

Mid Infrared Spectroscopy and Multivariate Analysis Evaluation of Adulteration in Whey Protein Powder

Jonathan Andrade^{a,b} (dr.andradej@gmail.com), Cristina Guimarães Pereira^a (crisgp13@yahoo.com.br), Thamiris Ranquine^a (thamirisranquine@gmail.com), Maria Jose Valenzuela Bell^a (mariajose.bell@ufjf.edu.br), and Virgilio Anjos^a (virgilio.anjos@ufjf.edu.br)

^a Grupo de Engenharia e Espectroscopia de Materiais, Departamento de Física, Universidade Federal de Juiz de Fora, Rua José Lourenço Kelmer, s/n - São Pedro, 36036-900, Juiz de Fora, MG, Brazil.

^b Ryerson University, Faculty of Chemistry and Biology, 350 Victoria Street, Toronto, ON, M5B 2K3, Canada.

DOI: <https://doi.org/10.34019/2674-9688.2020.V2.29759>

Abstract

Mid infrared (MIR) spectroscopy was combined with multivariate approaches Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression to assess modifications in spectral profile of whey protein concentrate (WPC) powder due to changes in formulation level using caffeine, creatine and lactose, simulating fraud. Adulterations were made by replacing WPC in different levels from 5 to 50% (w/w - 5% steps) with three adulterants in separate. The spectra comparison of the samples allowed the identification of peaks associated to characteristic chemical bonds of each adulterant. PCA was carried out and 89% of the total variability of the spectral data was explained by three principal components, which allowed the confirmation of variables influencing each sample mixture and validating the spectral observations. Above 20% decrease in WPC content (20% adulteration), it was possible to differentiate all the three substances used. Predictions of percentage of WPC substitution were made through PLS regressions. The best prediction models were: lactose > creatine > caffeine. However, predictions resulted in overall good accuracy, low relative errors and coefficients of determination of fitting of calibration and validation curves above 0.97 in all cases. Therefore, techniques employed here aid the quality assessment of food products as alternative analytical tools.

Keywords: Adulteration; Chemometrics; Food Quality; Infrared Spectroscopy; Whey Protein.

1. Introduction

Whey protein concentrate (WPC) is a considered value-added product and is quickly acquiring a commodity status [1]. Its application in food areas is quite broad once provide several functional properties including gelation, high solubility, foaming/emulsification, sensory characteristics and water holding capacity being therefore used to improve textural attribute [2].

Other important uses for WPC are as nutritional supplement. Sports drinks and protein fortified beverages are examples of products claiming health benefits. Its high content of protein on a dry basis (65 - 80%) turns the supplement a rich source of bioactive peptides which may play a role in the dietary management of chronic diseases [3,4]. However, due to its wide versatility of applications and nutritional value, WPC is an expensive product and may have its quality compromised through adulterations [1,5,6]. These procedures can occur by changing the composition of a food product either by the addition of poor-quality own-product material, product-foreign material, or adulteration with other constituents [7]. In WPC, the ingredient authentication would not only limit fraud but also help ensure manufacturers in obtaining the desired functionality in whey products [8].

Among the innovative technologies useful in the food authentication is the Fourier Transform Mid-Infrared Spectroscopy (FTIR-MIR) which recently has been applied for the qualitative or quantitative determination of protein ingredients in food products [9,10], however researches attesting the authenticity of WPCs are still scarce up to now. In FTIR-MIR technique little sample preparation is needed, the analysis is fast, the use of hazardous solvents is minimized and also it has the advantages of being non-destructive, suitable for using in an industrial setting, with relatively low financial cost of obtaining and running the equipment [11,12]. Others alternatives of non-destructive optical spectroscopy techniques are protein tagging, FT-NIR, Raman and time-resolved fluorescence [13-16].

In this work, the authenticity of WPC was studied by FTIR-MIR spectroscopy combined with the multivariate approaches Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression. Adulterations were carried out through the replacement of WPC by three powdered components: lactose (milk carbohydrate), creatine (amino acid derivative, with claims to aid the increase of muscle force and strength) and caffeine (stimulant and thermogenic substance). The statistical approaches will allow the characterization of the mixtures pointing out variables influencing the fraud and prediction of the percentage of adulterations in WPC by these substances. The association of these tools can result in a powerful analytical technique for the characterization of whey protein-based products.

2. Material and Methods

2.1. Sample preparation

Two batches of whey protein concentrate powder were acquired from Infinitypharma (Nova Iguaçu, RJ, Brazil) and adulterations were formulated throughout the substitutions by anhydrous caffeine (Fagron do Brasil Farmacêutica Ltda., São Paulo, SP, Brazil), lactose monohydrate (Henrifarma Produtos Químicos e Farmacêuticos, São Paulo, SP, Brazil) and creatine (Infinitypharma, Nova Iguaçu, RJ, Brazil). The experiment consisted of 130 tablets

(130 mg each) comprising 10 tablets of the pure components (duplicates of WPC1, WPC2, caffeine, creatine and lactose) and the other 120 of duplicates of the adulteration mixtures in proportions of 5 to 50% (w/w) of substitution (with steps of 5%).

To prepare the tablets, the constituents were weighted in their respective proportion and homogenized with mortar and pestle. An automatic hydraulic press (Atlas™ Power Hydraulic Press T25, SPECAC INC, Fort Washington, PA, USA) was set to 3 tons, where the weighted powders were submitted to a hard-press treatment during 1 second. The resulting tablets showed 10 mm of diameter and 1 mm of thickness.

2.2. FTIR spectral investigation

MIR spectra of each tablet were acquired in duplicate using a FT-MIR Spectrometer Vertex 70 (Bruker Optik GmbH, Ettlingen, Germany) equipped with an ATR accessory and OPUS 6.5 software. The spectra were collected through 64 scans in the absorbance mode from 4000 to 400 cm^{-1} range with a resolution of 4 cm^{-1} . Analyses were carried out at room temperature (20 °C) and the background was collected before each sample measurement. With ATR, the spectrum is acquired when pressure is applied over the sample. Using tablets, we were able to standardize the pressure applied resulting in comparable spectra [17](Bruker, 2009). Prior to data analysis, background subtraction was carried out to eliminate background noises from the environment and results obtained were averaged within each sample. Pre-processing has continued by selecting areas of interest and normalizing the data (0-1) to standardize the variables, reducing the influence of the ones with biggest scales and highlighting nuances. Further, the normalized spectra were submitted to multivariate analysis.

2.3. Multivariate Approaches

PCA was applied to the normalized spectral sets of pure WPC and the mixtures with lactose, creatine and caffeine to characterize each substance and the behavior of the adulterated formulations. The wavenumber range applied was from 1800 to 400 cm^{-1} in a spectral resolution of 4 cm^{-1} with a total of 727 data points for each sample.

PLS was applied to calibrate the spectral profiles and predict the percentages of adulteration. According to Christy et al. [18], compared with other methodologies, PLS calibration gives more precise results. For each model created to predict lactose, creatine and caffeine added to WPC, the samples were divided into calibration and validation datasets. The samples were randomly chosen within the range of 0 to 50% of substitution, where the calibration set consisted of 7 samples while the validation dataset consisted of the other 4 which were not part of the calibration set [19].

The optimum number of latent variables were obtained using full cross-validation (leave-one-out) method and the prediction performances were evaluated throughout the root mean square error (RMSE - which represents the standard deviation of the residuals), the relative prediction errors (RE %), and the coefficient of determination (R^2) of both calibration and validation data sets. In general, as low as the RMSE and RE% values can be, and R^2 as close as possible to 1, the better will be the model predictions. Also, the overall bias (BIAS), the standard error of prediction (SEP) and the residual predictive deviation (RPD) over the validation dataset were calculated to evaluate the accuracy of the predictive model [20,21].

