

EFFICIENCY OF DIFFERENT BIOINFORMATICS TOOLS IN METABOLITE PROFILING OF WHOLE COW'S MILK USING SYNTHETICALLY WATER-REMOVED ¹H NMR SPECTRA: A COMPARATIVE STUDY

LAREF NORA*

Address(es):

Center university of Relizane Ahmed Zabana, Algeria.

*Corresponding author: nora.laref@cu-relizane.dz

doi: 10.15414/jmbfs.2020.10.2.241-244

ARTICLE INFO

Received 25. 11. 2019

Revised 18. 6. 2020

Accepted 25. 6. 2020

Published 1. 10. 2020

Regular article



ABSTRACT

Large water peak obscures the signals corresponding to important metabolites, and thus hinders in obtaining complete information from samples. This study investigates the efficiency of three bioinformatics tools (Galaxy, Chenomx, and MetaboHunter) on the synthetically water-removed NMR spectra of whole cow's milk and compares the results obtained. Three samples of whole cow's milk were collected from Oran region (Algeria) and kept at -18°C until analysis. ¹H NMR spectra of the samples in DMSO d6 were recorded at ambient temperature without any pretreatment or purification, and the spectra were first processed with MestreNova to remove water and solvent artificially, followed by metabolite profiling using the above mentioned three bioinformatics tools. Detection of several metabolites, such as taurine, glycine, choline, threitol, niacinamide, and 1,3- dimethylurate, etc., was possible using the bioinformatics tools post water and solvent removal. In addition, dry content matter estimation revealed M3 as the richest milk among the three test samples. Although the bioinformatics tools identified many milk metabolites, there are differences in the detection efficiency probably because of the separate algorithm and different file format used by them.

Keywords: Whole cow's milk; Water removal; Bioinformatics; Metabolomics

INTRODUCTION

Milk, an exceptionally complex biological fluid of either human or bovine origin, contains mostly water, carbohydrates (mainly lactose), fat, proteins (casein micelles), minerals and vitamins. It comes in the form of an emulsion or colloid of fat globules in water containing the above mentioned metabolites. In dairy industry, milk is the most common target for adulteration, and the common milk adulterants include tap water, whey, synthetic milk, synthetic urine, urea and hydrogen peroxide. Considering the large scale consumption of milk and milk-based products, it is important to have appropriate quality checking procedures in place. On the other hand, since milk composition depends on several factors such as species, breed, genetics, somatic cell count, feed, season and lactation stage, identification of milk metabolites can provide information about those factors. The quality of milk often reflects the metabolic activity of the mammary gland. Moreover, early detection of metabolic disorder of a milk-producing animal is possible by screening metabolite-profile of its milk (Andreotti *et al.*, 2002; Belloque, 2008 ; Griffin and Roberts, 1985).

Metabolomics refers to identification and quantification of metabolites. Since milk originate from different cell types and metabolisms in the organism, the profiling of milk metabolites has often been carried out using high-throughput metabolomics methodologies such as high-resolution proton nuclear magnetic resonance (¹H NMR) spectroscopy, liquid chromatography mass spectrometry (LC-MS), and gas chromatography mass spectrometry (GC-MS). Despite less sensitivity, NMR spectroscopy has been emerged as the most preferred metabolomics method to analyze food and pharmaceutical products. NMR spectroscopy requires minimal sample preparation and detects all mobile hydrogen containing compounds. Since the method is non-destructive in nature, same sample can be used for multiple NMR experiments. Importantly, NMR spectroscopy can directly measure the molar concentration of a certain metabolite from the intensity of NMR resonances. Owing to these advantages, NMR spectroscopy has emerged as an attractive technique in the field of food science and pharmaceuticals. (Belloque, 1999 ; Eads and Bryant, 1986 ; Lamanna *et al.*, 2011).

Many reports are available in literature, particularly, on the analysis of milk metabolites using NMR spectroscopy since it allows easy sample handling and simultaneously detects a large number of compounds with a minimal amount of sample (Belloque, 1999 ; Lamanna *et al.*, 2011 ; Santos *et al.*, 2016). Apart from choosing relevant nuclei (particularly for milk analysis ¹H, ¹³C and ¹⁵P), one

can select a particular method among many different NMR methods, such as 1D, 2D-¹H-COSY, -HOHAHA or-DQ, 2D-¹H-NOESY, heteronuclear 2D and 3D, relaxation method, and time-domain NMR spectroscopy, etc., suitable to achieve different objective(s) in milk and dairy research. The application of NMR spectroscopy in milk and dairy research includes both qualitative and quantitative detection of milk compounds (Van Hekken and Dudley, 1997), isotopic analysis (Remaud *et al.*, 1997), *in vivo* reaction monitoring (Deborde *et al.*, 1998), physical state of dairy components (Breitschuh, 1997 ; Windhab, 1997), structural characterization of milk components (Molinari *et al.*, 1996), and conformational changes in milk protein (Belloque and Smith, 1998).

