Theor Chem Acc (2011) 130:1221–1229
DOI 10.1007/s00214-011-0972-2

REGULAR ARTICLE

Diagnosis of inflammatory lesions by high-wavenumber FT-Raman spectroscopy

Luis Felipe das Chagas e Silva de Carvalho · Érika Tiemi Sato · Janete Dias Almeida · Herculano da Silva Martinho

Received: 27 September 2010 / Accepted: 31 May 2011 / Published online: 25 June 2011

Abstract The Raman spectroscopy technique has been extensively used for biological sample characterization. In particular, the fingerprint spectral region (800–1,800 cm^{-1}) has been shown to be very promising for optical biopsy purposes. However, limitations for the widespread use of Raman-based optical biopsy technique still persist. For example, fluorescence when one uses visible light (400–700 nm) spectral sources is often present and appears to affect the mid-IR/Raman region more than the high-wavenumber region (2,800–3,600 cm^{-1}). But, both the higher wavenumber spectral region and the mid-IR/Raman region can be fluorescence-free when one uses lasers sources, which do not cause fluorescence, for example, 1,064, 830 or 785 nm sources. In addition, the Raman spectral signal of inflammatory infiltrates can influence the biopsy diagnoses and is one important source of misdiagnosis of normal versus pathological tissues. The present work seeks to evaluate whether the Raman spectra in the high-wavenumber spectral region can be used to distinguish between oral inflammatory fibrous hyperplasia (IFH) lesions and normal (NM) tissues and hence be used as a new diagnostic tool. Thirty spectra of oral IFH lesions and NM tissues from biopsies of 12 patients were analyzed using both principal components analysis (PCA) and a binary logistic regression (BLR) model. It was found that the high-wavenumber region Raman spectra can be used to discriminate between NM tissue and oral IFH tissues by analyzing the 2,800–3,050 cm^{-1} (CH_2 and CH_3 vibrations of lipids and proteins) and 3,050–3,600 cm^{-1} (CH, OH, and NH vibrations of proteins and water) spectral intensities. A simple classification model based on the relative areas of the above cited regions resulted in concordant pairs of 95.3%. Considering the standard errors in the model parameters, it was found that the sensitivity (*Se*) and specificity (*Sp*) fall in the interval $87\% < Se < 100\%$ and $73\% < Sp < 93\%$, respectively. In addition, it has been found that the Raman scattering cross-sections in the NH, OH, and CH stretching region are more intense than in the mid-IR/Raman (fingerprint) region.

Keywords Raman spectroscopy · Optical biopsy · Inflammatory lesions

Abbreviations

BLR	Binary linear regression
FT	Fourier transform
IFH	Inflammatory fibrous hyperplasia
NIR	Near infrared
NM	Normal
PCA	Principal component analysis
Se	Sensitivity
Sp	Specificity
UV	Ultraviolet

1 Introduction

Among common diseases, cancer is one of the most devastating. When detected in its early stages, it is well known

Dedicated to Professor Akira Imamura on the occasion of his 77th birthday and published as part of the Imamura Festschrift Issue.

L. F. das Chagas e Silva de Carvalho · É. T. Sato · H. da Silva Martinho (✉)
Centro de Ciências Naturais e Humanas, Universidade Federal do ABC—UFABC, Santo André, São Paulo 09090-400, Brazil
e-mail: herculano.martinho@ufabc.edu.br

L. F. das Chagas e Silva de Carvalho · J. D. Almeida
Departamento de Biociências e Diagnóstico Bucal, Faculdade de Odontologia de São José dos Campos—FOSJC, Universidade Estadual Paulista Júlio de Mesquita Filho—UNESP, São José dos Campos, São Paulo 12245-000, Brazil

that its treatment can be very successful. The number of worldwide cancer deaths is projected to increase to 22 million in 2030 influenced in part by an increasing and aging global population. In 1975, it was estimated that the global cancer burden was 5.0 million (see page 14 of Ref. [1]). In most developed countries, cancer is the second highest cause of death after heart disease, and epidemiological studies point to this trend emerging in the less developed world. Already more than half of all cancer cases occur in developing countries. Cancer prevention and early detection are essential components of all cancer control plans because many deaths can be avoided by changing one's lifestyle (no smoking, limiting one's alcohol in take, exercising regularly, and getting enough sleep every night) and having regular physical checkups and cancer screening. Hence, there is a need for more research in the development of new more cost effective and earlier diagnosis methods and techniques. New approaches to obtain early diagnosis of various types of cancer, and breast cancer in particular, are essential for the long-term survival of patients with a high risk for certain types of cancers. Genetic tests are being developed, but they are not cheap, and it is not obvious that they should be used for the general population. Hence, one needs to develop relatively cheap prescreens which can determine which patients should be subsequently genetically tested.

New developments in medical imaging technologies, such as ultrasonography, computer tomography, and magnetic resonance imaging, have improved the quality, selectivity, sensitivity, and quickness of the cancer diagnosis. However, the gold standard method for discriminating normal and altered tissues still is the histopathological analysis performed on a biopsy. Considerable time and cost are required to obtain the biopsy sample, and the specialized knowledge of a pathologist is needed for diagnosis. In order to overcome these problems and costs, various attempts to achieve a minimally invasive clinical diagnosis of cancers have been made, such as those using biomarkers (see Chap. 4 of Ref. [1]).

