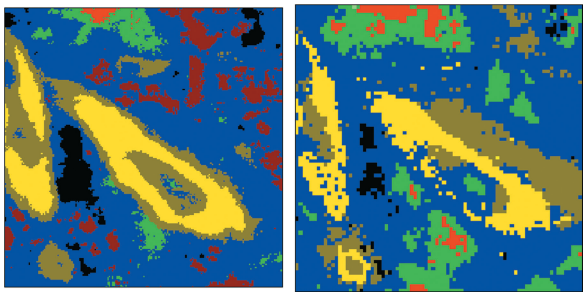
The Analyst b712958k

Raman mapping and FTIR imaging of lung tissue: congenital cystic adenomatoid malformation

Christoph Krafft,* Daniela Codrich, Gloria Pelizzo and Valter Sergo

Raman and FTIR images are collected from lung tissue sections. Segmentation by cluster analysis and mean spectra give detailed information about biochemical changes in malformations.



Please check this proof carefully. **Our staff will not read it in detail after you have returned it.**

Translation errors between word-processor files and typesetting systems can occur so the whole proof needs to be read. Please pay particular attention to: tabulated material; equations; numerical data; figures and graphics; and references. If you have not already indicated the corresponding author(s) please mark their name(s) with an asterisk. Please e-mail a list of corrections or the PDF with electronic notes attached – do not change the text within the PDF file or send a revised manuscript.

Please bear in mind that minor layout improvements, e.g. in line breaking, table widths and graphic placement, are routinely applied to the final version.

We will publish articles on the web as soon as possible after receiving your corrections; no late corrections will be made.

Please return your **final** corrections, where possible within **48 hours** of receipt by e-mail to: proofs@rsc.org

Reprints—Electronic (PDF) reprints will be provided free of charge to the corresponding author. Enquiries about purchasing paper reprints should be addressed via: http://www.rsc.org/Publishing/ReSourCe/PaperReprints/. Costs for reprints are below:

Reprint costs

No of pages	Cost for 50 copies	Cost for each additional 50 copies
2–4	£180	£115
5–8	£300	£230
9–20	£600	£480
21–40	£1100	£870
>40	£1700	£1455

Cost for including cover of journal issue:
£50 per 50 copies

Q1

Raman mapping and FTIR imaging of lung tissue: congenital cystic adenomatoid malformation

Christoph Krafft,*†[a] Daniela Codrich,[b] Gloria Pelizzo[b] and Valter Sergo[a]

Received 22nd August 2007, Accepted 17th December 2007
First published as an Advance Article on the web ??????
DOI: 10.1039/b712958k

Congenital cystic adenomatoid malformations (CCAMs) are benign masses of non-functional lung tissue developing from an overgrowth of the terminal bronchioles with subsequent suppressing of alveolar growth. For the first time CCAMs are studied by Raman mapping and Fourier transform infrared (FTIR) imaging. Both vibrational spectroscopic methods are able to analyze the biochemical composition of tissues and their pathological changes at the molecular level. Cryosections were prepared on calcium fluoride substrates from CCAMs and from normal lung tissue of two infant patients who underwent surgery. Raman maps were collected at a step size of 100 μm in order to assess the whole tissue section and at a smaller step size of 10 μm in order to resolve details in selected areas. FTIR images were collected in the macroscopic and microscopic modes. Data sets were segmented by cluster analysis and the mean spectra of each cluster were compared. At low lateral resolution a lower red blood cell content and higher lipid content were found in CCAMs than in normal lung tissue. At higher lateral resolution, accumulations of lipids and glycogen were identified in CCAMs. The lipid aggregates contain a high concentration of phosphatidylcholine. It is discussed how the combined application of Raman mapping and FTIR imaging might improve the differential diagnosis of lung malformations and how both modalities might be applied to other bioanalytical and biomedical problems in the future.

Introduction

Congenital cystic adenomatoid malformations (CCAMs) are developmental abnormalities of the lung that typically appear in infancy and childhood and are one of the most frequent causes of neonatal respiratory distress. Cystic structures arise from an overgrowth of the terminal bronchioles with subsequent suppression of alveolar growth. CCAMs are usually unilateral and involve only one lobe of the lung. They are detected in chest images as abnormal air, air/fluid-filled cysts, or fluid-filled/solid-appearing cysts, and large masses may cause mediastinal shift. Around 80–85% of all cases are diagnosed in the first two years of life,[1] but CCAM has also been described in older patients as well.[2] Prenatal identification of lung abnormalities has increased with routine prenatal ultrasonography. Genetics, prenatal evaluation and treatment have recently been reviewed.[3] Although five pathological types of CCAM have been described,[4] the more useful prenatal characterization is to divide them by gross anatomy into cystic types, which contain one or more large cysts (macrocystic), and solid types which contain both solid regions and cysts (microcystic).[5] In most cases, a fetus with CCAM is closely monitored during pregnancy due to the unpredictable growth pattern of CCAM, and the abnormal mass of lung tissue is excised *via* surgery after birth. In cases complicated by hydrops, prenatal intervention has been described as an experimental option to correct hemodynamic complications of CCAM *in utero*.[6] The fetal therapy is still burdened by a high mortality rate and technical issues. As far as the prognosis is concerned, it is important to diagnose the exact type of malformation in order to exclude associated anomalies as well as the development of malignancies in later life. CCAM may act as a predisposing condition for the development of bronchioloalveolar carcinoma and mesenchymal tumors.[7] In order to avoid risks of infections, pulmonary compression and malignant transformation, segmentectomy or lobectomy is the treatment of choice to remove the abnormal mass of lung tissue within the first year of age,[8] even in asymptomatic patients.[9,10] Early resection also maximizes compensatory lung growth.

Raman and infrared (IR) spectroscopy are optical techniques that allow probing molecular vibrations of samples without external labels or extensive preparation. Raman spectroscopy excites vibrations by inelastic scattering of monochromatic light. IR spectroscopy excites vibrations by absorption of IR radiation. Different functional groups in molecules have different vibrational energies, and these can then be used to identify them. Practically, both IR and Raman spectra usually appear as graphs reporting intensities *versus* wavenumber values that are directly proportional to the energy of the vibrations. Hence, the intensity/wavenumber couples provide information about the quantity and type of a functional group. Without going into detail about physical selection rules, both techniques

[a]*CENMAT, Department of Materials and Natural Resources, University Trieste, Via Valerio 2, 34127 Trieste, Italy.*
E-mail: christoph.krafft@tu-dresden.de
[b]*Department of Pediatric Surgery, IRCCS Burlo Garofolo, 34137 Trieste, Italy*
† Present address: Bioanalytical Chemistry, Dresden University of Technology, 01062 Dresden, Germany.

probe vibrational properties, but they are not redundant and to a certain extent they are complementary. Combining the vibrational spectroscopic data with lateral resolution gives powerful imaging techniques that have already been applied to numerous biomedical problems.[11] The fundamental principle of both methods in this field is that the spectra represent the biochemical composition and structure of the samples. As the transition between normal and pathological cells and tissues changes their composition and structure, these methods offer prospects to complement the standard diagnostic tools such as histopathology or immunohistochemistry.

The pathogenesis of CCAM is not well defined. The cellular components have been evaluated by immunohistochemistry.[12] The biochemical composition of CCAM has not been studied in detail so far. Therefore, the scope of the current study is to investigate the changes in the biochemical composition between normal lung tissue and CCAM. Raman and FTIR imaging are applied as innovative experimental tools. The methods and results of an accompanying paper about normal lung tissue[13] are transferred to collect and analyze macroscopic images with lateral resolutions of 63 and 100 μm. Furthermore, microscopic images with lateral resolutions near 10 μm are presented in this study in order to identify more details. It will be discussed how the combined application of Raman and FTIR imaging might improve the differential diagnosis of lung malformations in the future.

Experimental

Tissue specimens

All experimental procedures were approved by the local ethical committee. Lung tissue specimens were obtained from two infant patients undergoing lobectomy because of respiratory distress due to extended solid-type CCAM. The excised lobes contained also a fraction of normal tissue which served as controls. Small tissue blocks of *ca.* 1 cm^3 volume were snap-frozen in liquid nitrogen without fixing or embedding medium. Tissue sections of 20 μm thickness were cut using a cryotome, transferred on slides made of calcium fluoride (VUV grade) and passively dried without further treatment. Photographs of unstained sections were taken using a Leica microscope with a low (1.6 ×/NA 0.05) or a higher (10 ×/NA 0.3) magnification objective and a digital camera accessory.

