Optical spectroscopy and multivariate analysis of biomedical optics

J.L. Pichardo-Molina and C. Araujo-Andrade
Centro de Investigaciones en Óptica, A.C.,
Loma del Bosque No.115, Col. Lomas del Campestre, 37150, León, Gto. México.

G. Gutiérrez-Juárez
Instituto de Física de la Universidad de Guanajuato,
Apartado Postal E-143, 37150 León, Gto. México.

Recibido el 9 de mayo de 2007; aceptado el 5 de marzo de 2008

Although some optical spectroscopy methods were introduced more than ten decades ago, they are still finding new applications in many areas of science. Specifically, medicine and biology are two areas of research where optical methods may facilitate and improve the study and characterization of tissue and biological molecules in order to improve medical diagnosis. Optical spectroscopy can aid in the study and detection of some diseases faster than standard laboratory techniques. This work demonstrates applications of Micro-Raman spectroscopy and multivariate analysis to biomedical problems such as: breast cancer detection, toxoplasmosis study through indirect antibody detection, and discrimination between antibody isotypes (IgG and IgM).

Keywords: Colostrum; multivariate; NIR; Raman; serum.

Algunas espectroscopias ópticas existen desde hace más de diez décadas; sin embargo, en recientes años estas técnicas se han aplicado en distintas áreas del conocimiento. Por otro lado, la medicina y la biología son dos áreas de investigación donde la óptica puede facilitar el estudio y caracterización de tejidos y moléculas biológicas con el fin de mejorar el diagnóstico clínico. Las espectroscopias ópticas pueden ayudar en el estudio y detección de algunas enfermedades de manera mas rápida que las técnicas estándar de laboratorio. En este trabajo, mostramos algunas aplicaciones de la espectroscopia Raman y del análisis multivariante en algunos problemas específicos del área biomédica, por ejemplo: detección de cáncer de mama, diagnóstico de la toxoplasmosis e identificación de isotipos (IgG e IgM).

Descriptores: Calostro; multivariado; NIR; Raman; suero.

PACS: 87.64.Je; 89.20.-a; 87.14.-g

1. Introduction

Medicine and biotechnology are two areas where the constant development of new and more sophisticated techniques is required in order to make better and faster decisions related to clinical diagnosis. Physicians base their diagnoses principally on the use of three kinds of information: clinical, laboratory, and image analysis. Consequently, the new proposed techniques should improve upon the current procedures for early detection of disease.

The medical community is not only interested in techniques that improve the clinical diagnosis; they are also interested in noninvasive techniques that are less expensive, and especially in those techniques that do not require a sophisticated technician. In the last three decades, the most promising noninvasive techniques for characterization and analysis have been related to the use of visible or near infrared light. Light has been used to investigate materials and to obtain corresponding information due to the interaction of radiation with the material under study. Optical techniques are based primarily on the simple idea that light passes through the material in small quantities (transmission, diffusion, reflection, or dispersion) and emerges with information about the material through which it has passed (tissue, serum, etc.).

Following this direction, optical techniques such as absorption (UV-Vis, NIR, MIR), fluorescence, reflectance (UV-Vis, NIR, MIR), Raman, and Micro-Raman spectroscopies have shown great potential for biomedical applications, for example glucose determination [1–6], tissue and biological sample characterization [7-10], monitoring of cell proliferation [11], quantitative analysis of serum samples [12], measurement of carotenoids in the skin and retina [13, 14], measurements of brain activity [15], to assess neural activation during object processing in infants [16], studies on immunoassays [17], *in vivo* disease diagnosis [18], breast cancer detection in tissue or in serum blood [19–23], and also alcohol testing and beverage identification have all been studied using optical spectroscopies and multivariate techniques [24, 25].

Raman and Micro-Raman spectroscopy provide information on molecular structure by means of the normal vibrational modes of the molecule under study. The interaction of the light with the sample causes an energy exchange between photons and molecules, and this interaction causes the scattering of the photons, in which the majority change direction (Rayleigh scattering), but only a few (1 in 1×10^8 photons) change in frequency (Raman scattering) [26, 27]. In addition, infrared spectroscopy obeys Beer's law, by which light is absorbed through the sample and the transmitted or reflected light is detected at the same frequency as the incident light.

The principal objective of this paper is to show potential applications of optical spectroscopies, specifically Raman spectroscopy, to biomedical problems.

In the next section, we describe the Raman spectroscopy technique applied in the analysis of biological samples. Section 2.1 describes the experimental setup, the protocol for sample preparation, and the acquired data. Raman spectra (RS) were analyzed through multivariate analysis (three multivariate techniques were used in this study), and the methods are described in Sec. 2.2; for more details on these methods, we encourage the reader to consult the corresponding references. In Sec. 3, three different biomedical problems are discussed, including the discrimination between of serum samples of healthy subjects and breast cancer patients, antibody detection, and antibody isotype identification.

Our preliminary results show how Raman spectroscopy can be a useful technique in the study of biomedical problems.

