doi: 10.1255/mrfs.7

Differentiation of the botanical origin of honeys by fast, non-targeted ^{1}H-NMR profiling and chemometric tools as alternative authenticity screening tool

N. Gerhardt[a], M. Birkenmeier[a], T. Kuballa[b], M. Ohmenhaeuser[c], S. Rohn[d] and P. Weller[a]*
[a]Institute for Instrumental Analytics, Mannheim University of Applied Sciences, Germany. Corresponding Author: p.weller@hs-mannheim.de
[b]Chemisches und Veterinaeruntersuchungsamt (CVUA) Karlsruhe, Germany. Thomas.Kuballa@CVUAKA.BWL.DE
[c]Chemisches und Veterinaeruntersuchungsamt (CVUA) Freiburg, Germany. Marc.Ohmenhaeuser@cvuafr.bwl.de
[d]Institute of Food Chemistry, University of Hamburg, Germany. rohn@chemie.uni-hamburg.de

The globalization of trade of foods with its overwhelming variety has led to an increased demand for authentic products by all parts of the food value chain. In particular high-priced products are commonly adulterated, mislabeled or completely substituted. Honey is a typical example for such foods, which show an increasing number of adulterations, mostly related to the declaration of the botanical origin. A non-targeted ^{1}H-NMR-based screening, combined with multivariate statistical analysis was applied as a fast and simple comprehensive approach to verify the botanical origin of honey samples. The NMR fingerprints of honey sample were processed by taylor-made chemometric tools, based on principal component analysis (PCA) and linear discriminant analysis (LDA) in custom MATLAB routines. The results obtained by PCA-LDA showed very good discrimination between the different honey types with 98.9 % correct overall classification rate of the samples. Hence, this NMR based screening approach could be an effective alternative to traditional, laborious methods.

Introduction

According to current European Union regulations (Reg. (EU) 1169/2011), a detailed labelling of honey products with complete information about their botanical and geographical origin is mandatory. These aspects are particularly relevant in terms of both product quality and authenticity. Honey is a complex matrix, consisting of sugars, proteins, amino acids, organic acids, which varies with botanical origin, climate, soil and even storage conditions. This complex combination of parameters defines the character of the product and as so, its authenticity. Yet, there are no chemical markers or set of markers for authenticity, that are accessible by conventional, target-based analyses. Hence, an analytical approach covering a multitude of parameters in parallel on the one hand, paired with strong discrimination power on the other hand is required here. Specifically high-priced types of honey are a potential target for adulteration. A proof of such is complex, which is why a number of analytical techniques have already been discussed as potential tools for the authenticity assessment of honey. These investigations typically involve the use of chromatographic methods, often in combination with mass spectrometry or UV detection,[1-3] IR-based spectroscopic techniques (FT-MIR/NIR),[4-6] or sensor array systems (electronic nose or E-nose).[7-8] Particularly chromatographic methods for the analysis of such complex products as honey are time-consuming and usually require laborious sample preparation, often with chemical derivatization. The conventional univariate techniques used in quality control of honey are commonly melissopalynological analysis, physical-chemical methods (e.g. 5-hydroxymethyl furfural (HMF), enzyme activity, moisture and mono- and disaccharides) and the sensory evaluation. In particular, the analysis of pollen used for the differentiation of the botanical origin requires specifically trained personnel, which are commonly available only in specialized laboratories. Thus, there is an urgent need for new analytical tools that allow both a rapid, reliable and ideally, a non-targeted authentication of the botanical and geographical origin of honey. Non-targeted approaches do not focus on individual compounds or signals, but take into account all information that is obtained from the analysis in a way that can be compared to a fingerprint analysis.