FTIR imaging

FTIR images were recorded using an FTIR spectrometer Bruker IFS66/S equipped with a mercury cadmium telluride (MCT) focal plane array (FPA) detector of 64 × 64 pixels (Bruker Optik, Germany). Coupling to the macro chamber IMAC™ (Bruker) gave a field of view of 4 × 4 mm and each pixel corresponded to an area of 63 × 63 μm. Using the microscope Hyperion with 15× magnification, the field of view was reduced to 270 × 270 μm and each pixel corresponded to an area of 4.2 × 4.2 μm. The actual lateral resolution is limited by diffraction. According to Abbe's Law it can be calculated for the most intense protein band at 1650 cm^{-1} using a Cassegrain objective with a numerical aperture of 0.4 as 9 μm. Images of 4096 IR spectra in the wavenumber range 950–3900 cm^{-1} at 4 cm^{-1} spectral resolution were acquired by co-adding 11 interferograms. Ratios of the single beam spectra of each sample FTIR image against the spectra of a background FTIR image of neat calcium fluoride slides were obtained and converted to absorbance. Acquisition of each image took approximately 3 min.

Raman mapping

Raman images were recorded at 785 nm laser excitation in the mapping mode, *i.e.* Raman spectra were sequentially collected by moving the sample with a motorized stage at a step size of 100 or 10 μm. The Raman system inVia (Renishaw, UK) was coupled to a microscope Leica DMLM. A thermoelectrically cooled charge coupled device (CCD) camera was used for detection. The laser radiation of 25 mW intensity was focused onto the sample with a 50 ×/NA 0.75 objective. The size of the rectangular laser spot was approximately 2 × 7 μm. Spectra were obtained in the 600–1800 cm^{-1} region using the synchro mode of the instrument software WiRE™ 2.0 (Renishaw). In the synchro mode, the grating with 1200 lines per mm is continuously moved to obtain Raman spectra of extended spectral regions. Before data collection, the instrument was calibrated using the well-documented Raman bands of 4-acetamidophenol.[14] Approximately 100 spectra were collected per hour.

Data processing

Details of the procedures, that were performed using the software CytoSpec[15] and in-house developed functions under a Matlab platform (The Mathworks Inc., USA), have been published in an accompanying paper.[13] Briefly, data pre-processing separates spectral variances due to physical effects from the chemical information. Elastic light scattering, which changes the baseline, and sample thickness, which determines the overall signal intensity, belong to the most significant physical effects. Spectral variances due to scattering are reduced by subtracting a multi-point linear baseline from each spectrum. The baseline segments were obtained by linear interpolation between the selected points. Spectral variances due to the thickness are reduced by multiplicative signal correction.[16] Further pre-processing of Raman spectra included subtraction of background originating from the optical elements in the laser light delivery pathway and correction of spikes due to cosmic rays. Subsequently, spectra with intensities below a threshold were removed from the FTIR and Raman data sets. These spectra corresponded to sample regions without tissue. The pre-processed data sets were subjected to a *k*-means cluster analysis using a Euclidian distance metric. This algorithm segments the data sets into a predefined number of groups (also referred to as clusters) according to their spectral similarity. The number of clusters was chosen as the optimum number to ensure that every tissue type and disease state is recognized. The averaged spectrum of each cluster was calculated and imported to the program GRAMS (Thermo Fisher, USA) for analysis.

Results

Raman maps and FTIR images of lung tissue and CCAM

Vibrational spectroscopic images can be constructed in a univariate mode where chemical maps (also called functional

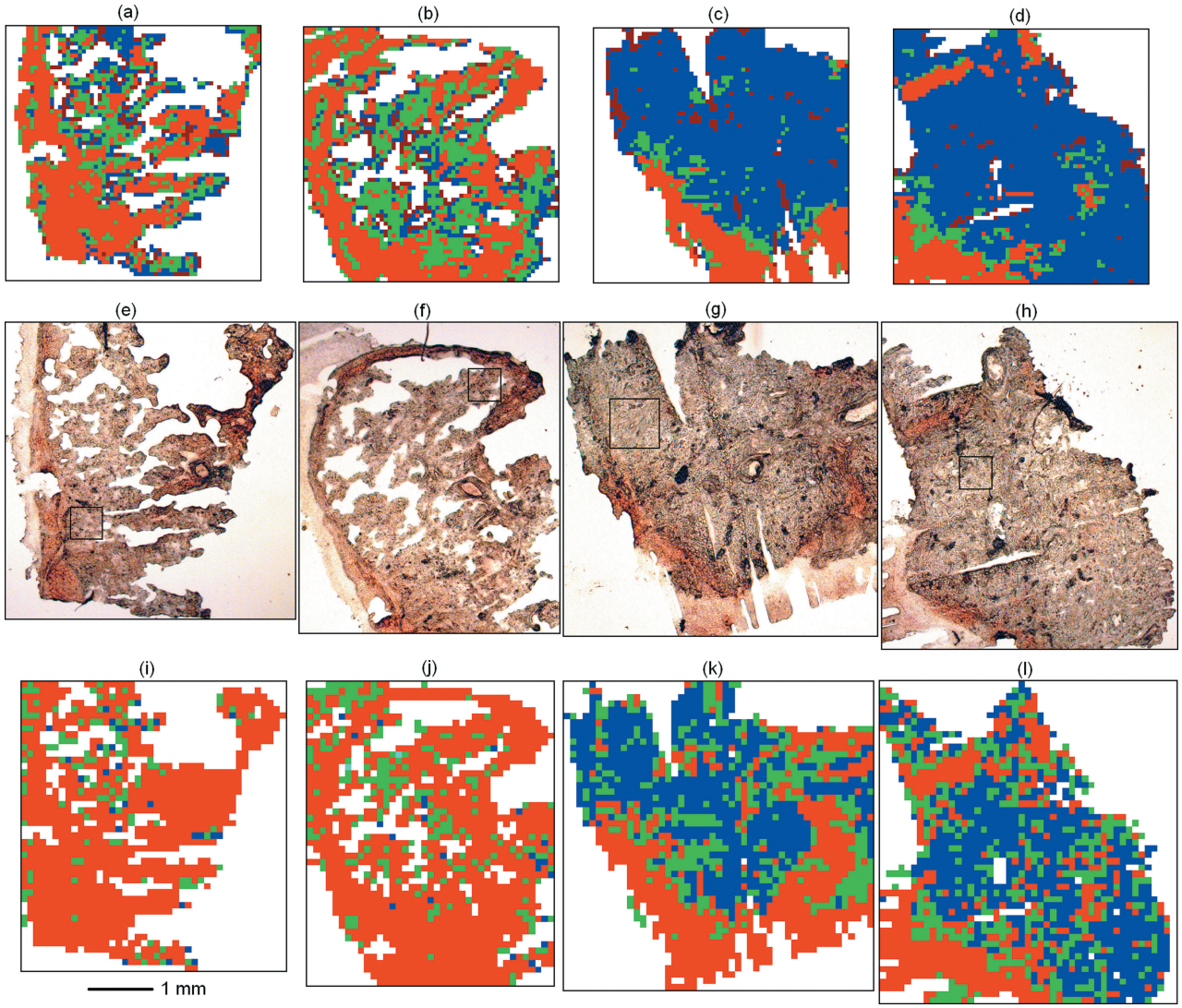


Fig. 1 FTIR images (a–d), photographs (e–h) and Raman maps (i–l) of lung tissue sections. Data of the first (a,e,i) and second (b,f,j) columns were obtained from normal samples of patients 1 and 2, respectively. Data of the third (c,g,k) and forth (d,h,l) columns were obtained from CCAM of patients 1 and 2, respectively. The color code represents the segmentation into four (a–d) and into three (i–l) groups based on cluster analyses of spectra. Boxes in (e–h) indicate the positions of the microscopic images shown later in Fig. 4.

group maps) based on peak intensities, peak areas or peak ratios are generated. While these methods provide information on the distribution and relative concentration of a particular functional group and hence a specific major biomolecule, they are not very useful in terms of classifying anatomical and histopathological features within tissue sections with strongly overlapping spectral contribution from various biomolecules. The spectral information, which is usually distributed over a broad spectral region, can be utilized more efficiently using a multivariate image reconstruction. The FTIR images and Raman maps in this study were processed using unsupervised *k*-means cluster analyses. Fig. 1 compares FTIR images (a–d) and Raman maps (i–l) with photographs (e–h) of lung tissue sections. Data of the first (a,e,i) and third (c,g,k) columns were obtained from normal samples and CCAM of patient 1, respectively. Data of the second (b,f,j) and fourth (d,h,l) columns were obtained from normal samples and CCAM of patient 2, respectively. Fig. 1(e) and 1(f) show typical sponge-like morphologies of normal lung tissue. A dried film of liquid is visible at the left margins. Fig. 1(g) and 1(h) show morphological changes in solid-type CCAM. The tissue sections have fewer holes which is consistent with non-functional, non-aerating, abnormal masses of lung tissue.

