LWT - Food Science and Technology 49 (2012) 55–63

Contents lists available at SciVerse ScienceDirect

LWT - Food Science and Technology

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

Identification of components of Brazilian honey by ^{1}H NMR and classification of its botanical origin by chemometric methods

Elisangela F. Boffo [a,*], Leila A. Tavares [b], Antonio C.T. Tobias [c], Márcia M.C. Ferreira [d], Antonio G. 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 Ciências Matemáticas e Naturais, Universidade Federal do Espírito Santo, CEP 29932-540, São Mateus, ES, Brazil
[c] Curso de Agronomia e Medicina Veterinária, Centro Regional Universitário do Espírito Santo do Pinhal "UNIPINHAL", CEP 13990-000, Espírito Santo do Pinhal, SP, Brazil
[d] Instituto de Química, Universidade Estadual de Campinas, P.O. Box 6154, CEP 13084-971, Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 30 August 2011
Received in revised form
21 April 2012
Accepted 24 April 2012

Keywords:
Honey
^{1}H nuclear magnetic resonance
K-nearest neighbor
Soft independent modeling of class analogies
Partial least squares – discriminant analysis

ABSTRACT

The potential of NMR spectroscopy to differentiate honeys concerning to the nectar employed in its production was evaluated. The application of chemometric methods to ^{1}H NMR spectra has allowed to discriminate the honeys produced in the state of São Paulo, being identified the signals of responsible substances for the discrimination. Application of PCA and HCA methods to ^{1}H NMR data have resulted in the natural clustering of the samples. Wildflower honeys were characterized by higher concentration of phenylalanine and tyrosine. Citrus honeys showed higher amounts of sucrose than other compounds, while eucalyptus honeys had higher amount of lactic acid than the others. Assa-peixe honeys showed spectra similar to eucalyptus and citrus. Sugar-cane honeys showed some signals similar to eucalyptus and citrus honeys, but also showed the tyrosine and phenylalanine signals. Adulterated honeys showed 5-hydroxymethylfurfural, citric acid and ethanol signals. KNN, SIMCA and PLS-DA methods were used to build predictive models for honey classification. In the commercial honeys prediction KNN, SIMCA and PLS-DA models correctly classified 66.7; 22.2 and 72.2% of the samples, respectively.

1. Introduction

Honey is the natural product obtained by honeybees from the nectar of flowers or from secretions of living parts of plants or excretions of plant sucking insects, which the bees collect and transform by combining with specific substances of their own and store in the honeycomb to ripen and mature (Brasil, Instrução Normativa n° 11, 2000). The composition of honey consists of varying proportions of sugars, water, amino acids, oil, mineral salts and especial enzymes produced by bees (Enrich, Boeykens, Caracciolo, Custo, & Vázquez, 2007).

For the general quality control of honey according to the current standards of the Codex Alimentarius (Codex Standard for Honey, 2002) and the European Union (EU-Council Council Directive, 2002), several physical and chemical measurements have to be determined based on their composition.

Sugars are the main constituents of honey, comprising about 95% of honey dry weight. The relative amount of the two monosaccharides, fructose (F) and glucose (G), as well as, the fructose–glucose and glucose–water ratios are useful for the classification of unifloral honeys. For example, the G + F minimum value for blossom honeys should be 60 g/100 g, while for honeydew honeys it is 45 g/100 g (EU-Council Council Directive, 2002).

The honeys' color depends on the how old the honey is and the kind of flower that supplies the nectar. The determination of color is a useful classification criterion for unifloral honeys. For example, alfafa produces a white honey, heather a reddish-brown, acacia and citrus, a straw color. Honey color is related with its flavor. Light colored honey is mild whereas darker types have stronger flavors. Light honeys generally fetch the highest prices. Nevertheless, in Germany, Austria and Switzerland, dark honeys are especially appreciated. Dark colored honeys are reported to contain more phenolic acid derivatives but less flavonoids than light colored ones (Bogdanov, Ruoff, & Oddo, 2004). The most commonly used methods are based on optical comparison, using simple color grading after Pfund or Lovibond (Fell, 1978).

Hydroxymethylfurfural (HMF) is an important indicator for evaluation of storage time and heat damage. It is a sugar break-down product and increases with temperature and storage time while fresh honeys contain only traces of HMF (Zappalà, Fallico, Arena, & Verzera, 2005). Diastase activity in honey is also affected

* Corresponding author. Instituto de Química, Universidade Federal da Bahia, CEP 40170-115, Salvador, BA, Brazil. Tel.: +55 71 3283 6812; fax: +55 71 3137 4117.
E-mail addresses: eboffo@ufba.br, efboffo@yahoo.com.br (E.F. Boffo).

0023-6438
doi:10.1016/j.lwt.2012.04.024

by storage time and temperature. The diastase enzyme facilitates conversion of starch to maltose and is added by bees during honey production. However, its natural levels are variable in honeys depending on floral source. A lower value in diastase activity is a useful quality indicator (Bogdanov et al., 2004).

The electrical conductivity parameter was included recently in the new international standards for honey by Codex Alimentarius in 2001 and European Commission in 2002 (Bogdanov et al., 2004). It was introduced for differentiation between honeydew and blossom honey. The electrical conductivity of mixed blossom-honeydew honeys lies between 0.5 and 0.8 mS/cm. While the values of pure blossom honeys are below 0.5 mS/cm with many exceptions (Bogdanov & Gfeller, 2006). Etzold and Lichtenberg-Kraag (2008) showed be possible to distinguish between honeydew and blossom honey mixed with honeydew combining electrical conductivity data and FTIR.

All honeys are acidic due to the presence of organic acids that contribute to honey flavor and stability against microbial spoilage. Generally, the pH-value lying between 3.5 and 5.5. According to Sanz, Gonzalez, Lorenzo, Sanz, and Martínez-Castro (2005) and Krauze and Zelewski (1991) free acidity, total acidity and pH have presented some classification power for the discrimination between unifloral honeys.

Honey is 100% natural and nothing should be extracted or added to it. In some cases it is contaminated by the addition of sugar and the search for competitively priced products sometimes drives certain importers to acquire falsified honey. Moreover, some type of honeys can demand a higher price than other ones, and in order to prevent fraudulent labeling, a means of differentiating between honeys from different kinds must be developed (Devillers, Morlot, Pharm-Delegue, & Doré, 2004).

Nowadays, most of the analytical techniques intensively used involve some kind of sample pre-treatment. Moreover, the choice of methods and protocols often depends on the type of compound under investigation, making the overall characterization process laborious, time consuming and not completely reproducible. The advantages of the NMR technique with respect to other analytical methods are the non-invasive approach, the relatively easy and quick data acquisition (Caligiani, Acquotti, Palla, & Bocchi, 2007) and the possibility to provide information on a wide range of metabolites in a single experiment (Lolli, Bertelli, Plessi, Sabatini, & Restani, 2008). Finally, the sample preparation is almost negligible.

