Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

Discriminating adenocarcinoma from normal colonic mucosa through deconvolution of Raman spectra

Patricia Cambraia Lopes
Joaquim Agostinho Moreira
Abilio Almeida
Artur Esteves
Ivan Gregora
Martin Ledinsky
Jose Machado Lopes
Rui Henrique
Albino Oliveira

SPIE

Discriminating adenocarcinoma from normal colonic mucosa through deconvolution of Raman spectra

Patricia Cambraia Lopes,[a] Joaquim Agostinho Moreira,[a] Abilio Almeida,[a] Artur Esteves,[a] Ivan Gregora,[b] Martin Ledinsky,[b] Jose Machado Lopes,[c] Rui Henrique,[d,e] and Albino Oliveira[f]

[a]IFIMUP and IN–Institute of Nanoscience and Nanotechnology, Universidade do Porto, Departamento de Física e Astronomia da Faculdade de Ciências, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
[b]Academy of Sciences of the Czech Republic, Institute of Physics, Prague, Czech Republic
[c]Portuguese Oncology Institute, Department of Medical Oncology, Porto, Portugal
[d]Portuguese Oncology Institute, Department of Pathology, Porto, Portugal
[e]University of Porto, Department of Pathology and Molecular Immunology, Institute of Biomedical Sciences Abel Salazar (ICBAS), Porto, Portugal
[f]São Sebastião Hospital, Pathological Anatomy Laboratory, Espinho, Vila da Feira, Portugal

Abstract. In this work, we considered the feasibility of Raman spectroscopy for discriminating between adenocarcinomatous and normal mucosal formalin-fixed colonic tissues. Unlike earlier studies in colorectal cancer, a spectral deconvolution model was implemented to derive spectral information. Eleven samples of human colon were used, and 55 spectra were analyzed. Each spectrum was resolved into 25 bands from 975 to 1720 cm^{-1}, where modes of proteins, lipids, and nucleic acids are observed. From a comparative study of band intensities, those presenting higher differences between tissue types were correlated to biochemical assignments. Results from fitting procedure were further used as inputs for linear discriminant analysis, where combinations of band intensities and intensity ratios were tested, yielding accuracies up to 81%. This analysis yields objective discriminating parameters after fitting optimization. The bands with higher diagnosis relevance detected by spectra deconvolution enable to confine the study to some spectral regions instead of broader ranges. A critical view upon limitations of this approach is presented, along with a comparison of our results to earlier ones obtained in fresh colonic tissues. This enabled to assess the effect of formalin fixation in colonic tissues, and determine its relevance in the present analysis. [DOI: 10.1117/1.3658756]

Keywords: biomedical optics; Raman spectroscopy; deconvolution; data processing.

Paper 11376RR received Jul. 17, 2011; revised manuscript received Oct. 14, 2011; accepted for publication Oct. 14, 2011; published online Nov. 23, 2011.

1 Introduction

1.1 *Motivation*

Colorectal adenocarcinoma is among the most prevalent type of cancer in developed countries. In terms of incidence, gastrointestinal (GI) neoplasms have been reported to be the leading oncological pathologies by the Portuguese Oncology Institute (IPO-Porto), representing 23% of all diagnosed cases in 2008. Colorectal cancer accounts itself for 14% and 10% incidence rates in men and women, respectively.[1] In the United States of America, colorectal cancer ranks third in incidence and mortality, being represented by 10% and 9% rates, respectively, for both men and women in 2009.[2]

About 96% of all colorectal cancers are adenocarcinomas, which arise from the epithelial lining of glandular tissue in the colonic mucosa, its outermost layer, and typically evolve from benign neoplasms known as adenomatous polyps, through the adenoma-carcinoma sequence, a process that can take 8 to 10 years.[3,4] The early screening of this pathology is a crucial factor toward the improvement of clinical outcomes, by increasing the probability of tumor control with treatment, thus avoiding dissemination of the cancerous cells to distant tissues and organs (metastasis) and enabling its complete surgical removal. A colonoscopy is currently the gold standard for local screening of colorectal cancer. This endoscopic procedure relies on direct white-light observation of the whole large intestine, at the same time enabling endoscopic mucosal resection of suspect lesions and potentially premalignant polyps (polypectomy), by colonoscopy-directed biopsy. The biopsy specimens are, in turn, subject to histopathological analysis, the gold standard procedure for diagnosis.

As an alternative to endoscopic procedure, noninvasive imaging tools are under research. Among them, virtual colonoscopy reconstructs the interior of the colon from computed-tomography (CT) or nuclear magnetic resonance scans. So far this technique has some drawbacks, such as the use of ionizing radiation in the case of CT scans and being unable to identify lesions smaller than 5 mm.[5] Moreover, a real limitation of the CT technique is that structural information is collected by essentially absorption and scattering within the tissues, while the information regarding chemical composition is lost. Furthermore, a standard colonoscopy is required whenever there is evidence for some disorder, as a means to perform accurate diagnosis. Nevertheless, the endoscopic inspection of oncological pathology in the colon is far from being trivial. Although most colorectal cancers have a distinct appearance with light endoscopy, the existence of malignant neoplasms with atypical patterns have been

Address all correspondence to: Abilio Almeida, University of Porto, Faculty of Sciences, Rua do Campo Alegre, 687 Porto, 4169-007 Portugal; Tel: 220402348; E-mail: amalmeid@fc.up.pt.

1083-3668/2011/16(12)/127001/12/$25.00

reported.[6] Moreover, in some common diseases the visual identification of premalignant manifestations is significantly hampered, as it is the case of small and flat adenomas, in the presence of inflamed mucosa of chronic ulcerative colitis, and in differentiating adenomatous (dysplastic) from benign (hyperplastic) or non-neoplastic (metaplastic, inflammatory) polyps.[4,6] Similar issues are present in the case of gastric cancer.[7]

There has been a remarkable effort by the scientific community in the research of novel efficient methods that aim at providing *in vivo* real-time diagnosis of GI lesions.[6,8–10] Examples of such optical adjuncts to a colonoscopy are fluorescence spectroscopy and imaging, light-scattering spectroscopies, as is the case of Raman spectroscopy (RS), optical coherence tomography, chromoendoscopy, narrow-band imaging, confocal fluorescence endoscopy, immunophotodetection, and more recently, terahertz imaging and spectroscopy.[4,11–13] These techniques may yield invaluable tools if they are to provide the physicians the ability for accurate identification of suspect lesions *in situ*, alongside high sampling yields, enabling margin delineation and reducing the diagnosis waiting time and the costs and risks associated with unnecessary biopsies.[14]

RS has been reported to be the spectroscopic technique which provides the most detailed information about both the chemical composition and structure of tissues, yielding much more significant spectral features than those obtained from fluorescence or elastic scattering.[9] In fact, there has been notable interest in the research of RS applications for cancer diagnosis, most notably for *ex vivo* studies of neoplasms affecting the breast, lung, skin, uterus, and GI tract. RS relies on the detection of inelastically scattered light by the molecular optical vibrations. The frequency and profile of Raman bands are highly specific to both molecular structure and normal vibrations of the chemical species present in the scatter system. RS is harmless and suitable for *in vivo* applications, as nonionizing and nonmutagenic radiation wavelengths, ranging from visible to near-infrared (NIR), are used for excitation. In a traditional microscope setup and due to the absorption of visible light by the living tissues and the inherently weak Raman cross-section of biomolecules, only signals coming from near the illuminated surface (~0.5 to 1 mm) are collected, which is advantageous for epithelial tissue screening, avoiding superposition of information from internal layers.[9,15] In addition, usually RS requires no sample preparation or use of dyes, tracers, or targeting drugs, in contrast to other optical techniques, which renders its diagnostic capabilities independent of the degree of selective localization of such agents within lesions.

Despite all the advantages, RS is technically challenging to implement at a clinical level, mostly due to the spectral contamination of a Raman signal by strong broadband fluorescence background, caused by tissue endogenous fluorophores. Nevertheless, RS has highly profited from technological advances in lasers, spectrometers with increasing resolution, high sensitivity detector devices with fast response, as well as dedicated optical fiber probes with suppression of fluorescence in the fiber optics, which further encourage *in vivo* applications.[16,17] Moreover, it has been suggested that the use of typical laser powers of 100 to 250 mW focused to spot sizes of 500 μm, used to record Raman spectra of biological tissues, should not imply significant damage caused by heating.[14] However, special care is needed in order to avoid tissue damage if RS is to be used toward clinical applications, which can be better achieved by following ANSI established standards for laser safety.

