Multivariate analysis of Raman spectra applied to microbiology: Discrimination of microorganisms at the species level

P. Mobili, A. Londero, G. De Antoni, and A. Gómez-Zavaglia
Centro de Investigación y Desarrollo en Criotecnología de Alimentos, Conicet La Plata, Universidad Nacional de La Plata, 1900 La Plata, Argentina.

C. Araujo-Andrade*, H. Ávila-Donoso, R. Ivanov-Tzonchev, and I. Moreno
Unidad Académica de Física de la Universidad Autónoma de Zacatecas, Calz. Solidaridad S/N, esq. Paseo de la Bufa. Zacatecas, 98060, Zac. México.

C. Frausto-Reyes
Centro de Investigaciones en Óptica, A.C. Unidad Aguascalientes, Ags. México.

Recibido el 19 de febrero de 2010; aceptado el 16 de junio de 2010

In this work, multivariate methods such as: principal component analysis (PCA) and partial least square-discriminant analysis (PLS-DA), were applied for the analysis and interpretation of Raman spectra, collected from microorganisms of different species. The main objective was to develop a methodology for a rapid and free of chemical-reagents discrimination and classification of microorganism at the species level. The raw Raman spectra of microorganisms were recorded in the spectral range of 2000 to 200 cm^{-1}. However, a detailed analysis of the results obtained by means of PCA, showed that the spectral region from 1700 to 1500 cm^{-1}, provides chemical and biochemical information highly correlated with the species of the microorganisms analyzed in this study, allowing a clear discrimination among species. Also, in order to evaluate the capability of multivariate methods to develop a classification rule, PLS-DA in a leave-one-out-cross-validation method (LOOCV) was used for the calibration and validation of a classification model, as a first approach. The results obtained for this method, showed an acceptable classification among the strains under study. On the other hand, taken into account the complexity of microorganisms' communities and the experimental procedures for their identification, discrimination and classification, the non-destructive and versatility of Raman spectroscopy and the capability of the multivariate methods for the analysis of spectral data, result useful tools for the classification and discrimination of this kind of samples.

Keywords: Raman spectroscopy; multivariate analysis; microorganisms.

En el presente trabajo, se describe el uso de los métodos multivariantes tales como: análisis de componentes principales (PCA) y de análisis discriminante por mínimos cuadrados parciales (PLS-DA), para el análisis e interpretación de espectros Raman, colectados de un grupo de microorganismos pertenecientes a diferentes especies. El principal objetivo de este estudio fue, sentar las bases para el desarrollo de una metodología experimental y de análisis, que permitan la discriminación y clasificación de microorganismos a nivel de especie de una forma rápida y libre de agentes reactivos. Los espectros Raman de los microorganismos, fueron colectados para su análisis en el rango de los 2000 a los 200 cm^{-1}. Sin embargo, los resultados obtenidos mediante PCA, mostraron que la región comprendida de los 1700 a los 1500 cm^{-1}, provee información química, bioquímica y estructural, estrechamente relacionada con la especie de los microorganismos analizados en este estudio, lo cual permite su discriminación. Adicionalmente, como una primera aproximación y para evaluar la capacidad de los métodos multivariantes en la clasificación taxonómica de microorganismos, la información contenida en los espectros Raman fue utilizada para calibrar un modelo mediante PLS-DA, posteriormente validado mediante el método (LOOCV). Los resultados obtenidos por la metodología empleada, muestran una clasificación aceptable en términos de la especie para las cepas analizadas en este estudio. Por otro lado, tomando en cuenta la complejidad de las comunidades de microorganismos y los métodos experimentales actuales para su identificación, discriminación y clasificación, la versatilidad y características no-destructivas de la espectroscopia Raman, así como la capacidad de los métodos multivariantes para el análisis de datos espectrales, podemos concluir que la conjunción de ambas técnicas, representa una herramienta con un gran potencial para la discriminación y clasificación de este tipo de muestras biológicas.

Descriptores: Espectroscopia Raman; análisis multivariante; microorganismos.

PACS: 87.64.K; 78.30.Er; 02.50.Sk

1. Introduction

Microbial contamination is not only a medical problem but also plays a large role in pharmaceutical clean-room production and food-processing technology. For all these fields, a fast and nonambiguous identification of pathogenic microorganisms is required. Standard bacterial identification and classification methods are generally based on their morphology, biochemical reactions and their ability to grow in various media under different conditions [1-8]. However, these methods are time consuming, and require training and expertise. For this reason, the search for alternative techniques for a quick and reliable identification, discrimination or classification of microorganisms is therefore an intensively active research field [9].