The currently available screening and diagnostic methods have their shortcomings that make fast, effective, and efficient detection of cancer in the general population impossible, most importantly in developing countries. These limitations are due to the fact that all methods are based on subjective interpretations of morphological abnormalities. For example, the cytological analysis of the cervix, which is the elementary principle of the Pap smear and colposcopy for cervical cancer diagnosis, has a high false negative rate of $\sim 50\%$ [2]. This fact is inherent to the subjective histological grading of pathologies. It is common that those atypical cells may be associated with inflammatory infiltrates and that the slice may be misinterpreted by the pathologist as simple inflammation instead of a malignancy or neoplasticity or vice versa. Data from the World Cancer Report 2008 (see Chap. 4 of Ref. [1]) revealed that tests with the best overall serum biomarker (prostate-specific antigen) presented 90% sensitivity and 25% specificity. The same resource indicated that sensitivity (Se) and specificity (Sp) for the Pap smear test that detects cervical intraepithelial cancer fall in the 47–62% and 60–95% ranges, respectively. Oral cytology is one of the most accurate conventional methods with a sensitivity of 92% and a specificity of 94%.

Recently, there has been developed new diagnosis tools based on photonic technology. One of these tools is the optical biopsy. Optical biopsy refers to techniques where the light-tissue interaction is analyzed and information concerning the state of the tissue is obtained both "in vivo" or "ex vivo". Optical spectroscopy techniques such as infrared absorption, fluorescence, and Raman scattering are also employed. Holmstrup et al. [3] described many molecular interaction features in cells and tissues that cannot be accessed by conventional histopathology that can be probed by optical techniques. For the purposes of the present work, it is important to introduce a general comparison of the Se and Sp for several optical biopsy methods and conventional histopathological methods as well. Lieber et al. [4] demonstrated the application of Raman spectroscopy and its auto-fluorescence-free background for pediatric Wilms' tumor diagnosis. The fluorescence-free background spectra were able to discriminate normal kidney from Wilms' tumor with Se = 81% and Sp = 100%. The Raman spectra obtained Se = 93% and Sp = 100% specificity. Mo et al. [5] studied the high-wavenumber Raman spectroscopy spectral region and concluded that the diagnostic algorithms based on principal components and linear discriminant analysis together with the leave-one-patient-out cross-validation method furnished Se = 93.5% and Sp = 97.8% for dysplasia tissue identification. Backhaus et al. [6] developed a simple and rapid method for the detection of breast cancer with IR-spectroscopy. With cluster analysis (a method of unsupervised learning), they were able to achieve a Se = 98% and a Sp = 95%. Griebe et al. [7] analyzed biological markers which play an evolving role in the diagnosis of Alzheimer disease (AD). The FT-IR spectroscopy data showed Se = 88.5% and Sp = 80%. Another example of a new tool for use in optical diagnosis is elastic scattering spectroscopy, which is a point-contact technique where one collects broadband optical spectra sensitive to absorption and scattering within the tissue. Using this technique, Austiwik et al. [8] obtained Se = 69% for detection of clinically relevant metastases and Sp = 96%.

Raman spectroscopy techniques are of special interest due to their high sensitivity in the detection of biochemical and molecular variations in tissues [9]. Usually, the

spectral region between 500 and 1,800 cm^{-1} (fingerprint region) has the most relevant biochemical information concerning biological tissues [9, 10]. Many of vibrational band frequencies of amino acids, nucleic acids, proteins, lipids, glucose, and other carbohydrates fall in this region [11]. This spectral window was extensively studied by Fourier transform-Raman (FT-Raman) (1,064 nm laser excitation) and near infrared (NIR) Raman (785 and 830 nm laser excitation) techniques [10–12]. The FT-Raman spectra of biological samples are almost auto-fluorescence-free, minimizing the need for preprocessing the data [12]. The auto-fluorescence emitted by tissues severely interferes with the discrete spectra when the excitation falls in the visible region [13]. This problem can be minimized by using excitation sources close to the infrared. However, several technological obstacles restrict the wide clinical application of Raman spectroscopy as an optical biopsy technique [8].

In the actual technological stage, the available spectroscopy instrumentation (detectors, gratings, lasers, etc.) with best performance operates in the visible spectral region (400–750 nm). Near-IR (750–1,000 nm) operation is also possible with some loss of quantum efficiency in detection systems. However, exciting biological samples in the visible range, it is not always suitable for usual Raman measurements due to the strong auto-fluorescence that arises in biological samples. The strong fluorescence signal masks the majority of Raman bands and decreases the signal-to-noise (S/N) and/or signal-to-background (S/B) ratio of the measurements and is hence a major obstacle to the widespread use of Raman spectroscopy as an optical biopsy. To overcome these problems, one possibility is to use an infrared radiation source, specially a Nd:YAG laser at 1,064 nm. For reasonable spectra with this excitation to be obtained with the FT-Raman technique, a longer acquisition time is required (typically 1,000 times longer than required for dispersive techniques). Another option is to apply ultraviolet (UV) radiation, where usually it is possible to obtain Raman spectra with very good quality due to the elimination of fluorescence and a Raman intensity magnification. Nevertheless, the instrumentation for UV is limited, expensive and can cause much damage to tissues, especially for "in vivo" applications [14]. One interesting possibility is to use the 2,800–3,600 cm^{-1} spectral region, since this region is almost fluorescence-free and the bands in this region have a larger Raman scattering cross-section than those in either the mid-IR/Raman or terahertz regions. In this spectral region, one could employ a 785 nm-based optical fiber system, which has several advantages, such as lower acquisition times, good signal-to-noise (and S/B) ratio, no (or less) tissue damage, less fluorescence contributing as compared to the visible light, and one is able to manufacture a more compact and portable system.