A cluster analysis of the combined FTIR images was performed in the interval 950–1480 cm^{-1}. The color code in Fig. 1(a–d) represents the segmentation into four classes or clusters. Regions appear white where spectra were removed before at positions without tissue. The percentages of the cluster assignments are given in Table 1**(a)**. In Fig. 1(a) and 1(b) most spectra were assigned to the red cluster (49 and 46%, respectively), followed by the green cluster (24 and 34%), the blue cluster (19 and 12%) and the brown cluster (7 and 7%). Spectra of the brown cluster are mainly located at the margins. Fig. 1(c) and 1(d) are dominated by the blue cluster (70 and 74%, respectively). Here, spectra of the red cluster (17 and 15%) are located mainly near the dried liquid film. Spectra of the green

Table 1 Cluster assignments (in %) of FTIR images (IR) and Raman maps (RA) in **(a)** Fig. 1 and in **(b)** Fig. 4. For brevity, the clusters of all spectra besides the red, green and blue clusters are pooled in the cluster called 'rest'. Dashes indicate that the Raman maps in Fig. 1 were segmented in just three clusters. The values 0 indicate that less than 0.05% of the spectra were assigned to this cluster

		Red		Green		Blue		Rest	
Color		IR	RA	IR	RA	IR	RA	IR	RA
(a)	Fig. 1(a/i)	49	85	24	12	19	3	7	—
	Fig. 1(b/j)	46	81	34	17	12	2	7	—
	Fig. 1(c/k)	17	38	8	24	70	38	5	—
	Fig. 1(d/l)	15	31	8	25	74	43	3	—
(b)	Fig. 4(a/i)	74	54	25	40	1	6	0	0
	Fig. 4(b/j)	75	65	24	34	1	1	0	0
	Fig. 4(c/k)	0.5	2.6	4	13.4	55	56	40.5	28
	Fig. 4(d/l)	0.4	11.3	6	16	76	71	17.6	1.7

cluster (8 and 8%) are found at the transition between the red and the blue clusters. The percentages of the spectra in the brown cluster are reduced to 5 and 3%.

The Raman maps in Fig. 1 consist of 42 × 46 (i), 39 × 48 (j), 49 × 49 (k) and 47 × 51 (g) spectra using a step size of 100 μm. The dimension of each Raman map is slightly larger than the 4 × 4 mm FTIR image of the same tissue section. However, the number of spectra was lower in the Raman maps than in the FTIR images where each data set was composed of 64 × 64 = 4096 spectra and each spectrum corresponded to an area of 63 × 63 μm. After pooling the four Raman maps, a cluster analysis was performed in the spectral interval 1190–1750 cm^{-1}. As a cluster corresponding to margins (brown in FTIR images) was not identified in the Raman maps, the number of clusters was reduced to three in order to compare the segmentation with the FTIR result. The colors in the Raman maps represent similar clusters as in the FTIR images. In Fig. 1(i) and 1(j) most spectra were assigned to the red cluster (85 and 81%, respectively), followed by the green cluster (12 and 17%) and the blue cluster (3 and 2%). In Fig. 1(k) and l(l) spectra of the red cluster (38 and 31%) are mainly located near the outer regions, and spectra of the blue cluster (38 and 43%) and of the green cluster (24 and 25%) are found in the core regions. The percentages are included in Table 1**(a)**.

IR spectra of CCAM

The mean spectra of each cluster in Fig. 2 were extracted from the FTIR image in Fig. 1(c). Each mean spectrum represents all spectra in a cluster and is used to interpret the biochemical differences between clusters. As the IR spectra of the other FTIR images were grouped to the same clusters, the results are almost identical. Analysis of the spectra allows obtaining information about the underlying spectral changes in each cluster. The IR spectra of the red cluster (trace 1) and the green cluster (trace 2) are overlaid in Fig. 2(a) after normalization to equal intensities at 1656 cm^{-1}. At first sight both spectra are similar. Spectral contributions of proteins dominate which are assigned to vibrations of the peptide backbone (amide III at 1304, amide II at 1542, amide I at 1656, and amide A at 3300 cm^{-1}) and of amino acid residues (1391, 1454, 2872, and 2959 cm^{-1}). Further spectral contributions at 1085 and 1242 cm^{-1} are assigned to lipids, nucleic acids and carbohydrates. Bands assignments according to the literature[17,18] are summarized in Table 2**(a)**.

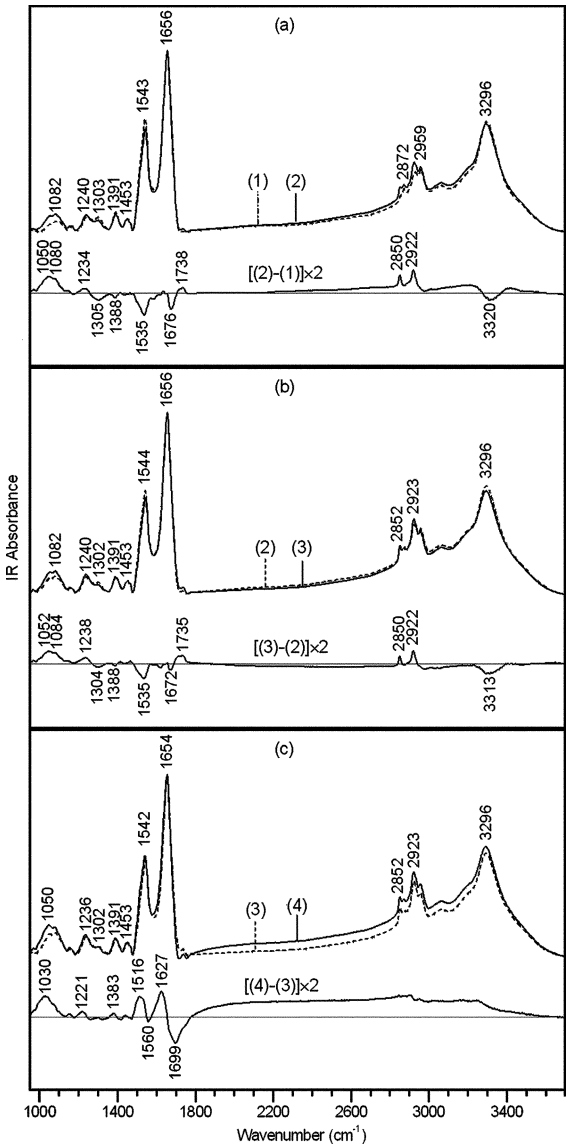


Fig. 2 IR spectra from 950 to 3700 cm^{-1} averaged from the red (trace 1), green (trace 2), blue (trace 3) and brown (trace 4) clusters of the FTIR image in Fig. 1(c). Difference spectra were calculated for comparison. Spectra were baseline corrected and normalized for display.