NMR is a powerful technique used to obtain structural information (Blau et al., 2008; Valente et al., 2008), and therefore it can help to understand the structure of components in complex systems such as food (Cazor, Deborde, Moing, Rolin, & This, 2006). The ^{1}H NMR spectroscopy can also be considered a fingerprinting technique (Bertram et al., 2005).

The richness of information, however, makes the spectra too complex to be analyzed or compared by eye. Multivariate analysis is therefore applied directly to the spectral data to extract the useful information. Several papers have been demonstrating the high efficiency these methods coupled to spectroscopy to classify honey samples or to detect some adulteration.

Combination of NMR data and chemometric analysis can also give interesting results for authentication purposes related to food in general, such as it has already been demonstrated in other similar works (Beretta, Caneva, Regazzoni, Bakhtyari, & Facino, 2008; Boffo, Ferreira, & Ferreira, 2009; Boffo, Tavares, Ferreira, & Ferreira, 2009; Consonni & Cagliani, 2008; Prestes et al., 2007; Schievano, Peggion, & Mammi, 2010). Chemometrics and FTIR spectroscopy (Kelly, Downey, & Fouratier, 2004; Sivakesava & Irudayaraj, 2001) and HPLC (Cotte et al., 2004) also have been successfully applied to the honey study.

In this study we present the investigation of a combined NMR and chemometric data analysis approach to describe the variability in the composition of honey samples and to identify the chemical compounds responsible for the discrimination among sample clusters. A database consisting of spectra from authentic samples describing the regular range of product variation was built. The classification methods, KNN (*K*-Nearest Neighbor), SIMCA (Soft Independent Modeling of Class Analogies) and PLS-DA (Partial Least Squares – Discriminant Analysis) were used to classify the commercial honeys of the state of São Paulo into three categories: wildflower, eucalyptus and citrus honeys. These methods were compared with objective to determinate the classification model that shows better prediction ability.

2. Materials and methods

2.1. Honey samples

Forty-six honey samples obtained from flowers of different plants, such as: citrus (*Citrus* sp.) – 13 samples, eucalyptus (*Eucalyptus* sp.) – 14 samples, assa-peixe (*Vernonia* sp.) – two samples, wildflower – 14 samples, and produced in the sugar-cane (*Saccharum* sp.) plantation [bee colonies placed near recently cut sugar-cane, and the bees collected the sap that oozed from the cut cane stems] – two samples, as well from bees fed with a sucrose solution (one sample) were studied. Some of these samples were provided by the beekeepers and the others were bought in markets in the state of São Paulo. All samples were collected in the years from 2004 to 2006.

All honeys collected were stored at room temperature (18–23 °C) from the time of acquisition to spectral analysis (max. six months). Given that the honey samples were stored in the dark in screw-cap jars at moderate temperatures, it is unlikely that any significant change would have occurred during storage. However, because this methodology would be applied to honey samples of indeterminable age, such variability may increase the robustness of the discriminating models developed.

2.2. NMR spectroscopy

The samples were prepared, in triplicate, dissolving 150 mg of honey in 450 μL of D_2O. Fifty microliter of a solution of TMSP (sodium-3-trimethylsilyl-2,2,3,3-d_4 propionate), 0.16 g/100 mL, prepared in D_2O was used as internal reference for chemical shift (δ 0.0). D_2O (99.9%) and TMSP (98%) were from Cambridge Isotope Laboratories, Inc. (USA).

All NMR experiments were recorded at room temperature using a Bruker DRX400 spectrometer operating at 9.4 T, equipped with 5-mm direct and inverse detection probes and observing ^{1}H at 400.2 MHz and ^{13}C at 100.6 MHz.

^{1}H NMR spectra (low power water signal suppression) were acquired using spectral width of 4664 Hz; 65,536 data points; pulse width of 8.5 μs; relaxation delay of 1.5 s; acquisition time of 7.0 s and 64 scans. Each ^{1}H NMR spectrum was acquired in 9 min and 7 s. Spectra were processed using 32,768 data points, by applying an exponential line broadening of 0.3 Hz for sensitivity enhancement before Fourier transform and were accurately phased and baseline adjusted. Phase correction was performed manually for each spectrum, and the baseline correction was applied over the entire spectral range, using a simple polynomial curve fit included in TopSpin® software.

^{13}C NMR spectra were acquired using spectral width of 27,027 Hz; 65,536 data points; pulse width of 6.0 μs; relaxation delay of 0.1 s; acquisition time of 1.4 s; and 32,768 scans. Each ^{13}C NMR spectrum was acquired in 12 h and 31 min. Spectra were processed using 65,536 data points and applying an exponential line broadening of 1.0 Hz.

Two dimensional NMR experiments were acquired using the standard spectrometer library pulse sequences.

$^{1}H-^{1}H$ *g*COSY and TOCSY (mixing time of 120 ms) experiments were obtained with spectral widths of 4664 Hz in f_1, 32 scans per t_1 increment and relaxation delay of 1.2 s *g*COSY experiment was acquired in 5 h and 10 min. TOCSY experiment was acquired in 5 h and 49 min. One-bond $^{1}H-^{13}C$ *g*HSQC experiment was acquired with an evolution delay of 1.7 ms for an average $^{1}J_{C,H}$ of 145 Hz. Spectral width of 22,140 Hz in f_1, 24 scans per t_1 increment and relaxation delay of 1.0 s were recorded. *g*HSQC experiment was acquired in 5 h and 4 min. The long-range $^{1}H-^{13}C$ *g*HMBC experiment was recorded setting the evolution delay of 62.5 ms for $^{LR}J_{C,H}$ for coupling constants of 8 Hz. Spectral width of 22,645 Hz in f_1, 64 scans per t_1 increment and relaxation delay of 1.0 s were used. *g*HMBC experiment was acquired in 17 h and 13 min. All spectra were acquired with spectral widths of 4664 Hz in f_2, $4k \times 256$ data matrices.

2.3. Chemometric analyses

Chemometrics is defined by the International Chemometrics Society as "the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods" (Hibbert, Minkkinen, Faber, & Wise, 2009).

