*Quest Journals*
*Journal of Research in Pharmaceutical Science*
*Volume 8 ~ Issue 6 (2022) pp: 15-28*
*ISSN(Online) : 2347-2995*
www.questjournals.org

**Research Paper**

# Chemometric study of a mixture of amino acids by spectrophotometry in a flow system

Lidiane Cristina Nunes[1], Efraim Lázaro Reis[1,*], César Reis[1], Clausius Duque Gonçalves Reis[2], Paulo Henrique Fidêncio[3],Juracir Silva Santos[4]

[1]*Department of Chemistry, Federal University of Viçosa, Viçosa, M. G., Brazil*
[2]*Institute of Exact Sciences, Federal University of Viçosa, Rio Paranaíba, M. G., Brazil*
[3]*Department of Chemistry, Federal University of the Jequitinhonha and Mucuri Valleys, Diamantina, M. G., Brazil*
[4]*Federal Institute of Education, Science and Technology of Bahia, Senhor do Bonfim, BA, Brasil*

***ABSTRACT***
*The purpose of this work was to perform a simultaneous study of a mixture of amino acids using multivariate calibration methods. The methodology is based on the reaction between amino acids and sodium 1,2-naphthoquinone-4-sulfonate (NQS). A series of 32 synthetic mixtures with different concentrations of phenylalanine, histidine, proline, tryptophan and isoleucine was prepared according to a $2^5$ factorial design. The spectra were obtained by reacting in a flow-through system; the solutions were added continuously, employing a peristaltic pump with a fixed flow rate of 0.40 mL $min^{-1}$. Since NQS decomposes in alkaline media, the reaction was developed online by inserting NQS and the buffer solution ($Na_2CO_3$/NaOH) at pH 10. For the reaction development the reaction coil was placed in a thermostated bath and kept at 70 ºC. The final solution was collected in a quartz flow cell and read in a UV-VIS spectrophotometer. Two chemometric techniques, principal component analysis (PCA) and partial least squares regression (PLS) were applied. The multivariate model was developed from spectra recorded between 290 and 590 nm and preprocessed with the first derivative. The prediction errors were less than 6% for all species under study when using the spectra preprocessed with the first derivative. In the analysis of real samples, some inconsistencies were observed, mainly due to the presence of non-modeled substances. Overall, the results showed a good predictive ability of the multivariate models based on UV-VIS spectrophotometry. For the analysis of real samples, however, there is the need to consider the presence of other substances that are not modeled precisely because we do not know their nature.*

***KEYWORDS:*** *Amino Acids, Flow System, Principal Component Analysis, Partial Least Squares Regression, UV spectrophotometry.*

***R**eceived 28 May, 2022; **R**evised 05 June, 2022; **A**ccepted 07 June, 2022 © The author(s) 2022.*
***P**ublished with open access at **www.questjournals.org***

## I. INTRODUCTION

Amino acids are basic structural units of proteins of fundamental importance in nutrition, because they are closely related to the vital processes of cells and, consequently, of the entire organism [1]. Proteins come from exogenous sources or synthesized by the body. When the amino acid cannot be synthesized by the body or is synthesized in insufficient amounts, it is called essential, since it must be obtained from exogenous sources. Lysine, tryptophan, histidine, phenylalanine, leucine, isoleucine, threonine, methionine and valineare considered essential amino acids. The others are synthesized from common products of intermediary metabolism, which is why they are classified as non-essential [2].

The determination of amino acids is usually performed by a high performance liquid chromatography (HPLC) technique, which has good selectivity, but the process may be inadequate to analyze a large number of samples because of the time consumed; the official methods of pre- or post-column derivatization are time consuming and expensive.

Amino acids are not directly detectable by spectroscopic techniques (UV-VIS and fluorimetry); a chemical procedure for derivatization of these is necessary. Amino and carboxylic groups offer possibilities for derivatization, chromatography provides selectivity while derivatization enhances sensitivity [3, 4].

A spectrophotometric method for amino acid determination based on the reaction with sodium 1,2-naphthoquinone-4-sulfonate (NQS) in basic medium was developed with a continuous flow analysis system. The general equation of reaction involved is the derivatization of the amino acids with the NQS as shown in Figure 1.

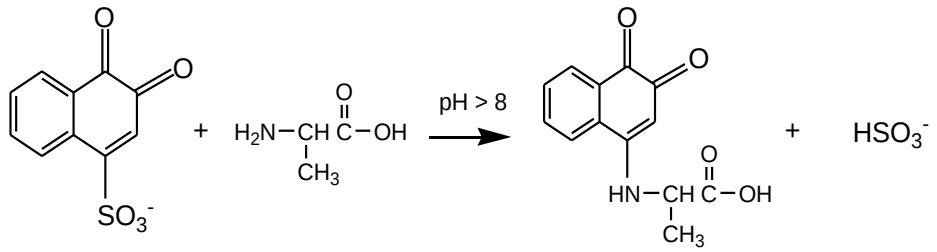



**Figure 1**. General equation of the reaction between NQS and an amino acid.

The sodium 1,2-naphthoquinone-4-sulfonate (NQS) reagent has the advantages of being water-soluble, reacts with primary and secondary amino groups. The main problem with this reagent is instability in alkaline medium. This reagent has been used in the determination of aliphatic amines in alcoholic beverages by liquid-liquid extraction with a flow injection system [5].

Amino acids, in general, are hardly analyzed by molecular spectrophotometry using univariate methods, because their spectra in the UV-VIS region are severely overlapped [6]. Thus, using derivative spectrophotometry associated with multivariate calibration methods, it is possible to determine components in mixture, in addition to detecting anomalous samples, even if having the information that the derivation of the spectra does not increase the information content of the original spectrum, however it allows an interpretation more complete, as it shows more clearly its maximums, minimums and inflection points, as long as it is possible to locate them.

In this sense, the evaluation of multivariate calibration techniques for the spectrophotometric determination of amino acid mixtures based on their reactions with chromophore agents in a flow system constitutes the proposal of this work. Two multivariate statistical techniques were used, principal components analysis (PCA) and the partial least squares method (PLS) which are based on the concentration of total information in a response matrix.

## II. MATERIAL AND METHODS

### Preparation of Reagent and Amino Acid Solutions

In this work, synthetic mixtures of five amino acids were used for the study with the reaction with NQS: phenylalanine (FlukaBiochemika), tryptophan (Sigma Aldrich), proline (FlukaBiochemika), isoleucine (FlukaBiochemika) and histidine (Sigma Aldrich). All amino acid stock solutions were prepared at a concentration of 0.01 mol $L^{-1}$. Samples of protein food supplements were also studied.

A buffer solution of 0.1 mol $L^{-1}$ sodium carbonate / 0.75 mol $L^{-1}$ sodium hydroxide was prepared and the pH adjusted to 10. A stock solution of sodium 1,2-naphthoquinone-4-sulfonate (NQS) was prepared at a concentration of 0.02 mol $L^{-1}$.

### Equipment, Accessories and Reaction Parameters

In the experiments carried out, an Ismatec peristaltic pump, 8-channel ICP model, provided with Tygon® pumping tubes, was used as the propulsion unit. For transmission lines and helical tubular reactors, polyethylene tubes with an internal diameter of 0.35 mm were used.

UV-VIS absorption spectra were obtained using a HITACHI U-2000 spectrophotometer and quartz cuvettes for a 1 cm optical path flow.

All weighingswere performed on a Bel digital analytical balance, model 210A.

The pH measurements were performed using a MS TecnologiapHmeter, model mPA210 with a combined pH electrode.

The bath for thermostatization of the reaction system, maintained at 70 ºC, was obtained with a flask, with the sides thermally insulated, on a magnetic stirrer brand Tecnal, model Te-0853, which has a temperature sensor attached.

