduction.

The main intention of this study is to establish a relationship between the quality control attributes and production processability parameters described above, by NIR.

3.1. Method development

In order to find a suitable approach to qualify the *G. biloba* extract several multivariate classification methods were tested. These methods, also known as pattern-recognition methods combine the different chemical properties and processability behaviours of Ginkgo (see Table 1) with different chemometric methods—i.e., supervised and unsupervised classification methods. The difference among them is related with the prior knowledge that we have about the sample categories. Supervised classification methods can only be developed when we have a training set of samples with known categories, while unsupervised methods do not require previous knowledge [2].

Three types of techniques were used: identification supervised methods, PCA and cluster unsupervised methods.

Identification supervised methods were performed constructing a library of reference spectra, according to the pre-defined classes. After developing a method, the "unknown" samples were compared to that library, and were given the class name of the closest match with one of the reference spectra groups. The similarity level is based on spectral distances. This enables us to evaluate to which degree a substance is identical to one of the reference spectra classes, but does not allow the graphical representation of the groups. Spectral distances were calculated using the factorization algorithm. Using this algorithm, spectra is represented as a linear combination of the factor spectra

Table 1
Ginkgo biloba batches used in method development and validation, its properties and production behaviours

Batch	Water dispersion	Hidroalcoholic solubility	Water content (%)	Production behaviour
Method development				
1	+	+	1.4	N
2	+	+	1.2	N
3	+	+	1.9	N
4	+	+	1.6	N
5	−	−	2.7	GC
6	−	+	2.7	GC
7	+	+	2.6	GC
8	+	−	2.3	GC
9	+	+	1.7	N
10	+	+	1.4	N
11	−	−	2.9	C
12	+	+	1.4	N
13	+	+	1.3	C
14	+	+	1.8	GC
15	+	+	2.6	GC
16	+	+	1.7	N
17	+	+	1.3	N
18	+	+	2.8	GC
19	+	+	1.4	N
20	+	−	2.9	GC
21	+	+	1.2	N
22	+	+	1.3	N
23	+	+	1.4	N
Method validation				
1	+	+	1.2	N
2	+	+	2.0	N
3	+	+	2.9	GC
4	+	−	2.6	GC
5	+	+	1.4	N
6	+	+	1.2	N
7	−	+	2.9	GC

Within specification result (+), out of specification result (−), normal (N), granulation and compression problems (GC), only compression problems (C).

and the resulting coefficients are used to calculate the spectral distance [13].

Principal component analysis and cluster methods were performed as before, after preprocessing of the spectra in order to achieve the best discrimination between classes. These two classification techniques do not require prior knowledge about the spectra. They have the advantage of grouping samples according to their similarities. These groups are called classes or clusters. As we had previously knowledge about spectra classes, we tried to develop PCA and cluster methods, which form group spectra as we expected. Model performance was evaluated on a validation group by comparing the predicted classes to the true classes of the validation samples.

The scaling to first range algorithm was used to calculate the spectral distances in cluster methods. Scaling to first range computes the spectra difference separately for each frequency range chosen, and each frequency range can be weighted differently according to the importance degree we gave it [13]. Cluster hierarchical dendrograms were performed using the Ward's algorithm, which instead of determining the spectral distance, determines the growth of heterogeneity H, and tries to find as homogeneous groups as possible. This means that only two groups are merged when the smallest growth in heterogeneity factor H is observed [13].

Six classification studies were made according to the samples chemical properties or processability behaviour (see Table 2).

3.1.1. First study

Four classes were defined based on water dispersion and hidroalcoholic solubility properties: *G. biloba* batches with positive and negative results in the water dispersion and hidroalcoholic solubility tests. Classes are described as: D+S+, D+S−, D−S+ and D−S−.

Fig. 1 represents the pre-processed average spectra of each class, along the spectral range from 5600 to 6200 cm^{-1}, were major spectral differences are seen.