Another restriction to the wide application of the Raman-based optical biopsy is due to the prealtered, premalignant, and/or inflammatory tissues role as a misclassification source [16–19]. In fact, the development of a malignancy is related to an inflammatory process that occurs simultaneously and generally surrounding the neoplastic process [16–19].

The Raman spectra of inflammatory infiltrates in the fingerprint region (500–1,800 cm^{-1}) can cause errors in the discrimination of the normal and pathological tissue spectra [17–19]. One alternative is to analyze the bands in the high frequency region (2,800–3,600 cm^{-1}) that are related to OH, CH, CH_2, CH_3, and NH stretch vibrations due to lipids, proteins, carbohydrates, among other species present. Nazemi and Brennan III [15] evaluated lipid concentrations in the human coronary artery by high-wavenumber Raman spectroscopy. The reported results showed accurate compositional information of the human coronary artery and raised the possibility of determining this information in vivo via small-profile cardiovascular catheters. This region also has an important advantage related to being almost fluorescence-free when using fiber optic probes, which have fluorescence problems in the mid-IR/Raman region. Hence, the Raman spectra in the high frequency region can be very suitably accessed via fiber optic probes in this region. Nijsen et al. [16] employed the high wave-number shift region in a Raman spectroscopic study on basal cell carcinoma. By incorporating a threshold for the probability of prediction in the classification model, tumor tissue could be distinguished with a 100% prediction accuracy and noninvolved tissue with a 99% prediction accuracy.

In the present study, we have investigated the vibrational spectra of normal (NM) and oral inflammatory fibrous hyperplasia (IFH) tissues of buccal mucosa in the high-wavenumber (2,800–3,600 cm^{-1}) region using a 1,064 nm laser source. The characterization of these two processes by Raman spectroscopy is expected to be useful to reduce the false negative rates and also to shed light on the nature of the inflammatory process. IFH is a good inflammatory tissue prototype. The epithelium of the oral mucosa shows changes induced by the inflammatory process located in the lamina propria [3]. The alterations of IFH are observed in the epithelium and connective tissue. Usually, the IFH in the oral mucosa is caused by some previous trauma, for example due to the habitual biting of the mucosa. Thus, an inflammatory response is started at this site, which leads to increased tissue damage caused by this inflammation that ceases when the inflammatory factor is also interrupted [17].

2 Materials and methods

This research was first approved by the local ethical research committee 067/2006/PH-CEP and then carried out according to the ethical principles established by the Brazilian Health Ministry. Patients were informed concerning the subject of the research and they gave their permission/consent for the collection and use of their tissue samples for research purposes.

2.1 Sample preparation

Samples of IFH and NM were obtained from biopsies performed at the Department of Bioscience and Oral Diagnosis—UNESP/BRAZIL. Biopsies from twelve patients, six samples of each kind of tissue, were analyzed. After the surgical procedure, the tissue samples were immediately snap frozen and stored in liquid nitrogen (77 K) in cryogenic vials prior to FT-Raman spectra recording.

2.2 FT-Raman spectroscopy

The Raman spectra were measured at five different sites of location totalizing 30 spectra for each kind of tissue. Soon after the procedure, all samples were fixed in 10% formaldehyde solution for further histopathological analysis. A Bruker RFS 100/S FT-Raman spectrometer was used with a 1,064 nm Nd:YAG laser operating as the excitation light source. The laser power at the sample was kept at 230 mW on a laser spot of 0.3 mm diameter furnishing a power density of 330 W/cm^2, which is typical in optical biopsy studies for tissues. For example, in another study published by us [18], a similar power density level was employed using an optical fiber system (110 mW over 200 μ laser spot). Due to the relevance of the power level irradiation on the tissue samples, the initial Raman-based optical biopsy studies raised and discussed this important issue. For example, in the work of Gniadecka et al. [19], the authors used a 1,064 nm excitation laser source over a 100 μ diameter area and observed no spectral differences for laser powers between 30 and 300 mW. In fact, powers as high as 700 mW have been employed in this kind of studies (see e.g., Carter et al. [21]). This was not the case when dealing with cells in micro-Raman experimental setups [11, 13]. Beyond the intrinsic fragility of the cell, the highly focused beam strongly increased the power density. For example, Dasgupta et al. [20] have shown that the long exposition times of erythrocytes to Raman tweezers light (1.6 MW/cm^2, almost 5,000 times greater than our case) is able to degrade the hemoglobin. I would add a bit about the power, power density, and the exposure time. Also in X-ray analysis, the question of degradation of the sample is mentioned and discussed. Hence, the need to control the accumulated exposure time, which can also lead to effects, which one does not see or shorter exposure times, or pulsed laser systems. The system/tissues have time to transfer some of the energy and momentum to the environment, water for example. Vibrational energy transfer to other modes and to the environment is good in this regard. The spectrometer resolution was set to 4 cm^{-1}. Each spectrum was recorded with 300 scans which took approximately 6 min. For FT-Raman data collection, all samples that were stored at liquid nitrogen temperatures (80 K) were brought to room temperature and kept moistened with a 0.9% physiological solution to preserve their structural characteristics, before being placed in a windowless aluminum holder for the Raman spectra collection. The spectra were collected in this way; for both the epithelial and connective tissues.

