Available online at www.sciencedirect.com
SCIENCE DIRECT
Talanta 68 (2006) 1005–1012
Talanta
www.elsevier.com/locate/talanta

Monitoring substrate and products in a bioprocess with FTIR spectroscopy coupled to artificial neural networks enhanced with a genetic-algorithm-based method for wavelength selection

Vanina G. Franco [a], Juan C. Perín [b], Víctor E. Mantovani [a], Héctor C. Goicoechea [a,*]

[a] *Laboratorio de Control de Calidad de Medicamentos, Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe S3000ZAA CC 242, Argentina*
[b] *Cátedra de Química Orgánica, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe S3000 CC. 242, Argentina*

Received 9 May 2005; received in revised form 4 July 2005; accepted 4 July 2005
Available online 10 August 2005

Abstract

An experiment was developed as a simple alternative to existing analytical methods for the simultaneous quantitation of glucose (substrate) and glucuronic acid (main product) in the bioprocesses Kombucha by using FTIR spectroscopy coupled to multivariate calibration (partial least-squares, PLS-1 and artificial neural networks, ANNs). Wavelength selection through a novel ranked regions genetic algorithm (RRGA) was used to enhance the predictive ability of the chemometric models. Acceptable results were obtained by using the ANNs models considering the complexity of the sample and the speediness and simplicity of the method. The accuracy on the glucuronic acid determination was calculated by analysing spiked real fermentation samples (recoveries ca. 115%).

Keywords: Bioprocess; Mid-infrared spectroscopy; Multivariate calibration; Genetic algorithm; Glucuronic acid; Glucose; Gluconic acid

1. Introduction

An important branch of Biotechnology is currently devoted to the development of processes for probiotics production on different scales. An attractive bioprocess is constituted for the glucose fermentative degradation produced by a combination of a yeast (*Schizosacaromyces pombe*) and a bacteria (*Acetobacter xylimun*), usually named *Kombucha*, which constitutes a generous probiotic producer [1–3]. This system yields a series of compounds like organic acids, which are very important, since some of them have health beneficial properties. Among those acids the following can be mentioned: glucuronic, gluconic, acetic, lactic, succinic, mannonic, propionic and ascorbic acids. The system also produces vitamins B_1, B_2, B_3, B_6 and B_{12} [3]. However, most properties of Kombucha are attributed to the acidic composition of the beverage. Its detoxifying property is presumably due to the capacity of glucuronic acid to bind to toxin molecules and increase their excretion from organism by the kidneys or the intestines [2]. Thus, several pathologies produced by the accumulation of toxins in the body may be relieved this way.

Several reports have been presented regarding the determination of glucuronic and gluconic acids in fermentations and foods, but they are exclusively based on separative techniques like high performance liquid chromatography (HPLC) and capillary electrophoresis (CE) [4–8]. On the other hand, Fourier transform infrared (FTIR) spectroscopy shows an enormous potential for the quantitative multicomponent analysis of complex samples, and in recent times, has been exploited as an alternative technique for fermentation monitoring. Recently presented examples of the use of the more information-rich mid-infrared spectra

* Corresponding author. Tel.: +54 342 4575205; fax: +54 342 4575205.
E-mail address: hgoico@fbcb.unl.edu.ar (H.C. Goicoechea).

0039-9140/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2005.07.003

are the measurement of substances like glucose, acetic and lactic acids, lactose, galactose and acetone–butanol [9–11].

Since FTIR spectroscopy is a direct probe of molecular vibrations, each molecule has unique absorption spectrum. FTIR spectroscopy provides high information spectra and the Lambert–Beer law is obeyed, making it possible to quantify the concentration of several components also in a complex sample using multivariate data analysis, consequently avoiding time consuming separation steps prior measurement. On the other hand, chemometric applications to improve information obtained from modern instrumental data (i.e. spectral and electrochemical data) in different fields have acquired a routine character [12–22]. Partial least-squares (PLS) has become a usual tool for multivariate calibration because of the quality of the obtained calibration models, the ease of its implementation and the availability of software [23,24]. It allows a rapid determination of components, usually with no need of a prior separation for analysis. An additional advantage of such multivariate methods is that calibration can be performed by ignoring the concentrations of all other components except the analyte of interest. This makes these methods especially appealing for the determination of the active components in complex samples, whose components may show analytical signals, which are severely overlapped with those from the analytes. Pertinent examples of the application of FTIR coupled to multivariate methods are found in the literature [25–27].

Even though PLS assumes a linear relationship between the measured sample parameters and the intensity of its absorption bands, several authors have postulated that small deviations from linearity are acceptable as they can readily be suppressed by including additional modeling factors [28–30]. However, in the presence of substantial non-linearity, PLS tends to give large prediction errors and calls for more suitable models. Analogous considerations can be made when modeling complex and overlapped signals. Intrinsically non-linear calibration techniques such as non-linear partial least-squares (NPLSs) [28,31,32], locally weighted regression (LWR) [28,32], alternating conditional expectations (ACE) [28] and artificial neural networks (ANNs) [33–38] are applicable in the latter cases. However, it is important to state that these methods are computationally more complex than linear methods, and they heavily depend on the amount and quality of data available [34].