Before the chemometric analyses, the ^{1}H NMR spectra were corrected by shifting to right or left as needed, using the TMSP signal as reference. The resulting spectra were converted into

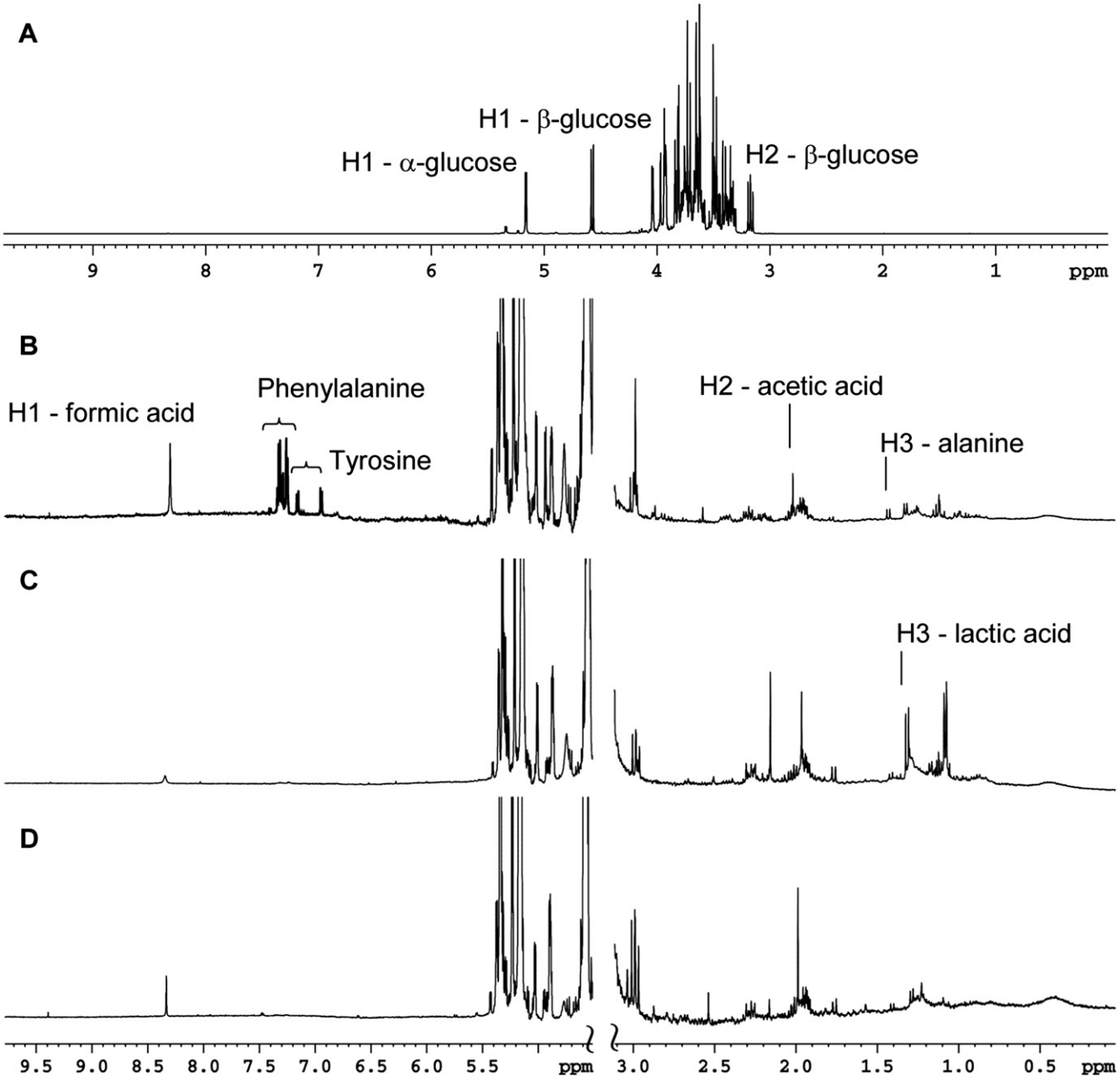

Fig. 1. (A) Complete ^{1}H NMR spectrum of citrus honey sample with water suppression. The principal resonances are from either glucose or fructose. (B–D) Expansions of δ 0.00–3.10 and of δ 4.50–9.70 regions of wildflower, eucalyptus and citrus honeys spectra, respectively. These regions have been magnified along the vertical scale to show the presence of minor constituents in honeys (D_2O).

JCAMP format to build the data matrix, using Origin® software (v. 5.0, Microcal, USA).

Pirouette® versions 3.11 and 4.0 (Infometrix Inc., Bothell, Washington, USA) were the software used for data analysis. The data matrix was built with 4644 variables (columns) and 138 spectra (lines – 46 samples in triplicate).

Each ^{1}H NMR spectrum was normalized using area normalization (the area under the sample profile is set equal to one) and first derivative was taken (to correct minor variations in the spectra baseline) prior to PCA and HCA analyses. The data were also autoscaled, i.e., each variable was mean-centered and scaled to unit variance. In HCA, the Euclidean distances among samples are calculated and transformed into similarity indices ranging from 0 to 1 by using the incremental linkage method.

PCA and HCA analysis were applied in two studies. One to verify the behavior and discrimination of all honey samples. In this study was included some honey types such as assa-peixe and those produced by feeding the bees with a sucrose solution (sugar-cane) and placing the beehive in the sugar-cane plantation. They are commercialized by few producers and, for this reason, only a small amount of these honey types was analyzed (five samples). Moreover, two samples considered adulterated (eucalyptus and citrus honeys) were analyzed, too. Another PCA and HCA analysis were made using only samples included in the classification study, shown below.

The KNN, SIMCA and PLS-DA training sets were built with citrus, eucalyptus and wildflower authentic honeys (21 samples prepared in triplicate, seven samples for each honey type, $X = (63 \times 4644)$). In the prediction of their class identities were used 18 commercial samples (7, 6 and 5 samples for wildflower, eucalyptus and citrus, respectively).

KNN, SIMCA and PLS-DA methods were used in order to attain classification rules for predicting the nectar source used for the honeys production. In KNN, the Euclidean distance was used as the criterion for calculating the distance between samples from the training set, and the optimum number of nearest neighbors (K) was selected by taking into account the success in classification with different K values. For all neighbors tested (1–10) none of the samples were misclassified, therefore $K = 1$ was selected, considering that there was only seven different samples in each class.

For SIMCA model, the number of principal components (PCs) used in each class model was determined using local scope and 95% confidence level, 4 PCs were selected for wildflower and eucalyptus categories and 5 PCs for citrus.

In PLS-DA model, the optimum number of PCs was chosen based on predicted residual sum of squares (PRESS), which should be minimized, along with the R^2 values from regression. The predictability of the model was tested by computing the standard error of calibration (SEC) and standard error of validation (SEV). Step-validation (leave-three-out procedure) was used to estimate the performance of the model developed. For PLS-DA model, 4 PCs were selected for wildflower category and 3 PCs for eucalyptus and citrus.