In the last decades, the use of Raman spectroscopy for biological purposes has been increased because of its great benefits such as its high sensitivity to subtle molecular (biochemical) changes, as well as its capability for non-invasive sensing. In addition, the Raman technique offers generally narrow bandwidths, minimal sample preparation and is easily interfaced to fiber-optics for remote analysis; therefore, interfacing to separation techniques or remote sensing is usually straightforward. Furthermore, the spatial resolution of Raman microspectroscopy in the low micrometer scale and its ability to probe samples under *in-vivo* conditions, allow new insights into living single cells without the necessity of using fixatives, markers or stains [10-16].

In addition, Raman spectroscopy is claiming an important position in the identification of microorganisms, particularly in the past few years. The main reasons for this aweakening are the capacity to register spectra directly from microcolonies growing on the culture plate, and the ability to measure in hydrated samples without water interference.

Furthermore, the use of chemometric methods for the analysis and interpretation of spectra, have allowed for discrimination of microorganisms at the specie level, providing clustering patterns congruent with the phylogenetic trees constructed from rDNA sequence analysis [17].

Depending on the purpose of the data analysis, and the previous knowledge of the samples under study, the chemometric methods are usually classified as: unsupervised and supervised. The unsupervised methods such as Principal component analysis (PCA), factor analysis (FA), and clusters analysis (CA), as its name suggests, does not require any previous knowledge of the samples under study, and can provide patterns, groupings, detection of outliers, etc. On the other side, the supervised methods such as multiple linear regression (MLR), principal component regression (PCR), Partial least square regression (PLS) and linear discriminant analysis (LDA) among others, require a set of well-characterized samples, and are used for pattern recognition purposes. The supervised methods always comprise a two-stage process: a) calibration and b) validation. In the calibration stage, each sample or object is identified as a member or not of a determined class or group, according to the previous characterization of the samples or objects (reference method). The goal is to calibrate a prediction/classification model (training set), which will be used to classify or predict the class or value of a set of new and non-characterized samples or objects. In the validation stage, the model calibrated is tested and validated. There are several methods to validate a model, but the most commons are:

1) those where a set of new, independent and well-characterized samples or objects are classified/predicted using the model to be tested (validation set). According to the experts, this is the best way to validate a model, and is used when there exists a sufficient quantity of samples or measures [18,19],
2) using the leave-one-cross-validation method (LOOCV). This validation method is used when the number of samples or measurements is not enough to have an independent training and validation set. In this validation method, the training set, itself, is used to validate de model, taken into account certain considerations. We will return to this point latter. [18,19].

The aim of this paper is to describe how these methods can be applied for the classification of microorganisms, in particular those with probiotic properties, which are generally associated with health benefits [20-22]. Taking into account that some probiotic products are composed by complex microbiotas containing lactobacilli, lactococci and also yeasts, the development of a method allowing a clear differentiation of potentially probiotic species results a valuable tool for quick identification of unknown strains isolated from these complex environments [23-27].

2. Materials and methods

2.1. Bacterial strains and growth conditions

Twelve strains of lactobacilli of the same species (CIDCA 83111; CIDCA 83113; CIDCA 83115; CIDCA 8321; CIDCA 8325; CIDCA 8335; CIDCA 8344; CIDCA 8345; CIDCA 8347; CIDCA 8348; ATCC 8007; JCM 5818), and four belonging to two close related species (ATCC 8287; JCM 1059; CIDCA 8322 and CIDCA 8328), were cultured in MRS broth [28] (Biokar Diagnostics, Beauvais, France) at 30°C for 48 hr.

The microorganisms were harvested in the stationary phase, collected by centrifugation (10000 g at 10°C for 10 min), washed twice with phosphate buffered saline (PBS, pH 7), lyophilized in a FD4 Heto freeze drier (Lab Equipment, Denmark) and conserved at room temperature.

2.2. Spectroscopic instrumentation and data acquisition

The Raman Spectra of the lyophilized samples were collected by placing them onto an aluminum substrate and then under a Leica microscope (DMLM) integrated to the Raman system (Renishaw 1000B). Multiple scans were conducted in different points of the sample by moving the substrate on an X-Y stage. The Raman system was calibrated using the first-order phonon of Si at 520 cm^{-1}, and further improved by use of samples of chloroform ($CHCl_3$) with bands at 261, 364 and 667 cm^{-1} and cyclohexane (C_6H_{12}) with bands at 383, 426, 801, 1028, 1157, 1265, 1347, 1443 cm^{-1}. The wavelength of excitation was 830 nm and the laser beam was focused (spot size of approximately 2.0 μm at high magnification) on the surface of the sample with a 50X objective. The laser power irradiation over the samples was 45 mW. Each spectrum was registered with an exposure of 30 seconds, two accumulations, and collected in the 2000-200 cm^{-1} range with 2 cm^{-1} spectral resolution.

Raman spectra were acquired via GRAMS software (version 3.04, Thermo Galactic, USA). Multivariate data analysis and pre-processing of data were performed on the Raman spectra, using Origin (version 6.0, Microcal^{TM}, USA) and The Unscrambler^{TM} software (version 8.0, CAMO, Norway).

