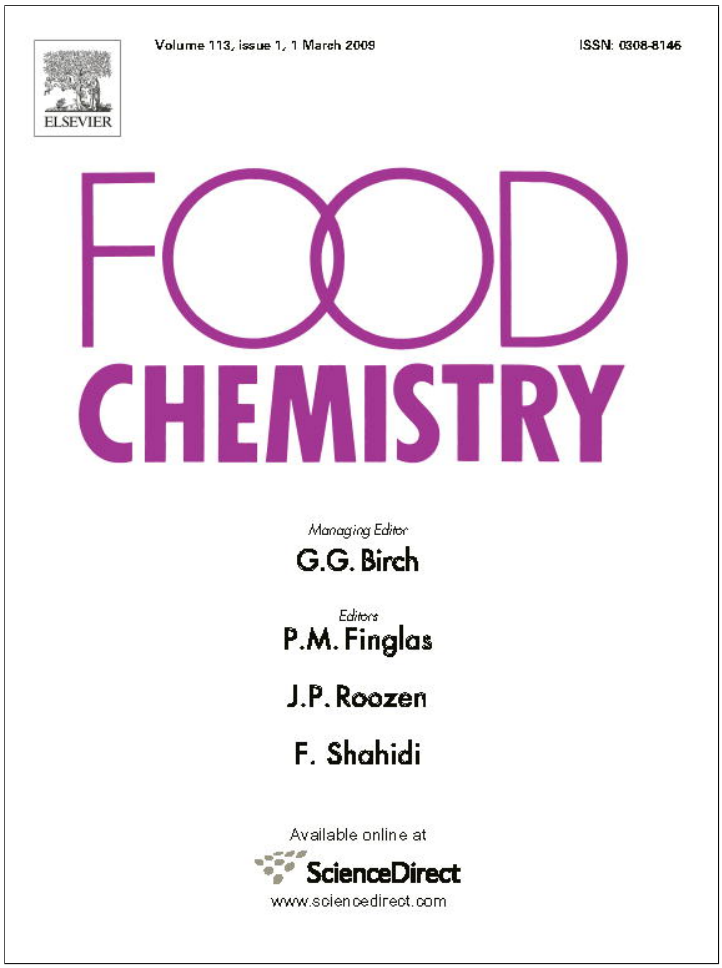
View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Biblioteca Digital do IPB

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Food Chemistry 113 (2009) 246–252

Contents lists available at ScienceDirect

Food Chemistry

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

Analytical Methods

UV spectrophotometry method for the monitoring of galacto-oligosaccharides production

Luís G. Dias [a], Ana C.A. Veloso [a,b], Daniela M. Correia [a], Orlando Rocha [b,c], Duarte Torres [b,c], Isabel Rocha [b,c], Lígia R. Rodrigues [b,c], António M. Peres [a,d,*]

[a] CIMO – Mountain Research Centre, Escola Superior Agrária de Bragança, Instituto Politécnico de Bragança, Quinta de Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal
[b] IBB – Institute for Biotechnology and Bioengineering, Center of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[c] BIOTEMPO – Biotechnology Consulting Ltd., Avepark, Zona Industrial da Gandra, Apartado 4152, 4806-909 Caldas das Taipas, Portugal
[d] LSRE – Laboratory of Separation and Reaction Engineering, Escola Superior Agrária de Bragança, Instituto Politécnico de Bragança, Quinta de Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal

ARTICLE INFO

Article history:
Received 14 March 2008
Received in revised form 23 June 2008
Accepted 28 June 2008

Keywords:
Fermentation processes
Galacto-oligosaccharides
UV spectrophotometer
Partial least squares regression
Artificial neural network

ABSTRACT

Monitoring the industrial production of galacto-oligosaccharides (GOS) requires a fast and accurate methodology able to quantify, in real time, the substrate level and the product yield. In this work, a simple, fast and inexpensive UV spectrophotometric method, together with partial least squares regression (PLS) and artificial neural networks (ANN), was applied to simultaneously estimate the products (GOS) and the substrate (lactose) concentrations in fermentation samples. The selected multiple models were trained and their prediction abilities evaluated by cross-validation and external validation being the results obtained compared with HPLC measurements. ANN models, generated from absorbance spectra data of the fermentation samples, gave, in general, the best performance being able to accurately and precisely predict lactose and total GOS levels, with standard error of prediction lower than 13 g kg^{-1} and coefficient of determination for the external validation set of 0.93–0.94, showing residual predictive deviations higher than five, whereas lower precision was obtained with the multiple model generated with PLS. The results obtained show that UV spectrophotometry allowed an accurate and non-destructive determination of sugars in fermentation samples and could be used as a fast alternative method for monitoring GOS production.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

In the last decade an increasing attention has been paid to the development of dietary supplements for functional foods that positively affect the microbial composition of the gut, namely probiotics and prebiotics (Chockchaisawasdee, Athanasopoulos, Niranjan, & Rastall, 2005). Prebiotics are non-digestible food ingredients and are able to stimulate the specific growth/activity of a restricted number of colonic bacteria. This family of compounds includes several oligosaccharides (namely fructo-, gluco-, galacto-, isomalto-, xylo-, and soyo-oligosaccharides), lactulose, lactosucrose, among others (Fric, 2007). Galacto-oligosaccharides (GOS) are carbohydrate-based food ingredients that can enhance health-related physiological activities, which can provide protection from infection; decrease the number of potentially pathogenic bacteria; facilitate the normal functions of the gut; stimulate the absorption of some minerals and decrease blood lipids content (Anthony, Merriman, & Heimbach, 2006; Baminger, Haltrich, Kulbe, & Nidetzky, 2001; Bruzzese, Volpicelli, Squaglia, Tartaglione, & Guarino, 2006; Fric, 2007; Palframan, Gibson, & Rastall, 2002; Pennacchia, Vaughan, & Villani, 2006; Petzelbauer, Zeleny, Reiter, Kulbe, & Nidetzky, 2000; Simmering & Blaut, 2001; Tzortzis, Goulas, Baillon, Gibson, & Rastall, 2004; van den Broek, Hinz, Beldman, Vincken, & Voragen, 2008; Vos et al., 2006). In addition, their stability under acidic conditions makes them ideal substances for application as ingredients in a wide range of food products. Their excellent taste quality, low caloric value and relatively high sweetness make GOS interesting functional sweeteners (Remaud-Simeon, Willemot, Sarçabal, de Montalk, & Monsan, 2000; Splechtna et al.,2001, 2006).