2. Materials and methods

2.1. Micro-Raman experimental setup

A linearly polarized radiation of 514.5 nm from a 2.6W water-cooled argon laser (Spectra Physics, Stabilite 2017) was used as an excitation source. The laser light was focused on the sample with a 40X microscope objective. RS were recorded with a monochromator (Jobin Yvon, HR 460) equipped with an air-cooled CCD (256 pixels $\times$ 1024 pixels). Grams software (version 3.04) was used to obtain the spectra. To reject Rayleigh emission light and plasma frequencies of the laser, a holographic Super Notch-Plus filter (Kaiser Optical Systems, HSPF-31453) and an interference filter (Melles Griot, 03 IFS 004), respectively, were used. The Raman system was calibrated using the 520 cm^{-1} Raman line of a silicon wafer.

To collect the RS of the biological samples, we utilized the following protocol: a drop of each sample was placed on an aluminum substrate; after evaporation, the solid residues were analyzed in different zones. Several RS were acquired in each drop, where each selected zone was studied in detail using a microscope. All samples were analyzed on the same day, and under for same experimental setup and conditions.

2.2. Multivariate methods

To analyze the spectral data obtained, multivariate methods were used. These methods have been developed to deal with a large and complex amount of information in which two or more variables are analyzed simultaneously. These methods are used in many different fields of research from optical spectroscopy to neuroscience.

Today, multivariate methods are used to analyze data in qualitative and quantitative applications. In the following sections, we describe the multivariate methods used in the present study.

2.2.1. Principal component analysis

Principal component analysis (PCA) is a multivariate technique that acts 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 (PCs) [28–30]. Mathematically, PCA is a linear transformation,

$$PC = XW, \tag{1}$$

where the rows of the matrix X represent each RS and the columns of the matrix transformation W are the loading vectors, while the columns of PC represent the new set of variables called "scores".

PCs are 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, when spectroscopic data are analyzed, each spectrum contains a large number of variables, in this case, the Raman frequencies. The principal goal of PCA is to obtain the information about the structure of the spectroscopic data, looking for differences between samples in such way that it is possible to group the data.

In fact, when PC scores are plotted, for example PC1 vs PC2 or any combination of the PCs, they can reveal relationships between samples (grouping). It is important to remember that PCA does not act in a supervised manner, which means that PCA does not know a priori the number of kinds (groups) of samples under study. PCA provides insight into how much variance is explained by each PC, and how many PCs should be kept in order to maintain the maximum information from the original data without adding noise to the current information [31, 32]. From the perspective of physical or chemical information, when PC loadings are plotted as a function of the variables (Raman frequencies in this case), the plot reveals which variables account for the greatest differences, in other words, which bands show the greatest differences between RS of the samples in study. Raman bands are related to the normal vibrational modes of the molecules in the samples, and vibrational modes can be associated with one or several specific markers.

2.2.2. Linear discriminate analysis

Linear discriminate analysis (LDA) is a multivariate technique that acts in a supervised manner, meaning that we know *a priori* how many groups there are and which samples correspond to each group. Sometimes, though not as a general rule, the PC scores are analyzed with LDA. Then, as in the PCA method, LDA reduces the dimensionality of the data set by finding an alternative set of coordinates termed "canonical components", or DAs. DAs are linear combinations of the original variables (PC scores). The alternative set is obtained through maximizing the variance between the samples of different groups and minimizing the variance between samples

of the same group [29, 33, 34]. When the canonical component scores are plotted, they reveal relationships existing between the samples, such as natural clustering of the data. This technique provides insight into how effective a pattern recognition algorithm is in classifying the data.

2.2.3. Partial least square regression

Partial Least Square Regression (PLSR) was developed in the 1960's by Herman Wold as an econometric technique, but some of its most avid users are chemical engineers and chemometricians. Partial Least Square (PLS) is a method for constructing predictive models when the factors are highly collinear. The PLS method emphasizes the prediction of the responses, and not the understanding of the underlying relationship between variables. PLS uses factor analysis to compress the size of the spectra and to remove redundant information. Generally, PLS uses information about the property of interest (*i.e.*, the analyte concentration) along with sample variance in the compression process to create factors that are correlated with the property of interest. However, when prediction is the goal and there is no practical need to limit the number of measured factors, PLS is a highly recommended tool [32, 35–37].

3. Micro-Raman applications

In the following subsections, we present the principal results of some applications to biomedical problems using Micro-Raman spectroscopy; these results are based on current projects in our laboratory. In each case, the medical considerations, data acquisition, and data analysis are discussed.

3.1. Serum analysis of breast cancer patients

Breast cancer is the most common form of malignant tumor found among women in the western world. Each year, new cases of breast cancer are detected around the world, and according to the American Cancer Society, 178,000 new cases of breast cancer will be diagnosed for 2007 with a mortality rate of 26.9 per 100,000 women [38]. In México, the number of breast cancer cases per year are not well documented; however, the National Institute of Statistics Geography and Informatics (INEGI) reported 11,242 new cases of breast cancer in 2002 with a mortality rate of 15.2 per 100,000 women [39]; thus, approximately 7,600 Mexican women die each year due to this disease. Early detection can increase the chances of survival for patients, and new efficient or complementary techniques can be helpful to physicians dealing with this difficult task.