A second derivative was used as pre-processing technique in the identification method along the spectral range from 5250 to 6350 cm^{-1}. From the 209 spectra used in identification method, 8 can be confused with other reference spectra. This happens because spectral differences between classes are not enough to allow a separation based on an identification threshold. Threshold for a reference spectra is calculated as the maximum distance as listed in the average report plus a fraction of the standard deviation, that was set to 0.25 using OPUS [13]. This was the identification method with best results, but even so, this four-class method does not allow a perfect class by class separation.

Table 2
Classification studies performed and the classes used

Classification parameters		Study number	Classes
	Chemical Properties	1	D+S+ D+S- D-S+ D-S-
		2	D+ D-
		3	S+ S-
		4	W1-2 W2-3
	Processability behavior	5	N B
		6	N GC C

Water dispersion (D), hidroalcoholic solubility (S), normal (N); bad (B), which includes both groups: granulation and compression problems (GC); compression problems (C); water content between 1% and 2% (w1–2) and between 2 and 3 (W2–3).

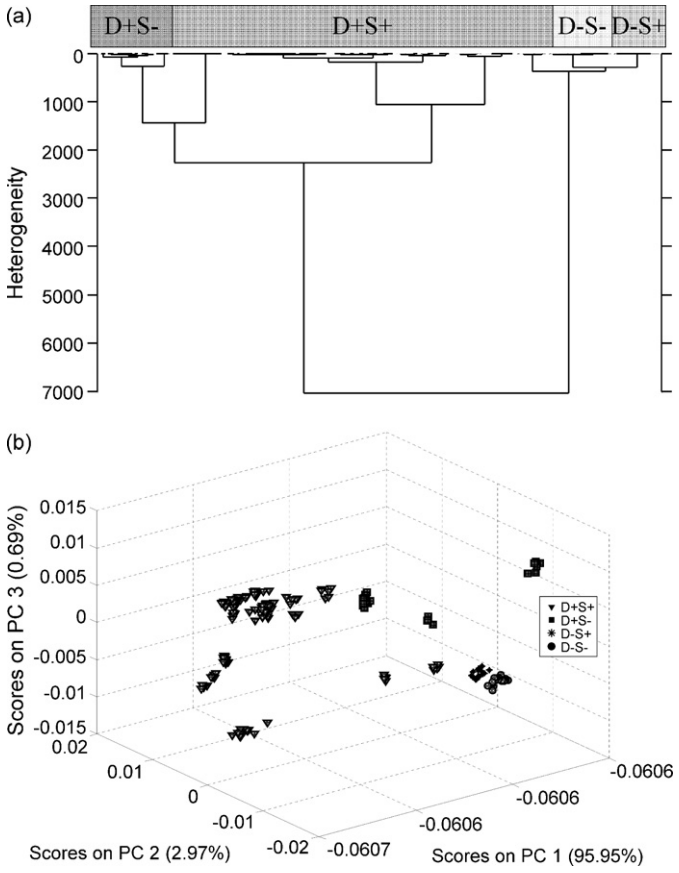

Fig. 2. Cluster from four classes (a) and 2b-PCA from four classes (b).

Cluster method developed uses first derivative and vector normalization as pre-processing techniques. This method separates the spectra in four individualized classes, by the following heterogeneity order: D+S−, D+S+, D−S− and D−S+ (Fig. 2a).

Based on these clustering results, and after the analysis of Fig. 1, it seems possible to separate the *G. biloba* extract according to its water dispersion properties (see below study 2).

Applying a PCA to the spectra, it was possible to distinguish each spectral class from each other using a first derivative and multiplicative scatter correction (MSC) as pre-processing in the spectral range from 4210 to 6350 cm^{-1} (Fig. 2b). The PCA score plot gave four individual clusters, and it is possible to form two major clusters by grouping the two D+ classes and the other joining the two D− classes. On the other hand, it looks that S+ to S− group separation is not possible. In order to have more results and, best responses, we tried to separate the *G. biloba* extract based only on its water dispersion or hidroalcoholic solubility properties (studies 2 and 3).