Despite many advantages of NMR as a tool to study metabolites, large water peak poses practical problem, particularly while using ¹H NMR for milk component analysis. Since milk constitutes of large amount of water, without water suppression/removal, it is difficult to detect other signals characteristics to important metabolites present in milk. To avoid this issue, ¹³C or ¹⁵P NMR spectroscopy has usually been used to analyze milk; however, these require pretreatment of milk samples. Many methods, such as extraction of triglyceride, removal of fat, pH adjustment or addition of MnCl₂, etc., have been applied to pretreat milk sample before recording ¹³C or ¹⁵P NMR spectrum (Eads and Bryant, 1986; Hu *et al.*, 2004). But pretreatment may result in structural changes in milk components. In this scenario, liquid-state ¹H NMR is a more non-destructive technique despite the issue of huge water peak obscuring signals characteristics to important milk metabolites. There are two ways one can tackle this problem — water suppression during recording spectra on NMR spectrometer (e.g., presaturation method) or artificial water removal while data analyzing using different NMR data processing tools such as MestreNova, nmrPipe, and TopSpin, etc. Water suppression (such as with presaturation method) during recording NMR spectra is certainly beneficial from the signal-to-noise point of view as suppression of large water peak increases receiver gain and hence sensitivity. However, choosing and employing an efficient water suppression technique often requires good knowledge in NMR spectroscopy as well as good skill in handling NMR spectrometer. Moreover, good water suppression depends on parameters that fluctuate from one sample to another. This hinders potential application of large scale data collection using auto-sampler. Solvent suppression at the processing step using software overcomes these difficulties at the cost of overall sensitivity (due to low receiver gain in spectrometer). Bioinformatics tools such as Metabohunter, Chenomx, and Galaxy, etc. rely both on efficient water suppression (by the software like

MestreNova) and good signal-to-noise ratio of the spectra. Choosing an appropriate bioinformatics tools can, therefore, be a daunting task since each tool uses special algorithm and file format. This study attempts to investigate constituents of milk by analyzing 1D liquid-state ¹H NMR spectra of three different whole cow's milk samples using three different bioinformatics tools (Metabohunter, Chenomx, and Galaxy) and to compare the data obtained by those three tools.

MATERIALS AND METHODS

Milk collection

Three samples of whole cow's milk were collected locally from cows raised in a farm situated in the Algerian City (Oran). That milk was put in hygienic bottle and then put in the freezer at -18 °C until analysis.

Sample Preparation

Three milk samples stored at -18 °C were first defrosted and brought to ambient temperature. Each milk sample (0.01 mL) was taken in a 5 mm standard NMR tube and DMSO-d6 with TMS was added. Simple hand-mixing was performed by inverting the NMR tube upside-down for three or four times.

NMR

1D ¹H NMR spectra of the whole cow's milk samples were recorded without any purification at ambient temperature on a liquid-state 400 MHz Bruker NMR spectrometer. Each spectrum was recorded with 65536 time-domain data points (real plus imaginary) corresponding to acquisition time of 4.09 sec. To improve signal to noise ration 16 scans were used along with $\pi/6$ pulse for excitation (pulse sequence zg30 in Bruker spectrometers).

NMR data processing workflow

The raw NMR data from spectrometer were first processed using TopSpin or MestreNova before using three different bioinformatics tools to assign peaks for metabolome studies, and the distribution analysis was performed using the software JMP 13 (SW). Fig. 1 summarizes the workflow of NMR data analysis using three different bioinformatics tools — the interface graphic of Galaxy <https://galaxy.workflow4metabolomics.org/>, the interface graphic of MetaboHunter <http://www.nrcbioinformatics.ca/metabohunter/>, and Chenomx software.

