Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1758 (2006) 874–882
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.elsevier.com.locate/bbamem

# Diagnosing benign and malignant lesions in breast tissue sections by using IR-microspectroscopy

Heinz Fabian [a,*], Ngoc Anh Ngo Thi [a], Michael Eiden [b], Peter Lasch [a], Jürgen Schmitt [b], Dieter Naumann [a]

[a] *Robert Koch-Institute, P25, Nordufer 20, 13353 Berlin, Germany*
[b] *Synthon KG, Karl Christ Strasse 39, 69118 Heidelberg, Germany*

Received 3 March 2006; received in revised form 16 May 2006; accepted 18 May 2006
Available online 23 May 2006

**Abstract**

The collection of IR spectra through microscope optics and the visualization of the IR data by IR imaging represent a visualization approach, which uses infrared spectral features as a native intrinsic contrast mechanism. To illustrate the potential of this spectroscopic methodology in breast cancer research, we have acquired IR-microspectroscopic data from benign and malignant lesions in breast tissue sections by point microscopy with spot sizes of 30–40 μm. Four classes of distinct breast tissue spectra were defined and stored in the data base: fibroadenoma (a total of 1175 spectra from 14 patients), ductal carcinoma in situ (a total of 1349 spectra from 8 patients), connective tissue (a total of 464 spectra), and adipose tissue (a total of 146 spectra). Artifical neural network analysis, a supervised pattern recognition method, was used to develop an automated classifier to separate the four classes. After training the artifical neural network classifier, infrared spectra of independent external validation data sets ("unknown spectra") were analyzed. In this way, all spectra (a total of 386) taken from micro areas inside the epithelium of fibroadenomas from 4 patients were correctly classified. Out of the 421 spectra taken from micro areas of the in situ component of invasive ductal carcinomas of 3 patients, 93% were correctly identified. Based on these results, the potential of the IR-microspectroscopic approach for diagnosing breast tissue lesions is discussed.


*Keywords:* Infrared microspectroscopy; Infrared imaging; Artifical neural network; Breast cancer; Cancer diagnostic

## 1. Introduction

Traditionally, the diagnosis of breast tumors is based on visual inspection of stained tissue sections by an experienced pathologist. The standard histopathological procedures have in common that they "tag" and visualize the distribution and structure of cellular components in tissue sections using standard light microscopy-based techniques. Infrared microspectroscopy is a relatively new tool for examining tissue sections, which requires neither sample staining nor fixation [1–5]. In combination with advanced data processing techniques, this new approach has the potential to complement established strategies of tissue section diagnosis [3–6]. In a conventional IR imaging system the spectra are collected using an IR microspectrometer with a globar as IR-light source and a single-point detector element. Point-by-point mapping experiments of tissue sections have proved that the IR approach enables the segmentation of major tissue structures based on their IR spectra and the diagnosing of pathologic tissue alterations [6–11]. However, the collection of individual spectra over large areas of tissue is very time consuming (hours) and the spatial resolution is limited because of the low throughput of IR light through small apertures sizes (20–30 μm). The situation improved considerably by a new generation of IR microspectrometers in which the optical aperture is omitted and the single-element detector is replaced by a multi detector array [5]. The latter approach enables to sample larger tissue areas very rapidly (in minutes) and at a spatial resolution of 5–10 μm, which is at the diffraction limit in the mid-IR region.

In the mid-1990s, we have carried out IR-spectroscopy and IR-microscopy studies on human breast tumors, xenografted breast tumors, and breast tumor cell lines, which demonstrated the importance of correct sampling and knowledge of sample histology

* Corresponding author. Fax: +49 30 45472606.
*E-mail address:* fabianh@rki.de (H. Fabian).

0005-2736/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2006.05.015

in order to avoid incorrect interpretation of spectral differences [7,12]. For example, differences on a microscopic level between adjacent small areas (30 $\mu m^2$) of xenografted tumors (transplanted tumors grown in nude mice) have been detected, which could be attributed to local variations in the collagen content [7]. The data suggested that previously reported spectral changes observed by others between malignant and normal tissues, which were attributed to structural changes in DNA, may have an alternative explanation, and may in fact be due to variations in connective tissue content on a microscopic level. Moreover, comparative studies of human breast tumor cell lines demonstrated that related cell types can be differentiated based on their infrared spectra [12]. The studies suggested that infrared methods may have the potential to differentiate not only between major tissue constituents present in breast tissue but also, e.g., between malignant and benign breast tumors.

Recently, we have applied mid-IR microspectroscopic imaging to malignant breast tumor tissue sections [13]. Most types of breast cancer start in the linings of the ducts and lobules of the breast. Typically, cancer-related changes in the breast involve only subtle alterations in the biochemical and morphological composition of the tissue. Moreover, these changes occur at the microscopic level. This requires to collect high-quality spectra and to examine tissue sections with high spatial resolution. As illustrated in our previous paper, breast tissue microheterogeneity is a particularly severe problem for IR microspectroscopy in case of invasive ductal carcinomas. An invasive carcinoma is the progressively destructive phase where the tumor invades the surrounding breast tissue. This implies the need for collecting IR spectra at the level of individual cells to minimize the possibility of "contamination" of the spectra. This "contamination" problem was found to be much less severe in case of the ductal carcinoma in situ. A ductal carcinoma in situ (DCIS) represents the beginning stage of tumor development. The DCIS is characterized by neoplastic cell proliferation inside the breast ducts, whereas the organoid structure remains preserved. Another breast lesion, which is also located in the ducts, is a fibroadenoma. Fibroadenoma is a benign tumor, which is very common and, fortunately, does not tend to become malignant. As a very first step in the evaluation of the diagnostic power of IR-microspectroscopy, we have recently compared the IR spectral characteristics of a fibroadenoma and of a malignant tumor (DCIS) [14]. Our own initial data suggested that the two breast tumor types can be differentiated based on their IR spectra. The work presented here on an extended pool of breast tumor tissue samples applies automated classification methods in order to evaluate the diagnostic potential of the IR approach.

