Talanta 106 (2013) 169–173

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Short communication

Quantification of *Coffea arabica* and *Coffea canephora* var. robusta in roasted and ground coffee blends

Laura Ruth Cagliani [a], Gloria Pellegrino [b], Graziella Giugno [b], Roberto Consonni [a,*]

[a] National Council of Research, Institute for the Study of Macromolecules, NMR laboratory, v. Bassini 15, 20133 Milan, Italy
[b] Lavazza spa, st. di Settimo 410, 10156 Turin, Italy

ARTICLE INFO

Article history:
Received 17 August 2012
Received in revised form
29 November 2012
Accepted 3 December 2012
Available online 23 December 2012

Keywords:
Roasted coffee blends
Arabica
Robusta
NMR
OPLS

ABSTRACT

This study reports direct quantification of arabica in roasted and ground coffee blends of *Coffea arabica* and *Coffea canephora* var. robusta. ^{1}H–NMR analysis of water extracts of coffee blends were combined with multivariate statistical analysis to obtain an OPLS model with high predictive capability. This approach allowed to evaluate the composition of coffee blends of unknown arabica and robusta content, on the basis of multiple chemical components. Differences in geographical origin of the analyzed samples did not affected the compositional determination of coffee blends. This approach represents a valid tool in authentication procedures of arabica and robusta blends of roasted and ground coffee.

1. Introduction

Authentication is one of the most challenging issue in food quality control. In the last years, the interest toward the development of new analytical methods for testing the authenticity of food products has experienced a wide growth, in particular for those endowed with high market prices. Among the highly valuated products, coffee represents the second commodity in the world from the economical point of view, involving Africa, Asia and Latin America as the main exporting partners, including additionally more than 70 coffee-producing nations, while the export involves more than 150 countries (http://www.faostat.fao.org). Notwithstanding more than 100 different species have been recognized so far, *Coffea arabica*, *Coffea canephora* var. robusta, *Coffea liberica* and *Coffea excelsa* remain the most popular species. From the commercial point of view, only *C. arabica* and *C. canephora* var. robusta (commonly known as arabica and robusta respectively) represent the two most relevant and widely cultivated species, accounting for 56% and 44% respectively the world's production in 2011, reaching 134 millions of bags. Differences in environmental, growing conditions, methods of processing and drying between arabica and robusta species, led to a sweet and floral/fruity pronounced flavor profile in arabica, normally cultivated on mountain slopes and causing a higher price for this latter species. Conversely robusta is more resistant to disease, grows on flat lands leading to an easier mechanized production and resulted in strong and cocoa flavors. The price gap between the two species has significantly widened in the last years and as a consequence of this, a growing financial incentive to unlawfully replace of high quality arabica coffee with the cheaper robusta is increased. The requirement of a simple, reliable and quick analytical method able to evaluate the ratio between the two species in coffee blends is therefore economically very important. Green coffee beans are relatively simple to differentiate on the basis of their size, shape and color; in case of roasted beans, a size based discrimination could be performed as well. Conversely, the identification and the quantification of arabica and robusta in roasted and ground coffee blends results very challenging. The estimation of arabica fraction can be performed only by sensory or chemical means. Volatile components, even though constitute only 0.1% of the total components, revealed significant changes in the final aroma of coffee [1]. When sensory assessments are concerned, professional taster can estimate the coffee blends composition with an error of about 20% depending on the varieties of coffee and roasting conditions [2]. Conversely, in the field of chemical analysis of roasted coffee, several approaches have been applied with the only aim of differentiating arabica and robusta species, by considering single class of chemical compound such as caffeine [3,4], amino acids [5], chlorogenic acids [6], saccharides [7] and metal content [8]. Lipid fraction was also investigated as well by monitoring fatty acids [9,10], sterols [11,12], diterpenes [13,14] and tochoferols [15,16].

In 1991 Mariani and Fedeli [11] investigated Δ^5 avenasterol and 24-methylen–cholesterol by HPLC analysis, quantified arabica and robusta amount in roasted coffee blends declaring that the

* Corresponding author. Tel.: +39 2 236 99578; fax: +39 2 236 99620.
E-mail address: roberto.consonni@ismac.cnr.it (R. Consonni).

