

# Determination of amylose content in starch using Raman spectroscopy and multivariate calibration analysis

Mariana R. Almeida · Rafael S. Alves ·  
Laura B. L. R. Nascimbem · Rodrigo Stephani ·  
Ronei J. Poppi · Luiz Fernando C. de Oliveira

Received: 2 December 2009 / Revised: 10 February 2010 / Accepted: 10 February 2010 / Published online: 6 March 2010  
© Springer-Verlag 2010

**Abstract** Fourier transform Raman spectroscopy and chemometric tools have been used for exploratory analysis of pure corn and cassava starch samples and mixtures of both starches, as well as for the quantification of amylose content in corn and cassava starch samples. The exploratory analysis using principal component analysis shows that two natural groups of similar samples can be obtained, according to the amylose content, and consequently the botanical origins. The Raman band at  $480\text{ cm}^{-1}$ , assigned to the ring vibration of starches, has the major contribution to the separation of the corn and cassava starch samples. This region was used as a marker to identify the presence of starch in different samples, as well as to characterize amylose and amylopectin. Two calibration models were developed based on partial least squares regression involving pure corn and cassava, and a third model with both starch samples was also built; the results were compared with the results of the standard colorimetric method. The

samples were separated into two groups of calibration and validation by employing the Kennard-Stone algorithm and the optimum number of latent variables was chosen by the root mean square error of cross-validation obtained from the calibration set by internal validation (leave one out). The performance of each model was evaluated by the root mean square errors of calibration and prediction, and the results obtained indicate that Fourier transform Raman spectroscopy can be used for rapid determination of apparent amylose in starch samples with prediction errors similar to those of the standard method.

**Keywords** Corn starch · Cassava starch · Fourier transform Raman · Principal component analysis · Partial least squares

## Introduction

Starch is a biopolymer with a very complex structure, formed by glycosidic linkages between glucose units, and presenting functional properties which make the polymer very helpful for the paper industry, the textile industry, and mainly the food industry. The ratio between amylose and amylopectin is fundamental in determining the functional properties of different starches, and this ratio depends on the botanical source, from cereals such as corn, wheat, and rice, or from root plants such as potato or tapioca [1]. This polysaccharide exists as grains whose format and size depend on the botanical source, and is basically composed of two macromolecules: amylose and amylopectin. Amylose is a molecule which is essentially linear, formed by D-glucose residues linked by  $\alpha(1-4)$  bonds in a helicoidal structure; inside the helix it contains hydrogen atoms, which characterizes amylose as a hydrophobic species. This hydrophobic characteristic is very important in the biolo-

**Electronic supplementary material** The online version of this article (doi:10.1007/s00216-010-3566-2) contains supplementary material, which is available to authorized users.

M. R. Almeida · R. S. Alves · L. F. C. de Oliveira (✉)  
Núcleo de Espectroscopia e Estrutura Molecular (NEEM),  
Departamento de Química, Universidade Federal de Juiz de Fora,  
36036-330 Juiz de Fora, MG, Brazil  
e-mail: luiz.oliveira@ufjf.edu.br

L. B. L. R. Nascimbem · R. J. Poppi  
Laboratório de Quimiometria em Química Analítica (LAQQA),  
Instituto de Química, UNICAMP,  
cx. Postal 6154,  
13083-970 Campinas, SP, Brazil

R. Stephani  
Gemacom Comércio e Serviços LTDA,  
36092-050 Juiz de Fora, MG, Brazil

gical medium, and allows the complexation with free fatty acids, and also with alcohols or iodine. On the other hand, amylopectin is a highly branched polymer formed by D-glucose units linked in an  $\alpha(1-4)$  fashion, and containing only 5–6%  $\alpha(1-6)$  bonds in the structure [2].

The food industry is the greatest consumer of starch, where it is used mainly because of the viscosity, gelation power, adhesion, and retrogradation trend, among other properties, which are influenced mainly by the amylose to amylopectin ratio [3]. Amylose is one of the components responsible for the grain structure, and its quantification is very important to understand starch behavior. The amylose content is commonly measured by methods which involve iodine reaction, such as potentiometric, amperometric, and spectrophotometric methods, all of them involving the starch-iodine complex [4].

Most of the physical-chemical characteristics and properties of starch can be explained by the amylose content; this fact supports the need for analytical techniques which are fast and efficient for use in routine analysis for industry, reducing time and reagent consumption. The multivariate methods associated with vibrational spectroscopic techniques, such as IR spectroscopy and Raman spectroscopy, have been shown to be very good tools in the quality control analysis of the food industry, since the methods are efficient and have good credibility on the basis of the results [5, 6].

Raman spectroscopy offers some advantages for the characterization of food products when compared with IR spectroscopy. For instance, it is not destructive, does not require sample preparation, which takes time and consumes reagents, can determine more than one component at the same time, and water does not interfere in the analysis. The use of Raman spectroscopy has been described for the successful characterization of several natural products, such as cacao seeds [7], guarana extracts [8], and honey [9]. The use of Raman spectroscopy together chemometric analysis has also been applied to the determination of several products, such as the nutritional parameters of milk and children's foods [10], the degree of unsaturation of vegetal oils [11], and the adulteration of honey samples [12].