2.3. Multivariate methods

Principal component analysis (PCA) and partial least square-discriminant analysis (PLS-DA) were performed over the pre-processed Raman spectra in order to evaluate

(a) the spectral differences among species of microorganisms in the PC space and

(b) to develop models allowing discrimination and classification among species.

2.3.1. Principal component analysis (PCA)

PCA is a multivariate technique that operates in an unsupervised manner and is used to analyze the inherent structure of the data. PCA reduces the dimensionality of the data set by finding an alternative set of coordinates, the principal components (PC's) [18,19]. The general form of PCA model is:

$$X = TP^T + E \tag{1}$$

where X matrix is decomposed by PCA into two smaller matrices, one of scores (T) and other of loadings (P).

PC's correspond to a linear combination of the original variables, which are orthogonal to each other and designed in such a way that each one successively accounts for the maximum variability of the data set.

In other words, PCA involves a mathematical procedure that transforms a large number of correlated variables (*i.e.* Raman shifts) into a smaller number of uncorrelated variables called principal components. Numerically this means that

$$\sum_{i=1}^{I} t_{ia} t_{ib} = 0 \tag{2}$$

where t_a and t_b are the a^{th} and b^{th} columns of T matrix, respectively, and

$$\sum_{i=1}^{I} p_{ia} p_{ib} = 0 \tag{3}$$

where p_a and p_b are the a^{th} and ^{bth} rows of P matrix, respectively.

The first principal component (PC1) accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

When PC-Scores are plotted, for example PC1 *vs* PC2 or any combination of the PC's, they can reveal relationships between samples (grouping). It is important to remember that PCA does not act in a supervised manner, meaning that each number of the groups under study is not known *a priori*. PCA provides insight into the percentage of variance explained by each PC and how many PCs should be kept to maintain the maximum information from the original data without adding noise to the current information [18,19]. In addition, when the PC-loadings are plotted as a function of the variables, the plot reveals the most important diagnostic variables or regions related with the differences founded in the data set.

2.3.2. Partial least square-discriminant analysis (PLS-DA)

PLS-DA is a version of PLS in which one or several Y-variables are modeled simultaneously, thus taking advantage of possible correlations or colinearity between Y-variables.

In PLS, the response matrix ***X*** is decomposed in a fashion similar to PCA, generating a matrix of scores, ***T***, and loadings factors, ***P***. (these vectors can also be referred to as basis vectors)

$$X = TP^T + E \tag{4}$$

A similar analysis is performed for ***Y***, producing a matrix of scores, ***U***, and loadings, ***Q***.

$$Y = UQ^T + F \tag{5}$$

The goal of PLS is to model all the constituents forming ***X*** and ***Y*** so that the residuals for ***X*** block, ***E***, and the residuals for ***Y*** block, ***F***, are approximately equal to zero. An inner relationship is also constructed that relates the scores of ***X*** block to the scores of the ***Y*** block.

$$U = TW \tag{6}$$

The above model is improved by developing the so-called inner relationship. Because latent (basis) vectors are calculated for both blocks independently, they may have only a weak relation to each other. The inner relation is improved by exchanging the scores, ***T*** and ***U***, in an iterative calculation. This allows information from one block to be used to adjust the orientation of the latent vectors in the other block, and vice versa. An explanation of the iterative method is available in the literature [29-31]. Once the complete model is calculated, the above equations can be combined to give a matrix of regression vectors, one for each component in ***Y***:

$$\hat{B} = P(P^T P)^{-1} W Q^T \tag{7}$$

$$\hat{Y} = X\hat{B} \tag{8}$$

For a discriminant analysis approach, it is assumed that a sample has to be a member of one of the classes included in the analysis. Each class is represented by an indicator variable, that is, a binary variable with a value of 1 for members of the class and 0 for nonmembers. This way, by building