The development of the reaction of sodium 1,2-naphthoquinone-4-sulfonate (NQS) with amino acids is highly dependent on variables such as temperature, pH, reagent concentration and flow volume. A preliminary study to assess the effect of these variables was carried out by Saurina et al. (1993).

The pH value was studied in the range of 8 to 12. The maximum absorbance value was reached at pH 10.

From thestock solution of the NQS reagent, its concentration was varied in the range of 5.0 x $10^{-4}$ to 2.5 x $10^{-3}$mol $L^{-1}$. Although the absorbance value increases with increasing concentration value, a value of 0.001 mol $L^{-1}$was selected since the background noise increases significantly with increasing reagent concentration, despite the maximum absorbance not being reached. .

The combination of the length and internal diameter of the reactor and the total flow rate determines the residence time of the sample in the reaction coil [7]. The residence time and temperature are factors of great influence on the absorbance value because of the kinetic reaction between NQS and amino acids. The influence of these parameters on the absorbance was simultaneously studied. The maximum signal was reached with a long residence time (17-20 min) and low temperature (25-30 ºC), but under these conditions the sample processing proceeded very slowly. For low residence times, the absorbance increases considerably with increasing temperature. The final values selected were total flow rate of 1.10 mL $min^{-1}$; reaction coil 10.0 m x 0.35 mm and temperature of 70 ºC [5].

The scheme (Figure 2) was used to obtain the analytical signal based on the reaction of sodium 1,2 naphthoquinone-4-sulfonate (NQS) with amino acids. The solutions were added continuously using a peristaltic pump and the flow rate of all solutions was set at 0.40 mL $min^{-1}$.

Initially, the reagent (NQS) and the buffer solution ($Na_2CO_3$/NaOH) were inserted by confluence, thus generating the alkaline solution by mixing them in a reaction coil (1.00 m x 0.35 mm).

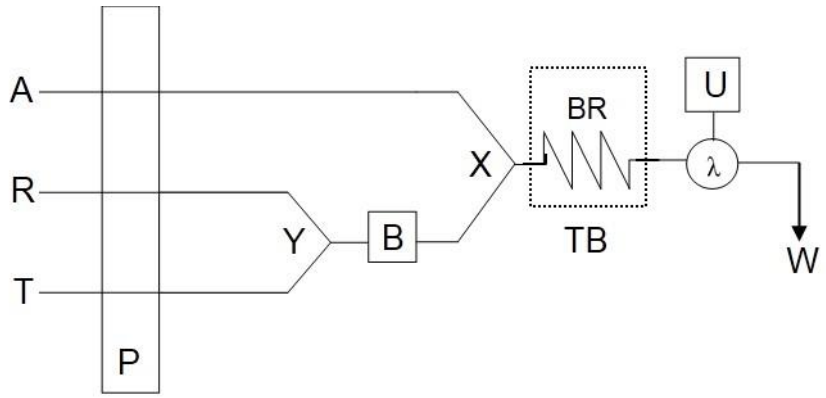



**Figure 2**. Scheme of the System used to obtain the analytical signal based on the reaction of sodium 1,2 naphthoquinone-4-sulfonate (NQS) with amino acids. A: Sample; R: Reagent NQS 1.00 x $10^{-3}$mol $L^{-1}$ in HCl 0.10 mol $L^{-}$1; T: Buffer: 0.10 mol $L^{-1}$ $Na_2CO_3$ / 0.75 mol $L^{-1}$NaOH; P: Peristaltic pump; Y: confluence between R and T; B: Reaction coil (1 m x 0.80 mm); X: Confluence between A and the mixture of R and T; TB: Thermostatized bath; BR: Reaction coil (10 m x 0.8 mm); λ: Spectrophotometric detection; U: Data acquisition; W: Discard.

The reaction was carried out in line considering that the NQS solution decomposes in an alkaline medium. The analytical reaction between the NQS and the amino acids was carried out in a reaction coil (10.0 m x 0.35 mm) submerged in a thermostated bath maintained at 70 ºC. The final solution was collected directly into the quartz flow cuvette for reading on the spectrophotometer.

The spectra were recorded in the range from 290 to 590 nm with a spectral resolution of 1 nm and a sweep speed of 300 nm $min^{-1}$, obtaining 301 absorbance values per spectrum. The spectrophotometer was interfaced to a microcomputer through an RS232C serial port. The spectrophotometer control and data acquisition were done through a program developed in Quick basic [8].

**Development of Calibration Models for the data obtained by UV-VIS Spectrophotometry**

The files with the spectral data stored in ASCII format were arranged in a data matrix that contained the values of the absorbance readings at each wavelength, in the Origin software [9].

For the development of calibration models, concentration ranges of amino acids were chosen from the analytical curves prepared for each amino acid at the wavelength of maximum absorption. Analytical curves were obtained for the ranges from 3.00 to 7.00 x $10^{-4}$mol $L^{-1}$ for phenylalanine; 3.00 to 6.00 x $10^{-4}$mol $L^{-1}$ for

histidine, 1.50 to 7.50 x $10^{-4}$mol $L^{-1}$ for isoleucine; 0.50 to 9.00 x $10^{-4}$mol $L^{-1}$ for proline and 1.00 to 9.00 x $10^{-4}$mol $L^{-1}$ for tryptophan, from stock solutions 0.01 mol $L^{-1}$ of each amino acid.

The development of the set for the multivariate calibration was prepared according to a $2^5$ factorial design, obtaining 32 mixtures with different proportions for the five amino acids.

The spectrophotometric measurements of the 32 synthetic mixtures were submitted to principal component analysis (PCA). For the elaboration of the multivariate model, the spectral data were centered on the mean. This operation was performed on the data matrix of the absorbance measurements.

After the pre-processing step, the data matrix was submitted to Principal Component Analysis. These procedures were performed using a routine of the PLS-ToolBoxchemometric package [10] in the MatLab environment [11].

The data for multivariate calibration correspond to the average of three scans of each of the 32 mixtures according to the experimental design used.

Data processing was performed using the partial least squares (PLS) method contained in the PLS-ToolBoxchemometric package [10]. Data were arranged in two matrices, one with concentration values of the 5 amino acids, in a total of 32 standards and the second with absorbance values in the range 290 to 590 nm in 1 nm intervals. As a result, a concentration matrix (32 lines by 5 columns) and a response matrix (32 lines and 301 columns) were obtained.

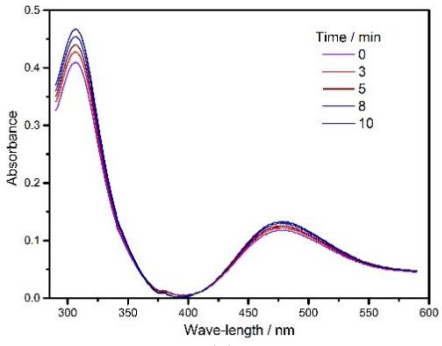
In order to verify an improvement in the efficiency of the multivariate models involving UV-VIS spectrophotometry, some pre-processing procedures of the original data were performed, making the first derivative of the spectra to improve the signal separation and Savitzky-Golay digital smoothing for reduction of spectral noise.

**Analysis of Actual Samples**

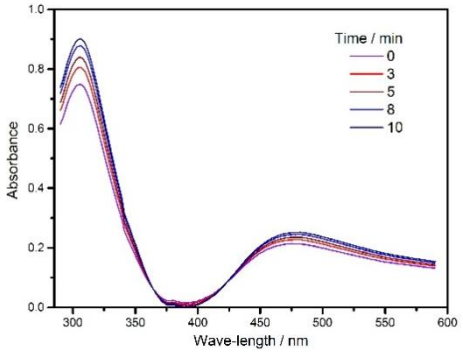
Four samples of protein food supplements, two in powder form and two in liquid form, were subjected to extraction. The equivalent of 3.00 g of solid sample and 3.00 mL of liquid sample were extracted with 50.00 mL of phenol solution (1% v/v) for 30 minutes with magnetic stirring. The extracts obtained were filtered through a 45 μm pore membrane; 5.00 mL of the filtered extract were transferred to a 50.00 mL flask and made up to volume with deionized water. Then 5.00 mL of this solution was diluted to 50.00 mL, completing the volume with deionized water. All assays were performed in triplicate.