All statistical analyses were carried out using OriginPRO 2015 (Originlab Corp., Northampton, MA, USA) software.

3. Results and Discussion

3.1. MIR spectroscopy

When the sample matrix is composed of many different organic compounds, the infrared absorption bands obtained are the result of the combination of vibrations of many different chemical groups at certain wavenumbers, turning the analysis of the spectra an arduous task. In an attempt to identify important chemical groups or even compounds that might influence the spectra, the interpretation details of the spectra of the pure components and their mixtures with WPC are described as follows.

3.1.1 WPC characterization

Figure 1 shows the normalized spectra of pure WPC in the wavenumber range of 3600 to 400 cm^{-1} .

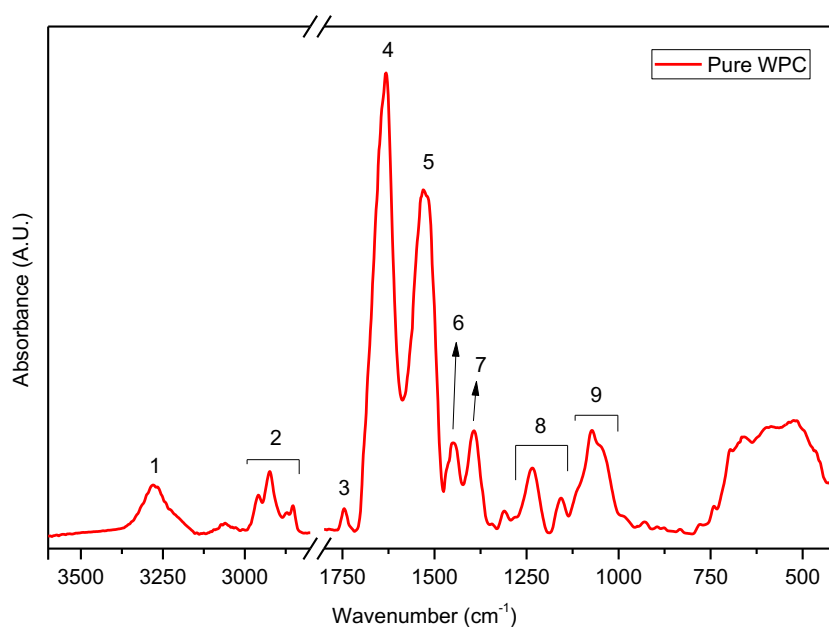


Figure 1 – Normalized MIR spectrum of pure WPC.

WPC spectra in the Figure 1 can be divided into two areas: the first one with high energy (3600 to 2800 cm^{-1}) and a second one (1800 to 900 cm^{-1}) which is considered a rich region presenting distinct bands associated with important functional groups [22,23]. In the area with high energy there are two peaks where the broad region around peak 1 (3279 cm^{-1}) is ascribed to stretching vibration of hydroxyl groups, commonly linked to water-binding, and the region 2, characteristic of stretching of C-H bonds [24].

The strongest peaks and main absorptions regions are within the wavenumber range of 1800 to 900 cm^{-1} . In the WPC characterization, the main concern is with absorption as result of vibrations in the amide bonds of amino acids and proteins, primarily Amide I (C=O

stretching vibration – peak 4, centered at approximately 1632 cm^{-1}), Amide II (N-H bending and C-N stretching vibrations – peak 5 at approximately 1531 cm^{-1}) and also Amide III (mainly C-N, stretching vibration – peak 8, around 1250 cm^{-1}) [8,25]. Peak 3 is associated to vibrational groups of C=O groups of lipids. The band at 1450 cm^{-1} (peak 6), corresponds to the bending of C-H bonds in CH₂ groups, while peak 7 (1392 cm^{-1}) to the stretching of C=O in carboxylic acid groups [8]. Finally, the vibration in the 1130 to 1000 cm^{-1} range (region 9) is linked to stretching vibrations of C-OH, C-C and C-O-C present in carbohydrates, mainly lactose [25,26].

The region below 1000 cm^{-1} (fingerprint region) is complex showing many bands that frequently overlap each other. This area and sometimes not used on spectral analysis due to its difficulty on assigning absorption bands once unique patterns are found in different organic molecules. However, they might be important on discrimination of substances.

3.1.2 Lactose adulteration

Regarding to pure lactose in Figure 2A can be observed a spectrum with several sharp and intense peaks, especially in the lower energy region. This feature makes the spectra of the lactose and others crystalline sugars complex and with many unassigned bands [27]. Table 1 (see **Supporting Information**) describes the chemical assignments from the highlighted bands in Figures 2A.

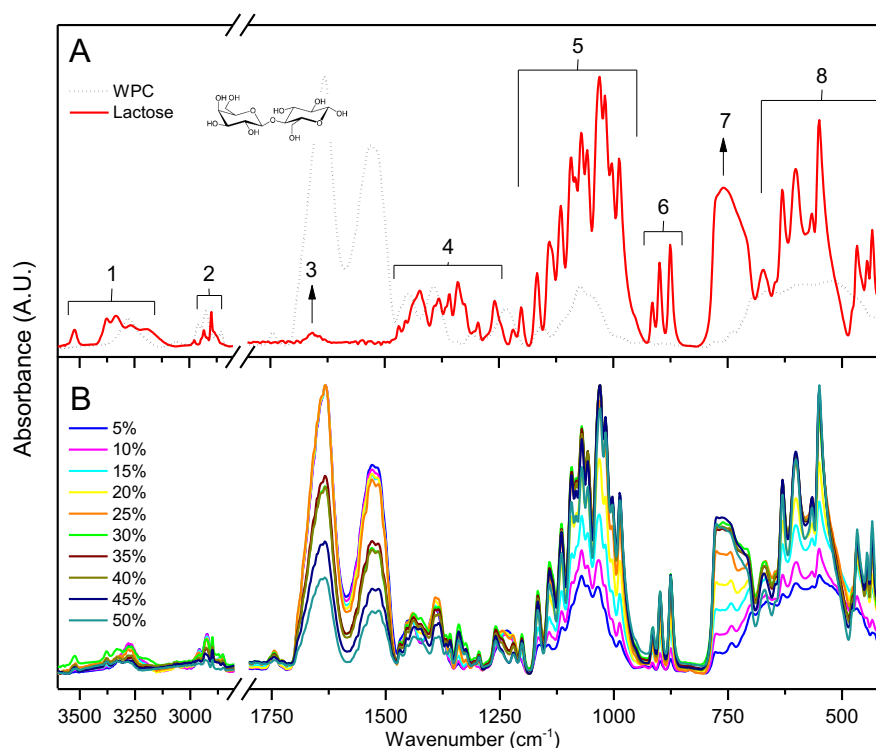


Figure 2 – Mid infrared spectra of (A) pure components: lactose and WPC; and (B) mixtures containing levels of WPC substituted by lactose (5 to 50% w/w).

