

Chemometric analysis applied in ^1H HR-MAS NMR and FT-IR data for chemotaxonomic distinction of intact lichen samples

Glaucia Braz Alcantara^{a,*}, Neli Kika Honda^b,
Márcia Miguel Castro Ferreira^c, Antonio Gilberto Ferreira^a

^a Departamento de Química, Universidade Federal de São Carlos, P.O. Box 676, CEP 13565-905, São Carlos/SP, Brazil

^b Departamento de Química, Universidade Federal de Mato Grosso do Sul, P.O. Box 549, CEP 79070-900, Campo Grande/MS, Brazil

^c Instituto de Química, Universidade Estadual de Campinas, P.O. Box 6154, CEP 13084-971, Campinas/SP, Brazil

Received 14 October 2006; received in revised form 12 March 2007; accepted 15 March 2007

Available online 18 March 2007

Abstract

This paper describes the potentiality of chemometric analysis applied in ^1H HR-MAS NMR and FT-IR data for lichen chemotaxonomic investigations. Lichens present a difficult morphologic differentiation and the chemical analyses are frequently employed for their taxonomic classification, mainly due to the secondary metabolites to be relatively constant for these organisms. The lichen chemotaxonomic classification is usually carried out by color reactions, chromatography, fluorescence and mass spectrometry analysis, where the identification is obtained by one or more techniques. There are some papers which use the carbohydrate content in chemotaxonomy investigation. However, the majority of these techniques involve laborious and time consuming sample pre-treatment. This work focuses on application of ^1H high resolution magic angle spinning – nuclear magnetic resonance (HR-MAS NMR) and Fourier transform infrared (FT-IR) associated with chemometric analysis to intact samples. In comparison to other traditional techniques, ^1H HR-MAS NMR and FT-IR allied with chemometrics provided a fast and economic method for lichen chemotaxonomy. Both methods were useful for lichen analysis and permitted the satisfactory distinction among families, genera and species, although better results were achieved for FT-IR data.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Lichens; High resolution magic angle spinning – nuclear magnetic resonance; Fourier transform infrared; Chemometrics

1. Introduction

Lichens comprise a diversity of about 13,500 species growing worldwide, distributed about 600 genera and 80 families which correspond to about 20% of known fungi [1]. The chemical analyses of lichens are frequently employed to their taxonomic classification, principally because their secondary metabolites are reasonably invariable within the species and also because most of them are exclusively found in lichens [2,3]. The lichen chemotaxonomic classification is usually made from color reaction, chromatography, fluorescence and mass spectrometry analysis. So, the identification is obtained by one or more of these techniques [4,5].

Carbohydrate compositions of lichens are considerably investigated by ^{13}C NMR [6–9], which demonstrate to be useful in

chemotaxonomic studies. Among polysaccharides, two classes are present: one of them is the heteropolysaccharides, including mannose-containing polysaccharides, and the other is the glucans [10]. Polyols such as arabinitol, D-mannitol, ribitol and other soluble carbohydrates are also abundant in lichens and may constitute about 5% of the dry weight of the thallus [11]. However, the majority of such analyses requires sample pre-treatment which are time and reagent consuming. Moreover, study of lichen carbohydrate compositions requires detailed ^{13}C NMR analysis. Therefore, faster methods are largely required and could present great importance for lichen chemotaxonomy.

In this context, direct investigations applied in intact lichen samples can be presented as a great alternative. High resolution magic angle spinning – nuclear magnetic resonance (HR-MAS NMR) and FT-IR (Fourier transform infrared) spectroscopy can be useful analytical techniques which reduce drawbacks connected to sample preparation and/or the time rate for the analysis.

* Corresponding author. Tel.: +55 16 33518066.

E-mail address: glabraz@yahoo.com.br (G.B. Alcantara).

HR-MAS NMR technique combines the typical advantages of solid and liquid-state NMR techniques and has recently been reported in literature as an analytical tool in study of some metabolites in animal tissue [12,13], vegetal material [14–16] and for solid food product investigations [17,18]. FT-IR has also been used in several analyses of intact materials, such as identification of seven species of *Lactobacillus acidophilus* group [19] and many analyses of food products [20,21]. Recently, FT-IR has been successfully applied for the identification and intra-species characterization of airborne filamentous fungi [22].