Table 2 Assignments of **(a)** IR bands and **(b)** Raman bands in spectra of normal lung tissue and CCAM compiled from previous results: IR,[17–19,24] RA.[20–23] The main contributors were tabulated

IR		RA	
Band/cm^{-1}	Assignment[a]	Band/cm^{-1}	Assignment[a]
(a)		**(b)**	
969	PC	700	Cholesterol
1026	Glycogen	717	PC [$\nu_s(N^+(CH_3)_3)$]
1030	Mucus	747	Hemoglobin
1060	Glycolipids [ν(COC)], cholesterol	852	Protein (Tyr), collagen
1080	Phospholipids [$\nu_s(PO_2^-)$], glycogen	874	PC [$\nu_{as}(N^+(CH_3)_3)$], collagen
1153	Glycogen	941	Protein [ν(C–C)], collagen
1234	Phospholipids [$\nu_{as}(PO_2^-)$], collagen	1002	Protein (Phe), hemoglobin
1303	Protein (amide III)	1063	Lipid
1391	Protein [$\nu(COO^-)$], cholesterol	1125	Lipid, hemoglobin
1453	Protein [$\delta(CH_3)$], collagen	1252	Protein (amide III), collagen, hemoglobin
1467	Lipid [$\delta(CH_2)$]	1270	Lipid [δ(=CH)]
1543	Protein (amide II)	1296	Lipid [$\delta(CH_2)$]
1656	Protein (amide I)	1370	Hemoglobin
1738	Lipid [ν(C=O)]	1438	Lipid [$\delta(CH_2)$], cholesterol
2850	Lipid [$\nu_s(CH_2)$]	1447	Protein [$\delta(CH_3)$]
2872	Protein [$\nu_s(CH_3)$]	1560	Hemoglobin
2922	Lipid [$\nu_{as}(CH_2)$]	1577	Hemoglobin
2959	Protein [$\nu_{as}(CH_3)$]	1620	Hemoglobin
3296	Protein (amide A)	1656	Protein (amide I), collagen, lipid [ν(C=C)]
3385	Glycogen	1738	Lipid [ν(C=O)]

[a] Abbreviations: phosphatidylcholine (PC), stretch vibration (ν), deformation vibration (δ), symmetric (s), antisymmetric (as), tyrosine (Tyr), phenylalanine (Phe).

As biomolecules are mainly composed of similar functional groups with C, H, O, N and P atoms and most IR absorption bands are broad, they strongly overlap and it is difficult to resolve each individual band. However, instead of determining the composition or structure of tissues by IR spectra, it is usually sufficient in most applications to analyze the spectral changes. In order to assess the small differences between traces (1) and (2), a difference spectrum was calculated and included in Fig. 2(a) at a two-times amplified scale. All positive difference bands at 1050 (vibration of C–O), 1080 and 1234 (symmetric and antisymmetric stretch vibrations of PO_2), 1738 (stretch vibration of C=O), 2850 and 2922 cm^{-1} (symmetric and antisymmetric stretch vibrations of CH_2) are assigned to lipids, in particular phospholipids. All negative difference bands at 1305, 1388, 1535, 1676 and 3320 cm^{-1} are assigned to proteins. In particular, the bands at 1305 and 1535 were reported to correlate with hemoglobin.[13,18,19]

The IR spectra of the green (trace 2) and the blue clusters (trace 3) are overlaid in Fig. 2(b). The band positions largely agree with Fig. 2(a). The difference spectrum between traces (3) and (2) contains the same bands as in Fig. 2(a), however, with slightly different relative intensities. This observation of qualitative similar difference spectra indicates that the lipid content successively increases in the order trace (1) < (2) < (3), whereas the hemoglobin content successively decreases in the order trace (1) > (2) > (3). As the red and green clusters that are represented by traces (1) and (2) are more abundant in FTIR images of normal lung tissue and the blue cluster which is represented by trace (3) is more abundant in FTIR images of CCAM, the total lipid content is higher in CCAM whereas the total hemoglobin content is higher in normal lung tissue.

The IR spectra of the blue (trace 3) and the brown (trace 4) clusters are overlaid in Fig. 2(c). The increase of the background and the shift of the amide II and amide I bands toward lower wavenumber values point to dispersion artifacts that have been described in detail for an FTIR image of a normal lung tissue section.[13] The features at 1516/1560 and at 1627/1699 in the difference spectrum between traces (4) and (3) are a consequence of these physical effects. The intense dispersion artifacts in the wavenumber interval 1480–1800 cm^{-1} were the main reason why the cluster analysis was restricted to the interval 950–1480 cm^{-1}. The difference band near 1030 cm^{-1} is assigned to the carbohydrate moiety of mucus.[20] The absence of differences near lipid and hemoglobin bands indicates that these constituents did not change between traces (3) and (4). Whereas the spectra of the brown cluster are mainly found near the margins in FTIR images of normal tissue [Fig. 1(a,b)], they are scattered throughout the sample in FTIR images of CCAM [Fig. 1(c,d)].

Raman spectra of CCAM

The Raman spectra in Fig. 3 represent the average of the red (trace 1), green (trace 2) and blue (trace 3) clusters in Fig. 1(i–l). Spectral variations are observed in particular in the interval 1500–1700 cm^{-1}. Spectral contributions in this interval are assigned to hemoglobin which is a major constituent of red blood cells.[13,18,21] Upon excitation of Raman spectra with 785 nm radiation, bands of red blood cells were enhanced and partly masked the weaker bands of other proteins, lipids, carbohydrates and nucleic acids. Therefore, a procedure was suggested to separate the spectral contributions of red blood cells from the remaining components.[13] Fig. 3(d) shows the resulting spectrum

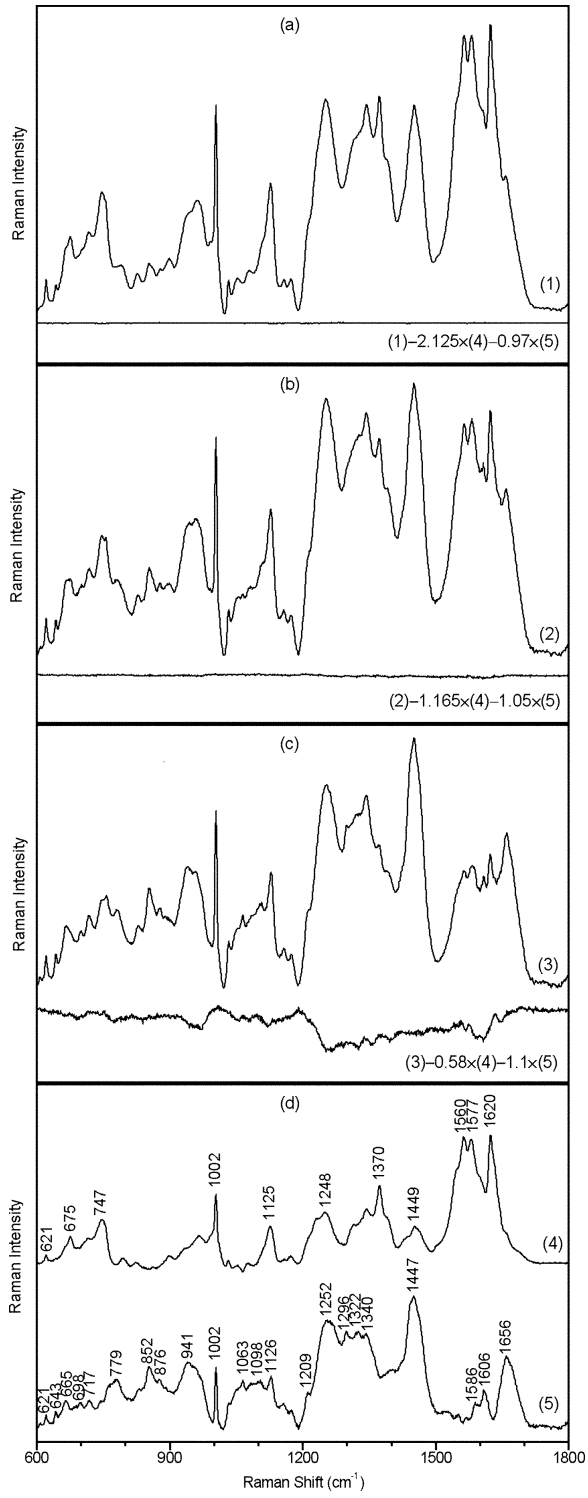


Fig. 3 Raman spectra from 600 to 1800 cm^{-1} averaged from the red (trace 1), green (trace 2) and blue (trace 3) clusters of the Raman maps in Fig. 1(i–l). Spectral contributions of red blood cells (trace 4) and of lung tissue without red blood cells (trace 5) were calculated from the red and green clusters. Difference spectra between Raman spectra and linear combinations of (4) and (5) were included.