1.2 *Outline and Trends in RS for Colorectal Cancer Diagnosis*

Raman scattering studies on colonic tissue to date have focused on probing fresh samples, using NIR excitation, for spectral analyses in the so-called fingerprint region, between ~800 and 1800 cm^{-1}, where the main vibrational signatures of proteins, lipids, carbohydrates, and nucleic acids are observed, with minimal influence from water.[17] First *in vivo* Raman spectra of human GI tissues measured during routine clinical endoscopy, were reported by Shim et al.,[18] with an acceptable signal-to-noise ratio (SNR) and short collection time (5 s), thus demonstrating the feasibility of minimally invasive *in vivo* RS, by means of a fiber optic probe incorporated into the instrument channel of an endoscope.[19] Together with Shim and co-workers, Molckovsky[14] proceeded with a proof-of-concept study, showing promising results in differentiating adenomatous from hyperplastic polyps, by principal component analyses (PCA) followed by linear discriminant analysis (LDA), as a classification tool applied to Raman spectra obtained in both *in vivo* and *ex vivo* conditions. Though LDA/PCA algorithms identified adenomas with high accuracy, classifiers developed *in vivo* and *ex vivo* showed poor correlation.

Ex vivo studies performed so far on RS in the field of colorectal cancer diagnosis have been also based, to our knowledge, on several multivariate statistical techniques of variable complexity, in order to implement a reliable spectral classification. Remarkable results were achieved by Stone et al.[20] using LDA, by Huang et al.,[21] and Widjaja et al.[22] with support vector machines, and by Chowdary et al.[23] applying Mahalanobis distance and spectral residuals as discriminating parameters, with each methodology being performed after data reduction by PCA. However, these methodologies give little insight on the origin of the contrast between different spectral classes, due to relying on changes in parameters space, and results from several independent studies are difficult to compare. Possible drawbacks related to tissue diagnosis based on statistical algorithms are the constant need to test unknown spectra against large sets of reference spectra, acquired with identical conditions for each tissue type, and the difficult handling and control of relevant variables from the standpoint of the user. Nevertheless, the need to apply statistical tools on the overall spectral range arises due to the fact that there are no bands whose presence definitely proves the existence and nature of a tumor.[15] In particular, spectral changes between epithelial neoplasms and normal tissue appear to be restricted to small intensity variations, rather than steep differences in spectral profiles, which is evident from the mean spectra from various organs reported by Stone et al.[20,24] On the contrary, the spectrum of breast carcinoma is markedly different from that of corresponding normal tissue since the latter is dominated by the spectrum of fat, a stronger Raman scatter than proteins, that is lacking in cancerous tissue.[20,25]

The present work is addressed to discriminating adenocarcinomatous from normal mucosal colonic tissues using Raman spectroscopy. In contrast to previously used methods toward analyzing Raman spectra of colonic tissues, we have used a classical model where spectral deconvolution is carried out by a sum

of damped oscillators.[26] The aim of this analysis is to compare spectral intensities associated with individual component bands, resolved from the Raman spectra obtained from adenocarcinomatous and normal mucosal colonic tissues, and subsequently investigate whether the differences in intensities are amenable of providing suitable means of discrimination between the tissue types. Linear discriminant analysis was used for this purpose. This study is intended to provide a first insight on the discriminating capabilities of this type of analysis, as well as a better understanding on the biochemical correlation to the most significant spectral changes between the two tissue types. Since tissue specimens examined in the present work were under formalin fixation, a comparative analysis between our results and those previously reported on fresh tissues is also presented.

2 Experimental and Data Analysis

2.1 *Sample Characterization and Processing*

Eleven tissue specimens were studied; each one from a different individual. Samples consisted of surgically resected transverse sections of human colonic tissues, with lateral dimensions ranging from 2 to 6 cm and thickness between 1.5 and 5 mm. Except for one entirely nonpathological sample, all the others contained both normal and affected tissues (adenocarcinoma). These specimens were gently sent from the Department of Pathology, Portuguese Oncology Institute of Porto, Portugal and Pathological Anatomy Laboratory, Espinho, Portugal, after histopathological analysis and formalin fixation. Histopathological diagnosis and margin delineation was performed for each specimen by the expert pathologists, via microscopical examination of their adjacent sections stained with hematoxylin and eosin (H&E). A total of 55 Raman spectra from colonic tissue samples were recorded, comprising 23 spectra from normal mucosa and 32 spectra from adenocarcinoma, which correspond on average to 2 or 3 spectra sampled per tissue specimen per region (normal or cancerous). Every spectrum was equally accounted for throughout the data analyses, as there was no specific reason to consider any kind of field effect. Cancer-assigned spectra were acquired in regions were macroscopic disease could be observed, whereas normal-assigned spectra were recorded at a distance not less than 0.5 cm from the margins. An H&E-stained adjacent section of a sample is illustrated in Fig. 1, where the corresponding points of interest for Raman measurements are depicted.

The fixative used after excision was 10% neutral buffered formalin (4% formaldehyde methanol-free aqueous solution), a most convenient way to store the specimens at room temperature. The fixation procedure preserves tissues by promptly interrupting its metabolism and stabilizing its structure, due to promoting cross-linkage of amine groups in tissue proteins.[27] The use of the fixative is needed to circumvent degradation due to the temporal delay between sample collection and Raman measurements. Prior to the measurements, samples were rinsed with running water and the excess liquid on the surface was absorbed.

2.2 *Micro-Raman Spectroscopy*

Ex vivo Raman measurements over the spectral range 800 to 1800 cm^{-1} were accomplished with a dispersive-type Renishaw microspectrometer, InVia Reflex model, equipped with a Renishaw RenCam CCD detector and a single 1200 lines/mm dispersion grating, where the display system comprises a Leica microscope. Excitation source was provided by a Renishaw diode laser, RL 785 type, emitting at 785 nm. Near-infrared illumination was chosen to reduce tissue autofluorescence. Due to the macroscopic dimensions of the samples, a ultralong working distance (N.A. 0.65) objective lens was used, with 50× magnification, to focus the laser beam to a spot size of 2 to 3 μm, and collect the backscattered photons in nonconfocal mode. The monochromator entrance slit was set to 65 μm, yielding a spectral resolution of 2 cm^{-1}. While the majority of spectra was recorded with ~1 $mW/\mu m^2$ laser intensity on the sample, with one accumulation using 100 s integration time, a few spectra were acquired with ~5 $mW/\mu m^2$ laser intensity, and one accumulation with 50 s integration time.

Fig. 1 Example of an H&E-stained adjacent section used for histopathological examination and margin delineation. The trace surrounds the cancerous region, whereas the stars indicate the positions of the measuring points, located near the outer intestinal lining: open star stands for normal mucosa and closed star for adenocarcinoma.

2.3 *Data Preprocessing*

In order to correct for overall intensity and background fluctuations on the total data set, the normalization criteria adopted comprises the following steps: 1. background subtraction, 2. spectral interpolation, and 3. intensity correction. Backgrounds were estimated using LabSpec® 5 spectroscopy software by fitting a fifth-order polynomial to each spectrum, as is common practice for tissue fluorescence elimination.[27] The fifth-order polynomial yields a good compromise for background subtraction in our spectra, since it has a high enough order to fit the background, without jeopardizing the quality of the Raman bands. Linear interpolation was performed with Igor Pro® 6.0 (WaveMetrics), in order to fix the same Raman shift vector

with equal spectral increment for all spectra ($\Delta\omega = 0.9\ \mathrm{cm}^{-1}$), without affecting spectral profiles. This procedure enabled to compute, e.g., mean spectrum of overall spectral sets, as well as using a faster fitting algorithm. Intensity correction was accomplished by total area normalization, i.e., by dividing spectral intensities by a scaling constant—the sum of all discrete intensity values in the spectral range 800 to 1800 cm^{-1}—which yielded an absolute value of 0.9 for the area under every normalized spectrum, corresponding to the constant spectral increment.

2.4 *Spectral Deconvolution and Classification*

Spectral deconvolution was performed in all normalized spectra in order to compare intensities of different Raman bands between tissue types. This analysis was performed by least-squares fitting of Eq. (1) to each spectrum, which describes a model involving a sum of independent damped and harmonic oscillators:[26]

$$I(\omega) = [1 + n(\omega, T)] \sum_{j=1}^{N} A_{0j} \frac{\omega \Omega_{0j}^2 \Gamma_{0j}}{\left(\Omega_{0j}^2 - \omega^2\right)^2 + (\omega \Gamma_{0j})^2}. \quad (1)$$

According to Eq. (1), the fitted curve *I(ω)* is given by a superposition of a set of *N* bands, with the *j*'th band representing an oscillator with a characteristic frequency Ω_{0j}, amplitude A_{0j}, and full width at half maximum Γ_{0j} (damping factor). *n(ω,T)* is the temperature-dependent Bose–Einstein factor. Oscillator bands are approximately of Lorentzian shape in the frequency range considered. This analysis was implemented by means of an improved in-built routine, running at Igor Pro® 6.0 (WaveMetrics).

