Food Chemistry 242 (2018) 196–204

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms

Pedro Sousa Sampaio[a,b,*], Andreia Soares[a], Ana Castanho[a], Ana Sofia Almeida[a], Jorge Oliveira[c], Carla Brites[a]

[a] Instituto Nacional de Investigação Agrária e Veterinária (INIAV), Av. da República, Quinta do Marquês, 2780-157 Oeiras, Portugal
[b] Faculty of Engineering, Lusophone University of Humanities and Technology, Campo Grande, 376, 1749-019 Lisbon, Portugal
[c] University College Cork, School of Engineering, Ireland

ARTICLE INFO

Keywords:
Multivariate models
Process Analytical Technologies
PLS
iPLS
siPLS
mwPLS

ABSTRACT

Determining amylose content in rice with near infrared (NIR) spectroscopy, associated with a suitable multivariate regression method, is both feasible and relevant for the rice business to enable Process Analytical Technology applications for this critical factor, but it has not been fully exploited. Due to it being time-consuming and prone to experimental errors, it is urgent to develop a low-cost, nondestructive and 'on-line' method able to provide high accuracy and reproducibility. Different rice varieties and specific chemometrics tools, such as partial least squares (PLS), interval-PLS, synergy interval-PLS and moving windows-PLS, were applied to develop an optimal regression model for rice amylose determination. The model performance was evaluated by the root mean square error of prediction (RMSEP) and the correlation coefficient (R). The high performance of the siPLS method (R = 0.94; RMSEP = 1.938; 8941–8194 cm^{-1}; 5592–5045 cm^{-1}; and 4683–4335 cm^{-1}) shows the feasibility of NIR technology for determination of the amylose with high accuracy.

1. Introduction

Rice (*Oryza sativa* L.), the world's main food crop, is constituted fundamentally by starch. Starch is a complex polysaccharide of α-D-glucose units exclusively, which are joined by a sequence of α-D-(1,4)-glucosidic linkages thus giving rise to linear or helical chains referred to as amylose. Although α-(1,6)-glucosidic linkages are much less frequent, they form branch points between the chains thereby creating highly branched domains, denominated amylopectin (Pandey et al., 2012). Starch biosynthesis in higher plants including rice is catalysed by four classes of enzymes, namely, ADP-Glc pyrophosphorylase (AGPase), starch synthase, starch branching enzymes and starch debranching enzymes. The enzyme granule bound starch synthase-I controls the synthesis of amylose in the rice endosperm, while soluble starch synthase, starch branching enzyme and starch debranching enzymes together control the synthesis of amylopectin (Bao, Sun, & Corke, 2002; Zhang et al., 2011). Amylose is considered to be the most important determinant of the eating quality of rice, and based on its content rice varieties can be classified as waxy (0–2%); very low (3–12%); low (13–20%); intermediate (21–25%) and high (> 26%) (Juliano et al., 1981). The fine structure of amylose, both molecular size and chain-length distribution, are also significant factors of the hardness of cooked rice (Li, Prakash, Nicholson, Fitzgerald, & Gilbert, 2016). Amylose content is correlated with the retrogradation behaviour, influencing the textural properties of cooked rice and the viscoelasticity dynamic of rice starch gel (Lu et al., 2009).

The classical method for amylose and amylopectin determination is the colour complex formed by iodine reaction coupled with potentiometric or amperometric titration. The method is based on the capacity inherent to amylose to accommodate polyiodide ions, chiefly I_5^-, within its helical structure. As amylopectin is unable to form such complexes because of its short chains and branch linkages interfering with the formation of stable structures, these complexes are specific for the amylose fraction (Hizukuri, 1996). However, the iodine affinity varies within species, hence compromising the accuracy of this method. A survey conducted by the international network for quality rice (INQR) showed that five different versions of the iodine binding method are currently in use and that the reproducibility was high within laboratories but low between laboratories (Fitzgerald et al., 2009). There are also other methods, such as differential scanning calorimetry (Sievert & Holm, 1993), potentiometry (Banks, Greenwood, & Muir, 1971), spectrophotometry (Morrison & Laignelet, 1983), and chromatography (Matheson & Welsh, 1988; Yun et al., 2013). The amylose can also be evaluated by the enzymatic method, developed by Megazyme