In this context, a non-targeted profiling of honey by nuclear magnetic resonance spectroscopy (NMR) is a very promising method, which typically requires low sample amounts as well as minimal sample pre-processing and has the potential of automation. The application of liquid-state high-resolution ^{1}H-NMR spectroscopy in food analysis, particularly for quality control of food has already been successfully employed, featuring the simultaneous detection of various components and high reproducibility.[9-10] Additionally, the combination of chemometrics with ^{1}H-NMR spectroscopy has shown to provide much more detailed and comprehensive results with a better understanding of the relatonship between variables and thus easier interpretation of data.[11] Recently, several studies concerning the use of NMR spectroscopy in combination with appropriate multivariate statistical analysis for the determination of the botanical origin of honey have been published.[12-17] All of these studies proved NMR-based screening to be a suitable tool for the rapid authenticity analysis of honey. However, no optimizations for an automated and standardized sample preparation procedure – particularly for routine analysis of a large number of samples – have been implemented so far in currently existing studies. This is in particular relevant for the transfer from a research state to routine application.

In the present study, a standardized, semi-automated sample preparation procedure of honey using ^{1}H-NMR profiling was

implemented for routine use in food control laboratories in order to ensure the required high reproducibility of ^{1}H-NMR spectra, in particular constant chemical shifts and to enable a time-saving process for sample weighing. This work also suggests a non-targeted fingerprinting approach of high-resolution ^{1}H-NMR spectra in combination with multivariate data analysis. Here, the global information from a ^{1}H-NMR spectrum is extracted by means of chemometric techniques such as principal component analysis (PCA) and linear discriminant analysis (LDA) in order to resolve complex chemical profiles and to detect sample-specific patterns. At this point, highly reproducible results are crucial for the subsequent chemometric analysis to avoid misinterpretation of the results, e.g. due to misalignments of homogeneous peaks or artifacts. With these prerequisites, a compensation of these types of error, usually performed by bucketing procedure, is no longer necessary, which means that less chemical information encoded in the original NMR spectra are lost, thus resulting in improved discrimination among samples.

Material and methods

Honey samples and chemicals

A total of 97 reference samples of monofloral honeys from various European countries were analyzed using ^{1}H-NMR. The samples were acacia honey (*Robinia pseudoacacia*), rapeseed honey (*Brassica napus*) and honeydew honey (forest flower honeys) from various countries. The samples were obtained by governmental food inspectors from Baden-Wuerttemberg, Germany, supermarkets or directly from bee keepers. The botanical origin of the honey samples was confirmed by microscopic pollen analysis. All samples were stored in the dark at room temperature (18 - 23 °C) in screw-cap jars before analysis. The NMR buffer was prepared by dissolving 7.5 g of NaH_2PO_4 and 1 g of H_3PO_4 (85 % w/v) in 10 mL of deionised water (Millipore, Bedford, MA, USA). The commercially available sodium salt of 3(trimethylsilyl)-propionate acid-d4 (TSP) was used as a chemical shift reference (0 ppm) (98 %, Sigma Aldrich, Germany). D_2O (99.9 %) was from Deutero (Deutero GmbH, Germany).

Standardized sample preparation procedure

In order to standardize and simplify the sample weighing procedure and to reduce the potential sources of error, an automated Mettler-Toledo weighing system with a liquid dosing unit was applied. A mixture of precisely 250 mg/g honey in demineralised water was prepared by automatically adding water to a defined amount of honey (400 - 800 mg). The samples were prepared in duplicate. Subsequently, the samples were centrifuged (30 min/4000 rpm at 10 °C) and 850 µL of the supernatant were mixed with 50 µL NMR buffer and 100 µL of an internal standard (0.1 % TSP in D_2O).

^{1}H-NMR analyses in general show a strong susceptibility to the pH value of the sample, which may induce non-linear chemical shifts of single analyte signals. The varying composition of honeys (e.g. amino acids, sugars) leads to differing parameters, such as ionic strength and pH and as such, to chemical shifts in the range of 0.02 - 0.05 ppm. While this would be irrelevant in terms of structure elucidation aspects, it would strongly inflict chemometric analysis by misalignments of homogeneous peaks, artifacts and subsequently, to misinterpretations. For these reasons, the pH value of all samples was adjustet to pH 3.10 using an automated microtitration unit (BtpH unit, Bruker), which ensures highly reproducible pH (±0.01 pH units or better). Spectra generated under adjusted pH conditions are more reproducible and more immediately comparable among each other, which renders complex spectra aligning procedures obsolete. For the measurement, 600 µL of the pH-adjusted solutions were transferred into 5 mm NMR tubes for NMR measurement.