Curve fitting was applied to the spectral range from 975 to 1720 cm^{-1}, where the most prominent and systematic bands were observed. This range was split into three separate subranges for easier implementation. Both normal and cancerous spectra were resolved into 25 component bands, in the 975 to 1720 cm^{-1} range. A constant baseline was used and kept unchanged, to yield better comparison between the fitting results from normalized spectra, and no smoothing was applied prior to the fitting. This baseline stands just as an additive constant to Eq. (1), without modifying the spectral profiles. Initial parameters were estimated by applying curve fitting to the average spectra from normal and cancerous sets. It was observed that the frequencies of resolved bands from the two average spectra were in very good agreement ($\pm 2\ \mathrm{cm}^{-1}$, data not shown), due to the high similarity between spectral profiles. Initial parameters were then introduced in the model, in order to fit the spectra with a higher SNR. It was found that individual bandwidths did not differ substantially from spectrum to spectrum after fitting optimization of these spectra, and thus, in a next step, we chose to apply average bandwidth values to every spectrum and to keep them fixed. In this way, we could implement a systematic analysis that otherwise would be cumbersome and subjective, having in mind that spectral deconvolution in the context of biologic spectra, containing multiple close to and overlapped bands, is not straightforward, especially when dealing with *in vivo* spectra having considerable noise due to constraints in acquisition time and laser power.

The band intensities (i.e., areas under the resolved curves) from normal tissue spectra were then compared against those from cancerous tissue spectra, independently of the specimen probed (extrinsic variability) and the site on each sample region (intrinsic variability). Statistical comparison was performed using PASW® Statistics 18.0. A student's t-test was used to compare band intensities between spectra from normal and cancerous tissues for each component band, after confirming normal distribution of the data samples by using the Kolmogorov–Smirnov test. A significance level of 5% was considered.

LDA was applied to intensity distributions, as well as intensity ratio distributions, in order to estimate the accuracy of discrimination between normal and cancerous tissues, based on the confined spectral information concerning those bands, that individual or in combination, yielded the most significant differences between them. An inbuilt routine of MATLAB® 7.1 was used for LDA. This multivariate tool performs supervised classification by taking into account the histopathological analysis corresponding to each measured spectrum, in order to estimate a discriminant function. This function is described by a linear combination of a set of parameters, which maximizes the ratio of intergroup variance to intragroup variance.[20]

3 Results and Discussion

3.1 *Average Spectral Profiles*

Average Raman spectra of normal and adenocarcinomatous tissues are presented in Fig. 2. It can be observed from Fig. 2 that the characteristic spectra of normal and cancerous tissues

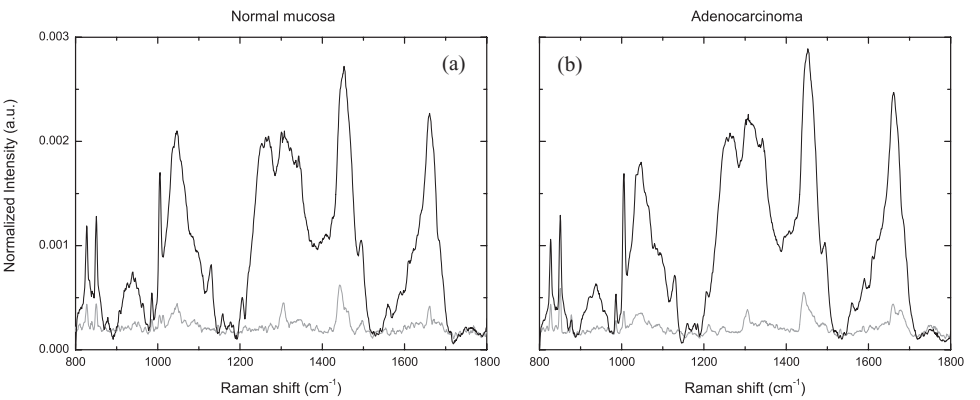

Fig. 2 Average Raman spectra (black lines) and standard deviation spectra (gray lines) of formalin-fixed colonic tissues. (a) Normal mucosa ($n = 23$); (b) Adenocarcinoma ($n = 32$).

are similar in shape and intensity. From visual inspection, spectral profiles seem to be in good agreement with those reported by Krafft et al.,[28,29] acquired with equipment similar to the one used in this work, and by Andrade et al.[30] with FT-Raman spectroscopy, both regarding spectra of fresh healthy tissues. Despite the similarities referred to above, some differences are evidenced in our spectra regarding the two bands located at ~1046 and 1495 cm^{-1}. They can be assigned to formalin vibrations,[31] and its presence in other formalin-fixed biologic tissues, has been documented elsewhere.[32] In comparison to our results, Raman spectra of fresh colonic tissues reported elsewhere appear to be much less intense in the spectral region around 1046 cm^{-1}, whereas the band at 1495 cm^{-1} is not observed at all. Though our study reveals that formalin fixation does not yield large changes toward the vibrational phenotype of fresh colonic tissues, spectral contamination stemming from formalin bands must be considered, as will be touched on in Sec. 3.2.

Also shown in Fig. 2 is the spectral variability of each group, represented in terms of the standard deviation of spectral intensities. Spectral variability has similar patterns for both normal and cancerous classes and appears to be more pronounced within the spectral ranges 1000 to 1100, 1290 to 1360, 1425 to 1500, and 1550 to 1700 cm^{-1}. These fluctuations may reflect the natural heterogeneity of each tissue type, which is reinforced by the similarity with standard deviation spectra reported elsewhere for the case of normal epithelial tissue,[30] especially in the former and latter spectral intervals. The peaks observed at approximately 1300 and 1445 cm^{-1} in our standard deviation spectra are mainly attributed to two lipid bands, which have characteristic steep and intense profiles, in contrast to the average spectra profiles, and which presence was observed only in a few spectra from both sorts of tissues.

Direct comparison between average spectra of normal versus cancerous tissues is depicted in Fig. 3. According to these results, an inversion in spectral intensity differences takes place around 1200 cm^{-1}, which evidences that the intensity of the Raman signal is higher in healthy tissue than in cancerous tissue below 1200 cm^{-1}, and vice versa above 1200 cm^{-1}. The latter region includes typical prominent bands arising from vibrational modes of proteins amide I and III, and CH_2 bending and twisting modes of proteins and lipids, together with other bands from nucleic acids and aromatic aminoacids, described elsewhere.[17,27,33] The magnitude of maximum intensity differences found in the regions of interest are as follows: 1010 to 1135 cm^{-1} (~15%), 1300 to 1370 cm^{-1} (~10%), 1435 to 1480 cm^{-1} (~8%), 1560 to 1600 cm^{-1} (~20%), and 1600 to 1690 cm^{-1} (10% to 20%).

In general, the observed higher intensity contribution to the average spectrum by cancerous tissue agrees with previous results for regions around 1330 and 1600 cm^{-1} or 1300 to 1345 cm^{-1} and 1530 to 1654 cm^{-1}.[22,23] A remarkable agreement can be seen between both our average spectra profiles and intensity differences, against those obtained by Stone et al.,[20] except for the range 1550 to 1620 cm^{-1} that yields an opposite difference. Difference spectra reported by Stone et al.[20] indeed resemble a similar pattern, mainly for maximum peak differences around 1320, 1336, 1443, and 1654 cm^{-1}, with the intensity differences representing up to 10% variation, which is in good agreement with our results. Regarding the lower frequency spectral region, the higher intensity registered for a normal tissue against a cancerous one is consistent with two aforementioned references.[20,22] However, we were unable to find a higher contribution from the phenylalanine symmetric ring-breathing mode [νs(C–C), 1005 cm^{-1}] in contrast to the former two reports, which is probably due to the nearby presence of a formalin band that is much broader.

At this stage, we could remark that beyond differences in spectral intensities referred to above, it is also apparent that spectral variability within both groups is not negligible. Moreover, if band intensities are to be used for tissue discrimination, the overall differences should prevail over the point-by-point variations that cannot be avoided in principle.