* Corresponding author at: Instituto Nacional de Investigação Agrária e Veterinária (INIAV), Av. da República, Quinta do Marquês, 2780-157 Oeiras, Portugal.
E-mail address: pnsampaio@gmail.pt (P.S. Sampaio).

http://dx.doi.org/10.1016/j.foodchem.2017.09.058
Received 14 February 2017; Received in revised form 11 August 2017; Accepted 12 September 2017
Available online 14 September 2017
0308-8146/

(Gibson, Solah, & McCleary, 1997). However this method is characterised by some drawbacks, such as the relatively high cost per sample and, mainly, it is hard testing a large number of samples (Hu, Burton, & Yang, 2010; Soong, Quek, & Henry, 2015). Despite the existence of other procedures, the colorimetric method still commonly used, and their accuracy was improved by using standards from specific rice varieties carrying the alleles of the Waxy gene responsible for amylose synthesis and calibration values obtained by separation of hydrodynamic volume and molecular weight of amylose by size exclusion chromatography (ISO 6647–1,2, 2015).

Near-infrared (NIR) spectroscopy is a promising technique with fast, easy-to-use, and nondestructive analytical potentials being widely accepted, requiring minimal or no sample preparation (Bart et al., 2007). It has become particularly popular in recent years in the pharmaceutical industry to assist the development of online Process Analytical Technologies (PAT) to achieve Quality by Design in manufacturing. However, its prediction accuracy depends on sample physical status, chemical components, temperature, colour, cleanliness, quantity used for measurement and above all, the statistical model used (Bagchi, Sharma, & Chattopadhyay, 2016). Apparent amylose content has been predicted by NIR spectroscopy using milled rice flour (Bao, Cai, & Corke, 2001; Delwiche, Bean, Miller, Webb, & Williams, 1995), milled whole grain (Delwiche et al., 1995; Shu, Wu, Xia, Gao, & McClung, 1999; Windham et al., 1997), or amylose and proteins in rice flour (Xie et al., 2014). However, those studies faced several drawbacks concerning the valuable rice amylose reference data and the model performance. The main difficulty of NIR spectroscopy with multivariate analysis is related to wavenumber or spectral region selection, especially when the spectra displays unresolved peaks or fails to identify important features. Several methods have been studied to select the optimal variables for multivariate calibration to remove irrelevant spectral variables and improve model performance; The multivariate calibration builds a predictive model relating measured quantities (wavenumbers) to properties of interest (concentration data). A variety of linear regression methods based on latent variables (LVs) have been developed to address this problem, such as partial least squares (PLS), but due to several drawbacks, such as the noise in spectral data, the calibration and prediction errors are high, and the model can be affected (Wold & Sjostrom, 2001). Meanwhile, spectral region selection, using appropriate algorithms, was reported to considerably improve the performance of the full-spectrum calibration techniques, avoiding non-modeled interferences and building a well-fitted model (Friedel, Patz, & Dietrich, 2013; Lee, Bawn, & Yoon, 2012; Nørgaard, Saudland, Wagner, Nielsen, & Munck & Engelsen, 2000). Studies then performed showed that it is fundamental to conduct a spectral region selection responsible for the property of interest to increase the prediction performance (Kalivas, 1997; Spiegelman et al., 1998). These methods can be classified into two classes: single wavelength selection and wavelength interval selection. Actually, several approaches have been proposed for selection of the optimal set of spectral regions, such an interval PLS (iPLS), synergy PLS (siPLS) and moving window PLS (mwPLS) (Friedel et al., 2013; Leardi & Nørgaard, 2004; Ma, Wang, Chen, Cheng, & Lai, 2017). The principle of iPLS consists of splitting the spectra into equal-width intervals and developing sub-PLS models for each one. The sub-intervals with the lowest value of the root mean squared error of prediction (RMSEP) are deemed to be the best. Many methods based on iPLS were developed to optimise the combination of the selected intervals, such as siPLS (Leardi & Nørgaard, 2004). The main advantage of this kind of method is that it uses a graphical display to focus on a choice of better sub-intervals and conduct comparisons among the prediction performance of local models and the full-spectrum model. Instead of just testing a series of adjacent but non-overlapping intervals, which would miss some more informative ones, mwPLS was proposed to overcome this drawback. It builds a series in a window that moves through the whole spectra and then chooses the informative intervals with low model complexity and low value of the sum of residuals. The mwPLS is a modelling technique that can be thought of as a series of diagnostic PLS regressions based on all continuous window size "H" in the parent data set. In effect, a window of size H is "moved" across the data set to collect modelling information. The model quality and number of latent variables (LVs) required for model production during this process can then be used to find the best spectral region(s) of size H. mwPLS is a promising procedure used to conduct consecutive wavelength selection for building an optimal calibration model; this method is proven to be effective for waveband selection in analysis of many objects (Chen, Yin, Tang, & Pan, 2017; Yun et al., 2013).