^{1}H nuclear magnetic resonance measurement at 400 MHz

All NMR measurements were performed on a Bruker Avance 400 Ultrashield spectrometer (Bruker BioSpin, Rheinstetten, Germany) equipped with a 5 mm SEI probe with Z-gradient coils, using a Bruker Automatic Sample Changer (B-ACS 120). ^{1}H-NMR spectra were acquired at 300.0 K without sample tube spinning. 32 scans and 4 preceeding dummy scans of 65 k points were acquired with a spectral width of 20.5504 ppm, a receiver gain of 16 and an acquisition time of 3.985 s. Water suppression was achieved using the NOESY-presaturation pulse sequence (Bruker 1D noesygppr1d pulse sequence) by irradiation at the water frequency (1881.78 Hz) during the recycle and mixing time delays. The data were acquired automatically using the ICON-NMR software (Bruker BioSpin, Rheinstetten, Germany). Aquisition time for each ^{1}H-NMR spectrum was 5 min. All NMR spectra were phased, baseline-corrected and calibrated via the TSP signal at 0.0 ppm.

Spectral preprocessing

For chemometry, MATLAB R2016a (The MathWorks, Natick, Massachusetts, USA) with Statistical Toolbox was used. Before chemometric analysis, the ^{1}H-NMR spectra were corrected by alignment, using the TSP signal as reference. The resulting spectra were converted into JCAMP format to build the data matrix in Matlab R2016a. The data matrix was built with 47001 variables (columns) and 194 spectra (lines – 97 samples in duplicate). Subsequently, preprocessing steps including Savitzky-Golay smoothing (5 point window, second order polynomial), baseline correction and interpolation were performed. Before multivariate analysis, all data were mean centered.

Statistical analysis

PCA was used for visualization, dimensionality reduction of the data and as a tool for a differentiation between different honey types without any prior knowledge of sample class. According to Jolliffe (2002), PCA can be characterized as an unsupervised linear mixture model that attempts to explain the variance within a dataset by a smaller number of mutually decorrelated principal components (PCs). Because of its applicability in dimensionality reduction, data visualization, clustering and sample group discrimination, PCA is often used as a starting point for data analysis, especially in a hypothesis free, exploratory experimental setup.

LDA was performed as a subsequent, supervised chemometric method in order to classify honey samples according to their botanical origin. According to Otto (1999) and Naes et al. (2002), this mathematical procedure maximizes the variance between groups and minimizes the variance within each group, in such a way that samples belonging or not to a specific group can be detected more easily than by PCA. A high dimensionality of the data, e.g. when the number of variables exceeds the number of samples, as it is the case in NMR analyses, may cause a substantial deterioration in the classification performance of the calculated LDA model. The latter requires a reduction of the system dimensionality in order to avoid overfitting. For this reason, LDA was applied to the PCA scores based on the first four extracted PCs.

Results and discussion

^{1}H NMR spectra of honey

^{1}H-NMR-based screenings allow rapid and easy extraction of global information in a single analysis while conserving the relationship of substances present in a complex matrix. Figure 1 shows the typical ^{1}H-NMR spectra of the honey samples in a buffered solution from different botanical origins including acacia, rapeseed and honeydew honey. As can be seen in this figure, three regions could be identified, corresponding to amino acids region (0 - 3 ppm), the sugar region (3 - 6 ppm) and the phenolics region (6 - 10 ppm). These spectra show the dominant resonances of major compounds (glucose, fructose, maltose and sucrose) and in contrast, less intensive resonances of the minor components (e.g. amino acids, organic acids, etc.), which also play important roles in the differentiation of honey.