However, analyses of materials without any pre-treatment usually originate complex spectra, which cannot be interpreted by visual examination. Therefore, chemometrics can be a useful tool to extract hidden information of data with high similarity, mainly those obtained by NMR and FT-IR. Without appropriate methods of data analysis, the spectroscopic details which potentially make these experimental techniques so powerful would become useless [23]. Chemometric analyses are particularly appropriate to areas such as food investigations [24–26], plant extracts [27] and clinical chemistry [28–30], where successful classification or diagnosis depend on simultaneous consideration of several variables.

This paper focuses on the chemotaxonomic distinction of intact lichen samples by the application of the chemometric methods on data originated by both techniques, ^1H HR-MAS NMR and FT-IR. Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used to explore the data and the classification chemometric method *K*th nearest neighbor (KNN) was applied to prediction of unknown samples using these spectroscopic techniques.

2. Experimental

2.1. Origin of samples and preparation

Eleven species of lichen samples from six genera and two different families were collected in Mato Grosso do Sul state – Brazil (Table 1). The samples employed in this study were

Table 1
Families, genera and species of the analyzed lichen samples

Families	Genera	Species	Codes ^a	
Physciaceae	<i>Dirinaria</i>	<i>aspera</i> (H. Magn.) Awasthi	Dirin	
	<i>Heterodermia</i>	<i>speciosa</i> (Wulf.) Trevisan	Hspec	
	<i>Pyxine</i>	<i>daedalea</i> Krog & R. Sant.	Pyx	
Parmeliaceae	<i>Hypotrachyna</i>	<i>dactylifera</i> (Vain.) Hale	Hypot	
	<i>Canoparmelia</i>	<i>cryptochlorophaea</i> (Hale) Hale	Canop	
	<i>Parmotrema</i>		<i>breviciliatum</i> (Hale) Hale	Pbrev
			<i>cornuta</i> (Lynge) Hale	Pcor
			<i>delicatulum</i> (Vain.) Hale	Pdel
			<i>dilatatum</i> (Vain.) Hale	Pdil
			<i>mesotropum</i> (Müll. Arg.) Hale	Pmes
	<i>tinctorum</i> (Nyl.) Hale	Ptinc		

^aSamples analyzed by ^1H HR-MAS NMR are with the prefix “HR” before code.

systematically selected to represent different families, genera and species. The sample selection was carried out taking into consideration their availability, known botanic identification and previous chemical studies. Although the number of lichen species in the world is very extensive, this preliminary work introduces a new perspective of chemotaxonomic analysis of lichens.

After identification performed by morphology, thin layer chromatographic and/or microcrystallization analyses, the samples were previously cleaned and then powdered in a cryogenic mill.

For HR-MAS NMR and FT-IR analyses, the lichens were used without any pre-treatment. The spectroscopic analyses were carried out in triplicate for each lichen species. Each replicate was analyzed with the introduction of a new material in the spectrometer.

2.2. ^1H HR-MAS NMR spectra

All ^1H HR-MAS NMR measurements were carried through on a Bruker Avance DRX 400 instrument (operating at 400.13 MHz for ^1H) equipped with a 4 mm HR-MAS probehead and zirconium rotor. A few drops of D_2O were added in the samples for field homogeneity adjustment. The spectra were collected using 5 kHz spinning speed without temperature regulation using the Carr–Purcell–Meiboom–Gill (CPMG) spin-echo pulse sequence. Water suppression was included in the CPMG sequence. The CPMG pulse sequence is as follows: $D - [-90^\circ - (\tau - 180^\circ - \tau)_n - \text{FID}]$, which $D = 1.0$ s to allow T_1 relaxation; $\tau = 1.250$ ms was fixed after optimisation to permit the broad signals attenuation (“ T_2 filter”) and refocusing of spin-coupled multiplets; n = a fixed loop of 150 cycles, giving a total spin–spin relaxation delay $2n\tau$ of 375 ms. Typically, 512 free induction decays (FIDs) were collected into 32 K data points using a 11.8 μs pulse width (90° pulse angle) and an acquisition time of 2.06 s. Prior to Fourier transformation (FT), the FIDs were zero-filled and an exponential weighing factor corresponding to line broadening to 1 Hz was applied. The acquired

NMR spectra were phase corrected and referenced using TMSP-2,2,3,3-D₄ (sodium-3-trimethylsilylpropionate) like an internal reference.