The object of this experiment was to develop a simple alternative to existing analytical methods for the simultaneous quantitation of glucose (substrate) and glucuronic and gluconic acids (main products) in the bioprocesses Kombucha by using FTIR spectroscopy coupled to multivariate calibration (PLS-1 and ANNs), and exploit the possibility of wavelength selection as a powerful tool to enhance the predictive ability. We have thus applied a very recently introduced method for wavelength selection based in a simple and fast genetic algorithm (GA), named RRGA (ranked regions GA) [39].

2. Experimental

2.1. Reagents

All chemicals were analytical-reagent grade and used without further purification. Nanopure water was used throughout. Kombucha was obtained at Kombucha Magic Mushroom Farm Inc.

2.2. Apparatus

FT-IR spectra were measured for all samples in a Perkin-Elmer Spectrum RX1 FT-IR spectrophotometer, using a barium fluoride cell with an optical path of 25 μm. The spectral range was between 450 and 4400 cm^{-1} at 2 cm^{-1} intervals (i.e., 1976 data points).

2.3. Software

All spectra were saved in ASCII format, and transferred to a PC Pentium 750 microcomputer for subsequent manipulation. All calculations were done using MATLAB 6.0 [40]. PLS-1 was implemented using the MVC1 MATLAB toolbox [41]. ANNs were implemented by using a home made MATLAB routine.

2.4. Batch reactor setup

An infusion of commercial black tea and glucose was used as medium, and a 3 L cylindrical vessel of thermal glass (3 L culture flask: Cole Palmer, Catalog Number E-29300-04) was used as reactor. After sterilization at 105° C, the reactor was charged with 2 L of medium to which 20.0 g of Kombucha wet mass was added. Then, the system was stored at 28° C. Details on the scheme of the reactor can be found in ref. [42].

2.5. Calibration, validation and real samples

In the presently studied fermentation samples, the analytes of interest are embedded in a complex mixture of a large number of components. Moreover, the analyst occasionally does not possess the information corresponding to the composition of the sample, although an estimation of the components can be made and an artificial calibration set can be built in order to take into account all possible variability sources on the instrumental response. In view of these considerations, two replicates of a 15 sample set were built to be used for the calibration step when applying chemometic tools. Each replicate of the artificial calibration set consisted of 15 samples corresponding to a central composite design with one central sample (in this case mixtures of the studied components in five concentration levels with one blank matrix sample), with the concentrations of the three analytes lying in the known linear absorbance-concentration range. The rest of the matrix components were added from a blank solution containing: acetic acid, 3.0 g L^{-1}, lactic acid, 0.6 g L^{-1} and

Table 1
Sample set used for calibrating PLS and training ANNs models

Sample	Glucuronic acid ($g\,L^{-1}$)	Gluconic acid ($g\,L^{-1}$)	Glucose ($g\,L^{-1}$)
1	0.07	0.60	5.56
2	9.63	0.60	5.56
3	4.81	0.04	5.56
4	4.81	1.16	5.56
5	4.81	0.60	0.62
6	4.89	0.60	10.51
7	1.98	0.32	3.09
8	7.72	0.32	3.09
9	1.98	0.92	3.09
10	7.72	0.92	3.09
11	1.98	0.60	8.65
12	7.72	0.32	8.65
13	2.05	0.92	8.65
14	7.72	0.92	8.65
15	4.81	0.60	5.56

ethanol, $0.5\,g\,L^{-1}$ prepared in the infusion of commercial black tea without glucose. The corresponding concentration of glucose and glucuronic and gluconic acids can be seen in Table 1. Both replicates were prepared to be used as training and monitoring sets when applying ANNs (see Table 1).

On the other hand, a nine samples validation set was prepared in order to validate the chemometric models. The analytes concentrations were chosen in a randomized way, but spanning the concentration range of the calibration set (see Table 3). Calibration and validation sets were prepared by placing suitable amounts of each compound and completing to the flask mark with the blank solution. All mixtures were measured in random order.

Finally, samples were taken from the fermentation process at two different times (initial and seven fermentation days), filtered and undergone to the measurement procedure. In order to perform a recovery study, real samples were spiked with glucuronic acid. This analyte was chosen for being the one at which the properties of the probiotic are attributed [2].