GOS are mainly produced by enzymatic synthesis with β-galactosidases with transgalactosylation activity, using lactose as substrate, in reactors operated in continuous or batch conversion modes (Albayrak & Yang, 2002; Aslan & Tanrıseven, 2007; Boon, van 't Riet, & Janssen, 2000; Chen, Ou-Yang, & Yeh, 2003; Chockchaisawasdee et al., 2005; Nakkharat, Kulbe, Yamabhai, & Haltrich, 2006; Petzelbauer, Splecthna, & Nidetzky, 2002; Sako,

* Corresponding author. Address: LSRE – Laboratory of Separation and Reaction Engineering, Escola Superior Agrária de Bragança, Instituto Politécnico de Bragança, Quinta de Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal. Tel.: +351 273303220; fax: +351 273325405.
E-mail address: peres@ipb.pt (A.M. Peres).

0308-8146/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2008.06.072

Author's personal copy

Matsumoto, & Tanaka, 1999; Splechtna, Nguyen, & Haltrich, 2007; Splechtna et al., 2001, 2006, 2007).

Up to date different methods have been used to quantify reaction mixture products. Concerning the quantification of lactose and GOS, the most common methods are: capillary electrophoresis with a UV-DAD detector (Chockchaisawasdee et al., 2005; Splechtna et al., 2001, 2006, 2007; van den Broek et al., 2008), high-performance anion-exchange chromatography with pulsed amperometric detection (Chockchaisawasdee et al., 2005; Nakkharat et al., 2006; Splechtna, Nguyen, & Haltrich, 2007; Splechtna et al., 2006, 2007), high-performance chromatography with a refractive index detector (Albayrak & Yang, 2002; Boon et al., 2000), thin layer chromatography (Aslan & Tanrıseven, 2007; Sanz, Côté, Gibson, & Rastall, 2005; Splechtna et al., 2001, 2006) and mass spectrometry (Sanz et al., 2005).

In this work, a bioreactor operating in a batch mode was used for routine GOS production using lactose as substrate and a yeast GOS-producer strain. The saccharides present in the reaction mixture were quantified using a hydrophilic interaction high-performance liquid chromatography technique with refractive index detection. For the lactose and total GOS quantifications, a novel UV spectrophotometric method, less expensive and faster (less than 5 min analysis time, including the filtration and dilution steps), was also developed, allowing a more rapid and equal accurate monitoring of the fermentation process in an almost real time basis, which is most relevant for industry applications. The ability of the new proposed method to estimate the lactose and global GOS concentration profiles was evaluated by means of two statistical techniques: partial least square regression (PLS) and artificial neural networks (ANN). The predictive abilities of the regression models were investigated using a full leave-one-out cross-validation and an external validation. It should be remarked that the objective was to develop a simple and precise UV spectrophotometric method to quantify the above-mentioned sugars, specially the total GOS content, produced in pre-established and comparable operating conditions, in such a way that could reproduce a usual industrial GOS production process.

To the best of the authors' knowledge, this is the first time that such an approach, based on UV spectra, is used to estimate and predict lactose and GOS contents on fermentation samples. A more complex and time-consuming approach has been described in the literature for the rapid estimation of global sugars in fruit juices; soft drinks and during wine making; using an UV spectrophotometric detection of by-products from UV photodegradation of carbohydrates (Roig & Thomas, 2003a, 2003b).

Other approaches also described in the literature are based on more complex analytical methodologies (near-infrared reflectance spectroscopy) that, together with empirical regression models are used for food composition estimation. Recently, Kulmyrzaev and co-workers (Kulmyrzaev, Karoui, de Baerdemaeker, & Dufour, 2007) have demonstrated the potential of more expensive techniques, such as infrared and fluorescence spectroscopy, along with chemometric tools, namely descriptive techniques such as principal component analysis (PCA) and predictive techniques such as factorial discriminant analysis (FDA), principal component regressions (PCR) and partial least square regressions (PLS), to measure and characterize nutritional components of foods, such as dairy products, muscle tissue, and cereals. On the other hand, Huang and collaborators (Huang, Kangas, & Rasco, 2007) have reviewed the basic theory of artificial neural networks (ANN) technology and its applications in food science. Indeed, in the last two decades, ANN have been applied in almost every aspect of food science, namely in food safety and quality analysis, including microbial growth modelling, spectroscopic data interpretation, and prediction of physical, chemical, functional and sensory properties of several food products during processing and distribution (Huang et al., 2007). However, it is still difficult to find works concerning their applications for the rapid monitoring of fermentation processes, namely in which concerns the prediction of substrates and products concentrations.

2. Experimental

2.1. Reagents

Lactose (>99.5% w/w) was purchased from Hi-media; ammonium sulfate (>99% w/w), potassium sulfate (>99% w/w), potassium hydrogen sulfate (>99% w/w) were purchased from Riedel-de-Haën; magnesium sulphate (>99.5% w/w), ammonium hydroxide (HPLC grade) were purchased from Sigma; calcium carbonate (>99% w/w) was purchased from Panreac; acetonitrile (HPLC grade) was purchased from Carlo Erba, ammonia (25% v/v, solution) was purchased from Merck; phosphoric acid ($\geqslant$85% v/v solution) was purchased from Fluka.

2.2. GOS production

GOS were produced by a yeast GOS-producer strain in a 5L Biostat MD fermenter from B. Braun Biotech (Germany) connected to a digital control unit. The medium composition used for GOS production was: 250 g kg^{-1} for lactose, 6 g kg^{-1} for ammonium sulphate, 3 g kg^{-1} for potassium sulphate, 1 g kg^{-1} for potassium hydrogen sulfate, 0.25 g kg^{-1} for magnesium sulfate and 0.2 g kg^{-1} for calcium carbonate. Ammonia and phosphoric acid solutions were used for pH control during the fermentation. Samples were collected periodically along the fermentation and, subsequently, filtered using cellulose acetate membranes (0.2 μm) to remove biomass.