2.3. FT-IR spectra

All FT-IR spectra were registered in a Bomem Hartmann & Braun MB-Series model 102 spectrometer, obtained at 4000 to 400 cm⁻¹ region with a spectral resolution of 4 cm⁻¹ and 16 scans. The KBr disks were prepared using powdered 1:99 mg samples of lichen and dry potassium bromide, respectively, which produced translucent pellets.

2.4. Data treatment

The dots from ¹H HR-MAS NMR and FT-IR spectra were transformed into ASCII files and the resulting data matrices were imported into the Origin software (v. 5.0, Microcal, USA). After this, the selected regions from both spectra were imported into Pirouette[®] (v.2.02, Infometrix, USA) for principle component analysis (PCA), hierarchical cluster analysis (HCA) and *K*th nearest neighbor (KNN). For the NMR data the region from 3.0 to 4.0 ppm was selected for statistical analysis because it presented the most important signals for chemometric distinction. In the FT-IR data, the water absorption band (ν 4000–3000 cm⁻¹) and the noisy regions (ν 2800–1800 cm⁻¹ and ν 900–400 cm⁻¹) were removed before statistical analysis, in order to guarantee secure results.

Data analyses on both generated data matrices (HR-MAS NMR and FT-IR) were performed by the Pirouette[®] software (v.2.02, Infometrix, USA). PCA and HCA methods were used for data exploration and the method KNN was used for classification.

3. Results and discussion

3.1. Spectral data for multivariate analysis

The HR-MAS NMR technique employs the magic angle spinning to minimize the dipolar couplings and chemical shift anisotropy effects which provide the broadening of signals due to the restricted molecular motion. In Fig. 1A and B, respectively, we can observe the comparison between ¹H NMR spectrum obtained without magic angle spinning, at 54.74°, and ¹H HR-MAS NMR spectrum. This result shows a reduced resonance linewidth and thus the spectral quality is significantly improved.

The use of CPMG pulse sequence in HR-MAS NMR analysis permitted to minimize the broadening of resonance signals due to high molecular weight components, which do not provide useful information for the multivariate analysis. Moreover, better resolution was achieved when CPMG pulse sequence was employed. In Fig. 1C, we can observe the ¹H HR-MAS NMR spectra obtained without the CPMG pulse sequence. The comparison between ¹H HR-MAS NMR spectra with and without the CPMG sequence (Fig. 1B and C, respectively) shows the gain of resolution when this sequence is used. Therefore, the

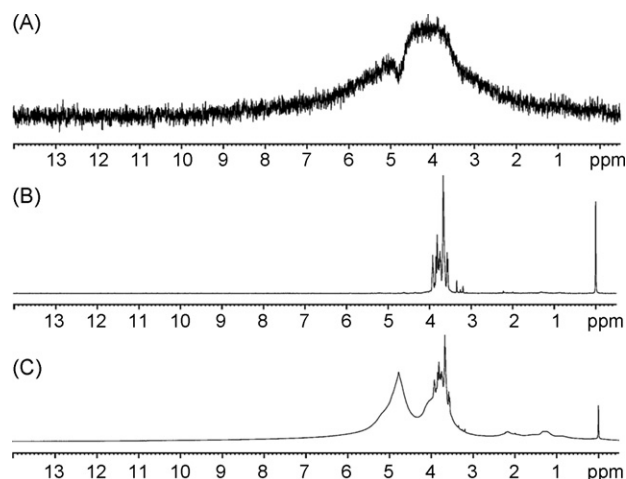


Fig. 1. Lichen ¹H NMR spectra: (A) without magic angle spinning; (B) HR-MAS with CPMG pulse; and (C) HR-MAS without CPMG pulse.

lichen analysis was put in effect with the HR-MAS technique added to the CPMG sequence.