The main objective of this study was to evaluate the capability of Micro-Raman spectroscopy and multivariate methods in discriminating between serum samples of patients with breast cancer and control subjects(healthy subjects).

3.1.1. Serum samples

We collected serum samples of eleven patients with confirmed clinical and histopathological diagnosis of breast cancer (including four patients with metastases) and twelve healthy control subjects. The serum was obtained by centrifuging the sample of blood obtained from each volunteer.

All patients and the control group were from the central region of Mexico and had similar ethnic and socioeconomic backgrounds. Raman spectra were collected using the experimental setup and protocol described above. Prior to the multivariate analysis, a baseline was corrected from Raman spectra using the commercial software Microcal Origin, in order to eliminate the fluorescence contribution of each spectrum. Next each spectrum was smoothed using the adjacent averaging method with ten points for the averaging, and then normalized by applying the maximum normalization transformation.

3.1.2. Breast cancer serum results

Figure 1 shows the mean Raman spectra of the control group and breast cancer patients. The Raman spectra were analyzed visually in order to find evident spectral differences between spectra, our main purpose being the identification of new bands or shift bands. However, after the visual analysis was completed, we concluded that the differences come principally from changes in the relative changes in Raman intensities. Next, the mean values of intensity of each band were calculated by group, and the results showed mean values of breast cancer always lower (in some bands one-third lower) than the mean value of controls. The mean value of Raman intensity can be useful for identification of breast cancer patients if the standard deviations of both groups do not overlap. However, after the standard deviation was calculated for each band, our hypothesis was refuted because Raman intensity by itself does not permit us to use the mean values to discriminate between RS. Nevertheless, to confirm if these changes in Raman intensity have significant statistical differences, the intensity of all Raman frequencies was analyzed using an independent t-test (by two groups) where $p < 0.05$ means that there are significant differences in intensity between the two kinds of samples on specific bands. Figure 2 shows a RS showing those spectral regions where significant differences are observed between the two groups of RS. The black line indicates those bands where $p > 0.05$, and the red line means those bands where $p < 0.05$. As can be seen, only small spectral ranges fail to present statistical differences between these two groups. This result means that there are real differences between the RS of serum samples, but it is still necessary to use a better analysis to find those bands that provide information related to the problem under study (breast cancer).

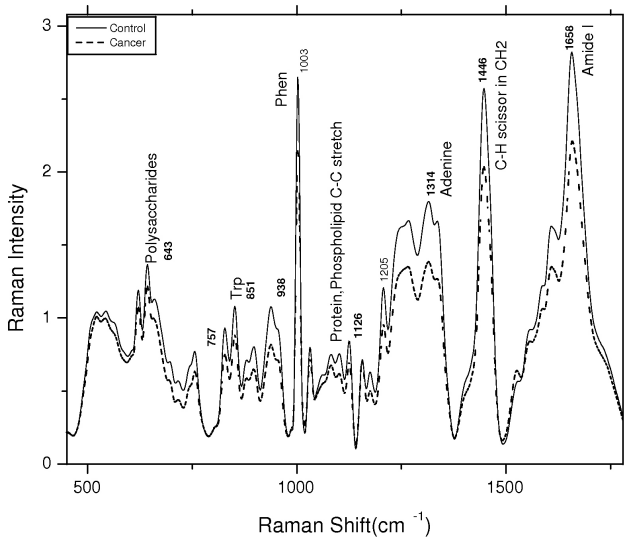

FIGURE 1. This figure shows the Raman spectra of control (continuous line) and breast cancer groups (dotted line). The major difference between these two spectra is due to the changes in Raman intensity. In fact, the mean Raman intensity of breast cancer patients is lower than that of controls, and this fact can be observed at 1658, 1446, 1314, and 1003 cm^{-1}.

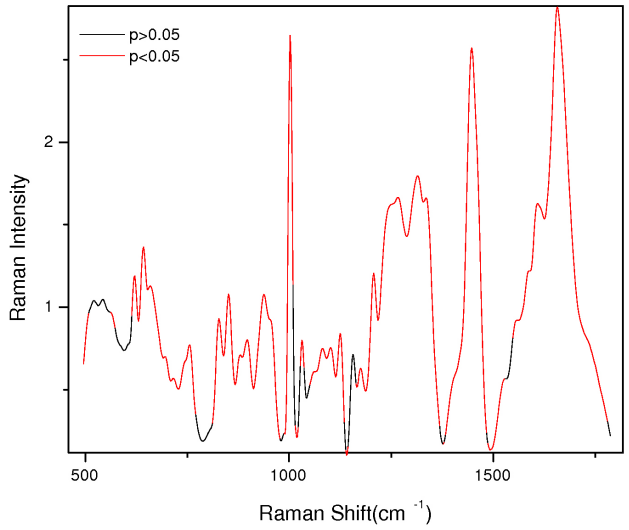

FIGURE 2. A statistical t-test was used to determine which regions of Raman spectra are statistically significant. In this figure, a characteristic Raman spectrum of a serum sample was plotted to show the results of the t-test in the following form: the continuous line represents those regions where p values are lower that 0.5, while the dotted line corresponds to those regions where p values are greater than 0.5. We observe that only small regions of the spectra fail to show significant differences between the spectra of control and breast cancer patients.