Finally, commercial samples were evaluated with regard to the nectar employed in their production.

3. Results and discussion

3.1. ^{1}H NMR spectra analysis

^{1}H NMR provides a simple method to obtain global information about complex samples in a single experiment maintaining the natural ratio of the substances. Fig. 1A represents a typical ^{1}H NMR spectrum of citrus honey in water solution. Different spectral regions are characterized by specific compound resonances, such as the aliphatic (δ 0.00–3.00), carbohydrates (δ 3.00–6.00) and aromatic (δ 6.00–10.00) regions.

In the carbohydrates region, dominant resonances of main monosaccharides (α- and β-glucopyranose, β-fructopyranose, α- and β-fructofuranose) were observed and specific signals of glucopyranose, δ 5.22 and 4.63 (α and β anomeric hydrogen, respectively) and δ 3.23 (H_2 of β-glucopyranose) were recognized. Those signals are practically equal to all honey analyzed and only small variations in the intensity were observed. The assignment of the major signals originated from those major constituents of the honeys is summarized in Table 1 (obtained from 2D NMR experiments, *g*COSY, TOCSY, *g*HSQC and *g*HMBC, and ^{13}C NMR spectrum).

Among all resonances of minor components, some compounds can be readily identified and resumed in Fig. 1B–D (region expansion of δ 0.00–3.10 and 4.50–9.70 of ^{1}H NMR spectra of (B) wildflower, (C) eucalyptus and (D) citrus honeys). The three honeys showed the signals of formic acid (singlet – δ 8.45), acetic acid (singlet – δ 2.00) and alanine (doublet – δ 1.46; J = 7.30 Hz). However, the wildflower honey presented in the region of δ 6.80–7.50 the aromatic signals of phenylalanine and tyrosine. The eucalyptus honeys showed a higher quantity of lactic acid (doublet

Table 1
^{1}H and ^{13}C NMR data for carbohydrates presents in the Brazilian honey.

Compound	δ^{13}C	δ^1H (mult., J in Hz)
α-Glucopyranose (α-glu)		
C^1H	94.6	5.22 (*d*, 3.70)
C^2H	74.0	3.54–3.49 (*m*)
C^3H	75.4	3.73–3.66 (*m*)
C^4H	72.1 or 72.2	3.42–3.35 (*m*)
C^5H	73.9	3.84–3.78 (*m*)
C^6H	63.3	3.73–3.67 (*m*)
$C^{6'}H$	63.3	3.85–3.79 (*m*)
β-Glucopyranose (β-Glu)		
C^1H	98.4	4.63 (*d*, 8.00)
C^2H	76.7	3.23 (*dd*, 8.00; 9.20)
C^3H	78.3	3.50–3.40 (*m*)
C^4H	72.1 or 72.2	3.46–3.35 (*m*)
C^5H	78.4	3.50–3.40 (*m*)
C^6H	63.4	3.77–3.72 (*m*)
$C^{6'}H$	63.4	3.90–3.85 (*m*)
β-Fructopyranose (β-FP)		
C^1H	66.5	3.57–3.52 (*m*)
$C^{1'}H$	66.5	3.72–3.66 (*m*)
C^2H	100.6	–
C^3H	70.2	3.79–3.74 (*m*)
C^4H	72.3	3.90–3.85 (*m*)
C^5H	71.8	3.99–3.97 (*m*)
C^6H	65.9	3.72–3.65 (*m*)
$C^{6'}H$	65.9	4.03–3.97 (*m*)
β-Fructofuranose (β-FF)		
C^1H	65.4	3.58–3.53 (*m*)
C^2H	104.1	–
C^3H	78.0	4.10–4.07 (*m*)
C^4H	77.0	4.10–4.07 (*m*)
C^5H	83.2	3.85–3.77 (*m*)
C^6H	65.0	3.82–3.75 (*m*)
$C^{6'}H$	65.0	3.68–3.62 (*m*)
α-Fructofuranose (α-FF)		
C^1H	65.6	3.65–3.62 (*m*)
C^2H	107.0	–
C^3H	84.5	4.10–4.07 (*m*)
C^4H	78.6	4.00–3.95 (*m*)
C^5H	83.8	4.07–4.02 (*m*)
C^6H	63.7	3.68–3.64 (*m*)
$C^{6'}H$	63.7	3.80–3.77 (*m*)

Abbreviations: *d* – doublet, *dd* – doublet of doublet, *m* – multiplet.

– δ 1.35; $J = 6.90$ Hz). The assignment of these signals and the other compounds in the honey are summarized in Table 2.

3.2. Honey discrimination with PCA and HCA methods

Usually, unsupervised methods such as Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) constitute the first step in data analysis. Without assuming any previous knowledge of sample class, these methods are enabling for the data visualization in a reduced dimensional space built on the dissimilarities between samples with respect to their biochemical composition. In this step, samples are identified in a pertinent space of reduced dimension. They were also used to select the optimal signal pre-treatment procedure.

Chemometric methods were applied directly to ^{1}H NMR spectra from honey samples. Two analysis were performed, one using all honey types (46 samples) and other including only wildflower, eucalyptus and citrus honeys (39 samples).

Table 2
^{1}H and ^{13}C NMR data for minor compounds presents in the Brazilian honey.

Position	δ^{13}C	δ^1H (mult., J in Hz)
Acetic acid		
CO_2H	179.8	–
CH_3	23.6	2.00 (*s*)
Alanine		
CO_2H	178.3	–
CH	53.1	3.75–3.85 (*m*)
CH_3	18.9	1.46 (*d*, 7.30)
Citric acid		
$C^{1,5}O_2H$	176.5	–
C^2H	46.1	2.79 (*d*, 15.50)
$C^{2'}H$	46.1	2.94 (*d*, 15.50)
C^3	76.1	–
C^4H	46.1	2.79 (*d*, 15.50)
$C^{4'}H$	46.1	2.94 (*d*, 15.50)
C^6O_2H	180.2	–
Ethanol		
CH_2	60.1	3.56–3.66 (*m*)
CH_3	19.6	1.15 (*t*, 7.10)
Formic acid		
HCO_2H	173.5	8.45 (*s*)
5-Hydroxymethylfurfural (HMF)		
C^2	154.3	–
C^3H	129.7	7.54 (*d*, 3.70)
C^4H	113.7	6.68 (*d*, 3.70)
C^5	164.1	–
CH_2	58.8	4.69 (*s*)
CHO	183.2	9.45 (*s*)
Lactic acid		
CO_2H	183.7	–
CH	70.6	4.30–4.40 (*m*)
CH_3	22.5	1.35 (*d*, 6.90)
Phenylalanine		
Aromatic ($C^{2,6}H$)	132.1	7.31 (*d*, 7.20)
Aromatic ($C^{3,5}H$)	131.8	7.41 (*d*, 7.20)
Aromatic (C^4H)	130.4	7.35–7.39 (*m*)
Tyrosine		
Aromatic (C^1)	129.4	–
Aromatic ($C^{2,6}H$)	133.5	7.18 (*d*, 8.40)
Aromatic ($C^{3,5}H$)	118.6	6.88 (*d*, 8.40)
Aromatic (C^4OH)	157.5	–

Abbreviations: *s* – singlet, *d* – doublet, *t* – triplet, *m* – multiplet.