The literature shows the application of spectroscopic techniques used in the analysis of starch. Fechner et al. [13] investigated the retrogradation of untreated wild-type starches (potato, corn, and wheat) by employing Raman spectroscopy; Zhabankov et al. [14] characterized some medical polysaccharides, including amylose and amylopectin, by Fourier transform (FT) IR (FTIR) and FT-Raman spectroscopies, whereas Schuster et al. [15] presented a novel approach to monitor starch hydrolysis by FT-Raman spectroscopy. Spectroscopic techniques and chemometric methods have also been applied to the study of starch. Kizil et al. [16] applied FTIR and FT-Raman methods, as well as

discriminant analysis, for the rapid characterization and classification of selected irradiated starch samples. Dupuy et al. [17] described the use of Raman spectroscopy to identify modified starches regarding their origin and type of modification, by using principal component analysis (PCA). Sohn et al. [18] employed FT-Raman and near-IR reflectance spectroscopy to compare calibration models for determination of rice cooking quality parameters such as apparent amylose and protein. It is important to note that none of these previous investigations were made using the analysis of amylose content in corn and cassava native starches.

The use of chemometric tools has shown several possibilities of analysis for the processing and interpretation of the analytical data which was not possible before for complex samples, mainly based on the exploratory analysis and development of regression models, also known as multivariate calibration. The exploratory analysis based on the PCA allows the transformation of complex analytical data to obtain an easier visualization of the more important and relevant information. The main purpose of exploratory analysis is to learn from the data about the relationships between variables and samples. The multivariate analysis involves the simultaneous determination of several different parameters to quantify some other variable which is interesting for the characterization of the system, offering the possibility of data analysis of spectra containing superposed signals and the simultaneous determination of several constituents. Partial least squares (PLS) regression is the most used method in the literature for the development of multivariate calibration models [19], based on the correlation of the spectroscopic information with the investigated property, in our case the analyte concentration.

In this work, we present the vibrational characterization of native starch samples of different botanical sources (corn and cassava) and the exploratory analysis using PCA for these samples, as well as the development of multivariate calibration models to quantify the amylose content in starch samples by employing the PLS algorithm. The main purpose of this investigation is to demonstrate that Raman spectroscopy can be used for the quality control of native starches by the food industry.

## Experimental

### Samples

Samples of corn and cassava powder native starches were provided by Gemacom Comércio e Serviços (Juiz de Fora, MG, Brazil). For the analysis by Raman spectroscopy, the samples underwent no pretreatment and were used as received. The mixtures of corn and cassava starches were

prepared by physical mixing in proportions ranging from 10 to 50% w/w. For the exploratory analysis, 34 samples of corn starch, 17 samples of cassava starch, and 13 samples of the mixtures were used. For the spectrophotometric amylose determination in starch samples, amylose and amylopectin (purchased from Sigma) were used as calibration standards in ratios of 0, 10, 20, 30, 40, and 50% w/w. Twenty-three samples of corn and 17 of cassava starches were used for the amylose content determination by the colorimetric method. All these samples were used for the construction of the PLS models for determination of amylose content, employing the Raman spectra data (Raman intensity).

#### Determination of amylose content

The amylose contents (% w/w) from 40 samples of native starches were determined by the spectrophotometric standard method (ISO 6647) [20]. A total of 100 mg of starch granules was homogenized with 1 ml of 95% ethanol and 9 mL of 1 M NaOH. The sample was heated for 10 min in a boiling-water bath to gelatinize the starch. After cooling, it was transferred into a volumetric flask and the volume was made up to 100 mL with water. Then 1 mL of 1 M acetic acid and 2 mL of iodine solution (0.2% I<sub>2</sub>, 2% KI) were added to a 5-mL aliquot. The solution was made up to 100 mL with water and allowed to stand for 10 min. Spectrophotometric quantification was performed at 620 nm, with a Shimadzu model UVPC 1601 spectrophotometer, using the multipoint working curve method with two repetitions and quartz cells of 10-mm path length. Two determinations were made on separate test portions taken from the same sample in each of the replicates. The apparent amylose content was calculated using an equation obtained from the standard curve using purified amylose and amylopectin extracted from potato tubers. The values of amylose obtained have to match the dependent variables used later for the construction and validation of the model calibration.

#### Raman instrumentation

Raman spectra of native starch samples were obtained with a Bruker RFS 100 FT-Raman spectrometer equipped with a germanium detector using liquid nitrogen as the coolant, and with 1,064-nm excitation from a Nd:YAG laser. A few milligrams of the sample were placed into a small aluminum sample cup, the laser light with a power of 100 mW was introduced and focused on the sample, and the scattered radiation was collected at 180°. For each spectrum an average of 1,024 scans were performed at a resolution of 4 cm<sup>-1</sup>, over the 3,500–400-cm<sup>-1</sup> range. A three-term Blackman-Harris apodization function and a

zero filling factor of 2 were used. The OPUS 6.0 (Bruker Optik, Ettlingen, Germany) software program was used for Raman data acquisition. All the spectra were obtained in duplicate to avoid any doubt about the intensity or the wavenumber of the vibrational bands observed in the spectra. The software package GRAMS/AI 7.02 (Thermo Galactic, USA) was used for import/export data of Bruker OPUS File Format Convert Driver for ASCII-XY. The intensities for each wavenumber obtained from the Raman spectra correspond to the independent variables used for the construction and validation of models.