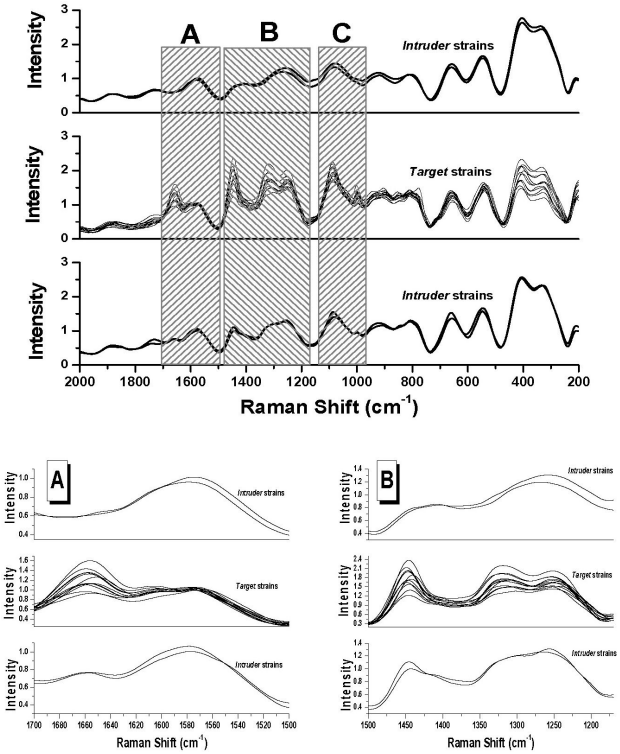

FIGURE 1. Raman spectra corresponding to all the strains of microorganisms under study (*target* strains and *intruder* strains). The gray frame shows the spectral regions where the main differences among spectra are located. A zoom view of rectangles A and B is also shown in the bottom.

a PLS model with indicator variables (***Y***), it is possible to directly predict the class membership from the ***X***-variables describing any given sample. The model output corresponds to the predicted value for an unknown sample. A correct prediction should have ideally, a ***Y*** value equal to 1 for the members of the class and 0 for the non-members. In practice, values of ≥ 0.5 are interpreted as indicating membership to the group being modeled, whereas values of ≤ 0.5, as indicating non-membership. All predicted values are accompanied by a deviation that is an estimation of how reliable the prediction is.

2.3.3. Leave-one-out-cross-validation (LOOCV)

Full-cross-validation or leave-one-out-cross-validation method consists in making as many sub-models as there are objects, each time leaving out just one of the objects and only using this for the testing. If there are ***n*** objects, each sub-model will thus be made on ***n-1*** samples. The square difference between the predicted and the ***Y***-value for each omitted sample is summed and averaged, giving the usual validation ***Y***-variance apparently in the exact same sense as for test set prediction.

LOOCV is the best one, and indeed the only alternative when there are not enough samples for a separate test set. For these reasons many text books and experts recommend full cross validation as a general approach to prediction testing,

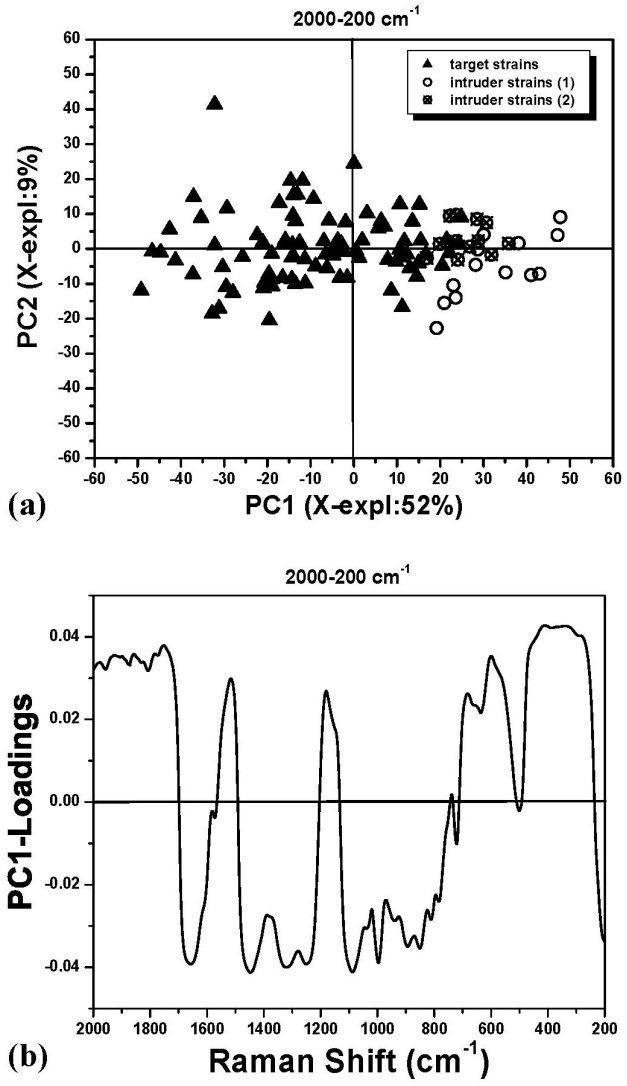

FIGURE 2. (a)Two-dimensional PCA-Scores plot of component 1 and component 2, obtained from the analysis conducted over the Raman spectra of *target* and *intruder* strains in the range from 2000 to 200 cm^{-1}. ▲ corresponds to *target* strains, ◯, to *intruder 1* strains and ⊗, to *intruder 2* strains. (b) One-dimensional PCA-loadings plot performed on the 2000-200 cm^{-1} range of the Raman spectra for the *target* and *intruder* samples under study. The most important diagnostic variables related with the spectral differences among lactobacilli species are shown.