The first study showed that is possible to discriminate a complex data set. PCA score plot (Fig. 2) presents 45.5% of the variability original information. PC1 describes 30.3%, while PC2 describes 15.2% of the total variability. In this plot, it can be observed a good discrimination between adulterated samples (positive scores values of PC1 – a cluster well defined) and the others. The wildflower honeys were also well discriminated to negative scores values in PC1 and PC2. However, it was not possible to distinguish satisfactorily the assa-peixe honeys from eucalyptus and citrus.

Loading values of the variables associated to the first two principal components showed that all the variables analyzed have a significant effect on PC1 to discriminate between authentic and adulterated honeys. Adulterated honeys showed the presence of 5-hydroxymethylfurfural (HMF) (Fig. 3F and G), citric acid (Fig. 3D) signals and the absence of amino acids signals usually found in the honeys. Citric acid was probably intentionally added to act as antioxidant, since it was not observed in the ^{1}H NMR spectra for the citrus honeys. The HMF has been very used as marker in adulteration of the honeys by addition of sucrose. However, it can be made by the exposition of honeys to high temperatures and also for a long period of time, storage under inadequate conditions, pH changes and other causes (Fallico, Zappalà, Arena, & Verzera, 2004; Tosi, Ré, Lucero, & Bulacio, 2004).

Assa-peixe honeys showed spectral region of δ 1.00–3.10 from ^{1}H NMR spectra similar to the eucalyptus and citrus ones (as lactic and acetic acid signals), justifying the position in the PCA scores plot. On the other hand, sugar-cane honeys showed some signals similar to the eucalyptus and citrus honey, in the same spectral region, but also presented the signals of the aromatic hydrogen of tyrosine and phenylalanine, such as the wildflower honeys, explaining its grouping in values near zero in PC2.

In order to increase the discrimination between honeys of the different botanical origin and to obtain classification models with high performance another study by PCA and HCA was made. In this case, spectra of five authentic samples of each honey type (wildflower, eucalyptus and citrus) were analyzed (as shown in Fig. 3A). The best discrimination was gotten when carbohydrates signals and non-informative ranges of the spectra were excluded; as shown in Fig. 3B.

In Fig. 4, PCA results related to the data matrix obtained from ^{1}H NMR spectra of honeys after the variable selection were reported. The first principal component (PC1) shows 24.0% of total variance while the second component (PC2) shows 17.2%; the two PCs

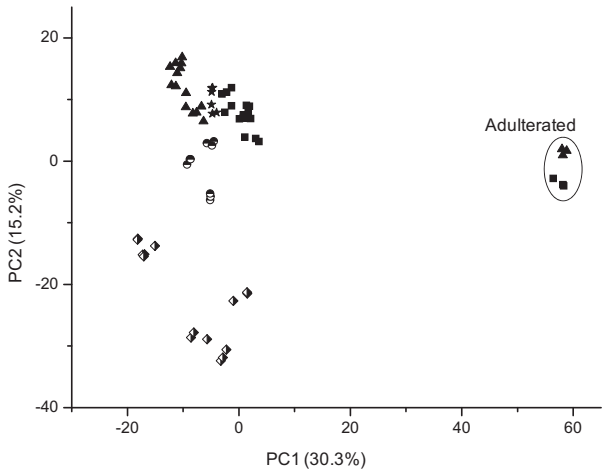

Fig. 2. PCA scores plot (PC1 × PC2) of all honey types analyzed by ^{1}H NMR (45.5% of the total variance): ◆ – wildflower; ▲ – eucalyptus; ■ – citrus; ◓ – sugar-cane; ★ – assa-peixe honeys.