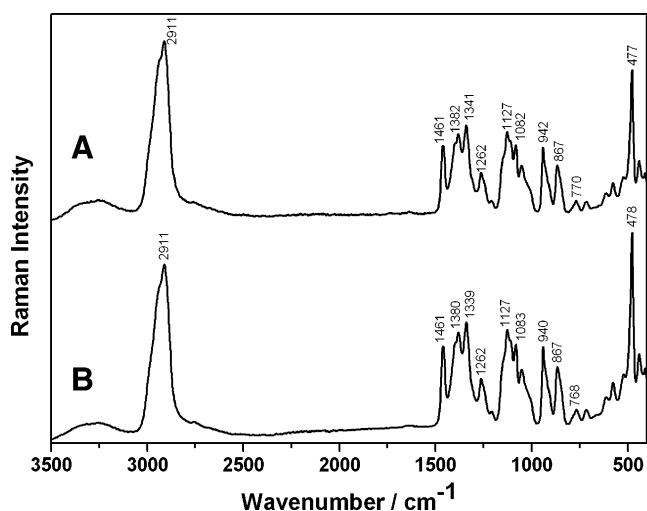
#### Chemometric analysis

The Raman spectra were obtained manipulated using a MATLAB 6.5 environment. For all PCA and PLS analyses the data were preprocessed using the mean center; the smoothed second-order derivative (Savitzky-Golay algorithm with 15 points and a second-order polynomial function) [21] was used to minimize problems due to baseline shifts of different starch samples. Other preprocessing data treatments were tested, but the second-order derivative was the method which best described the set, giving the best results in the PCA and PLS analyses. For the calibration models using PLS, the samples were separated into two groups by employing the Kennard-Stone algorithm [22], and the optimum number of latent variables was chosen by the root mean square error of cross-validation (RMSECV) obtained from the calibration set by internal validation (leave one out). Anomalous samples were discarded by the leverage and residue analysis; the occurrence of such samples among the calibration samples can produce models with a low capability of prevision and, when present in the validation samples, can influence the validation results, leading to results which are indicative that the model is not adequate. The performance of the models was evaluated by the root mean square error of calibration (RMSEC) and the root mean square error of prediction (RMSEP).

## Results and discussion

#### Vibrational spectra of the corn and cassava starch samples

Figure 1 shows typical Raman spectra for the starch samples of corn and cassava; the main vibrational bands are listed in Table 1 together with their respective tentative assignments, based on comparison with literature data [23–26]. The region at 2,900 cm<sup>-1</sup> is related to the symmetrical and antisymmetrical CH stretching; despite the intensity of such bands, the use of them is not so straightforward since most of organic compounds present such bands. There are



**Fig. 1** Fourier transform (FT) Raman spectra of corn starch (A) and cassava starch (B) samples excited at 1,064 nm

only small differences in the spectral range between 2,800 and 3,000  $\text{cm}^{-1}$  observed in the Raman spectra of corn and cassava starches; according to Kizil et al. [16], the intensity changes in this range can be mainly attributed to the variations in the amount of amylose and amylopectin

**Table 1** Raman wavenumbers and their respective tentative assignments based on literature data [23–26]

Wavenumber ( $\text{cm}^{-1}$ )	Assignment
2,911 vs	$\nu(\text{C-H})$
1,461 s	$\delta(\text{CH})+\delta(\text{CH}_2)+\delta(\text{C-O-H})$
1,396 s	$\delta(\text{C-C-H})$
1,380 s	$\delta(\text{C-O-H})$
1,340 s	$\nu(\text{C-O}); \delta(\text{C-O-H})$
1,263 m	$\delta(\text{C-C-H})+\delta(\text{O-C-H})+\delta(\text{C-O-H})$
1,207 w	$\nu(\text{C-C})+\nu(\text{C-O})$
1,127 s	$\nu(\text{C-O})+\nu(\text{C-C})+\delta(\text{C-O-H})$
1,109 s	$\nu(\text{C-C})+\nu(\text{C-O})+\delta(\text{C-O-H})$
1,083 s	$\nu(\text{C-O})+\nu(\text{C-C})+\delta(\text{C-O-H})$
1,053 s	$\nu(\text{C-O})+\nu(\text{C-C})+\delta(\text{C-O-H})$
941 s	$\delta(\text{C-O-C})+\delta(\text{C-O-H})+\nu(\text{C-O})$
868 s	$\delta(\text{C-C-H})+\delta(\text{C-O-C})$
769 w	$\delta(\text{C-C-O})$
718 w	$\delta(\text{C-C-O})$
615 w	$\delta(\text{C-C-O})$
577 w	$\delta(\text{C-C-O})+\tau(\text{C-O})$
524 w	$\delta(\text{C-C-O})+\tau(\text{C-O})$
478 vs	$\delta(\text{C-C-C})+\tau(\text{C-O})$
441 w	$\delta(\text{C-C-O})+\delta(\text{C-C-C})$
410 w	$\delta(\text{C-C-O})+\delta(\text{C-C-C})$