0039-9140/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2012.12.003

limit of detection for robusta addition was 30% and 15% considering the two markers respectively. This was one of the first attempt to evaluate the single species of coffee within a roasted coffee blend, followed few years later by Frega et al. [17]. Although these methods seemed to provide reliable results, pretreatment steps and elaborated methodologies make them scarcely reproducible and limited in frauds screening, being time consuming. Additionally these approaches evaluated single chemical components, that could be modified by fraudsters.

In these last years, analysis of multiple chemical components raised as alternative approach to investigations of single specific marker, resulting as a valid method for food quality assessment [18]. In this respect metabolomics has shown large potentiality in living systems, showing how the metabolic content plays a dominant role for a comprehensive analysis of all metabolite changes upon stimuli [19]. The metabolite content constitutes a sort of a fingerprint for each system under investigation, as well as for food products. Some studies focused on the application of different analytical techniques able to monitor in food matrices different compounds simultaneously. Among them, infrared spectroscopy [20–22] and GC–MS [23] were applied in the view of possible species differentiation of coffee; conversely NIR [24], EI-MS [25] were applied to study the percentage of robusta in coffee blends.

NMR spectroscopy represents an elective technique capable to monitor different class of chemical compounds simultaneously with a single experiment, without any purification or derivatization process of sample, as other techniques require; moreover the NMR analysis are very reproducible and are not time consuming. Particularly the combined use of NMR and chemometrics have been already applied successfully in food authentication [26–30].

Noticeably, NMR has been already applied to roasted coffee in different studies with the aim of chemical characterization [31–34], as well as monitoring roasting process [35,36]. Moreover a relevant paper was suggesting the identification of cluster discriminators of the roast degree in arabica and robusta coffee beans [37].

The aim of our study, presented for the first time in this paper, is the compositional evaluation of arabica content in roasted and ground coffee arabica and robusta blends. This was achieved by applying ^{1}H–NMR spectroscopy to water extracts of coffee blends, enabling the detection of different classes of chemical compounds simultaneously, with a single experiment. The advantage of the NMR measurement joined with chemometrics allowed us to establish an analytical procedure for the authentication of roasted and ground coffee blends.

2. Material and methods

2.1. Analytical set up

Certified arabica and robusta roasted coffee beans were derived from different geographical location and ground following the commercial requirements. Samples were processed under different roasting conditions.

Experimental blends were accurately prepared with arabica composition ranging between 0 and 100% in weight. A minimum of four replicates were considered in double for each blend obtained with different geographical origins of arabica and robusta (African, American and Asian samples) constituting the training set. Other roasted coffee blends of known composition were acquired in double and used as test set. A proper amount of roasted and ground coffee was dissolved in a buffered solution and after extraction was centrifugated. Supernatant solution was inserted into the NMR tube for analysis. ^{1}H–NMR spectra were acquired on a 500 MHz spectrometer (Bruker DMX, Bruker Biospin, GmbH Rheinstetten, Karlsruhe, Germany), with a spectral width of 6900 Hz over 32 k data points. Water suppression was achieved by applying a standard Bruker presaturation pulse program [38]. After Fourier transformation, all spectra were carefully phase and baseline adjusted with TOPSPIN software (Bruker Biospin software TOPSPIN 3.0®). Aligned spectra were subjected to "intelligent" bucket integration to conceal any peak position variation in the range from 0.15 to 10.00 ppm and buckets were normalized to the total integrals. The residual water signal between 4.85 and 4.75 ppm was set to zero constant value with ACD/NMR software (ACD Labs, version 11, Toronto, Ontario, Canada).

2.2. Statistical method

Orthogonal Projection to Latent Structures (OPLS) was performed with mean centering as data pretreatment. The OPLS multivariate projection method was applied to extract linear relationships from two data blocks X and Y and for removing the so-called structured noise [39,40], that contain only uncorrelated information. OPLS removes this structured noise by decomposing the systematic variation in the X block into two model parts: the predictive or parallel part, modeling the joint X–Y correlated variation, and the non-predictive or orthogonal part, not related to Y. The number of latent components can be determined by cross-validation techniques, and in this study, we used 7-fold cross-validation. In addition, a permutation test on the Y block was performed to safely overcome casualty or overfitting into model. Statistical data analysis was performed with the SIMCA-P+13 (Umetrics, Umea, Sweden) software.