3.2 *Band Intensities of Normal Versus Cancerous Tissues*

Spectral deconvolution was systematically performed in order to sort out the most significant differences between spectra obtained from normal mucosa and adenocarcinoma tissues, regarding the intensities of Raman bands. Figure 4 shows examples of the output obtained by curve-fitting analyses, as described in Sec. 2.4. The spectral ranges considered, I: 975 to 1150 cm^{-1}, II: 1190 to 1530 cm^{-1}, III: 1550 to 1720 cm^{-1}, were decomposed into 8, 11, and 6 bands, respectively. Region I includes the very distinct phenyl band (1005 cm^{-1}) and additional bands arising mainly from stretching modes of single bounds (C–C, C–O, C–N, O–P–O), that are mostly assigned to lipids and carbohydrates.[17,30,33,34] Amide III and δCH_2 bands are present in spectral region II, and amide I in region III.[14] From Fig. 4, it can be seen that the formalin band situated at 1495 cm^{-1} [δ(HCH), see Ref. 31] can be easily resolved, thus yielding little or negligible influence to nearby bands. For this reason, its presence (and thus the presence of formalin in tissues) does not represent a major issue in our opinion, regarding the spectral information that can be derived from the neighbor bands, if curve-fitting analysis is applied. The same conclusion is not straightforward for the broad and more intense formalin band at 1046 cm^{-1} [νa(OCO, see Fig. 4 and Ref. 31]. This band is actually decomposed into two bands, when dealing with significantly higher concentrations of formaldehyde,[31] but in the present case it can be represented by a single band.

For the subsequent analysis, two spectra from normal tissues and five from cancerous tissues were further discarded, since their appearance differed significantly from the majority of the measured spectra. The differences stem from the presence of fat-like very pronounced spectral bands, especially those appearing at 1304, 1446, and 1662 cm^{-1}. Since fat is a stronger Raman scattering material than protein, quantitative comparison of spectral band intensities when Raman spectra with high fat contribution is also considered, is very difficult to make if total area normalization is to be applied, as is the case in the present analysis.

Intensity distributions of component bands were obtained and a comparison was made between normal- and cancerous-assigned spectra. The intensity distributions that presented significant differences (t test) are associated with the following bands: 1019 cm^{-1} ($p = 0.001$), 1046 cm^{-1} ($p = 0.007$), 1065 cm^{-1} ($p < 0.001$), 1323 cm^{-1} ($p < 0.001$), 1345 cm^{-1} ($p < 0.001$), 1467 cm^{-1} ($p < 0.001$), 1588 cm^{-1} ($p = 0.004$),

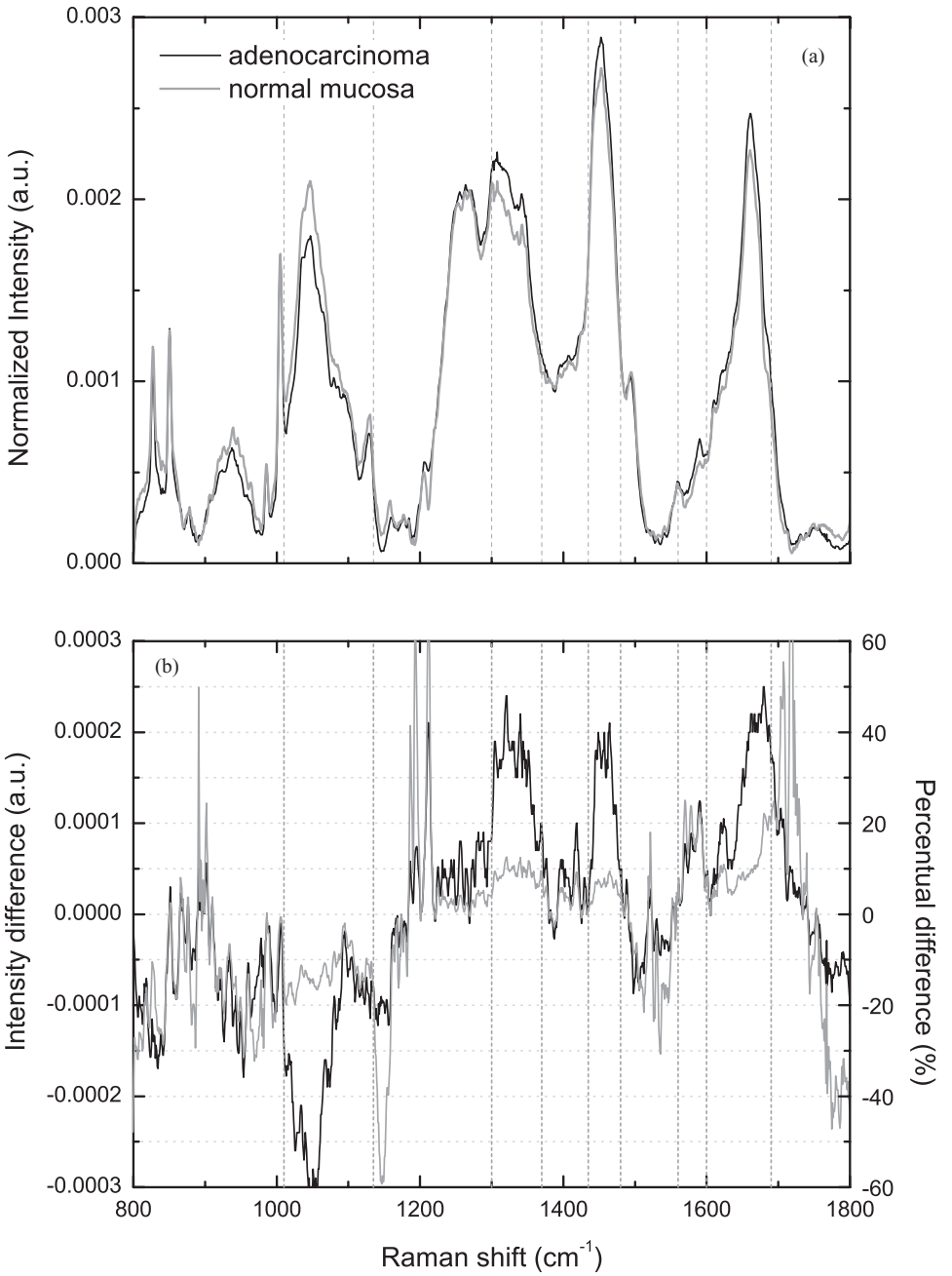

Fig. 3 Comparison between spectral profiles of cancerous versus normal colonic tissues. (a) Average Raman spectra of normal mucosa (gray line) and adenocarcinoma (black line); (b) Difference spectra (cancerous to normal) expressed in values of absolute intensity (left axis, black line) and percentage, relative to the normal average spectrum (right axis, gray line). Vertical lines correspond to the intervals 1010 to 1135, 1300 to 1370, 1435 to 1480, 1560 to 1600, and 1600 to 1690 cm^{-1} in units of Raman shift, respectively.

1662 cm^{-1} ($p = 0.001$), and 1685 cm^{-1} ($p < 0.001$). These intensity distributions are presented in Fig. 5, and the corresponding bands are evidenced by arrows in Fig. 4. As expected, overall intensities of curve-resolved bands in spectral region I tend to be higher for normal tissues than in cancerous tissues, and vice versa for the spectral regions II and III. It is worth noting though that there is a considerable variability, with consequent partial superposition of intensity values.

Assuming that the intensity of the Raman band is proportional to the number of oscillators, relative band intensities can give a measure of the relative amount of scattering species. For this purpose, tentative assignments of major Raman bands shown in Fig. 5 are presented in Table 1, following earlier reported results.[27,34,35] Our results suggest that nucleic acids (bands at 1323 cm^{-1}, 1345 cm^{-1}, and possibly 1662 cm^{-1}), aminoacids (hydroxyproline and aromatic phenylalanine, 1588 cm^{-1}), collagen (1467 cm^{-1}, 1662 cm^{-1}), and disordered protein conformations (1685 cm^{-1}) are more predominant in colonic adenocarcinoma, whereas lipids (1065 cm^{-1}) and carbohydrates (1019 cm^{-1}) are more