vs very strong, s strong, m medium, w weak

present in starches. On the other hand, the region between 1,500 and 1,200  $\text{cm}^{-1}$  is rich in structural information; the Raman spectra of carbohydrates present several vibrational features in this region. Most of the bands are due to coupled vibrations involving hydrogen atoms; for instance, one can see a feature at 1,461  $\text{cm}^{-1}$ , which corresponds to CH, CH<sub>2</sub>, and COH deformations. In the region between 1,380 and 1,400  $\text{cm}^{-1}$  one can observe the coupling of the CCH and COH deformation modes, whereas in the region between 1,340 and 1,200  $\text{cm}^{-1}$  bands containing the contributions of several vibrational modes are observed, such as CO and CC stretching and CCH, COH, and CCH deformations. According to Berzerdjeb et al. [23], the region between 1,200 and 800  $\text{cm}^{-1}$  is very characteristic of the CO and CC stretching and COC deformation modes, referring to the glycosidic bond; this region is also known as the fingerprint or anomeric region, and is very often cited in literature by other investigators, such as Nikonenko et al. [27], Yang and Zhang [28], Baranska et al. [29], Thygesen et al. [30], and Vandenebeele et al. [31]. In the anomeric region one can see the distinction between the  $\alpha$  and  $\beta$  configurations of the polysaccharide molecules. The vibrations originating from  $\alpha$ -1,4 glycosidic linkages can be observed as strong Raman bands in the 920–960- $\text{cm}^{-1}$  region, and thus the band observed at 940  $\text{cm}^{-1}$  was assigned to the amylose  $\alpha$ -1,4 glycosidic linkage. In the FT-Raman spectra of native starches, very subtle changes of the glycosidic  $\alpha$ -1,6 linkage band can also be observed, and it can be assigned to the presence of amylopectin [16], since the band at 920–960  $\text{cm}^{-1}$  appears with a small change in intensity, when compared with the pure amylose band. Vibrations in the 800–400- $\text{cm}^{-1}$  region are in general due to CCO and CCC deformations, and in this region one can see very strong coupling which is related to the glycosidic ring skeletal deformations.

The very intense Raman band at 475–485  $\text{cm}^{-1}$  has been used as a marker to identify the presence of starch in different samples, as well as to characterize amylose and amylopectin, the polysaccharides which are the constituents of starch [16]. The literature has plenty of discussions about the vibrational spectra of starches in this region. For instance, Veij et al. [32] assigned the band at 477  $\text{cm}^{-1}$  to the presence of starch in pharmaceutical formulations, where starch is used as an excipient. Fechner et al. [13] showed small differences in the range between 400 and 640  $\text{cm}^{-1}$  between some native starches. Gussen et al. [25] discussed the band shift observed in the Raman spectra of amylose and amylopectin, such as the backbone vibrational band around 477  $\text{cm}^{-1}$  and the bands at 410, 769, and 1,382  $\text{cm}^{-1}$ . Baranska et al. [29] assigned the band at 478  $\text{cm}^{-1}$  to the presence of starch in the Raman spectrum of carrot, by using a Raman mapping technique to investigate the distribution of starch, and Vandenebeele et

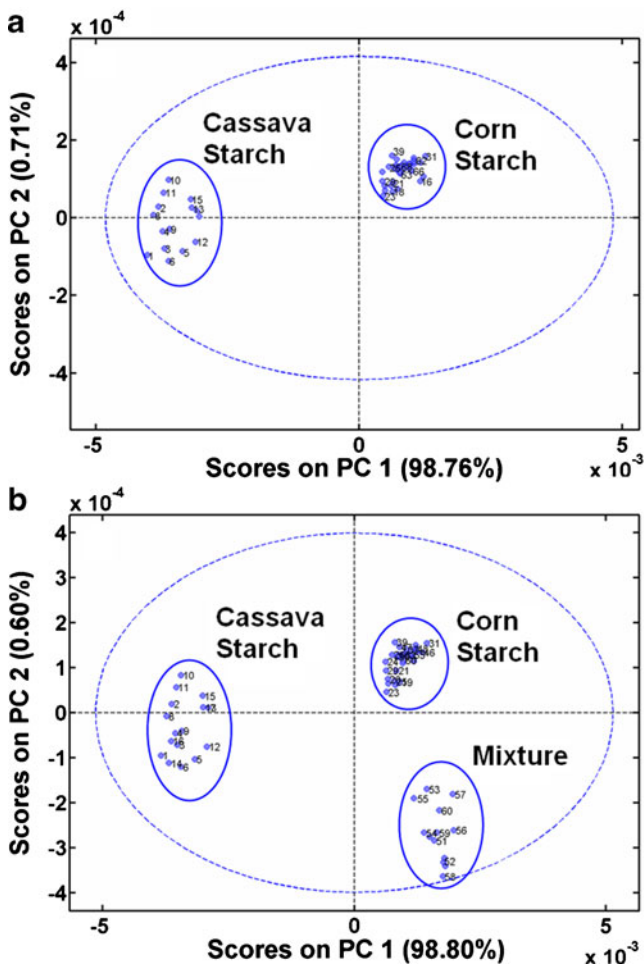
al. [31] used the band at  $477\text{ cm}^{-1}$  as a marker to distinguish between starch and gum. As discussed before, bands in the  $475\text{--}485\text{ cm}^{-1}$  region are related to the amount of amylose and amylopectin present in the starch samples; changes in band intensity or position are indicative of different amylose or amylopectin concentrations in the samples investigated.