## 2. Materials and methods

### 2.1. Tissue preparation

Human breast tumor tissue was obtained from the tumor bank of the Robert-Rössle-Clinic of Oncology, Humboldt-University, Charité, in Berlin-Buch. For sectioning the tissue blocks were water-mounted on a cryostat table at a temperature of −27 °C. Serial cryo-sectioning was performed at a pre-defined thickness of 7 μm. Slices were either thaw-mounted on 1 mm thick $CaF_2$ or $BaF_2$ windows for IR microscopy or were mounted on conventional glass slides for staining with hematoxylin and eosin for light microscopy. Regions of interest were identified using the hematoxylin and eosin (H&E) stained sections under a light microscope (Olympus BX50). Corresponding patterns on the unstained tissue mounted on $CaF_2$ or $BaF_2$ windows were identified utilizing the native and the Nomarski contrast mode, respectively.

### 2.2. Data acquisition

Conventional point-by-point mapping experiments in transmission mode were performed on two experimental set-ups, each consisting of a Bruker A 590 IR microscope attached to a Bruker IFS28B Fourier transform IR spectrometer. The IR microscopes were isolated in a vented Plexiglas housing to enable purging with dry air in order to reduce spectral contributions from atmospheric water vapour and $CO_2$. Selected areas of interest were mapped with aperture diameters of 30 or 40 μm. Typically, spectra at a spectral resolution of 6 $cm^{-1}$ were generated from 70 interferograms collected at each position.

Selected IR imaging experiments were performed in the application labs of Bruker-Optik GmbH in Ettlingen, Germany and of Perkin-Elmer in Seer Green, UK. The Bruker Hyperion IR imaging system was composed of a Bruker IFS-66 step-scan Fourier transform spectrometer coupled to a Bruker IRScope II IR microscope equipped with a 64 × 64 MCT focal plane array detector, as described previously [13,14]. Data were acquired in transmission mode from areas of 260 by 260 μm mapped to the individual elements of the array detector, resulting in 4096 individual pixel spectra. The spectral resolution was set to 6 $cm^{-1}$.

At Perkin-Elmer, the data were collected using the Perkin-Elmer Spotlight 300 IR microscope equipped with a 16-element linear array detector and a computer-controlled sample stage. Data were acquired in transmission mode and the sample was raster scanned to access an area of 1100 μm × 1060 μm. The spectral resolution was set to 8 $cm^{-1}$ and the acquisition time was approximately 50 minutes.

### 2.3. Data pre-processing and analysis

#### 2.3.1. Cluster analysis

The data were analyzed using routines of the CytoSpec IR imaging software (www.cytospec.com). Pre-processing of the rough data included a quality test to exclude single spectra of very low intensities. Such spectra, e.g., due to tissue gaps, were routinely excluded from further analysis. Spectra passing the test were subsequently vector normalized between 1000 and 1900 $cm^{-1}$. Cluster analysis was performed on first derivative spectra (9-point Savitzky–Golay algorithm). Spectral distances were calculated as so-called *D*-values (redefined Persons correlation coefficients) and the Ward's algorithm was used for hierarchical clustering [11].

#### 2.3.2. Artifical neural network analysis

Data pre-processing, network training, and classification was performed using Synthon's NeuroDeveloper software package specifically designed for ANN analysis of infrared spectral data. Data pre-processing included calculation of first derivatives (Savitzky–Golay algorithm, 5-point smoothing) and averaging two adjacent data points. Artifical neural network calculations were performed based on a three-layer feedforward network with the Rprop learning function[15]. For the analysis, the spectral data (1st derivatives) were split into a training data set (used to estimate the network model parameters), an internal validation data set (used to check the generalization ability of the network), and an external validation data set. The external validation data set can be considered as the best and strictest estimator for the generalization ability of the ANN-model. The high degree of spectral information redundancy over the entire spectral region was diminished by a feature selection algorithm. This improved the quality of the classifier considerably and reduced the complexity and dimensionality of the classification model.

## 3. Results and discussion

### 3.1. Imaging of benign breast tumor tissue sections

Fig. 1A shows the H&E stained section of a breast fibroadenoma. Three major tissue components are discernable: (I) ductal epithelium, (II) mantle tissue and (III) distant connective