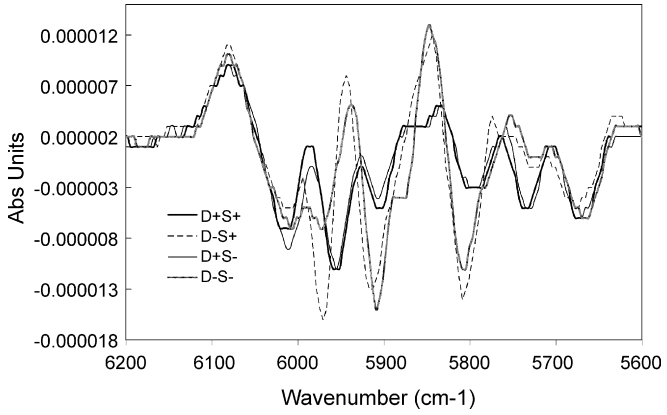

Fig. 1. Second derivative average spectra from four classes.

3.1.2. Second study

The second study was performed in the two classes with different water dispersion properties, i.e., D+ and D−.

The identification method with best discrimination results was obtained using the spectral range from 5300 to 6300 cm^{-1}, using the factorization algorithm and second derivative as pre-processing technique, as in the previous study. This method allows the perfect discrimination of each spectra class, which suggests the existence of different properties from batch to batch, based on the water dispersion parameter.

The cluster method developed also enabled the perfect separation between D+ and D− classes, as we had foreseen by the analysis of Fig. 2a.

A PCA score plot of the pre-processed spectra of two water dispersion groups was performed and supports the previous conclusions. Results confirm that the *G. biloba* spectra

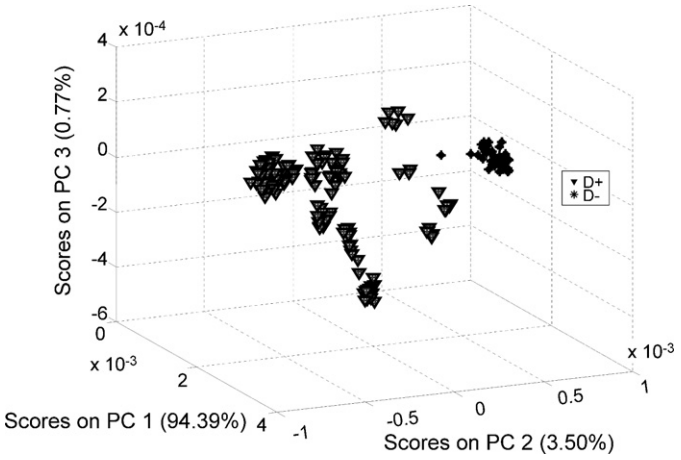

Fig. 3. PCA score plot for two classes based on water dispersion properties.

can be separated based on its water dispersion properties (Fig. 3).

3.1.3. Third study

The third study was performed in the two classes with different hidroalcoholic solubility properties, i.e., S+ and S−. The first derivative of the average spectra of both classes is represented in Fig. 4.

Besides some absorbance differences, it seems that there are no important differences to consider between the two spectral classes. The identification method developed does not allow the perfectly separation of the two classes, because all the S− spectra overlap with the S+ spectra. Applying the same pre-processing in the same spectral region, cluster method separates the two classes (Fig. 5a), but like in the identification method, the PCA method does not allow the perfect separation of the two, S− and S+ groups (Fig. 5b).

Results show that hidroalcoholic solubility parameter could be responsible for some of the differences verified in the incoming *G. biloba* batches, but could not be used to qualify this raw material.

