Theor Chem Acc (2011) 130:1239–1247
DOI 10.1007/s00214-011-1027-4

REGULAR ARTICLE

In and ex vivo breast disease study by Raman spectroscopy

L. Raniero · R. A. Canevari · L. N. Z. Ramalho · F. S. Ramalho · E. A. P. dos Santos · R. A. Bitar · K. J. Jalkanen · H. S. Martinho · A. A. Martin

Received: 23 March 2011 / Accepted: 17 August 2011 / Published online: 21 September 2011

Abstract In this work, Raman spectra in the 900–1,800 cm^{-1} wavenumber region of in vivo and ex vivo breast tissues of both healthy mice (normal) and mice with induced mammary gland tumors (abnormal) were measured. In the case of the in vivo tissues, the Raman spectra were collected for both transcutaneous (with skin) and skin-removed tissues. To identify the spectral differences between normal and cancer breast tissue, the paired *t*-test was carried out for each wavenumber using the whole spectral range from both groups. Quadratic discriminate analysis based on principal component analysis (PCA) was also used to determine and evaluate differences in the Raman spectra for the various samples as a basis for diagnostic purposes. The differences in the Raman spectra of the samples were due to biochemical changes at the molecular, cellular and tissue levels. The sensitivity and specificity of the classification scheme based on the differences in the Raman spectra obtained by PCA were evaluated using the receiver operating characteristic (ROC) curve. The in vivo transcutaneous normal and abnormal tissues were correctly classified based on their measured Raman spectra with a discriminant proportion of 73%, while the in vivo skin-removed normal and abnormal tissues were correctly classified again based on their measured Raman spectra with a discriminant proportion of 86%. This result reveals a strong influence due to the skin of the breast, which decreased the specificity by 11%. Finally, the results from ex vivo measurements gave the highest specificity and sensitivity: 96 and 97%, respectively, as well as a largest percentage for correct discrimination: 94%. Now that the important bands have been experimentally determined in this and other works, what remains is for first principles molecular-level simulations to determine whether the changes are simply due to conformational changes, due to aggregation, due to changes in the environment, or complex interactions of all of the above.

Keywords Breast cancer · Raman spectroscopy · Receiver operator characteristic curve · Principal component analysis · Quadratic discriminant analysis · Cancer detection · Cancer diagnostic · Optical biopsy

Abbreviations

PCA	Principal component analysis
DNA	Deoxyribonucleic acid
FWHH	Full width at half height
ROC	Receiver operating curve

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

L. Raniero · R. A. Canevari · E. A. P. dos Santos · K. J. Jalkanen · A. A. Martin (✉)
Laboratory of Biomedical Vibrational Spectroscopy (LEVB), Institute of Research and Development, IP&D, Universidade do Vale do Paraíba, UniVap, Avenida Shishima Hifumi, 2911, Urbanova, São José dos Campos, SP CEP 12244-000, Brazil
e-mail: amartin@univap.br

L. N. Z. Ramalho · F. S. Ramalho · R. A. Bitar
Departamento de Patologia e Medicina Legal, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Avenida Bandeirantes, 3900, Monte Alegre, Ribeirão Preto, SP CEP 14049-900, Brazil

K. J. Jalkanen
Department of Physics, Quantum Protein (QuP) Center, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

H. S. Martinho
Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Rua Santa Adélia, 166, Bangu, Santo André, SP CEP 09210-170, Brazil

BC	Breast cancer
CDC	Center for disease control
MRI	Magnetic resonance imaging
NTT	Normal tissue transcutaneous
CTT	Cancer tissue transcutaneous
NTWS	Normal tissue without skin
CTWS	Cancer tissue without skin
NTEV	Normal tissue ex vivo
CTEV	Cancer tissue ex vivo
DMBA	7,12-dimethylbenz(a)anthracene
FT	Fourier transform

1 Introduction

Breast cancer (BC) was one of four most common types of new cancer cases and deaths worldwide in 2007 [1] and in the USA, the second leading cause of deaths according to the Center for Disease Control (CDC) in Atlanta, Georgia [2]. In addition, the number of cases is expected to increase during the coming years due to an increase in life expectancy combined with other factors such as poor nutrition, physical inactivity, and an increase in exposure to radiation and chemicals in the environment among other risk factors. In Brazil, the estimated number of new cases of breast cancer expected in 2010 is 49,240 with an estimated risk of 51 cases per every 100,000 women [3].

Many BC studies that have been reported can be subdivided into prevention, diagnosis, and treatment. The breast self-examination is the first preventive care recommended for women over the age 20 to help find new lumps and other changes [4]. Nevertheless, this examination has limitations and screening mammograms are needed to provide a better diagnosis [5]. Indeed, mammography screening has shown positive results, reducing the mortality rate from breast cancer by 15–25% for women between the ages of 50 and 69 [6]. This technique has shown limitations for very dense breast tissue; a small tumor may not be visible until it becomes larger, increasing the probability of tumor metastasis. In this case, an additional ultrasound examination after a negative mammogram could be useful for cancer detection [7]. Another technique is magnetic resonance imaging (MRI), which has been used to detect additional foci of cancer that are occult on conventional imaging, but the use of MRI is controversial [8].

Although these methods detect the image of an abnormal breast tissue, the final diagnosis of BC must be confirmed through biopsy methods, such as fine-needle aspiration, surgical or core incision biopsies (with or without image guiding), radio-guided occult lesion localization, excision biopsies with wire localization, and mammotomy [9]. All techniques are done with direct access to the suspected breast lesion and require procedures that range from local anesthesia in a doctor's office to a hospital stay with the use of general anesthesia, both of which may be traumatic to the patients.

Among different diagnostic techniques, Raman spectroscopy has been shown to be a promising new tool for real-time diagnosis and results published to date have shown excellent potential to distinguish normal from neoplastic tissues with high specificity and sensitivity. For ex vivo breast malignancy analysis, the specificity and sensitivity values are around 96 and 94%, respectively [10–13].