2.3 Histopathological analysis

The stain used was hematoxylin and eosin (H&E), and the following staining protocol was adhered to for all samples: first, the sections were deparaffinized with xylene for 10 min, followed by re-hydration with absolute alcohol for 5 min, and washed with 95% alcohol for 2 min and then 70% alcohol for 2 min. After a quick wash in distilled water, the slides were taken into solution with hematoxylin and then counterstained with eosin. The histopathology of the NM samples showed normal epithelium, lamina propria

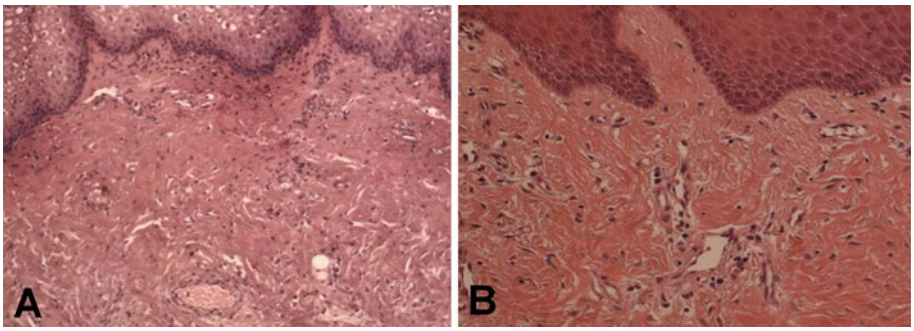

Fig. 1 Photomicrograph of histopathology sections showing: **a** normal mucosa (NM), **b** inflammatory fibrous hyperplasia (IFH) (H&E stain, 100×)

with an appearance of normality, and collagen fibers arranged in wavy bundles with typical cellular components (Fig. 1a). IFH tissues (Fig. 1b) showed epithelial changes, such as hydropic degeneration, exocytosis, spongiosis, acanthosis, and epithelial hyperplasia of cones [17]. The collagen fibers presented thick and irregular shapes. The diffuse inflammatory infiltrate is predominantly mononuclear and sometimes showed the presence of congestive blood vessels. Depending on the relative amount of inflammatory cells, the infiltrate can be classified as mild, moderate, or intense infiltrate.

2.4 Data analysis

All the spectra were baseline corrected (manual background fitting). For the statistical analysis, the spectra were also autoscaled (vector normalized). The spectral differences were analyzed using multivariate principal components analysis (PCA). PCA was performed over the range 2,800–3,600 cm^{-1} by computing the covariance matrix. The underlying data structure was summarized by clustering PC1, PC2, PC3, and PC4 scores at a 95% level of similarity using correlation distance measurements. The set of PCs that provided the best classification after visual inspection of scatter plots was fed into the binary logistic regression (BLR) algorithm [22] to determine the parameter equation that best differentiated the pathologic states. BLR provides a method for modeling a binary response variable, considering only two values of 0 or 1, and is based on the linear dependence between the logistic function of the probability of response 1 and the diagnosis variable. The BLR model equation is

$$\ln\left(\frac{p}{1-p}\right) = a + \sum_i b_i C_i,$$

where p is the probability of obtaining response 1, a and b_i are the model parameters, and C_i the diagnosis variables. All these steps were performed using the statistical software Minitab, version 14.20 Minitab Inc.®. State College, Pennsylvania, USA. The model's predictive ability was estimated by measuring the association between the response variable and predictive probabilities, p values, Somers' D parameter, Goodman–Kruskal gamma test (GKG), and Kendall rank correlation coefficient (KCC) [22].

3 Results and discussion

Figure 2 shows the Box Plot for the NM (Fig. 2a) and IFH (Fig. 2b) baseline corrected spectral data. The black lines correspond to the average spectrum, while the vertical gray ones are the regions between the first and third quartiles. The assignment of the main Raman bands is presented in Table 1, following Refs. [10] and [11]. It is important to note that the 2,800–3,050 cm^{-1} region (CH_2 and CH_3 stretch vibrations) showed greater intra-group variation for normal mucosa. However, the IFH spectra showed the highest intra-group variation in the 3,050–3,600 cm^{-1} region (CH, OH, and NH stretch vibrations). The relative area $A_{2800-3050}/A_{3050-3600}$ was ~3 for NM, while it decreases to ~1.5 for IFH. The two above cited regions present subtle spectral structures consisting of a set of superposed nearly overlapping bands that are related to CH_2 and CH_3 stretch and CH, OH, and NH stretch vibrations on different molecules,