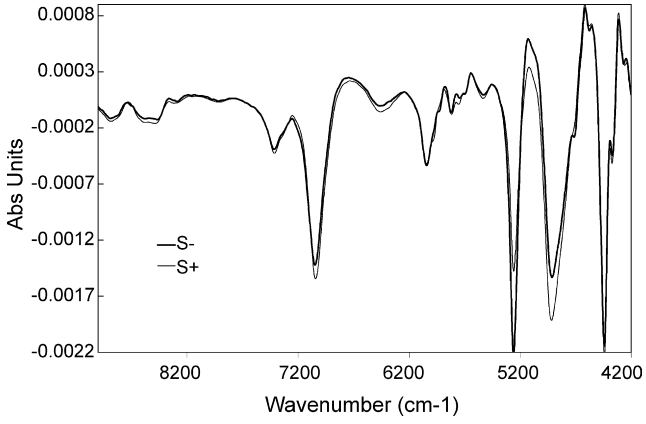

Fig. 4. First derivative of average spectra for two hidroalcoholic solubility classes.

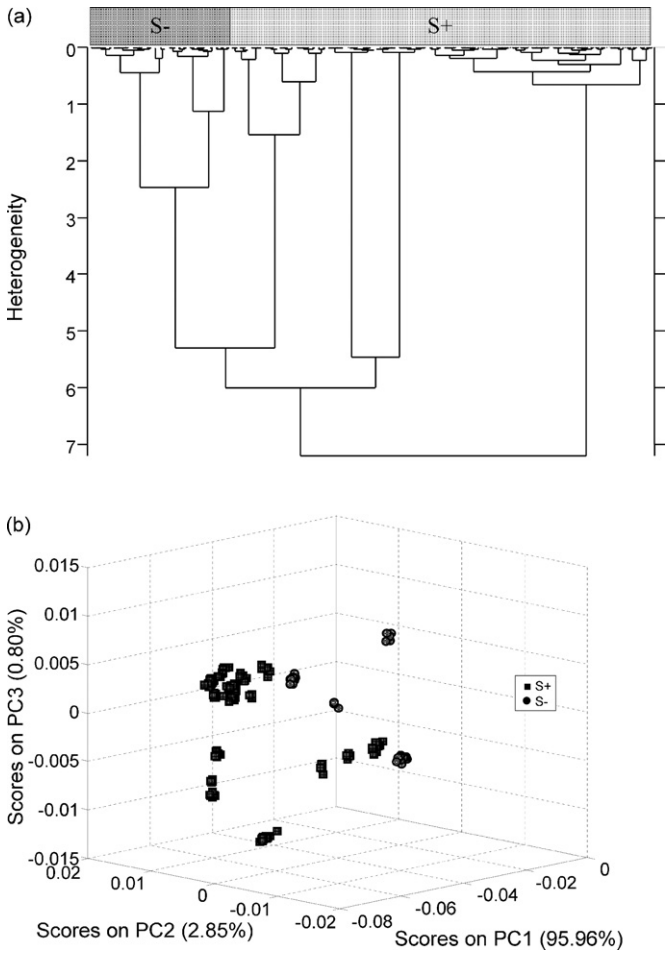

Fig. 5. Cluster (a) and PCA (b) from two hidroalcoholic solubility classes.

3.1.4. Fourth study

The fourth study relates the water content of the incoming *G. biloba* extract batches with its production behaviour. The water content level of the studied *G. biloba* batches is between 1.2% and 2.9% (Table 1). Analyzing Table 1, it is possible to see that almost all the batches with water content behind 2.0% have a good performance in the production; and that almost all batches with water content higher than 2.0% have a bad performance in the production. Batch 14 is an exception to that and should not be used alone for classification purposes.

In order to relate the water content with the production behaviour, two classes were built: W1–2, which includes batches with water between 1.0% and 2.0%; and W2–3, which includes batches with water between 2.0% and 3.0%.

Methods were developed around two spectral regions: 4500–5300 and 6600–7200 cm^{-1}; which are related with the near-infrared water absorption.

All methods developed enable a perfect identification and class by class separation. Fig. 6 represents the cluster and PCA score plot of those methods.