The principal goals of this study are:

1) to explore whether it is possible to discriminate the RS and decide which subjects are healthy or which have breast cancer;
2) to determine which bands make this discrimination possible, and identify those biological markers that may be associated with breast cancer.

The t-test showed that relevant information is contained in the Raman intensity, and this information should be analyzed using new methodologies that permit us to make the discrimination and also to identify those bands that contribute most to this discrimination. For this reason, multivariate analysis (MM), PCA and LDA algorithms were applied to the full range of the data set.

As was mentioned above, PCA is a MM that permits the user to reduce the number of variables while preserving the maximum original information. Then, after PCA was applied to RS data, the results show that, of the original variables, 720 wavelengths can be replaced by a new set of ten new variables. Also, the first significant loadings were plotted as a function of Raman frequency. This new plot shows a spectrum similar to the original Raman spectrum: in this case, the intensity of the peaks represent those frequencies where there are huge differences between the spectra of the two groups. In this case, the bands correspond to 731(no assignation), 851 (protein, Tyr), 1002 (Phe), 1157 (beta carotene, C-C skeletal stretch), 1318 (adenine), 1338(Trp, adenine, α helix, and phospho- lipids), 1450 (β sheet and phospho-lipids), 1523 (beta carotene), and 1656 cm^{-1} (C=O stretch, Trp, adenine, phospholipids).

After PCA was conducted for variable reduction, the LDA algorithm was used to analyze the new variables (PCs), and a cross-validation method was used in order to find the relationship between groups of samples. A cross-validation method is used when the number of samples is small to divide the data into groups, one to build the model (trainee data), and one more to prove the model (testing data). In cross-validation, a portion of the data is set aside as training data, leaving the remainder as testing data. In this approach one samples (testing data) at a time was left out.

Figure 3 shows the results obtained from the LDA, where hollow circles represent the samples corresponding to breast cancer patients, triangles represent those corresponding to the control group, full circles denote spectra corresponding to patients with metastasis, and the continuous line represents a borderline decision.

As can be seen, discrimination between the two groups is given by the first component, LDA1, and both groups have similar dispersion over this component. However, when component LDA2 is considered, it is clear that the breast cancer group has a greater dispersion, but it is hard to determine the real variable that produces this dispersion.

This methodology demonstrates that the combination of Raman spectroscopy and multivariate analysis has a sensitivity of 0.97 and specificity of 0.78. The conclusion of this section is that Raman spectroscopy was able to detect molecular markers correlated with the presence or absence of breast cancer in sanguineous serum, and a the protocol was able to identify the samples from breast cancer patients with a sensitivity of 0.97. These preliminary results need to be confirmed

with a large number of subjects. For more details from this study, we refer the readers to the Refs. 23.

3.2. Study of specific antibodies in colostrum samples

Antibodies are the principal mode of defense of our body against extra-cellular agents such as bacteria, viruses, and exotoxins. Antibodies are serum proteins that aid in the neutralization of pathogens or antigens in order to protect us. Detection of a patient antibody response to a pathogen is often the only means of diagnosing an infection. Knowing the type of antibody response that the patient is currently producing can help ascertain whether the infection is ongoing or resolved. We are particularly interested in the study of human colostrum, because colostrum is the principal source of antibodies that protect a newborn. There are many serological tests used to detect and identify specific antibodies in human fluids. However, these tests require prior sample preparation, and require considerable time to obtain results [40–48].

The goal of this study is to show the feasibility of using Raman spectroscopy and PCA as an auxiliary tool for analyzing colostrum samples and to identify those patients that have been in contact with *Toxoplasma gondii* (*T. gondii*).

3.2.1. Colostrum samples

Human colostrum samples were obtained from a group of pregnant women from the central region of Mexico with similar socioeconomic and ethnic lifestyles. For this study, we chose 11 colostrum samples, of which five were negative and six were positive for the presence of antibodies IgG, IgM, and IgA anti-T. gondii. Each colostrum sample was tested by an indirect ELISA test at the Institute of Medical Research of the University of Guanajuato. For the colostrum samples, a total of 165 Raman spectra were obtained from the 11 samples of colostrum; 75 spectra to negative samples and 90 spectra corresponded to positive samples. The Raman spectra were collected using the experimental setup and protocol described in Section 2.1.