2.6. Genetic algorithm implementation

It is now widely accepted that multivariate calibration techniques such as partial least-squares greatly benefit from appropriate sensor selection [43–46]. Models constructed using a suitably restricted subset of spectral data rather than full spectra are extremely successful in extracting information for the prediction of analyte concentrations in samples with complex background constituents [14]. Several selection methods have been reported. They include simulated annealing [47], artificial neural networks [48], genetic algorithms [49], and extensive searches using moving window strategies [50], among a great variety of other techniques (see ref. [51] and references therein). Comprehensive searches do not allow for the selection of multiple wavelength regions, because this activity is prohibitively time consuming for spectra composed of a large number of data points like the presently studied case. Genetic algorithms are promising numerical optimization techniques which mimic natural selection processes [52–54], and are particularly appealing for avoiding such extensive searches. They are being increasingly used for variable selection in diverse areas of chemistry: mass spectrometry [55], near-infrared (NIR) [56] and mid-infrared spectroscopies [57], quantitative structure activity relationship (QSAR) [58], classification problems [59,60], selection of the best principal components for multivariate models [61], etc.

The RRGA applied in the present report, was implemented starting with a population of 40 chromosomes, initialized with 20% of all wavelengths selected at random from the full spectral range. Each gene was chosen to encode regions of five consecutive data points. The single crossover scheme with 50% probability was employed for recombination (the alternative multiple crossover procedure gave similar results), and a probability of 0.05 was applied to mutations after offsprings were produced. The algorithm was stopped after 100 generations.

In order to define the objective function to be optimized, the calibration set was randomly divided into two subsets, one used for calibration (including 70% of the calibration samples) and the other one for monitoring (the remaining samples), and the process was repeated three times using different random seeds for partitioning the calibration set. The root mean square error of prediction (RMSEP) values were then calculated for a number of factors ranging from 1 to a certain maximum (estimated from full spectral cross-validation on the complete calibration data set). In each of these three calibration/prediction procedures, the minimum error ($RMSEP_{min}$) was first considered, but the selected RMSEP was the one which was not higher than the product $F_{\alpha, I_{cal}-A, I_{mon}-A} \times RMSEP_{min}$ ($F_{\alpha, I_{cal}-A, I_{mon}-A}$ is the statistical F ratio computed for $\alpha = 0.05$ with $I_{cal} - A$ and $I_{mon} - A$ degrees of freedom, I_{cal} and I_{mon} are the number of calibration and monitoring samples in the subsets, and A the number of spectral factors) [62]. This procedure mimics the usual selection of factors by cross-validation and avoids the inclusion of a factor-depending term into the objective function. The three selected RMSEP values were then averaged, defining the figure of merit to be minimized by the algorithm. After stopping the GA, the following steps were implemented:

(1) Repeat the above calculations a number of times (20 in the present work), registering a statistical histogram of the inclusion of a given sensor range in the final top chromosome. Each time the calculation is repeated, the random partitioning of the calibration set (see above) is performed again.
(2) Select the sensor ranges included in the histogram a certain percentage of times above a threshold, for example, 70%.
(3) Perform cross-validation using the sensors obtained in step (2) in order to re-estimate the number of factors.

(4) Repeat steps (1) through (3), reinitializing the GA with the sensors obtained in step (2) and setting the maximum number of factors as obtained in step (3). Continue until the sensors selected in step (2) stabilize.
(5) Use the wavelengths selected in step (2) for PLS/ANNs model building and prediction.

3. Results and discussion

3.1. Spectral behaviour of the analytes and matrix

Fig. 1A, shows the FTIR spectra corresponding to solvent, two individual pure components (glucose and glucuronic acid) and a 7 days ferment solution, all of them collected from aqueous solutions. The concentrations of the analytes were 10 g L^{-1}. As can be seen, an intense overlapping exists between the spectra. The solvent contribution is particularly important at 3920, 3490 and 3280 cm^{-1} (O–H stretching) and 1645 cm^{-1} (H–O–H bending) [63]. For the latter reason, only the region 1300–900 cm^{-1} is useful for the quantitative analysis of the target analytes, as can be seen in Fig. 1B. Fortunately, the most intense bands present at glucose and glucuronic acid are those located between 1200 and 900 cm^{-1} due to C–O–C stretching vibrations [11].