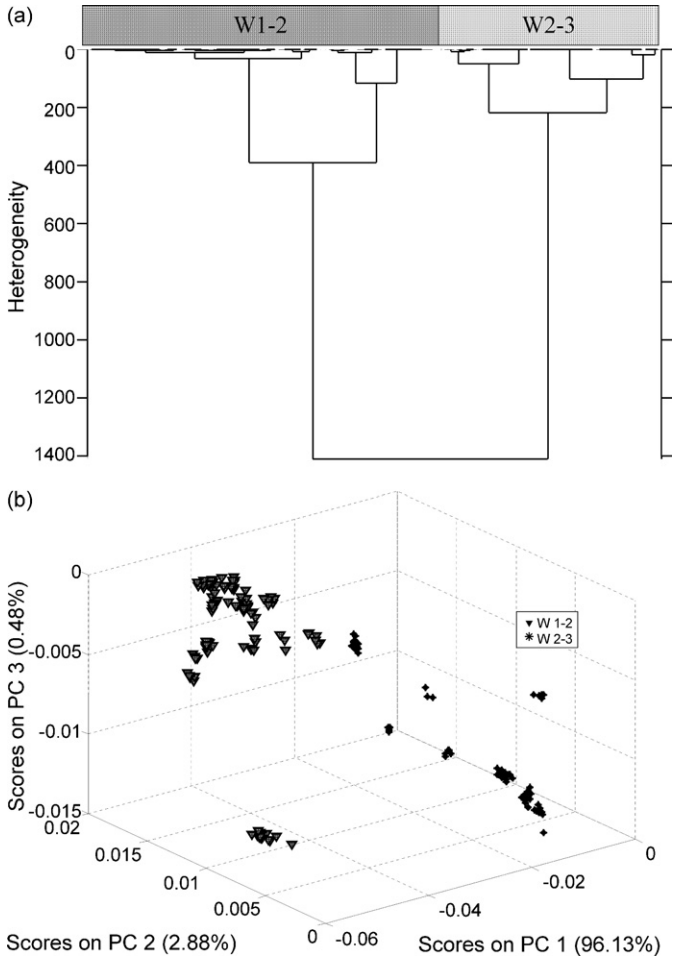

Fig. 6. Cluster (a) and PCA score plot (b) for two water content classes.