of red blood cells (trace 4) and the spectrum of the remaining lung tissue components (trace 5). Spectrum (4) was calculated from the differences between the spectra of the red and the green clusters in Raman maps of Fig. 1(i–l). Spectrum (5) was obtained by subtracting spectrum (4) from the spectra of the red and the green clusters in Raman maps of Fig. 1(i–l). All bands in spectrum (4) are assigned to hemoglobin in an aggregated, denatured state.[18] Bands in spectrum (5) are assigned to proteins (621, 643, 852, 941, 1002, 1209, 1252, 1322, 1340, 1447, 1586, 1606 and 1656 cm^{-1}), lipids (698, 717, 1063, 1126 and 1296 cm^{-1}) and nucleic acids (665, 779 and 1098 cm^{-1}). Band assignments according to refs 22–24 are summarized in Table 2**(b)**.

The absence of significant bands in difference spectra of Fig. 4(a) and 4(b) indicates that linear combinations of spectra (4) and (5) closely approximate spectra (1) and (2). Whereas the coefficients for spectrum (4) decrease from 2.125 in spectrum (1) to 1.165 in spectrum (2), the fit coefficients for spectrum (5) are close to 1. This result confirms that the main variations are differences in the red blood cell content. The fit coefficients in Fig. 4(c) decrease to 0.58 for spectrum (4) and increase to 1.1 for spectrum (5). The difference spectrum shows broad positive and negative bands which coincide with the positions of hemoglobin bands, *e.g.* near 1250 and 1600 cm^{-1}. No difference bands can clearly be assigned to proteins, lipids or other constituents. The Raman data confirm that the hemoglobin content is lower in CCAM than in normal lung tissue because the red and green clusters that are represented by traces (1) and (2) are more abundant in normal lung tissue and the blue cluster, which is represented by trace (3), is more abundant in CCAM. However, the Raman data did not confirm the correlation with the lipid content.

FTIR microscopic images and Raman microscopic maps of lung tissue and CCAM

Fig. 4 shows FTIR microscopic images that are composed of four (a,b,d) and nine (c) individual images covering areas of 540 × 540 μm and 810 × 810 μm, respectively. The positions of the Raman microscopic maps in Fig. 4 largely agree with the FTIR microscopic images. They consist of 49 × 49 (i), 63 × 63 (j), 80 × 80 (k) and 55 × 55 (l) spectra covering areas of 480 × 480 μm, 620 × 620 μm, 790 × 790 μm and 540 × 540 μm, respectively. The photographs are shown in Fig. 4(e–h) for comparison. Normal lung tissue sections (e,f) contain cracks and darker spots in a network-like pattern. CCAM tissue sections (g,h) look different. Here, several features are resolved at the microscopic level that will be identified by their spectral fingerprints. After pooling all 21 FTIR microscopic images, a cluster analysis of the combined data sets containing 86 016 spectra was performed in the wavenumber interval 950–1480 cm^{-1}. A cluster analysis of the combined Raman data sets containing 15 795 spectra was performed in the wavenumber interval 1190–1750 cm^{-1}. Compared with Fig. 1, the number of clusters was increased by three and the additional clusters were colored by yellow, olive and black. The cluster memberships that coincide with the clusters in Fig. 1 are coded in the same colors as before. The percentages of the cluster assignments are summarized in Table 1**(b)**.

In normal lung tissue of patient 1 (first row) and patient 2 (second row), most spectra were assigned to the red clusters (IR: 75/54%, Raman: 74/65%), followed by the green clusters (IR: 24/40%, Raman: 25/34%) and the blue clusters (IR: 1/6%,

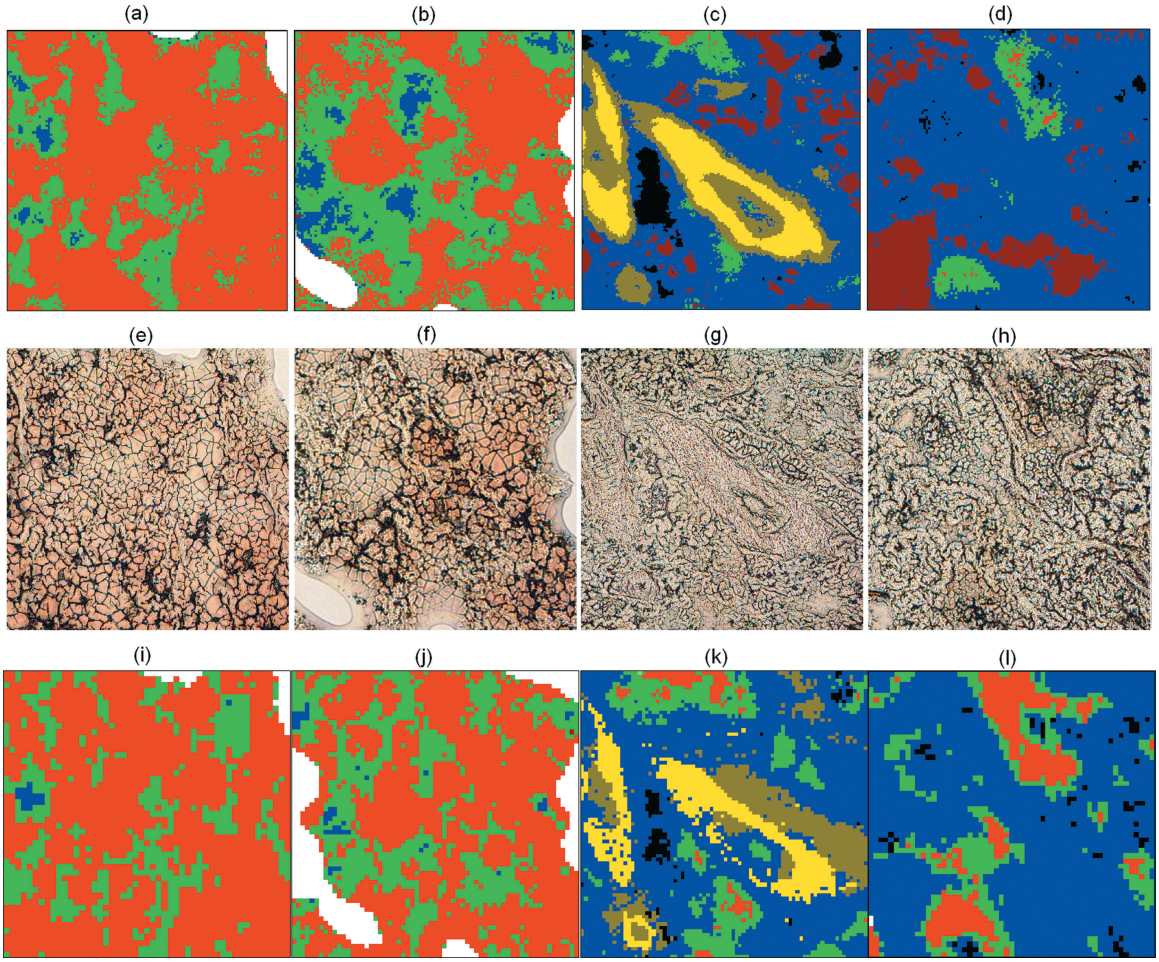


Fig. 4 FTIR microscopic images (a–d), photographs (e–h) and Raman microscopic maps (i–l) of lung tissue sections. Data of the first (a,e,i) and second (b,f,j) columns were obtained from normal samples of patients 1 and 2, respectively. Data of the third (c,g,k) and fourth (d,h,l) columns were obtained from CCAM of patients 1 and 2, respectively. The color code represents the segmentation into seven (a–d) and into six (i–l) groups based on cluster analyses of spectra.