### Sample classification

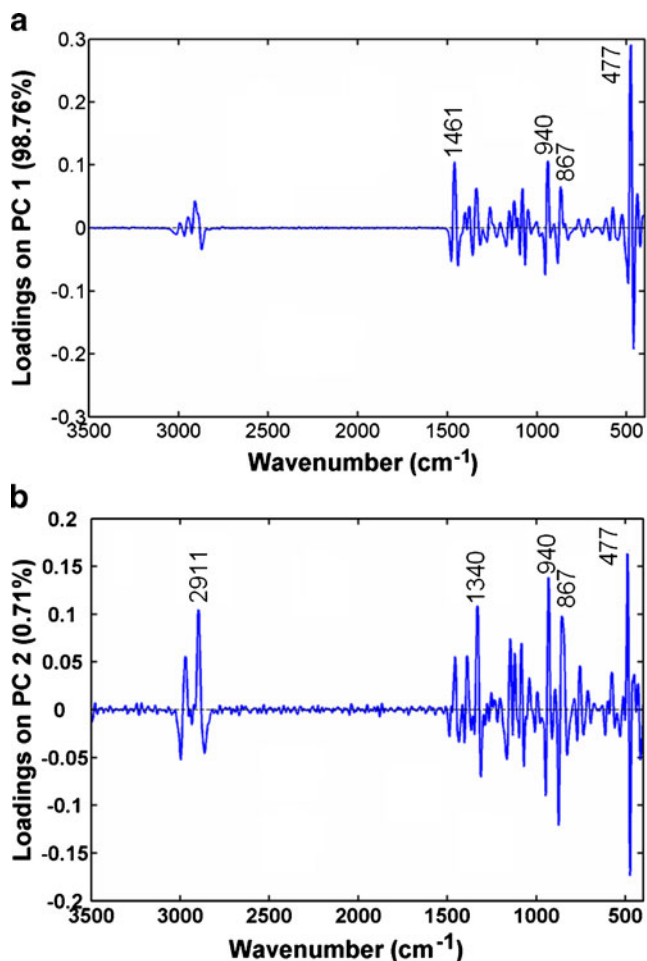
Chemometric multivariate analysis, based on PCA, was applied to the FT-Raman spectra of starches, to evaluate the ability to differentiate the botanical type (corn and cassava). In Fig. 2 one can see the scores of the PCA models for pure corn and cassava starches, and for the mixture of both starches. The first PCA model was built with 34 samples of corn starch and 17 samples of cassava starch. Three outlier samples were removed from the model; two of them belong to the cassava set, where one had a Raman spectrum with a

very high signal-to-noise ratio, and the other had a Raman spectrum with a different amylose content. The third sample removed belonged to the corn starch set, being a sample with amylose content lower than that of the others. The difference in amylose content depends mainly on the botanical origin, but can also be related to the growth state of the plant as well as the method of extraction [2]. Figure 2a shows the separation of samples into different groups, where the first two principal components described 99.4% of data variance. It was possible to investigate the interrelations among the samples on the basis of the score plots of the two principal components; for instance, the starches can be classified into two clusters, which are grouped according to their botanical origins, where it is possible to observe that the first principal component (98.7% of variance explained) is responsible for the separation between the starch samples. The variables which are mainly responsible for this differentiation can be observed by projecting the loadings for each variable of the first principal component and for the small contribution of the second principal component, as presented at Fig. 3. The spectral range from  $500\text{ to }450\text{ cm}^{-1}$  has the highest loadings and the most important variables for group separation; this spectral region involves the deformation modes of the C-C bond as well as the C-O bond torsion, and is correlated with different amylose contents in starches. The bands at  $870$ ,  $950$ , and  $1,460\text{ cm}^{-1}$  also have a contribution in the first principal component. The second principal component (Fig. 3b) shows the contribution of the  $2,900\text{ cm}^{-1}$  region, but also the contribution for the bands at  $477$ ,  $870$ ,  $950$ , and  $1,340\text{ cm}^{-1}$ . Figure 4 shows the second-derivative spectra of corn and cassava samples, showing clearly the same main Raman wavenumbers which are also important for the discrimination of starches samples, for instance, the ones at  $940$  and  $477\text{ cm}^{-1}$ , both related to the glycosidic bond and  $\delta(\text{C-C-C})+\tau(\text{C-O})$ , respectively. Figure 4 also shows the second-derivative spectra are in a good agreement with the PCA loading plot. The samples investigated have a lower amylose content in cassava than in corn, implying the data are on the negative side of the first principal component scores, whereas the corn samples are on the positive side because they have higher levels of amylose.