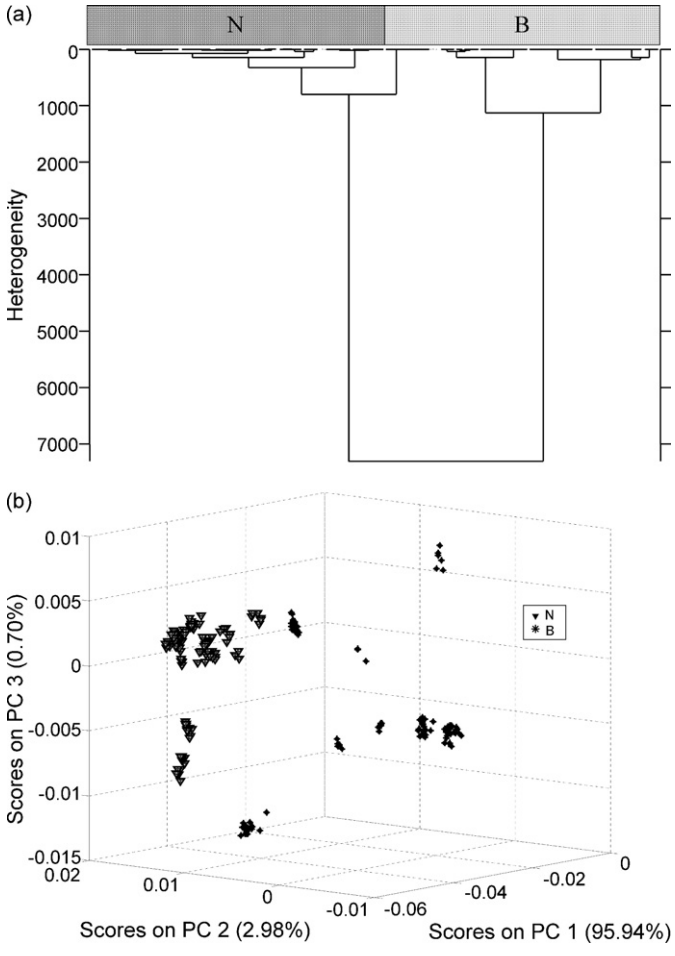

Fig. 8. Cluster (a) and PCA score plot (b) results for two production classes.

3.1.5. Fifth study

The fifth study was made based only on two-class production behaviour—normal (N) and bad (B). Normal class includes batches without any kind of problems; and the bad one includes batches with simultaneous granulation and compression problems.

Although spectral classes have major differences between the range of 5300 and 6300 cm^{-1}, as can be seen in Fig. 7, this spectral range was not sufficient to discriminate the classes by neither of the methods developed; all of them use a wider spectral range (see Table 3). Identification, cluster and PCA methods developed allow the perfect class by class separation and identification (Fig. 8).

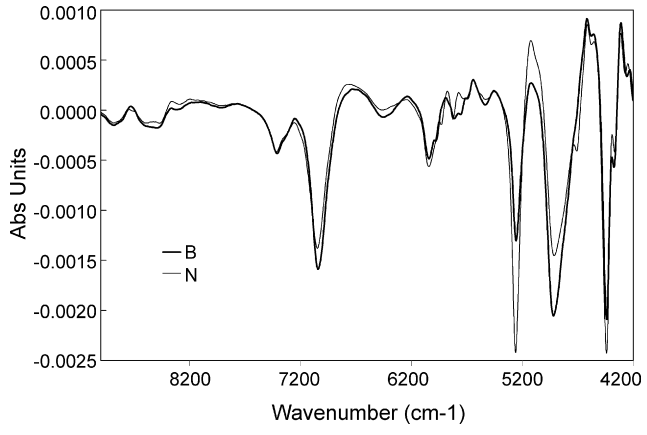

Fig. 7. First derivative of two production classes.

3.1.6. Sixth study

The sixth and last study was made dividing spectra in three classes according to production behaviour—normal (N), compression problems (C) and simultaneous granulation and compression problems (GC).

It was not possible to separate spectra by an identification method, because the three groups used do not have enough differences to allow this type of classification, even in the spectral range from 4950 to 6350 cm^{-1}.

The best clustering result is the same of the previous study, because spectra are the same, only considering three classes instead of two (Fig. 9a). The PCA score plot in Fig. 9b reflects the same as in the cluster: a unique and agglomerated cluster for the normal class and a mixed cluster for both compression and granulation plus compression class.

Classes C and GC are not distinguishable by PCA (Fig. 9b) because spectral differences are slight when compared to the normal class spectra, and sometimes the correct definition of

Table 3
Best results obtained in classification methods for each study

Study number	Method parameters	Classification method		
		Identification	Cluster	PCA
1	PP	2nd Der 5SP	1st Der + VN 9SP	1st Der + MSC 15SP
	SR	5250–6350	4600–5300; 5690–6200	4210–6350
2	PP	2nd Der + VN 9SP	2nd Der 9SP	2nd Der 15SP
	SR	5300–6300	5300–6300	5250–6350
3	PP	1st Der + VN 5SP	1st Der + VN 9SP	1st Der + Nor 15SP
	SR	4300–6950	4300–6950	4300–6950
4	PP	1st Der 9SP	1st Der + VN 13SP	1st Der + MSC 15SP
	SR	4480–5320	4500–5380; 6515–7250	4480–5320; 6610–7160
5	PP	1st Der + VN 5SP	1st Der + VN 9SP	1st Der + MSC 15SP
	SR	4300–6950	4560–6315	4200–6350
6	PP	2nd Der + VN 5SP	1st Der + VN 9SP	1st Der + MSC 15SP
	SR	5300–6300	4560–6315	4220–6350

Pre-processing (PP), spectral range (SR), derivative (Der), smoothing points (SP), vector normalization (VN), multiplicative scatter correction (MSC), mean center (MC).