3.2.2. Results of antibody detection

The Raman spectra of colostrum samples show small differences between positive and negative, and these are observed primarily due to changes in intensity and in the shift of several specific bands, such as 1681, 1162, 950, and 886 cm^{-1}. Figure 4 shows the mean Raman spectra. The band shift is on the order of the resolution of the experimental system, and therefore it is difficult to use this information as a classifier parameter in order to discriminate between patients exposed to this parasite from those patients that were not exposed. The principal goal of this section is to explore whether it is possible to use a chemometric technique such as PCA to discriminate between samples. As a first approximation, PCA was conducted on the set of raw data (RS): however, the results did not reveal a clear discrimination between samples, apparently because the small differences did not provide enough information. Nevertheless, a second derivative of RS was analyzed by PCA using the full cross-validation method. The derivative is a common transformation in spectroscopy, and is used to enhance differences among spectra, to resolve overlapping bands in qualitative analysis, and most importantly, to reduce the effects of interference from scattering, matrix, or other absorbing compounds in quantitative analysis. The second derivative of each spectrum was obtained based on the

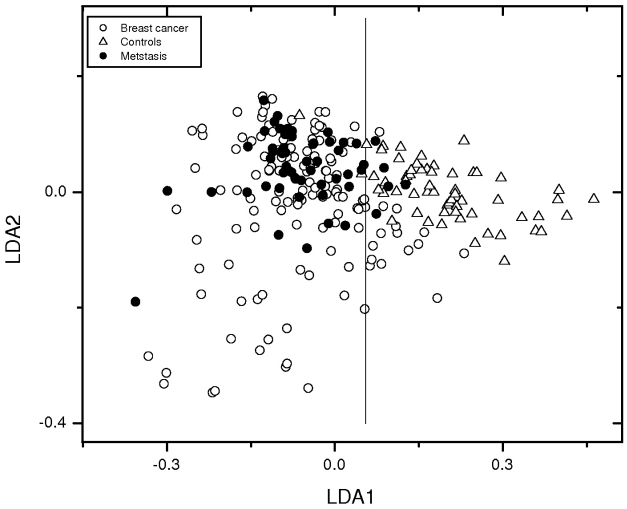

FIGURE 3. The two-dimensional PCA scatterplot graphically shows the discrimination between controls and breast cancer patients. The control group is represented by triangles and is distributed on the left side of the plot, while the breast cancer group is represented by circles and occupies the right side. Also, the plot shows that the distribution of the breast cancer patients has a huge dispersion in the PCA plane as compared with the controls. This indicates that the spectra of the control group have fewer variations in the measured parameters such as Raman intensity and Raman shifting.

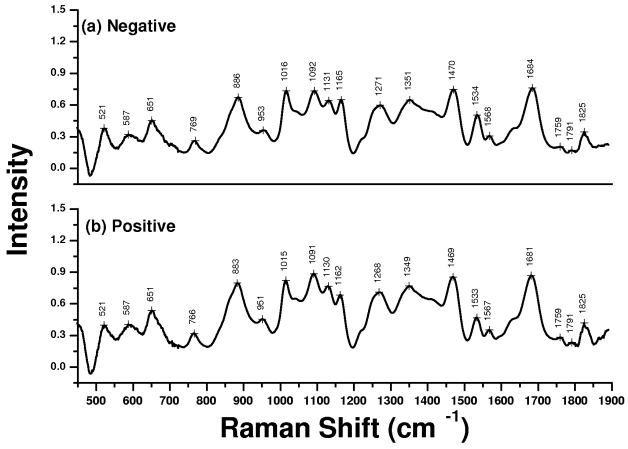

FIGURE 4. Mean Raman spectra of colostrum samples, a) negative samples and b) positive samples (IGg, and IGm). The band shift and variations in intensity are the major differences between positive and negative samples for the T. gondii-specific antibody. Some of the bands that present a shift, taken as a reference for the negative samples, are located at 886, 950, 1162, and 1681 cm^{-1}.

Savitzki-Golay algorithm using a second-order polynomial function with a five data-point window. The small changes in Raman spectra due to the presence of the specific antibodies is enhanced by the first and second derivative.

The new PCA analysis conducted on the second derivatives showed an improved discrimination between positive and negatives samples of colostrum. For this case, five PCs explained the maximum variance of the data set. After a careful revision of the PCA score plots, we found that PC2 is the component that explains the major differences associated with the absence or presence of anti-T. gondii antibodies. Figure 5a and 5b shows the two-dimensional score plots of PC1 vs PC2 and PC2 vs PC3. These plots show a good separation between positive and negative samples. In addition, subgroups can be observed in each group; when RS on the PC space are projected on the first component, an overlap of positive and negative samples can be observed, and the reason for the overlap is that positive and negative samples have common information. In fact, the first loading vector as a function of the wavenumber shows average information of the variations in the two spectra, and means that component one does not have information related to the specific antibodies.

As a matter of fact, when second loading vectors were plotted as a function of Raman frequency, higher variation between the two groups of samples was observed. Six bands marked the greatest differences, 1119, 1172, 1195, 1513, 1542 and 1558 cm^{-1}, and corresponded to (C-N), Tyr + Phe, Trp, aromatic amino acids, Tyr and amide II, respectively. These bands may be associated with those molecules that conform to the antibodies. In fact, P.C. Painter and J.L. Koening [49] studied human and rabbit antibodies IgG and IgM, identifying those principal bands of IgG, where some of these bands corresponded to those observed by PCA, 1119, 1172, and 1558 cm^{-1}.