According to Zhou et al. [26] the region with lower energy 1250 to 800 cm^{-1} is rich in intense bands and embodies characteristic peaks of C-O and C-C vibrations present in carbohydrates as lactose. Even though this region represents vibrations associated to the

same compound (lactose) in both WPC and the adulterant, their shapes are quite different (Figure 2B), which occurs due the lactose state presents in each sample. In the WPC, beside the whey proteins, lactose is also an important constituent, which is in a meta-stable glassy state (amorphous lactose) since its crystallization is impaired by the presence of proteins [4,28,29]. On the other hand, pure lactose (used as an adulterant) is mostly in its crystalline form, since is thermodynamically more stable than the amorphous state [30]. In general, the spectrum of amorphous lactose (present in WPC) could be distinguished from those of crystalline lactose (present in pure lactose) by peaks less defined and in a smaller number as occurred in Figure 1 [31,32].

In Figure 2B some modifications can be observed in the spectra due to the lactose substitution in the samples. By comparing the spectra of pure substances in Figure 2A it can be noted the most remarkable differences between the spectra occur mainly in the regions of 1800 to 1460 cm^{-1} (characteristic region in WPC) and 1200 to 400 cm^{-1} (where intense peaks were found in lactose). In the former region, the addition of lactose brings effects in the normalized spectra intensity of mixtures from 30% adulteration. Once this range is mainly related to the presence of protein (Amide I and II), the replacement of WPC by lactose has led to a decrease in the spectra intensity in this region. Also, the peak 3 (showed in the lactose spectrum – Figure 2A) was masked in Figure 2B by the high intensity of the amide peaks. The latter (below 1200 cm^{-1}), it can be noted that in the range of regions 5 and 8 (Figure 2A), there was an increase in the intensity of the peaks as soon as adulteration was carried out (Figure 2B). Regarding the shape of the peaks, the addition of only 5% of pure lactose into the WPC already altered the spectra and with increasing concentration, they became sharper. That fact may be related to the lactose state which is in a crystalline form in the pure lactose sample, as previously stated [27,32].

In the region 4 we can also visually realize wavenumbers evincing the adulterations, even though the peak intensities are lower in this range comparing to the others. The range from 1369 to 1248 cm^{-1} is dominated by small sharp peaks coming from deformation modes of the CH, OH and CH₂ groups present in the lactose samples, but not visualized in the WPC spectrum (Figure 2A).

Regarding area of peak 6 and peak 7, they have arisen in all adulterated samples (5 to 50%) (Figure 2B). However, they are characteristic of the pure lactose sample, not occurring in the spectrum of the WPC (Figure 2A). This fact suggests that these regions might be areas with significance when evaluating the adulteration of WPC by lactose.

3.1.3 Caffeine adulteration

The most relevant peaks of pure caffeine were found in the low energy region, between 1800 and 400 cm^{-1} (Figure 3A). The chemical structure of caffeine contains two rings fused (imidazole and pyrimidine) comprising 2 carbonyl groups and 10 C-N bonds, that can contribute to the vibrational intensity obtained. The chemical bonds related to the peaks in Figure 3A are described in Table 1 (see **Supporting Information**).

In Figure 3B it is possible to notice some differences between WPC and pure caffeine in the region above 2800 cm^{-1} , where no peaks were observed in the caffeine sample. The intense protein peaks of WPC amide I (1631 cm^{-1}) and amide II (1530 cm^{-1}) showed a decreasing trend as the substitutions went on. Even decreasing the intensity, these peaks mask the presence of smaller intensity peaks characteristic of caffeine in the same region,

with the exception of the caffeine peak 1 (1693 cm^{-1}), that changes the shape of the base of the Amide I peak even at lower levels of substitution, rising in accordance with caffeine concentration in the mixtures. The Amide I consists of combination of stretching vibrations of C=O and N-H chemical groups, commonly found in peptide bonds [33], while in caffeine, the peak 2 (1643 cm^{-1}) was assigned only to C=O stretch only [34] and might be related to Amide I as well. In fact, there is even a shift in the amide I peak from 1631 cm^{-1} in the pure WPC sample to 1643 cm^{-1} in the 50% mixture. Following similar pattern, the shape of the WPC Amide II peak (1530 cm^{-1}) had changed due to the decrease in its the intensity with the substitution, but also the contribution of the caffeine peak at 1544 cm^{-1} (peak 4). Another decrease in intensity observed involves the WPC's carbohydrate region surrounding the C-OH stretch at 1070 cm^{-1} (peak 9, Figure 1), which had the spectra changed as well due to the contribution of caffeine peaks in smaller wavenumbers.

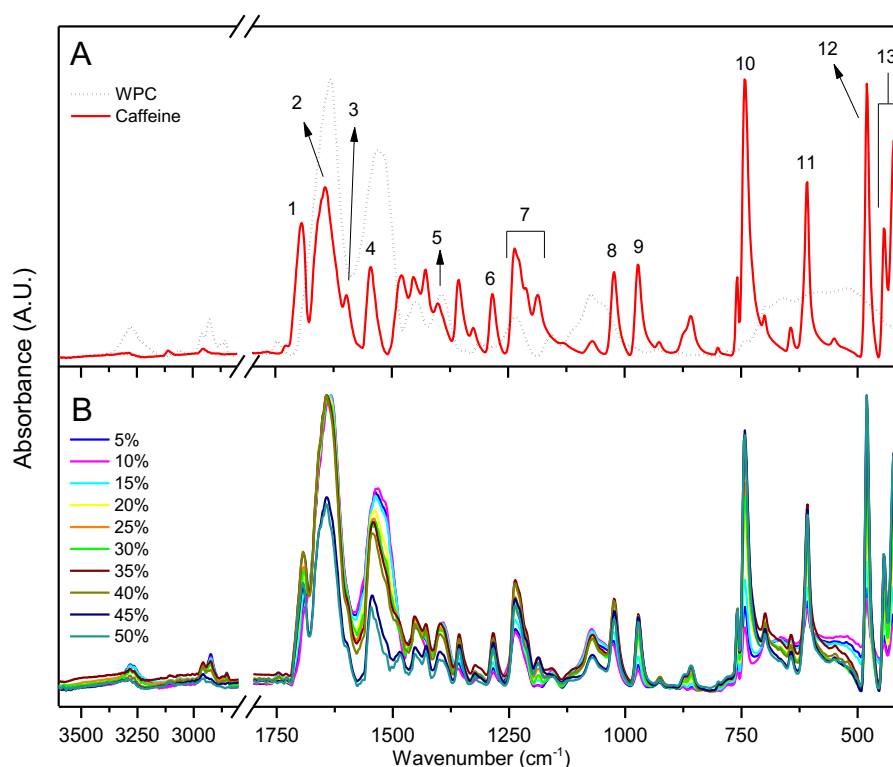


Figure 3 – Mid infrared spectra of (A) pure components: caffeine and WPC; and (B) mixtures containing levels of WPC substituted by caffeine (5 to 50% w/w).

The main changes in the spectra regarding the appearance of characteristic peaks of caffeine happened in the fingerprint region below 1000 cm^{-1} . It is noticeable the gradual increase of the peaks that showed high intensity in the pure caffeine sample, especially those related to imidazole and pyrimidine rings vibrations. According to [35], this area was significantly important for the use in quantification of caffeine in pharmaceutical formulations, and it might play an important role in this study as well.

3.1.4. Creatine adulteration

As in caffeine case, creatine revealed highly intense peaks in the wavenumber range below 1800 cm^{-1} , with some significant differences found when compared to WPC sample (Figure 4A – Table 1). Creatine is a nitrogenous organic acid with biological activity in the muscle and brain tissue, therefore, its infrared spectra is dominated by N-H, C=O and C-N bond vibrations [36].