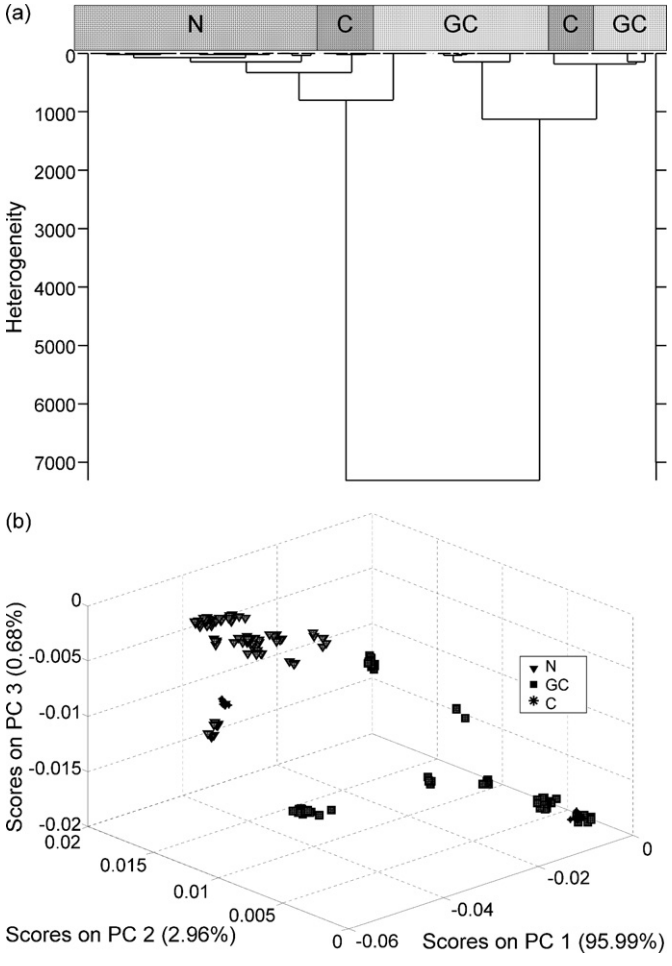

Fig. 9. Cluster (a) and PCA score plot (b) for three production classes.

production behaviour may not be consistent from batch to batch. Some *G. biloba* batches are used in more than one tablet manufacturing batch, its behaviour is not always exactly the same. These slight differences in raw material behaviour became the production behaviour class selection dubious, and for that reason it is better to join both classes with undesired behaviour, as done in the fifth study.

Table 3 describes the best classification results for each study and method type.

3.1.7. Validation results

In order to conclude about the best approach to classify the *G. biloba* extract, a validation set of seven batches (Table 1) was used to challenge all the 18 methods developed.

Best classification results, with no false positives or multiple classification responses, are obtained in the second, fourth and fifth study approach, in which we used, respectively, water dispersion properties, water content and processability behaviour based on two classes to classify *G. biloba* extract. Cluster and PCA methods gave in general better results than the identification methods, but the only method type without any dubious result was the cluster.

3.1.8. Relationship between quality control parameters, production processability and final product

From all the parameters investigated in the previous described studies, three of them show to be adequate to qualify *G. biloba*: (1) water dispersion, (2) water content and (3) production processability based on two classes. Besides that, by the analysis of Table 1, it is possible to see that all the batches with negative results in the water dispersion and hidroalcoholic solubility had bad production behaviour, so it is possible to conclude that those parameters both have influence on the production behaviour of *G. biloba*. Some of those batches with positive results in these two parameters had also bad production behaviour, but all of them were batches with a water content level larger than 2.0%. Finally, all the batches with a water content level larger than 2.0% were batches with bad production behaviour. The main conclusion is that the production processability of the incoming *G. biloba* batches is mainly influenced by the water content level.