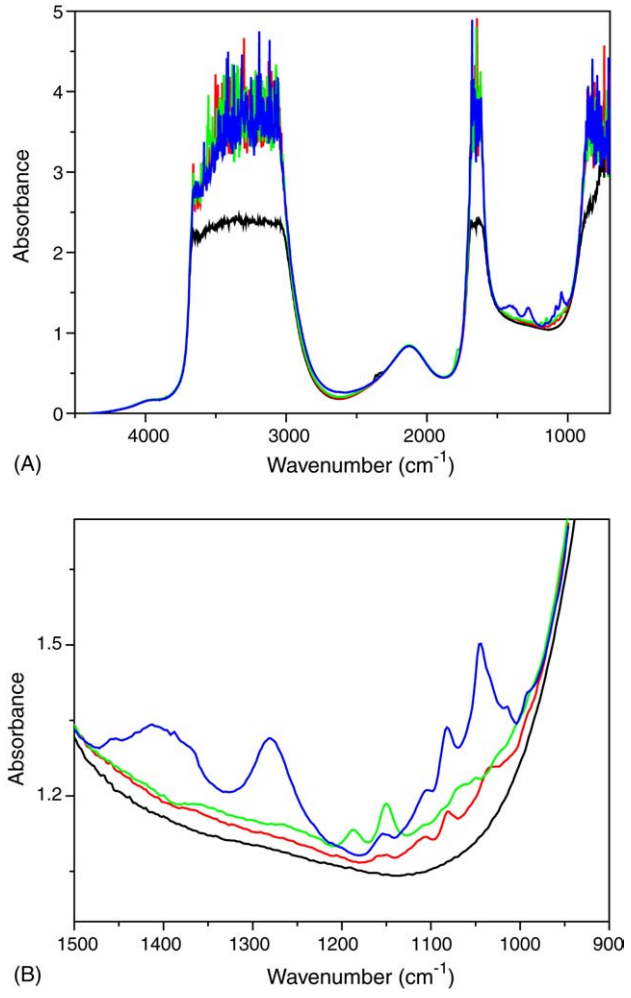

Fig. 1. (A) FTIR spectra in the complete spectral range (4500–700 cm^{-1}) corresponding to water (black solid line), water solution of glucuronic acid 10.0 g L^{-1} (green solid line), water solution of glucose 10.0 g L^{-1} (red solid line) and a 7 days ferment (blue solid line). (B) The four FTIR spectra of (A) in the working reduced range (1500–900 cm^{-1}). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

3.2. Application of multivariate tools

3.2.1. PLS-regression

PLS, a standard method for analysing multicomponent mixtures, has been systematically applied in a significant variety of samples. The presence of certain types of mild non-linearities can in principle be modelled by PLS using additional spectral factors [34]. In order to apply both PLS-1 and ANNs models, signals for the standard samples were recorded in the range 4400–450 cm^{-1}. These spectra were then subjected to PLS-1 analysis for a reduced region in which the problem of the solvent interference is alleviated (1400–900 cm^{-1}) and the one selected by using genetic algorithms. The corresponding statistical parameters for the PLS models are shown in Table 2. This table shows the spectral regions, the optimum number of factors used for calibration, the root mean square error for cross-validation (RMSECV), the relative error of prediction (REP%), and the square of the correlation coefficient for cross-validation (r^2) for the calibration set. The optimum number of factors – which allows one to model the system with the optimum data volume avoiding over fitting – was determined with the well known cross-validation procedure [12]. This procedure removes one training sample at a time and uses the remaining samples to build the latent factors and regression. As can be seen in Table 2, the calibration parameters are good enough for glucose and glucuronic acid when the GA-selected region is used for modelling, but not for gluconic acid. According to the selected regions, the glucose peak at ca. 1180 cm^{-1} and the glucuronic acid peak at ca. 1180 cm^{-1} are the most selective for the enhancement of the quality of the built models. On the contrary, for gluconic acid probably there is no sufficient sensitivity, and consequently its determination is not possible at the present conditions.

Table 3 shows the prediction results when applying the bests PLS-1 models to the nine validation samples. The obtained results for glucuronic acid (REP% = 15.5) can be considered acceptable if one takes into account the complexity of the studied system. On the other hand, bad results for glucose (REP% = 27.7) were obtained, and a completely unacceptable REP% = 122.0 for gluconic acid was found.

Fig. 2 shows the calibration spectra in the region 1400–900 cm^{-1}. This later figure also shows three spectra corresponding to real fermentation samples. Bad results were obtained when the determination of the three analytes in these samples was made. This fact could be explained considering that all the interferences were not correctly modelled when

Table 2
Spectral regions and statistical parameters for the calibration

	Glucuronic acid		Gluconic acid		Glucose	
	Total	GA	Total	GA	Total	GA
Spectral region (wavenumbers)	1400–900	1205–1165	1300–900	1082–1002	1300–900	1795–1781, 1077–1070
Factors	5	5	5	5	6	4
RMSECV	1.51	0.24	0.47	0.34	1.47	0.38
REP%	30	5.41	54.5	26.21	25.7	6.13
r^2	0.513	0.985	0.273	0.523	0.553	0.981

Table 3
Results obtained when applying PLS at the GA-selected wave numbers on the validation set