2.3. HPLC analysis

For the HPLC analysis a modular liquid chromatograph (Jasco) equipped with a Prevail Carbohydrate ES column (5 μm, 250 × 4.6 mm) from Alltech was used, at room temperature. Elution was achieved using a mixture of acetonitrile and 0.04% ammonium hydroxide in water (70:30 v/v) at a flow rate of 1.0 ml/min. The response of the refractive index detector was recorded and integrated using the Star Chromatography Workstation software (Varian). A calibration curve was obtained using lactose as external standard. The oligosaccharides concentrations are proportional to their peak areas and they were determined using the same proportionality constant as extracted from the calibration with lactose. It should be noted that the accuracy of this approximation was verified by checking the material balance as reported by other authors (Boon, Janssen, & van der Padt, 2000).

2.4. Spectrophotometric analysis

The spectra data (190–1100 nm, at intervals of 1 nm) of 53 samples withdrawn from three different fermentations carried out in a routine process of GOS production were registered using a SPECORD 200 spectrophotometer (Analytic Jena) and treated using the WinASPECT software. Before analysed, each sample was diluted in the proportion of 1:100 with deionised water (obtained from a TGI pure water system) and filtered through a 0.2 μm nylon filter (Whatman). Absorption was detected in a narrower wavelength interval (190–299 nm), when compared with the above-mentioned.

2.5. Data analysis

The lactose and total GOS concentrations, obtained from the chromatographic analysis, were used together with the spectral

Author's personal copy

absorbance data of the fermentation diluted samples to obtain the empirical PLS and ANN regression models. Since absorbance values are sensitive to base effects, the absorbance spectra were converted to difference spectra by subtracting the deionised water spectra recorded in each day. Moreover, the spectral data interval used for the data analysis was reduced to the UV range of 190–299 nm, since it corresponded to the interval where radiation absorption was detected. Furthermore, the spectra data were transformed using different mathematical treatments (none, first or second order derivatives) before use in order to establish and choose the best PLS or ANN multiple regression models. Also, the existence of chemical anomalous samples was investigated using the student's *t* statistic, obtained by calculating the quotient of the difference between the reference and the predicted value and the standard error of calibration (SEC). Samples with a *t* value higher than 2.5 were considered chemically anomalous, suggesting that the reference data were suspicious (Andrés et al., 2007; Ortiz-Somovilla, España-España, Gaitán-Jurado, Pérez-Aparicio, & De Pedro-Sanz, 2007).

2.5.1. Partial least square regression (PLS)

PLS is a widely used statistical tool for the establishment of empirical models based on absorbance and reference data. PLS regression is a bilinear modeling technique that extracts the most relevant information as a mathematical model of linear combinations of the spectral bands to predict a property of interest (chemical, physical or sensory attributes) of different samples. This technique is very useful when a whole range of spectral information is analyzed which is typically high dimensional compared to the number of observations, showing in general strong autocorrelation between bands (Helland, 2001; Kandaswamy, Bajwa, & Apple, 2005). Since the principal component scores are uncorrelated, the problem of multi-collinearity among the predictor variables is avoided. Two common PLS approaches can be used. In one case, the calibrations are generated for one component at a time, while in the other; it is possible to calibrate multiple components simultaneously.

In this work only multiple models have been studied in order to estimate and predict simultaneously the lactose and the total GOS content of each fermentation sample. The experimental concentration data together with the different types of derivative spectra treatments (none, first or second order derivatives) of the 110 wavelengths were processed in a commercial software (The Unscrambler software) using a multiple PLS method, with a maximum of 15 principal components (factors) being allowed. The number of PLS factors that significantly contributed to the model variability was identified through a full cross-validation "leave-one-out" method.

2.5.2. Artificial neural networks (ANN)

ANN are very sophisticated non-linear modeling techniques, capable of modeling particularly complex functions. The basic idea of ANN is to simulate the function of the human brain that has a basic unit called a neuron. Similar to a biological neuron, an artificial neuron receives a series of input information connected to a weight factor, which is adjustable during network training. Neurons form layers with intra or inter-layer connections, resulting in feedback or feedforward networks. Layers between the input and the output layers are called hidden layers. Usually, in order to determine the optimal number of hidden nodes, a trial and error strategy is employed. The learning process for developing a neural network can be either supervised or unsupervised if it needs or not target outputs as teacher, respectively. Two different ANN models can also be established according to whether it is intended to obtain calibration models for one or more components at a time: single or multiple ANN, respectively.

In this work a multiple feedfoward network trained by back-propagation, which is one of the best known supervised learning algorithms, is used. Two kinds of networks were studied: multilayer perceptrons (MLP) and radial basis function (RBF). The former is one of the most popular network architectures in use today. In the later training is much faster than in MLP, since the simple linear transformation in the output layer can be fully optimized using traditional linear modeling techniques.

The proximate analysis data (concentration) together with the different types of derivative spectra treatments (none, first or second order derivatives) were processed using a commercial software (STATISTICA Neural Networks software) with multiple (two) network outputs: lactose and total GOS contents. For each network tested an automatic search for an effective sub-set of the specified variables (110 wavelengths in the 190–299 nm range) was allowed. Moreover, the network complexity was also automatically determined by the program. From the multiple MLP or RBF networks tested, the one that gave the best performance concerning the simultaneous estimation and prediction of lactose and total GOS contents was retained.