Both spectral data (HR-MAS and FT-IR) have essentially presented carbohydrate signals once these compounds are predominantly found in intact material analysis. ¹H HR-MAS NMR lichen spectra presented the typical carbohydrate signals at about δ 3.0–4.0 ppm (Fig. 2A), which comprehend the polyol region (very common in lichens), with ¹³C NMR signals at about δ 66–76 ppm. The NMR assignments described in Fig. 2A correspond to arabinitol (Ara-H1, H5: δ 3.56 ppm; Ara-H2, H4: δ 3.93 ppm; Ara-H3: δ 3.62–3.71 ppm), mannitol (Man-H1, H6: δ 3.83 ppm; Man-H2, H5: δ 3.75 ppm; Man-H3, H4: δ 3.62–3.71 ppm) and non identified compounds (δ 3.18–3.37 ppm). All ¹H HR-MAS NMR spectra showed a very similar profile and just a few differences were detected at some constituent proportions. Therefore, only this region was used for chemometric analysis.

FT-IR spectra also presented a high similarity without any pronounced differences between families, genera and species. These spectra corresponded to the some characteristic absorption regions (Fig. 2B), such as: C–H stretching vibration (3000–2800 cm⁻¹); C=O stretching vibration (1700–1600 cm⁻¹); C–H bending vibration (1500–1200 cm⁻¹) and C–O stretching vibration from carbohydrates (1200–900 cm⁻¹). These regions of FT-IR spectra were applied to chemometric analysis. The observed absorptions concern to low molecular weight carbohydrate, specifically to polyols and monosaccharides.

3.2. Chemotaxonomic distinction with PCA and HCA methods

PCA was performed on the matrix data of ¹H HR-MAS NMR and FT-IR. Several pre-treatments were tested and the best results for the NMR data matrix were obtained when first derivative and normalization (vector length normalization) were applied to the samples and the resulted data was autoscaled (mean centered and scaled to unit variance). Chemometric analysis of FT-IR data presented the best results for princi-

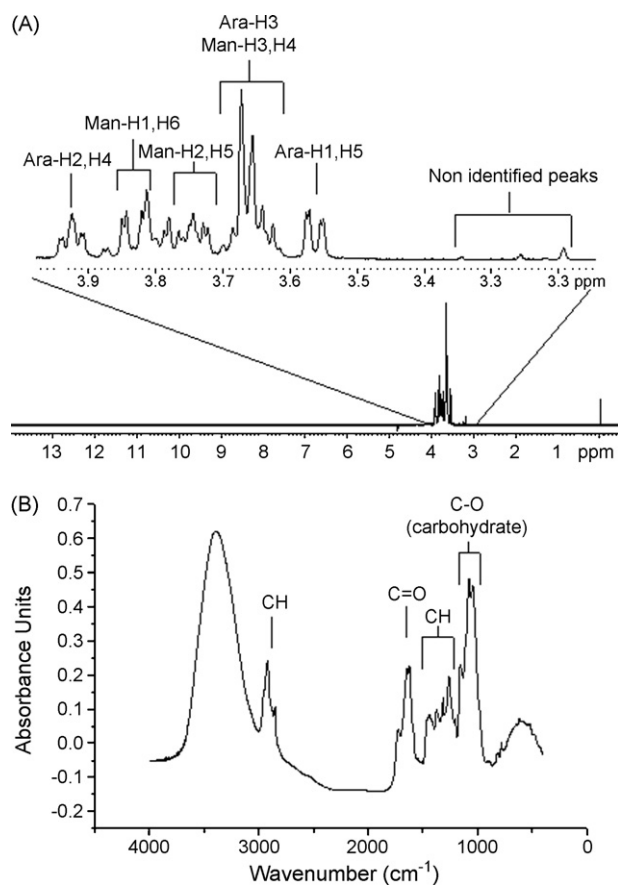


Fig. 2. Representative spectral profile for a lichen sample: (A) ^1H HR-MAS NMR spectrum with signal description of arabinitol (Ara), mannitol (Man) and non identified compounds; and (B) FT-IR spectrum with carbohydrate absorptions in prominence.

pal component analysis (PCA) with the first derivative applied on samples and autoscale preprocessing. The autoscale preprocessing applied on variables of spectroscopic data was very important due to the fact that, for the differences on intensities of signals (^1H HR-MAS NMR) or absorption bands (FT-IR), the same importance for all regions was attributed.