3. Results and discussion

As reported elsewhere [41,42], caffeine, trigonelline, N-methyl pyridine, chlorogenic acids, saccharides, amino acids, free quinic and other organic acids represent the most abundant chemical compounds in water extracts of both arabica and robusta species. Concerning chlorogenic acids, different isomers could be detected in water solution, such as caffeic and ferulic acids esterified with quinic acid in position 3, 4 and 5 of this latter moiety, as well as caffeine–chlorogenate complexes [31,43]. All of these components have been detected by one single ^{1}H–NMR experiment for each sample, with their natural occurring abundance and the assignment of the main components for these two species was indicated in Fig. 1. Noticeably, the comparison of the two reported spectra, suggests the main differences in the abundance of chemical compounds for arabica and robusta species. Acetate, trigonelline and formiate were present in larger amount in arabica variety, while robusta presented larger amount of caffeine, according to previous investigations [4]. The real advantage of using this spectroscopic technique relays on the capability to detect simultaneously different classes of chemical compounds, that all together contributed to characterize every single coffee blend. NMR data of the training set samples have been processed to build the OPLS model. The OPLS model resulted in one parallel (t1) and one orthogonal (to1) component and samples appeared well arranged according to their compositional properties, complying the percentage of arabica along the first parallel component (Fig. 2) regardless of the geographical differences among them. The overall goodness of fit (R^2X) resulted 62.9% and 12.8% for the predictive and the orthogonal component respectively with the overall cross validation coefficient Q^2 of 97.0%. To check whether this differentiation could have occurred by chance, Y scrambling validation on the corresponding PLS model was

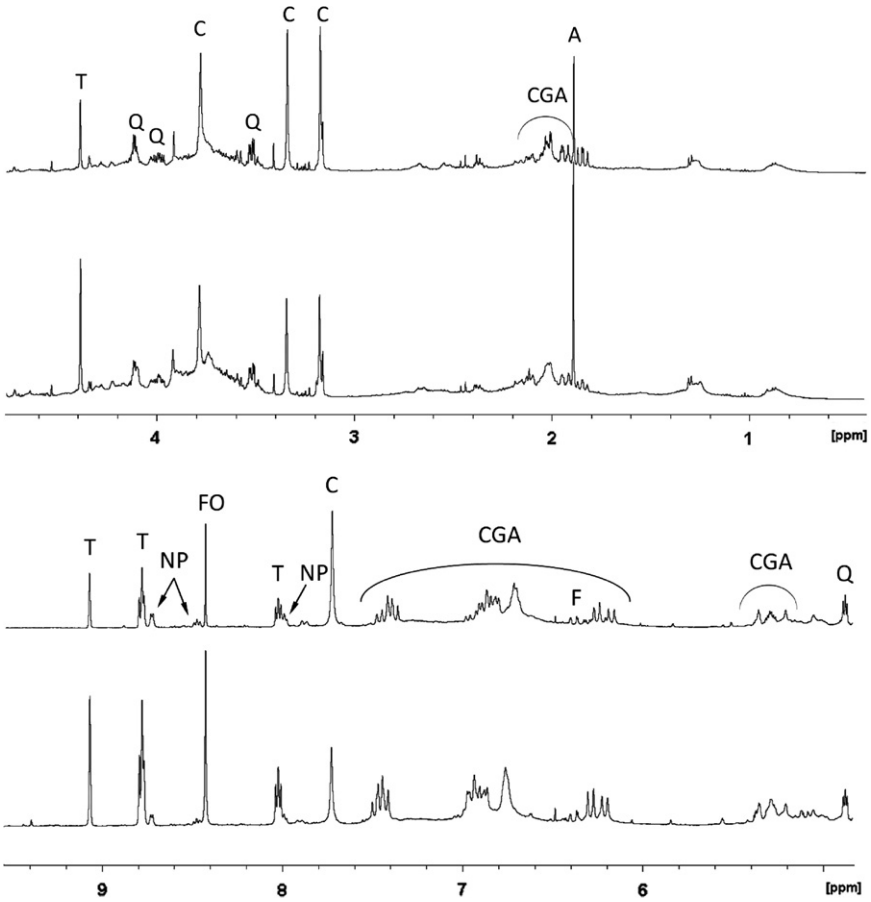

Fig. 1. ^{1}H NMR spectra of water extracts of *C. arabica* (bottom trace) and *C. canephora* var. robusta (top trace) with the assignment of main chemical compounds. Spectra are referenced and scaled against external standard. Letters stands for: A: acetate; CGA: chlorogenic acids; C: caffeine, Q: quinic acids; T: trigonelline; F: 2-furyl methanol; NP: N-methyl pyridine; and FO: formiate.