2.5.3. PLS and ANN models calibration and validation

For PLS and ANN multiple models development the experimental fermentation samples (53 samples corresponding to three independent fermentations) were divided into two or three groups, respectively. The PLS multiple model's first group, used for the establishment of the regression model, was constituted by 43 samples (training set which was also used as cross-validation set), while the second one (external validation set) consisted of 10 samples and was used to validate the PLS model. Regarding the ANN multiple model, the first group consisted in 33 samples (training set), the second in 10 samples (validation set) and the last group in other 10 samples (test or external validation set). The samples constituting each set group were randomly selected from the overall samples withdrawn from the three fermentations carried out for GOS production from lactose, using a yeast GOS-producer strain (Section 2.2). It should be remarked that the samples of the external validation set were the same for both PLS and ANN multiple models. Only one PLS or ANN multiple model was used for the prediction of both lactose and total GOS content, selected after a series of tryouts in order to choose the model with best predictive performance using raw and different derivative treatments applied to the spectra data (first and second order derivatives). The optimal number of PLS factors (latent values) and the best type and complexity of the ANN, for both constituents, were chosen in order to obtain the lowest standard error of calibration (SEC), of cross-validation (SECV) and validation (SEP) for the overall error between modeled and reference values as well as the highest regression coefficient of determination for calibration (R^2) and for cross-validation or validation (r^2) (Ortiz-Somovilla et al., 2007). Once the best PLS or ANN multiple model had been selected, with a specific combination of treatments (raw spectral data, first or second order derivatives), a comparison between the standard errors of calibration and validation of both models was carried out. Moreover, the residual predictive deviation (RPD) was also used to evaluate the predictive ability of both calibration models. This statistic is given by the relationship between the standard deviation (SD) of the population's reference values and the standard error of cross-validation (SECV) or of internal validation (SEP). RPD values higher than 3, show good predictive ability of the model (Kulmyrzaev et al., 2007; Ortiz-Somovilla et al., 2007). Finally, the predictive behavior of each multiple model was also evaluated using a test set (external validation) by calculating the standard error of prediction (SEP) and the coefficient of determination in the external validation (R^2_{EV}) (Ortiz-Somovilla et al., 2007).

Author's personal copy

3. Results and discussion

3.1. Reference values and UV spectra data

The reference values concerning the concentrations of lactose and total GOS in the fermentation samples, employed for developing the predictive PLS and ANN multiple models, were obtained using the chromatographic procedure described in Section 2.3. Table 1 shows the number of samples used (*N*), mean values, standard deviation (SD), as well as the range of reference values of the fermentation components analyzed (lactose and total GOS concentrations), used in the training, validation and test sets for the development of the PLS and ANN empirical models. As can be inferred from the data shown in the above-mentioned table, the reference data used presented a high variability for both lactose and total GOS contents which is required in order to obtain robust predictive multivariate models.

Concerning the spectra data obtained for the fermentation samples analysed it was observed that significant absorption only occurred in the range of 190–299 nm. Fig. 1 shows an example of the UV spectra (190–299 nm) recorded for two different fermentation samples, with different lactose and total GOS contents (in both samples the two compounds are present, and one had a high lactose content and the other a high total GOS content).

3.2. Prediction models

3.2.1. Characterization of PLS and ANN multiple predictive models

The spectra treatments referred in Section 2.5 were applied in order to obtain the PLS and ANN multiple regression equations. The results obtained with each one of the multiple regression models showed low cross-validation residuals (*t* values calculated for lactose and total GOS were lower than 2.5), thus no evidence of chemical anomalous samples was detected. Therefore, none of the 53 fermentation samples was omitted in this study.

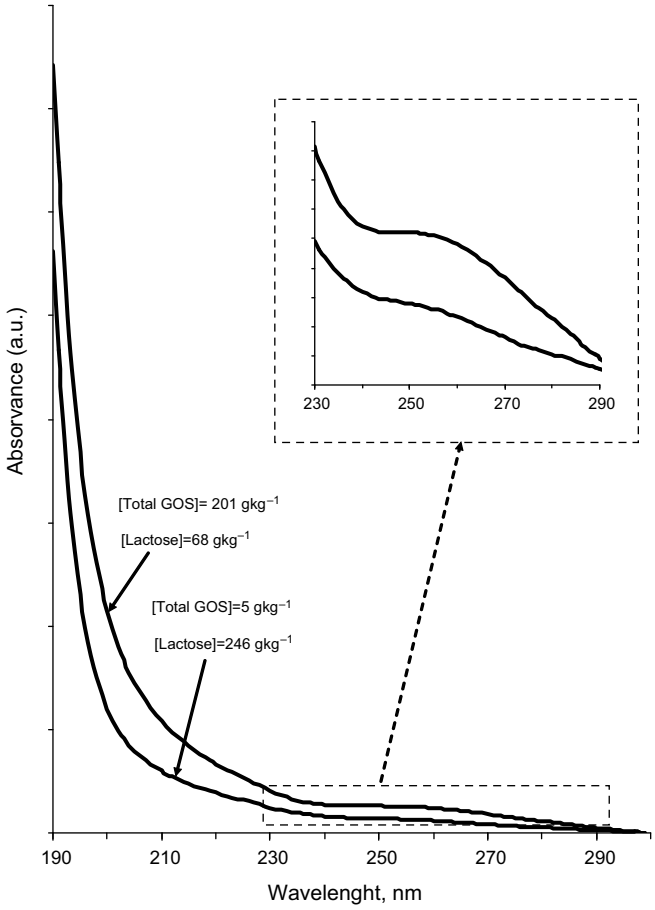


Fig. 1. UV spectra (190–299 nm) recorded for two fermentation samples, with different lactose and total GOS contents.

Table 1
Reference values (calibration, validation and test sets) used for the development of the PLS multiple regression model and of the ANN multiple regression model

		Lactose	Total GOS
Sets: PLS multiple regression model			
Training (calibration and cross-validation)	*N*	43	43
	Mean concentration (g kg^{-1})	129.1	99.9
	SD (g kg^{-1})	70.4	61.4
	Range (g kg^{-1})	5.9–245.6	0–242.0
Test	*N*	10	10
	Mean concentration (g kg^{-1})	141.2	94.0
	SD (g kg^{-1})	59.0	74.9
	Range (g kg^{-1})	38.5–239.6	7.8–255.0
Sets: ANN multiple regression model			
Training (calibration)	*N*	33	33
	Mean concentration (g kg^{-1})	129.6	98.5
	SD (g kg^{-1})	75.4	65.6
	Range (g kg^{-1})	5.9–245.6	0–242.0
Validation	*N*	10	10
	Mean concentration (g kg^{-1})	127.3	104.7
	SD (g kg^{-1})	54.3	47.8
	Range (g kg^{-1})	18.1–186.3	47.5–193.4
Test (external validation)	Mean concentration (g kg^{-1})	141.2	94.0
	SD (g kg^{-1})	59.0	74.9
	Range (g kg^{-1})	38.5–239.6	7.8–255.0

As already mentioned, for both PLS and ANN multiple models analyzed different mathematical treatments were applied to the UV spectra signals (none, first and second order derivatives). The best performances were achieved using the raw data (without any treatment) and the second order derivative treatment for the PLS and ANN multiple regressions, respectively.