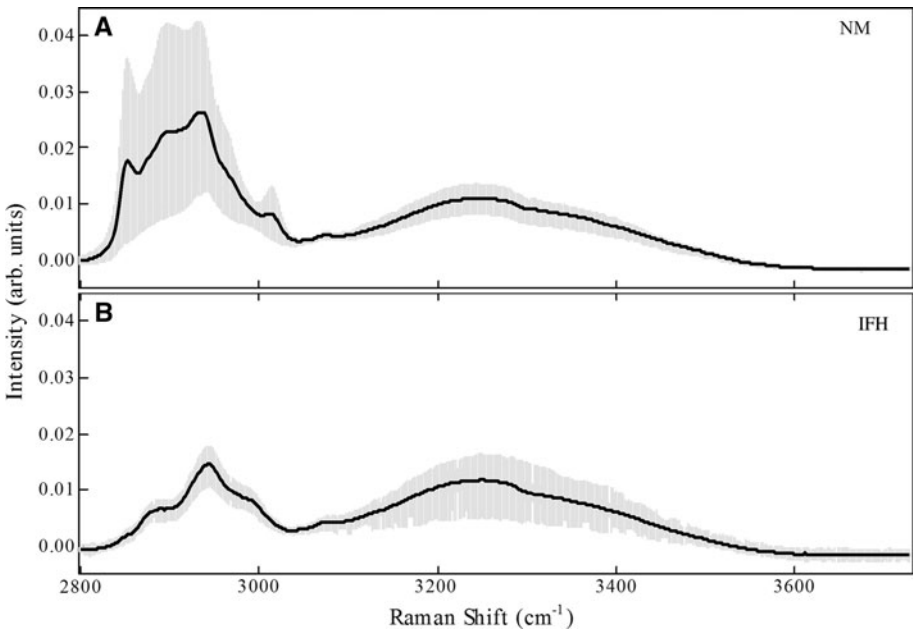

Fig. 2 **a** Box plot of IFH and **b** Box plot of NM, The *black line* is related to the average of the spectrum obtained and the region shaded in *gray* refers to the variation found in groups

Table 1 Assignment of the bands, which furnished variations in their loading plots based on Ref. [11].

Band (cm^{-1})	Assignment	% Concordant	Somers' D
1. 2,851	CH_2 symmetric stretch of lipids	91.7	0.83
2. 2,870	CH_3 assymetric stretch of lipids and proteins	87.2	0.75
3. 2,890	Overtone of the CH_2 deformation vibration near 1,450 cm^{-1}	94.2	0.88
4. 2,936	CH_2 antisymmetric stretch	98.1	0.96
5. 2,965	CH_3 antisymmetric stretch of lipids, fatty acids cholesterol and cholesterol ester	83.1	0.67
6. 2,988	CH_2 symmetric stretch	93.8	0.88
7. 3,011	Unsatured = CH stretch	93.3	0.87
8. 3,063	CH stretch of aromatic rings	75.2	0.51
9. 3,179	NH_2 symmetric stretch	63.4	0.28
10. 3,285	NH_2 antisymmetric stretch which overlaps with OH	75.8	0.52

The outcomes of the binary logistic regression tests are also displayed. The two best discriminating bands are shadowed

respectively. In order to evaluate whether some of these bands could contribute to the discrimination, we first evaluated which of these are the most relevant ones. The criterion was based on the analysis of the loading plots for the first 4 PCs. We consider as relevant those bands that presented positive or negative intensity variations compared with PC1. The most relevant bands identified are shown in Fig. 3 as vertical dash lines. Table 1 displays the 10 most relevant bands identified and the corresponding assignments.

We analyzed the contribution to the classification of each isolated band by considering its ability to correctly furnish the discrimination between NM and IFH in a binary logistic model. Table 1 summarizes the results considering as diagnostic variables PC1, PC2, PC3, and PC4. The p values less than 0.05 indicate that there is sufficient evidence that the corresponding variable contribution is not zero using an α-level of 0.05 [22]. The accuracy of the model could be evaluated by the percent of concordant pairs, while the Somers' D parameter provides an estimate of the rank of the correlation of the observed binary response variable and the predicted probabilities and is used as a quality indicator of the model fit. The best discrimination was observed using band 4 with 98.1% of concordant pairs.

Another way to determine the statistical relevance of these bands for discrimination is to analyze the relative variation in corresponding spectral areas. Applying the binary logistic regression model, we found that the best fit to this data set was

$$\ln\left(\frac{p}{1-p}\right) = 12(3) - 8(4)A_{2800-3050} - 4(1)A_{3050-3600}, \quad (1)$$