Sample	Glucuronic acid (g L^{-1})		Gluconic acid (g L^{-1})		Glucose (g L^{-1})	
	Actual	Found	Actual	Found	Actual	Found
1	2.48	2.42	0.25	0.48	2.48	2.62
2	2.48	2.13	0.49	0.66	2.48	2.06
3	2.48	2.99	0.77	1.02	2.48	2.22
4	4.96	3.89	0.25	0.98	4.96	4.21
5	4.96	4.06	0.49	0.95	4.96	4.84
6	4.96	4.85	0.77	−0.40	4.96	3.52
7	7.5	7.30	0.25	0.80	7.5	7.11
8	7.5	7.80	0.49	0.96	7.5	6.50
9	7.5	7.07	0.77	0.93	7.5	7.60
REP%		15.5		122.0		27.7

building the calibration set. But a principal component analysis (Fig. 3) showed similarity among the spectra plotted in Fig. 2. On the other hand, a residual versus PLS predicted values for all the three analytes as suggested in ref. [28] is shown in Fig. 4. In the latter figure it can clearly seen the presence of non-linearity. This fact could indicate the presence of other phenomena like non-linearities. Therefore, the use of a more robust chemometric method should be postulated.

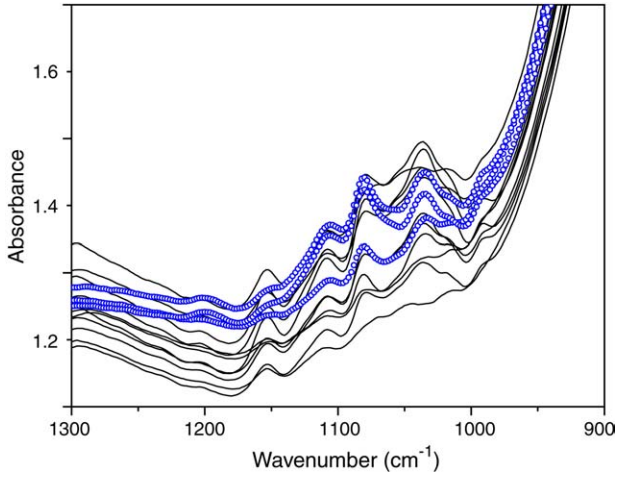

Fig. 2. FTIR spectra of calibration (black solid lines) and three real samples in the restricted region (1300–900 cm^{-1}) (blue circles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

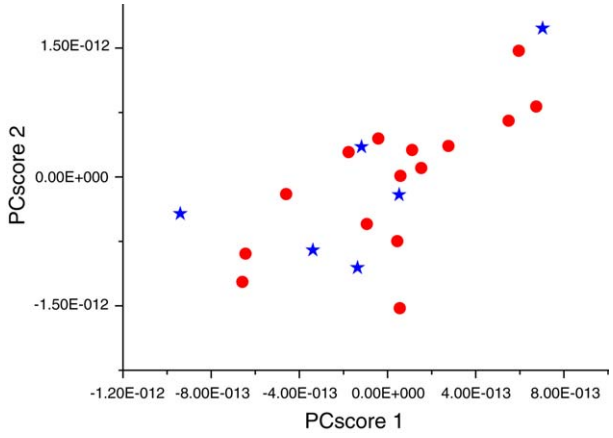

Fig. 3. Score plot for first and second principal components for the fifteen calibration samples (red circles) and real samples (blue stars). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

3.2.2. *Artificial neural networks*

ANNs are calibration methods especially created to model non-linear information, although they are also able to deal with a linear behaviour and can often improve the results in comparison with a linear model. The so-called multilayer feed-forward networks [34,64] are often used for prediction as well as for classification. In the present work we have used an ANN that consists of three layers of neurons or nodes, which are the basic computing units: the input layer with

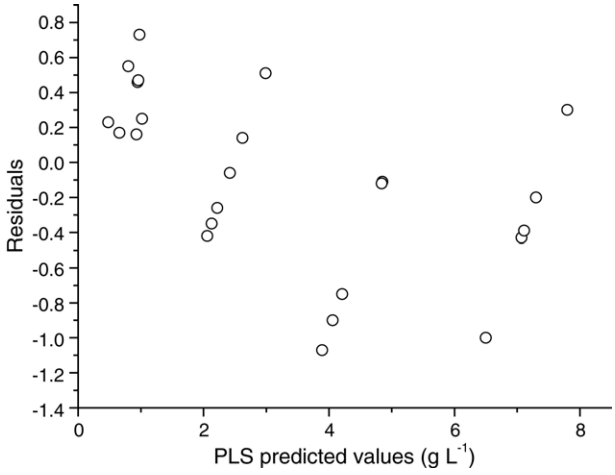

Fig. 4. Residual vs. PLS predicted values plot.