A new data set was inserted into the model to verify its separation capacity, built using Raman spectra of mixtures of corn and cassava starches in ratios ranging from 10 to 50% (w/w). The projection of the first two principal components (Fig. 2b) showed the formation of three clusters, with the new, third group being the mixture of corn and cassava starches. The recognition of the botanical origin of starch is very important because it enables the discovery of fraud, caused by improper mixing of starch products from different origins. Analy-



**Fig. 2** a Score plot of the first principal component (*PC1*) versus the second principal component (*PC2*) of the corn and cassava starch samples. b Score plot of *PC1* versus *PC2* of the corn starch, cassava starch, and mixtures of starch samples (see “Experimental”)



**Fig. 3** a Loading plot of PC1 versus Raman shift. b Loading plot of PC2 versus Raman shift

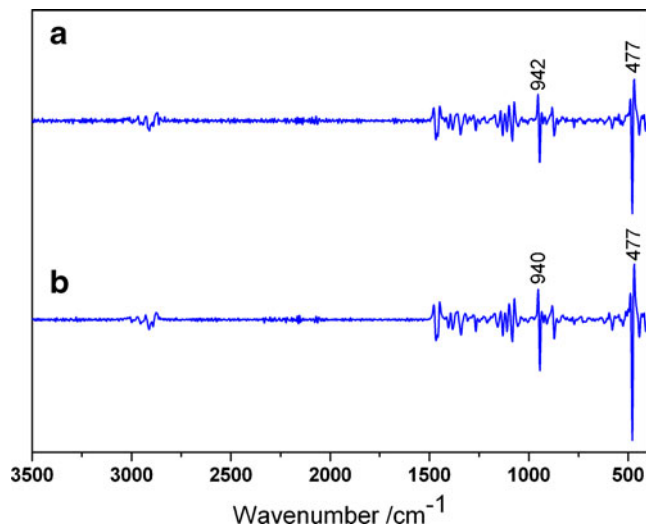
sis of Fig. 2 also shows the botanical origin can also be used to identify the groups; this is a very important finding, mainly for the development of the multivariate calibration models.

#### Models for amylose content determination

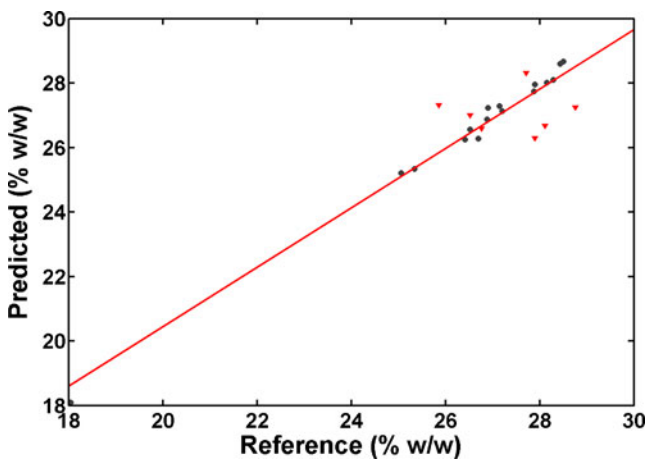
PLS is the method usually employed for multivariate calibration. All the relevant information contained in the Raman spectra were concentrated in a few latent variables, which were optimized to produce the best correlation with the desired property to be determined, in our case the amylose content in starch samples. The PLS model using 23 corn starch samples was constructed with the data obtained from the Raman intensity (variable  $X$ ) and the amylose content (variable  $Y$ ); the data were arranged in matrix form, 15 samples were used for the calibration set and eight samples for the validation set. The number of latent variables to be used in this model was chosen by leave-one-out cross-validation, resulting in the choice of four latent variables and an RMSECV of 2.14%; the high

RMSECV can be explained by the presence of the sample with 18% w/w amylose. Analysis of the chart leverage (Fig. S1) clearly shows that it has a high leverage and low residue, which implies the sample is important for the regression model and cannot be considered an anomalous sample (an outlier). The regression analysis (Fig. 5) shows the correlation between the values determined by the reference analysis method and the values predicted by Raman spectroscopy; the correlation coefficient obtained was 0.978. The results obtained from PLS and the criteria of the validation method, based on the relative error and the RMSEP, are shown in Table 2. The accuracy of the model can be evaluated by the RMSEC and the RMSEP; for the model built these are 0.17 and 1.15% w/w, respectively, and the prevision relative errors are lower than 6%, which is very similar to the standard method, showing that Raman spectroscopy can be applied for quantification of amylose in samples of corn starch.

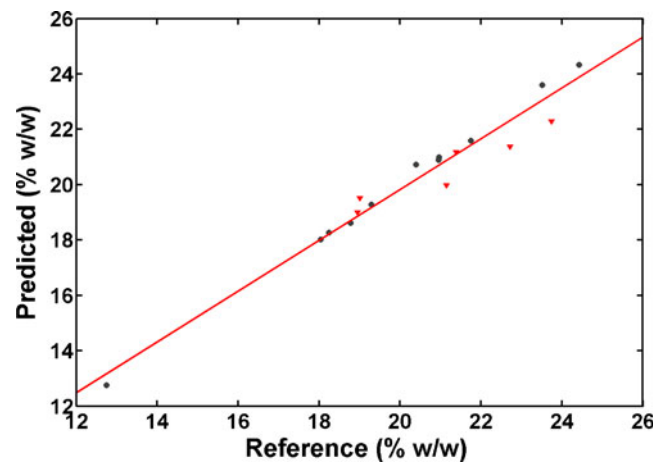
For the cassava starch samples a model containing 17 samples (11 samples were used for the calibration set and six for the validation set, Fig. 6) was built. Samples which have high leverage and high residue are considered outlier samples; in the calibration data one anomalous sample was observed (Fig. S2). This sample had a Raman spectrum with a very high signal-to-noise ratio when compared with other sample spectra, which seems to be the origin of the problem. To obtain the best number of latent variables, the leave-one-out cross-validation procedure was used. The best PLS regression results were obtained using Raman spectra with a model containing six latent variables, an RMSECV of 3.20%. The high RMSECV can be explained by the presence of a sample with high leverage; this sample can be classified as informative, since it has amylose content lower than the others (approximately 12%, com-