Raman spectroscopy can also be used during surgery to help doctors differentiate lesions borders, giving an appropriate margin of security for the medical procedure, especially when the tumor limits are not clearly visible. For breast-conserving surgeries, such as partial mastectomy, 20–50% of the patients require a second surgical procedure due to positive margins, which are an indication of incomplete cancer resection [14, 15]. For such applications, a laser light can be guided to and the subsequent Raman signal collected from the same region (tissue) of interest through the same optical fiber and thereby provide information about the tissue being examined and give real-time diagnosis [16]. For non-invasive procedures, this technique can be applied to diagnose skin cancer.

Basically, the Raman spectral differences between normal and abnormal tissues are related to the biochemical changes in the cells. Some biochemical changes are, for example, related to the different quantity of lipids and proteins between normal and cancer cells. Normal cells will store energy in the form of lipids that contain many calories in a small space (energy-efficient molecules), and most of them are composed of some sort of fatty acid arrangement. Malignant cells, due to their uncontrolled growth process, synthesize large amounts of proteins, which are essential for the modulation and maintenance of cellular activities. Therefore, it is reasonable to assume that cancer cells should have a higher protein concentration than normal cells, whereas normal cells should have a higher lipid concentration than cancer cells. The increased concentration of proteins could be responsible for some of the observed structural changes, as proteins, for example, like insulin, are known to form aggregates and fibrils at certain critical concentrations [17]. A variety of experimental techniques were used to monitor such processes initially, including both fluorescence and electronic circular dichroism (ECD) spectroscopies. Subsequently, the simulation of the vibrational spectra of the damaged nucleic acid bases [18–20] and the effects of the explicit biological environment have been shown to be important in understanding the structure, function, and the vibrational spectra of biomolecules in complex environments. Hence the need to include

the explicit solvent molecules and denaturing influences in modeling studies of vibrational spectra [21–23].

However, Raman was shown to be not sensitive enough to observe and understand the nucleation events that initiate the aggregation effects and subsequent fibril formation [17]. More recently, the chiral analog of infrared absorption, vibrational circular dichroism (VCD), has been shown to be very sensitive to nucleation events and hence fibril formation. In addition, there is an increased sensitivity because the signal due to the fibril appears to be enhanced relative to the monomer signal. The VCD spectra, being a vibrational spectroscopy, have a wealth of additional information over that of the ECD spectra, which is the electronic absorption equivalent of VCD [24].

The detection of the biochemical changes during the disease process is the main key of using Raman for cancer diagnosis. Recently, there has been some work in trying to account for dietary and drug effects in the use of Raman spectroscopy as a diagnostic tool for rheumatoid arthritis [25], which can, in theory, be eliminated by asking the patient to refrain from taking their prescription drugs and not eating for 24 h before one draws blood. Methods like the ones we developed in that and this work are fundamental and hence must be developed to document, take into account, and hopefully eliminate any such induced errors. Certainly, this comparison could be helpful to determine changes in the specificity and sensitivity of the transcutaneous cancer diagnoses, for example, for the oral region.

Although there are many articles focused on in vivo or ex vivo Raman analysis of cancer tissue, to the best of our knowledge, there have been no studies focusing on the influence of the skin, as well as the comparison between in vivo and ex vivo Raman measurements other than the effect in the high-frequency region, which will also be published in this special issue [26].

Therefore, in this article, the normal (NT) and cancer breast tissues (CT) of mice were studied by Raman spectroscopy in vivo transcutaneous (NTT—normal tissue or CTT—cancer tissue), after skin removal (NTWS or CTWS), and ex vivo tissue biopsy (NTEV or CTEV) measurements. The evaluation of spectral data was done by *t*-test and a quadratic discriminate analysis based on principal component analysis (PCA). The sensitivity and specificity were evaluated using the receiver operating characteristic (ROC) curve.

2 Experimental details

These studies were performed in accordance with the Guidelines for the Care and Use of Experimental Animals; in addition, the local Ethics Committee approved all animal procedures used in this study (A022/CEP/2006). The group studied was comprised of 20 young virgin Sprague–Dawley female rats, with an average weight of 175 ± 10 g and age of 40 days. Food and water were allowed ad libitum. Mammary gland tumors were induced in 15 rats by a single-dose administration of 50 mg/kg of DMBA (7,12-dimethylbenz(a)anthracene) diluted in 1 mL of soy oil given intragastrically by gavages [27]. The control group was composed of five rats, for which the gavage procedures were simulated with soy oil. The mammary glands (six pairs) of each rat were checked by visual and physical examinations (palpation) three times a week. Seven weeks after the initial drug injections, palpable tumors with diameters between 0.8 and 1.0 cm were detected in at least one of the six breasts from all rats in the DMBA group, while no palpable tumors were found in any of the rats in the control group.

In order to compare the changes in vibrational modes as a function of the experimental procedure used, Raman spectra were collected in three different ways, as shown in Fig. 1. In Fig. 1(a) a Raman probe was used to acquire in vivo spectra from normal and abnormal mammary tissues though the skin (transcutaneous); (b) in the same region, skin was locally removed surgically and the spectra were collected in vivo using the same Raman probe (skin removed); and (c) after the animal was sacrificed, a piece of breast tissue was removed (biopsy) from the same region and analyzed directly with the FT-Raman spectrometer. The Raman spectra for the healthy rat skin were also obtained by using this last procedure. The number of rats, the number of mammary glands (six pairs) for each rat, and the number of Raman spectra collected for each category is given in Table 1.

The biopsy samples, after the surgical procedure, were identified, snap-frozen, and stored in liquid nitrogen (77 K) in cryogenic vials (Nalgene®) before the FT-Raman spectra were measured. For ex vivo FT-Raman data collection, the samples that had been stored at liquid nitrogen temperature were brought to room temperature and kept moist in 0.9% physiological solution to preserve their structural characteristics and placed in an aluminum sample holder for the Raman spectra collection.