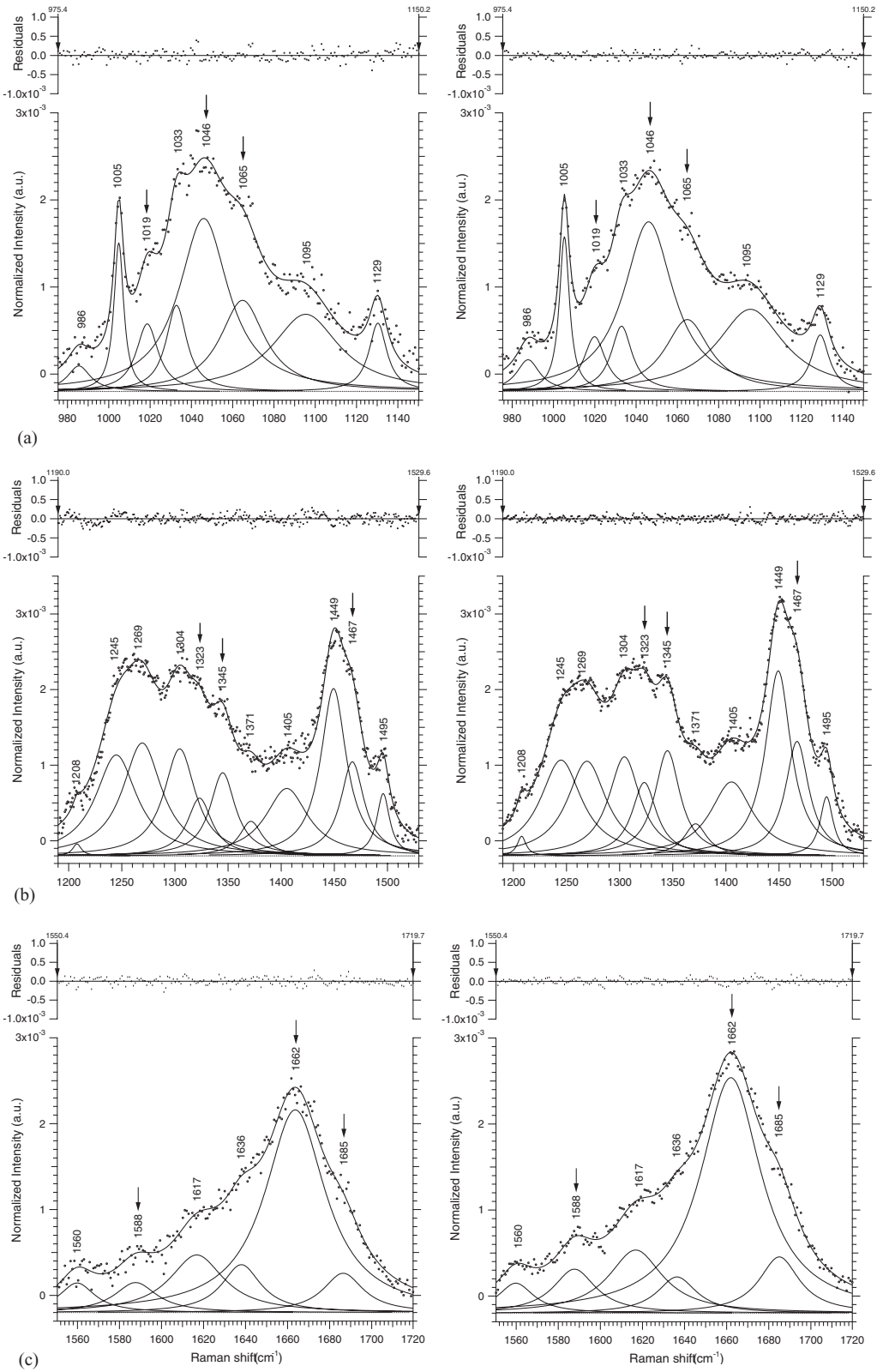

Fig. 4 Examples of the output of the fitting model applied to Raman spectra of normal colonic mucosa (left column) and adenocarcinoma (right column), using three separate spectral windows: (a), (b), and (c). Residuals represent differences between the fitted trace (solid curve) and the experimental points after normalization (dotted). The fitted trace is given by the superposition of all component bands plus a constant baseline. The arrows indicate those bands referred to in Fig. 5.

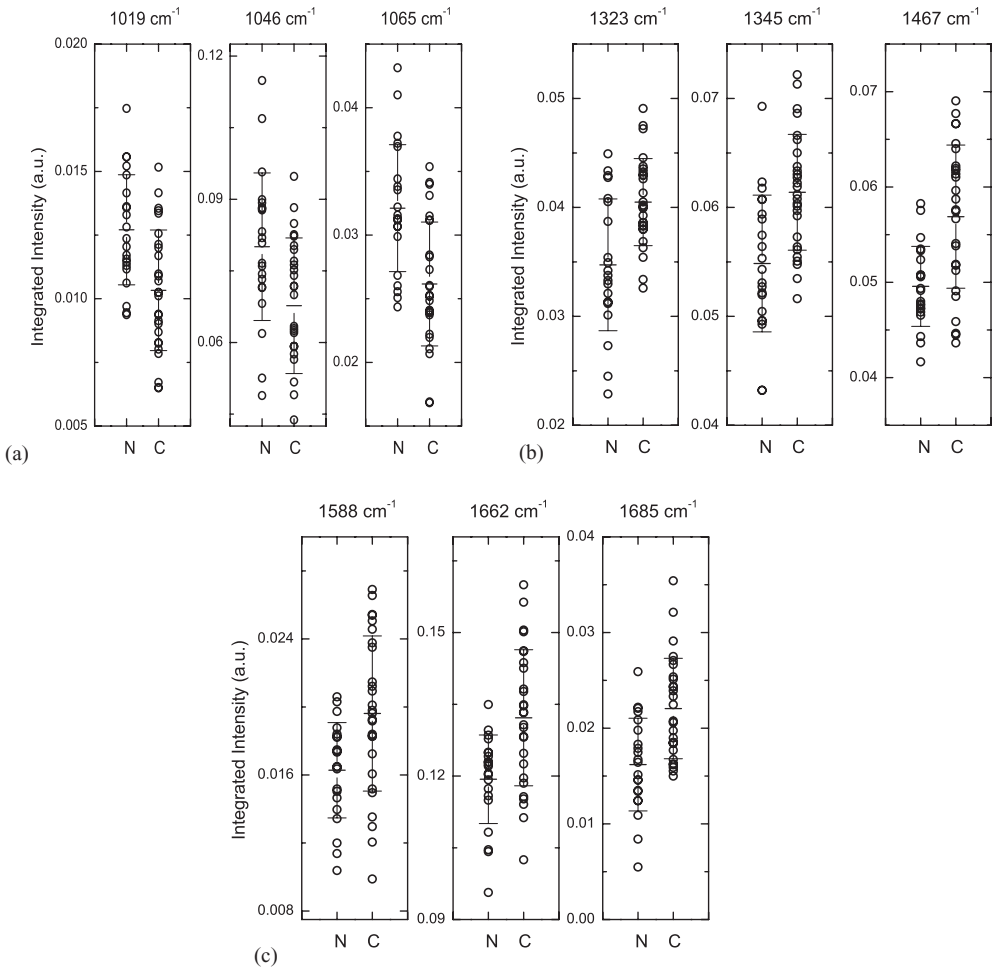

Fig. 5 Intensity distributions of curve-resolved bands, shown separately for the spectral sets of normal colonic mucosa (N) and adenocarcinoma (C). Each band is labeled by its characteristic Raman shift and bands are grouped for the spectral regions (a), (b), and (c), accordingly to Fig. 4.

abundant in normal colonic mucosa. In fact, the higher nucleic content in cancerous tissue is a general result for various neoplasms studied by RS,[36] showing that this technique is sensible to the increased nucleus-to-cytoplasm ratio, typical of neoplastic tissue with high proliferation rate.

The formalin band at 1046 cm^{-1} yields a significant higher intensity for the case of normal tissue (see Fig. 5), in contrast to what was observed for the other formalin band at 1495 cm^{-1} ($p = 0.154$, not shown). The former band appears to be superimposed in a spectral range containing tissue bands

Table 1 Tentative assignments of major vibrational modes contributing to specific curve-resolved bands, from Raman spectra of mucosal colonic tissue Refs. 27 and 34, and 35).