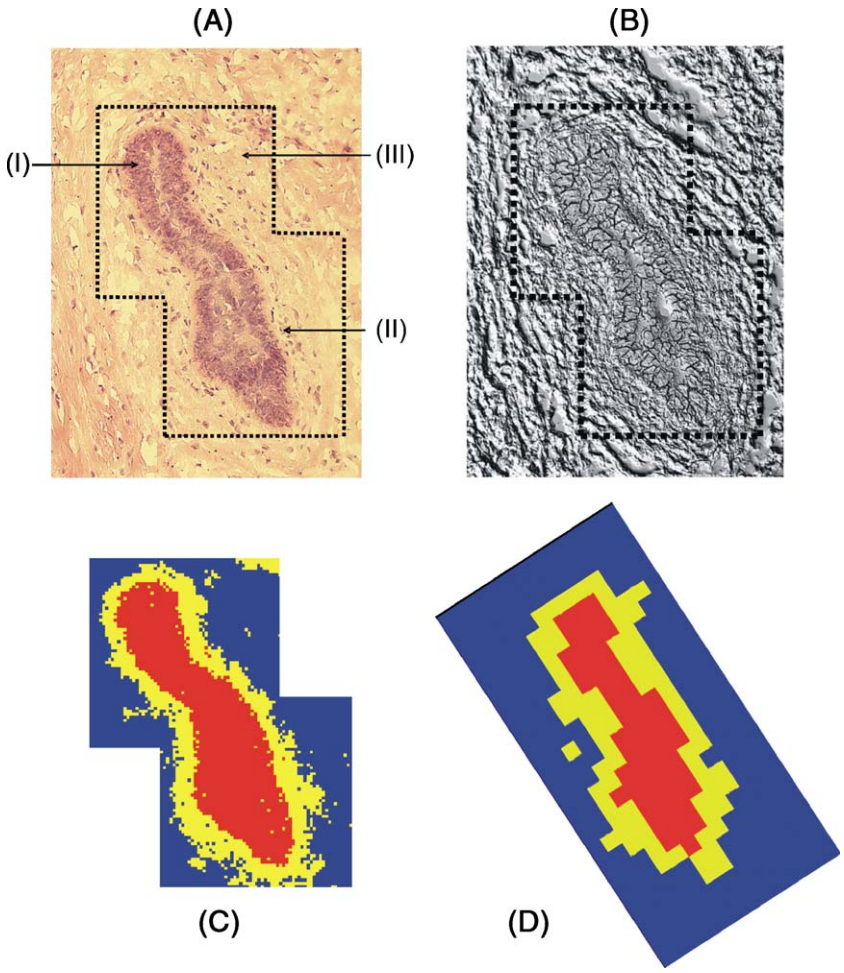


Fig. 1. Images of 7 μm thin cryo-sections of a human breast fibroadenoma. (A) Photomicrograph of a H&E stained tissue section showing three major tissue constituents: benign duct (I), mantle zone (II), and distant connective tissue (III). Magnification: ×100. (B) Photomicrograph of the corresponding serial section mounted on a $BaF_2$ window differentially interference contrasted according to the Nomarski technique (magnification: ×100). The dashed black lines embrace the two areas selected for IR examination. (C) IR image of the fibroadenoma section based on the results of cluster analysis using 2×4096 single pixel spectra obtained with an array detector system at a spatial resolution of 6–10 μm. Each block represents a spectrum acquired at this position. Pixels of the same color indicate that the spectra at these positions fall within the same cluster. Cluster analysis was performed by utilizing the spectral information in the regions 1000–1500 $cm^{-1}$ and 2800–3000 $cm^{-1}$. (D) IR image of the fibroadenoma based on cluster analysis using 26×13 point-by-point spectral data obtained at a spatial resolution of 30 μm.

tissue. Fig. 1B displays the Nomarski contrast image of the adjacent nonstained serial section used for IR analysis. Clear architectural features of the tissue for survey purposes remain visible. This includes the sharp marked ductal outline at the site of the basement membrane as one criterion for structural organization. The areas inside the two dashed black outline squares were examined via an IR microscope equipped with a Bruker focal plane array detector system.

IR spectral images preserves histo-morphological aspects from the light microscopy image. In addition, the IR approach enables the detection and quantification of the molecular biochemistry of microscopic samples, thus opening up a new dimension in the analysis of tissue sections. Several methods can be used to convert IR multispectral data into a two-dimensional image. A frequently used approach is known as functional group intensity or frequency shift mapping [3,4]. This procedure results in a chemically specific contrast of tissue sections, which easily permits the identification and visualization of a number of major tissue constituents, such as collagen, fat or cellular components. More sophisticated approaches, such as unsupervised cluster analysis, are now becoming routine. Cluster analysis groups spectra based upon the degree of similarity between spectra. Spectra (objects) that are similar to each other appear in one group (cluster) and spectra being different are grouped into different clusters. In this way, areas of tissue exhibiting similar biochemistry can be identified. For cluster imaging, the class membership is color encoded and plotted as a function of the $x,y$-coordinates within the tissue section [11].

By comparison of the IR image of the fibroadenoma tissue section based on the results of hierarchical cluster analysis of the 2×4096 single pixel spectra (Fig. 1C) with the H&E photomicrograph of the tissue section (Fig. 1A), the red colored regions can be assigned as the nuclear rich epithelial proliferation of the previously diagnosed breast fibroadenoma. The yellow area intermixed with some blue pixels as seen in Fig. 1C belongs to the mantle zone. The surrounding blue area represents distant connective tissue. The presence of some related red colored pixels within the right circumference of the yellow colored mantle zone

may indicate a micro area of tissue containing only few nuclei [14]. The latter feature cannot be detected by classical point IR microspectroscopy, where the spatial resolution is limited to ~20 μm or more (Fig. 1D). Individual breast tumor cells are only approximately 10 μm in diameter and, thus, the spectrum obtained from a microarea of ~30 μm in diameter is always the composite spectrum of the tumor cells themselves and other components surrounding the cells. This spectral "averaging" problem of all tissue components present in corresponding micro areas is much less severe if the tumor cells are tightly packed, such as in the nuclear rich breast duct epithelia. If one concentrates on these micro areas of a breast tissue section, then it is feasible to minimize the possibility of "contamination" of the spectra even when applying the classical point microspectroscopy approach.