a number of active neurons corresponding to the predictor variables in regression, and one hidden layer with a number of active neurons. The input and the hidden layer numbers are optimised during training, and the output layer has just one unit. The neurons are connected in a hierarchical manner, i.e., the outputs of one layer of nodes are used as inputs for the next layer and so on. In the hidden layer the sigmoid function $f(x) = 1/(1 + e^{-x})$ is used, and the output of the hidden neuron j, O_j, is calculated as:

$$O_j = f\left[\sum_{i=1}^{m}(s_i w_{ij} + w_{bj})\right] \quad (1)$$

where s_i is the input from neuron i in the layer above, to neuron j in the hidden layer, w_{ij} are the connection weights between neurons i and j, w_{bj} is the bias to neuron j and m is the total number of neurons in the layer above. Both in the input and output layers linear functions are used. In the presently used ANN, learning is carried out through the back-propagation rule [34,64]. It is important to stress that ANNs trained with this rule have a remarkable advantage: there is no need to know the exact form of the analytical function on which the model should be built. Thus, neither the functional type nor the number of parameter in the model needs to be given [64].

One of the two calibration sets, consisting of 15 samples was used to train the ANNs. The other calibration set was used as monitoring set. The nine samples validation set was used as the test set for checking the ANNs predictive ability and for comparison between both calibration models. The number of neurons in the input hidden layers was optimised by trial and error. The finally selected architecture for the three components is displayed in Table 4. The numbers between brackets indicate how many active neurons are employed in each layer. This means that the employed architecture has two input neurons, three hidden neurons and a single output neuron for glucuronic acid. In order to find the best model, each ANN was trained with the above-mentioned training set, but it was subsequently stopped before it learns idiosyncrasies present in the training data. This was achieved by searching the minimum value of the root mean square error for the monitoring set ($RMSEP_{mon}$). On the other hand, it is important to consider that the ANN should not be overtrained, fact that can be managed by taking into account that the number of objects should not be exceed that for the adjustable weights. In the presently trained for glucuronic acid ANNs, weights were ($3 \times 4 \times 1 = 12$). These figures were obtained after considering the number of input and hidden layers plus one bias neuron on each layer. It should be noted that the latter number computed for gluconic acid is equal to 16 indicating an overtraining, but the determination of this component is not advised considering the lack of accuracy calculated in the present study.

Table 4
Statistical parameters for the ANNs models

Parámetros	Glucuronic acid	Gluconic acid	Glucose
Architecture	[2,3,1]	[3,3,1]	[3,2,1]
Spectral region (cm^{-1})	1205–1165	1082–1002	1795–1781 1077–1070
$RMSEP_{cal}$	1.24	1.13	0.879
$RMSEP_{mon}$	1.34	1.05	1.08
r_{cal}^2	0.965	0.954	0.989
r_{mon}^2	0.971	0.956	0.978
Weights	12	16	12

ANNs were applied with significant improvement of the PLS obtained results. Table 5 shows the results obtained when this non-parametric multivariate model was applied on the nine samples validation set. As can be seen the improvement on the prediction observed when ANNs are applied (ca. 50%) explains the power of this method not only in modelling non-linearities but also in solving complex and overlapped signals. The agreement between the predicted and the actual concentrations, although is not excellent for glucuronic acid and glucose, demonstrate the potential of the method to simultaneously distinguish and quantify both components in the studied concentration range.

Finally, ANNs were applied on four real fermentation samples in order to predict the acid glucuronic concentration at the beginning of the process and after seven fermentation days. Besides, these samples were spiked with known amounts of the analyte in order to compute the recovery of the

Table 5
Results obtained when applying ANNs on the validation set

Sample	Glucuronic acid ($g\,L^{-1}$)		Gluconic acid ($g\,L^{-1}$)		Glucose ($g\,L^{-1}$)	
	Actual	Found	Actual	Found	Actual	Found
1	2.48	2.75	0.25	0.24	2.48	2.95
2	2.48	2.38	0.49	0.42	2.48	2.28
3	2.48	2.94	0.77	1.06	2.48	2.99
4	4.96	5.25	0.25	0.34	4.96	5.55
5	4.96	4.87	0.49	0.27	4.96	4.55
6	4.96	4.84	0.77	0.51	4.96	4.50
7	7.5	7.23	0.25	0.70	7.5	7.03
8	7.5	7.48	0.49	0.80	7.5	7.78
9	7.5	8.53	0.77	0.62	7.5	8.69
REP%		9.7		28.0		10.0

Table 6
Results obtained on the glucuronic acid spiked real samples by using the FTIR/ANNs model

Sample	Glucuronic acid ($g\,L^{-1}$)			
	Predicted in fermentation	Spiked	Predicted after addition	Recovery (%)
1[a]	4.1 (4)	8.6	14.5	121
	4.3 (4)	8.8	14.8	119
	4.0 (4)	8.4	14.3	123
2[b]	–	10.2	12.2	120
	–	10.5	12.8	122
	–	10.3	11.5	112
2[a]	4.0 (1)	10.4	15.4	110
	4.2 (1)	11.1	16.2	108
	4.1 (1)	10.1	15.7	115
3[b]	–	10.4	11.0	106
	–	10.2	9.9	97
	–	10.4	10.5	101
3[a]	5.4 (3)	10.5	19.9	138
	5.4 (3)	10.7	18.5	122
	5.4 (3)	10.1	20.5	150
4[b]	–	10.2	12.2	120
	–	10.2	13.9	136
	–	10.2	12.6	124
4[a]	5.5 (3)	10.7	14.2	81
	5.4 (3)	10.4	12.5	68
	5.2 (3)	10.1	11.7	64