PCA for HR-MAS NMR data presented the separation of the lichens into families (Fig. 3). Parmeliaceae family was located

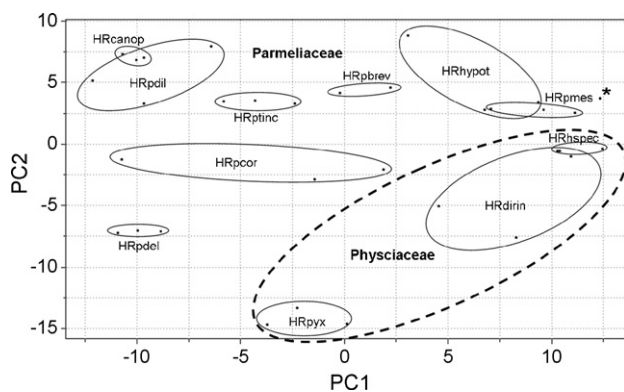


Fig. 3. PCA scores plot of all lichens analyzed by ^1H HR-MAS NMR with lichen family distinction (Physciaceae family inside hatched ellipse) – 34.57% of the total variance.

on the most positive side of PC2 axis. On the negative side of PC2 axis and the most positive side of PC1 axis the Physciaceae family was located (inside hatched ellipse). In this case, *Parmotrema delicatulum* (Parmeliaceae sample) was located on the most negative side of PC2 (outlier). Another outlier for a *P. breviciliatum* replicate was observed and it is represented by asterisk in Fig. 3. The examination of PC1 and PC2 loadings of ^1H HR-MAS NMR data suggested that this separation occurs due to spectral domains situated in δ 3.54–58 and δ 3.90–95 ppm (arabinitol) on the positive side of PC1 axis. On the negative side of PC2, the responsible loadings were the resonances at δ 3.18–3.37 ppm (non identified compounds). However, some overlapping of different species were observed for Canop and Pdil, Pmes and Hypot, Hspec and Dirin species (see Fig. 3) and thus represented an important limitation of the method.

HCA analysis was also performed on the matrix data of ^1H HR-MAS NMR. In this analysis, the Euclidean distance was used as metric and complete linkage method (farthest neighbor clustering – this technique assigns a sample to cluster whose farthest neighbor is closest to the sample) was employed. Comparison between PCA and HCA for NMR data presented a similar result in relation to family distinction. The dendrogram in Fig. 4 shows two main clusters with a similarity index of 0.217, separating between Physciaceae and Parmeliaceae families in HR-MAS NMR analysis. Only one sample, *Parmotrema delicatulum*, presented an unusual behavior.

FT-IR data treatment did not distinguish the lichen families. But, in Fig. 5, the PCA scores plot shows separation of genera. The *Parmotrema* genus (presented in hatched square) differed from the other genera from Parmeliaceae family on the most positive side of PC3. The assessment of PC1 and PC3 loadings of FT-IR data suggested that the absorption bands at about $1400\text{--}1200\text{ cm}^{-1}$ (C–H bending vibration) are responsible for this separation on the negative side of PC3.

Chemometric analysis for both data sets showed the distinction among species. The separation into six species of *Parmotrema* genus was found (Fig. 6A shows NMR results and Fig. 6B shows FT-IR results). The regions responsible for the distinction of species for NMR data in Fig. 6A are δ 3.18–3.37 (non identified compounds); 3.54–3.58 and 3.90–3.95 ppm (arabinitol). For FT-IR data, the loadings in Fig. 6B correspond to absorptions at $1400\text{--}1200\text{ cm}^{-1}$ (C–H bending vibration) and $1700\text{--}1600\text{ cm}^{-1}$ (C=O stretching vibration).

It is important to emphasize that the replicates grouping was better when FT-IR was employed. The less reproducibility observed in the HR-MAS technique depends dramatically on the sample insertion inside the HR-MAS rotor, because when we add deuterated water drops and put the rotor spacer, part of sample and water can be expelled; therefore, the triplicate grouping for HR-MAS NMR demonstrated less satisfactory results.