## III. RESULTS AND DISCUSSION

The stability of the reaction product between NQS and amino acids was evaluated as a function of time, using solutions with a concentration of 5.00 x $10^{-4}$mol $L^{-1}$ of each amino acid. The spectra were recorded at times 0, 3, 5, 8 and 10 minutes after the sample was collected in a quartz cuvette and can be seen in Figure 3(a-e) and Table 1.




(a)




(b)

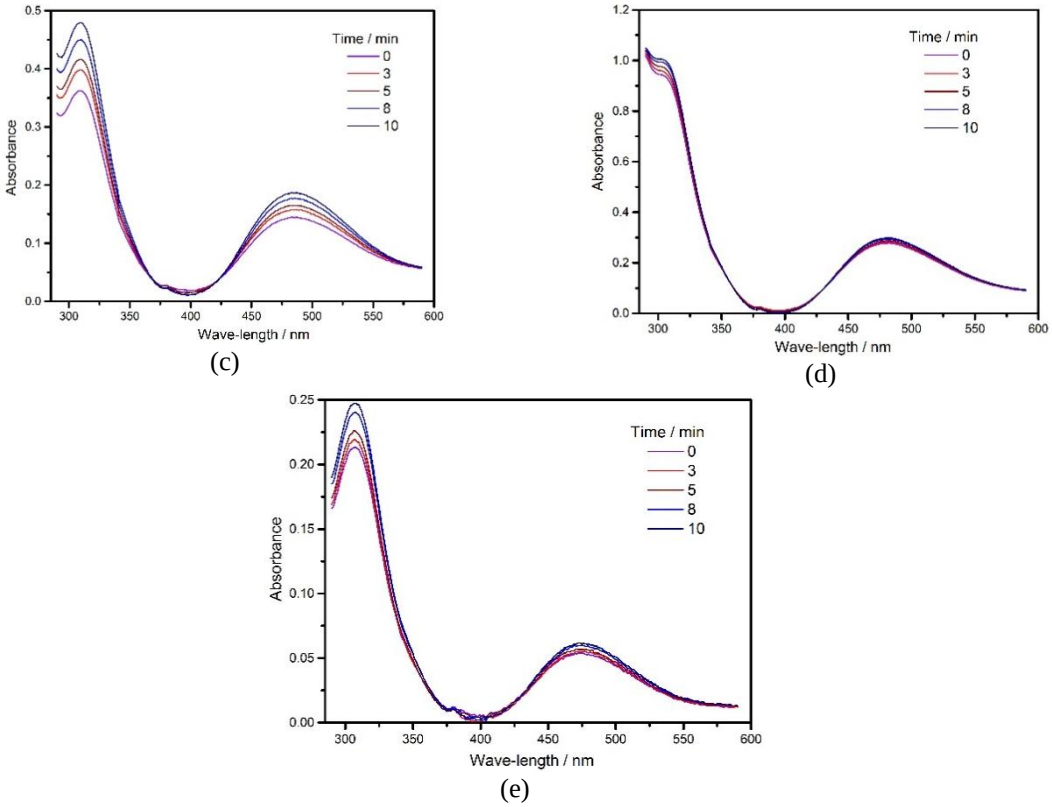



**Figure 3**. Spectra obtained for phenylalanine (a), histidine (b), proline (c), tryptophan (d) and isoleucine (e) solutions at a concentration of 5.00 x $10^{-4}$mol $L^{-1}$, in the range from 0 to 10 min.

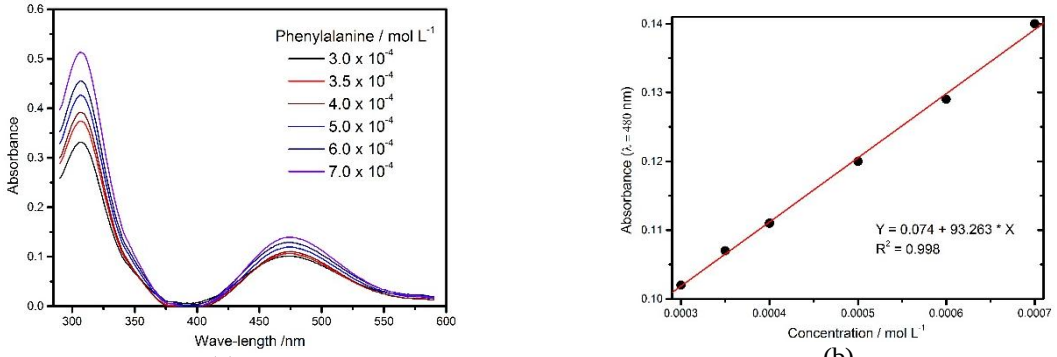
**Table 1**. Variation of absorbance as a function of time for the five amino acids.

| **Aminoacid** | **Absorbance / 480 nm** | | |
|---|---|---|---|
| | **0 minutes** | **10 minutes** | **Δ Absorbance** |
| Phenylalanine | 0.118 | 0.132 | 0.014 |
| Histidine | 0.213 | 0.251 | 0.038 |
| Proline | 0.143 | 0.185 | 0.042 |
| Tryptophan | 0.277 | 0.298 | 0.021 |
| Isoleucine | 0.053 | 0.061 | 0.008 |

By analyzing the graphs of Figure 3(a-e) and Table 1, it is verified that the compound is stable in the studied interval, therefore, the spectral measurements were performed in a discrete way, without loss of information in the time of 10 minutes.

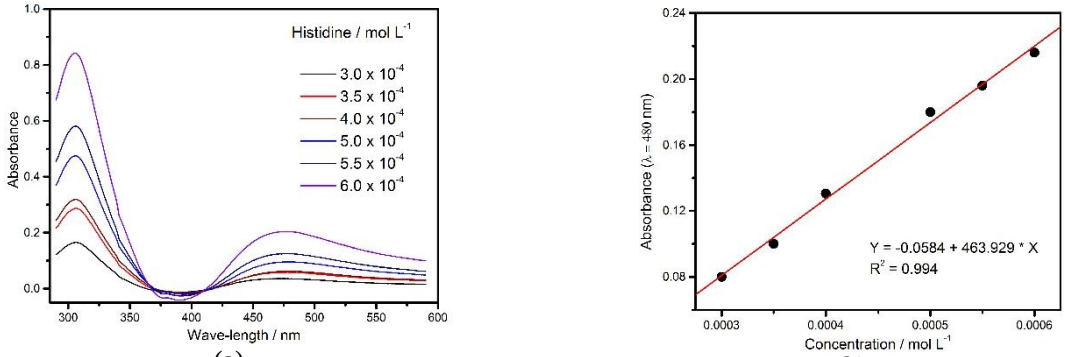
**Analytical Curves of Amino Acids**

For the development of calibration models, a spectrophotometric reading was performed for each amino acid, as can be seen in figures 4(a) to 8(a). From the absorption spectra obtained, the analytical curves were constructed, based on the regions of maximum absorption in the visible region, as can be seen in Figures 4(b) to 8(b).