### 3.2. IR spectra of major tissue constituents of benign breast tumor tissue sections

Fig. 2 shows representative IR spectra of major tissue constituents present in fibroadenoma tissue sections obtained by point microspectroscopy. Many of the IR spectral features of breast duct epithelia, connective tissue, and adipose tissue have been previously assigned to defined functional groups of proteins, nucleic acids, and lipids that make up the respective tissue [7–11]. Features between 1500 and 1700 $cm^{-1}$, assigned to the amide I and amide II bands of proteins, dominate the spectra of tumor cells and connective tissue. The most intense band near 1655 $cm^{-1}$ (amide I) arises primarily from the C═O stretching vibrations of the amide groups of the protein backbone, while the absorption band near 1542 $cm^{-1}$ (amide II) arises from amide N–H bending vibrations. The spectral characteristics of both bands are known to be sensitive to protein backbone conformation [16,17]. The broad feature centred between 3290 and 3320 $cm^{-1}$ is due to protein NH stretching vibrations. In connective tissue, the region between 1000 and 1340 $cm^{-1}$ is dominated by absorptions from collagen (e.g., at 1033, 1082, 1234 and 1336 $cm^{-1}$), arising from C–OH stretching vibrations of collagen carbohydrate moieties, C–N stretching vibrations of the collagen backbone and $CH_2$ wagging vibrations of collagen side chains [7]. Nucleic acids give rise to absorptions of medium intensity between 1220–1240 $cm^{-1}$ and at 1083 $cm^{-1}$, arising from $PO_2^-$ asymmetric and symmetric stretching vibrations, respectively [18]. Less intense absorptions attributed to in-plane double-bond vibrations of the bases occur between 1720 and 1630 $cm^{-1}$. Lipids give rise to a number of absorptions in IR spectra. The most intense of these absorptions are found in the range 2800–3000 $cm^{-1}$ and are assigned to asymmetric and symmetric stretching vibrations of $CH_2$ (near 2924 and 2853 $cm^{-1}$) and $CH_3$ (near 2959 and 2873 $cm^{-1}$) groups of the acyl chains. The positions and widths of the $CH_2$ and $CH_3$ absorptions provide information regarding the packing characteristics of the acyl chains, which in turn may be related to the fluidity of cell membranes [19]. Absorptions also arise from the headgroup and the interfacial region of lipids, including absorptions attributed to ester C═O stretching vibrations (1746 $cm^{-1}$), $PO_2^-$ asymmetric (1238 $cm^{-1}$), and symmetric (1080 $cm^{-1}$) stretching vibrations. In isolated phospholipid systems, the ester C═O stretching and $PO_2^-$ stretching vibrations result in absorptions of approximately equal intensity. The low intensity of the ester C═O absorption in Fig. 2A indicates that lipid-like materials contribute little intensity to fibroadenoma cell spectra. Further major absorptions in the spectra of lipids arise from $CH_2$ bending (1463 $cm^{-1}$) and ester CO–O–C stretching

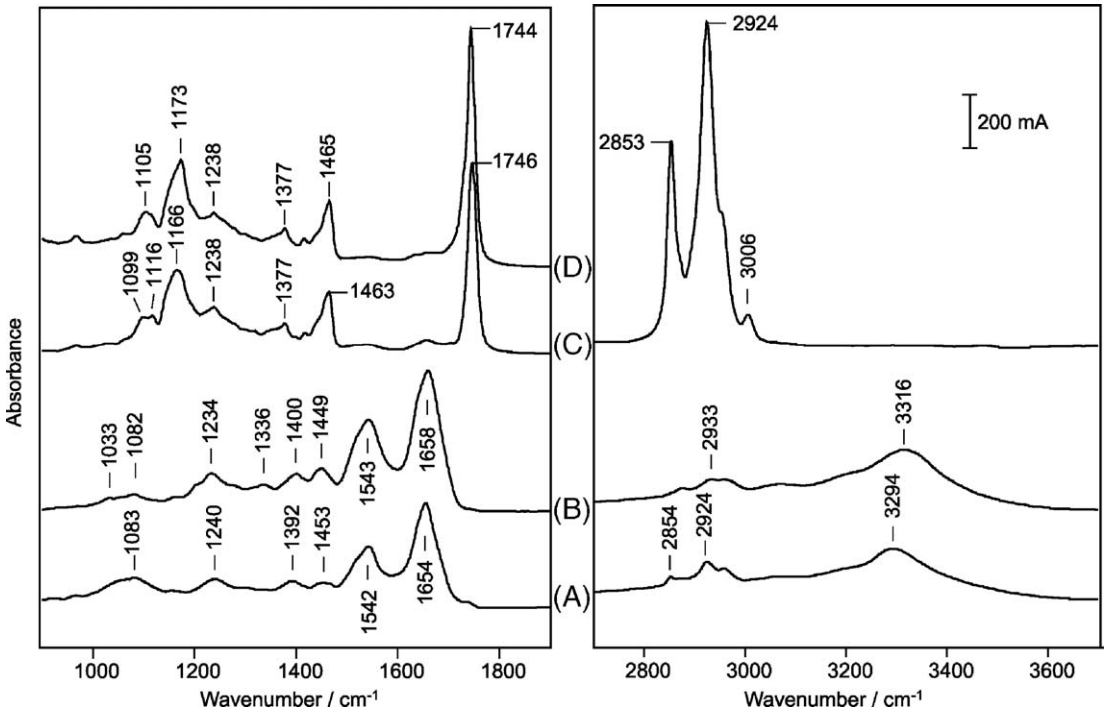


Fig. 2. IR spectra of major tissue constituents observed within benign breast tumor tissue sections acquired through a microscope that sampled an area of ~30 μm in diameter: (A) absorbance spectra of the epithelium of a fibroadenoma, (B) connective tissue, (C) adipose tissue, and (D) milk secretion in ducts.

(near 1165 $cm^{-1}$) vibrations. A further breast tissue constituent, which has been discussed previously [14], is milk secretion fluid. Its spectrum (Fig. 2D) differs in some spectral regions (e.g., between 1000 and 1200 $cm^{-1}$) from the spectrum of adipose tissue (Fig. 2C). The very intense $CH_2$ and $CH_3$ absorptions of milk secretion fluid are not shown for brevity.