Assignment of ^{1}H-NMR signals was performed according to the existing literature.[4,9,11] As expected, the most intensive and dominant signals in ^{1}H-NMR spectra of honey stem from the sugar regions (δ 3.0 - 5.5 ppm), typically α- and β-glucose and fructose. Further saccharides including xylose, D(+)-galactose and disaccharides maltose and sucrose are also detected in this region. These signals are virtually identical in all honey samples analyzed with only very minor variations in intensity. The minor signals from organic carboxylic acids or amino acids (citric acid: δ 2.65 ppm, acetic acid: δ 2.08 ppm, proline: δ 2.30 - 3.37 ppm, alanine: δ 1.48 ppm, phenylalanine δ 7.3 - 7.41 ppm, tyrosine: δ 6.88 - 7.18 ppm), and a number of other compounds (e.g. formic acid, ethanol or hydroxymethylfurfural) show significant variations in the intensity among the honey samples with the different botanical origins. This indicates that the relevant discriminating information is hidden not only in the major sugar signals, but also in the minor component signals.

Discrimination of botanical origins by PCA-LDA

The basic approach for the discrimination of the botanical origins of honey is the chemometric evaluation of the full ^{1}H-NMR spectra by a combination of PCA, followed by a LDA. The calibration model employed for the ^{1}H-NMR dataset was generated by dimensionality reduction and extraction of the most relevant information by PCA. This allows to describe a complex data set, such as NMR spectra by a few numbers of PCs. However, PCA is a non-supervised technique and does not predefine classes. Therefore, in a subsequent step, a LDA was performed on the PCA scores in order to calculate discriminant functions for the classification of honey samples in the correct groups and to find the directions to maximise class separability whilst aiming to minimize dispersion within each class. In order to investigate for the functional groups or classes of compounds (e.g. organic acids, carbohydrates region and aromatic signals) that are most responsible for class separation among samples, different spectral regions (δ 0 - 3 ppm, δ 3 - 6 ppm and δ 6 - 10 ppm) were subjected to PCA-LDA calculations.

The quality of the calculated LDA model (applied to the PCA scores) was evaluated by analyzing the percentage of correctly classified samples. In addition, the predictive ability was evaluated by k-fold cross-validation (k = 10). Figure 2 shows the LDA score plot obtained by using the first four PCs as variables. In total, three discriminant functions were calculated for the LDA model. As can be seen, a good degree of discrimination among unifloral honey samples was achieved when the whole spectral range (δ 0-10 ppm) was used. The discriminant model allowed the correct classification of all samples into their respective botanical origin groups with a success rate of 98.9 %.

The first three PCs explain 89.9 % of the total variability. The corresponding loading plots of the first three PCs (Figure 2) suggest that the discrimination between various unifloral botanical species is strongly correlated with the variations in the saccharide content. Accordingly, the anomeric sugar region between 4.5 und 5.5 ppm, with all the anomeric and part of the side chain protons of saccarides is the part of the ^{1}H-NMR spectrum with the main information content. Yet, there are no specific saccharide markers that were characteristic of the other botanical species. This underlines