[a] Seven days of fermentation.
[b] Initial.

analytical methodology. Table 6 shows the results obtained before and after the spiking. Recoveries over 100% in most of the cases are indicative of a positive systematic error. Nevertheless, considering the complexity of the sample and the speediness and simplicity of the method, the results should be considered acceptable. In addition, the automation possibility is other advantage of the present method.

4. Conclusion

FTIR spectroscopy combined with chemometric data evaluation becomes an excellent tool for evaluating complex systems like the presently studied one. Region selection through a novel genetic algorithm was able to enhance the predictive ability of artificial neural networks. Two compounds were measured simultaneously with a reasonably good accuracy, though simplifying and speeding up the analysis.

Acknowledgments

Financial support from the Universidad Nacional del Litoral and the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) are gratefully acknowledged. The authors thank Dr. A.C. Olivieri for the ANNs MATLAB routine.

References

[1] G.W. Frank, Kombucha, first ed., Ennsthaler D., Steyr., 1993, pp. 19–46.
[2] C. Dufresne, E. Farnworth, Food Res. Int. 33 (2000) 409.
[3] P.J. Blanc, Biotechnol. Lett. 18 (1996) 139.
[4] C. Chen, B.Y.J. Liu, Appl. Microbiol. 89 (2000) 834.
[5] I. Kimura-Takagi, M. Nagaoka, H. Shibata, S. Hashimoto, S. Ueyama, R. Aiyama, T. Yokokura, Chromatography 22 (2001) 85.
[6] A. Meyer, C. Raba, K. Fischer, Anal. Chem. 73 (2001) 2377.
[7] Y.S. Fung, H.O. Tung, Electrophoresis 22 (2001) 2242.
[8] T. Soga, M. Serwe, Food Chem. 69 (2000) 339.
[9] D.L. Doak, J.A. Philips, Biotechnol. Prog. 15 (1999) 529.
[10] P. Fayolle, P. Daniel, G. Corrieu, Vib. Spectrosc. 14 (1997) 247.
[11] M. Kansiz, J.R. Gapes, D. Mac Naughton, B. Lendl, K.C. Schuster, Anal. Chim. Acta 438 (2001) 175.
[12] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1998) 1193.
[13] R.G. Brereton, Analyst 125 (2000) 2125.
[14] H. Martens, T. Naes, Multivariate Calibration, Wiley, Chichester, 1989.
[15] K.R. Beebe, B.R. Kowalski, Anal. Chem. 61 (1989) 1007A.
[16] A. Espinosa-Mansilla, A. Muñoz de la Peña, F. Salinas, M. Martínez Galera, Anal. Chim. Acta 276 (1993) 141.
[17] H.C. Goicoechea, A.C. Olivieri, Anal. Chem. 71 (1999) 4361.
[18] K. Wiberg, A. Hagman, P. Burén, S.P. Jacobsson, Analyst 126 (2001) 1142.
[19] A. Guiberteau Cabanillas, T. Galeano Díaz, M.N. Mora Diez, F. Salinas, J.M. Ortiz Burguillos, J.C. Viré, Analyst 125 (2000) 909.
[20] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. De la Pezuela, Anal. Chim. Acta 333 (1996) 147.
[21] N.R. Marsilli, M.S. Sobrero, H.C. Goicoechea, Anal. Bioanal. Chem. 376 (2003) 126.