### 3.3. Collection of point spectra from benign and malignant breast lesions

A comparison with the FPA spectra revealed that, on the average, the signal-to-noise ratio of the point spectra from different tissue constituents was 5–10 times higher than those of the single pixel spectra obtained by the focal plane array system. Similar results have been obtained for DCIS tissue sections [13]. Different breast tumor types are known to express only subtle differences in the intrinsic biochemical composition and organization of the corresponding tissue. As a consequence, only minor spectroscopic differences between different tumor types can be expected. Although the conventional point mapping approach can fail to detect small numbers of tumor cells owing its low spatial resolution capabilities, it offers the possibility to acquire high-quality spectra. This prompted us to test whether fibroadenoma and ductal carcinoma in situ can be differentiated based on their IR point spectra. For this purpose, tissue sections from 22 patients were examined by using IR microspectroscopy. Being aware of problems associated with microheterogeneity of tissues, only spectra of well-defined tumor areas were pre-selected for this study. After the point-by-point mapping experiments, each mapping data set was evaluated by cluster analysis first. By comparing the cluster image with the H&E-stained photomicrograph of the corresponding tissue section, areas characteristic of major tissue components were identified. Spectra characteristic of the corresponding cluster (tissue component) were selected by a routine of the CytoSpec software and stored. Spectral data that could suffer from a spectral "averaging" problem associated with different tissue components present in the corresponding micro-areas, such as the spectra of the mantle zone of the fibroadenoma shown in Fig. 1D (yellow area), were excluded from further analysis.

### 3.4. Differentiation between fibroadenoma and ductal carcinoma in situ

Four classes of distinct breast tissue spectra were defined and stored in the data base: spectra from fibroadenoma (FA), ductal carcinoma in situ (DCIS), connective tissue (a total of 464 spectra), and adipose tissue (a total of 146 spectra). Artifical neural network (ANN) analysis, a supervised pattern recognition method, was used to develop an automated classifier to separate the four classes. Supervised methods make use of the fact that clinical information concerning the samples from which the spectra were obtained is available. This information is used to train the classifier in order to predict the class identity of unknown spectra afterwards. The quality of the training process can be assessed online by classifying data of a so-called internal validation subset. The spectra of those data sets are not used for teaching but are continuously analyzed during teaching. When the teaching process is finished, the classifier is challenged by a third independent data set for external validation. It is important to note that this latter subset of spectra ("unknown spectra") is kept totally separate form the training and internal validation procedure. The benign tumor spectra from 4 patients (a total of 386 spectra) and the malignant tumor spectra from 3 patients (a total of 421 spectra) were selected for testing the classifier performance (external validation). All the remaining reference spectra of the data base (631 spectra from 10 patients with FA, 743 spectra from 5 patients with DCIS) were combined in a network training and a network internal validation set, respectively. Half of the spectra of connective tissue and fat, respectively, were selected for training. The second half of the data set was splitted into spectra selected for internal and external validation, respectively. Two spectral windows were chosen for network analysis: (i) The range 900–1500 $cm^{-1}$ alone (region 1), and (ii) region 1 in combination with range 2800–3100 $cm^{-1}$. Region 1 was chosen because our previous analysis of a limited data set indicated that the two breast tumor types can be differentiated based on their IR spectra in this spectral range [14].

A combined ANN analysis of the data set containing all 4 classes of breast tissue revealed that all spectra of connective tissue and adipose tissue were classified correctly (data not shown). This is not surprising because their spectral features are clearly different from each other and from those of tumors as well (see Fig. 2). While all tumor spectra used for internal validation were classified correctly as fibroadenoma or DCIS in accordance with the histopathological examination, some problems arised during external validation. The test of the established ANN model with independent fibroadenoma spectra resulted in a correct classification of all FA spectra, but only 85% of the DCIS spectra of the test data set were correctly classified. This was not surprising, because different breast tumor types are known to express only subtle differences in the intrinsic biochemical composition and organization of the corresponding tissue. As a consequence, only minor spectroscopic differences between the two tumor types can be expected. In such cases, a toplevel/sublevel strategy for ANN analysis of the data set turned out to be useful [15]. On the toplevel, only spectra representing 3 classes of tissue (in the present case connective tissue, adipose tissue, and fibroadenoma plus DCIS) were analyzed. After training and internal validation of the toplevel net, all 3 classes of spectra were classified correctly. Afterwards, a sublevel data set representing only fibroadenoma and DCIS spectra was generated. During training of the subnet, the algorithm is now only searching for spectral features that permit to differentiate between the two tumor types.

The spectral information of the two classes discriminating DCIS and FA in the sublevel network was extracted by transforming the training dataset into principal components. Principal Component Analysis (PCA) is a statistical method used for the extraction of a set of uncorrelated variables which explain most of the variance in the dataset. The method comprises most of the information of the dataset within those variables (called principal components) and thus, can be used for data compression and unsupervised feature extraction tasks [15]. In our case, PCA of the training spectra representing DCIS and FA resulted in 45 principal

Table 1
Classification results of neural network analysis of breast tumor spectra

| Tissue | Training | Internal validation | External validation | |
|---|---|---|---|---|
| | | | 900–1500 and 2800–3100 $cm^{-1}$ | 900–1500 $cm^{-1}$ |
| FA | 631 | 158 (158) | 386 (386) | 386 (380) |
| DCIS | 743 | 185 (185) | 421 (391) | 421 (384) |

The numbers represent the total number of spectra used for training, internal validation, and external validation of the network. The numbers in parentheses are the spectra classified by ANN analysis in accordance with the histopathological classification of the corresponding tumor. FA = fibroadenoma, DCIS = ductal carcinoma in situ. Note that the spectra of the external validation data set were not included into the training/internal validation process.

components with eigenvalues greater than one. Although the computed principal components cannot be interpreted easily due to their nature as linear combinations of the original values, their information content regarding the separability of "DCIS" from "FA" is higher those of the original data. Subsequent ANN training was performed on the generated principal components using resilient backpropagation as learning function [13]. Although principal components greater than 10 may contain very little spectral information, automated search for optimal neural network topology revealed that a feed-forward network using finally a subset of 26 principal components as input neurons, 20 units in the hidden layer and 2 output units with the class assignments "DCIS" and "FA" performed best.