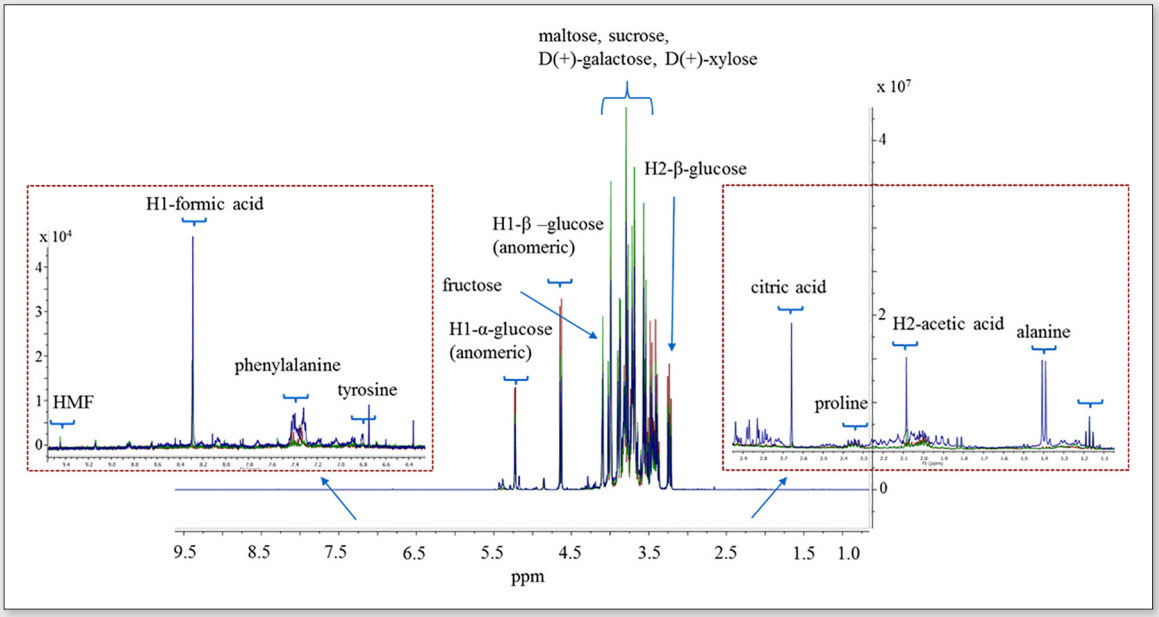

Figure 1. Characteristic 1H-NMR spectra of acacia, rapeseed and honeydew honeys with water suppression. Spectral expansion of δ 0-3 ppm (aliphatic) and δ 6-10 ppm (aromatic) regions show the presence of minor compounds in honeys.

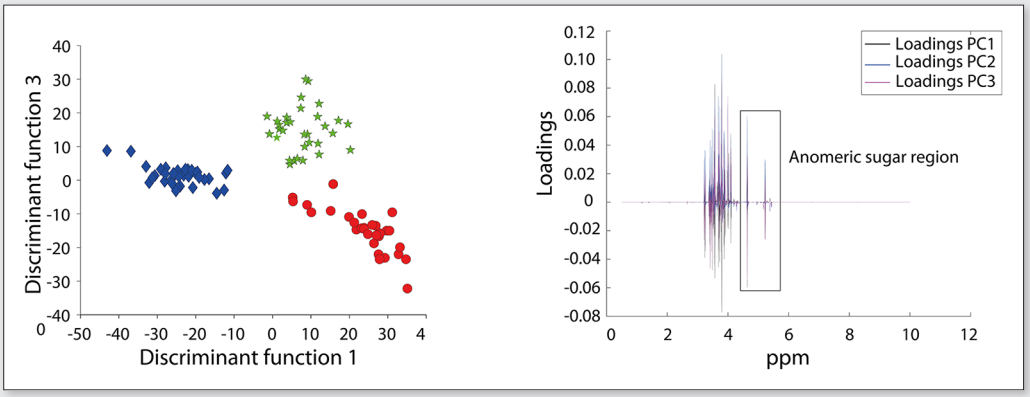

Figure 2. LDA score plot of unifloral honeys of different botanical origin obtained by using first four PCs as variables (spectral region: δ 0-10 ppm) with corresponding PCA loadings plot for the first three PCs describing 89.9 % of the total variance.