Raman: 1/1%). As fewer than 0.05% of the spectra were assigned to other clusters, the percentage values were rounded to zero. In particular, there was almost no evidence for a brown cluster in FTIR microscopic images that represents mucus at the tissue margins.

In CCAM of patient 1 (third row) and of patient 2 (fourth row), most spectra were assigned to the blue clusters (IR: 55/76%, Raman: 56/71%). Additional clusters form the second largest groups (IR: 40.5/17.6%, Raman: 28/1.7%), and only few spectra were assigned to the green (IR: 4/6%, Raman: 13.4/16%) and red (IR: 0.5/0.4%, Raman: 2.6/11.3%) clusters. Whereas the red, green and blue clusters were already identified in FTIR images and Raman maps in Fig. 1, additional clusters were assigned to new features in microscopic images of the CCAMs. The IR and Raman spectra of these features are described next.

IR spectra of microscopic features in CCAMs

The IR spectra of normal lung tissue and of CCAMs that represent the red, green and blue clusters in FTIR microscopic images are virtually identical with the IR spectra in Fig. 2. Therefore, they are not displayed. Fig. 5 shows the mean IR spectra of the yellow cluster (trace 1), the black cluster (trace 3) and the brown cluster (trace 4), which differs significantly from the brown cluster before (Fig. 2, trace 4).

As the yellow (13%) and olive clusters (15%) were exclusively found in Fig. 4(c), the IR spectra were extracted from this FTIR microscopic image. The IR spectrum of the yellow cluster (trace 1) is overlaid with the spectrum of the blue cluster (trace 2) in Fig. 5(a). The two-times amplified difference spectrum indicates numerous positive and negative bands most notable at 1232, 1337, 1450 cm^{-1} and at 1054, 1095, 1739, 2851, 2921 cm^{-1}, and near amide II, I and A. The negative difference bands are consistent with elevated lipid content in the blue cluster which represents CCAM. The positive difference bands point to collagen in smooth muscle tissues that are present in tissue of the respiratory tract in blood vessels and in bronchioles. Inspection of the photograph confirms the assignment to blood vessels because the yellow cluster surrounds the olive cluster which contains a channel-like feature in the center. Similar spectral contributions of smooth muscle have previously been identified in macroscopic FTIR images of lung tissue.[13] Elevated IR bands of smooth muscles were also detected in IR spectra of the olive cluster, however, with lower intensities than in trace 1 (not shown).

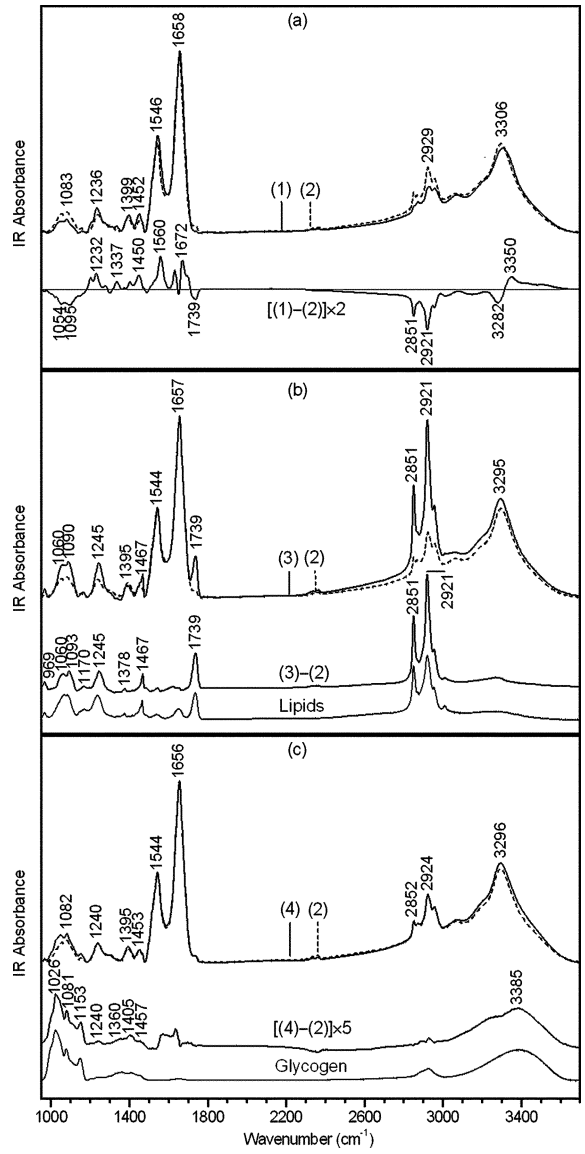


Fig. 5 IR spectra from 950 to 3700 cm^{-1} averaged from the yellow (trace 1), blue (trace 2), black (trace 3) and brown (trace 4) clusters of the FTIR microscopic images of CCAM in Fig. 4(c,d). Difference spectra and spectra of reference compounds were included for comparison.

The IR spectrum of the black cluster (trace 3) is overlaid with the spectrum of the blue cluster (trace 2) in Fig. 5(b). The difference spectrum indicates only positive bands which points to an increased concentration of a tissue component. All bands at 969, 1060, 1093, 1170, 1245, 1378, 1467, 1739, 2851 and 2921 cm^{-1} coincide with an IR spectrum of a lipid extract from bovine brain (purchased from Sigma Aldrich, Germany) shown in Fig. 5(b) which identified this component as lipids. Pronounced bands due to phosphodioxy groups at 1093 and 1245 cm^{-1} are consistent with phospholipids. Relative intensity changes between the CH_2 valence vibration bands at 2851 and 2921 cm^{-1} and the bands in the region 950–1800 cm^{-1} are consistent with longer fatty acid chains in the CCAM lipid aggregates. IR spectra of the black cluster were detected in Fig. 4(c) (2.9%) and in Fig. 4(d) (1.1%) but not in FTIR microscopic images of normal lung tissue in Fig. 4(a,b).

The IR spectrum of the brown cluster (trace 4) is overlaid with the IR spectrum of the blue cluster (trace 2) in Fig. 5(c). Dispersion effects are smaller in trace 4 than in the IR spectrum of the FTIR macroscopic image (Fig. 2, trace 4). The five-times amplified difference spectrum also indicates only positive difference bands which points to an increased concentration of another tissue component. The high similarity with an IR spectrum of glycogen [Fig. 5(c)] with key bands at 1026, 1081, 1153 and 3385 cm^{-1} identified this component as glycogen.[19] IR spectra of the brown cluster were detected in Fig. 4(c) (5.7%) and in Fig. 4(d) 16.5% but not in FTIR microscopic images of normal lung tissue in Fig. 4(a,b).

Raman spectra of microscopic features in CCAM

The Raman spectra of normal lung tissue and CCAM that represent the red, green and blue clusters in Raman microscopic maps in Fig. 4 contain the same spectral properties as the Raman spectra in Fig. 3. Therefore, they are not displayed. Raman spectra that correspond to IR spectra of the brown cluster in Fig. 5(c) were not found. However, Raman spectra that correspond to IR spectra of yellow and black clusters were identified in Fig. 4(k) (26.2 and 1.8%, respectively) and in Fig. 4(l) (1.7%). They are shown in Fig. 6.

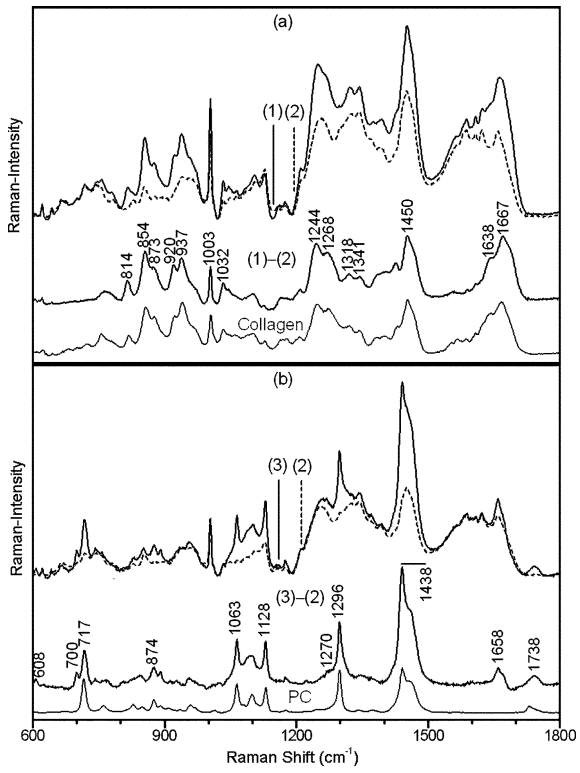


Fig. 6 Raman spectra from 600 to 1800 cm^{-1} averaged from the yellow (trace 1) and black (trace 3) clusters of the Raman microscopic maps of CCAM in Fig. 4(k). A linear combination of red blood cells and the remaining components was calculated (trace 2). Difference spectra and spectra of reference compounds were included for comparison.