Colostrum samples were studied using Raman spectroscopy and multivariate methods to determine which samples correspond to patients with anti-T-gondii antibodies. Also, multivariate PCA permitted the identification of those bands that show the greatest variance between RS. Good discrimination was obtained by analyzing the second derivative of the Raman spectra of colostrum samples; however, it was necessary to conduct additional testing for a more complete study to determine if the collected Raman spectra information offered direct or indirect information related to the presence of specific antibodies. For full information about this work, we refer the reader to the Ref. 50.

3.3. Antibody isotype identification

The goal of this section is to show the capacity of Raman spectroscopy and multivariate analysis for discriminating between antibody isotypes. This last application relates to the ability of Raman spectroscopy and multivariate analysis to distinguish between serum samples from volunteers who were in contact with *Toxoplasma gondii*, and to generate specific antibody isotypes, specifically IgG and IgM against T. gondii.

The study was performed using the RS data of Subsection 1 of Sec. 3.2, but for the present analysis, only those RS corresponding to positive samples were analyzed. Three samples were positive to IgG anti-T. gondii, and three samples positive to IgM anti-T. gondii. As was mentioned previously, each sample was first analyzed with an ELISA test to determine the isotype.

For this study, we analyzed the second derivative of ninety RS from positive samples, 43 RS corresponding to samples with IgG, and 47 corresponding to samples with IgM. In this case, the multivariate analysis was conducted using PCA and LDA, as described in Section 2.4.

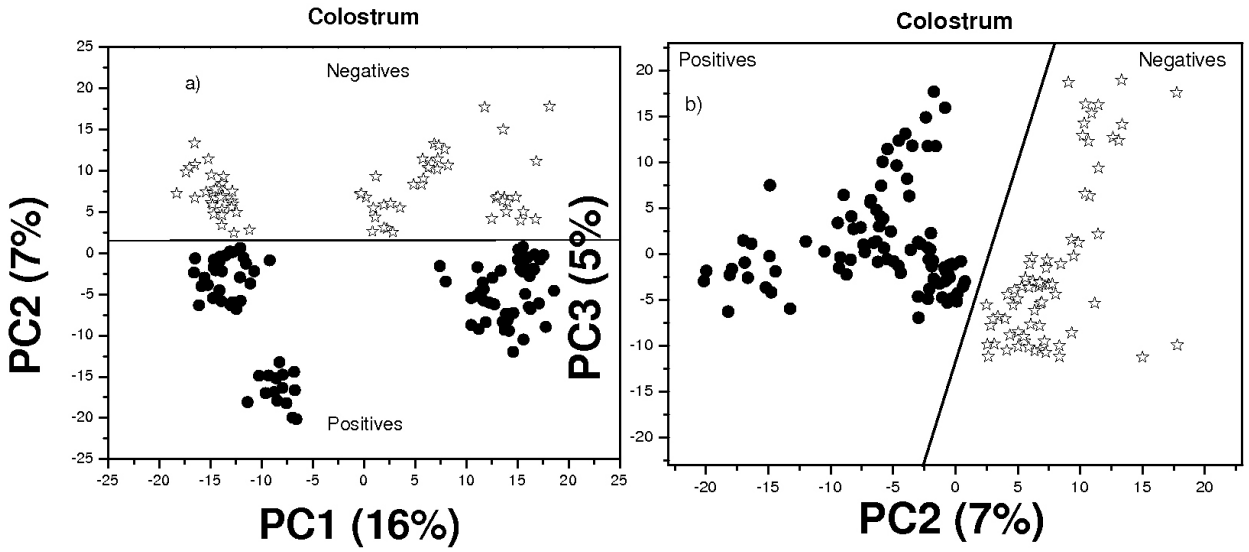

FIGURE 5. PC2 is the component that permits discrimination between positive and negative samples, demonstrating that the PC1 component does not have information related to the presence of specific antibodies.