TABLE I. X-explained variance values obtained from PCA scores analysis for raw Raman spectra in the selected ranges.

Range/ cm^{-1}	PC-number	X-explained variance (%)
2000-200	PC1	52
	PC2	9
1700-1500	PC1	72
	PC2	10
1500-1170	PC1	61
	PC2	16

claiming that this should give the most comprehensive testing of the model [18,19].

3. Results and discussion

3.1. Raman spectral features

Our first analysis was addressed in to identify the spectral features which characterize the twelve potentially probiotic strains of lactobacilli, described in the Sec. 2.1, belonging to the same species (from here and thereafter denominated in the text as *target* species), from the other four strains belonging to two close related species (denominated from here and thereafter as *intruder* species), also described in the Sec. 2.1.

Figure 1 shows the pre-processed average raw Raman spectra in the 2000-200 cm^{-1} region, of the sixteen strains belonging to the *target* and *intruder* species. The fluorescence contribution was removed from the Raman spectra by approximating a polynomial function to the spectra. This function was hand-made adjusted for a best fitting and then subtracted from the spectra. The spectral features in Raman spectra, shows clear differences among the *target* and *intruder* species in the whole range analyzed. These differences were related with the presence or absence of some bands and with the relative intensity of others.

In Fig. 1, the framed regions correspond to the main differences observed in the Raman spectra (Scripts A, B and C). Features corresponding to *target* strains were clearly different from the others strains belonging to the *intruder* species. These differences involved mainly the following bands: nucleic acid bases ring stretchings and amide I (script A), CH_2 and CH deformations (script B) and, amide III and the various oligo- and polysaccharides of the cell wall (script C) [32,33].

Some minor differences were also observed among the *target* strains, especially in the regions A and C of Fig. 1, which were assigned to the amide III (*ca.* 1250 cm^{-1}) and amide I (*ca.* 1660 cm^{-1}) bands of proteins, respectively [32,33].

The differences between *intruder* species were less evident and were mainly related with the shape and relative intensities of the bands belonging to the three regions indicated in Fig. 1.

As explained in previous section, Raman spectra are also a source of abundant chemical and structural information. In order to take advantage of this information, a multivariate analysis was carried out over the raw Raman spectra.

3.2. Principal component analysis (PCA)

PCA was performed over the Raman data sets, with the aim of comparing in an unsupervised manner the inherent structure of the spectral data in terms of similarities and differences.

PCA was conducted over the raw Raman spectra on the range (2000-200 cm^{-1}). Figure 2a depicts the PCA-scores plot of PC2 *vs* PC1 obtained from this analysis. A separation among *target* and *intruder* species can be observed along the PC1-axis, which explains the 52% of the total variance in the data set. In spite of that, some overlap of samples was also observed. In Fig. 2b are plotted the PC1-loadings values, which represent the regions of the Raman spectra where the differences among species are more evident. According to Fig. 2b, the main differences among the microorganisms' species analyzed, are in agreement with those observed and depicted in Fig. 1, localized in the 1500-1170 and 1700-1500 cm^{-1} regions.

TABLE II. Data set preparation and results of performance of PLS-DA.

Parameter	**Raman data /cm^{-1}**		
Range	2000-200	1500-1170	1700-1500
N^{o}Samples (*target*/*intruder*)	16 (12/4)	16 (12/4)	16 (12/4)
N^{o}Spectra collected	119	119	119
N^{o}Spectra removed of the analysis.	4	4	4
Mathematical treatment	Raw	Raw	Raw
Pre-processing	BLR^a, MSC^b, VN^c	BLR^a, MSC^b, VN^c	BLR^a, MSC^b, VN^c
PLS-factors	2	4	2

^{a}BLR: Base line remove. ^{b}MSC: Multiplicative scatter correction. ^{c}VN: Vector normalization.

TABLE III. Results obtained from PLS-DA over the different Raman ranges analyzed.

Range	**2000-200 cm^{-1}**		**1700-1500 cm^{-1}**		**1500-1170 cm^{-1}**	
	Calibration	Validation	Calibration	Validation	Calibration	Validation
$\mathbf{r}^2$	0.77	0.72	0.88	0.87	0.87	0.85
RMSEC/RMSEP	0.26	0.29	0.19	0.20	0.20	0.22
SEC/SEP	0.26	0.29	0.19	0.20	0.20	0.22