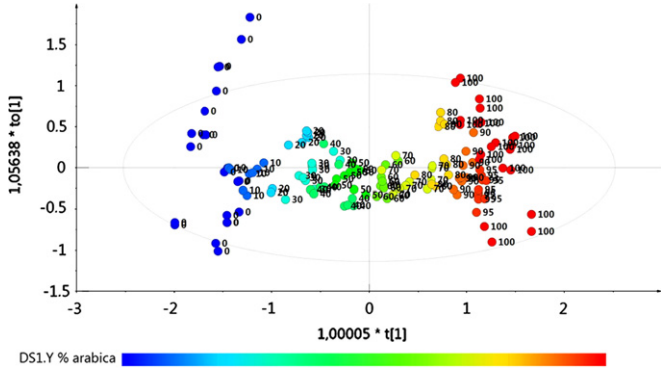

Fig. 2. OPLS score plot of water extracts of *C. arabica* and *C. canephora* var. robusta for training set samples.

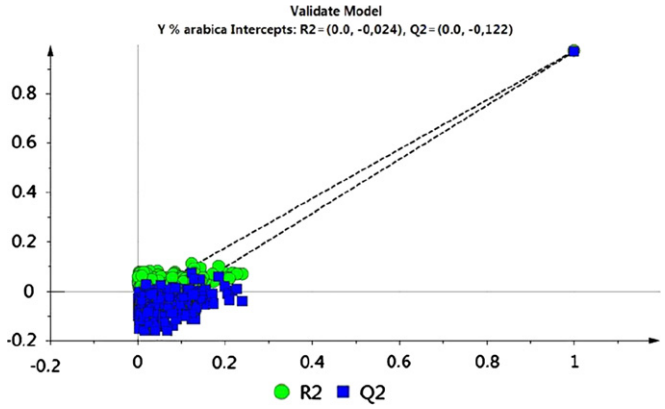

Fig. 3. Permutation test performed with 200 rounds of random permutations of the *Y* variable performed on the training set samples.

performed with 200 rounds of random permutation of the *Y* variable. The substantial decrease of both parameters Q^2 and R^2 (vertical axis interception point of the Q^2 and *R*2 regression line resulted both with negative values) enforced the statistical validity of the obtained model (Fig. 3). A plot of predicted versus reference percentage composition in arabica was performed on the training set blends samples: the R2 (linear regression coefficient) value was 0.972, indicating a very good linear correlation (Fig. 4). On the basis of this model, test set samples constituted by arabica and robusta blends of known composition obtained either by commercial market or by laboratory preparation were investigated to check the predictive performance of the OPLS model. The real and OPLS predicted percentage composition in arabica for test set blends were reported in Table 1. The very good agreement between predicted and real percentage composition gained by the OPLS model is evident. Noteworthy, differences in roasting conditions and geographical origins of samples did not impaired the prediction capability of the OPLS model. The prediction versus reference blend percentage composition of arabica evaluated for test set samples depicted in Table 1, revealed *R*2 of 0.998, showing a very good linear correlation. This high linear correlation coefficient resulted comparable respect

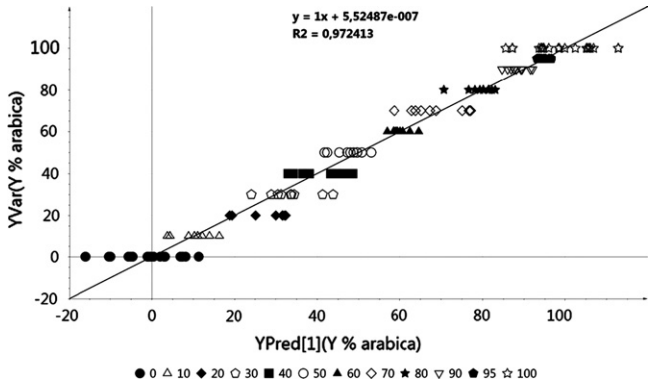

Fig. 4. Predicted (*Y*Pred) versus reference (*Y*Var) percentage composition in arabica obtained for training set samples.

Table 1
OPLS prediction of arabica percentage composition for test set roasted coffee blends samples. Medium OPLS predicted and real values are reported.