Concerning the PLS multiple model, nine factors were selected for the estimation/prediction of lactose and total GOS concentrations, using 110 selected wavelengths in the UV range of 190–299 nm. This model was chosen after the comparison between the total SEP values calculated as the root square of the sum of the quadratic SEP obtained for lactose and total GOS, using the different PLS multiple models tested (from 1 to 15 factors). This procedure was employed to decide the optimal factor number (number of principal components, PCs) of the final PLS multiple model. As can be inferred from the results presented in Fig. 2, the above-mentioned selected PLS multiple model (with nine factors) was the one that allowed the best predictive performance.

Regarding the ANN multiple model developed simultaneously for the analyzed variables a RBF network with 82 inputs (82 of the 110 wavelengths) and 13 hidden layers was selected (obtained using K-Means for defining radial neuron weights; K-Nearest Neighbor for defining radius; and pseudo-inverse training algorithms), showing very good performance for the two variables studied (regression ratios equal to 0.08 and 0.19 for total GOS and lactose, respectively). This network was chosen after a series of tryouts that allowed finding the best network using a search strategy where a balance between network performance and

Author's personal copy

type-complexity (MLP or RBF-number of input variables and hidden nodes) was considered in order to ensure network diversity. For each attempt at least 75 networks with different complexities were tested, being retained the best one found. In order to partially illustrate the selection procedure adopted to choose the best type-complexity network a comparison between the total SEP values calculated for the 10 best networks retained for each of the two types of networks considered in this study is presented in Fig. 2. The analysis of the results presented shows that the aforementioned RBF chosen network (with 82 input variables and 13 hidden nodes) was the one with the best predictive ability, showing the lowest total SEP values, which is the statistical parameter that is usually adopted when comparing performances of differently configured ANN models (Gallardo, Alegret, & del Valle, 2004).

3.3.2. Evaluation of predictive models

Table 2 shows the statistics of the PLS and ANN multiple equations designed to simultaneously predict the lactose and total GOS concentrations in fermentation samples, based on the raw UV spectra and on the second order derivative treatment of the UV spectra, respectively. Results are shown for the training, validation and test sets of samples, and include the standard errors of calibration, of cross-validation and of prediction (SEC, SECV and SEP, respectively), the coefficients of determination in calibration, in cross-validation and in external validation (R^2, r^2 and R^2_{EV}, respectively), as well as the residual predictive deviation (RPD).

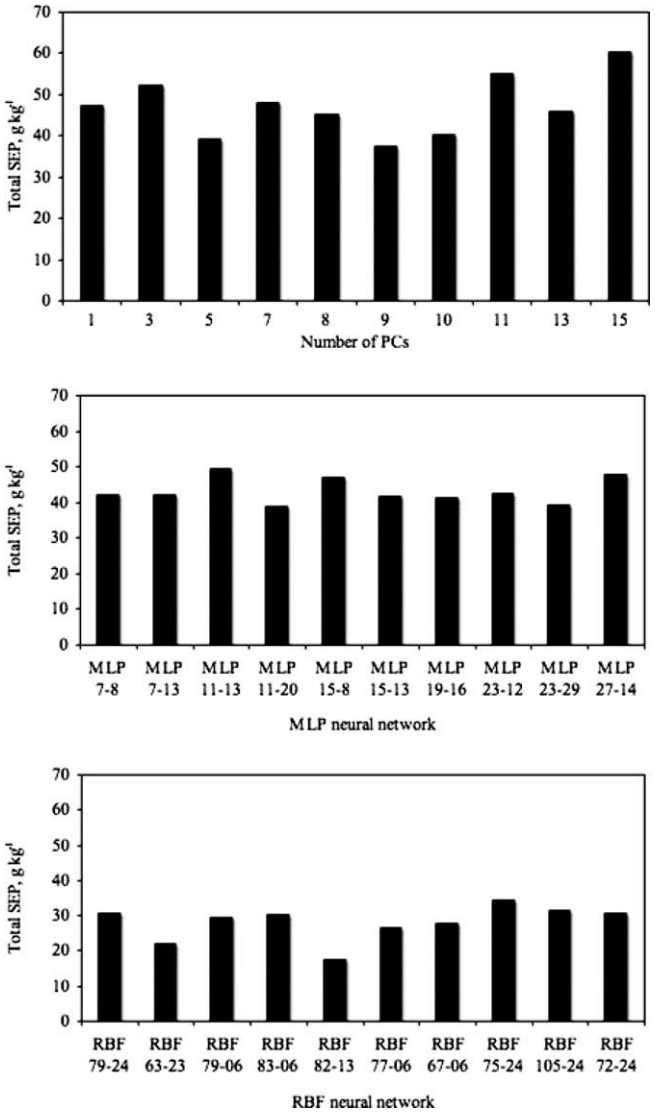


Fig. 2. Illustration of one of the criteria (total SEP values) used in the selection procedure adopted for deciding the best PLS and ANN multiple models, using raw data (without any treatment) and second order derivative treatment, respectively.

Generally, the results obtained show that the ANN multiple model developed presents a more satisfactory performance concerning the prediction of lactose and total GOS contents, than that achieved with the PLS multiple model. Indeed, the coefficients of determination (R^2, r^2 and R^2_{EV}) obtained for lactose and total GOS in the training, validation and test sets were equal to or greater than 0.93 and 0.85 for ANN and PLS, respectively (Table 2), showing that UV absorbance data can explain a significant proportion of the concentration data variability. It has been reported that the accuracy of a model presenting determination coefficient values in the range of 0.7–0.89 or equal to or greater than 0.9 is good or very good, respectively (Gaitán-Jurado, Ortiz-Somovilla, España-España, Pérez-Aparicio, & De Pedro-Sanz, 2008). Therefore, from the results obtained in this study it can also be inferred that, globally, the ANN multiple model established is very accurate while the results obtained with the PLS multiple model show a satisfactory accuracy.