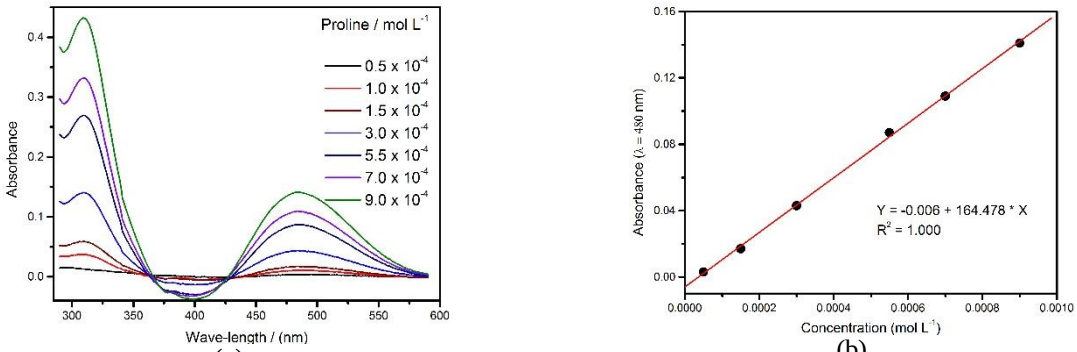

(a) (b)

**Figure 4**. (a) PheylalanineUV-VIS absorption spectra; (b) Analytical curve, absorbance as a function of phenylalanine concentrations.



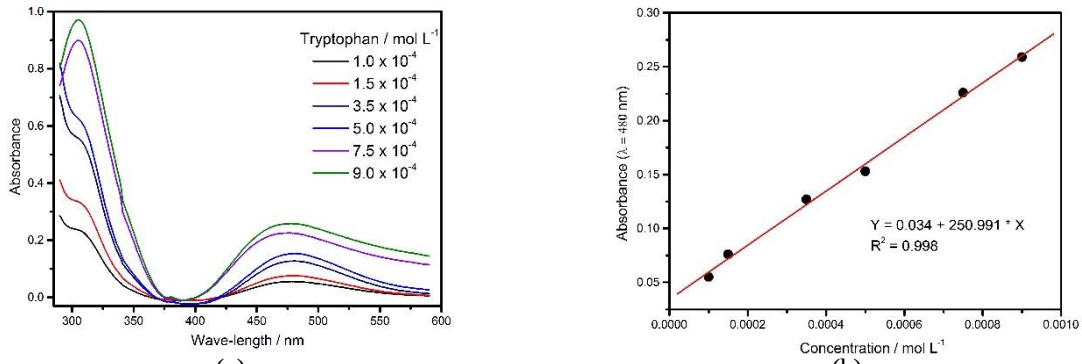

(a) (b)

**Figure 5**. (a) Histidine UV-VIS absorption spectra;(b) Analytical curve, absorbance as a function of histidine concentrations.



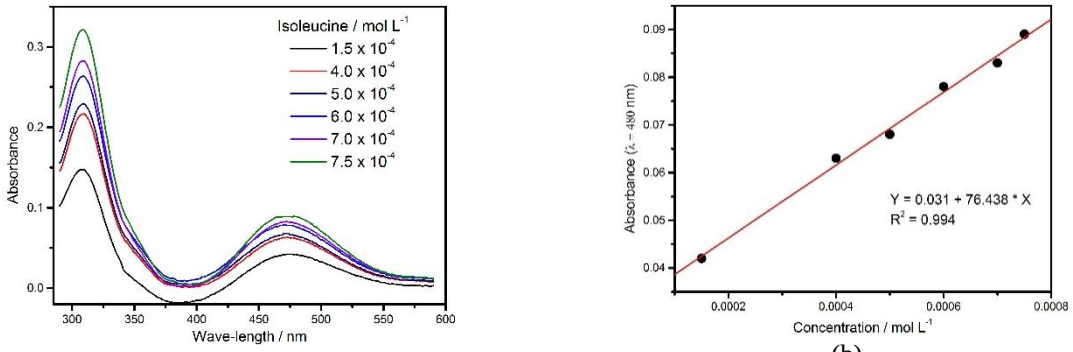

(a) (b)

**Figure6.**(a) ProlineUV-VIS absorption spectra;(b) Analytical curve, absorbance as a function of proline concentrations.




(a) (b)

**Figure7.**(a) Tryptophan UV-VIS absorption spectra;(b) Analytical curve, absorbance as a function of tryptophan concentrations.




(a) (b)

**Figure8.**(a) Isoleucine UV-VIS absorption spectra;(b) Analytical curve, absorbance as a function of isoleucine concentrations.

The analytical curves showed regression coefficients of 0.998 for phenylalanine,0.994 for histidine,0.998 for tryptophan,1.000 for proline and 0.994 for isoleucine.

The two levels of concentration used to carry out the experimental design were selected from the analytical curves, taking the lowest level (-) as the lowest concentration and the highest level (+) as the highest concentration (Table 2).

According to the $2^5$ factorial design, 32 synthetic mixtures were prepared.The concentration levels that were studied and the factorial design with the respective concentration values can be seen in Tables 2 and 3, respectively.

**Table 2.**Concentration values from the analytical calibration curves for levels (-) and (+).

| **Level** | **Concentration (x $10^{-4}$ mol $L^{-1}$)** | | | | |
|---|---|---|---|---|---|
| | **Phenylalanine** | **Histidine** | **Proline** | **Tryptophan** | **Isoleucine** |
| (-) | 3.00 | 3.00 | 0.50 | 1.00 | 1.50 |
| (+) | 7.00 | 6.00 | 9.00 | 9.00 | 7.50 |

**Table 3.**Composition of mixtures used in the calibration and validation stages for the development of multivariate models.

| **Mixtures** | **Concentrations (x $10^{-4}$mol $L^{-1}$)** | | | | |
|---|---|---|---|---|---|
| | **Phenylalanine** | **Histidine** | **Proline** | **Tryptophan** | **Isoleucine** |
| 1 | 7.00 | 6.00 | 9.00 | 9.00 | 7.50 |
| 2 | 7.00 | 6.00 | 9.00 | 9.00 | 1.50 |
| 3 | 7.00 | 6.00 | 9.00 | 1.00 | 7.50 |
| 4 | 7.00 | 6.00 | 9.00 | 1.00 | 1.50 |
| 5 | 7.00 | 6.00 | 0.50 | 9.00 | 7.50 |
| 6 | 7.00 | 6.00 | 0.50 | 9.00 | 1.50 |
| 7 | 7.00 | 6.00 | 0.50 | 1.00 | 7.50 |
| 8 | 7.00 | 6.00 | 0.50 | 1.00 | 1.50 |
| 9 | 7.00 | 3.00 | 9.00 | 9.00 | 7.50 |
| 10 | 7.00 | 3.00 | 9.00 | 9.00 | 1.50 |
| 11 | 7.00 | 3.00 | 9.00 | 1.00 | 7.50 |
| 12 | 7.00 | 3.00 | 9.00 | 1.00 | 1.50 |
| 13 | 7.00 | 3.00 | 0.50 | 9.00 | 7.50 |
| 14 | 7.00 | 3.00 | 0.50 | 9.00 | 1.50 |
| 15 | 7.00 | 3.00 | 0.50 | 1.00 | 7.50 |
| 16 | 7.00 | 3.00 | 0.50 | 1.00 | 1.50 |
| 17 | 3.00 | 6.00 | 9.00 | 9.00 | 7.50 |
| 18 | 3.00 | 6.00 | 9.00 | 9.00 | 1.50 |

| | | | | | |
|---|---|---|---|---|---|
| 19 | 3.00 | 6.00 | 9.00 | 1.00 | 7.50 |
| 20 | 3.00 | 6.00 | 9.00 | 1.00 | 1.50 |
| 21 | 3.00 | 6.00 | 0.50 | 9.00 | 7.50 |
| 22 | 3.00 | 6.00 | 0.50 | 9.00 | 1.50 |
| 23 | 3.00 | 6.00 | 0.50 | 1.00 | 7.50 |
| 24 | 3.00 | 6.00 | 0.50 | 1.00 | 1.50 |
| 25 | 3.00 | 3.00 | 9.00 | 9.00 | 7.50 |
| 26 | 3.00 | 3.00 | 9.00 | 9.00 | 1.50 |
| 27 | 3.00 | 3.00 | 9.00 | 1.00 | 7.50 |
| 28 | 3.00 | 3.00 | 9.00 | 1.00 | 1.50 |
| 29 | 3.00 | 3.00 | 0.50 | 9.00 | 7.50 |
| 30 | 3.00 | 3.00 | 0.50 | 9.00 | 1.50 |
| 31 | 3.00 | 3.00 | 0.50 | 1.00 | 7.50 |
| 32 | 3.00 | 3.00 | 0.50 | 1.00 | 1.50 |

**Multivariate Analysis**

The amino acids phenylalanine, histidine, proline and tryptophan, with the exception of isoleucine, have similar structures as can be seen in Figure 9.