Band position (cm^{-1})	Major assignments
1019	ν(C–O) ribose; glycogen
1065	skeletal ν(C–C) trans in lipids
1323	Guanine (B, Z marker); CH_3CH_2 wagging and deforming in collagen and purine bases of nucleic acids
1345	CH_3, CH_2 wagging collagen; guanine (DNA, RNA)
1467	$\delta_{as}(CH_3)$ and $\delta(CH_2)$ of proteins
1588	Phenylalanine, hydroxyproline
1662	Amide I (proteins); Nucleic acid modes; ν(C = C) *cis* fatty acids
1685	Amide I (disordered structure; non hydrogen bonded)

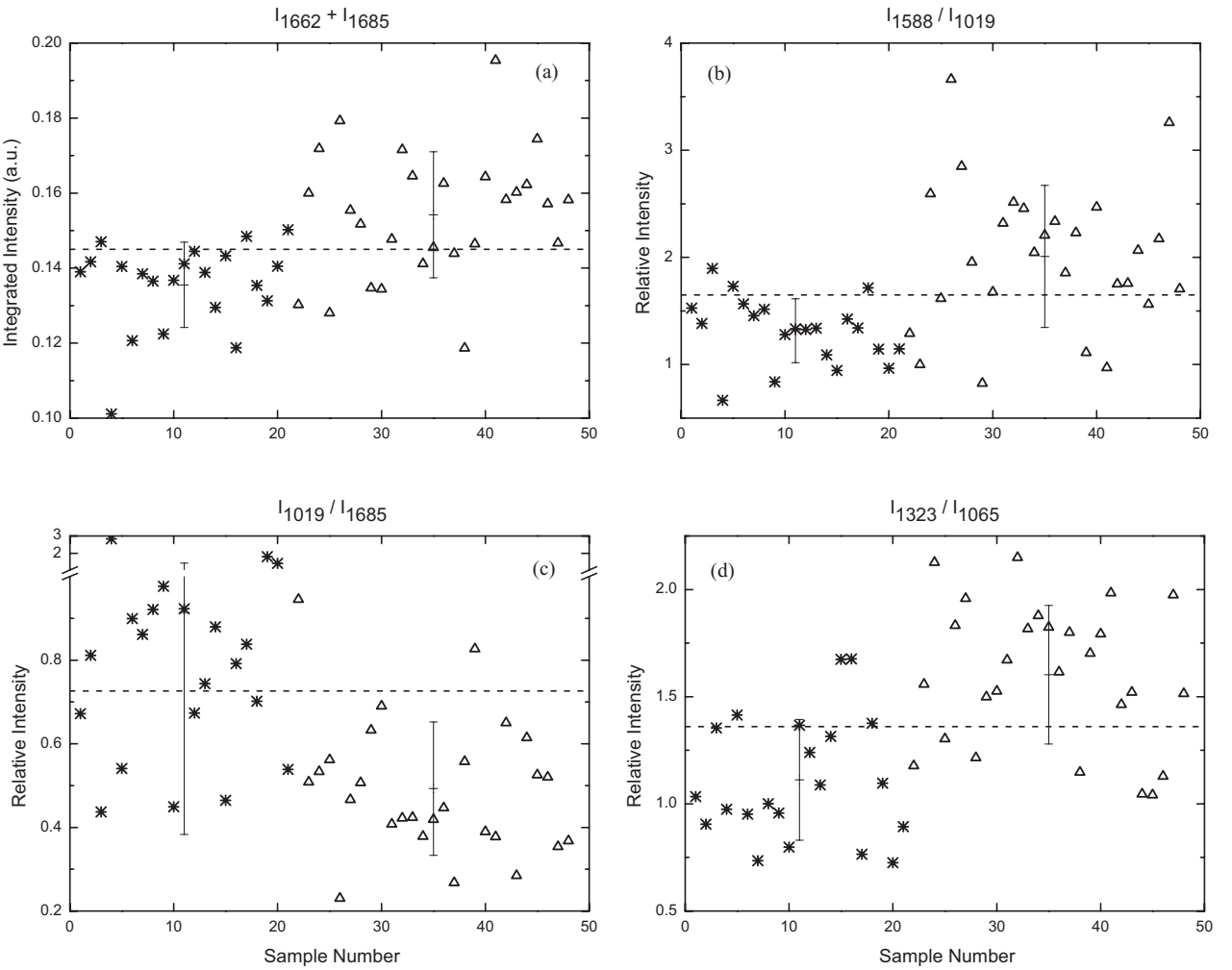

Fig. 6 Plots of distributions of summed intensities (a) and intensity ratios [(b), (c), and (d)] of curve-resolved bands, for the spectral sets of normal colonic mucosa (*) and adenocarcinoma (Δ). Bands are labeled by their characteristic mean value of Raman shift. Dashed lines separate overall data into the two groups, which were obtained after applying LDA, for each distribution.

with significantly lower intensity, and thus can possibly mask near and overlapped tissue-bands, in contrast to the other formalin band. As such, we believe that the influence of the formalin band at 1046 cm^{-1} in that part of the spectrum could not be totally accounted for by the present analysis. What this result suggests is that we could not resolve overlapped bands close to 1046 cm^{-1}, which may present themselves different spectral features between the tissue types. This possible issue may justify the significant difference observed for the band at 1495 cm^{-1}. If the influence of a formalin band resolved from some spectrum is to be neglected, then the subtraction of this band to the overall spectrum should yield a spectrum approximately resembling that of fresh tissue. This intensity distribution issue can thus only be clearly understood after performing similar experiments on fresh tissues.

3.3 *Diagnosis Potential of RS Using Combined Band Intensities*

Intensity distributions obtained by curve-fitting analysis were subject to a classification algorithm, LDA, as a means to estimate the sensitivity, specificity, and overall accuracy that can be achieved with our results. By sensitivity, we mean the fraction of true-positive events obtained by the model, relative to the total amount of positives (i.e., histopathology-certified cancerous states). While the specificity is the fraction of true-negative events (to total number of negative samples), overall accuracy is the ratio of true to total events, obtained by the discriminant classification. After applying LDA to the raw intensity distributions shown in Fig. 5, we were able to obtain accuracies as good as 75% (1065, 1323 cm^{-1}) or 73% (1662 cm^{-1}), and as bad as 65% (1685 cm^{-1}) or 60% (1449 cm^{-1}).

In order to enhance sensitivity (Se.), specificity (Sp.), and overall accuracy (Ac.) we also applied LDA to different sets of parameters. Better discriminating accuracies were achieved using summed intensities of two neighboring bands, as was the case for the parameter $I_{1662}+I_{1685}$ [Ac. 79%, Se. 74%, Sp. 86%—see Fig. 6(a)] or $I_{1449}+I_{1467}$ (Ac. 77%, Se. 74%, Sp. 81%). In addition, we exploited several combinations of intensity ratios between two different bands, despite using the absolute intensity values alone. An intensity ratio may be interpreted as a relative intensity of one band to another, which can provide a more precise indication of the relative concentrations of different tissue constituents. Furthermore, intensity