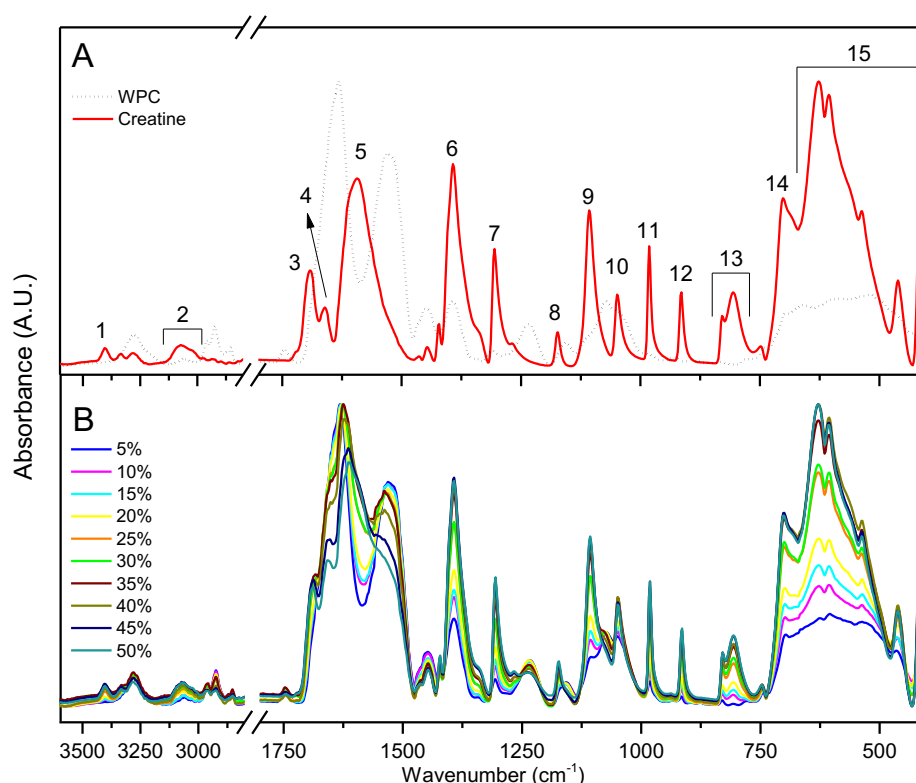


Figure 4 – Mid infrared spectra of (A) pure components: creatine and WPC; and (B) mixtures containing levels of WPC substituted by creatine (5 to 50% w/w).

The major differences observed in the mixtures of WPC and creatine (Figure 4A and 4B) are the link of the region comprising the Amide I and II peaks, the range surrounding 1070 cm^{-1} band (C-OH stretch of carbohydrates) and the highly intense peaks characteristic of creatine, in special below 1000 cm^{-1} . The presence of creatine in the formulated mixtures gradually changed the shape of these areas as seen in Figure 3B. In the first case, the Amide I and II bands became broader than pure WPC sample, in addition to the reduction of intensity of these bands and rising of the absorption peak corresponding to the peak 5 of creatine (1593 cm^{-1}) in between them, which agrees with the rising of the creatine level in the mixtures. In case of the region between 1130 and 1000 cm^{-1} , the band changed even in concentrations as low as 5% of creatine due to the peaks 8 and 9 at 1107 cm^{-1} and 1047 cm^{-1} . Peaks as 1305 cm^{-1} (7), 981 cm^{-1} (11), 914 cm^{-1} (12) and the broad bands below 850 cm^{-1} comprising the region 13, peak 14 and region 15 (Figure 3A), also showed a rising trend in the mixtures.

With the characterization of the differences and similarities found, the visual observations

obtained provide insights of certain regions that might influence the differentiation of the samples. Thus, due to the size of the datasets, it is essential that chemometric analyses support the spectral results. Unsupervised exploratory approaches such as PCA can be used to validate the observations. In addition, in order to perform a quantification analysis, multivariate calibration is required [37].

3.2 Principal components analysis

Once the main differences within the samples were observed in the spectral range below 1800 cm^{-1} , the normalized spectra of pure WPC and the mixtures of caffeine, creatine and lactose, from 1800 to 400 cm^{-1} , were submitted to principal component analysis (Figure 5). To avoid biased discrimination of the samples, the spectra of the pure adulterants were not included in this analysis, emphasizing the differences in relation to pure WPC.

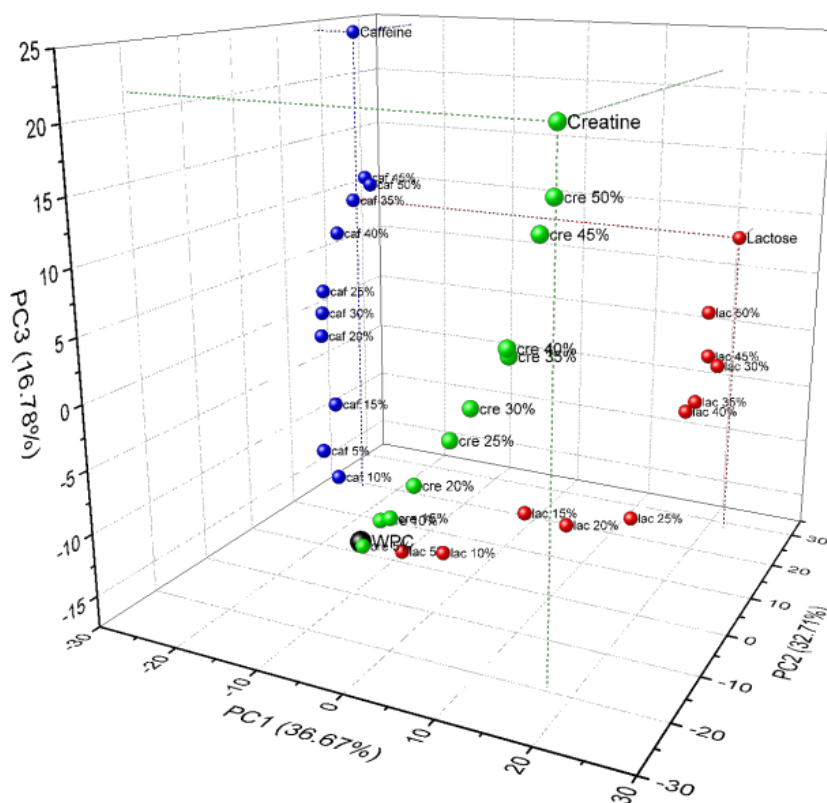


Figure 5 – Score plot result from PCA of pure WPC, Caffeine (caf), Creatine (cre), Lactose (lac) and different levels of substitution of WPC in weight.

Figure 6 presents the same plot of Figure 5 by pairs of principal components. From the figure approximately 89% of the total variance of the spectral data was explained by the first three principal components. PC1 accounted for 38.81% of the variability of the data and highlight the differences of pure WPC and caffeine treatment against the others, since they were positioned in opposition from creatine and lactose samples, in special the highly concentrated ones (Figures 6A and 6B). Among creatine and lactose substitutions, it can be seen that all creatine samples showed similarities with low levels of substitution of lactose, up to 20%. The higher levels of lactose substitution were positioned in the most positive end of the PC1 axis. All caffeine substituted samples were correlated in relation to PC1. On the

other hand, PC2 explained 33.31% of the variance, with creatine substituted samples being positioned in its positive scale while caffeine and lactose mostly in the negative side. Although in the positive scale, closer to the origin of PC2 axis, the WPC sample showed similarities with the lowest levels of substitutions of these substances. The group of samples containing of creatine levels above 20% revealed less differences within them. The wavenumbers that led to the positioning of the lowest levels may indicate the main areas of the spectra responsible for the differentiation of the creatine substituted samples.