**Fig. 4** Second-derivative spectra of corn starch (A) and cassava starch (B) samples



**Fig. 5** Reference values versus FT-Raman predictions of the amylose content in corn starch. Circles calibration samples, triangles validation samples



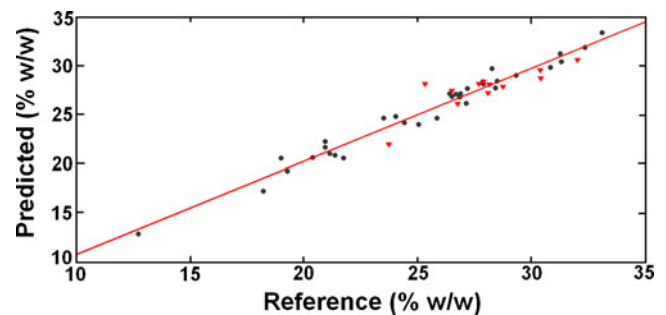
**Fig. 6** Reference values versus FT-Raman predictions of the amylose content in cassava starch. Circles calibration samples, triangles validation samples

**Table 2** Results for the validation set modeled with partial least squares (PLS) regression for starch samples

Model	Colorimetric results (% w/w)	PLS model (% w/w)	Relative error (%)
Corn	27.90±0.30	26.23	-5.86
	27.90±0.98	27.36	-1.94
	27.71±0.20	28.31	2.18
	25.86±0.41	27.35	5.78
	26.77±0.57	26.53	-0.87
	28.11±0.50	27.18	-5.46
	28.11±0.32	26.62	-5.29
	26.52±0.04	28.89	1.40
Cassava	21.38±0.04	21.17	-0.93
	21.14±0.09	19.97	-5.51
	22.72±0.01	21.37	-5.92
	23.74±0.08	22.29	-6.09
	19.01±0.27	19.51	2.64
	18.95±0.98	19.00	0.30
	23.74±0.08	21.92	-7.64
Corn and cassava	27.90±0.98	28.36	1.66
	27.71±0.20	28.12	1.47
	28.15±0.32	28.08	-0.24
	26.77±0.57	26.09	-2.54
	28.76±0.17	27.85	-3.15
	28.11±0.50	27.22	-3.16
	25.34±0.34	28.14	11.06
	26.52±0.04	27.42	3.38
	32.04±0.12	30.57	-4.58
	30.41±0.13	29.54	-2.87
30.42±0.24	28.70	-5.64	

pared with the approximately 20% average content). In the case of a linear model this same sample can provide additional information for the calibration set, which implies an improvement to the model; for instance, the sample has no negative influence on the cassava model. All these observations can be proved by the calculation of the RMSEC, the RMSEP, and the relative errors. The PLS calibration model constructed was used to determine the amylose content in other cassava starch samples. Figure 6 shows a plot of the amylose content predicted by the PLS model, using the Raman data, versus the values determined by the reference analysis method, which gives a correlation coefficient of 0.953 and an RMSEC of 0.33%, whereas the predictions for the verification samples of validation are indicative that the results are quite acceptable, RMSEP of 1.50% (w/w). Table 2 summarizes the prediction capabilities of the PLS model built for cassava starches and their relative error values.

Comparison of the two models clearly shows the model for corn starch samples has a better regression and prediction than the model for cassava samples; this fact can only be explained in terms of the larger number of corn



**Fig. 7** Reference values versus FT-Raman predictions of the amylose content in the corn and cassava starches. Circles calibration samples, triangles validation samples

starch samples contained in the model, when compared with number of cassava samples, which contributes to a better set of results.

A model using samples of both corn and cassava starches was also developed; this model was possible owing to the observation in the exploratory analysis, where the first principal component (98.7% of variance explained) is mainly responsible for the separation between the starch samples, and also this separation is related to the amylose content in these samples. For PLS modeling, the data matrix was divided into two data sets, one of calibration and one of validation, which ensures that there are samples of corn and cassava starches in the two data sets (validation and calibration). The optimum model dimension was determined by the minimum RMSECV (3.40%) for the calibration samples, and the latent variables were extracted by the leave one sample out at a time cross-validation method, resulting in the choice of six latent variables. The plot of the values predicted by the PLS model against the values measured by the colorimetric method in Fig. 7 shows the very good performance of the PLS model, with a correlation coefficient of 0.924, demonstrating good agreement between the data obtained from the conventional method and those predicted by the PLS model. The RMSEC and RMSEP obtained for the PLS model were 0.70 and 1.30% w/w, respectively. Table 2 shows the results from the PLS model and the relative errors obtained in the predictions using both corn and cassava samples; the small relative errors indicate that Raman spectroscopy can be applied to the analysis of new starch samples not used in the PLS model developed, and also the possibility of using this regression model for samples of corn and cassava starches, thus obtaining results which are very similar to the those of models built separately.