A Bruker RFS 100/S FT-Raman spectrometer equipped with a Nd:YAG laser with a wavelength of 1,064 nm as the excitation light source was used. The spectrometer resolution was set to 4 cm^{-1}, and 600 scans (~8 min) were recorded and averaged to generate the Raman spectra (900–1,800 cm^{-1}). In vivo measurements were taken using a fiber-optic probe (RamProbe, Brucker®) that delivers a power of 120 mW to the sample, and the Raman scattering was collected from approximately 1 mm^3 of tissue volume probing a large amount of cells (~10^6) that was composed of at least 80% neoplastic cells in the abnormal tissue. An

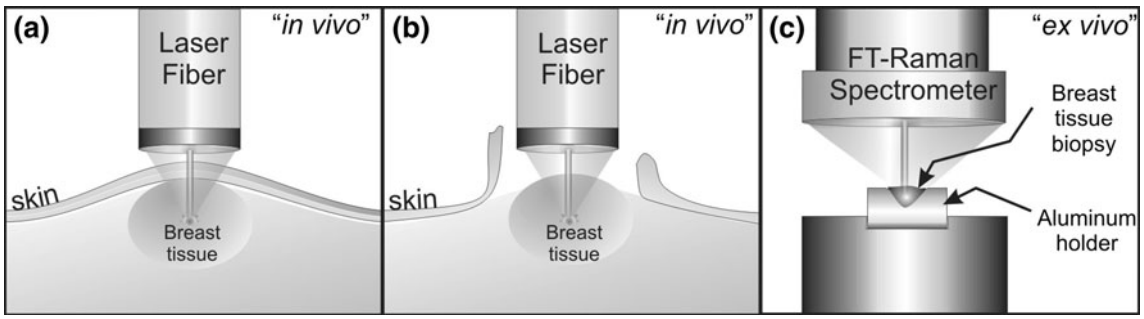

Fig. 1 Measurement details: **a** in vivo spectra with a Raman probe (transcutaneous); **b** in vivo spectra acquired without skin with Raman probe; **c** ex vivo spectra from biopsy tissue

Table 1 Number of Raman spectra collected from each rat for each category

Group	Number of rats	Number of mammary glands	Number of spectra
In vivo transcutaneous			
NTT	5	12	15
CTT	15	12	27
Normal skin	5	12	15
In vivo skin removed			
NTWS	5	17	69
CTWS	15	12	58
Ex vivo tissue biopsied			
NTEV	5	4	18
CTEV	15	8	30

aluminum sample holder was used for ex vivo measurements using the FT-Raman spectrometer. Measurements were taken in the frequency range from 900 to 1,800 cm^{-1}. The laser spot size was 200 μm in diameter and the power was kept below 110 mW at the sample to better preserve the integrity of the samples and prevent photodecomposition caused by the laser beam irradiation.

A baseline for all spectra was generated with a Matlab routine using a fifth-order polynomial to subtract the background due to tissue fluorescence [28]. Subsequently, all spectra were normalized by the peak at ~1,445 cm^{-1}. The Raman shifts found for the ~1,445 cm^{-1} band between normal and cancer tissue were determined by a Gaussian fitting of this mode from the mean spectra. Minitab® software (Release 14.20) was used to carried out paired *t* test for each wavenumber using the whole spectral range from both groups, which calculates the probability that two samples are from the same population, assuming the populations have the same mean. In order to identify the spectral differences between normal and cancer breast tissues, the average spectra for both groups were calculated with the standard deviation and plotted in Figs. 2, 3, 4, 5. The *p* values of *t* test were also plotted together to highlight the important regions in the spectra.

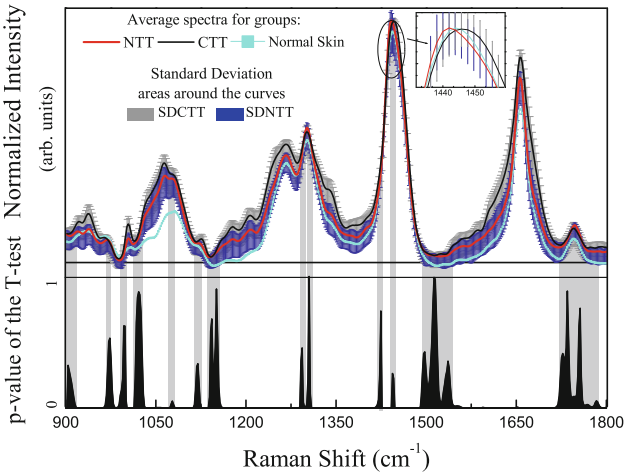

Fig. 2 Averages curves for normal and abnormal tissues for transcutaneous measurement group

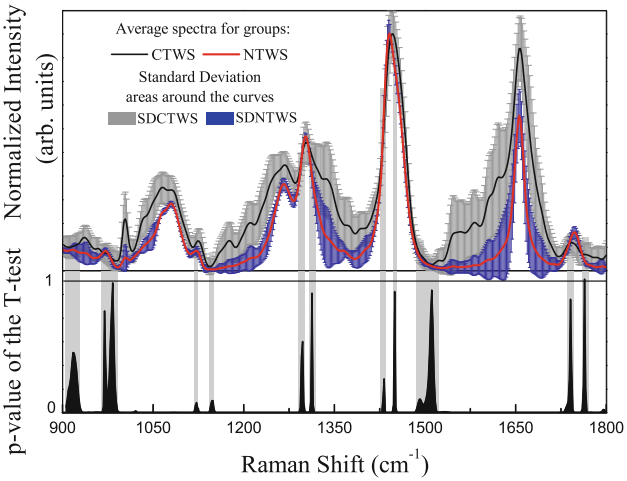

Fig. 3 Averages curves for normal and abnormal tissues for without skin measurement group

For the quadratic discriminant analyses (QDA), the background corrected spectra were vector-normalized (mean subtraction and divided by the standard deviation), mean-centered, and statistically analyzed by using the