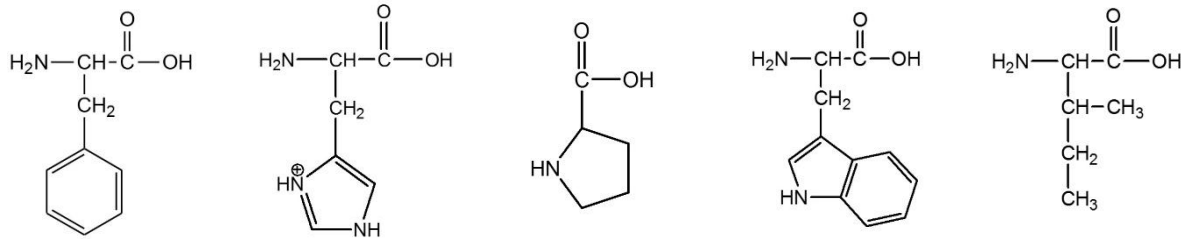



**Figure 9.** Chemical structures (a) phenylalanine, (b) histidine, (c) proline, (d) tryptophan and (e) isoleucine.

The structural similarity of these amino acids makes the individual and mixture spectra very similar.These characteristics make it impossible to differentiate and/or classify the different amino acids just by the simple visual analysis of the spectra (Figure 10a).

In order to circumvent the interference problems, the UV-VIS absorption spectra were processed in the first derivative.Since the derivation increases the signal to noise, a Savitzky-Golay digital smoothing procedure was performed simultaneously (Figure 10b).

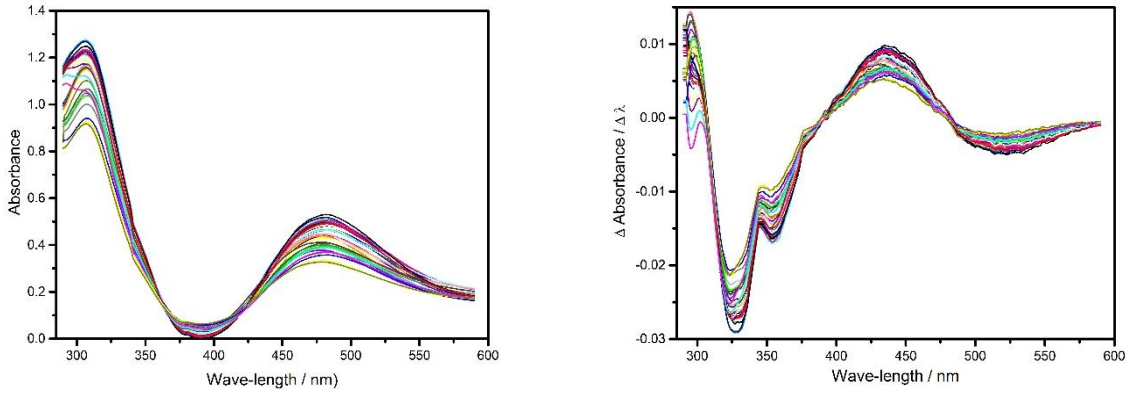



**Figure 10.**UV-VIS absorption spectra of 32 synthetic mixtures of phenylalanine, histidine, proline, tryptophan and isoleucine obtained at pH 10.00 used in the multivariate model, without pre-processing (a) and with pre-processing using the first derivative (b).

Observing the original spectra of the mixture (Figure 10a) and the spectra in the form of the first derivative (Figure 10b) there is an increase in spectral information due to the emergence of new peaks, which can be used in chemometric processing.

**Principal Component Analysis (PCA)**

The observation of the set of spectra (Figure 11) allows us to identify some differences. The real samples, named 34, 35 and 36, present concentration below the levels studied in this work and the real sample 33 presents a spectrum similar to the synthetic mixtures used in the planning, although it presents different characteristics.

The dataset obtained with the spectrophotometric measurements was submitted to chemometric treatment by principal component analysis (PCA), a relevant tool for visualization that allows the graphic representation of the results of the factorial design in different groups.

The model was built with the first derivative of the absorbance data as a function of wavelength using second-order polynomials and digital smoothing Savitzky-Golay type with 5 points. This model allows small spectral differences to be highlighted.

In addition to the spectral dataset being subjected to transformation by the first derivative, a pre-processing procedure for mean-centered data was performed, using the PLS-Toolbox program (EigenVector, 1998) in a MatLab environment (Math Works, 1993).

The values of the derivatives of the absorbances as a function of the wavelength of the set of 36 spectra were arranged in a matrix form and the data were submitted to analysis of the principal components.

This differentiation can be observed from the score chart (Figure 12). This parameter, which represents the coordinate of each sample in the axes system formed by the new principal components (PC1 vs. PC2) indicates that samples 33 to 36 are differentiable from the rest of the set since they are far from all mixtures synthetic materials used in planning, despite having some similar characteristics identified by the analysis of PC2.

In the first instance, it is possible to assume that discrepancies of this type are due to the presence of excipients, species that were not considered in the modeling and introduce spectral changes that make it impossible to adequately predict the species of interest. The analyzed samples present, in addition to the amino acids considered in this study, many others, which may be influencing the results obtained.

Based on the results obtained by the PCA, it appears that the model developed with the PLS cannot be applied to any of the samples. Since these have differentiable characteristics from the set of synthetic mixtures, the predictions will be compromised.

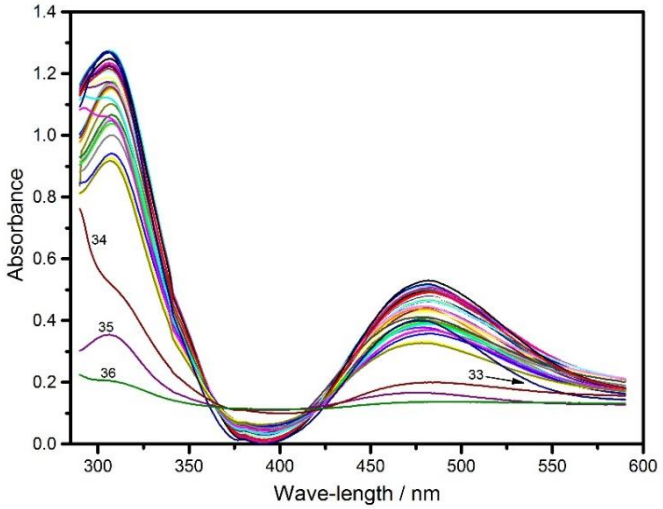



**Figure 11.**Spectra of the 32 solutions of the factorial design and of the 4 solutions of the real samples, used in the analysis of the principal components.

The analysis of PC1 versus PC2 scores shows a separation of the samples by their constituents. The analysis of the principal components made it possible to visualize through a graphic representation of the results of the factorial design in different groups, according to the maximum and minimum values of concentration of the amino acid solutions.