Briefly, the time domain data or the fid files, generated in the spectrometer, were fed to the MestreNova software to perform Fourier Transformation (FT) after synthetically removing water and solvent peaks. It should be noted that no water suppression technique was used while data collection. The frequency domain spectra were then saved in appropriate formats in order to determine the composition of whole milk using the three different bioinformatics tools. The MNova files obtained using MestreNova were directly used in Chenomx software. MestreNova can also generate JCAMP files. However, the software Galaxy needs spectra in Zip format. So, the JCAMP files, generated by MestreNova, were converted to Zip files using Bruker TopSpin software. Metabohunter, on the other hand, needs input files in Excel format — which were generated after peak picking the spectrum using MestreNova software and then saving the peak positions in Excel file format. The same Excel files were then fed in JMP 13 software to obtain the distribution of chemical shift (ppm) with intensity (Fig. 1).

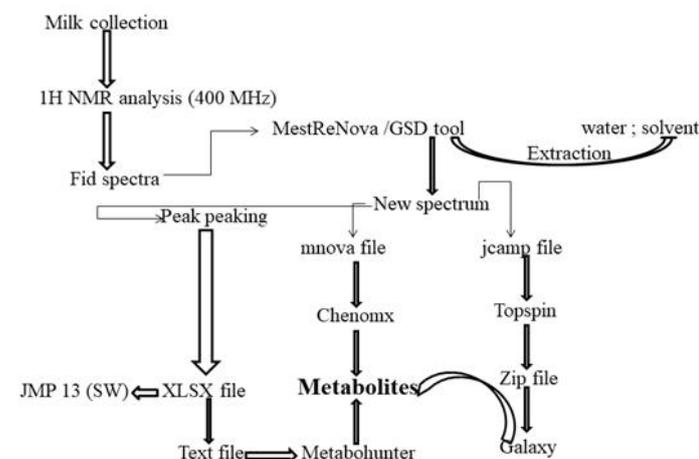


Figure 1 Metabolomics analysis workflow

RESULTS AND DISCUSSION

Fig. 2 illustrates ¹H NMR spectra of whole cow's milk after water and solvent peaks removal synthetically using MestreNova. As evident from the figure, ¹H NMR spectra of all the milk samples were solely dominated by the large water peak. However, GSD tool in MestreNova software efficiently removed the broad peak generated by water and the peak by DMSO d6. As a result, the important peaks corresponding to the milk constituents, those were previously obscured by water and solvent became clearly visible. Therefore, it is possible to characterize raw milk directly using NMR spectroscopy without prior purification or any pretreatment. Table 1 lists chemical shifts (in ppm) of H₂O (A, M1, M3 at 3.70, 3.94 and 3.56 ppm, respectively) and DMSO-d6 for each milk sample.

Table 1 ¹H NMR chemical shifts (in ppm) of water and DMSO-d6 in different whole cow's milk samples

	Samples		
	A	M1	M3
Water	3.700	3.490	3.569
	2.493	2.500	2.495
	2.498	2.505	2.500
DMSO-d6	2.503	2.507	2.506
	2.507	2.512	2.511
	2.513	2.517	2.516