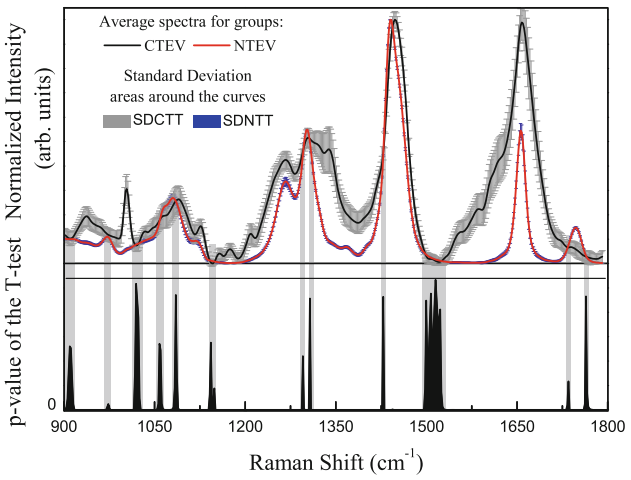

Fig. 4 Averages curves for normal and abnormal tissues for ex vivo biopsy measurements group

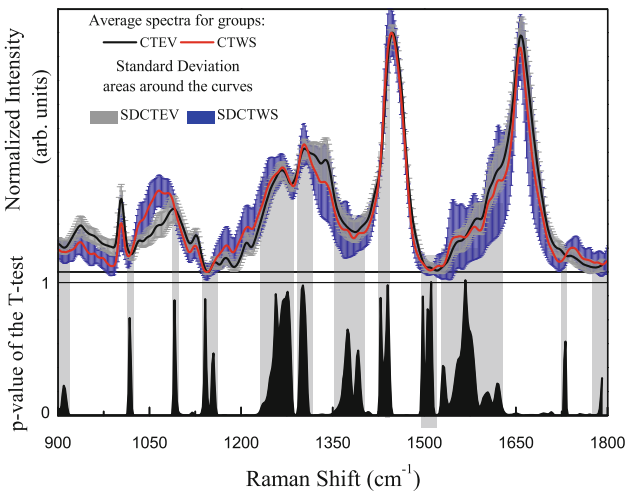

Fig. 5 Averages curves for abnormal tissues for in vivo (after skin removal—CTWS) and ex vivo biopsy (CTEV) measurements group

multivariate statistical tool box of the Minitab® software (Release 14.20). To reduce the dimensionality of the data set, the principal component analysis (PCA) was used in the Raman spectral range from 900 to 1,800 cm^{-1}. The information from this analysis was obtained through four principal components (PC1, PC2, PC3, and PC4). The QDA were done with leave-one-out cross-validation using PC1-PC4, according to pathological classification reviewed by two pathologists (Fig. 6).

The relative operating characteristic curves (ROC) were calculated using the values of squared distance of observation to the group center given by QDA. The ROC curve is the plot of the sensitivity versus 1−specificity, which gives the relationship between the rate of false negatives and the rate of false positives [29, 30].

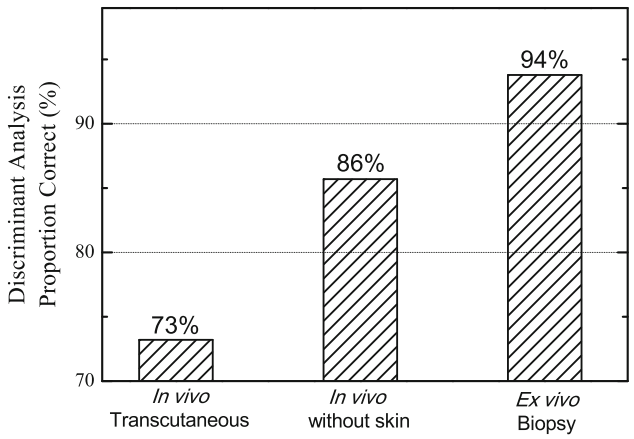

Fig. 6 Discriminant analysis with cross-validation for normal and abnormal tissues for in vivo transcutaneous, in vivo without skin, and ex vivo biopsy groups

3 Results and discussion

In order to analyze the differences between NTT and CTT spectra, a statistics-based *t*-test was carried out for each wavenumber using the whole spectral range from both groups. Figure 2 shows the mean spectra (with standard deviation) for normal (NTT) and cancer (CTT) tissues for in vivo transcutaneous measurements, as well as the Raman spectra for the healthy rat skin, obtained as shown in Fig. 1a. The dark filled area in Fig. 2 represents the higher values of t-test, where the regions are quite similar statistically. Table 2 reports the major assignment for cancer and normal tissues to guide the biochemical interpretation of Raman spectra [31–39].

Despite the great histological difference that exists between normal and neoplastic tissues, the biochemical differences between them are less evident through the skin. In addition, due to the fact that the normal tissue and breast lesions are located at a certain depth under the skin, the spectral profiles of these tissues also carry vibrational information from the skin components, masking the peculiar differences between them. The region around the peak at 1,748 cm^{-1} observed in connective tissue (collagen III) and lipids, for example, has shown no significant difference between CTT and NTT, which also indicates an influence of the skin. Indeed, the biochemical components of skin are very similar to NTT as shown in this figure, which reinforces the hypothesis dissembling the CTT spectra. However, a peak shift was observed at 1,440 cm^{-1} in the normal tissue (NTT) and at 1,445 cm^{-1} in the cancer tissue (CTT), suggesting that CH_2 deformation is from different functional groups, where peak at 1,440 cm^{-1} suggests the presence of lipids and 1,445 cm^{-1} indicates the dominance of proteins [31, 37, 40, 41].