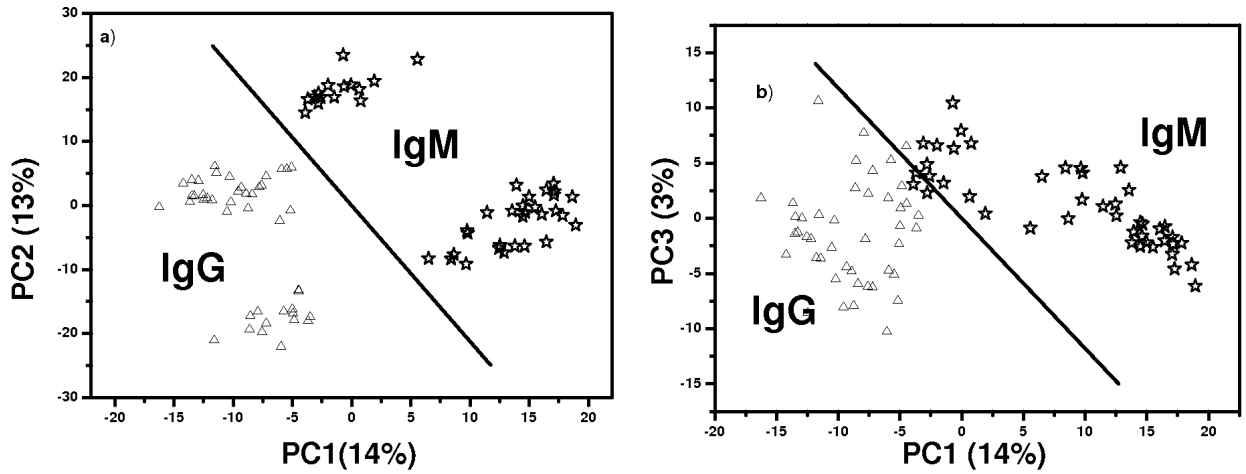

FIGURE 6. Two-dimensional plots of PC1 Vs PC2 and PC1 Vs PC3 show discrimination between specific isotypes IgG and IgM T-gondii; the positive and negative samples are arranged in two groups in each case. However, only the first component, PC1, retains the information that permits discrimination between the isotypes.

3.3.1. Isotype identification results

According to the measurement protocol and the experimental setup described in the sections above, the second derivatives of the 90 Raman spectra of six positive colostrum samples were analyzed using PCA and LDA, 43 RS corresponding to IgM samples and 47 to IgG samples.

The score plots show a good discrimination between IgG and IgM antibodies to anti-*T. gondii*, as can be seen in Fig. 6a and 6b. The PCA analysis shows that three PCs are the optimum number needed to explain the maximum variances in the data set, where PC1 explains 14%, PC2 explains 13%, and PC3 explains 3%. The line shown in the score plots is the calculated borderline decision. After analyzing the score plots, we concluded that PC1 is the component that explains the differences between antibody isotypes, and the loading plot of PC1 shows that too many slight differences exist as band shift and intensity variation along the entire spectral range between the IgG and IgM spectra.

However, at this moment, the number of samples analyzed for antibody isotype is small. Therefore, we intend to increase the number of samples to see if those bands observed are completely responsible for discrimination of the isotypes, and to decide if the intensity and the shift of these bands are related to antibody isotypes. Nevertheless, the results seem very promising for the identification of isotypes using a nondestructive technique that does not require sample preparation.

4. Conclusions

Raman and infrared spectroscopies open up new possibilities in clinical diagnosis, especially with the rapid evolution of optoelectronic devices. Many recent studies have demonstrated the feasibility of optical spectroscopies in many applications of medicine and biotechnology. Currently, Raman spectroscopy has been used in the study and identification of breast cancer biopsies or serum samples, detection of calcified atherosclerotic tissue, and has identified differences between Alzheimer's disease and normal temporal gray cortex. In this work, our team was engaged in the study of biological fluids in order to study breast cancer and antibody detection. Our results show that Raman spectroscopy and multivariate techniques are very promising methods for discrimination between control subjects and breast cancer patients. It was also possible to identify the presence of specific antibodies and to discriminate between two isotypes. The potential of these techniques opens up new possibilities for noninvasive techniques for clinical application in the near feature.

Acknowledgements

The authors wish to thank the Consejo Nacional de Ciencia y Tecnología (CONACYT) and the Consejo Nacional de Ciencia y Tecnología de Guanajuato(CONCyTEG) for financial support under Grant No. 54467 and Anexo 07-04-K662-80A01.

1. K.J. Jeon, in D. Hwang, S. Hahn, and G. Yoon, *J. Biomed. Opt.* **11** (2006) 014022-1.
2. J.L. Lambert, C.C. Pelletier, and M. Borchert, *J. Biomed. Opt.* **10** (2005) 031110-1.
3. M. Meinke, I. Gersonde, M. Friebel, J. Helfmann, and G. Müller, *Appl. Spectros.* **59** (2005) 826.
4. C.C. Pelletier, J.L. Lambert, and M. Borchert, *Appl. Spectros.* **59** (2005) 1024.
5. L. Zhang, G.W. Small, and M. Arnold, *Anal. Chem.* **75** (2003) 5905.