Based on the toplevel/sublevel strategy, all of the fibroadenoma spectra and 93% of the DCIS spectra were correctly classified (Table 1). This result was obtained by analyzing the spectral information contained in both, the range 900–1500 $cm^{-1}$ and in the CH stretching region (2800–3100 $cm^{-1}$). The accuracy of classification of fibroadenoma was slightly reduced when the latter region was omitted from data analysis (right column in Table 1), suggesting that lipid-like spectral features might be important for classification. It is interesting to note that practically all the misclassified point spectra of DCIS were members of the spectral data set acquired from a single tissue section (29 out of 220 spectra). This suggests that the small number of misclassified spectra may represent an intrinsic variability of the corresponding section, rather than a weakness of the established ANN classifier.

To get possible insights into the global chemical changes associated with the two breast lesions, we have compared the mean IR spectra of fibroadenoma and of ductal carcinoma in situ (Fig. 3A). The majority of the spectral features of the difference spectrum (at 1173, 1380, 1466, 2852 and 2923 $cm^{-1}$ in Fig. 3B) correlates with major absorption bands of fat (Fig. 1C). This suggests that the microareas diagnosed as DCIS have more lipid-like material than those diagnosed as fibroadenoma. As discussed previously, only spectra of well-defined tumor areas have been preselected for this comparative study. Thus, it seems unlikely that the higher content of fat-like features observed in the DCIS spectra is simply a consequence of the spectral "averaging" problem associated with the collection of point spectra from microareas of 30–40 μm in diameter. This suggestion was corroborated by a comparative analysis of the FPA single pixel spectra acquired from the fibroadenoma shown in Fig. 1 and a DCIS [13] (data not shown). Although the signal-to-noise ratio of

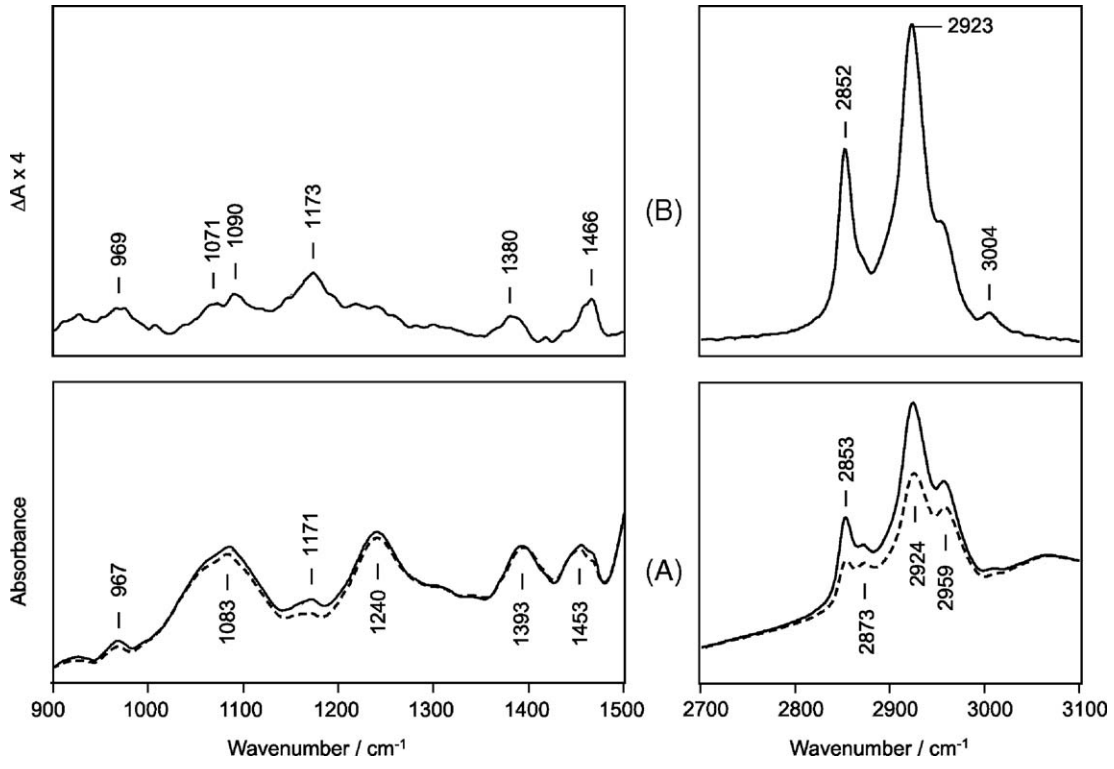


Fig. 3. (A) Mean IR spectra of breast ducts representative of the epithelium of a fibroadenoma (dashed line) and of the in situ component of an invasive ductal carcinoma (solid line). The spectra are shown after vector normalization in the range 1800–900 $cm^{-1}$. (B) IR difference spectrum (mean DCIS spectrum minus mean FA spectrum). Note that the absorbance scale for the difference spectrum in part B was expanded by a factor of 4 compared to the scale in part A.

the single pixel spectra was approximately one order of magnitude lower than those of the spectra obtained by point-by-point mapping, the mean FPA spectra also revealed a higher content of fat-like features in DCIS. In order to test whether fat-like IR spectral features alone permit to differentiate DCIS and FA, we have trained an ANN classifier using only the spectral information contained in the range 2800–3100 $cm^{-1}$. The external validation with independent DCIS spectra revealed that almost half of the DCIS spectra were misclassified. This demonstrates that a complex pattern of features across a wide spectral range (900–1500 $cm^{-1}$ in combination with 2800–3100 $cm^{-1}$), rather than specific fat-like features alone, are necessary for differentiation of fibroadenoma from ductal carcinoma in situ.