3.3. Chemotaxonomic classification with KNN method

Chemotaxonomic prediction of lichen samples was performed by the KNN method, in which an unknown pattern was classified according to the majority of the votes of its K th nearest neighbors in the n -space [32]. The same prepro-

Table 2
Lichen classification by KNN method for ^1H HR-MAS NMR and FT-IR data sets

Unknown samples	Classification by KNN		True classes
	HR-MAS NMR	FT-IR	
US1a	Hspec	Hspec	Hspec
US1b	Hspec	Hspec	Hspec
US1c	Hspec	Hspec	Hspec
US2a	Dirin	Dirin	Dirin
US2b	Dirin	Dirin	Dirin
US2c	Dirin	Dirin	Dirin
US3a	Pbrev	Pbrev	Pbrev
US3b	Canop	Pbrev	Pbrev
US3c	Pbrev	Pbrev	Pbrev
US4a	Pcor	Pcor	Pcor
US4b	Pcor	Pcor	Pcor
US4c	Pcor	Pcor	Pcor
US5a	Pdil	Pdil	Pdil
US5b	Pmes	Pdil	Pdil
US5c	Pdil	Pdil	Pdil
US6a	Pmes	Pmes	Pmes
US6b	Pmes	Pmes	Pmes
US6c	Pmes	Pmes	Pmes
US7a	Pbrev	Ptinc	Ptinc
US7b	Pmes	Ptinc	Ptinc
US7c	Pmes	Ptinc	Ptinc
US8a	Pyx	Pyx	Pyx
US8b	Pyx	Pyx	Pyx
US8c	Pyx	Pyx	Pyx

Although FT-IR data treatment did not separate lichen families, when the KNN method was performed, we observed an excellent chemotaxonomic classification. This can be explained by the fact that, in KNN, the prediction model is constructed by using the physical closeness of samples in space. Whereas, in PCA, the samples are evaluated in order to reveal any natural grouping without using class membership information in calculations. So, if the class of samples is more compact, small prediction errors are observed in KNN. Therefore, the best grouping of triplicate for FT-IR analysis implicated excellent classification.

4. Conclusions

Metabolic profiling by ^1H HR-MAS NMR and FT-IR together with chemometric analysis contributed to the lichen chemotaxonomic characterization. In comparison to other traditional techniques, ^1H HR-MAS NMR and FT-IR allied with chemometrics provided a fast and economic method for lichen identification. Both methods were useful for lichen analyses and permitted the satisfactory distinction among families, genera and species, although better results were achieved for FT-IR data. This result indicates chemometrics as a powerful tool for chemotaxonomic analysis.

Acknowledgements

This work has been financially supported by FAPESP, CNPq, CAPES and FINEP.