Sample	OPLS predicted % in arabica	Medium OPLS predicted % in arabica	Real % in arabica
1	99.83 101.71	100.77	100
2	100.78 101.03	100.90	100
3	95.14 98.14	96.64	95
4	89.94 90.79	90.37	90
5	77.78 82.15	79.97	80
6	67.62 66.80	67.21	70
7	57.24 56.85	57.04	60
8	50.87 53.87	52.37	55
9	46.79 46.70	46.75	50
10	42.36 45.32	43.84	45
11	33.02 24.74	28.88	30
12	11.69 10.23	10.96	10
13	5.47 6.73	6.10	5
14	−2.04 −2.23	−2.14	0
15	−4.81 −2.17	−3.49	0

to the values obtained by other authors monitoring the robusta variety content obtained by NIRS [24] and ESI FT-ICR MS [25]. In the present study, room temperature water extracts were performed differently to Garret et al. work [25], in which hot water extracts were considered; according to Pizarro et al. [24] we preferred the use of a regression approach endowed with the removal of structured noise and non significant information from spectra, obtaining highly accurate predictions.

4. Conclusions

In the present study, for the first time, ^{1}H–NMR and chemometrics has been applied on roasted and ground coffee blends, giving a practical answer to the commercial relevant problem of coffee authentication. The ^{1}H–NMR spectra of extracts of roasted coffee, obtained with water at room temperature, report a complete set of information of the water soluble chemical compounds; these data have been used to train an OPLS model with very high predictive capability. With the present approach, samples did not required any chemical derivatization or separation techniques, thus enabling direct measurement on sample as it is. Differently from methods based on single marker detection for coffee blend assessment, presenting the disadvantage of being more prone to frauds and to erroneous evaluations due to considerable sample preparation, our approach, by considering several chemical compounds simultaneously, prevents possible misled declaration or labeling about the composition. The OPLS predicted values of arabica content in roasted coffee blends reached a considerable high accuracy, in spite of different geographical origin and roasting conditions, making this approach suitable for authentication of unknown arabica and robusta coffee blends. This method has been patented [44].

Acknowledgments

“A. De Marco” Foundation is acknowledged for funding support.