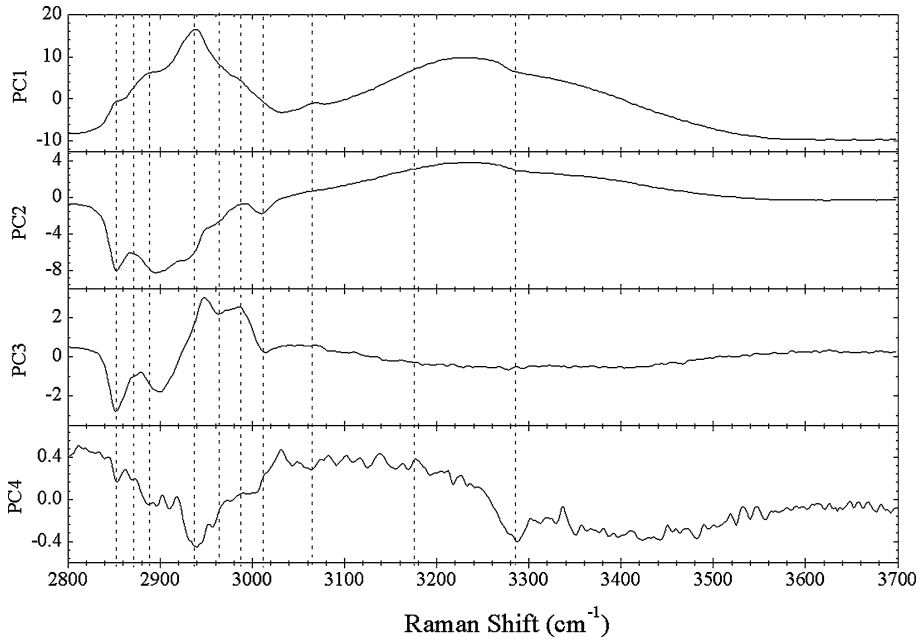

Fig. 3 Loading plot of PC1-PC4. Observe the *lines* indicating the 10 spectral regions contributed most to the diagnosis between the groups NM and IFH

where $A_{2800–3050}$ and $A_{3050–3600}$ are the integrated areas of the 2,800–3,050 cm^{-1} and 3,050–3,600 cm^{-1} spectral regions, respectively. The numbers in parenthesis indicates the standard error (SE) in the corresponding parameter. In this case, the percent concordant pairs were 95.3% and the Somers' D parameter was 0.91. The Se and Sp were 97 and 87%, respectively. The GKG test value was 0.91, which indicates a relatively good positive association. The KCC was 0.46, which indicates a fairly good but imprecise model. This outcome is a manifestation of the low-*N* of the present study. In a previous work [23] on the same tissues, we found a similar discriminative power using the 530–580 cm^{-1} part of the fingerprint region. It is important to note that the discrimination power was similar to that obtained using the specific bands given in Table 1. The best fit model to the $A_{2800–3050}$ single area furnished 81.0% of concordant pairs, Somers' D parameter 0.74, GKG equal to 0.74, and KCC of 0.38. For the $A_{3050–3600}$ area, it was found 71.7%, 0.44, 0.44, and 0.22 for the concordant pairs, Somers' D parameter, GKG, and KCC, respectively. Thus, it is clear that the best discrimination could be obtained when considering the two areas in the model represented by Eq. 1.

Figure 4 shows the scattering plot of the two areas compared with the logistic model. The straight line corresponds to the random case $p = 0.50$. It is very clear that the data separates into two classes. The hatching area indicates the upper and lower limits of the model considering the SE in the parameters. Thus, we can state that Se and Sp will fall in the intervals from $87\% < Se < 100\%$ to $73\% < Sp < 93\%$, respectively. By increasing the sample number, one would certainty narrows these intervals.

Figure 3 show that more subtle information could be extracted from these results. In fact, there are a set of overlapping bands originating from specific locations, which are not equally contributing to the diagnosis. One interesting example is the structured water around proteins.

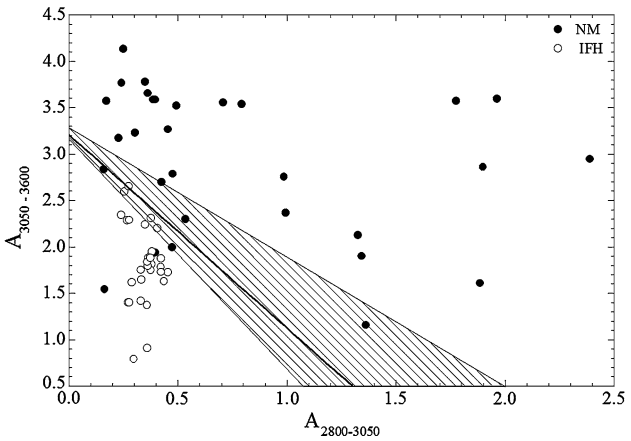

Fig. 4 Plot of $A_{2800–3050}$ and $A_{3,050–3,600}$ areas scaled to the logistic model. The *straight line* corresponds to the random case $p = 0.50$

Water and other biomolecules inside the cells behave differently (have different properties) from (than) the water outside [26, 27]. The viscosity of the intracellular water, for example, is higher and its ability to spread is smaller, as observed by nuclear magnetic resonance (NMR) [24]. As water molecules occupy specific sites and form localized clusters with structures that are determined by their hydrogen-bonding capabilities they are able to influence the vast majority of physical and chemical properties of the tissues, thus the water may be self-structured differently in normal and inflammatory tissues [24].