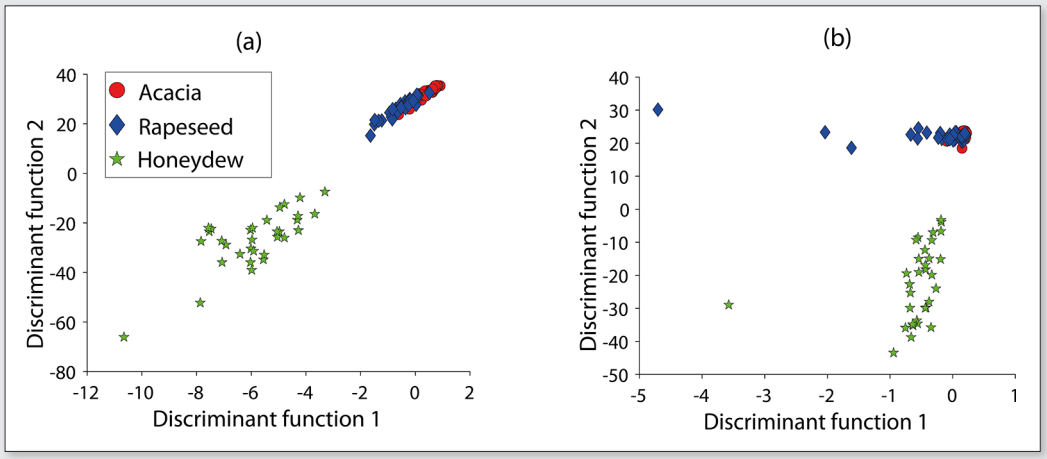

Figure 3. LDA score plot of unifloral honeys of different botanical origin obtained by using first four PCs as variables: (a) spectral region δ 6-10 ppm; (b) spectral region δ 0-3 ppm.

the importance of a multivariate approach. This assumption is confirmed by a PCA-LDA analysis, where the carbohydrates region was excluded and only the spectral regions of aliphatic (δ 0 - 3 ppm) and aromatic (δ 0 - 3 ppm) signals were considered in the classification model, respectively. As Figure 3 shows the discrimination power of minor components alone is not sufficient enough, compared to the whole spectra region (δ 0 - 10 ppm), in particular to distinguish between botanical species of higher similarity, such as acacia and rapeseed honey. These results suggest that the global NMR fingerprint approach, in which the whole information is extracted from the spectra, is suitable for a satisfactory discrimination and classification of unifloral honeys – especially due to complexity of natural substances – on the basis of their botanical origin.

Conclusion

In this study, ^{1}H-NMR-based screening and chemometric analysis was evaluated for its suitability in the differentiation of the floral origin of honeys on the example of acacia, rapeseed and honeydew. It could be demonstrated that a reproducible sample weight and a precise adjustment of the pH value of the diluted honey-water solution is crucial to avoid chemical shift variability across spectra that could inflict the chemometric analysis. The combination of PCA for data reduction, followed by a LDA proved to be a powerful chemometric tool to classify the honey samples accordingly to their botanical origin. The best discrimination quality was obtained when the whole spectra region was included in the classification model, whereby the carbohydrates signals showed the strongest contribution. As a summary, authenticity of the analysed monofloral honeys can be verified without the need for a microscopic pollen analysis. Thus, ^{1}H-NMR fingerprinting represents a suitable and comprehensive approach for the discrimination of complex samples and for the evaluation of authenticity. Moreover, due the possibility for automation and the low cost per analysis required for the screening, ^{1}H-NMR profiling has the potential to be used in routine work by food control laboratories.

Acknowledgement

The authors would like to acknowledge the Chemisches und Veterinaeruntersuchungsamt (CVUA) Karlsruhe and Freiburg for providing instrumental support and for supplying honey samples. We gratefully acknowledge Center for Applied Research in Biomedical Mass Spectrometry (ABIMAS) for the financial support.

References

1. A. Cavazza; C. Corradini; M. Musci; P. Salvadeo. High-performance liquid chromatographic phenolic compound fingerprint for authenticity

assessment of honey. *J Sci Food Agr* **93**(5), 1169–1175 (2013). DOI: 10.1002/jsfa.5869