Table 2 Peak assignments of the Raman spectra of normal tissue and breast carcinoma

Peak position (cm^{-1})	Major assignment
1,734–1,754	Lipids
1,654–1,656	Amide I (C=O stretching mode of proteins, a-helix conformation) and lipid (C=C stretch)
1,510–1,572	Nucleic acid mode (C=C and C=N stretch)
1,445	Protein (CH_2 deformation)
1,440	Lipid (CH_2 deformation)
1,299–1,303	Lipid, fatty acids, collagen, and amide III (protein)
1,265–1,267	Amide III (collagen and protein) and lipid (C–H)
1,207	Hydroxyproline, tyrosine
1,175–1,176	Cytosine and guanine
1,083	C–N stretching mode of proteins (and lipid mode to lesser degree)
1,078	C–C or C–O stretch (lipid), C–C or PO_2 stretch (nucleic acids)
1,064	Skeletal C–C stretch lipids
1,031	C–H in-plane bending mode of phenylalanine
1,001	Symmetric ring breathing mode of phenylalanine
935	C–C stretching mode of proline, valine, and protein backbone (a-helix conformation)/glycogen

Figure 3 shows the mean Raman spectra of normal (NTWS) and malignant breast tissue (CTWS) without skin influence, obtained as shown in Fig. 1b. Compared to the transcutaneous Raman data (Fig. 2), the peaks in the malignant tissue became broader (higher width at half height) and the peak around 1,748 cm^{-1} shows decreased intensity for CTWS group. The Raman modes at 1,301 and 1,748 cm^{-1} in the normal tissue spectra correspond to strong lipid bands, and the modes at 935 and 1,083 cm^{-1} correspond to weak protein bands. In the case of the malignant breast spectra, lipid bands are slightly weaker (1,301 and 1,748 cm^{-1}) compared with protein bands (935 and 1,083 cm^{-1}) as well as nucleic acid bands (1,078 cm^{-1}, 1,175–1,176 cm^{-1}, and 1,510–1,572 cm^{-1}). These features suggest the dominance of collagen, non-collagenous proteins, and nucleic acid in pathological tissues compared with normal tissues (Table 1). This result correlates well with that of Manoharan et al. [42] who using NIR Raman spectroscopic reported the dominance of lipid features in the normal breast tissue spectra and protein signatures in breast lesions. The ratios of peak areas of amide I peak to those of the δCH_2 modes and the peak position of the δCH_2 bands were their discriminating parameters. They achieved nearly 100% accuracy in differentiating normal tissue from pathological tissue. These findings were reinforced by Shafer–Peltier et al. [37] who developed a Raman microspectroscopic model of the human breast to extract information about the morphological and chemical components present in normal and abnormal tissue. The results from the model fitting have shown lipid and protein predominance for normal and malignant tissues, respectively. Another Raman microspectroscopic study from Yu et al. [43] showed a relative increase in nucleic acids and a decrease in lipid contents in malignant conditions. Yan et al. [44] also reported changes in the phosphate backbone modes, which was associated with damaged DNA and an uncontrolled fissiparity of cancer cell [44, 45].

The Raman modes at 935 cm^{-1}, 1,001–1,031 cm^{-1}, and 1,207 cm^{-1} correspond to the bands of the proline, valine, phenylalanine, hydroxyproline, and tyrosine amino acids, respectively, which characterize the primary structure of proteins. The breast pathological tissues are mainly composed of collagen [46]. Proline, valine, and phenylalanine are the main amino acids in collagen [47]. The intensity clearly increases in these bands in the tumoral tissues compared with the intensity of normal tissues. The origin of this intensity variation probably relies on the different collagen amounts present in normal and pathological tissues [46]. Moreover, the relative abundance of collagen increases in the carcinogenic process of skin [48], lung [49], breast [37], and epithelial cancers in general [34]. In this special issue of TCA, a new form of collagen has been reported theoretically by Bohr and Olsen that may be partially responsible for the observed changes in the collagen bonds due to either changes in concentration, aggregation state, or conformational changes induced by one of these two due to the changing environment in the diseased state and its progression [50].

In fact, phenylalanine is present in various neoplasm processes that are characterized by uncontrolled cell growth. For breast cancers in particular, due to the desmoplastic reaction, also called reactive fibrosis, deposition of abundant collagen occurs as a stromal response to an invasive carcinoma. Structures relatively remote from the cancer itself may be involved, such as Cooper's ligaments and duct structures between the tumor and the nipple [32, 46]. Thus, the Raman spectra of breast carcinoma are expected to show more intense proline, valine, and phenylalanine collagen bands than other tissues. Thus, the observed spectral features can provide vital clues in understanding the differences in biochemical composition of tissues, which is the basis of optical spectroscopic diagnosis.

The ex vivo measurements showed the highest amplitude difference between normal (NTEV) and abnormal (CTEV) tissues (Fig. 4). This finding has also been reported previously by Thakur et al. [51] who demonstrated that Raman spectroscopy measurements of mammary gland tissues from mice injected with 4T1 tumor cells can distinguish mammary tumors from other physiological states

of the mammary glands. The modes at 1,064, 1,078, 1,300, 1,440, and 1,748 cm^{-1} indicate higher lipid contribution in the NTEV than in the CTEV. This finding corroborates with our brief description that normal cells have a higher concentration of lipids than cancer cells. The fact that this effect could not be observed in the other two experimental procedures could be explained by the possible contribution of normal cells in the CTT and CTWS spectra masking this effect. Those contributions from the normal cells come from the characteristic of the in vivo technique itself. The Raman signal could either be collected close to the lesion border or by probing a normal portion of the tissue due to the penetration depth of the laser. For the ex vivo experiment, the biopsy was removed where most of cells originated from the uncontrolled growth of cancer cells.