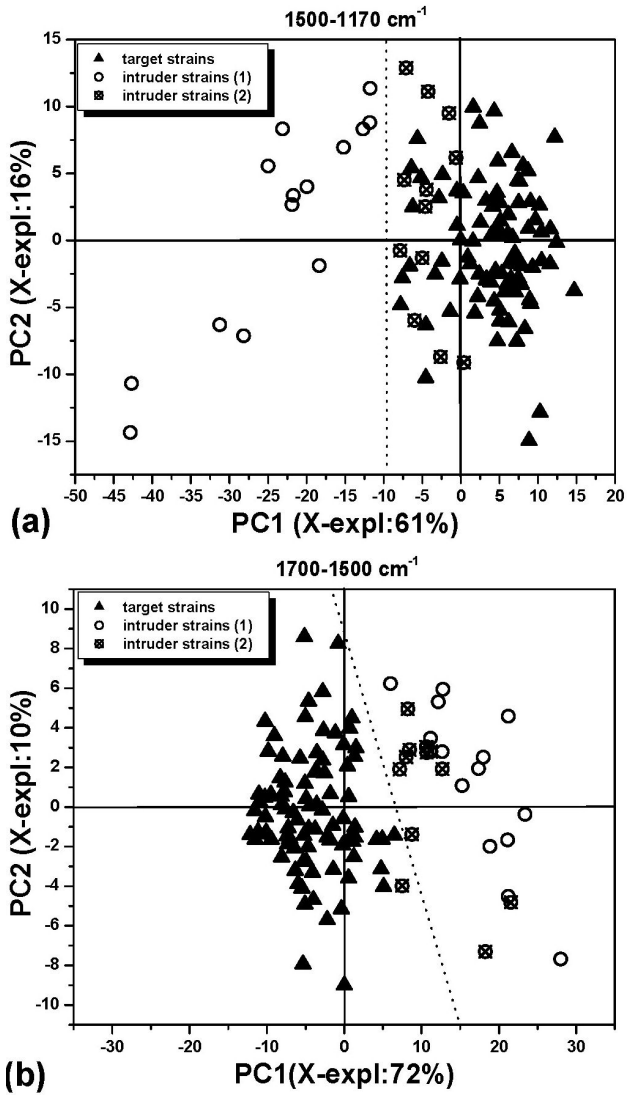

FIGURE 3. Two-dimensional PCA-Scores plots performed on the Raman spectra of the lactobacilli samples under study: **(a)** in the 1500-1170 cm^{-1}region. (**b)** in the 1700-1500 cm^{-1} region. ▲ corresponds to *target* strains, ◯, to *intruder 1* strains and ⊗, to *intruder 2* strains.

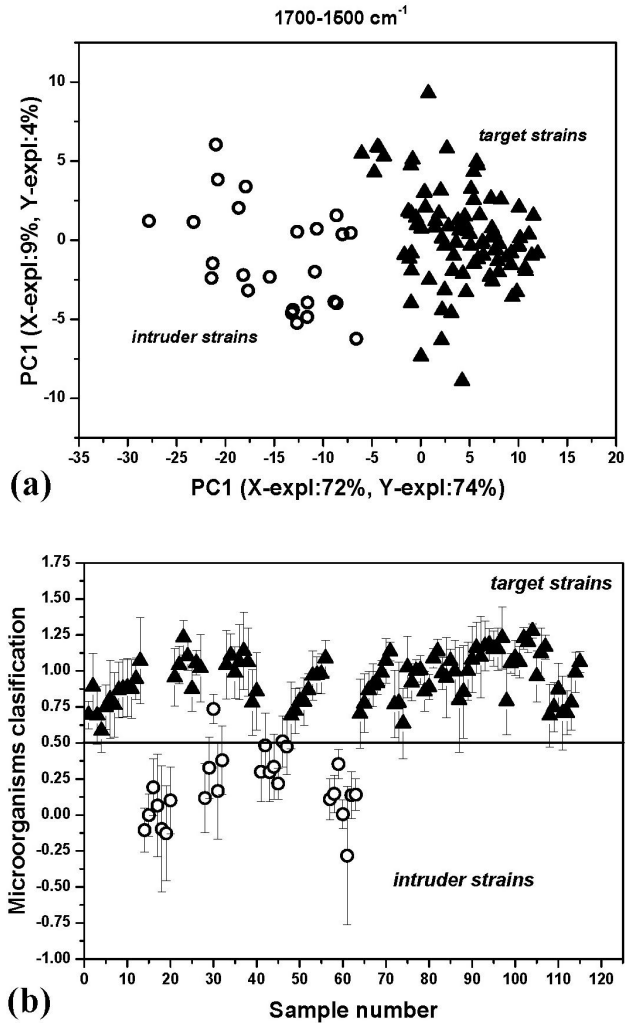

FIGURE 4. (a) Discriminant scores plot of PC2 *vs* PC1 in the 1700-1500 cm^{-1} range. (b) Leave-one-out-cross-validation, for the classification model calibrated with the Raman spectra in the range of 1700-1500 cm^{-1}. ▲ corresponds to *target* strains, ◯, to *intruder* strains.