We investigate also, in a theoretical way the relationship between this quality control and production processability features with some of the final product (viz. coated tablets) analytical results: (1) water content, (2) dissolution profile and (3) hardness. Neither one of these three parameters shows to have a relation with the quality control parameters or production processability problems described before. This is due probably to the fact that batch to batch differences of *G. biloba* are not so great to influence the final product analytical results.

4. Conclusions

Some *G. biloba* properties measured in raw material have been shown to have great influence in granulation, and compression problems during tablet manufacturing process. Variability in those properties has a minor effect in the final product quality, but even so it is very important to qualify *G. biloba*, in order to avoid problems during the manufacturing process.

From all the parameters investigated, the dispersion in water, moisture content and production processability based on two classes of behaviours show to be adequate to qualify *G. biloba*, using a cluster based method. The methods developed to qualify *G. biloba* based on the two classes separation (normal and bad production behaviour), revealed to be the most effective to classify batches for validation of the proposed strategy. The water content level of *G. biloba* is the quality control parameter with the greater influence in its subsequent processability, but even so, it is not sufficient by itself to classify *G. biloba*, because the correspondence between the two water content classes used, and the two production behaviour classes observed is not linear (Table 1). The main conclusion is that the production processability of the incoming *G. biloba* batches is mainly influenced by the water content level of the bulk raw material, and that the best way to qualify this raw material is the two production class method approach, complemented with those based on the water content level. NIR spectroscopy has revealed to be a powerful technique for predicting raw material processability behaviour, with no need for additional quality control tests.

This could be considered a process analytical technology application since it: (1) establishes the source of variability in a qualitative way, (2) explains its propagation to the final product quality attributes and (3) lays the basis for a control strategy to be applied in the manufacturing process.

Acknowledgments

The authors wish to thanks to Laboratórios Atral S.A., for supporting this works and for permission to publish the above results. S.R. gratefully acknowledges financial support from the Portuguese Foundation for Science and Technology (research grant no. SFRH/BDE/15521/2004).

References

[1] H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise, Near-Infrared Spectroscopy: Principles, Instruments, Applications, Wiley-VCH, Weinheim, Germany, 2002.
[2] G. Reich, Adv. Drug Deliv. Rev. 57 (2005) 1109–1143.
[3] E.W. Ciurczak, J.K. Drennen III, Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy, Marcel Dekker, Inc., New York, 2002.
[4] K. Krämer, S. Ebel, Anal. Chim. Acta 420 (2000) 155–161.
[5] M.S. Kemper, L.M. Luchetta, J. Near Infrared Spectrosc. 11 (2003) 155–174.
[6] J. Luypaert, et al., Talanta 72 (3) (2007) 865–883.
[7] T. Herring, LCGC N. Am. 22 (5) (2004) 456–462.
[8] T.A. van Beek, J. Chromatogr. A 967 (2002) 21–55.
[9] Ginkgo Dry Extract, Purified and Standardised. Pharmeuropa, The European Pharmacopeia Forum. vol. 18, no. 3, EDQM, strasbourg, July 2006, pp. 436–440.
[10] Y. Woo, H. Kim, J. Cho, H. Chung, J. Pharm. Biomed. Anal. 21 (1999) 407–413.
[11] W. Yoon, R. Jee, A. Charvill, G. Lee, A. Moffat, J. Pharm. Biomed. Anal. 34 (2004) 933–944.
[12] PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance Guidance for Industry, US Food and Drug Administration, Rockville, 2004, http://www.fda.gov/cder/guidance/6419fnl.pdf.
[13] OPUS, Optic Users Software Manual, Version 4.2, Bruker Optics GmbH, Germany, 2002.