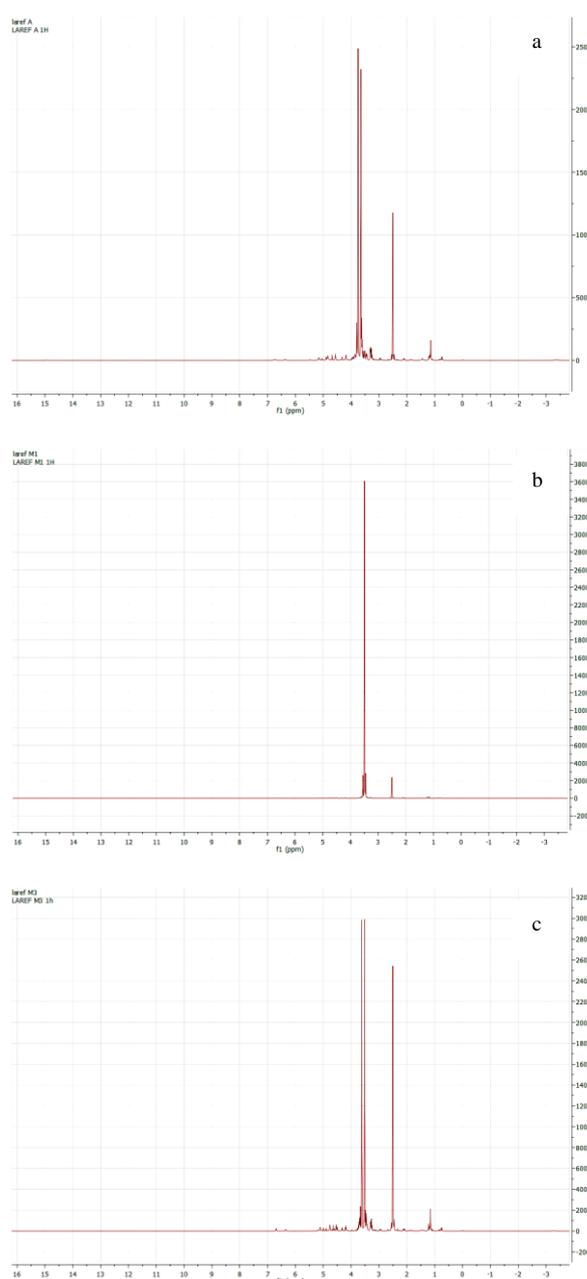


Figure 2 ¹H NMR spectra of different whole cow's milk samples after GSD tool treatment (a: sample A; b: sample M1; c: sample M3).

The bioinformatics tools enabled identification of different constituents in three milk samples after processing the spectrum with MestreNova for water and solvent removal. Table 2 lists different compounds found in the three test milk samples. A wide range of organic compounds were identified. Among them alanine, pyridoxamine-5-phosphate, taurine and threitol ...etc, were found in the test samples post-water removal. Hence, these were used to investigate the compositional differences among the test samples. Other substances, including glycerol, carnitine, choline, citrate and glycine were also identified. The findings are in accordance with the earlier reports (Palma *et al.*, 2017; Zhao *et al.*, 2017). Table 2 shows that not all the milk samples have similar metabolites composition. For example, certain compounds, such as arabitol, cellibiose and glycocholate, etc., were only found in the sample A, 3-hydroxyisovalerate, acetic acid, alanine and ethacrynic acid were only found in the sample M1 and carnitine, citrate and malic acid were only found in the sample M3. The variability in the metabolite profiles of different milk samples can be attributed to the different sources of metabolites. In other words, milk metabolites originate from multiple cell types or metabolisms in the organism.

Ketosis, a metabolic disorder, commonly occurs in dairy cows. High glycerophosphocholine (GPC) level and high ratio of glycerophosphocholine (GPC) and phosphocholine (PC) in milk indicate lower risk of ketosis and thus help in selecting healthy cows for breeding purpose. Klein *et al.*, 2012 first showed the correlation between ketosis risk and GPC and PC levels in milk using 600 MHz triple channel (¹H, ¹³C and ³¹P) NMR spectrometer. In the present study, the bioinformatics tool Galaxy detected GPC in the sample A and PC in the sample M3 (Table 2). However, no quantitative analysis was performed to obtain the levels of those metabolites. This study shows that each bioinformatics tool has identified a set of substances since each tool uses a special algorithm and file format.

Furthermore, distribution analysis was performed to obtain information about metabolites in different test samples (Fig. 3) using peak intensity. Most of the metabolites were found in the region between 3 ppm and 5 ppm. But certain metabolites showed very small signals in the region from 4 ppm to 6 ppm. Although those signals were very weak in MestreNova, peak peaking helped in visualizing certain substances in this region.

Table 2 Metabolites detected in whole cow's milk using bioinformatics tools

	Samples		
	A	M1	M3
1,3-Dimethylurate	**		**
1,3-Dimethylurea		***	
3-Hydroxyisovalerate		**	
6-Phosphogluconic acid trisodium salt	***		***
Acetic Acid		***	
Acetyl phosphate lithium, potassium salt			***
Acetylcholine chloride			***
Alanine		***	*
Arabitol	*		
Carnitine			*
Cellibiose	**		
Choline		**	**
Citrate			*
Ethacrynic acid		***	
Ethanol		***	
Glutathione reduced		***	
Glutathione, oxidized		***	
Glycerol	*		
Glycerphosphocholine	*		
Glycine		*	*
Glycocholate	**		
Guanidino acetic Acid	*		
Maleic Acid		***	
Malic Acid			*
Methionine methylsulfonium iodide	***		
Methyl Alcohol			***
N,N-DimethylGlycine	*		*
Niacinamide	**	**	**
Phosphocholine	*		*
Propylene Glycol	*		
Pyridoxamine-5-phosphate	***	***	***
Sodium glycolate	***		
Succinylacetone		**	
Tartaric Acid	*	***	*
Taurine	*	*	*
Threitol	***		
Threo-isocitric acid, monopotassium	*	***	*
TMAO		***	
Xylitol	*		*