During our research, we have also examined tissue sections diagnosed as invasive ductal carcinomas (IDC). An invasive carcinoma is the phase where the tumor infiltrates in between surrounding breast tissue, which is often made up by adipose tissue. An inspection of the IR spectra acquired by classical point microspectroscopy not seldomly revealed large differences from spectrum to spectrum, which could be correlated with variations in the amount of fat present in the corresponding microareas. This suggests that the data set of individual point spectra of sections diagnosed as IDC might not adequately represent the "pure" IR features of IDC. A spectrum obtained from a spot of ~30 μm in diameter is always the composite spectrum of individual breast tumor cells themselves and of other components surrounding the cells. If the cells are not as tightly packed as in nuclear rich breast ducts, the tumor cell signal may be overwhelmed by the spectral contributions of other components surrounding the cells, which may vary from spot to spot. Spectral differences among individual data sets unrelated to the disease state may cause the classifier to behave unpredictably and, thus, the point spectra acquired from IDC sections could not be used for ANN analysis. While classical point IR microspectroscopy is a useful practical approach for examination of lesions located in breast ducts, IR imaging approaches offering higher spatial resolution seems to be mandatory for the characterization of invasive breast lesions.

Some state-of-the-art IR imaging instruments now permit to collect single pixel spectra with high spectral quality and low pixel-to-pixel variability [20]. The spatial resolution of the instruments is determined by the optical configuration of the system, which is usually set to 5–10 μm. As an example, recently we had

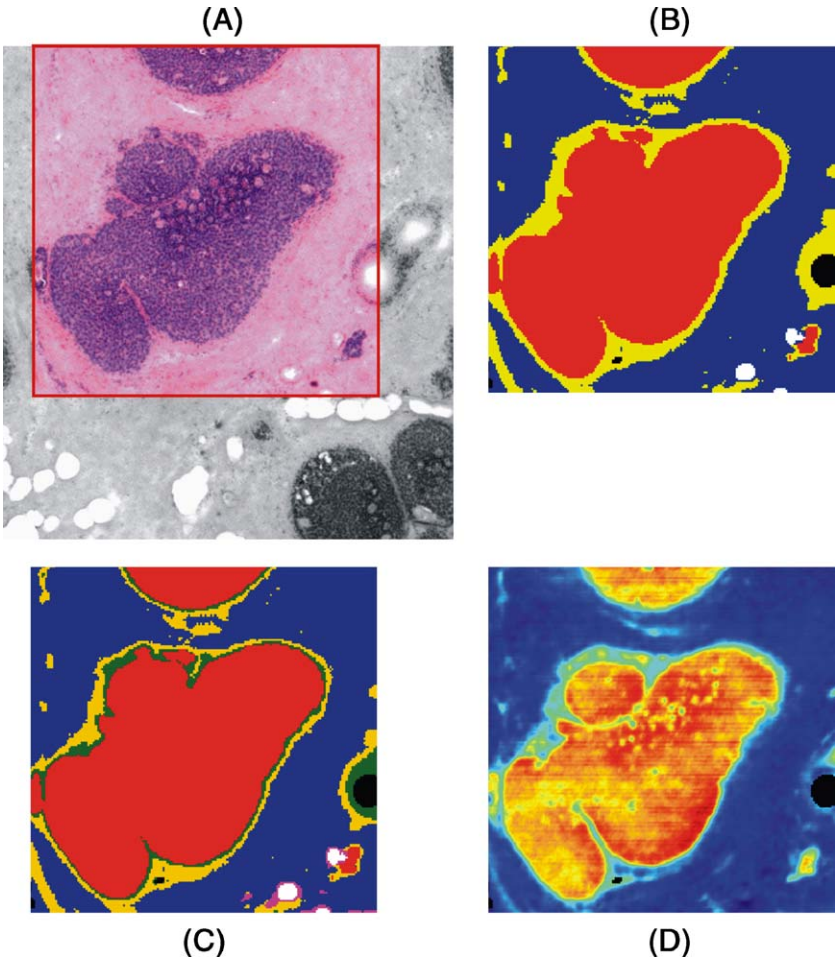


Fig. 4. Images of a breast tissue section containing a ductal carcinoma in situ. (A) Visible image of a cryosection of unstained breast tissue, overlayed by a photomicrograph of the H&E stained adjacent section of the specimen. The red line defines the area selected for IR examination. (B) IR image of clustering results generated by using four clusters. (C) IR image of clustering results generated by using six clusters. (D) IR image based on IR band intensities at 1080 $cm^{-1}$. The intensity values (red, high; blue, low) were taken from the absorbance spectra after normalization with respect to the integrated intensity between 1000 and 1900 $cm^{-1}$. Individual spectra with very low absorbance values are indicated by black spots.