Additionally, from the analysis of the other statistics presented on Table 2 (SEC, SECV, RPD and SEP), it can be stated that both multiple models present similar and satisfactory performances regarding the estimation of lactose and total GOS concentrations (similar SEC values), although as already stated, regarding the predictive characteristics the ANN multiple model is considerably more accurate (lower SEP values and higher RPD values), especially for the prediction of the total GOS concentration.

Furthermore, the quality of the results obtained in this work with both PLS and ANN multiple models is similar to that reported by Roig and Thomas (Roig & Thomas, 2003a, 2003b) for the rapid estimation of global sugars in fruit juices soft drinks and during wine making using however, a more complex UV spectrophotometric detection by means of by-products of UV photodegradation of carbohydrates. Moreover, the coefficients of determination obtained in the present work are of the same order of magnitude, although lower, than those reported by Rodriguez-Saona and collaborators (Rodriguez-Saona, Fry, McLaughlin, & Calvey, 2001) for the rapid analysis of sugars in fruit juices by FT-NIR spectroscopy together with PLS regression models. So, taking into account the more experimental simplicity of the proposed analytical technique and the good prediction performance in the prediction of major components of the complex medium where it is applied, it can

Table 2
Statistics of selected PLS multiple and ANN multiple equations for calibration (SEC, R^2), cross-validation (SECV, r^2 and RPD) and external validation (SEP, R^2_{EV}) sets

Compound	Calibration set		Cross-validation set		Test set (external validation)		
	SEC (g kg^{-1})	R^2	SECV (g kg^{-1})	r^2	RPD	SEP (g kg^{-1})	R^2_{EV}
PLS multiple regression model							
Lactose	12.94	0.97	20.65	0.92	3.44	20.65	0.95
Total GOS	15.04	0.94	25.01	0.85	2.48	24.91	0.98
	Calibration set		Validation set		Test set (external validation)		
	SEC (g kg^{-1})	R^2	SEP (g kg^{-1})	r^2	RPD	SEP (g kg^{-1})	R^2_{EV}
ANN multiple regression mode							
Lactose	14.14	0.97	9.87	0.96	5.50	12.11	0.93
Total GOS	15.81	0.94	5.29	0.99	9.04	12.97	0.94

Author's personal copy

be concluded that the approach now developed is an alternative and an improved tool for fast fermentation monitoring, specially when compared with the standard methods usually applied for the quantification of GOS.

Even though the satisfactory results obtained with both the PLS and ANN multiple models, which can be also inferred from the analysis of Fig. 3, a validation process was carried out in order to test the acceptance of the UV method as an alternative methodology for lactose and total GOS quantification. Therefore, a comparison between the lactose and total GOS concentrations predicted by the multivariate-UV spectrophotometric method with those obtained by the HPLC method, considered as the reference procedure, was carried out, by testing if the slope and intercept values of the regression models obtained could be considered equal to the theoretical expected ones: one and zero, respectively (Roig and Thomas, 2003a, 2003b). The results obtained regarding this validation procedure are listed in Table 3 for the best PLS and ANN multiple regression models obtained.

From the previous results gathered on Table 3, it is possible to conclude that, except for the prediction of total GOS content (external validation) with the PLS multiple model, regarding the two parameters under study, there is no statistical evidence, at a 5% significance level, that the slope and the intercept of the regression lines are different from the theoretical expected values. This leads to the acceptance of the UV method as an alternative quantification procedure for lactose and total GOS concentrations in the studied fermentation samples.

Although the slopes of the regression lines obtained in training and validation sets were high (equal or greater than 0.91), for the test set, only with the ANN model an acceptable slope value could be obtained (0.88) for the total GOS concentration prediction, showing, as expected from the previous results of this study, that this model is the most accurate.

However, it should be recognized that the robustness and performance of both UV-PLS and UV-ANN multiple models can be increased if a larger experimental database is used. In fact, only three

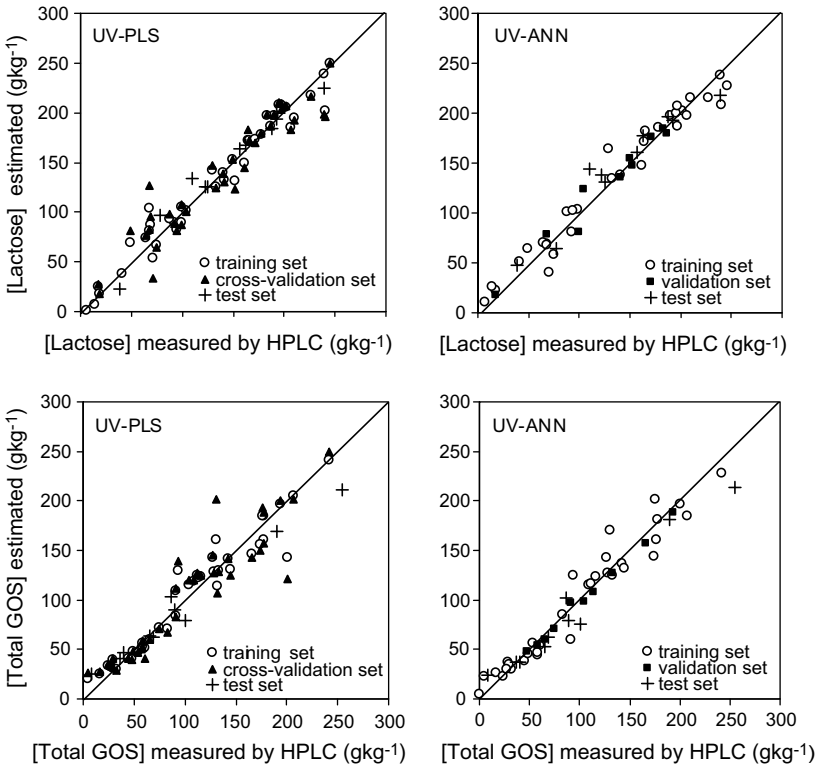


Fig. 3. Estimated lactose and total GOS concentrations using UV-PLS and UV-ANN multiple models versus experimental data obtained by HPLC.