In order to evaluate if these differences were specific to the microorganisms' species, PCA was conducted independently in both ranges (1500-1170 and 1700-1500 cm^{-1}). Fig. 3a depicts the PCA-scores in the 1500-1170 cm^{-1} region and Fig. 3b, the ones in the 1700-1500 cm^{-1} region. In Fig. 3a, an evident separation between one of the *intruder* species (identified with the hollow circle) and *target* strains, can be observed along the PC1-axis, which explains 61% of total variance. However, the other *intruder* species (identified with the cross-circle), cluster together with the *target* strains. This behavior can be explained taking into account the spectral similarities between the *target* strains and one of the *intruder* species in the amide II and amide III regions (1500-1170 cm^{-1} region) depicted more clearly in the zoom in Fig. 1 [32,33].

In Fig. 3b, discrimination between *target* strains and *intruder* species can be observed along PC1-axis, which explains 72% of variance. Taking into consideration the percentage of explained variance in the X-data set for PC1 (Table I) and the PCA-scores for each region analyzed, it is evident that the best region to build a model aiming to discriminate among *target* and *intruder* samples, is the one of 1700-1500 cm^{-1}. In this range, the main differences were observed in the bands at 1580 cm^{-1} (arising from the adenine and guanine ring stretching) and at 1660 cm^{-1} (arising from amide I vibrations) [32,33].

In addition, the results obtained, also suggest that PC1 in the 1700-1500 cm^{-1} range, is the component containing species specific information. Since the Amide I band (arising from the C=O stretching vibrations) is the predominant band in this region, these results indicate that the discrimination between *target* and *intruder* strains is somehow influenced by the protein structure. Further studies should investigate

in which way the secondary structure of proteins determine these species differences.

3.3. Partial least square-discriminant analysis (PLS-DA)

PLS-discriminant analysis was used to develop the classification rules for unknown samples in real applications.

As it was explained before, this method operates in a supervised manner, meaning that a prior knowledge of the class membership is required. In this study, two classes: *target* and *intruder* species, were defined. A correct prediction should have ideally, a Y value equal to 1 for the samples belonging to the *target* samples and 0, for the *intruder* samples.

The PLS-DA analysis was carried out over the preprocessed raw Raman spectra in a full cross validation method (leave-one-out-cross validation) due to the limited number of samples [18,19].

The analysis was performed using the raw Raman data set. The spectral range used to calibrate the classification model was that among 1700-1500 cm^{-1}, which according with the PCA-scores, is the best region to discriminate among *target* and *intruder* species. On the other hand, according to the PCA results, the 2000-200 cm^{-1} and 1500-1170 cm^{-1} regions, contains an excess of non-species specific information that does not allow a clear discrimination between both groups, however two classification models were calibrated for these two regions, in order to compare the classification ability of each model calibrated in these three regions. The number of samples, number of spectra collected, number of spectra removed from the analysis, range, mathematical treatment, data pre-processing and PLS- factors are shown in the Table II.

Figure 4a depicts the discriminant scores plot of PC2 *vs* PC1 in the 1700-1500 cm^{-1} range, which revealed a clear separation between *target* and *intruder* samples. The values obtained using the prediction model calibrated for this range in a full-cross-validation, are shown in Fig. 4b In spite of the occurrence of 4 sample mismatches (three *intruder* strains were classified as uncertain and one, as *target* strain), a good discrimination between *target* and *intruder* groups was observed. The results obtained from PLS-DA over all the Raman ranges analyzed are summarized in the Table III.

4. Conclusion

In this work, an approach based on Raman spectroscopy in combination with PCA and PLS-DA was developed for rapid differentiation of microorganisms with probiotic properties. In this sense, the development of a supervised multivariate method allowing the discrimination of microorganisms at a species level is certainly the main achievement of this work. This approach can be considered as a very promising tool for the discrimination and classification of unknown samples.

The predictive ability of the method used in this work allowed the classification of lactobacilli strains with the only requirement of being a pure strain.

Taking into account the complexity of certain microorganisms' communities and the influence of the culture conditions in the microorganisms' composition, the identification of a given species (*target* species) in such a complex environment definitely saves time and efforts.

Finally, the use of Raman spectroscopy and chemometrics shows great potentiality for the identification and classification of any bacterial species in the food industry and those isolated from different environments. This way, the use of this successful approach on complex system would exploit the advantages of using Raman spectroscopy to detect and differentiate microorganisms.

Acknowledgements

This work was supported by CONACYT-CONICET (México, Argentina) (bilateral project res. resolución N^o 962/07-05-2009), CONACYT under Project: Apoyos para investigadores nacionales para el fortalecimiento de actividades de tutoría y asesoría de estudiantes de nivel licenciatura 2008, PROMEP under Project: Apoyo a la incorporación de nuevos PTC, folio UAZ-PTC-092, Agencia Nacional de Promoción Científica y Tecnológica (Project PICT/2006/68), Argentina, CYTED (Ciencia y Tecnología para el Desarrollo) Network P108RT0362. AGZ is member of the research career CONICET (National Research Council, Argentina). GDA is member of the research career CIC-PBA (Commission for Scientific Research, Buenos Aires). A.L. and P.M. are doctoral and postdoctoral fellows from CONICET, respectively.