Cibulsky and Sadlej [25] observed that spectra in the OH stretch vibrational region are related to local structures and interactions of the hydrogen-bonds networks, so each cluster has a characteristic Raman spectrum. Since water is one important constituent, which has vibrational bands falling in this spectral window, we put forth the hypothesis that the origin of the discriminating power of the high-wavenumber region is related to the hydropic degeneration process in IFH tissues as seen by the changes in the OH water bands in this region for IFH versus NM tissues. In this way, Gniadecka et al. [26] concluded that an increased amount of the non-macromolecule bound, tetrahedral water was found in photoaged and malignant tumors of the skin. This phenomenon may be caused by alterations in protein structure and decreased protein–water interactions. They also observed spectral changes in proteins; collagen together with glycosaminoglycan is known to control resistance to transport of molecules and fluid. Degradation of collagen structure may result in higher resistance to fluid and macromolecules transport in tumor tissue. Despite the abundance of free tetrahedral water, the tumor tissue may show decreased fluid transfer. These results confirm the hypothesis that confined water was altered in pathologic tissues and due to some molecular arrangements of this water; it can influence the inflammatory response. The recent work of Bohr and Olsen [27] has shown that the collagen structure could be a variant of the well-known supercoiled triple helix. The proposed closed-packed triple helix structure presents a new feature, the existence of a central channel with negatively charged walls. This central channel offers the possibility of ion transport and may cast new light on various biological and physical phenomena [27]. As pointed out by the authors, the width of the channel in this calculated structure is ~ 2 Å, too small for water to be transported. Leikin et al. [28] have shown that hydration forces effects resulting from energetic cost of water rearrangement near collagen surface display spectral features in the Raman spectra of collagen. We argue that these hydration forces could increase the width of collagen channel enabling water confinement inside collagen channel. The lipids and proteins variation observed could be explained considering the acanthosis process in the IFH epithelium. In this process, the thickness of the epithelium

increases due to the growing of the spinous layer. The spinous layer has a more homogeneous composition than the connective tissue being almost lipid-free [17].

There are few studies in the literature studying protein–water interactions by high-wavenumber Raman spectroscopy from a fundamental point of view. In our opinion, the question is more complex and could be related to the water confinement regime itself. The structured water could, e.g., increase tunneling pathways between electrons among proteins boosting the transfer process (see Ref. [29]) Also, the proton mobility is strongly increased in confined water as compared to bulk (see Refs. [30, 31]). Both processes play an important role in the redox reactions producing reactive oxygen species and nitric oxide, which are at the top of the inflammatory cascade.

4 Conclusions

In conclusion, the results of the present work demonstrated the capacity of the high-wavenumber region (2,800–3,600 cm^{-1}) to provide relevant pieces of information concerning the discrimination between normal and inflammatory tissues by directly probing the 2,800–3,050 cm^{-1} (CH_2 and CH_3 stretch vibrations) and 3,050–3,600 cm^{-1} (CH, OH, and NH stretch vibrations) spectral frequencies and intensities. The discriminative power is similar to that obtained when probing specific bands (Table 1) and applying complex multivariate analysis (PCA procedure). The relevance of these findings relies on the fact that this almost fluorescence-free spectral region can furnish reliable discriminative information about inflammatory process in a very simple fashion which directly benefits the design of portable systems, e.g., 785 nm-based setups. In addition, the misleading or confusing discrimination that arises from prealtered, premalignant, and/or inflammatory tissues could be fixed/remedied. It is important to stress that more experimental effort needs to be devoted to increase the number of patients and also the pathology variety. Moreover, detailed studies need be performed to discriminate each contribution to the tissue spectra and molecular complex simulations of the Raman spectra will be most useful to either confirm or refute the many hypothesis and even speculation which appear in the literature.

Acknowledgments The authors would like to thank the Brazilian agencies FAPESP, CAPES and CNPq for the financial support and Airton Abrão Martin of LEVB-UNIVAP. We are also grateful to K. J. Jalkanen for the critical reading of the manuscript.