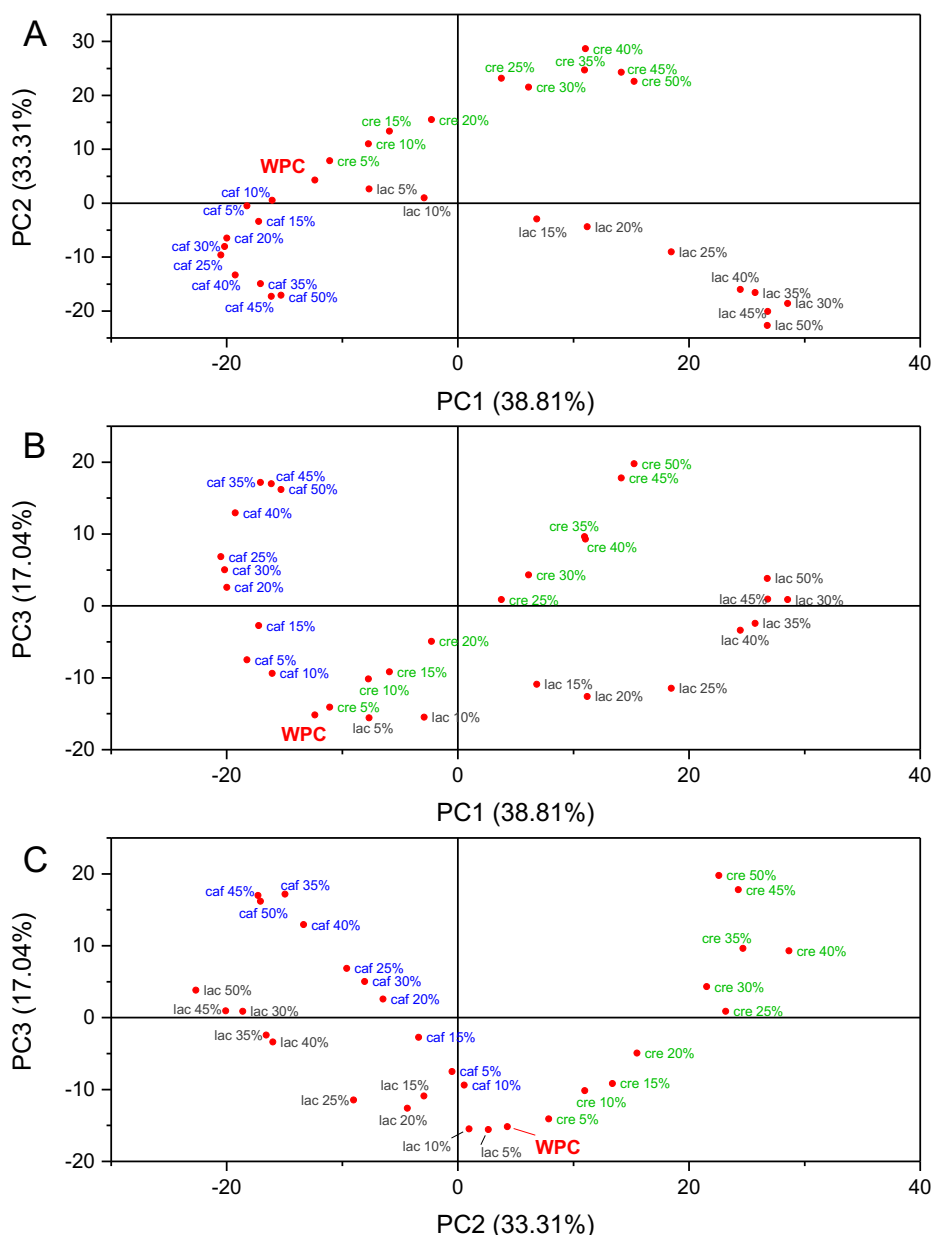


Figure 6 – Score plots of principal component analysis of the normalized spectral data of the mixtures of caffeine, creatine and lactose with WPC, within the range of 0 to 50% of substitution (w/w). (A) PC1 x PC2; (B) PC3 x PC1; (C) PC2 x PC3.

Explaining approximately 17% of the total variability of the data, PC3 showed the WPC sample in opposition of the highest levels of substitution. Substitution levels above 15% of caffeine showed more similarities with fewer samples of the other substances used. The score plot (Figure 6C) showed only three lactose concentrations (30, 45 and 50%) and

concentrations above 20% of creatine showed in the positive scale of PC3. The decrease of intense protein related peaks might be influencing this principal component, which can be clarified by analyzing the loading plots (Figure 7).

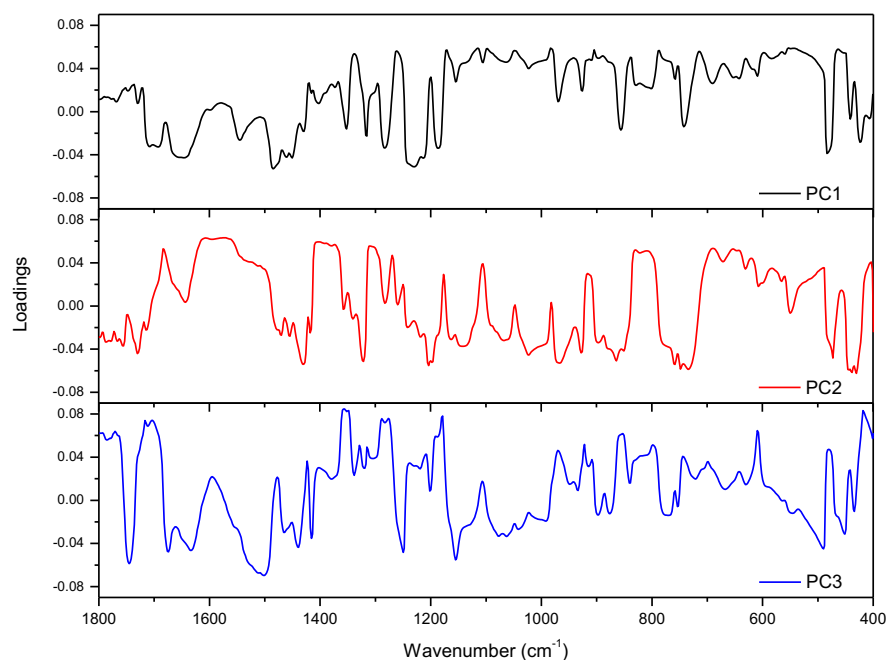


Figure 7 – Loading plot of the three first principal components explaining approximately 89% of total variability of the spectral data.

The regions where PC1 weighted positively referred mainly to areas where intense changes in the spectra were observed due to peaks characteristic of lactose and creatine. The range 1175 to 970 cm^{-1} is characterized by the presence of lactose in the samples, where vibrations of C-O, C-O-C, C-OH and C-C groups are found [38]. In the interval 730 to 490 cm^{-1} , positive loadings were also found in PC1. This range showed peaks in both creatine and lactose samples, which contributed for their correlations in the score plot (Figure 6A). PC1 weighted negatively in fewer regions, with highlight in peaks intense only when caffeine is present in the mixtures, such as the stretching of C=O at 1693 cm^{-1} and vibrations of imidazole and pyrimidine rings at 1236 and 480 cm^{-1} .