Figure 5 shows the differences for abnormal tissues between spectra from the in vivo measurement group (after skin removal—CTWS) and the ex vivo biopsy (CTEV) measurement group. Because it is very often the presence of normal cells in vivo tissue that makes it difficult to obtain an accurate pathological grading of abnormal tissue, this comparison is relevant to develop diagnostic algorithms using ex vivo biopsy material and apply them in vivo tissue. The Raman modes at 1,064 and 1,740 cm^{-1} in the in vivo abnormal tissue spectra clearly indicate the dominance of lipid features in relation to the ex vivo group. In addition, the 1,078, 1,175, and 1,548 cm^{-1} peaks in the in vivo spectra are strong nucleic acid bands. The stronger (more intense) Raman modes at $\sim$1,207 cm^{-1} due to hydroxyproline and tyrosine in the in vivo group spectra are indicative of larger amounts of these amino acids compared with the ex vivo group. In the case of the ex vivo spectra, lipid bands are slightly weaker compared with the proline, valine, and other protein residue bands at 935 cm^{-1}, phenylalanine at 1,004 cm^{-1}, tryptophan at 1,339 cm^{-1}, and the amide I protein bands at 1,657 cm^{-1}. These features suggest that the in vivo group abnormal tissue spectrum shows a greater contribution from normal tissues compared with the ex vivo *group abnormal tissue* spectrum. Therefore, the discrepancies between in vivo and ex vivo methodology suggest a mixed spectrum for the former case, which has biochemical components of normal and cancer cells, while the latter case has a more pure spectrum of cancer cells.

3.1 Statistical analysis

The principal components analysis was carried out in the range of 900 to 1,800 cm^{-1} by a covariance matrix. The quadratic discriminant analyses (QDA) were calculated using PC1–PC4, according to pathological classification. The main difference between QDA and linear discriminant analysis (LDA) is that there is no assumption that the groups have equal covariance matrices. QDA were done with leave-one-out cross-validation, which provides a more accurate estimation of the correct classification rate. This method leaves out one observation of the data at a time that is used to construct the classification rule. After this classification rule is formed, the data that were left out were classified by the rule that is compared to the original classification to see whether the rule determined without it can properly predict it. Here, it is important that the characteristic information content in the data left out is available in the other data for it to be correctly predicted. Hence, there must be a certain amount of redundancy in the data. Figure 6 shows the discriminant analysis for normal and abnormal tissues for in vivo transcutaneous, in vivo without skin, and ex vivo biopsy groups.

Among groups analyzed, the differences between in vivo transcutaneous and in vivo/ex vivo were 13 and 21%, respectively. The different percentage of discriminant could be better explained when the specificity and sensitivity were compared as shown in Fig. 7, which was done for the results from cross-validation analysis. In this graphic, the arrows represent the value of sensitivity and specificity calculated from QDA classification summary: numbers of true negatives, true positives, false negatives, and false positives.

Probably due to the influence of the skin and mixing spectra characteristic (normal and abnormal), the specificity and sensitivity for the in vivo transcutaneous diagnosis had the lowest values, 84 and 77%, respectively. When the influence of the skin is removed from the Raman data (in vivo without skin measurements), the percentage of correctly identifying healthy tissues increased by 10%, whereas the sensitivity did not improve. This latter result is probably due to the contribution of normal cells in the CTWS spectra. For the ex vivo biopsy Raman data, the

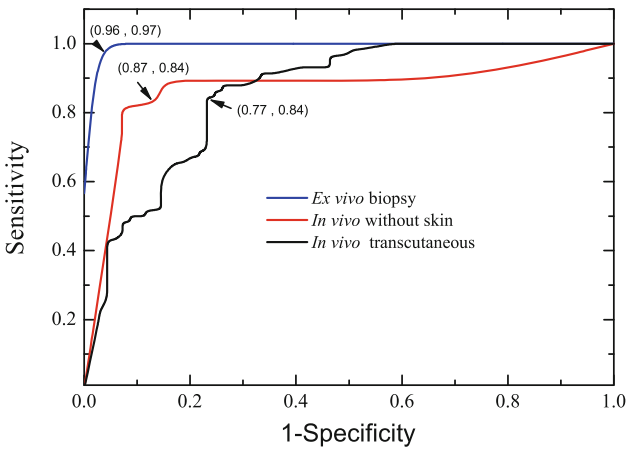

Fig. 7 ROC curve for specificity and sensitivity for in vivo transcutaneous, in vivo without skin, and ex vivo biopsy groups

sensitivity and specificity are approximately 13 and 19% higher than in the transcutaneous group, respectively.

The data analyses show many potential for in vivo transcutaneous application. Raman spectroscopy could certainly be applied for breast cancer using a minimally invasive procedure by passing the optical fiber through as one uses a fine hollow needle. However, for skin, colon, and oral lesion, this technique for cancer diagnoses could be more useful [52, 53]. For in vivo without the influence of skin, the optical fiber could be used to collect signals from endoscopies during laparoscopic surgery, or combined with other minimally invasive techniques. The sensitivity and specificity of diagnoses are mostly influenced by the region of the analyzed lesion, for example, if the fiber sticks close to a lesion border, the parameter could easily decrease an average of 10%.

4 Conclusions

The Raman spectra of transcutaneous normal and abnormal tissues in the 900 to 1,800 cm^{-1} region were used to correctly classify these tissues with a discriminatory proportion of 73%. This result reveals a strong skin influence, decreasing the analysis specificity by about 10%. For spectra acquired in situ without skin, the peaks are broad (higher width at half height) and the major contributions are from collagen, lipid, protein, and DNA bands, and the relatively strong amide I protein band, which suggests that malignant tissues could have higher protein concentration than normal tissues and normal tissues could have a higher lipid concentration than malignant tissues. The results from ex vivo measurements showed the highest values of specificity and sensitivity as well as proportion of correct discriminant analysis. What remains is for first principles ab initio and semi-empirical molecular-level simulations now to further validate, confirm, and better understand these molecular-level changes that have been shown in this work to be the molecular biomarkers of breast cancer. Whether they are only a sign of the disease or whether they are a contributing factor additionally needs to be elucidated. Clearly, the combination of experimental and theoretical Raman (and infrared, and perhaps now VCD) spectroscopy and the corresponding imaging is a way to not only monitor disease, but also its initiation events and progression.

Acknowledgments The authors would like to thank FAPESP and CNPq for financial support for the projects 01/14384-8 and 301362/2006-8, and 302761/2009-8, respectively, and to FAPESP for project 16782-2/2009 which allowed Prof. Jalkanen to visit LEVB at UniVaP for the period from June 2010 to May 2011 from the Quantum Protein (QuP) Center at the Technical University in Denmark (DTU), Kgs. Lyngby, Denmark.