References

1. Boyle P, Levin B (2008) World cancer report 2008. International Agency for Research on Cancer, Lyon
2. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, Matchar DB (2000) Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities. Ann Int Med 132:810–819
3. Holmstrup P, Vedtofte P, Reibel J, Stoltze K (2007) Oral premalignant lesions: is a biopsy reliable? J Oral Path Med 36:262–266
4. Lieber CA, Kabeer MH (2010) Characterization of pediatric Wilms' tumor using Raman and fluorescence spectroscopies. J Pediat Surg 45:549–554
5. Mo J, Zheng W, Low JJH, Ng J, Ilancheran A, Huang Z (2009) High wavenumber Raman spectroscopy for in vivo detection of cervical dysplasia. Anal Chem 81:8908–8915
6. Backhaus J, Mueller R, Formanski N, Szlama N, Meerpohl HG, Eidt M, Bugert P (2010) Diagnosis of breast cancer with infrared spectroscopy from serum samples. Vib Spectrosc 52:173–177
7. Griebe M, Daffertshofer M, Stroick M, Syren M, Ahmad-Nejad P, Neumaier M, Backhaus J, Hennerici MG, Fatar M (2007) Infrared spectroscopy: a new diagnostic tool in Alzheimer disease. Neurosci Let 420:29–33
8. Austwick MR, Clark B, Mosse CA, Johnson K, Chicken DW, Somasundaram SK, Calabro KW, Zhu Y, Falzon M, Kocjan G (2010) Scanning elastic scattering spectroscopy detects metastatic breast cancer in sentinel lymph nodes. J Biomed Opt 15:047001
9. Penteado SCG, Fogazza BP, Carvalho CS, Arisawa EAL, Martins MA, Martin AA, Martinho HS (2008) Diagnosis of degenerative lesions of supraspinatus rotator cuff tendons by Fourier transform-Raman spectroscopy. J Biomed Opt 13:014018
10. Naumann D (2001) FT-infrared and FT-Raman spectroscopy in biomedical research. Appl Spectrosc Rev 36:239–298
11. Naumann D (2000) Infrared spectroscopy in microbiology. In: Meyers RA (ed) Encyclopedia of analytical chemistry. Wiley, Chichester
12. Martinho HS, Yassoyama MCBM, de Oliveira PA, Bitar RA, Santo AME, Arisawa EAL, Martin AA (2008) Role of cervicitis in the Raman-based optical diagnosis of cervical intraepithelial neoplasia. J Biomed Opt 13:054029
13. de Veld D, Bakker Schut T, Skurichina M, Witjes M, Van der Wal J, Roodenburg JL, Sterenborg HJ (2005) Autofluorescence and Raman microspectroscopy of tissue sections of oral lesions. Las Med Sci 19:203–209
14. Martins MAS, Ribeiro DG, Pereira dos Santos EA, Martin AA, Fontes A, Martinho HS (2010) Shifted-excitation Raman difference spectroscopy for in vitro and in vivo biological samples analysis. Biomed Opt Express 1:617–626
15. Nazemi JH, Brennan JF III (2009) Lipid concentrations in human coronary artery determined with high wavenumber Raman shifted light. J Biomed Opt 14:034009
16. Nijssen A, Maquelin K, Caspers P, Schut T, Neumann M, Puppels G (2007) Discriminating basal cell carcinoma from perilesional skin using high wave-number Raman spectroscopy. J Biomed Opt 12:034004
17. Firoozmand LM, Almeida JD, Cabral LAG (2005) Study of denture-induced fibrous hyperplasia cases diagnosed from 1979 to 2001. Quint Int 36:825–829
18. García-Flores AF, Raniero L, Canevari RA, Jalkanen KJ, Bitar RA, Martinho HS and Martin AA (2011) High-wavenumber FT-Raman spectroscopy for in vivo and ex vivo measurements of breast cancer. Theo Chem Acc. doi:10.1007/s00214-011-0925-9
19. Gniadecka M, Wulf HC, Mortensen NN, Nielsen OF, Christensen DH (1997) Diagnosis of basal cell carcinoma by Raman spectroscopy. J Raman Spectrosc 28:125–129
20. Dasgupta R, Ahlawat S, Verma RS, Uppal A, Gupta PK (2010) Hemoglobin degradation in human erythrocytes with long-duration near-infrared laser exposure in Raman optical tweezers. J Biomed Opt 15:055009

21. Carter EA, Williams AC, Barry BW, Edwards HG (2000) Probing diseased skin with FT-Raman spectroscopy. Proc SPIE 3257:72–77. doi:10.1117/12.306085
22. Bewick V, Cheek L, Ball J (2005) Statistics review 14: logistic regression. Crit Care 9:112–118
23. Carvalho LFCS, Bitar RA, Arisawa EAL, Brandãoo AAH, Honório KM, Cabral LAG, Brandão AAH, Honório KM, Martin AA, Martinho HS, Almeida JD (2010) Spectral region optimization for Raman-based optical biopsy of inflammatory lesions. Photomed Las Surg 28:111–117
24. Chaplin M (2006) Do we underestimate the importance of water in cell biology? Nat Rev Mol Cell Biol 7:861–866
25. Cybulski H, Sadlej J (2007) On the calculations of the vibrational Raman spectra of small water clusters. Chem Phys 342:163–172
26. Gniadecka M, Nielsen OF, Wulf HC (2003) Water content and structure in malignant and benign skin tumours. J Mol Struct 661:405–410
27. Bohr J, Olsen K (2010) The close-packed triple helix as a possible new structural motif for collagen. Theor Chem Acc. doi: 10.1007/s00214-010-0761-3
28. Leikin S, Parsegian VA, Yang WH, Walrafen GE (1997) Raman spectral evidence for hydration forces between collagen triple helices. Proc Natl Acad Sci USA 94:11312–11317
29. Lin J, Balabin IA, Beratan DN (2005) The nature of aqueous tunneling pathways between electron-transfer proteins. Science 310:1311–1313
30. Rasaiah JC, Garde S, Hummer G (2008) Water in nonpolar confinement: from nanotubes to proteins and beyond. Ann Rev Phys Chem 59:713–740
31. Jalkanen KJ, Degtyarenko IM, Nieminen RM, Cao X, Nafie LA, Zhu F, Barron LD (2008) Role of hydration in determining the structure and vibrational spectra of L-alanine and N-acetyl L-alanine N′-methylamide in aqueous solution: a combined theoretical and experimental approach. Theor Chem Acc 119:191–210