The objective of this study was to test the various methods proposed to develop multivariate models to select the most appropriate to obtain reliable and accurate measurements of amylose in rice. A large set of rice varieties was used to challenge the various models. PLS, iPLS, siPLS and mwPLS procedures for NIR quantitative analysis of amylose were investigated and compared. The different steps required for model calibration were analysed. The number of PLS factors and the number of region intervals was optimised according to the root mean square error in the calibration set. The performance of the final model was evaluated according to the RMSEP and the correlation coefficient (R) with the prediction set. The model thus created can be considered a way to obtain a fast, non-destructive, accurate and reproducible methodology for amylose determination in different rice varieties (after a suitable milling procedure), providing a modern gold standard for laboratory and industrial analysis amenable for the development of PATs for the rice industry.

2. Materials and methods

2.1. Rice sample

For this study, sixteen rice varieties (including *Indica* and *Japonica* sub species) from a Portuguese Rice Breeding Program were grown at three different sites along the basins of 3 different rivers with very different micro-climates (Alcácer do Sal, Salvaterra de Magos and Montemor-o-Velho, Portugal) along four seasons (2012–2015), providing 168 samples. Also, 11 standard rice varieties, sourced from the International Rice Research Institute, Los Baños, Philippines, (IRRI), characterised by different amylose content, were also used: IR 65; IR 24; IR 64; WU BAI LI; IRRI109; IRRI134; IRRI138; IRRI148; IRRI149 and IRRI151.

2.2. Rice flour sample preparation

About 20 g of rice was ground to flour in a Cyclone Sample Mill (Falling number 3100, Perten, Sweden) equipped with a 0.8 mm screen.

2.3. Amylose determination

Amylose of rice was determined using the standard iodine colorimetric method according to ISO 6647–2: (2015). The absorbance was measured using a spectrophotometer (Hitachi; Japan) at 720 nm. Amylose content was quantified using a standard curve created from absorbance values of 4 calibrated samples from standard rice varieties carrying one of the five alleles of the Waxy gene, which is the gene responsible for amylose synthesis (IR8, IR24, IR64, IR65) obtained from IRRI. Pure amylose (potato origin) (Sigma-Aldrich, Germany) was also evaluated. The amylose content was evaluated in duplicate for each sample of rice, and the reference value corresponds to the average.

2.4. Instrumentation and measurements

The samples containing approximately 25 cm^3 of rice flour were loaded in a circular sample cup and pressed slightly to obtain a similar packing density. Sample spectra were collected using an NIR