References

[1] C. Sanz, L. Maetzu, M.J. Zapalena, J. Bello, C.J. Cid, Sci. Food Agric. 82 (2002) 840–847.
[2] T. Wermelinger, L. D'Ambrosio, B. Klopprodgge, C. Yeretzian, J. Agric. Food Chem. 59 (2011) 9074–9079.
[3] C.-L. Ky, J. Louarn, S. Dussert, B. Guyot, S. Hamon, M. Noirot, Food Chem. 75 (2001) 223–230.
[4] S. Casal, M.B.P.P. Oliveira, M.R. Alves, M.A. Ferreira, J. Agric. Food Chem. 48 (2000) 3420–3424.
[5] S. Casal, M.Rui Alves, E. Mendes, M.B.P.P. Oliveira, M.A. Ferreira, J. Agric. Food Chem. 51 (2003) 6495–6501.
[6] A. Farah, T. De Paulis, L.C. Trugo, P.R. Matrin, J. Agric. Food Chem. 53 (2005) 1505–1513.
[7] W.J. Rogers, S. Michaux, M. Bastin, P. Bucheli, Plant Sci. 149 (1999) 115–123.
[8] M.J. Martin, F. Pablos, A.G. Gonzalez, Food Chem. 66 (1999) 365–370.
[9] M.Rui Alves, S. Casal, M. Oliveira, M. Ferreira, J. Am. Oil Chem. Soc. 80 (2003) 511–517.
[10] M.J. Martin, F. Pablos, A.G. Gonzàlez, M.S. Valdenebro, M. Leòn-Camacho, Talanta 54 (2001) 291–297.
[11] C. Mariani, E. Fedeli, La Rivista Italiana delle Sostanze Grasse 63 (1991) 111–115.
[12] M. Valdenebro, M. Leòn-Camacho, F. Pablos, A.G. Gonzàlez, M.J. Martin, Analyst 124 (1999) 999–1002.
[13] A.B. Rubayiza, M. Meurens, J. Agric. Food Chem. 53 (2005) 4654–4659.
[14] D. Pacetti, E. Boselli, M. Balzano, N.G. Frega, Food Chem. 135 (2012) 1569–1574.
[15] A.G. Gonzàlez, F. Pablos, M.J. Martin, M. Leòn-Camacho, M. Valdenebro, Food Chem. 73 (2001) 93–101.
[16] R.C. Alves, S. Casal, M.R. Alves, M.B. Oliveira, Food Chem. 114 (2009) 295–299.
[17] N. Frega, F. Bocci, G. Lerker, J. High Res. Chromatogr. 17 (1994) 303–307.
[18] R. Consonni, L.R. Cagliani, in: S.L. Taylor (Ed.), Advances in Food and Nutrition Research, Elsevier Inc., University of Nebraska, Lincoln, USA, 2010, pp. 87–165.
[19] O. Fiehn, Comp. Funct. Genomics 2 (2001) 155–168.
[20] E.K. Kemsley, S. Ruault, R.H. Wilson, Food Chem. 54 (1995) 321–326.
[21] I. Esteban-Díez, J.M. González-Sáiz, C. Pizarro, Anal. Chim. Acta 514 (2004) 57–67.
[22] G. Downey, R. Briandet, R.H. Wilson, E.K. Kemsley, J. Agric. Food Chem. 45 (1997) 4357–4361.
[23] A.M.C. Hovell, E.J. Pereira, N.P. Arruda, C.M. Rezende, Anal. Chim. Acta 678 (2010) 160–168.
[24] C. Pizarro, I. Esteban-Díez, J.M. González-Sáiz, Anal. Chim. Acta 585 (2007) 266–276.
[25] R. Garret, B.G. Vaz, A.M.C. Hovell, M.N. Eberlin, C.M. Rezende, J. Agric. Food Chem. 60 (2012) 4253–4258.
[26] M. Koda, K. Furihata, F. Wei, T. Miyakawa, M. Tanokura, J. Agric. Food Chem. 60 (2012) 1158–1166.
[27] L. Mannina, F. Marini, M. Gobbino, A.P. Sobolev, D. Capitani, Talanta 80 (2010) 2141–2148.
[28] R. Consonni, L.R. Cagliani, J. Agric. Food Chem. 56 (2008) 6873–6880.
[29] M. Cuny, E. Vigneau, G. Le Gall, I. Colquhoun, M. Lees, D.N. Rutledge, Anal. Bioanal. Chem. 390 (2008) 419–427.
[30] R. Consonni, L.R. Cagliani, F. Benevelli, M. Spraul, E. Humpfer, M. Stocchero, Anal. Chim. Acta 611 (2008) 31–40.

[31] M. Bosco, R. Toffanin, D. De Palo, L. Zatti, A. Segre, J. Sci. Food Agric. 79 (1999) 869–878.
[32] L. Navarini, R. Gilli, V. Gombac, A. Abatangelo, M. Bosco, R. Toffanin, Carbohydr. Polym. 40 (1999) 71–81.
[33] O. Frank, S. Blumberg, G. Krümpel, T. Hofmann, J. Agric. Food Chem. 56 (2008) 9581–9585.
[34] S. Kreppenhofer, O. Frank, T. Hofmann, Food Chem. 126 (2011) 441–449.
[35] A. Ciampa, G. Renzi, A. Taglienti, P. Sequi, M. Valentini, J. Food Qual. 33 (2010) 199–211.
[36] F. Wei, K. Furiata, M. Koda, F. Hu, T. Miyagawa, M. Tanokura, J. Agric. Food Chem. 60 (2012) 1005–1012.
[37] N.C. Bicho, A.E. Leitão, J.C. Ramalho, F.C. Lidon, Eur. Food Res. Technol. 233 (2011) 303–311.
[38] A. Bax, J. Magn. Reson. 65 (1985) 142–145.
[39] J. Trygg, S. Wold, J. Chemom. 16 (2002) 119–128.
[40] O. Cloarec, E. Dumas, J. Trygg, A. Craig, R.H. Barton, J.C. Lindon, J.K. Nicholson, E. Holmes, Anal. Chem. 77 (2005) 517–526.
[41] R. Consonni, L.R. Cagliani, C. Cogliati, Talanta 88 (2012) 420–426.
[42] F. Wei, K. Furiata, F. Hu, T. Miyagawa, M. Tanokura, J. Agric. Food Chem. 59 (2011) 9065–9073.
[43] N. D'Amelio, L. Fontanive, F. Uggeri, F. Suggi-Liverani, L. Navarini, Food Biophys. 4 (2009) 321–330.
[44] R. Consonni, L.R. Cagliani, Patent no. MI2012A001281, Italy.