Table 3
Validation procedure of the PLS and of the ANN multiple models PLS and ANN multiple models by comparison between the experimental UV-raw spectra data and the UV-second order derivative spectra data, respectively, and the HPLC (reference method) results

	Lactose			Total GOS		
	Calibration	Validation	External validation	Calibration	Validation	External validation
PLS multiple model						
Slope	0.97	0.96	0.94	0.94	0.98	0.78
Intercept	4.36	2.96	10.4	5.99	0.71	16.2
CI slope	[0.91, 1.02]	[0.87, 1.05]	[0.77, 1.11]	[0.86, 1.01]	[0.85, 1.11]	[0.68, 0.88]
CI intercept	[−4.0, 12.7]	[−10.5, 16.4]	[−15.4, 36.3]	[−2.8, 14.8]	[−14.3, 15.7]	[4.5, 27.8]
ANN multiple model						
Slope	0.97	0.98	0.91	0.94	0.96	0.88
Intercept	4.34	3.56	18.8	5.53	0.47	4.92
CI slope	[0.90, 1.03]	[0.82, 1.13]	[0.71, 1.11]	[0.86, 1.03]	[0.90, 1.02]	[0.68, 1.09]
CI intercept	[−5.5, 14.2]	[−17.7, 24.8]	[−11.6, 49.3]	[−4.4, 15.5]	[−6.3, 7.3]	[−13.3, 23.1]

CI: *t*-test 95% confidence interval.

Author's personal copy

fermentation runs of GOS production have been carried out for this particular study. Therefore, better prediction performances can be expected with the inclusion of new data from other fermentations carried out in the same operation conditions, namely if a specific experimental design is used in order to take into account all possible situations that could occur in a pre-established GOS industrial production process.

4. Conclusions

The UV method together with chemometric models of calibration was successfully used for lactose and total GOS concentration quantification along the GOS production fermentation. This work showed that this methodology is an acceptable analytical method for a fast, simple and inexpensive monitoring of total GOS production in a pre-defined fermentation process, allowing to promptly verifying if the fermentation is running as expected, or if some correction action is needed, which is crucial when an industrial GOS production is envisaged, being a possible alternative to the standard analytical methods usually used. Concerning the chemometric methods analyzed, in general, the ANN multiple model showed greater robustness presenting the best global prediction performance. Nevertheless, the quality of the results obtained with the PLS multiple model proved that, in some cases, this method could also be used. However, further studies are needed in order to better sustain these conclusions.

Acknowledgements

The authors gratefully acknowledge the financial support of this study by Project Biolife-PRIME 03/347 of Agência da Inovação – Progama IDEIA (Potugal). The author Duarte Torres also acknowledges to Fundação para a Ciência e a Tecnologia (Portugal) for the PhD Grant received (reference SFRH/BDE/15510/2004).

References

Albayrak, N., & Yang, S.-T. (2002). Production of galacto-oligosaccharides from Lactose by *Aspergillus orizae* β-galactosidase immobilized on cotton cloth. *Biotechnology and Bioengineering, 77*(1), 8–19.

Andrés, S., Murray, I., Navajas, E. A., Fisher, A. V., Lambe, N. R., & Bünger, L. (2007). Prediction of sensory characteristics of lamb meat samples by near infrared reflectance spectroscopy. *Meat Science, 76*, 509–516.

Anthony, J. C., Merriman, T. N., & Heimbach, J. T. (2006). 90-Day oral (gavage) study in rats with galacto-oligosaccharides syrup. *Food and Chemical Toxicology, 44*, 819–826.

Aslan, Y., & Tanrıseven, A. (2007). Immobilization of Pectinex Ultra SP-L to produce galacto-oligosaccharides. *Journal of Molecular Catalysis B: Enzymatic, 45*, 73–77.

Baminger, U., Haltrich, D., Kulbe, K. D., & Nidetzky, B. (2001). Production of a lactose-free galacto-oligosaccharide mixture by using selective enzymatic oxidation of lactose into lactobionic acid. *Enzyme and Microbial Technology, 29*, 434–440.

Boon, M. A., Janssen, A. E. M., & van der Padt, A. (2000). Effects of temperature and enzyme origin on the enzymatic synthesis of oligosaccharides. *Enzyme and Microbial Technology, 26*, 271–281.

Boon, M. A., van 't Riet, K., & Janssen, A. E. M. (2000). Enzymatic synthesis of oligosaccharides: Product removal during a kinetically controlled reaction. *Biotechnology and Bioengineering, 70*(4), 411–420.

Bruzzese, E., Volpicelli, M., Squaglia, M., Tartaglione, A., & Guarino, A. (2006). Impact of prebiotics on human health. *Digestive and Liver Disease, 38*(2), S283–S287.

Chen, C. W., Ou-Yang, C.-C., & Yeh, C.-W. (2003). Synthesis of galacto-oligosaccharides and transgalactosylation modeling in reverse micelles. *Enzyme and Microbial Technology, 33*, 497–507.

Chockchaisawasdee, S., Athanasopoulos, V. I., Niranjan, K., & Rastall, R. A. (2005). Synthesis of galacto-oligosaccharide from lactose using β-galactosidase from *Kluyveromyces lactis*: Studies on batch and continuous uf membrane-fitted bioreactors. *Biotechnology and Bioengineering, 89*(4), 434–443.

Fric, P. (2007). Probiotics and prebiotics – renaissance of a therapeutic principle. *Central European Journal of Medicine, 2*(3), 237–270.

Gaitán-Jurado, A. J., Ortiz-Somovilla, V., España-España, F., Pérez-Aparicio, J., & De Pedro-Sanz, E. J. (2008). Quantitative analysis of pork dry-cured sausages to quality control by NIR spectroscopy. *Meat Science, 78*(4), 391–399.

Gallardo, J., Alegret, S., & del Valle, M. (2004). A flow-injection electronic tongue based on potentiometric sensors for the determination of nitrate in the presence of chloride. *Sensors and Actuators B, 101*, 72–80.

Helland, I. S. (2001). Some theoretical aspects of partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 58*, 97–107.