Most of the intense positive loadings obtained in the PC2 are related to peaks exclusive from creatine, in special the centered at 1593 cm^{-1} , 1392 cm^{-1} , 1305 cm^{-1} and the region surrounding the peak at 806 cm^{-1} . The decrease in WPC's Amide II band (1530 cm^{-1}) had also shown positive loadings, which correlates with the substitution in weight promoted. Another range revealing high loadings is between 710 and 560 cm^{-1} , where peaks associated with creatine and lactose are prominent. Notwithstanding, negative loadings occurred mainly where intense peaks of lactose were present, with few breaks in between, where characteristic creatine peaks changed the loadings to positive. It is possible to notice that in the regions where negative loadings were found, caffeine peaks agreed with the lactose ones, therefore, influencing the positioning of these two adulterants in opposition of creatine in the score plots (Figure 6A and 6C).

Lastly, negative loadings of PC3 correlates to regions of the spectra where higher intensity of signal was found in pure WPC sample or in those with low levels of substitution,

while positive loadings reflected regions where the adulterated samples showed up. Two regions that deserve attention here: between 865 and 792 cm^{-1} , with positive loadings, where two small broad peaks are present in caffeine and creatine samples (centered at 858 and 806 cm^{-1} , respectively) seemed to matter; and from 1100 and 985 cm^{-1} , where the intense lactose characteristic peaks weighted negatively. These observations suggest that PC3 is much more related to the substitution levels, than to a discrimination of a particular substance.

When analyzing the principal components in combination, it is clear that WPC sample revealed strong similarities with lowest levels of substitution of each substance used. Overall, it was noted a well-defined differentiation in relation to concentrations above 20% of each adulterant. Previous research has shown that these three adulterants could be differentiated using different spectroscopic analysis, with similar discrimination obtained [33], however, no prediction models were studied.

3.3 PLS regressions

PLS is based on two matrixes, being one made of the predictor values (X matrix) and the other with the observed values (Y matrix), which the calibration relies on. In this case, the percentage of substitution of WPC, from 0 to 50% (w/w), was used as Y matrix, while the X matrix consisted of the selected normalized MIR spectra data, therefore, the models were based on 727 variables. The calibration was promoted using full cross validation in all cases.

The calibration and validation regressions can be visualized in Figure 8 and the PLS performance and accuracy parameters were calculated and presented in Table 2. The RMSE, RE% and R2 were calculated for both calibration and validation datasets, while BIAS, SEP and RPD only to the validation sets.

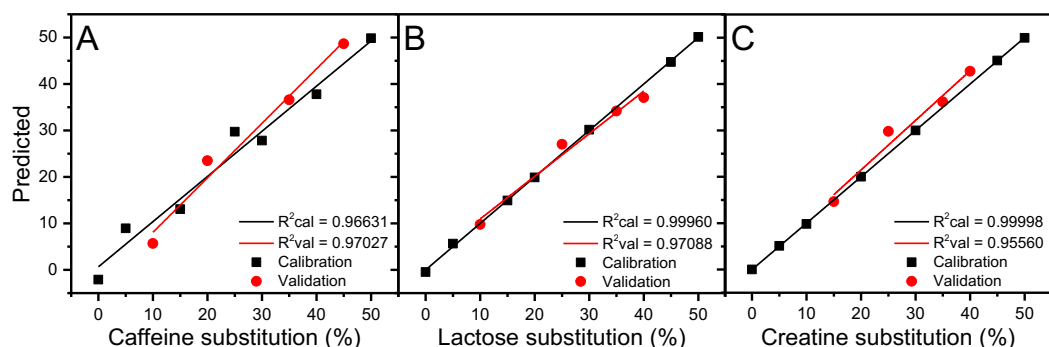


Figure 8 – PLS calibration and validation regressions of predictions of the substitutions with (A) caffeine, (B) lactose and (C) creatine.

Table 2. PLS predictions of the percentage of substitution of caffeine, lactose and creatine in whey protein powder.

Prediction of	Calibration $n = 6$			Validation $n = 4$					
	R^2	RMSE	RE%	R^2	RMSE	BIAS	SEP	RPD	RE%
Caffeine	0.96631	2.822	9.74	0.97027	3.415	1.101	3.233	4.808	11.15
Lactose	0.99960	0.324	1.10	0.97088	1.828	-0.496	1.760	7.517	6.14
Creatine	0.99998	0.078	0.27	0.95560	2.827	2.082	1.912	5.799	9.33

The number of factors that resulted in better prediction model in relation to the caffeine treatment was the one with 2 factors. These factors explained 98.2% and 97.2% of the cumulative variance of the spectral data and the percentage of substitution, respectively. Although the coefficient of determination of both calibration and validation curves were good, with values approximately 0.97, it was not possible to obtain a good superposition of the curves, which reflected in higher deviations and the highest relative error amongst the model predictions (Figure 8A, Table 1).

Overall, the results obtained for the creatine and lactose predictions were better than the caffeine. In the creatine and lactose predictions, the best predictions were obtained with the models with 4 factors, which explained above 99.8% of the cumulative variance of the data matrixes. The coefficient of determination of the calibration curves were above 0.99, while the validation above 0.95. Although in creatine treatment, the relative errors obtained were a bit higher than in lactose substitution, there are overlays of the calibration and validation curves, indicating better performance. Regarding the RMSE values, that represents the standard deviation of the residues of prediction, it can be noted that the validation dataset resulted in higher values, which was expected, due to the reduced number of samples used.

Among the wavenumbers that most influenced the prediction models, the WPC's Amide I and II peaks (1640 and 1530 cm^{-1}) were valued in all the three prediction models, since, in all cases, these peaks showed a reduction trend following the substitutions. In addition to these peaks, the fingerprint region (below 1000 cm^{-1}) also played an important role in all cases. In the caffeine model, the peaks at 742, 480 and 424 cm^{-1} weighted significantly to the good performance of the model. Creatine on the other hand showed several more peaks, with highlight to the region between 720 and 520 cm^{-1} , that showed high intensity in the mixtures. Finally, lactose substitution was the one with higher number of variables affecting the PLS model. The range 1140 to 980 cm^{-1} , the broad peak surrounding 770 cm^{-1} , and the range 630 to 530 cm^{-1} significantly improved the performance of the model.

In relation to the accuracy parameters, the models calculated revealed low BIAS values, with low contribution to the calculated SEP. Lactose substitution prediction was the one showing negative BIAS, consequently, with a slight underestimation of the values in general. The results obtained for the RPD indicates the overall accuracy of the models, where the lactose prediction can be classified as very good, followed by creatine (good) and caffeine (fair) models [21]. Therefore, although some imprecision was found in the models obtained, the methodologies and techniques employed showed great potential in the quantitative assessment of formulations of WPC.

4. Conclusions

In this work, the mid infrared spectral analysis coupled to multivariate statistical approaches allowed the characterization (qualitatively and quantitatively) of mixtures of different compounds in WPC. PCA allowed the segregation of the samples according to the model molecules used as adulterants in three different groups. The differentiation of the mixtures was feasible when the level of substitution was higher than 20% w/w. The lactose and creatine prediction model were the ones that more regions of the spectra correlated with the adulteration levels, resulting in overall good values in relation to the performance and accuracy parameters. Although caffeine model did not perform well as the others, it came out

in fair results. Therefore, the techniques employed can be of great importance to assess the quality of dried protein-based products, as alternatives to the classical non-environmental friendly chemical methodologies.

Supporting Information

Table 1 – IR band assignments for pure lactose, caffeine and creatine. Molecular structure and MIR spectra are presented in figures 2, 3 and 4.