## Conclusions

FT-Raman spectroscopy has been successfully applied in the classification and determination of amylose content in corn and cassava starch samples. The advantages of using Raman spectroscopy are very clear: samples do not need to undergo pretreatment, so resulting in a shorter time of analysis, and the technique is not destructive, which avoids the use of other chemicals.

Analysis of the scores of the principal components revealed a division into two different groups of starches, separated according to their amylose content and thus to the botanical source. The analysis of the loadings pointed out the region of the spectrum that contributes to this separation, and the band at approximately  $480\text{ cm}^{-1}$ , regarding the ring vibration and related to the amylose content in starch, showed a higher loading.

The models developed for the quantification of amylose content have shown that the use of Raman spectroscopy, together with multivariate calibration regression based on PLS, allows the quantification of amylose in starch samples with prediction errors similar to those of the standard method (colorimetry).

The results clearly show that Raman spectroscopy and chemometrics can be used to identify differences between corn and cassava starches, as well as between mixtures of both starches (cassava and corn).

**Acknowledgements** The authors thank CNPq, CAPES, and FINEP (Brazilian agencies) for financial support.

## References

- Biliaderis CG (1991) *Food J Physiol Pharmacol* 69:60–78
- Bul on A, Colonna P, Planchot V, Ball S (1998) *Int J Biol Macromol* 23:85–112
- Daiuto ER, Cereda MP, Carvalho LJCB (2002) *Braz J Food Technol* 5:217–223
- Mahmood T, Turner MA, Stoddard F (2007) *Starch/Starke* 59:357–365
- Morgano MA, Moriya C, Ferreira MMC (2003) *Braz J Food Technol* 6:77–83
- Leal ES, Okada FM, Zamora PP (2008) *Quim Nova* 31:1621–1625
- Edwards HGM, Villar SEJ, De Oliveira LFC, Le Hyaric M (2005) *Anal Chim Acta* 538:175–180
- Edwards HGM, Farwell DW, De Oliveira LFC, Alia JM, Le Hyaric M, De Almeida MV (2005) *Anal Chim Acta* 532:177–186
- de Oliveira LFC, Colombara R, Edwards HGM (2002) *Appl Spectrosc* 56:306–311
- Moros J, Garrigues S, Guardia M (2007) *Anal Chim Acta* 593:30–38
- Barthus R, Poppi RJ (2001) *Vib Spectrosc* 26:99–105
- Paradkar MM, Irudayaraj J (2001) *Food Chem* 76:231–239
- Fechner PM, Wartewig S, Kleinebudde P, Neubert RHH (2005) *Carbohydr Res* 340:2563–2568
- Zhbankov RG, Firsov SP, Korolik EV, Petrov PT, Lapkovski MP, Tsarenkov VM, Marchenwka MK, Ratajczak H (2000) *J Mol Struct* 555:85–96
- Schuster KC, Ehmoser H, Gapes JR, Lendl B (2000) *Vib Spectrosc* 22:181–190
- Kizil R, Irudayaraj J, Seetharaman K (2002) *J Agric Food Chem* 50:3912–3918
- Dupuy N, Laureyns J (2002) *Carbohydr Polym* 49:83–90
- Sohn M, Himmelsbach DS, Barton FE (2004) *Cereal Chem* 81:429–433
- Brereton R (2000) *Analyst* 125:2125–2154
- International Organization for Standardization (1987) ISO 6647. International Organization for Standardization, Geneva
- Savitzky A, Golay MJE (1964) *Anal Chem* 36:1627–1639
- Kennard RW, Stone LA (1969) *Technometrics* 11:137–148
- Benzerdjeb AM, Mokhtari INT, Rahal MS (2007) *Spectrochim Acta Part A* 68:284–299
- Mokrane A, Michel PF, Cartier A, Rivail JL (1997) *J Mol Struct (Theochem)* 395:71–80
- Gussem KD, Vandenabeele P, Verbeke A, Moens L (2005) *Spectrochim Acta Part A* 61:2898–2908



26. Synytsya A, Copíkova J, Matejka P, Machovic V (2003) *Carbohydr Polym* 54:97–106
27. Nikonenko NA, Buslov DK, Sushko NI, Zhabankov RG (2005) *J Mol Struct* 752:20–24
28. Yang L, Zhang LM (2009) *Carbohydr Polym* 76:349–361
29. Baranska M, Schulz H, Baranski R, Nothnagel T, Christensen LP (2005) *J Agric Food Chem* 53:6565–6571
30. Thygesen LG, Løkke MM, Micklander E, Engelsen SB (2003) *Trends Food Sci Technol* 14:50–57
31. Vandenaabeele P, Wehling B, Moens L, Edwards H, De Reu M, Van Hooydonk G (2002) *Anal Chim Acta* 407:261–274
32. Veij M, Vandenaabeele P, De Beer T, Remon JP, Moens L (2009) *J Raman Spectrosc* 40:297–307