6. G. Yoon, A. K. Amerov, K. J. Jeon, and K. Yoen-Joo , *Appl. Opt.* **41**(2002) 1469.
7. P. Matousek *et al.*, *Appl. Spectros.* **60** (2006) 758 .
8. K. Maruo *et al.*, *Appl. Spectros.* **60** (2006) 441.
9. N.K. Afseth, V.H. Segtnan, and J.P. Wold, *Appl. Spectros.* **60** (2006) 1358.
10. E. Salomatina, B. Jiang, J. Novak, and A.N. Yaroslavsky, *J. Biomed. Opt.* **11** (2006) 064026-1.
11. K. W. Short, S. Carpenter, J.P. Freyer, and J.R. Mourant, *Biophys. J.* **88** (2005) 4274.
12. D. Rohleder *et al.*, *J. Biomed. Opt.* **10** (2005) 031108-1.
13. I.V. Ermakov, M. Sharifzadeh, M. Ermakova, and W. Gellermann, *J Biomed. Opt.* **10** (2005) 064028-1.
14. W. Gellerman and P.S. Bernstein, *J. Biomed. Opt.* **9** (2004) 75.
15. R.N. Aslin and J. Mehler, *J. Biomed. Opt.* **10** (2005) 011010-1.
16. T. Wilcox, H. Bortfeld, R. Woods, E. Wruck, and D.A. Boas., *J. Biomed. Opt.* **10** (2005) 011010.
17. J. Choi *et al.*, *J. Biomed. Opt.* **9** (2004) 221.
18. J.T. Motz *et al.*, *J. Biomed. Opt.* **10** (2005) 031113.
19. P.R. Bargo *et al.*, *J. Biomed. Opt.* **10** (2005) 034018.
20. E.B. Hanlon *et al.*, *Phys. Med. Biol.* **45** (2000) R1.
21. R.R. Alfano *et al.*, *Lasers in Life Sci.* **4** (1991) 23.
22. N. Stone, C. Kendall, N. Shepherd, P. Crow, and H. Barr, *J. Raman Spectrosc.* **33** (2002) 564.
23. J.L. Pichardo-Molina *et al.*, Published on line in *Laser Med. Sci.* **22** (2007) 229.
24. T.D. Ridder, S.P. Hendee, and C.D. Brown, *Appl. Spectros.* **59** (2005) 181.
25. O. Barbosa-Garcia *et al.*, *Spectrochim. Acta. A* **66** (2007) 129.
26. M.J. Pelletier, *Appl. Spectros.* **57** (2003) 20A.
27. S. Jiménez-Sandoval, *Microelectron. J.* **31** (2000) 419.
28. R.G. Brereton, *Chemometrics, Data analysis for the laboratory and chemical plant* (Wiley, New York, 2003).
29. C.M. Bishop, *Pettern Recognition and machine learning* (Springer, Singapore ,2006).
30. I.T. Jolliffe, *Principal componnet analysis* (Springer, New York, 2004).
31. S. Wold, K. Esbensen, and P. Geladi, *Chemomet. Intell. Lab. Syst.* **2** (1987) 37.
32. H. Martens and T. Næs, *Multivariate Calibration* (Wiley, Chichester, UK, 1989).
33. T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning* (Springer, New York, 2001)
34. T. Isaksson and A.H. Aastveit, *Classification methods*, Edit. JM Chalmers,and PR Griffiths, Vibrational spectroscopies (John Wiley & Sons, Chichester, 2002).
35. C. Araujo-Andrade, I. Campos-Cantón, J.R. Martínez, G. Ortega Zarzosa y F. Ruiz, *Rev. Mex. Fís.* **51** (2005) 67.
36. A. Lorber, L.E. Wangen, and B.R. Kowalski, *J. Chemometr.* **1** (1987) 19.
37. P. Geladi and B.R. Kowalski, *Anal. Chim. Acta.* **185** (1986) 1.
38. http://www.cancer.org/downloads/STT/CAFF2007PWSecured.pdf
39. http://www.generoysaludreproductiva.gob.mx/
40. A.B. Sabin and H.A. Feldman, *Science* **108** (1948) 660.
41. B.C. Walton, B.M. Benchoff, and W.H. Brooks, *Am. J. Trop. Med. Hyg.* **15** (1966) 149.
42. A. Balsari *et al.*, *J. Clin. Pathol.* **33** (1980) 640.
43. K. Hedman, M. Lappalainen, I. Seppala, and O. Makela, *J. Infect. Dis.* **159** (1989) 736.
44. O. Liesenfeld, J.G. Montoya, S. Kinney, C. Press, and J.S. Remington, *J. Infect. Dis.* **183** (2001) 1248.
45. J.G. Montoya, O. Liesenfeld, S. Kinney, C. Press, and J.S. Remington, *J. Clin. Microbiol.* **40** (2002) 2504.
46. B.R. Dannemann, W.C. Vaughan, P. Thulliez, and J.S. Remington, *J. Clin. Microbiol.* **28** (1990) 1928.
47. J.S. Remington, R. McLeod, P. Thulliez, and G. Desmonts, J. S. Remington and J. Klein, eds., W.B. Saunders, (Philadelphia, 2001) p.205.
48. P. Thulliez, F. Daffos, and F. Forestier, *Scand. J. Infect. Dis. Suppl.* **84** (1992) 18.
49. Paul Painter and J.L. Koenig, *Biopolymers.* **14** (1975) 457.
50. C. Araujo-Andrade, J. L. Pichardo-Molina, G. Barbosa-Sabanero, C. Frausto-Reyes and A. Torres-López, *Journal of Biomedical Optics* **12** (2007) 034006-1.