	Region	Wavenumber (cm ⁻¹)	Chemical assignment	References
Lactose – Figure 2A	1	3560 - 3200	ν OH region	Wiercigroch et al. [27]
	2	2934 and 2899	ν CH	Wiercigroch et al. [27]
	3	1659	δ OH – crystal water	Listiohadi & Hourigan [31]
	4	1480 to 1250	ω CH ₂ ; ρ CH ₂ ; ρ CH; ρ OH	Wiercigroch et al. [27]
	5	1200 - 1000	ν COC; ν CO; ν C-OH; ν CC	Kędzińska-Matysek, Matwijczuk, & Florek [38] Listiohadi & Hourigan [31] Wiercigroch et al. [27]
	6	925 - 850	COH deformation; ν CC	Wiercigroch et al. [27]
	7	760	τ CO; τ COHO; τ HOH	Wiercigroch et al. [27]
	8	800 - 400	deformation of endocyclic OCC, OCO and COO	Wiercigroch et al. [27]
Caffeine – Figure 3A	1	1693	ν C=O	Rodrigues, Cardoso, Andrade, Donnici, & Sena [34]
	2	1643	ν C=O	Rodrigues et al. [34]
	3	1596	ν C=N	Paradkar & Irudayaraj [39] Schulz & Baranska [40]
	4	1544	ν N-C=O	Paradkar & Irudayaraj [39] Schulz & Baranska [40]
	5	1400	ν C-N imidazole ring; in-plane δ CH ₃	Palo et al. [35]
	6	1284	ν C-N	Edwards, Munshi, & Anstis [41]
	7	1236	ν imidazole and pyrimidine rings; δ N-H	Ucun, Sağlam, & Güçlü [42]
	8	1022	ν C=O ketonic; ν C-N	Palo et al. [35] Paradkar & Irudayaraj [39]
	9	972	ν imidazole and pyrimidine rings; ρ CH ₃	Kesimli, Topaçli, & Topaçli [43]
	10	742	ν imidazole and pyrimidine rings; δ and ν CH ₃	Edwards et al. [41]
	11	610	τ N-C-N out-of-plane (pyrimidine)	Palo et al. [35] Ucun et al. [42]
	12	480	δ out-of-plane (pyrimidine); δ C-N-C (imidazole)	Palo et al. [35] Ucun et al. (2007)
	13	445 and 420	δ in-plane pyrimidine ring	Palo et al. [35] Ucun et al. [42]

Region	Wavenumber (cm ⁻¹)	Chemical assignment	References	
Creatine – Figure 4A	1	3404	unknown; ν OH region	Wiercigroch et al. [27]
	2	3072	unknown	---
	3	1690	ν C=O	Podstawka, Światłowska, Borowiec, & Proniewicz [44]
	4	1660	ν C=N	Podstawka et al. [44]
	5	1593	ν NH ₂	Podstawka et al. [44]
	6	1392	ν C=O (-COOH)	T. Wang et al. [8]
	7	1305	ν CH ₃ ; ν NH ₂ ; ρ NH ₂	Podstawka et al. [44]
	8	1172	δ N-CH ₃ ; ν C-N-C	Podstawka et al. [44]
	9	1107	δ N-CH ₃	Podstawka et al. [44]
	10	1049	δ R-NH ₂	Podstawka et al. [44]
	11	981	ν CH ; ρ CH ₂	Podstawka et al. [44]
	12	916	ν C-COH	Podstawka et al. [44]
	13	850 to 770	ν C=N and ρ R-NH ₂	Gangopadhyay et al. [45] Podstawka et al. [44]
	14	702	ν NH ₂	Podstawka et al. [44]
	15	670 to 400	unknown, skeletal	Podstawka et al. [44]

ν : stretch, δ : deformation or bending, ρ : rocking, ω : wagging.

Acknowledgments

This work was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES Foundation (Ministry of Education of Brazil, Brasília – DF), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG Foundation, Brazil) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil). The authors also thank CAPES Foundation regarding the bursaries provided to JA, CGP and TR. The authors declare that they have no competing commercial or financial interests.

References

- [1] Garrido, B. C.; Souza, G. H. M. F.; Lourenço, D. C.; Fasciotti, M., *Journal of Proteomics*, 2016, 147, 48–55. <https://doi.org/10.1016/j.jprot.2016.03.044>
- [2] Dissanayake, M.; Kelly, A. L.; Vasiljevic, T., *Journal of Agricultural and Food Chemistry*, 2010, 58(11), 6825–6832. <https://doi.org/10.1021/jf1009796>
- [3] Patel, S., *J. of Func. Foods*, 2015, 19, 308–319. <https://doi.org/10.1016/j.jff.2015.09.040>
- [4] Tunick, M. H., In *Whey processing, functionality and health benefits* (1st ed.) 2008 Ames, Iowa: Wiley-Blackwell.
- [5] De Ceglie, C.; Calvano, C. D.; Zambonin, C. G., *Food Chemistry*, 2015, 176, 396–402.

<https://doi.org/http://dx.doi.org/10.1016/j.foodchem.2014.12.038>

- [6] Geyer, H.; Parr, M. K.; Koehler, K.; Mareck, U.; Schänzer, W.; Thevis, M., *Journal of Mass Spectrometry*, 2008, 43(7), 892–902. <https://doi.org/10.1002/jms.1452>
- [7] Pustjens, A. M.; Weesepeel, Y.; van Ruth, S. M. In *Innovation and Future Trends in Food Manufacturing and Supply Chain Technologies* 2016, pp. 3–20. Elsevier. <https://doi.org/10.1016/B978-1-78242-447-5.00001-0>
- [8] Wang, T.; Tan, S. Y.; Mutilangi, W.; Aykas, D. P.; Rodriguez-Saona, L. E., *Journal of Food Science*, 2015, 80(10), C2111–C2116. <https://doi.org/10.1111/1750-3841.13006>
- [9] Andrade, J.; Pereira, C.; Ranquine, T.; Azarias, C.; Bell, M. J. V.; Anjos, V. C., *Journal of Spectroscopy*, 2018 (Article ID 1381864), 9. <https://doi.org/10.1155/2018/1381864>
- [10] Chang, S. K. C.; Zhang, Y., In S. Nielsen (Ed.), *Food analysis*, 2017 (pp. 315–331), Springer, Cham.
- [11] Danezis, G. P.; Tsagkaris, A. S.; Camin, F.; Brusica, V.; Georgiou, C. A., *Trends in Analytical Chemistry*, 2016, 85, 123–132. <https://doi.org/10.1016/j.trac.2016.02.026>
- [12] Rodriguez-Saona, L. E.; Allendorf, M. E., *Annual Review of Food Science and Technology*, 2011, 2(1), 467–483. <https://doi.org/10.1146/annurev-food-022510-133750>
- [13] Paiva, I. M.; Anjos V. C.; Bell, M. J. V.; Furtado, M. A. M., *Revista do Instituto de Laticínios Cândido Tostes*, 2015, 70, 192-199.
- [14] Luiz, L. C.; Bell, M. J. V.; Rocha, R. A.; Leal, N. L.; Anjos, V. C., *Journal of Spectroscopy*, 2018, 5152832, 1-6.
- [15] Rocha, R. A.; Paiva, I. M.; Anjos, V.; Furtado, M. A. M.; Bell, M. J. V., *Journal of Dairy Science*, 2015, 98(6), 3559-3567.
- [16] Brandao, M. P.; Neto, M. G.; Anjos, V. C; Bell, M. J. V., *Food Control*, 2017, 81, 168-172.
- [17] Bruker. (2009). A225 Platinum ATR Accessory User Instructions (1st ed.). Ettlingen, Germany.: Bruker Optik GmbH.
- [18] Christy, A. A.; Kasemsuran, S.; Du, Y.; Ozaki, Y., *Analytical Sciences*, 2004, 20(6), 935–940. <https://doi.org/10.2116/analsci.20.935>
- [19] Kaylegian, K. E.; Houghton, G. E.; Lynch, J. M.; Fleming, J. R.; Barbano, D. M., *J. of Dairy Science*, 2006, 89(8), 2817–2832. [https://doi.org/10.3168/jds.S0022-0302\(06\)72555-3](https://doi.org/10.3168/jds.S0022-0302(06)72555-3)
- [20] Andrade, J.; Pereira, C. G.; Almeida Junior, J. C.; Viana, C. C. R.; Neves, L. N. O.; Silva, P. H. F.; Bell, M. J. V.; Anjos, V. C., *LWT*, 2019, 99, 166–172.
- [21] Williams, P., in P. Williams & K. Norris (Eds.), *Near Infrared Technology in the Agriculture and Food Industries 2*, 2001, (2nd ed., pp. 146–169). St. Paul, MN: Am. Assoc. Cereal Chem.
- [22] Li, B.; Wang, H.; Zhao, Q.; Ouyang, J.; Wu, Y., *Food Chemistry*, 2015, 181, 25–30. <https://doi.org/10.1016/j.foodchem.2015.02.079>