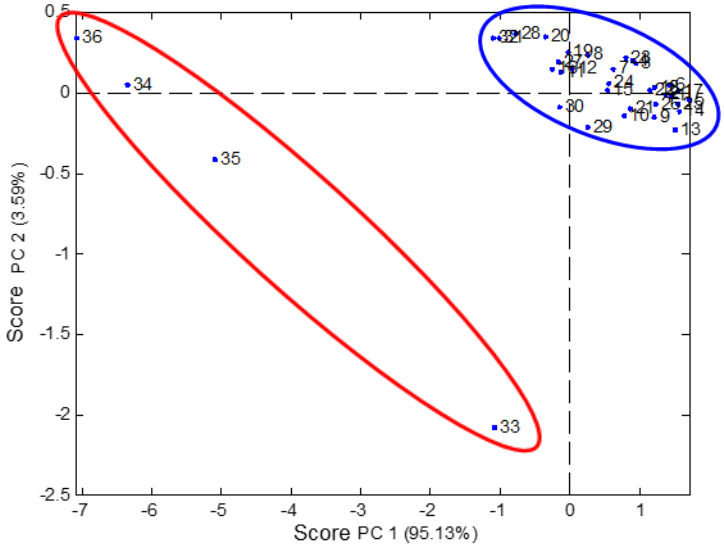



**Figure12.** PC1 versus PC2 scores for the 36 mixtures.

Based on the results obtained by the PCA, it appears that the model developed with the PLS cannot be applied to any of the samples. Since these have differentiable characteristics from the set of synthetic mixtures, the predictions will be compromised.

**Calibration using the Partial Least Squares Method (PLS)**

The absorbance values of the 32 mixtures as well as the concentration values were arranged in a matrix form using the partial least squares method (PLS) constructed with the data corresponding to the absorbance values (independent variables) and the concentration data (variables dependents). The determination was specific for each amino acid. Each sample in the set is represented by 301 absorbance values, recorded in each spectrum.

The best model is always chosen in order to optimize the calibration fit and its predictive capacity simultaneously, so that neither overfit nor underfitting of the selected model occurs.

Two models were built with the PLS, in the first one, the data were not submitted to a pre-processing; in the second model, the data were submitted to Savitzky-Golay digital smoothing and first-order derivative preprocessing.

**Table 4.**Results of prediction of concentrations of synthetic amino acid mixtures, without data processing.

| Sample | Concentrations (x $10^{-4}$mol $L^{-1}$) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Phenylalanine | | | Histidine | | | Proline | | | Tryptophan | | | Isoleucine | | |
| | Actual* | Det* | Error* | Actual | Det | Error | Actual | Det | Error | Actual | Det | Error | Actual | Det | Error |
| **1** | 7.00 | 7.48 | 6.86 | 6.00 | 5.98 | 0.33 | 9.00 | 8.68 | 3.56 | 9.00 | 8.79 | 2.33 | 7.50 | 6.52 | 13.06 |
| **2** | 7.00 | 7.17 | 2.43 | 6.00 | 5.95 | 0.83 | 9.00 | 8.91 | 1.00 | 9.00 | 8.89 | 1.22 | 1.50 | 1.16 | 22.60 |
| **3** | 7.00 | 7.29 | 4.14 | 6.00 | 6.06 | 1.00 | 9.00 | 8.70 | 3.33 | 1.00 | 1.24 | 24.00 | 7.50 | 7.31 | 2.53 |
| **4** | 7.00 | 6.45 | 7.86 | 6.00 | 6.06 | 1.00 | 9.00 | 9.66 | 7.33 | 1.00 | 1.04 | 4.00 | 1.50 | 2.31 | 54.00 |
| **5** | 7.00 | 6.98 | 0.28 | 6.00 | 5.95 | 0.83 | 0.50 | 0.81 | 62.00 | 9.00 | 8.82 | 2.00 | 7.50 | 7.70 | 2.67 |
| **6** | 7.00 | 5.95 | 15.00 | 6.00 | 5.97 | 0.50 | 0.50 | 0.44 | 12.00 | 9.00 | 8.64 | 4.00 | 1.50 | 1.95 | 30.00 |
| **7** | 7.00 | 6.94 | 0.86 | 6.00 | 6.04 | 0.67 | 0.50 | 0.76 | 52.00 | 1.00 | 1.37 | 37.00 | 7.50 | 7.65 | 2.00 |
| **8** | 7.00 | 6.97 | 0.43 | 6.00 | 6.00 | 0.00 | 0.50 | 0.53 | 6.00 | 1.00 | 0.98 | 2.00 | 1.50 | 1.44 | 4.00 |
| **9** | 7.00 | 7.16 | 2.28 | 3.00 | 3.10 | 3.33 | 9.00 | 9.02 | 0.22 | 9.00 | 9.17 | 1.89 | 7.50 | 7.29 | 2.80 |
| **10** | 7.00 | 6.37 | 9.00 | 3.00 | 3.07 | 2.33 | 9.00 | 8.75 | 2.78 | 9.00 | 9.15 | 1.67 | 1.50 | 1.80 | 20.00 |
| **11** | 7.00 | 7.21 | 3.00 | 3.00 | 2.94 | 2.00 | 9.00 | 8.95 | 0.56 | 1.00 | 0.95 | 5.00 | 7.50 | 7.82 | 4.27 |
| **12** | 7.00 | 7.04 | 0.57 | 3.00 | 2.95 | 1.67 | 9.00 | 9.31 | 3.44 | 1.00 | 0.79 | 21.00 | 1.50 | 0.90 | 40.00 |
| **13** | 7.00 | 6.99 | 0.14 | 3.00 | 3.05 | 1.67 | 0.50 | 0.37 | 26.00 | 9.00 | 9.16 | 1.78 | 7.50 | 7.66 | 2.13 |
| **14** | 7.00 | 6.58 | 6.00 | 3.00 | 2.92 | 2.67 | 0.50 | 0.66 | 32.00 | 9.00 | 9.05 | 0.55 | 1.50 | 1.91 | 27.33 |
| **15** | 7.00 | 7.53 | 7.57 | 3.00 | 2.91 | 3.00 | 0.50 | 0.40 | 20.00 | 1.00 | 0.96 | 4.00 | 7.50 | 7.20 | 4.00 |
| **16** | 7.00 | 6.49 | 7.28 | 3.00 | 2.88 | 4.00 | 0.50 | 0.97 | 94.00 | 1.00 | 0.71 | 29.00 | 1.50 | 2.07 | 38.00 |
| **17** | 3.00 | 3.19 | 6.33 | 6.00 | 5.94 | 1.00 | 9.00 | 9.42 | 4.67 | 9.00 | 8.80 | 2.22 | 7.50 | 7.82 | 4.27 |
| **18** | 3.00 | 3.24 | 8.00 | 6.00 | 5.94 | 1.00 | 9.00 | 8.96 | 0.44 | 9.00 | 9.24 | 2.67 | 1.50 | 1.45 | 3.33 |
| **19** | 3.00 | 2.88 | 4.00 | 6.00 | 5.92 | 1.33 | 9.00 | 8.76 | 0.00 | 1.00 | 0.97 | 3.00 | 7.50 | 7.42 | 1.09 |
| **20** | 3.00 | 3.27 | 9.00 | 6.00 | 5.94 | 1.00 | 9.00 | 9.08 | 2.67 | 1.00 | 0.93 | 7.00 | 1.50 | 1.77 | 18.00 |
| **21** | 3.00 | 2.58 | 14.00 | 6.00 | 5.93 | 1.20 | 0.50 | 0.92 | 0.89 | 9.00 | 8.57 | 4.78 | 7.50 | 8.37 | 11.60 |
| **22** | 3.00 | 3.20 | 6.67 | 6.00 | 6.04 | 0.67 | 0.50 | 0.11 | 84.00 | 9.00 | 9.37 | 4.11 | 1.50 | 1.38 | 8.00 |
| **23** | 3.00 | 2.75 | 8.33 | 6.00 | 6.23 | 3.83 | 0.50 | 0.10 | 78.00 | 1.00 | 1.26 | 26.00 | 7.50 | 6.51 | 13.20 |
| **24** | 3.00 | 3.17 | 5.67 | 6.00 | 5.99 | 0.16 | 0.50 | 0.27 | 80.00 | 1.00 | 1.00 | 0.00 | 1.50 | 1.29 | 14.00 |
| **25** | 3.00 | 3.83 | 27.67 | 3.00 | 2.88 | 4.00 | 9.00 | 9.17 | 46.00 | 9.00 | 8.96 | 0.44 | 7.50 | 7.67 | 2.27 |
| **26** | 3.00 | 2.86 | 4.67 | 3.00 | 3.17 | 5.67 | 9.00 | 8.71 | 1.89 | 9.00 | 9.06 | 0.66 | 1.50 | 1.34 | 10.67 |
| **27** | 3.00 | 2.36 | 21.33 | 3.00 | 3.16 | 5.33 | 9.00 | 8.46 | 3.22 | 1.00 | 0.88 | 12.00 | 7.50 | 7.04 | 6.13 |
| **28** | 3.00 | 2.78 | 7.33 | 3.00 | 3.02 | 0.67 | 9.00 | 9.14 | 6.00 | 1.00 | 1.21 | 21.00 | 1.50 | 1.43 | 4.67 |
| **29** | 3.00 | 3.18 | 6.00 | 3.00 | 3.02 | 0.67 | 0.50 | 0.64 | 1.56 | 9.00 | 9.02 | 0.22 | 7.50 | 7.35 | 2.00 |
| **30** | 3.00 | 3.18 | 6.00 | 3.00 | 3.01 | 0.33 | 0.50 | 0.45 | 28.00 | 9.00 | 9.15 | 1.67 | 1.50 | 0.96 | 36.00 |
| **31** | 3.00 | 3.5 | 16.67 | 3.00 | 3.05 | 1.67 | 0.50 | 0.47 | 10.00 | 1.00 | 1.06 | 6.00 | 7.50 | 7.68 | 2.40 |
| **32** | 3.00 | 3.53 | 17.67 | 3.00 | 2.89 | 3.67 | 0.50 | 0.67 | 6.00 | 1.00 | 0.75 | 25.00 | 1.50 | 1.81 | 20.67 |