The Raman spectrum of the yellow cluster (trace 1) is overlaid with the linear combination $0.4H + 0.55R$ (trace 2) of the spectral contributions of hemoglobin H and the remaining components R [Fig. 6(a)]. The spectra H and R were similar, as

spectra (4) and (5) in Fig. 3. They were calculated from Raman spectra of the red, green and blue clusters of Fig. 4(c) because the yellow cluster is only present here. The difference between spectra (1) and (2) was included in Fig. 6(a). The absence of negative difference bands and minimal positive difference bands in the intervals 650–750 cm^{-1} and 1500–1600 cm^{-1} indicate that spectral contributions of lipids, DNA and hemoglobin are well compensated. The remaining protein bands closely resemble the spectral signature of meninges [Fig. 6(a)] – the skin-like linings of the brain – which are known to contain a high collagen content.[18] Apparently, smooth muscles of blood vessels contain a significant content of a similar collagen type. Key marker bands of collagen that distinguish this protein from other proteins are located at 814, 854, 873, 920, 937, 1244, 1638 and 1667 cm^{-1}. They are due to the unusual amino acid composition with high content of hydroxyproline and proline and to the unusual coiled coil triple-helix secondary structure.

The Raman spectrum of the black cluster (trace 3) is overlaid with the same linear combination $0.4H + 0.55R$ (trace 2) of the spectral contributions of hemoglobin H and the remaining components R [Fig. 6(b)]. The difference spectrum between spectra (3) and (2) also contains no significant negative bands. The absence of positive differences near the prominent protein, hemoglobin and DNA bands indicate that these components did not change in spectra (3) and (2). The high similarity of the difference spectrum with a Raman spectrum of phosphatidylcholine (PC) in Fig. 6(b) suggests that this lipid constitutes the main difference. Further difference bands at 608 and 700 cm^{-1} are assigned to cholesterol and difference bands at 1270 and 1660 cm^{-1} to unsaturated fatty acids. The Raman spectra of lipids have previously been presented in detail.[22] The different relative intensities of the bands of the choline head group at 718 and 875 cm^{-1} and the fatty acid chains at 1065, 1130, 1298 and 1440 cm^{-1} are consistent with longer fatty acid side chains in the lipid aggregates of CCAMs. The side chains of the synthetic PC (purchased from Sigma Aldrich, Germany) were derived from palmitic or hexadecanoic acid which are 16 residues long.

Discussion

In this study, we have described the use of FTIR imaging and Raman mapping in the analysis of CCAMs. Data sets were obtained from four tissue sections that allow comparing (i) the biochemical composition of CCAM and normal lung tissue, (ii) the morphology at macroscopic and microscopic resolutions and (iii) the information content of Raman and IR spectroscopy. Although it was demonstrated that hierarchical cluster analysis (HCA) gives better segmentation,[25] k-means clustering was used in this study because this algorithm can be applied more easily to large data sets. The FTIR images in Fig. 4(a–d) are composed of altogether 86 016 IR spectra. As the time required for HCA increases with the square of the number of spectra, it would take 100 times longer (more than 450 hours or 19 days) than the time required for the same analysis in ref. 25. Another problem is the required memory: the size of the distance matrix in HCA, which also increases with the square of the data points, would be 30 Gbyte which clearly extends the memory of state-of-the-art computers.

Comparison of CCAM and normal lung tissue

The red clusters represent the most abundant spectra in FTIR images and Raman maps of normal lung tissue at microscopic and macroscopic resolutions, whereas the blue clusters represent the most abundant spectra in FTIR images and Raman maps of CCAMs. The green clusters have spectral properties that lie between the red and the blue clusters. As the spectral contributions of red blood cells are at a maximum in the spectra of the red clusters and at a minimum in the spectra of the blue clusters, the abundances indicate that the content of red blood cells is decreased in CCAMs. However, this feature does not seem to be very specific and might also be valid for other tissue types within the lung. More specific features are the elevated content of lipids and glycogen in CCAMs that are inhomogenously distributed throughout the tissue sections. Normal lung tissue is rather homogenous and main variances can be assigned to variations in the red blood cell content.

The overlap, that means blue clusters in normal lung tissue and red clusters in CCAMs, and the percentage values differ between FTIR images and Raman maps because the cluster analysis is an unsupervised algorithm which segments the spectra according to their similarity. A supervised classification model is expected to improve the distinction between CCAM and normal lung tissue. Such a classification approach requires training data that define each tissue class and the limits between the red, green and blue groups. However, more independent data sets have to be collected before such a model can be trained and validated.

Comparison of macroscopic and microscopic resolutions

Low lateral resolution of 63 and 100 μm gave a coarse overview of the chemical composition of the tissue section. Assessing the whole tissue section with a higher lateral resolution would substantially increase the accumulation time and the data size, *e.g.* a ten-fold increase of the resolution requires collecting ten times more spectra both in the x and y dimensions, and the total number of spectra increase by a factor of 100. Whereas single data sets have a typical size of 20 Mbyte, such a procedure gives data sets of 2 Gbyte that are difficult to analyze using standard computer equipment and software. Therefore, data sets at microscopic resolutions were recorded from tissue areas below 1 mm^2. At lateral resolutions near 10 μm, smaller features could be identified such as lipid aggregates, blood vessels and glycogen accumulations. In particular, the chemical composition of CCAMs was inhomogenous on the microscopic level which was not recognized at the macroscopic level. The cluster assignments of FTIR images and Raman maps coincided better in Fig. 4 at high resolution than in Fig. 1 at low resolution. The distinction between normal tissue and CCAMs tends to be better, and the clusters tend to be more continuous in Fig. 4 than in Fig. 1. Therefore, it seems more favorable to select training spectra for a supervised classification approach from microscopic data sets. The resulting model should be able to classify both microscopic and macroscopic data. Comparison of images at high and low magnification is a common procedure in histopathology to assess stained tissue sections. The current study transferred this principle to FTIR imaging and Raman mapping.

Comparison of FTIR imaging and raman mapping

Data acquisition by FTIR imaging is more rapid and yields spectra with higher signal-to-noise ratios because the tissue sections are globally illuminated and the spectra are registered in parallel by a multichannel detector. In contrast, the laser is focused onto a small spot in Raman mapping and the spectra are sequentially registered which is a more time-consuming procedure. Furthermore, most biomolecules show weak Raman signals which requires long exposure times for each spectrum or which gives low signal-to-noise ratios in individual spectra at exposure times of several seconds. Averaging Raman spectra was applied in this study to improve the spectral quality. Total acquisition time in Raman mapping is determined by the product of the number of spectra and the time per spectrum. Under the applied experimental conditions, total acquisition times for Raman maps of 1872–6400 spectra were between 19 and 64 h, whereas the data collection of nine FTIR images plus one additional background image could be completed within 30 min.

FTIR imaging was more sensitive for the detection of lipids and mucus in macroscopic images and glycogen in microscopic images. These components are characterized by a high content of polar C–O, O–H and C–H groups that show usually stronger IR bands than Raman bands. Raman spectroscopy was more sensitive to detect hemoglobin and to determine the head group of lipids. Raman spectral contributions of hemoglobin are enhanced by a resonance effect. The increase in sensitivity is connected with a decrease in specificity because the enhanced hemoglobin bands partly masked the Raman bands of the other, non-enhanced constituents. Most vibrations that are specific for the head groups of lipids are located below 1000 cm^{-1}. This spectral region could not be studied by IR spectroscopy because the tissue sections were prepared on calcium fluoride substrates that show only low transmission of IR radiation below 1000 cm^{-1}. Therefore, phosphatidylcholine in lipid aggregates could be more favorably be identified by Raman bands at 718 and 875 cm^{-1}.