Huang, Y., Kangas, L. J., & Rasco, B. A. (2007). Applications of artificial neural networks (ANNs) in food science. *Critical Reviews in Food Science and Nutrition, 47*, 113–126.

Kandaswamy, J., Bajwa, S. G., & Apple, J. K. (2005). Chemometric modeling of fat, cholesterol and caloric content of fresh and cooked ground beef with NIR reflectance spectroscopy. In *Proceedings of the* ISA/IEEE 2005 – sensors for industry conference, sicon'05, art. no. 4027454 (pp. 52–58).

Kulmyrzaev, A. A., Karoui, R., de Baerdemaeker, J., & Dufour, E. (2007). Infrared and fluorescence spectroscopic techniques for the determination of nutritional constituents in foods. *International Journal of Food Properties, 10*, 299–320.

Nakkharat, P., Kulbe, K. D., Yamabhai, M., & Haltrich, D. (2006). Formation of galacto-oligosaccharides during lactose hydrolysis by a novel β-galactosidase from the moderately thermophilic fungus *Talaromyces thermophilus*. *Biotechnology Journal, 1*, 633–638.

Ortiz-Somovilla, V., España-España, F., Gaitán-Jurado, A. J., Pérez-Aparicio, J., & De Pedro-Sanz, E. J. (2007). Proximate analysis of homogenized and minced mass of pork sausages by NIRS. *Food Chemistry, 101*, 1031–1040.

Palframan, R. J., Gibson, G. R., & Rastall, R. A. (2002). Effect of pH and dose on the growth of gut bacteria on prebiotic carbohydrates *in vitro*. *Anaerobe, 8*, 287–292.

Pennacchia, C., Vaughan, E. E., & Villani, F. (2006). Potential probiotic *Lactobacillus* strains from fermented sausages: Further investigations on their probiotic properties. *Meat Science, 73*, 90–101.

Petzelbauer, I., Splecthna, B., & Nidetzky, B. (2002). Development of an ultrahigh-temperature process for the enzymatic hydrolysis of lactose. III. Utilization of two thermostable β-glycosidases in a continuous ultrafiltration membrane reactor and galacto-oligosaccharide formation under steady-state conditions. *Biotechnology and Bioengineering, 77*(4), 394–404.

Petzelbauer, I., Zeleny, R., Reiter, A., Kulbe, K. D., & Nidetzky, B. (2000). Development of an ultra-high-temperature process for the enzymatic hydrolysis of lactose. II. Oligosaccharide formation by two thermostable β-glycosidases. *Biotechnology Bioengineering, 69*, 140–149.

Remaud-Simeon, M., Willemot, R.-M., Sarçabal, P., de Montalk, G. P., & Monsan, P. (2000). Glucansucrases: Molecular engineering and oligosaccharide synthesis. *Journal of Molecular Catalysis B: Enzymatic, 10*, 117–128.

Rodriguez-Saona, L. E., Fry, F. S., McLaughlin, M. A., & Calvey, E. M. (2001). Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy. *Carbohydrate Research, 336*, 63–74.

Roig, B., & Thomas, O. (2003a). Rapid estimation of global sugars by UV photodegradation and UV spectrophotometry. *Analytica Chimica Acta, 477*, 325–329.

Roig, B., & Thomas, O. (2003b). UV monitoring of sugars during wine making. *Carbohydrate Research, 338*, 79–83.

Sako, T., Matsumoto, K., & Tanaka, R. (1999). Recent progress on research and application of non-digestible galacto-oligosaccharides. *International Dairy Journal, 9*, 69–80.

Sanz, M. L., Côté, G. L., Gibson, G. R., & Rastall, R. A. (2005). Prebiotic properties of alternansucrase maltose-acceptor oligosaccharides. *Journal of Agricultural and Food Chemistry, 53*, 5911–5916.

Simmering, R., & Blaut, M. (2001). Pro- and prebiotics the tasty guardian angels? *Applied Microbiology and Biotechnology, 55*, 19–28.

Splechtna, B., Nguyen, T.-H. , & Haltrich, D. (2007). Comparison between discontinuous and continuous lactose conversion processes for the production of prebiotic galacto-oligosaccharides using β-galactosidase from *Lactobacillus reuteri*. *Journal of Agricultural and Food Chemistry, 55*, 6772–6777.

Splechtna, B., Nguyen, T.-H., Steinböck, M., Kulbe, K. D., Lorenz, W., & Haltrich, D. (2006). Production of prebiotic galacto-oligosaccharides from lactose using β-galactosidases from *Lactobacillus reuteri*. *Journal of Agricultural and Food Chemistry, 54*, 4999–5006.

Splechtna, B., Nguyen, T.-H., Zehetner, R., Lettner, H. P., Lorenz, W., & Haltrich, D. (2007). Process development for the production of prebiotic galacto-oligosaccharides from lactose using β-galactosidase from *Lactobacillus* sp. *Biotechnology Journal, 2*, 480–485.

Splechtna, B., Petzelbauer, I., Baminger, U., Haltrich, D., Kulbe, K. D., & Nidetzky, B. (2001). Production of a lactose-free galacto-oligosaccharide mixture by using selective enzymatic oxidation of lactose into lactobionic acid. *Enzyme and Microbial Technology, 29*, 434–440.

Tzortzis, G., Goulas, A. K., Baillon, M.-L. A., Gibson, G. R., & Rastall, R. A. (2004). *In vitro* evaluation of the fermentation properties of galacto-oligosaccharides synthesised by α-galactosidase from *Lactobacillus reuteri*. *Applied Microbiology and Biotechnology, 64*, 106–111.

van den Broek, L. A. M., Hinz, S. W. A., Beldman, G., Vincken, J.-P., & Voragen, A. G. J. (2008). *Bifidobacterium* carbohydrases-their role in breakdown and synthesis of (potential) prebiotics. *Molecular Nutrition and Food Research, 52*(1), 146–163.

Vos, A. P., Haarman, M., Buco, A., Govers, M., Knol, J., Garssen, J., et al. (2006). A specific prebiotic oligosaccharide mixture stimulates delayed-type hypersensitivity in a murine influenza vaccination model. *International Immunopharmacology, 6*, 1277–1286.