[22] A. Marchesini, M. Wiliner, V. Mantovani, J. Robles, H.C. Goicoechea, J. Pharm. Biomed. Anal. 31 (2003) 39.
[23] B.K. Lavine, Anal. Chem. 72 (2000) 91R.
[24] P. Damiani, G. Escandar, A. Olivieri, H. Goicoechea, Curr. Pharm. Anal. 1 (2005) 145.
[25] T. Scheper, B. Hitzmann, E. Stark, R. Faurie, P. Sosnitza, K.F. Reardon, Anal. Chim. Acta 400 (1999) 121.
[26] J.W. Hall, B. McNeil, M.J. Rollins, I. Draper, B.G. Thompson, G. Macaloney, Appl. Spectrosc. 50 (1996) 102.
[27] P.J. Brimmer, J.W. Hall, Can. J. Appl. Spectrosc. 38 (1993) 155.
[28] V. Centner, O.E. de Noord, D.L. Massart, Anal Chim. Acta 376 (1998) 153.
[29] P.J. Gemperline, J.R. Long, V.G. Gregoriou, Anal. Chem. 63 (1991) 2313.
[30] S. Sekulic, M.B. Seasholtz, Z. Wang, B.R. Kowalski, S.E. Lee, B.R. Holt, Anal. Chem. 65 (1993) 835A.
[31] P.J. Gemperline, J.R. Long, V.G. Gregoriou, Anal. Chem. 63 (1991) 2313.
[32] S. Wold, N. Kettaneh-Wold, B. Skagerberg, Chemom. Intell. Lab. Syst. 7 (1989) 53.
[33] S. Wold, Chemom. Intell. Lab. Syst. 14 (1992) 71–84.
[34] F. Despagne, D.L. Massart, Analyst 123 (1998) 157R.
[35] P.J. Gemperline, J.R. Long, V.G. Gregoriou, Anal. Chem. 63 (1991) 2313.
[36] R.M. Carvalho, C. Mello, L.T. Kubota, Anal. Chim. Acta 420 (2000) 109.
[37] Q. Li, X. Yao, X. Chen, M. Liu, R. Zhang, X. Zhang, Z. Hu, Analyst 125 (2000) 2049.
[38] L. Hadjiiski, P. Geladi, P. Hopke, Chemom. Intell. Lab. Syst. 49 (1999) 91.
[39] H.C. Goicoechea, A.C. Olivieri, J. Chemom. 17 (2003) 338.
[40] MATLAB 6.0, The MathWorks Inc., Natick, Massachusetts, USA, 2000.
[41] A.C. Olivieri, H.C. Goicoechea, F.A. Iñón, Chemom. Intell. Lab. Syst. 73 (2004) 189.
[42] H.C. Goicoechea, D. Eluk, M. Kubescha, J. Ferraro, H. Miglietta, B. Rodil, V.E. Mantovani, Chem. Educator [Online] 5 (2000) 67; doi:10.1007/s008997990367a.
[43] A.S. Bangalore, R.E. Shaffer, G.W. Small, M.A. Arnold, Anal. Chem. 68 (1996) 4200.
[44] M.J. McShane, G.L. Cote, C.H. Spiegelman, Appl. Spectrosc. 51 (1997) 1559.
[45] C.H. Spiegelman, M.J. McShane, M.J. Goetz, M. Motamedi, Q.L. Yue, G.L. Cote, Anal. Chem. 70 (1998) 35.
[46] J.H. Jiang, R.J. Berry, H.W. Siesler, Y. Ozaki, Anal. Chem. 74 (2002) 3555.
[47] H. Swierenga, F. Wülfert, O.E. de Noord, A.P. de Weijer, A.K. Smilde, L.M.C. Buydens, Anal. Chim. Acta 411 (2000) 121.
[48] R. Todeschini, D. Galvani, J.L. Vilchez, M. del Olmo, N. Navas, Trends Anal. Chem. 18 (1999) 93.
[49] R. Leardi, J. Chemom. 15 (2001) 559.
[50] J. Ferré, F.X. Rius, Anal. Chem. 70 (1998) 1999.
[51] M.C. Ugulino Araújo, T.C. Bezerra Saldanha, R.K. Harrop Galvão, T. Yoneyama, H. Caldas Chame, V. Visani, Chemom. Intell. Lab. Syst. 57 (2001) 65.
[52] L. Davis, Genetic Algorithm and Simulated Annealing, Pitman, London, 1987.
[53] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1988.
[54] R. Leardi, J. Chemom. 14 (2000) 643.
[55] D. Broadhurst, R. Goodacre, A. Jones, J.J. Rowland, D.B. Kell, Anal. Chim. Acta 348 (1997) 71.
[56] Q. Ding, G.W. Small, M.A. Arnold, Anal. Chem. 70 (1998) 4472.
[57] P.A. da Costa Filho, R.J. Poppi, Quim. Nova 25 (2002) 46.
[58] K. Hasegawa, T. Kimura, K. Funatsu, Quant. Struct. Act. Relat. 18 (1999) 262.
[59] Q. Guo, W. Wu, D.L. Massart, C. Boucon, S. de Jong, Chemom. Intell. Lab. Syst. 61 (2002) 123.
[60] F. Ros, M. Pintore, J.R. Chrétien, Chemom. Intell. Lab. Syst. 63 (2002) 15.
[61] A.S. Barros, D.N. Rutledge, Chemom. Intell. Lab. Syst. 40 (1998) 65.
[62] E.K. Kemsley, Trends Anal. Chem. 17 (1998) 24.
[63] B. Stuart, D.J. Ando, Biological Applications of Infrared Spectroscopy, John Wiley & Sons, Chichester, 1997.
[64] J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design, second edition, Wiley-VCH, Weinheim, 1999.