* Actual: Actual concentration;Det: Determined concentration;Error: Error in percentage

**Tabela 5.**Results of prediction of concentrations of synthetic amino acid mixtures, with data processed with first derivative

| Sample | Concentrations (x $10^{-4}$mol $L^{-1}$) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Phenylalanine | | | Histidine | | | Proline | | | Tryptophan | | | Isoleucine | | |
| | Actual* | Det* | Error* | Actual | Det | Error | Actual | Det | Error | Actual | Det | Error | Actual | Det | Error |
| **1** | 7.00 | 6.99 | 0.14 | 6.00 | 6.00 | 0.00 | 9.00 | 9.01 | 0.11 | 9.00 | 9.01 | 0.11 | 7.50 | 7.49 | 0.13 |
| **2** | 7.00 | 7.02 | 0.28 | 6.00 | 6.01 | 0.16 | 9.00 | 9.01 | 0.11 | 9.00 | 9.00 | 0.00 | 1.50 | 1.58 | 5.33 |
| **3** | 7.00 | 6.99 | 0.14 | 6.00 | 5.98 | 0.33 | 9.00 | 9.01 | 0.11 | 1.00 | 1.00 | 0.00 | 7.50 | 7.41 | 1.20 |
| **4** | 7.00 | 7.01 | 0.14 | 6.00 | 6.01 | 0.16 | 9.00 | 8.99 | 0.11 | 1.00 | 1.00 | 0.00 | 1.50 | 1.57 | 4.67 |
| **5** | 7.00 | 7.02 | 0.28 | 6.00 | 6.01 | 0.16 | 0.50 | 0.53 | 6.00 | 9.00 | 8.99 | 0.11 | 7.50 | 7.60 | 1.33 |
| **6** | 7.00 | 7.01 | 0.14 | 6.00 | 6.01 | 0.16 | 0.50 | 0.49 | 2.00 | 9.00 | 8.99 | 0.11 | 1.50 | 1.57 | 4.67 |
| **7** | 7.00 | 6.99 | 0.14 | 6.00 | 5.99 | 0.17 | 0.50 | 0.51 | 2.00 | 1.00 | 0.99 | 1.00 | 7.50 | 7.42 | 1.07 |
| **8** | 7.00 | 7.00 | 0.00 | 6.00 | 6.00 | 0.00 | 0.50 | 0.50 | 0.00 | 1.00 | 1.00 | 0.00 | 1.50 | 1.50 | 0.00 |
| **9** | 7.00 | 7.00 | 0.00 | 3.00 | 2.99 | 0.33 | 9.00 | 9.00 | 0.00 | 9.00 | 9.02 | 0.22 | 7.50 | 7.51 | 0.13 |
| **10** | 7.00 | 7.01 | 0.14 | 3.00 | 3.01 | 0.33 | 9.00 | 8.99 | 0.11 | 9.00 | 9.00 | 0.00 | 1.50 | 1.46 | 2.66 |
| **11** | 7.00 | 6.99 | 0.14 | 3.00 | 3.00 | 0.00 | 9.00 | 8.98 | 0.22 | 1.00 | 0.99 | 1.00 | 7.50 | 7.60 | 1.33 |
| **12** | 7.00 | 6.98 | 0.28 | 3.00 | 3.00 | 0.00 | 9.00 | 9.02 | 0.22 | 1.00 | 1.00 | 0.00 | 1.50 | 1.48 | 1.33 |
| **13** | 7.00 | 7.00 | 0.00 | 3.00 | 3.00 | 0.00 | 0.50 | 0.52 | 4.00 | 9.00 | 9.00 | 0.00 | 7.50 | 7.54 | 0.53 |
| **14** | 7.00 | 6.99 | 0.14 | 3.00 | 2.99 | 0.33 | 0.50 | 0.51 | 2.00 | 9.00 | 8.98 | 0.22 | 1.50 | 1.46 | 2.67 |
| **15** | 7.00 | 7.00 | 0.00 | 3.00 | 2.98 | 0.67 | 0.50 | 0.50 | 0.00 | 1.00 | 1.01 | 1.00 | 7.50 | 7.52 | 0.26 |
| **16** | 7.00 | 6.99 | 0.14 | 3.00 | 3.00 | 0.00 | 0.50 | 0.50 | 0.00 | 1.00 | 1.01 | 1.00 | 1.50 | 1.47 | 2.00 |
| **17** | 3.00 | 2.99 | 0.33 | 6.00 | 5.99 | 0.17 | 9.00 | 9.01 | 0.11 | 9.00 | 9.01 | 0.11 | 7.50 | 7.49 | 0.13 |
| **18** | 3.00 | 3.00 | 0.00 | 6.00 | 5.99 | 0.17 | 9.00 | 8.97 | 0.33 | 9.00 | 9.01 | 0.11 | 1.50 | 1.48 | 1.33 |
| **19** | 3.00 | 2.97 | 1.00 | 6.00 | 6.00 | 0.00 | 9.00 | 8.98 | 0.22 | 1.00 | 0.99 | 1.00 | 7.50 | 7.45 | 0.67 |
| **20** | 3.00 | 3.00 | 0.00 | 6.00 | 6.00 | 0.00 | 9.00 | 8.97 | 0.33 | 1.00 | 1.00 | 0.00 | 1.50 | 1.48 | 1.33 |
| **21** | 3.00 | 2.99 | 0.33 | 6.00 | 6.00 | 0.00 | 0.50 | 0.49 | 2.00 | 9.00 | 9.00 | 0.00 | 7.50 | 7.5 | 0.00 |
| **22** | 3.00 | 3.00 | 0.00 | 6.00 | 6.00 | 0.00 | 0.50 | 0.50 | 0.00 | 9.00 | 9.00 | 0.00 | 1.50 | 1.50 | 0.00 |
| **23** | 3.00 | 2.99 | 0.33 | 6.00 | 6.01 | 0.16 | 0.50 | 0.51 | 2.00 | 1.00 | 0.97 | 3.00 | 7.50 | 7.51 | 0.93 |
| **24** | 3.00 | 3.02 | 0.67 | 6.00 | 6.00 | 0.00 | 0.50 | 0.49 | 2.00 | 1.00 | 1.02 | 2.00 | 1.50 | 1.49 | 0.67 |
| **25** | 3.00 | 2.99 | 0.33 | 3.00 | 3.01 | 0.33 | 9.00 | 9.03 | 0.33 | 9.00 | 9.01 | 0.11 | 7.50 | 7.56 | 0.80 |
| **26** | 3.00 | 3.02 | 0.67 | 3.00 | 3.00 | 0.00 | 9.00 | 9.00 | 0.00 | 9.00 | 8.98 | 0.22 | 1.50 | 1.48 | 1.33 |
| **27** | 3.00 | 3.02 | 0.67 | 3.00 | 3.02 | 0.67 | 9.00 | 9.00 | 0.00 | 1.00 | 1.00 | 0.00 | 7.50 | 7.56 | 0.80 |
| **28** | 3.00 | 3.01 | 0.33 | 3.00 | 3.00 | 0.00 | 9.00 | 9.03 | 0.33 | 1.00 | 0.98 | 2.00 | 1.50 | 1.54 | 2.67 |
| **29** | 3.00 | 3.02 | 0.67 | 3.00 | 3.00 | 0.00 | 0.50 | 0.50 | 0.00 | 9.00 | 8.98 | 0.22 | 7.50 | 7.48 | 0.26 |
| **30** | 3.00 | 2.97 | 1.00 | 3.00 | 3.00 | 0.00 | 0.50 | 0.51 | 2.00 | 9.00 | 9.01 | 0.11 | 1.50 | 1.49 | 0.67 |
| **31** | 3.00 | 3.00 | 0.00 | 3.00 | 2.9 | 0.33 | 0.50 | 0.5 | 0.00 | 1.00 | 1.0 | 1.00 | 7.50 | 7.4 | 0.26 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 9 | | | 0 | | | 1 | | | 8 | |
| **32** | 3.00 | 3.02 | 0.67 | 3.00 | 3.00 | 0.00 | 0.50 | 0.52 | 4.00 | 1.00 | 1.00 | 0.00 | 1.50 | 1.48 | 1.33 |