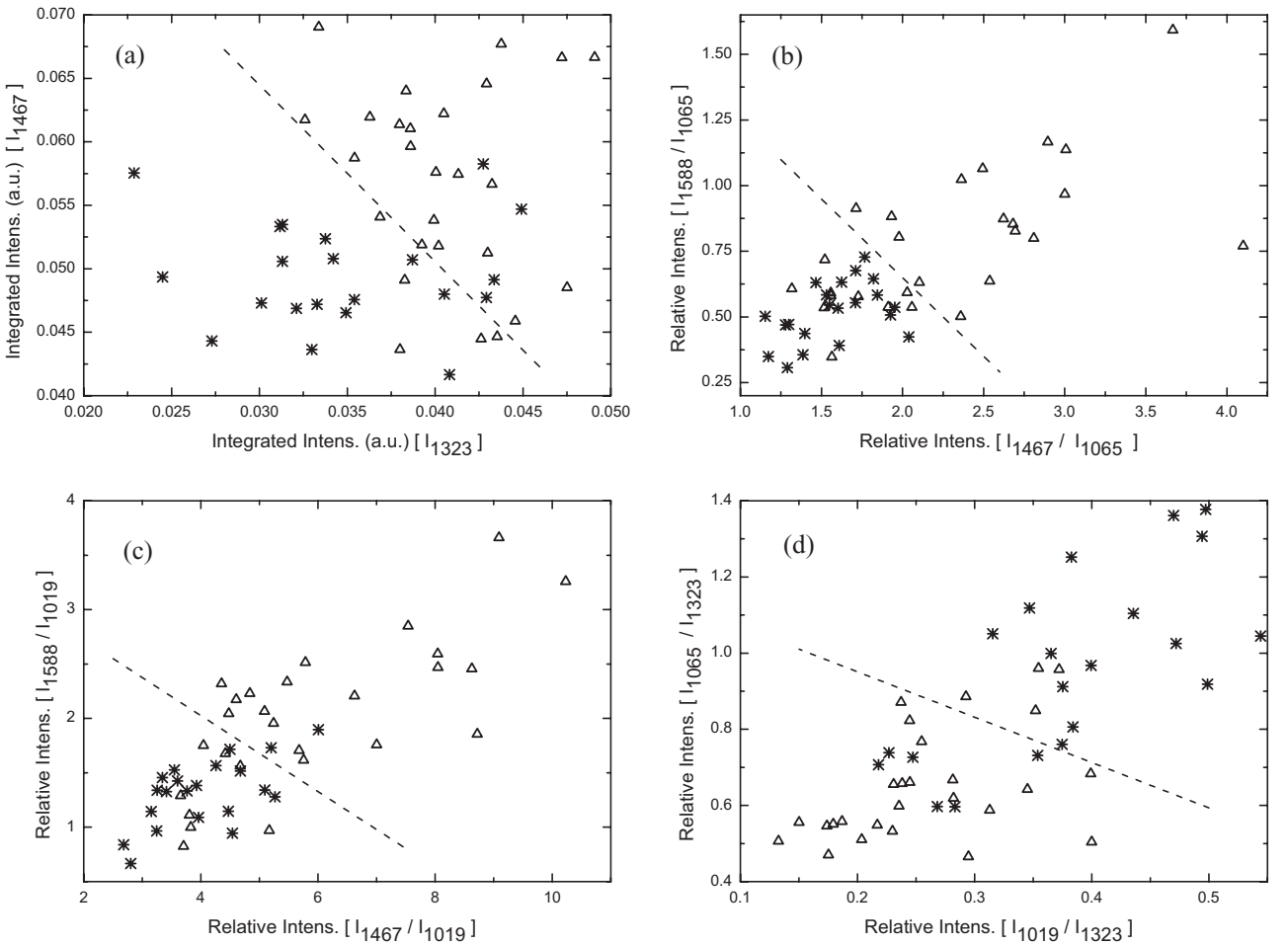

Fig. 7 Two-dimensional plots of intensities (a) and intensity ratios [(b), (c), and (d)] of curve-resolved bands, for each colonic spectrum from normal mucosa (*) and adenocarcinoma (Δ). Bands are labeled by their characteristic mean value of Raman shift. Dashed lines separate overall data into the two groups, which were obtained after applying LDA, for each 2D distribution.

ratios yield more interesting parameters from the standpoint of classification, because they represent dimensionless quantities, which are independent of the total spectral area considered and less influenced by the type of normalization. In this way, we obtained discriminating accuracies up to 79%. In Figs. 6(b)–6(d), we present a selection of the intensity ratio distributions, which yield better discrimination: I_{1588}/I_{1019} (Ac. 79%, Se. 74%, Sp. 86%), I_{1019}/I_{1685} (Ac. 79%, Se. 93%, Spec. 62%), and I_{1323}/I_{1065} (Ac. 75%, Se. 74%, Sp. 76%), respectively. We note that the error bar (standard deviation of the distribution) depicted in Fig. 6(c) for the normal set appears to be very big toward the lower values of relative intensity, which may not be realistic but is a result of the fact that three of the distribution values are considerably higher than the rest.

Finally, we also explored the possibility of applying a two-parameter discriminant analyses, rather than the single-parameter approach, that was previously used to generate a decision threshold value. In Fig. 7 we present a selection of the most relevant combinations of intensities [Fig. 7(a)] and intensity ratios [Figs. 7(b)–7(d)] from the standpoint of tissue discrimination. In these two-dimensional plots, each spectrum is represented in terms of two parameters. Spectral classification based on the pair (I_{1323}, I_{1467}) yielded 81% accuracy [Se. 81%, Sp. 81%—Fig. 7(a)], and 79% accuracy was obtained for the other three pairs: (I_{1467}/I_{1065}, I_{1588}/I_{1065}) [Se. 63%, Sp. 100%—Fig. 7(b)], (I_{1467}/I_{1019}, I_{1588}/I_{1019}) [Se. 70%, Sp. 90%—Fig. 7(c)] and (I_{1019}/I_{1323}, I_{1065}/I_{1323}) [Se. 85%, Sp. 71%—Fig. 7(d)]. A similar result to the latter was also found for the pair (I_{1019}/I_{1345}, I_{1065}/I_{1345}).

The two-dimensional (2D) plots with normalization with respect to the intensity of one particular band [I_{1065}, I_{1019}, or I_{1323}—Figs. 7(b)–7(d), respectively], show evidence for the clustering of normal and cancerous points in opposite directions. This feature gives rise to extreme regions of maximum probability for true classification, as well as an intermediate region with an inherent uncertainty in tissue classification, due to the partial superposition.

Classification results obtained for three sets of parameters are detailed in Table 2. The pair of parameters (I_{1323}, I_{1467}) yielded both sensitivity and specificity of 81%, whereas the pair of parameters (I_{1467}/I_{1065}, I_{1588}/I_{1065}) and parameter I_{1019}/I_{1685} represented the highest specificity (100%) and the highest sensitivity (93%) obtained, respectively. In general, it was noted that the sensitivity appeared to be more times poorer than the specificity in the situations studied, i.e., the number of false negatives appeared to be inferior to the number of false positives in

Table 2 Spectral classification by linear discriminant analysis, applied to curve-resolved band intensities or intensity ratios, using two variables [(I_{1323}, I_{1467}), Fig. 7(a); (I_{1467}/I_{1065}, I_{1588}/I_{1065}), Fig. 7(c)] or a single variable [(I_{1019}/I_{1685}), Fig. 6(c)]. Overall accuracies were 81%, 81%, and 79%, respectively.

		Spectral classification					
		(I_{1323}, I_{1467})		(I_{1467}/I_{1065}, I_{1588}/I_{1065})		I_{1019}/I_{1685}	
		Normal	Cancer	Normal	Cancer	Normal	Cancer
Histopathologic	Normal ($n = 21$)	17	4	21	0	13	8
diagnosis	Cancer ($n = 27$)	5	22	10	17	2	25
	Sensitivity (%)	81		63		93	
	Specificity (%)	81		100		62	

many cases. It was also evident that whereas overall accuracies up to 81% were found, in general the sensitivity will increase at the expense of a poorer sensibility and vice versa. This happens because the dispersion of values is considerable throughout the different combinations that were studied. It would be desirable that this feature could be studied in more detail, through performing extensive studies on a great number of randomly sampled spectra from each tissue type as well as optimizing acquisition conditions, in order to obtain an improved understanding of the nature of this dispersion and the limitations of the current approach.

Previous studies on curve-fitting analysis for the characterization of Raman spectra from normal versus cancerous lung[37] and breast[35] tissues reported higher differences in the intensity distributions of curve-resolved bands, but the differences between the average spectra were also significantly higher in comparison to our results. From the intensity distributions reported by Chowdary et al.[35] one can see that the differences between malignant and normal breast tissues are higher than if we compare malignant to benign tissue, due to a much similar tissue structure between the two in the latter case, which is somehow closer to our situation.

It should be noted from our results that the parameters that yielded enhanced accuracies in differentiating adenocarcinoma from normal mucosa were all derived from combinations of the intensities corresponding to those bands that were pointed out in Sec. 3.2 (see Table 1), which reinforces the fact that these bands represent the most relevant features of the Raman spectra from normal and cancerous colonic tissues, from the standpoint of tissue discrimination. This type of study may be useful in order to target a specific region of the Raman spectra with higher diagnostic relevance, instead of the whole 800 to 1800 cm^{-1} as it is current practice, because the use of a smaller spectral window can be profitable for an *in vivo* application, since it may allow to obtain better signal-to-noise ratios for short acquisition times.

4 Conclusions

A first attempt toward the application of deconvolution of Raman spectra for the purpose of discriminating cancerous colonic tissue from normal mucosa was described. It was found that mean spectral profiles from normal and cancerous tissue are similar in shape, whereas small intensity differences can be observed ($< \sim 20\%$), where normal tissue spectra reveals higher Raman intensity than cancerous tissue spectra up to 1200 cm^{-1}, and vice versa above that value. However, there is considerable variability in intensity values of Raman spectra within each group, which is in accordance with previous studies, and is also evident from the band intensity distributions obtained by curve-fitting analysis. Nevertheless, even with a small set of measured spectra, there is a clear indication from our results that only some of the component bands play an important role for tissue discrimination, those are placed at: 1019, 1065, 1323, 1345, 1449, 1467, 1588, 1662, and 1685 cm^{-1}. Several combinations of intensities or intensity ratios concerning these bands yielded parameters that classified the two tissue types with overall accuracies of 79%, when linear discriminant analysis was applied to the corresponding intensity distributions, and the maximum accuracy of 81% was obtained using the pair (I_{1323}, I_{1467}). The considerable dispersion in intensity distributions may pose the higher limit of achievable accuracy, and thus its origin and nature should be further investigated.

Two distinct formalin bands were identified at 1046 and 1495 cm^{-1}. Our results suggest that the latter band can be accounted for when curve deconvolution is applied, thus not hampering Raman studies performed on formalin-fixed colonic tissues, comparatively to fresh tissues. As in the case of the lower frequency band, some caution has to be taken and the complete influence of this band cannot be accounted for, thus it would be desirable to perform similar studies in the lower frequency region using fresh tissue samples.