* Galaxy ; ** Chenomx ; *** Metabolhunter

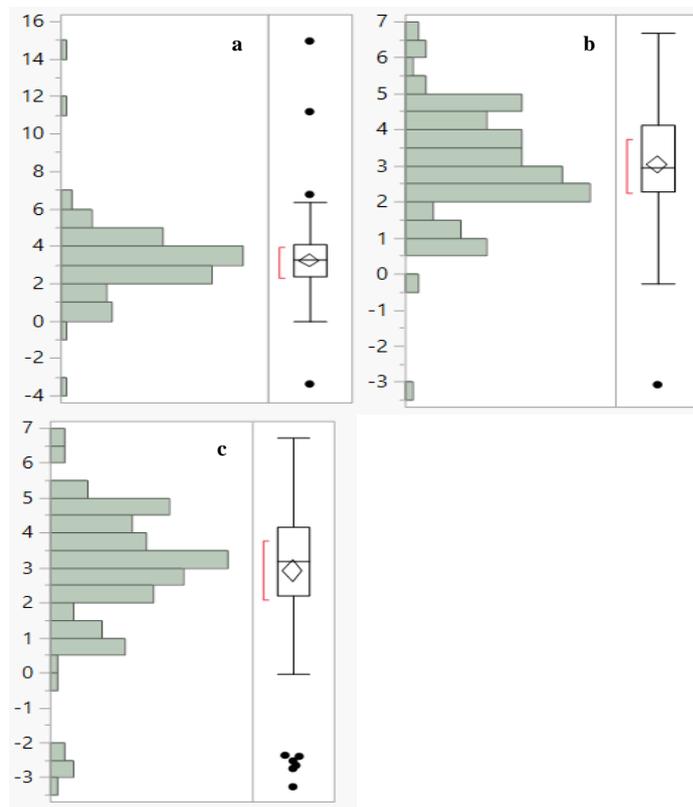


Figure 3 Chemical shifts distribution of metabolites identified in the whole cow's milk samples (a: sample A; b: sample M1; c: sample M3).

To know the richness of a milk sample, one needs to estimate the dry matter content (i.e., milk without water) since milk contains 87.7% of water in general. The dry matter content of milk was estimated by summing up the peak intensities, obtained after water and solvent suppressions, of all the metabolites. According to the sum of the peak intensities of the metabolites, M3 (10831.5) is the richest milk sample, followed by the samples M1 (8908) and A (8450.8).

CONCLUSION

In an attempt to show the efficiency of three different bioinformatics tools in analyzing milk constituents, without purification or pretreatment, using liquid-state 1D ¹H NMR spectroscopy, this study finds that MetaboHunter, Galaxy and Chenomx can identify many milk metabolites post water and solvent removal with MestreNova. The results, obtained from analyzing the data with the three tools, also show M3 as the richest milk sample among the three test samples. Despite the advantage of milk analysis using ¹H NMR without purification, these bioinformatics tools show certain limitations in identifying milk metabolites. Separate algorithms and file formats used by different tools could be the reason for obtaining dissimilar results in terms of metabolite identification in some cases. Therefore, while this study shows the advantage of milk analysis directly using ¹H NMR followed by data analysis with bioinformatics softwares, it also suggests the need of an improved algorithm that can identify the major milk metabolites easily and more efficiently.

REFERENCES

Andreotti, G., Lamanna, R., Trivellone, E., Motta, A. (2002) : ¹³C NMR Spectra of TAG: An Easy Way to Distinguish Milks from Different Animal Species. *J. Ame Oil Chemists' Society* 79 : 123-127. <https://doi.org/10.1007/s11746-002-0445-2>.
 Belloque, J., Smith, G.M. (1998) : Thermal denaturation of β-lactoglobulin. NMR study, *J. Agri. Food Chemi.* 46 : 1805-1813. <https://doi.org/10.1021/jf9709313>.
 Belloque, J. (2008) : High-Resolution NMR of Milk and Milk Proteins. *Modern Magnetic Reso.* 1631-1635. https://doi.org/10.1007/1-4020-3910-7_182.
 Belloque, J., Ramos, M. (1999) : Application of NMR spectroscopy to milk and dairy products. *Tren. Food Sci.Tech.* 10 :313-320. [https://doi.org/10.1016/S0924-2244\(00\)00012-1](https://doi.org/10.1016/S0924-2244(00)00012-1).
 Breitschuh, B., Windhab, E.J. (1997) : Parameters influencing cocrystallization and polymorphism in milk fat. *J. Ame Oil Chemists' Society.* 75 : 897-904. <https://doi.org/10.1007/s11746-998-0264-8>.