References

1. Global Cancer Facts & Figures 2007 (2007) American Cancer Society http://ww2.cancer.org/downloads/STT/Global_Cancer_Facts_and_Figures_2007_rev.pdf. Accessed 15 March 2011
2. Jemal A, Siegel R, Xu J, Ward E (2010) Cancer statistics, 2010. CA Cancer J Clin 60:277–300
3. Brazilian National Cancer Institute, Incidence of Cancer in Brazil (2008) Estimativa | 2010 Incidência de Câncer no Brasil [estimate/2010—incidence of cancer in Brazil] (2009) Instituto Nacional de Câncer (INCA) updated with 2010 statistics from the following online documents now available http://www.inca.gov.br/estimativa/2010/estimativa20091201.pdf. Accessed 15 March 2011
4. Breast cancer (2011) American Cancer Society. http://www.cancer.org/acs/groups/cid/documents/webcontent/003090-pdf.pdf. Accessed 15 March 2011
5. Wright T, McGechan A (2003) Breast cancer: new technologies for risk assessment and diagnosis. Mol Diagnosis 7:49–55
6. World Health Statistics 2008 (2008) Breast cancer: mortality and screening. WHO Press, World Health Organization, Switzerland. http://www.who.int/whosis/whostat/EN_WHS08_Full.pdf. Accessed 15 March 2011, pp 21–23
7. Nothacker M, Duda V, Hahn M, Warm M, Degenhardt F, Madjar H, Weinbrenner S, Ute-Susann A (2009) Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review. BMC Cancer 9:335
8. Houssami N, Hayes DF (2009) Review of preoperative magnetic resonance imaging (MRI) in breast cancer should MRI be performed on all women with newly diagnosed, early stage breast cancer? CA Cancer J Clin 59:290–302
9. Bach-Gansmo T, Tobin D (2009) In Hayat MA (ed) Methods of cancer diagnosis, therapy, and prognosis—breast carcinoma, vol 1. Springer Science, Heidelberg, Germany, pp 100–175
10. Pawlukojć A, Leciejewicz J, Ramirez-Cuesta AJ, Nowicka-Scheibe J (2005) L-Cysteine: neutron spectroscopy, Raman, IR and ab initio study. Spectrochim Acta A Mol Biomol Spectrosc 61(11–12):2474–2481
11. Naumann D (2001) FT-infrared and FT-raman spectroscopy in biomedical research. Appl Spectrosc Rev 36:239–298
12. Shi Y, Wang L (2005) Collective vibrational spectra of α- and γ-glycine studied by terahertz and Raman spectroscopy. J Phys D 38:3741–3745
13. Hall JA, Knaus JV (2003) An atlas of breast disease. The Parthenon Publishing Group, London
14. Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Lyons J, Hicks D, Fitzmaurice M, Dasari RR, Crowe JP, Feld MS (2006) In vivo margin assessment during partial mastectomy breast surgery using Raman spectroscopy. Cancer Res 66:3317–3322
15. Bigio IJ, Bown SG, Briggs G, Kelley C, Lakhani S, Pickard D, Ripley PM, Rose IG, Saunders C (2000) Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results. J Biomed Opt 5:221–228
16. Motz JT, Gandhi SJ, Scepanovic OR, Haka AS, Kramer JR, Dasari RR, Feld MS (2005) Real-time Raman system for in vivo disease diagnosis. J Biomed Opt 10:031113
17. Nielsen L, Frokjaer S, Brange J, Uversky VN, Fink AL (2001) Probing the mechanism of insulin fibril formation with insulin mutants. Biochemistry 40:8397–8409
18. Jalkanen KJ, Jürgensen VW, Claussen A, Rahim A, Jensen GM, Wade RC, Nardi F, Jung C, Degtyarenko IM, Nieminen RM, Herrmann F, Knapp-Mohammady M, Niehaus TA, Frimand K, Suhai S (2006) Use of vibrational spectroscopy to study protein and DNA structure, hydration, and binding of biomolecules: a