∗. Telephone: +52 (492)9241314 ext. 2201; Fax: +52 (492) 39 407; e-mail: caraujo@fisica.uaz.edu.mx

1. E.W. Koneman, S.D. Allen, W.M. Janda, P.C. Schreckenberger, and W.C. Win, *Color atlas and textbook of diagnostic microbiology*, 5th edn. (Lippincott-Raven, New York, 2001).
2. S.F. Al-Khaldi and M.M. Mosoba, *Nutrition* **20** (2004) 32.
3. Y.W. Tang *et al., J. Clin. Microbiol.* **36** (1998) 3674.
4. Y.W. Tag, G.W. Procop, and D.H. Persing, *Clin. Chem.* **43** (1997) 2021.
5. D.N. Fredricks and D.A. Relman, *J. Clin. Microbiol.* **36** (1998) 2810.
6. K.B. Mullis and F.A. Faloona, *Methods Enzymol*. **155** (1987) 49.
7. P. Grabar and P. Burtin, *Immunoelectrophoresis* (Elsevier, Amsterdam. 1964)
8. M.L. Bishop, J.L. Duben-Engelkirk, and E.P. Fody, *Clinical chemistry: principles, procedures and correlations*, 4th edn. (Lippincott, Williams and Wilkins, London. 2000).

9. S.P. Mulvaney and C.D. Keating, *Anal. Chem.* **72** (2000) 153R.
10. C. Araujo Andrade, J.L. Pichardo Molina, G. Barbosa Sabanero, C. Frausto Reyes, and A. Torres Lopez, *Journal of Biomedical Optics* **12** (2007) 034006.
11. C. Araujo-Andrade, J.L. Pichardo-Molina, G. Barbosa-Sabanero, and C. Frausto-Reyes, *Proc. SPIE* **6863**, (2008) 68630P.
12. G.J. Puppels *et al.*, *Nature* **347** (1990) 301.
13. C. Sandt, T. Smith-Palmer, J. Pink, L. Brennan, and D. Pink, *Journal of Applied Microbiology* **103** (2007) 1808.
14. M. Harz *et al.*, *The Analyst* **130** (2005) 1543.
15. N.P. Ivleva, M. Wagner, H. Horn, Reinhard, and C. Haisch, *Anal. Bioanal. Chem.* **393** (2009) 197.
16. S. Jimenez Sandoval, *Microelectronic Journal* **31** (2000) 419.
17. C. Kirshner *et al.*, *J. Clin. Microbiol.* **39** (2001) 1763.
18. K.H. Esbensen, in: *Multivariate Data Analysis - In Practice*, 5th Edition (CAMO Process AS, Esbjerg, Denmark, 2005); ISBN : 82-993330-3-2
19. H. Martens and T. Naes, *Multivariate Calibration* (Wiley, Chichester, England, 1989).
20. C.J. Saloff Coste, *kefir* (Danone World Newsletter, No. 11. 1996).
21. G.L. Garrote, A.G. Abraham, and G.L. De Antoni, *J. Food Prot.* **63** (2000) 364.
22. M.A. Golowczyc, P. Mobili, G.L. Garrote, A.G. Abraham, and G.L. De Antoni, *Int. J. Food Microbiol.* **118** (2007) 264.
23. G. Ottogalli, A. Galli, P. Resmini, and G. Volonterio, *Annali di Microbiologia ed Enzimologia.* **23** (1973) 109.
24. R.C. Witthuhn, T. Schoeman, and T.J. Britz, *Int. J. Dair. Technol.* **57** (2004) 33.
25. I. Molska, I. Moniuszco, K.A.M. Komorows, and V. Merilanen, *Acta Alimentaria Polónica.* **9** (1983) 80.
26. A. Zourari and E.M. Anfantakis, *Le Lait.* **68** (1988) 373.
27. G.L. Garrote, A.G. Abraham, and G.L. De Antoni, *Lebensmitt. Wiss. Technol.* **30** (1997) 77.
28. J.C. de Man, M. Rogosa, and M.E. Sharpe, *J. Appl. Bacteriol.* **23** (1960) 130.
29. P.C. Hansen, *SIAM* (Philadelphia, 1998).
30. P. Geladi and B.R. Kowalsky, *Anal. Chim. Acta* **185** (1986) 1.
31. P. Geladi and B.R. Kowalsky, *Anal. Chim. Acta* **185** (1986) 19.
32. J. De Gelder, K. De Gussem, P. Vandenabeele, P. De Vos, and L. Moens, *Anal. Chim. Acta* **585** (2007) 234.
33. K. Maquelin *et al.*, *J. Microbiol. Methods* **51** (2002) 255.