Smooth muscle tissue of blood vessels was detected by both methods with similar sensitivity, which is indicated by equal shapes of the yellow clusters. This tissue class differs from lung tissue and from CCAMs by a low content of lipids and hemoglobin, and a high content of proteins. A major protein is collagen which is known to adopt a three-stranded coiled coil helical secondary structure. Furthermore, the amino acids glycine, proline and hydroxyproline are more frequent in collagen than in other proteins. However, the detection of smooth muscles was not important for the distinction between lung tissue and CCAMs.

Conclusions

Our FTIR imaging and Raman mapping study is the first to reveal details on the chemical composition and the distribution of biomolecules in normal lung tissue and CCAMs. The combined application of FTIR imaging and Raman mapping was used to compare, to complement and to confirm the findings. Unlike Raman and IR spectroscopy, histopathology does not provide this information. This method requires preparatory procedures to highlight the tissue morphology. Embedding in paraffin, subsequent washing with polar and non-polar solvents and staining with contrast enhancing dyes alters the chemical composition, *e.g.* treatment with xylol not only removes paraffin but also virtually all lipids in the tissue sections, treatment with water or ethanol removes glycogen. The biochemical information obtained by Raman and IR spectroscopy can be utilized by magnetic resonance (MR) spectroscopy. This modality has previously detected increased contents of lipids and choline in brain tumors.[26] Based on the current results, MR spectroscopy might also detect these chemical changes in CCAMs. It has recently been reported that the detection of choline-containing compounds in fetuses by *in vivo* MR spectroscopy is feasible.[27] This result clearly demonstrates how vibrational spectroscopic imaging can complement the existing diagnostic techniques. To improve the differential diagnosis, other lung malformations will be studied in forthcoming papers. As the general strategy of combining FTIR imaging and Raman mapping has been shown to be superior, this approach will be applied to other bioanalytical and biomedical problems as well.

Acknowledgements

The authors thank Professor Reiner Salzer (Dresden University of Technology, Germany) for the opportunity to use the FTIR imaging system. This work has partly been funded by the INFRAEUR/BINASP project.

References

1 J. G. Nuchtern and F. J. Harberg, *Semin. Pediatr. Surg.*, 1994, **3**, 233–243.
2 B. J. Oh, J. S. Lee, C. M. Lim and Y. Koh, *Respirology*, 2006, **11**, 496–501.
3 R. D. Wilson, H. L. Hedrick, K. W. Lichty, A. W. Flake, M. P. Johnson, M. Bebbington and N. S. Adzick, *Am. J. Med. Genet., Part A*, 2006, **140**, 151–155.
4 J. T. Stocker, in *Pulmonary pathology*, ed. D. H. Dail and S. P. Hammer, Springer, New York, 1994, pp. 174–180.
5 N. S. Adzick and M. R. Harrison, *World J. Surg.*, 1993, **17**, 342–349.
6 N. S. Adzick and Y. Kitano, *Semin. Pediatr. Surg.*, 2003, **12**, 154–167.
7 C. Granata, C. Gambini, T. Balducci, P. Toma, A. Michelazzi, M. Conte and V. Jasonni, *Pediatr. Pulmonol.*, 1998, **25**, 62–66.
8 N. S. Adzick, *Clin. Perinatol.*, 2003, **30**, 481–492.
9 J. M. Laberge, P. Puligandla and H. Flageole, *Semin. Pediatr. Surg.*, 2005, **14**, 16–33.
10 J. K. Calvert and K. Lakhoo, *J. Pediatr. Surg.*, 2007, **42**, 411–414.
11 C. Krafft and V. Sergo, *Spectrosc. Int. J.*, 2006, **20**, 195–218.
12 R. A. Morotti, J. Cangiarella, M. C. Gutierrez, J. Jagirdar, F. Askin, G. Singh, S. A. Profitt, S. E. Wert, J. A. Whitsett and M. A. Greco, *Hum. Pathol.*, 1999, **30**, 618–625.
13 C. Krafft, D. Codrich, G. Pelizzo and V. Sergo, *Vib. Spectrosc.*, submitted .
14 The McCreery group: http://www.chemistry.ohio-state.edu/~rmccreer/shift.html (accessed 8 January 2008).
15 Cytospec website: http://www.cytospec.com (accessed 8 January 2008).
16 H. Martens, J. P. Nielsen and S. B. Engelsen, *Anal. Chem.*, 2003, **75**, 394–404.
17 D. Naumann, in *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers, John Wiley and Sons, Chichester, 2000, pp. 102–131.
18 C. Krafft, S. B. Sobottka, G. Schackert and R. Salzer, *J. Raman Spectrosc.*, 2006, **37**, 367–375.
19 L. Chiriboga, P. Xie, H. Yee, V. Vigorita, D. Zarou, D. Zakim and M. Diem, *Biospectroscopy*, 1998, **4**, 47–53.
20 L. Chiriboga, P. Xie, V. Vigorita, D. Zarou, D. Zakim and M. Diem, *Biospectroscopy*, 1998, **4**, 55–59.

21 B. R. Wood, P. Caspers, G. J. Puppels, S. Pandiancherri and D. McNaughton, *Anal. Bioanal. Chem.*, 2007, **387**, 1691–1703.
22 C. Krafft, L. Neudert, T. Simat and R. Salzer, *Spectrochim. Acta, Part A*, 2005, **61**, 5129–1535.
23 D. Borchman, D. Tang and M. C. Yappert, *Biospectroscopy*, 1999, **5**, 151–167.
24 W. L. Peticolas, *Methods Enzymol.*, 1995, **246**, 389–416.
25 P. Lasch, W. Haensch, D. Naumann and M. Diem, *Biochim. Biophys. Acta*, 2004, **1688**, 176–186.
26 P. E. Sijens, P. C. Levendag, C. J. Vecht, P. van Dijk and M. Oudkerk, *NMR Biomed.*, 1996, **9**, 65–71.
27 M. S. Clifton, B. N. Joe, A. S. Zektzer, J. Kurhanewicz, D. B. Vigneron, F. V. Coakley, K. K. Nobuhara and M. G. Swanson, *J. Pediatr. Surg.*, 2006, **41**, 768–773.

The Royal Society of Chemistry

Proofs for Correction analyst

Dear Author,

Paper No. b712958k

Please check the proofs of your paper carefully, paying particular attention to the numerical data, tables, figures and references.

When answering the queries below please ensure that any changes required are clearly marked **on the proof**. There is no need to e-mail your answers to the queries separately from the rest of your proof corrections.

Editor's queries are marked like this [Q1, Q2, ...], and for your convenience line numbers are indicated like this [5, 10, 15, ...].

Many thanks for your assistance.

Query	Remarks
Q1 For your information: you can cite this article before you receive notification of the page numbers by using the following format: authors, Analyst, 2008, DOI: 10.1039/b712958k.	
Q2 In Table 1(a), the IR totals for Fig. 1(a/i) and for Fig. 1(b/j) are 99%, as is the RA total for Fig. 1(d/l). Is this OK? Please clarify.	
Q3 In the first sentence of this paragraph, should '47 × 51 (g)' read '47 × 51 (l)'?	
Q4 Should this be Fig. 3(a) and 3b rather than 4(a) and 4(b)?	
Q5 Should this be Fig. 3(c) rather than 4(c)?	
Q6 For Table1(b), the values in the table do not appear to match with the explanation in this first sentence. Should the sentence be something along the lines of 'In normal lung tissue of patient 1 (first row) and patient 2 (second row), most spectra were assigned to the red clusters (IR: 74/75%, Raman: 54/65%), followed by the green clusters (IR: 25/24%, Raman: 40/34%) and the blue clusters (IR: 1/1%, Raman: 6/1%).'? Please clarify.	
Q7 Ref. 13: can this reference be updated yet? Please supply details to allow readers to access the reference (for references where page numbers are not yet known, please supply the DOI).	