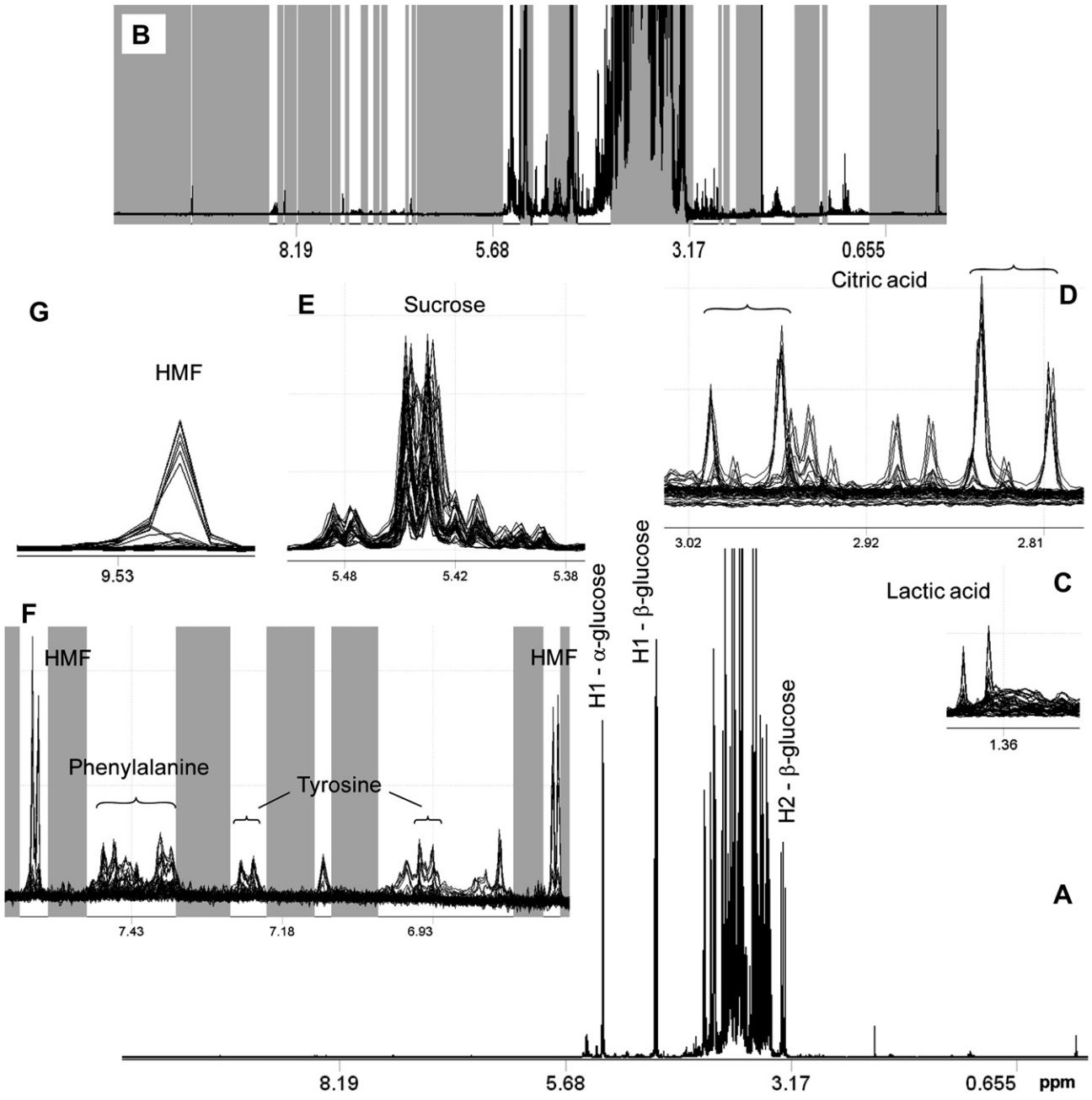

Fig. 3. (A) ^{1}H NMR spectra, with water suppression, of all honey analyzed; (B) selected regions used in statistical analyses (in white); (C–H) spectral expansions showing different quantities of honey constituents.

together show 41.2% of the original information. In this scores plot, very low sample variability between replicates is confirmed by observing the close proximity of the observations, thus supporting both the strong reproducibility of the NMR method and the sample homogeneity. The samples are grouped into three clearly distinct clusters according to the nectar used in their production: wildflower, eucalyptus and citrus. This discrimination was a direct consequence of the differences in their chemical composition.

The variables responsible for sample discrimination could be visualized on the loadings graphic and honey spectra. Samples located at negative scores of PC1 and PC2 (wildflower honeys) were richer in phenylalanine and tyrosine (Fig. 3F) than the others. The variable with high positive values on PC2 related to citrus honeys group showed higher amounts of sucrose (Fig. 3E) than the others. On the other hand, the variable with positive values on PC1 and negative values on PC2 related to eucalyptus honeys showed higher quantity of lactic acid than the others (Fig. 3C).

The dendrogram obtained from HCA is shown in Fig. 5. With a similarity index of 0.218 three main clusters were identified. This separation agreed well with the PCA results. Besides, at about 75% similarity, the replicates can be easily identified.

For subsequent classification analysis, only wildflower, eucalyptus and citrus honeys were evaluated.

3.3. Honey classification with KNN, SIMCA and PLS-DA methods

Using the KNN method, an unknown sample is classified according to the majority vote of its nearest neighbors in the multidimensional space. If there is a tie, the closer neighbors are given priority and proximity is measured using inter-sample distance.

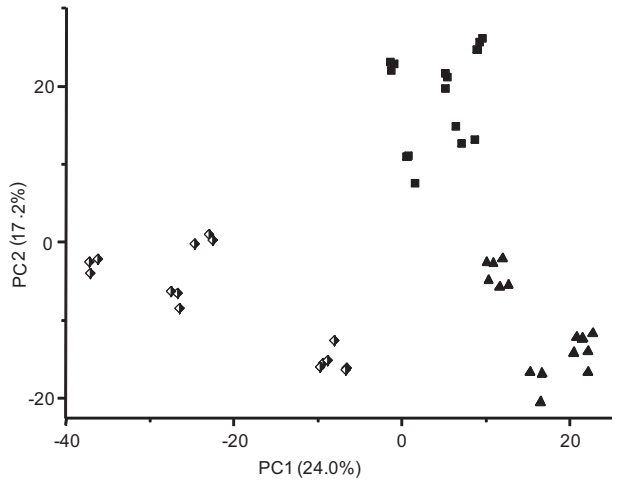

Fig. 4. PCA scores plot (PC1 × PC2) showing the discrimination between honey samples (41.2% of the total variance): ◆ – Wildflower; ▲ – eucalyptus and ■ – citrus honeys.

The method is self-validating because in the training set, each sample is compared with all the others in the set but not with itself. The best value of *K* can be chosen based on the results from the training set alone.

The SIMCA method builds a PCA model to each class and can be used to determine whether a new sample fits into a predetermined class, whether it does not fit in any of the classes or it indeed fits into more than one class.

The PLS-DA method is a variant of standard PLS regression in which the block of *Y*-variables consist of a set of binary indicator variables (one for each class) denoting class membership. For each binary class, a column of *Y* is generated by assigning a value of 0 or 1 to each sample, according to its class category. The set of predicted values by the model are rounded to either 0 or 1, and the true and predicted class memberships are then compared to evaluate how successful the model is at classifying the given samples.

Using these concepts, KNN, SIMCA and PLS-DA models were built with spectra of seven authentic samples of each honey type.

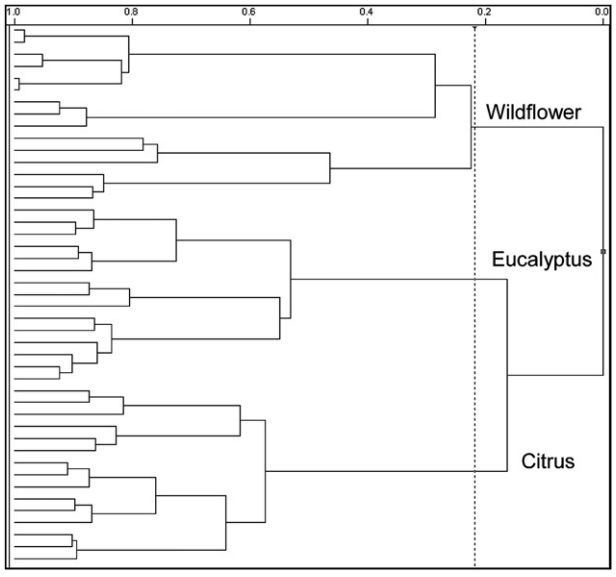

Fig. 5. HCA dendrogram obtained from ^{1}H NMR spectra from different honey types (similarity index: 0.218).

These samples were the same samples analyzed using PCA and HCA methods (Figs. 4 and 5).

Step-validation was used to select the optimal complexity of the SIMCA model, which resulted to be 4 principal components for wildflower and eucalyptus categories and 5 PCs for citrus. The variance explained was 82.1%, 69.3% and 68.3% for class 1 (wildflower), 2 (eucalyptus) and class 3 (citrus), respectively.