2. I.K. Karabagias; A. Badeka; S. Kontakos; S. Karabournioti; M.G. Kontominas. Characterization and classification of Thymus capitatus (L.) honey according to geographical origin based on volatile compounds, physicochemical parameters and chemometrics. *Food Res Int* **55**, 363–372 (2014). DOI: 10.1016/j.foodres.2013.11.032
3. J. Zhou; Y. Lihu; Y. Li; L. W. Chen, Z. J. Liming. Floral classification of honey using liquid chromatography-diode array detection-tandem mass spectrometry and chemometric analysis. *Food Chem* **145**, 941–949 (2014). DOI: 10.1016/j.foodchem.2013.08.117
4. C. Herrero Latorre; R. M. Pena Crecente; S. Garcia Martin; J. Barciela Garcia. A fast chemometric procedure based on NIR data for authentication of honey with protected geographical indication. *Food Chem* **141**(4), 3559–3565 (2013). DOI: 10.1016/j.foodchem.2013.06.022
5. S. Gok; M. Severcan; E. Goormaghtigh; I. Kandemir; F. Severcan. Differentiation of Anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate analysis. *Food Chem* **170**, 234–240 (2015). DOI: 10.1016/j.foodchem.2014.08.040
6. T. Woodcock; G. Downey; C. Odonnell. Near infrared spectral fingerprinting for confirmation of claimed PDO provenance of honey. *Food Chem* **114** (2), 742–746 (2009). DOI: 10.1016/j.foodchem.2008.10.034
7. A. Zakaria; A. Y. Shakaff; M.J. Masnan; M.N. Ahmad; A. H. Adom; M. N. Jaafar; S. A. Ghani; A. H. Abdullah; A. H. Aziz; L.M. Kamarudin. A biomimetic sensor for the classification of honeys of different floral origin and the detection of adulteration. *Sensors (Basel, Switzerland)* **11**(8), 7799–7822 (2011). DOI: 10.3390/s110807799
8. T. Dymerski; J. Gebicki; W. Wardencki; J. Namiesnik. Application of an electronic nose instrument to fast classification of Polish honey types. *Sensors (Basel, Switzerland)* **14**(6), 10709–10724 (2014). DOI: 10.3390/s140610709
9. L. Mannina; A. Sobolev; S. Viel. Liquid state 1H high field NMR in food analysis. *Prog Nucl Mag Res Sp* **66**, 1–39 (2012). DOI: 10.1016/j.pnmrs.2012.02.001
10. R. Godelmann; F. Fang; E. Humpfer; B. Schutz, B.; M. Bansbach; H. Schafer; M. Spraul. Targeted and nontargeted wine analysis by (1)H NMR spectroscopy combined with multivariate statistical analysis. *J. Agric. Food Chem.* **61**(23), 5610–5619 (2013). DOI: 10.1021/jf400800d
11. Y. Monakhova; D. Rutledge; A. Roßmann; H.U. Waiblinger; M. Mahler; M. Ilse; T. Kuballa; D. W. Lachenmeier. Determination of rice type by 1 H NMR spectroscopy in combination with different chemometric tools. *J. Chemometrics* **28**(2), 83–92 (2014). DOI: 10.1002/cem.2576
12. E. F. Boffo; L.A. Tavares; M.C. Ferreira; A.G. Ferreira. Identification of components of Brazilian honey by 1H NMR and classification of its botanical origin by chemometric methods. *LWT-Food Sci Technol* **49**(1), 55–63 (2012). DOI: 10.1016/j.lwt.2012.04.024
13. E. Schievano; C. Finotello; J. Uddin; S. Mammi; L. Piana. Objective Definition of Monofloral and Polyfloral Honeys Based on NMR Metabolomic Profiling. *J. Agr. Food. Chem.* **64**(18), 3645–3652 (2016). DOI: 10.1021/acs.jafc.6b00619
14. M. Ohmenhaeuser; Y.B. Monakhova.; T. Kuballa; D.W. Lachenmeier. Qualitative and Quantitative Control of Honeys Using NMR Spectroscopy and Chemometrics. *Anal Chem* **2013**(4), 1–9 (2013). DOI: 10.1155/2013/825318
15. M. Spiteri; E. Jamin; F. Thomas; A. Rebours; M. Lees; K. M. Rogers; N. R. Douglas. Fast and global authenticity screening of honey using (1) H-NMR profiling. *Food Chem* **189**, 60–66 (2015). DOI: 10.1016/j.foodchem.2014.11.099
16. R. Consonni; L. R. Cagliani; C. Cogliati. Geographical discrimination of honeys by saccharides analysis. *Food Control* **32**(2), 543–548 (2013). DOI: 10.1016/j.foodcont.2013.01.038
17. X. Zheng; Y. Zhao; H. Wu; J. Dong; J. Feng. Origin Identification and Quantitative Analysis of Honeys by Nuclear Magnetic Resonance and Chemometric Techniques. *Food Anal. Methods* **9**(6), 1470–1479 (2016). DOI: 10.1007/s12161-015-0325-1