- Deborde, C., Rolin, D.B., Bondon, A., de Certaines, J.D.; Boyaval, P. (1998): In vivo nuclear magnetic resonance study of citrate metabolism in *Propionibacterium freudenreichii* subsp. *Shermanii*. *J. Dairy Res.* 65 : 503-514.
- Eads, T.M., Bryant, R.G. (1986): High-Resolution Proton NMR Spectroscopy of Milk, Orange Juice, and Apple Juice with Efficient Suppression of the Water Peak. *J. Agri. Food Chemi.* 34 : 834-837. <https://doi.org/10.1021/jf00071a017>.
- Griffin, M.C., Roberts, G.C.K. (1985): A ¹H-n.m.r study of casein micelles. *Biochemi J.* 119 228 : 273-276.
- Hu, F., Furihata, K., Ito-Ishida, M., Kaminogawa, S., Tanokura, M. (2004): Nondestructive Observation of Bovine Milk by NMR Spectroscopy: Analysis of Existing States of Compounds and Detection of New Compounds. *J. Agri Food Chemi.* 52 : 4969-4974. <https://doi.org/10.1021/jf049616o>.
- Klein, M.S., Buttchereit, N., Miemczyk, S.P., Immervoll, A.K., Louis, C., Wiedemann, S., Junge, W., Thaller, G., Oefner, P.J., Gronwald, W. (2012): NMR Metabolomic Analysis of Dairy Cows Reveals Milk Glycerophosphocholine to Phosphocholine Ratio as Prognostic Biomarker for Risk of Ketosis. *J. Proteo Res.* 11 : 1373-1381. <https://doi.org/10.1021/pr201017n>.
- Lamanna, R., Braca, A., Di Paolob, E., Imparato, G. (2011): Identification of milk mixtures by ¹H NMR Profiling. *Magn reso chemi.* 49 : 22-26. <https://doi.org/10.1002/mrc.2807>.
- Molinari, H., Ragona, L., Varani, L., Musco, G., Consonni, R., Zetta, L., Monaco, H.L. (1996): Partially folded structure of monomeric bovine beta-lactoglobulin. *FEBS Lett.* 381 : 237-243.
- Palma, M., Hernandez-Castellano, L., Castro, N., Arguēllo, A., Capote, J., Matzapetakis, M., Martinho de Almeida, A. (2016): NMR-metabolomics profiling of mammary gland secretory tissue and milk serum in two goat breeds with different levels of tolerance to seasonal weight loss. *Molec BioSystems.* 12 : 2094-2107. <https://doi.org/10.1039/C5MB00851D>.
- Remaud, G.S., Martin, Y.L., Martin, G.G., Naulet, N., Martin, G.J. (1997): Authentication of mustard oils by combined stable isotope analysis (SNIF-NMR and IRMS), *J. Agri. Food Chemi.* 45 : 1844-1848. <https://doi.org/10.1021/jf960448c>.
- Santos, P.M., Pereira-Filho, E.R., Colnago, L.A. (2016): Detection and quantification of milk adulteration using time domain nuclear magnetic resonance (TD-NMR), *Microchemi J.* 124 : 15-19. <https://doi.org/10.1016/j.microc.2015.07.013>.
- Van Hekken, D.L., Dudley, R.L. (1997): Analysis of Modified Whole Casein with Different Phosphorus Contents Using Phosphorus-31 Nuclear Magnetic Resonance and Fourier Transform Infrared Spectroscopy, *J. Dairy Sci.* 80 , 11. [https://doi.org/10.3168/jds.S0022-0302\(97\)76237-4](https://doi.org/10.3168/jds.S0022-0302(97)76237-4).
- Zhao, Y., Chen, H., Feng, J., Chen, Z., Cai, S. (2017): ¹H NMR-based compositional identification of different powdered infant Formulas. *Food Chemistry.* 230 : 164-173. <https://doi.org/10.1016/j.foodchem.2017.03.020>.