The PLS-DA loadings for the calibration models were similar to those observed in the PCA analysis. The R^2, SEC and SEV for the PLS-DA calibration models were 0.96, 0.04 and 0.13, respectively, for class 1. For class 2, R^2, SEC and SEV values were 0.92, 0.09 and 0.18, respectively. For class 3, R^2, SEC and SEV values were 0.92, 0.08 and 0.20, respectively. The calibration statistics indicated that the model developed could be acceptable to classify new samples.

Summary classification results following the application of KNN, SIMCA and PLS-DA to the prediction set of commercial samples are shown in Table 3. In the KNN classification one wildflower honey was misclassified as eucalyptus and four samples were misclassified in the citrus group. One eucalyptus honey sample was misclassified as citrus. In the SIMCA classification, five wildflower, five eucalyptus and three citrus honey samples do not belong to any of the predefined classes. In the PLS-DA classification, four wildflower and one eucalyptus honey do not belong to any of the predefined classes, and only one wildflower sample was misclassified as citrus.

Fig. 6 shows the predicted data *y* for the commercial samples and their classification as (A) wildflower, (B) eucalyptus and (C) citrus class. The data support the information in Table 3. For the honeys marketed as wildflower, two samples were correctly classified, one was misclassified as citrus and four as not belonging to any class. For samples marketed as eucalyptus, five were classified correctly and one as not belonging to any class. The honeys marketed as citrus were all classified correctly.

Those results show that in the commercial honeys prediction (18 samples) such as wildflower, eucalyptus and citrus honeys, KNN model correctly classified 28.6; 83.3 and 100% of the samples, respectively; SIMCA model correctly classified 28.6; 0 and 40%, respectively and PLS-DA model correctly classified 28.6; 100 and 100%, respectively.

This performance shows the PLS-DA approach to be superior to that reported for KNN and SIMCA methods. By applying PLS-DA, a model describing the maximum separation of predefined classes was obtained. Moreover, these results show the honeys from citrus group to be the most compact one.

The results of this study suggested that NMR spectroscopy coupled with multivariate methods hold the necessary information for a successful classification of honey samples of eucalyptus, citrus and wildflower types. When using PLS-DA classification model to predict honey samples, high classification rates were achieved. However, taking into account the relatively low number of samples used and the data set structure one needs to be cautious about the ability to extrapolate the classification model to predict new

Table 3
Predicted class obtained by KNN, SIMCA and PLS-DA models applied to the prediction honey data.

True class	Predicted class										
	KNN			PLS-DA				SIMCA			
	W	E	C	W	E	C	NM	W	E	C	NM
Wildflower (W)	2	1	4	2	0	1	4	2	0	0	5
Eucalyptus (E)	0	5	1	0	5	0	1	0	1	0	5
Citrus (C)	0	0	5	0	0	5	0	0	0	2	3

NM – no match.

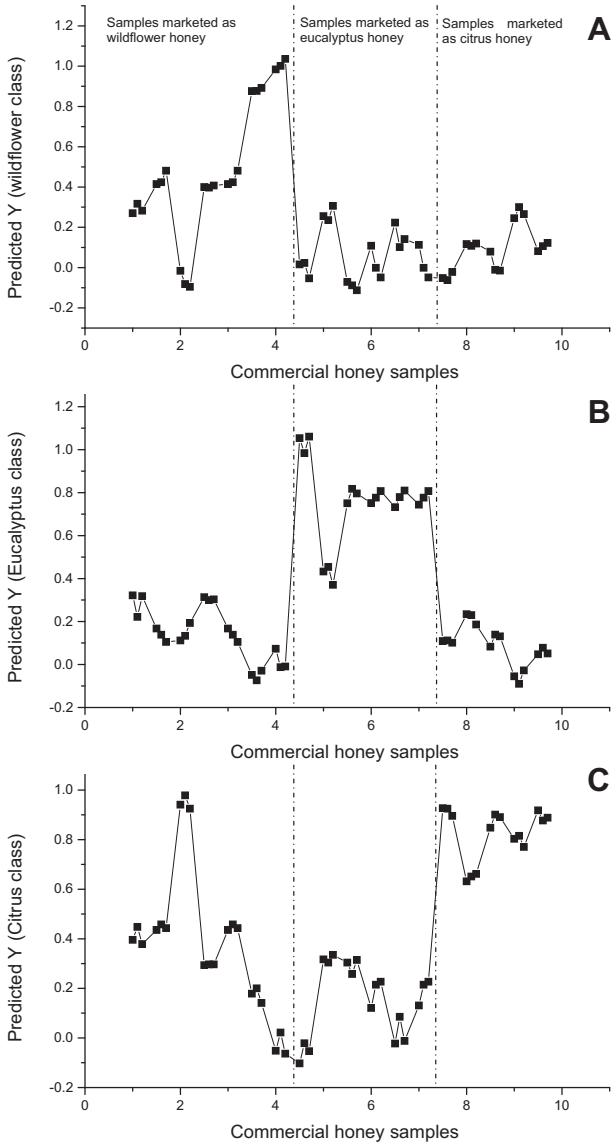

Fig. 6. Samples predicted plot using PLS-DA method to (A) wildflower, (B) eucalyptus and (C) citrus class.

samples in routine analysis. Therefore, it will be necessary to incorporate more samples to develop a more robust method to be commercially used by the industry as an application.

4. Conclusions

The application of chemometric methods to ^{1}H NMR spectra allowed to discriminate the eucalyptus, citrus and wildflower honeys produced in the state of São Paulo, being identified the signals of responsible substances for the discrimination.

Moreover, the chemometric methods for pattern recognition had shown that it is possible to classify the commercial honey samples according to the nectar they are generated from. KNN, SIMCA and PLS-DA pattern recognition models had correctly classified all samples through validation set. However, the PLS-DA method demonstrated the high efficiency in NMR data analysis with the aim of classification capability.

The PCA analysis also allowed discriminating the honeys that showed some kind of adulteration and identifying the type of compounds involved.

^{1}H NMR spectroscopy is a valid tool for food characterization and the combination with chemometric techniques largely improves the capability of sample classification. The simple sample preparation and the high quality of results obtained represent a valid alternative to other complex and time-consuming analyses.

Acknowledgments

The authors are grateful to CAPES, CNPq, FAPESP and FINEP for financial support.

References

Beretta, G., Caneva, E., Regazzoni, L., Bakhtyari, N. G., & Facino, R. M. (2008). A solid-phase extraction procedure coupled to ^{1}H NMR, with chemometric analysis, to seek reliable markers of the botanical origin of honey. *Analytica Chimica Acta, 620*, 176–182.