References

- [1] T.H. Nash III, *Lichen biology*, Cambridge University Press, Cambridge, 1996, 303 p.
- [2] W. Quilhot, G. Leighton, E. Flores, E. Fernandes, W. Pena, G. Guzman, *Acta Farm. Bonaerense* 6 (1) (1987) 15–22.
- [3] V.L. Eifler-Lima, A. Sperry, S. Sinbandhit, J. Boustie, S. Tomasi, E. Schenkel, *Magn. Reson. Chem.* 38 (2000) 472–474.
- [4] N.K. Honda, W. Vilegas, *Quím. Nova* 21 (6) (1998) 110–125.
- [5] E.R. Carbonero, G.L. Sasaki, P.M. Stuelp, P.A.J. Gorin, S.M. Woranovicz-Barreira, M. Iacomini, *FEMS Microbiol. Lett.* 194 (1) (2001) 65–69.
- [6] I. Yokota, S. Shibata, S. Saitô, *Carbohydr. Res.* 69 (1979) 252–258.
- [7] A.Z.A. Teixeira, M. Iacomini, P.A. Gorin, *Carbohydr. Res.* 266 (2) (1995) 309–314.
- [8] P.M. Stuelp, A.M.A. Carneiro Leão, P.A.J. Gorin, M. Iacomini, *Carbohydr. Pol.* 40 (1999) 101–106.
- [9] E.R. Carbonero, A.V. Montai, P.M. Stuelp, S.M. Woranovicz-Barreira, P.A.J. Gorin, M. Iacomini, *Phytochem.* 61 (2002) 681–686.
- [10] E.R. Carbonero, A.V. Montai, C.G. Mellinger, S. Eliasaro, G.L. Sasaki, P.A.J. Gorin, M. Iacomini, *Phytochem.* 66 (2005) 929–934.
- [11] M.E. Hale Jr., *The biology of lichens*, third ed., Edward Arnold, Baltimore, 1983, 190 p.
- [12] L.L. Cheng, I. Chang, B.L. Smith, R.G. Gonzalez, *J. Magn. Res.* 135 (1998) 194–202.
- [13] D. Moka, R. Vorreuther, H. Schicha, M. Spraul, E. Humpfer, M. Lipinski, P.J.D. Foxall, J.K. Nicholson, J.C. Lindon, *Anal. Commun.* 34 (1997) 107–109.
- [14] A.M. Gil, I.F. Duarte, I. Delgadillo, I.J. Colquhoun, F. Casuscelli, E. Humpfer, M. Spraul, *J. Agric. Food Chem.* 48 (2000) 1524–1536.
- [15] Q.X. Ni, T.M. Eads, *J. Agric. Food Chem.* 41 (1993) 1035–1040.
- [16] A. Broberg, L. Kenne, M. Pedersen, *Planta* 206 (2) (1998) 300–307.
- [17] M.A. Brescia, G. Di Martino, C. Fares, N. Di Fonzo, C. Platani, S. Chelli, F. Reniero, A. Sacco, *Cereal Chem.* 79 (2) (2002) 238–242.
- [18] A. Sacco, I.N. Bolsi, R. Massini, M. Spraul, E. Humpfer, S. Cheli, *J. Agric. Food Chem.* 46 (1998) 4242–4249.
- [19] W. Luginbühl, J. Jimeno, U. Zehntner, *LWT* 39 (2006) 152–158.
- [20] L.M. Reid, C.P. O'Donnell, G. Downey, *Trends Food Sci., Tech.* 17 (2006) 344–353.
- [21] R. Schneider, F. Charrier, M. Moutounet, R. Baumes, *Anal. Chim. Acta* 513 (2004) 91–96.
- [22] G. Fischer, S. Braun, R. Thissen, W. Dott, *J. Microbiol. Methods* 64 (2006) 63–77.
- [23] P.S. Belton, I.J. Colquhoun, E.K. Kemsley, I. Delgadillo, P. Roma, M.J. Dennis, M. Sharman, E. Holmes, K. Nicholson, M. Spraul, *Food Chem.* 61 (2) (1998) 207–213.
- [24] Y.W. Lai, E.K. Kemsley, R.H. Wilson, *J. Agric. Food Chem.* 42 (1994) 1154–1159.
- [25] M. Defernez, E.K. Kemsley, R.H. Wilson, *J. Agric. Food Chem.* 43 (1996) 109–113.
- [26] J.T.W.E. Vogels, L. Terwel, A.C. Tas, F. Van den Berg, F. Dukel, J. Van der Greef, *J. Agric. Food Chem.* 44 (1996) 175–180.
- [27] J.L. Ward, C. Harris, J. Lewis, M.H. Beale, *Phytochem.* 62 (2003) 949–957.
- [28] S.L. Howells, R.J. Maxwell, A.C. Peet, J.R. Griffiths, *Magn. Reson. Med.* 28 (1992) 214–236.
- [29] E. Holmes, T.M. Tsang, S.J. Tabrizi, *J. Am. Soc. Exp. NeuroTher.* 3 (2006) 358–372.
- [30] F. Szabo de Edelenyi, A.W. Simonetti, G. Postma, R. Huo, L.M.C. Buydens, *Anal. Chim. Acta* 544 (2005) 36–46.
- [31] G.B. Alcantara, A. Barison, A.G. Ferreira, N.K. Honda, M.M.C. Ferreira, 28th Reunião Anual da Sociedade Brasileira de Química (2005) PN-177.
- [32] B.R. Kowalski, C.F. Bender, *Anal. Chem.* 44 (8) (1972) 1405–1411.