- [23] Wang, X.; Esquerre, C.; Downey, G.; Henihan, L.; O'Callaghan, D.; O'Donnell, C., *Food Analytical Methods*, 2018, 11(5), 1380–1389. <https://doi.org/10.1007/s12161-017-1114-9>
- [24] O'Loughlin, I. B.; Kelly, P. M.; Murray, B. A.; FitzGerald, R. J.; Brodtkorb, A., *International Journal of Dairy Technology*, 2018, 68(3), 349–356. <https://doi.org/10.1111/1471-0307.12239>
- [25] Lei, Y.; Zhou, Q.; Zhang, Y. L.; Chen, J. B.; Sun, S. Q.; Noda, I., *Journal of Molecular Structure*, 2010, 974(1–3), 88–93. <https://doi.org/10.1016/j.molstruc.2009.12.030>
- [26] Zhou, Q.; Sun, S. Q.; Yu, L.; Xu, C. H.; Noda, I.; Zhang, X. R., *Journal of Molecular Structure*, 2006, 799(1–3), 77–84. <https://doi.org/10.1016/j.molstruc.2006.03.025>
- [27] Wiercigroch, E.; Szafraniec, E.; Czamara, K.; Pacia, M. Z.; Majzner, K.; Kochan, K.; Malek, K., *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2017, 185, 317–335. <https://doi.org/10.1016/j.saa.2017.05.045>
- [28] Morgan, F.; Appolonia Nouzille, C.; Baechler, R.; Vuataz, G.; Raemy, A., *Le Lait*, 2005, 85(4–5), 315–323. <https://doi.org/10.1051/lait:2005017>
- [29] Thomas, M. E. C.; Scher, J.; Desobry, S., *Journal of Dairy Science*, 2004, 87(5), 1158–1166. [https://doi.org/10.3168/jds.S0022-0302\(04\)73264-6](https://doi.org/10.3168/jds.S0022-0302(04)73264-6)
- [30] Murphy, B. M.; Prescott, S. W.; Larson, I.; Measurement of lactose crystallinity using Raman spectroscopy, 2005, 38, 186–190. <https://doi.org/10.1016/j.jpba.2004.12.013>
- [31] Listiohadi, Y.; Hourigan, J., *Dairy Science & Technology*, 2009, 89, 43–67. <https://doi.org/10.1051/dst:2008027>
- [32] Nørgaard, L.; Hahn, M. T.; Knudsen, L. B.; Farhat, I. A.; Engelsen, S. B., *International Dairy Journal*, 2005 15(12), 1261–1270. <https://doi.org/10.1016/j.idairyj.2004.12.009>
- [33] Pereira, C. G.; Andrade, J.; Ranquine, T.; Moura, I. N.; Alves da Rocha, R., Furtado, M. A. M.; Bell, M. J. V.; Anjos, V. C., *LWT - Food Science and Technology*, 2018, 97 180-186. <https://doi.org/10.1016/j.lwt.2018.06.050>
- [34] Rodrigues, N. V. S.; Cardoso, E. M.; Andrade, M. V. O.; Donnici, C. L.; Sena, M. M. *Journal of the Brazilian Chemical Society*, 2013 24(3), 507–517. <https://doi.org/10.5935/0103-5053.20130066>
- [35] Palo, M.; Kogermann, K.; Genina, N.; Fors, D.; Peltonen, J.; Heinämäki, J.; Sandler, N., *Journal of Drug Delivery Science and Technology*, 2016, 34, 60–70. <https://doi.org/10.1016/j.jddst.2016.02.007>
- [36] Jerônimo, D. P.; De Souza, R. A.; Da Silva, F. F.; Camargo, G. L.; Miranda, H. L.; Xavier, M.; Sakane, K. K.; Ribeiro, W., *Annals of Biomedical Engineering*, 2012, 40(9), 2069–2077. <https://doi.org/10.1007/s10439-012-0549-9>
- [37] Varmuza, K.; Filzmoser, P., *Introduction to multivariate statistical analysis in chemometrics*. 2016 Boca Raton, FL: Taylor & Francis Group.
- [38] Kędzierska-Matysek, M.; Matwijczuk, A.; Florek, M.; Bałowska, J.; Wolanciuk, A.; Matwijczuk, A.; Chruściel, E.; Walkowiak, R.; Karcz D.; Gładyszewska, B., *Application of FTIR spectroscopy for analysis of the quality of honey*. *BIO Web of Conferences*, 2018, 10 (January), 02008. <https://doi.org/10.1051/bioconf/20181002008>

- [39] Paradkar, M. M.; Irudayaraj, J., *Journal of Food Science*, 2002 67(7), 2507–2511. <https://doi.org/10.1111/j.1365-2621.2002.tb08767.x>
- [40] Schulz, H.; Baranska, M., *Vibrational Spectroscopy*, 2007, 43(1), 13–25. <https://doi.org/10.1016/j.vibspec.2006.06.001>
- [41] Edwards, H. G. M.; Munshi, T.; Anstis, M., *Spectrochimica Acta - Part A: Mol. and Biomol. Spectroscopy*, 2005, 61(7), 1453–1459. <https://doi.org/10.1016/j.saa.2004.10.022>
- [42] Uzun, F.; Sağlam, A.; Güçlü, V., *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 2007, 67(2), 342–349. <https://doi.org/10.1016/j.saa.2006.07.029>
- [43] Kesimli, B.; Topaçli, A.; Topaçli, C., *Journal of Molecular Structure*, 2003, 645(2–3), 199–204. [https://doi.org/10.1016/S0022-2860\(02\)00561-6](https://doi.org/10.1016/S0022-2860(02)00561-6)
- [44] Podstawka, E.; Światłowska, M.; Borowiec, E.; & Proniewicz, L. M., *Journal of Raman Spectroscopy*, 2007, 38(3), 356–363. <https://doi.org/10.1002/jrs.1653>
- [45] Gangopadhyay, D.; Sharma, P.; Singh, S. K.; Singh, P.; Tarcea, N.; Deckert, V.; Jürgen, P.; Singh, R. K., *Chem. Phys. Lett.*, 2015, 618, 225–230. <https://doi.org/10.1016/j.cplett.2014.11.021>