* Actual: Actual concentration;Det: Determined concentration;Error: Error in percentage

The internal cross-validation method was used. In this procedure, a spectrum is taken from the calibration set and used as a prediction procedure. This operation is repeated as many times as necessary to ensure that all elements of the set participate as a predictor. The concentration of the sample that was taken is predicted and by comparison with the expected value, the errors generated by cross-validation are estimated for each factor studied. This procedure is repeated for all samples and an average error is calculated.

The predictive capacity of the model developed with the first derivative can be evaluated from the results presented in Table 6. It is observed that this model allows predictions consistent with the known values, with relative errors below 6.0%.

**Table 6.**Mean of prediction errors (%) for the studied amino acids.

| Aminoacids | Model 1 | Model 2 |
|---|---|---|
| Phenylalanine | 7.595 | 0.284 |
| Histidine | 1.813 | 0.145 |
| Proline | 21.24 | 0.957 |
| Tryptophan | 8.069 | 0.458 |
| Isoleucine | 13.36 | 1.370 |

A comparison between the means of prediction errors for each factor is shown in Table 6, it can be seen that the means of prediction errors for each amino acid studied are in the range of 1.81% and 21.24% for the data without pre-processing (model 1) and below 1.37% for data with pre-processing (model 2) showing the best results.

## IV. CONCLUSION

It was possible to demonstrate by the work carried out that the use of spectrophotometric methods in conjunction with chemometric methods can result in analytical applications of great potential, both for qualitative and quantitative determinations.

For the spectrophotometric determination of amino acids, by their reaction with NQS, two multivariate statistical techniques were used, principal component analysis (PCA) and the partial least squares method (PLS). The intense overlap of the spectral bands of the amino acid mixtures makes it impossible for conventional spectrophotometric analysis techniques to be performed. The use of multivariate techniques allowed to circumvent this problem, providing good results.

For the analysis of the principal components, the model with the derived data was applied, observing the separation in two groups, one with the 32 solutions of the factorial design and the other with the 4 solutions of the real samples. Based on what was observed in the analysis of the principal components, it was found that the model developed with partial least squares cannot be applied to any of the 4 real samples, since they present differentiable characteristics from the set of 32 synthetic mixtures of the factorial design.

For the partial least squares method, using cross-validation in the 32 samples of the factorial design, the results of the quantification of each amino acid for the data without pre-processing were not satisfactory, with prediction errors above 90% for proline. On the other hand, for the data processed with the first derivative, results were obtained, with forecast errors below 6%.

## ACKNOWLEDGEMENTS

The authors thank the National Council of Technological and Scientific Development (CNPq) for the master's scholarship granted to the student L. C. Nunes and the Research Support Foundation of the State of Minas Gerais(FAPEMIG) for the financial support to the research Project.

## REFERENCES

[1]. SIMAS, R. C. **Determinação de proteína bruta e aminoácidos em farelo de soja por espectrofotometria no infravermelho próximo**. 2005. 119 f. Dissertação (Mestrado em Química) – UNICAMP, Campinas, 2005.

[2]. TANNUS, A. F. S. **Nitrogênio urinário e tecidual em ratos desnutridos com suplementação de glutamina**. 2001. 96 f. Dissertação (Mestrado em Biociências Aplicadas à Clínica Médica) - Universidade de São Paulo. Faculdade de Medicina de Ribeirão Preto, SP, 2001.

[3]. Xu, W.; Zhong, C.; Zou, C.; Wang, B.; Zhang, N. **Amino Acids**, 52, 1071–1088, 2020.

[4]. Sharer, J.D.; De Biase, I.; Matern, D.; Young, S.; Bennett, M. J.; Tolun, A. A. **Genetics in Medicine**,20, 1499–1507, 2018.
[5]. SAURINA, J. HERNÁNDEZ-CASSOU, S. **AnalyticaChimica Acta**, 1993, 283, 414.
[6]. SCARMINIO, I. S.; ISHIKAWA, D. N.; BARRETO, W. J.; PACZKOWSKI, É. L.; ARRUDA, I. C.**Química Nova**, v. 21, n. 5, p.590-596, 1998.
[7]. PALGROSSI, F. S.; PEDROTTI, J. J.; GUTZ, I. G. R. **Química Nova,** v. 24, n. 5, p. 689-692, 2001.
[8]. QUICKBASIC 4.5. Disponível em: https://www.qbasic.net. Acesso em 15/02/2021.
[9]. MICROCAL, **Origin 9.0**, Microcal, Northampton, MA, 2010.
[10]. EIGENVECTOR RESEARCH. (1998). **PLS_ToolBox**. versão 2.0.
[11]. MATH WORKS, **Matlab**, MathWorks, Natick, MA, 1993.