Bertram, H. C., Kristensen, N. B., Malmendal, A., Nielsen, N. C., Bro, R., Andersen, H. J., et al. (2005). A metabolomic investigation of splanchnic metabolism using ^{1}H NMR spectroscopy of bovine blood plasma. *Analytica Chimica Acta, 536*, 1–6.

Blau, L., Menegon, R. F., Ferreira, E. I., Ferreira, A. G., Boffo, E. F., Tavares, L. A., et al. (2008). Synthesis and total ^{1}H- and ^{13}C-NMR assignment of cephem derivatives for use in ADEPT approaches. *Molecules, 13*, 841–854.

Boffo, E. F., Ferreira, M. M. C., & Ferreira, A. G. (2009). SNIF-NMR and chemometric methods applied to ^{1}H NMR in the study of Brazilian brandy authenticity. In M. Guojonsdottir, P. Belton, & G. Webb (Eds.), *Magnetic resonance in food science: Challenges in a changing world* (pp. 143–150). Cambridge: Royal Society of Chemistry.

Boffo, E. F., Tavares, L. A., Ferreira, M. M. C., & Ferreira, A. G. (2009). Classification of Brazilian vinegars according to their ^{1}H NMR spectra by pattern recognition analysis. *LWT – Food Science and Technology, 42*, 1455–1460.

Bogdanov, S., & Gfeller, M. (2006). Classification of honeydew and blossom honeys by discriminant analysis. *ALP Science, 500*, 1–9.

Bogdanov, S., Ruoff, K., & Oddo, L. P. (2004). Physico-chemical methods for the characterisation of unifloral honeys: a review. *Apidologie, 35*, S4–S17.

Brasil, Ministério da Agricultura, Pecuária e Abastecimento, Instrução Normativa n° 11. (2000). *Regulamento Técnico de Identidade e Qualidade do Mel*. Diário Oficial da União. (23 de outubro).

Caligiani, A., Acquotti, D., Palla, G., & Bocchi, V. (2007). Identification and quantification of the main organic components of vinegars by high resolution ^{1}H NMR spectroscopy. *Analytica Chimica Acta, 585*, 110–119.

Cazor, A., Deborde, C., Moing, A., Rolin, D., & This, H. (2006). Sucrose, glucose, and fructose extraction in aqueous carrot root extracts prepared at different temperatures by means of direct NMR measurements. *Journal of Agricultural and Food Chemistry, 54*, 4681–4686.

CODEX STAN 12 Revised. (2002). Codex standard for honey. *Codex Alimentarius Commission, 30*, 748–753.

Consonni, R., & Cagliani, L. R. (2008). Geographical characterization of polyfloral and acacia honeys by nuclear magnetic resonance and chemometrics. *Journal of Agricultural and Food Chemistry, 56*, 6873–6880.

Cotte, J. F., Casabianca, H., Giroud, B., Albert, M., Lheritier, J., & Grenier-Loustalot, M. F. (2004). Characterization of honey amino acid profiles using high-pressure liquid chromatography to control authenticity. *Analytical and Bioanalytical Chemistry, 378*, 1342–1350.

Devillers, J., Morlot, M., Pharm-Delegue, M. H., & Doré, J. C. (2004). Classification of monofloral honeys based on their quality control data. *Food Chemistry, 86*, 305–312.

Enrich, C., Boeykens, S., Caracciolo, N., Custo, G., & Vázquez, C. (2007). Honey characterization by total reflection X-ray fluorescence: evaluation of environmental quality and risk for the human health. *X-Ray Spectrometry, 36*, 215–220.

Etzold, E., & Lichtenberg-Kraag, B. (2008). Determination of botanical origin of honey by Fourier-transformed infrared spectroscopy: an approach for routine analysis. *European Food Research and Technology, 227*, 579–586.

EU-Council Council Directive. (2002). 2001/110/EC of 20 December 2001 relating to honey. *Official Journal of the European Communities, L10*, 47–52.

Fallico, B., Zappalà, M., Arena, A., & Verzera, A. (2004). Effects of conditioning on HMF content in unifloral honeys. *Food Chemistry, 85*, 305–313.

Fell, R. D. (1978). The color grading of honey. *American Bee Journal, 18*, 782–789.

Hibbert, D. B., Minkkinen, P., Faber, N. M., & Wise, B. M. (2009). IUPAC project: a glossary of concepts and terms in chemometrics. *Analytica Chimica Acta, 642*, 3–5.

Kelly, J. F. D., Downey, G., & Fouratier, V. (2004). Initial study of honey adulteration by sugar solutions using mid-infrared (MID) spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry, 52*, 33–39.

Krauze, A., & Zelewski, R. I. (1991). Classification of honeys by principal component analysis on the basis of chemical and physical parameters. *Zeitschrift für Lebensmitteluntersuchung und Forschung A, 192*, 19–23.

Lolli, M., Bertelli, D., Plessi, M., Sabatini, A. G., & Restani, C. (2008). Classification of Italian honeys by 2D HR-NMR. *Journal of Agricultural and Food Chemistry, 56*, 1298–1304.

Prestes, R. A., Colnago, L. A., Forato, L. A., Vizzotto, L., Novotny, E. H., & Carrilho, E. (2007). A rapid and automated low resolution NMR method to analyze oil quality in intact oilseeds. *Analytica Chimica Acta, 596*, 325–329.

Sanz, M. L., Gonzalez, M., Lorenzo, C., Sanz, J., & Martínez-Castro, I. (2005). A contribution to the differentiation between nectar honey and honeydew honey. *Food Chemistry, 91*, 313–317.

Schievano, E., Peggion, E., & Mammi, S. (2010). ^{1}H nuclear magnetic resonance spectra of chloroform extracts of honey for chemometric determination of its botanical origin. *Journal of Agricultural and Food Chemistry, 58*, 57–65.

Sivakesava, S., & Irudayaraj, J. (2001). Prediction of inverted cane sugar adulteration of honey by Fourier transform infrared spectroscopy. *Journal of Food Science, 66*, 972–978.

Tosi, E. A., Ré, E., Lucero, H., & Bulacio, L. (2004). Effect of honey high-temperature short-time heating on parameters related to quality, crystallisation phenomena and fungal inhibition. *LWT – Food Science and Technology, 37*, 669–678.

Valente, A. M. M. P., Boffo, E. F., Costa, F. G., Melo, F. M. P., Melo, I. S., & Ferreira, A. G. (2008). Complete NMR assignments of a new prodigiosin isolated from *Streptomyces violaceusniger violaceusniger*. *Annals of Magnetic Resonance, 7*, 1–11.

Zappalà, M., Fallico, B., Arena, A., & Verzera, A. (2005). Methods for the determination of HMF in honey: a comparison. *Food Control, 16*, 273–277.