References

1. Instituto Português de Oncologia do Porto, Registo Oncológico 2008 (2009).
2. American Cancer Society, Cancer Statistics (2009).
3. American Cancer Society, Colorectal Cancer Facts & Figures (2008-2010).
4. J. C. Taylor, C. A. Kendall, N. Stone, and T. A. Cook, "Optical adjuncts for enhanced colonoscopic diagnosis," *Br. J. Surg.* **94**(1), 6–16 (2007).
5. C. L. Kay, D. Kulling, R. H. Hawes, J. W. Young, and P. B. Cotton, "Virtual endoscopy-comparison with colonoscopy in the detection of space-occupying lesions of the colon."*Endoscopy* **32**(3), 226–232 (2000).

6. R. S. Kwon, D. V. Sahani, and W. R. Brugge, "Gastrointestinal cancer imaging: deeper than the eye can see," *Gastroenterology* **128**(6), 1538–1553 (2005).
7. T. Kawabata, T. Mizuno, S. Okazaki, M. Hiramatsu, T. Setoguchi, H. Kikuchi, M. Yamamoto, Y. Hiramatsu, K. Kondo, and M. Baba, "Optical diagnosis of gastric cancer using near-infrared multichannel Raman spectroscopy with a 1064-nm excitation wavelength," *J. Gastroenterol.* **43**(4), 283–90 (2008).
8. H. Barr, T. Dix, and N. Stone, "Optical spectroscopy for the early diagnosis of gastrointestinal malignancy," *Lasers Med. Sci.* **13**(1), 3–13 (1998).
9. R. S. Dacosta, B. C. Wilson, and N. E. Marcon, "New optical technologies for earlier endoscopic diagnosis of premalignant gastrointestinal lesions," *J. Gastroenterol. Hepatol* **17**, S85–S104 (2002).
10. L. M. W. K. Song and B. C. Wilson, "Endoscopic detection of early upper GI cancers," *Best Practice & Research Clinical Gastroenterology* **19**(6), 833–856 (2005).
11. R. S. DaCosta, B. C. Wilson, and N. E. Marcon, "Optical techniques for the endoscopic detection of dysplastic colonic lesions," *Curr. Opin. Gastroenterol.* **21**, 70–79 (2004).
12. A. Dhar, K. S. Johnson, M. R. Novelli, S. G. Bown, I. J. Bigio, L. B. Lovat, and S. L. Bloom, "Elastic scattering spectroscopy for the diagnosis of colonic lesions: initial results of a novel optical biopsy technique," *Gastrointest. Endosc.* **63**(2), 257–261(2006).
13. G. Reese, C. Reid, R. Goldin, M. A. Tran-Dang, A. Fitzgerald, P. Tekkis, and V. P. Wallace, "Using terahertz pulsed imaging (TPI) to identify colonic pathology," in *IEEE 33rd International Conference on Infrared, Millimeter and Terahertz Waves, IRMMW-THz 2008* (2008).
14. A. Molckovsky, L. M. W. K. Song, M. G. Shim, N. E. Marcon, and B. C. Wilson, "Diagnostic potential of near-infrared Raman spectroscopy in the colon: Differentiating adenomatous from hyperplastic polyps," *Gastrointest. Endosc.* **57**(3), 396–402 (2003).
15. B. Schrader, B. Dippel, I. Erb, S. Keller, T. Lochte, H. Schulz, E. Tatsch, S. Wessel, "NIR Raman spectroscopy in medicine and biology: results and aspects," *J. Mol. Struct.* **480**, 21–32 (1999).
16. E. B. Hanlon, R. Manoharan, T. W. Koo, K. E. Shafer, J. T. Motz, M. Fitzmaurice, J. R. Kramer, I. Itzkan, R. R. Dasari, and M. S. Feld, "Prospects for *in vivo* Raman spectroscopy." *Phys. Med. Biol.* **45**, R1–R59 (2000).
17. W. Petrich, "Mid-infrared and Raman spectroscopy for medical diagnostics," *Appl. Spectrosc. Rev.* **36**(2), 181–237 (2001).
18. M. G. Shim, L. M. W. K. Song, N. E. Marcon, and B. C. Wilson, "*In vivo* near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy," *Photochem. Photobiol.* **72**(1), 146–50 (2000).
19. M. G. Shim and B. C. Wilson, "Development of an *in vivo* Raman spectroscopic system for diagnostic applications," *J. Raman Spectrosc.* **28**(2–3), 131–42 (1997).
20. N. Stone, C. Kendall, J. Smith, P. Crow, and H. Barr; "Raman spectroscopy for identification of epithelial cancers," *Faraday Discuss.* **126**, 141–157 (2004).
21. Z. Huang, W. Zheng, E. Widjaja, J. Mo, and C. Sheppard, "Classification of colonic tissues using Raman spectroscopy and multivariate techniques," *Proc. SPIE* **6093**, 60930Q (2006).
22. E. Widjaja, W. Zheng, and Z. Huang, "Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines," *Int. J. Oncol.* **32**(3), 653–662 (2008).
23. M. V. P. Chowdary, K. K. Kumar, K. Thakur, A. Anand, J. Kurien, C. M. Krishna, and S. Mathew, "Discrimination of normal and malignant mucosal tissues of the colon by Raman spectroscopy," *Photomed. Laser Surg.* **25**(4), 269–274 (2007).
24. N. Stone, C. Kendall, N. Shepherd, P. Crow, and H. Barr, "Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers," *J. Raman Spectrosc.* **33**, 564–573 (2002).
25. K. E. Shafer-Peltier, A. S. Haka, M. Fitzmaurice, J. Crowe, J. Myles, R. R. Dasari, and M. S. Feld, "Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis *in vivo*," *J. Raman Spectrosc.* **33**(7), 552–563 (2002).
26. W. Hayes and R. Loudon, *Scattering of Light by Crystals*, John Wiley & Sons, New York (1978).
27. A. Mahadevan-Jansen and R. R. Richards-Kortum, "Raman spectroscopy for the detection of cancers and precancers," *J. Biomed. Opt.* **1**(1), 31–70 (1996).
28. C. Krafft, D. Codrich, G. Pelizzo, and V. Sergo, "Raman and FTIR microscopic imaging of colon tissue: a comparative study," *J. Biophotonics* **1**(2), 154–169 (2008).
29. D. Codrich, "Applications and limits of Raman spectroscopy in the study of colonic and pulmonary malformations," PhD thesis, Universita' Degli Studi di Trieste (2007).
30. P. O. Andrade, R. A. Bitar, K. Yassoyama, H. Martinho, A. M. E. Santo, P. M. Bruno, and A. A. Martin, "Study of normal colorectal tissue by FT-Raman spectroscopy," *Anal. Bioanal. Chem.* **387**(5), 1643–1648 (2007).
31. N. Lebrun, P. Dhamelincourt, C. Focsa, B. Chazallon, J. L. Destombes, and D. Prevost, "Raman analysis of formaldehyde aqueous solutions as a function of concentration," *J. Raman Spectrosc.* **34**(6), 459–464 (2003).
32. Z. Huang, A. McWilliams, S. Lam, J. English, D. I. McLean, H. Lui, and H. Zeng, "Effect of formalin fixation on the near-infrared Raman spectroscopy of normal and cancerous human bronchial tissues," *Int. J. Oncol.* **23**(3), 649–656 (2003).
33. M. G. N. Shim, "Analysis of biological tissue with *ex vivo* and *in vivo* Raman spectroscopy," Master's thesis, University of Toronto (1996).
34. Z. Movasaghi, S. Rehman, and I. U. Rehman, "Raman spectroscopy of biological tissues," *Appl. Spectrosc. Rev.* **42**(5), 493–541 (2007).
35. M. V. P. Chowdary, K. K. Kumar, S. Mathew, L. Rao, C. M. Krishna, and J. Kurien, "Biochemical correlation of Raman spectra of normal, benign and malignant breast tissues: A spectral deconvolution study," *Biopolymers* **91**(7), 539–546 (2009).
36. M. D. Keller, E. M. Kanter, and A. Mahadevan-Jansen, "Raman spectroscopy for cancer diagnosis," *Spectroscopy* **21**(11), 33–41 (2006).
37. S. Kaminaka, T. Ito, H. Yamazaki, E. Kohda, and H. Hamaguchi, "Near-infrared multichannel Raman spectroscopy toward real-time *in vivo* cancer diagnosis," *J. Raman Spectrosc.* **33**(7), 498–502 (2002).