combined theoretical and experimental approach. Int J Quantum Chem 106:1060–1098
19. Van Mourik T, Danilov VI, Dailidonis VV, Kurita N, Wakabayashi H, Tsukamoto T (2010) A DFT study of uracil and 5-bromouracil in nanodroplets. Theor Chem Acc 125:233–244
20. González-Ramírez I, Roca-Sanjuán D, Climent T, Serrano-Pérez JJ, Merchán M, Serrano-Andrés L (2011) On the photoproduction of DNA/RNA cyclobutane pyrimidine dimers. Theor Chem Acc 128:705–711
21. Deplazes E, van Bronswijk W, Zhu F, Barron LD, Ma S, Nafie LA, Jalkanen KJ (2008) A combined theoretical and experimental study of the structure and vibrational absorption, vibrational circular dichroism, Raman and Raman optical activity spectra of the L-histidine zwitterions. Theor Chem Acc 119:155–176
22. Jalkanen KJ, Degtyarenko IM, Nieminen RM, Nafie LA, Cao X, 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 study. Theor Chem Acc 119:191–210
23. Rai AK, Fei W, Lu Z, Lin Z (2009) Effects of microsolvation and aqueous solvation on the tautomers of histidine: a computational study on energy, structure and IR spectrum. Theor Chem Acc 124:37–47
24. Measey TJ, Schweitzer-Stenner R (2010) Vibrational circular dichroism as a probe of fibrillogenesis: the origin of the anomalous intensity enhancement of amyloid-like fibrils. J Am Chem Soc 133:1066–1076
25. Carvalho CS, Martin AA, Santo AME, Andrade LEC, Pinheiro MM, Cardoso MAG and Raniero L (2011) A rheumatoid arthritis study using Raman spectroscopy. Theor Chem Acc. doi:10.1007/s00214-011-0905-0
26. García-Flores AF, Raniero L, Canevari RA, Jalkanen KJ, Bitar RA, Martinho HS and Martin AA (2011) High wavenumber Raman spectroscopy for in and ex vivo measurements of breast cancer. Theor Chem Acc. doi:10.1007/s00214-011-0925-9
27. Barros ACSD, Muranaka ENK, Mori LJ, Pelizon CHT, Iriya K, Giocondo G, Pinotti JA (2004) Induction of experimental mammary carcinogenesis in rats with 7,12-dimethylbenz(a)anthracene. Revista do Hospital das Clínicas 59:257–261
28. Lieber CA, Mahadevan-Jansen A (2003) Automated method for subtraction of fluorescence from biological Raman spectra. Appl Spectrosc 57:1363–1367
29. Lugosi L, Molnár I (2000) Evaluation of medical diagnostic tests: application of Bayes theorem, ROC-curve and Kappa-test. Orv Hetil 141:1725–1728
30. Albeck MJ, Børgesen SE (1990) ROC-curve analysis. A statistical method for the evaluation of diagnostic tests. Ugeskr Laeger 152:1650–1653
31. Venkatakrishna K, Kurien J, Pai KM, Valiathan M, Kumar NN, Krishna CM, Ullas G and. Kartha VB (2001) Optical pathology of oral tissue: a Raman spectroscopy diagnostic method. Curr Sci 80:665–669. http://www.ias.ac.in/currsci/mar102001/665.pdf
32. Bitar RA, Martinho HS, Tierra-Criollo CJ, Ramalho LNZ, Netto MM, Martin AA (2006) Biochemical analysis of human breast tissues using Fourier-transform Raman spectroscopy. J Biomed Opt 11:054001
33. Penteado SC, Fogazza BP, Carvalho CS, Arisawa EÂ, Martins MA, Martin AA, Martinho HS (2008) J Biomed Opt 13:014018
34. Stone N, Kendall C, Smith J, Crow P, Barr H (2004) Raman spectroscopy for identification of epithelial cancers. Faraday Discuss 126:141–157
35. Movasaghi Z, Rehman S, Rehman IU (2007) Raman spectroscopy of biological tissues. Appl Spectrosc Rev 42:493–541
36. Mahadevan-Jansen A, Mitchell MF, Ramanujam N, Malpica A, Thomsen S, Utzinger U, Richards-Kortum R (1998) Development of a fiber optic probe to measure NIR Raman spectra of cervical tissue in vivo. Photochem Photobiol 68:427–431
37. Shafer-Peltier KE, Haka AS, Fitzmaurice M, Crowe J, Myles J, Dasari RR, Feld MS (2002) Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo. J Raman Spectrosc 33:552–563
38. Moreno M, Raniero L, Arisawa EÂL, Santo AME, dos Santos EAP, Bitar RA, Martin AA (2010) Theor Chem Acc 125:329–333
39. Marzullo ACM, Neto OP, Bitar RA, Martinho SH, Martin AA (2007) FT-Raman spectra of the border of infiltrating ductal carcinoma lesions. Photomed Laser Surg 25:455–460
40. Socrates G (2004) Infrared and Raman characteristic group frequencies: tables and charts, 3rd edn. Wiley, Chichester, p 336
41. Rehman S, Movasaghi Z, Tucker AT, Joel SP, Darr JA, Ruban AV, Rehman IU (2007) Raman spectroscopic analysis of breast cancer tissues: identifying differences between normal, invasive ductal carcinoma and ductal carcinoma in situ of the breast tissue. J Raman Spectrosc 38:1345–1351
42. Manoharan R, Wang Y, Feld MS (1996) Histochemical analysis of biological tissues using Raman spectroscopy. Spectrochimica Acta A 52:215–249
43. Yu G, Xu XX, Niu Y, Wang B, Song ZF, Zhang CP (2004) Studies on human breast cancer tissues with Raman microspectroscopy. Spectrosc Spectr Anal 24:1359–1362
44. Yan XL, Dong RX, Wang QG, Chen SF, Zhang ZW, Zhang XJ, Zhang L (2005) Raman spectra of cell from breast cancer patients. Spectrosc Spectr Anal 25:58–61
45. Yu C, Gestl E, Eckert K, Allara D, Irudayaraj J (2006) Characterization of human breast epithelial cells by confocal Raman microspectroscopy. Cancer Detect Prev 30:515–522
46. Hellman S, Harris JR, Canellos GP, Fisher B (2004) Cancer of the breast. In: DeVita VT, Hellman S, Rosenberg SA (eds) Cancer: principles & practice of oncology. Lippincott Williams & Wilkins, Philadelphia, pp 914–970
47. Lehninger L, Nelson DL, Cox M (2004) Lehninger principles of biochemistry, 4th edn. W. H. Freeman & Co, New York, p 1100
48. Hata TR, Scholz TA, Ermakov IV, McClane RW, Khachik F, Gellermann W, Pershing LK (2000) Non-invasive Raman spectroscopy detection of carotenoids in human skin. J Invest Dermatol 115:441–448
49. Kaminaka S, Yamazaki H, Ito T, Kohda E, Hamaguchi HO (2001) Near-infrared Raman spectroscopy of human lung tissues: possibility of molecular-level cancer diagnosis. J Raman Spectrosc 32:139–141
50. Bohr J, Olsen K (2011) The close-packed triple helix as a possible new structural motif for collagen. Theor Chem Acc. doi:10.1007/s00214-010-0761-3
51. Thakur JS, Dai H, Serhatkulu GK, Naik R, Naik VM, Cao A, Pandya A, Auner GW, Rabah R, Klein MD, Freeman C (2007) Raman spectral signatures of mouse mammary tissue and associated lymph nodes: normal, tumor and mastitis. J Raman Spectrosc 38:127–134
52. Oliveira AP, Bitar RA, Silveira L, Zângaro RA, Martin AA (2006) Near-infrared Raman spectroscopy for oral carcinoma diagnosis. Photomed Laser Surg 24:348–353
53. Stone N, Stavroulaki P, Kendall C, Birchall M, Barr H (2000) Raman spectroscopy for early detection of laryngeal malignancy: preliminary results. Laryngoscope 110:1756–1763