the opportunity to acquire IR imaging data of breast tissue sections using a Perkin-Elmer Spotlight 300 system. The sample was raster scanned over an area of $1.10 \times 1.06$ mm$^2$. Each pixel sampled of an area of $6.25 \times 6.25$ μm at the sample plane. Fig. 4A shows a visible image of a cryosection of unstained breast tissue, overlayed by a photomicrograph of the H&E stained adjacent section of the specimen. The H&E image reveals a ductal carcinoma in situ (stained in purple) in the center and on the upper side, surrounded by connective tissue (stained in pink). Colorless areas point to fat or tissue gaps. The red line indicates the area selected for IR examination. Cluster analysis of the 29920 single pixel spectra permitted identification of four major clusters of spectra. Fig. 4B displays an IR image based on cluster analysis of the $176 \times 170$ single pixel spectra. Spectra of the DCIS are color encoded by red, while the blue regions correspond to collagen rich connective tissue. Spots dominated by fat are shown in white (lower right corner). Comparison of the cluster image with the histological data reveals that the yellow color encodes single pixel spectra which arise primarily from spots at the borderline between tumor and connective tissue. Thus, the four major clusters of spectra represent specific tissue structures of the section.

The clustering process is unsupervised (i.e. objective), but the definition of the number of clusters is a subjective step in this procedure, and may become more and more ambiguous in terms of known histopathology. However, images generated by superimposing an increasing number of single color cluster images permits one to identify whether there are differences between spectra belonging to a major clusters. In case of the studied DCIS section, maps obtained for 5 (not shown) and 6 clusters (Fig. 4C) only revealed some substructures in the borderline regions between tumor and connective tissue, while the tumor regions itself were encoded by the same color. The latter result shows that the single pixel spectra of the tissue regions diagnosed as DCIS are practically identical. In other words, no tissue substructures within the tumor regions could be detected at a spatial resolution of 6–7 μm. This corroborates our previous study [14], which suggested that point microscopy with spot sizes of 20–40 μm may be a useful practical approach for acquiring IR spectra of assemblies of tumor cells in breast ducts.

A procedure, which easily permits the identification and visualization of major tissue constituents, is functional group intensity mapping. For example, the contrast image shown in Fig. 4D is based on IR band intensities at 1080 cm$^{-1}$, which are known to be primarily characteristic of nuclear material (DNA, RNA). The locations, sizes, and shapes of the red colored areas in Fig. 4D are in good accordance with the H&E stained tumor areas shown in Fig. 4A. Thus, plotting the intensity at around 1080 cm$^{-1}$ as a function of the x,y-coordinates within the tissue section is an easy and fast way for discriminating nuclear material from connective tissue. The specific spectroscopic image contrast, arising from the inherent chemical composition of the tissue rather than from extrinsic reagents, offers some interesting possibilities for bioanalytical applications. For example, laser micro-dissection techniques are established methodologies to remove nuclear-rich material from tissue sections. In order to identify the locations of the nuclear-rich areas, typically conventional staining procedures have to be applied. In future, a combination of IR imaging techniques with laser micro-dissection technologies might provide a novel means of identifying micro areas of interest without the need for tissue staining.

## 4. Conclusions

The present study demonstrates that benign and malignant lesions in breast tumor tissue sections, such as fibroadenoma and ductal carcinoma in situ, can be differentiated based on their IR spectra. Despite only small differences between the spectra of the two lesions, artifical neural networks strategies were able to recognize spectral features that permitted us to develop an automated classifier to separate fibroadenomas from ductal carcinomas in situ. The classification was based on a complex pattern of features across a wide spectral range. This argues against the concept that the simple analysis of distinct spectral markers, such as suggested by other authors [21] could serve for the differentiation of benign from malignant breast tissue. Moreover, this work provides further insights into problems associated with the variability of the samples on a microscopic level. For instance, the spectral "averaging" problem prevented us from obtaining "pure" spectra of invasive ductal carcinomas, even when the spectra were acquired using aperture diameters of only 30–40 μm. Thus, although promising, the present study represents only a first step in the evaluation of the diagnostic power of the IR approach. A significantly extended data base of spectra of histopathologically well-defined tissues, which spans the natural variability of healthy and tumor breast tissues, is required in order to fully explore the possibilities of IR microspectroscopy. Here, IR microspectrometers equipped with multi detector arrays have paved the way to minimize problems associated with tissue microheterogeneity and data acquisition time, an essential prerequisite for future studies.

## Acknowledgments

The efforts and advice of our clinical collaborator Dr. W. Haensch (Robert-Rössle-Clinic of Oncology, Berlin) for evaluating the tissue samples is gratefully acknowledged. In addition, we thank Mrs. I. Wendler for providing the breast tumor samples and for initial assistance in preparation of the tissue slices. We are grateful to Dr. R. Spragg (Perkin Elmer) and Dr. M. Boese (Bruker-Optik GmbH) for their contributions in acquiring IR imaging data in the applications labs in Seer Green, UK and Ettlingen, Germany, respectively. This work was supported by a grant of the Deutsche Krebshilfe to H.F. (70-2420-Fa).

## References

[1] D.L. Wetzel, S.M. LeVine, Microspectroscopy—Imaging molecular chemistry with infrared microscopy, Science 285 (1999) 1224–1225.
[2] R. Dukor, Vibrational spectroscopy in the detection of cancer, in: J.M. Chalmers, P.R. Griffiths (Eds.), Handbook of Vibrational Spectroscopy, vol. 5, Wiley, Chichester, 2002, pp. 3335–3361.
[3] M. Diem, M. Romeo, S. Boydston-White, M. Miljkovoc, C. Matthaus, A decade of vibrational micro-spectroscopic imaging of human cells and tissue (1994–2004), Analyst 129 (2004) 880–885.
[4] A. Boskey, R. Mendelsohn, Infrared analysis of bone in health and disease, J. Biomed. Opt. 10 (2005) 031102-1–031102-9.

[5] I.W. Levin, R. Bhargava, Fourier transform infrared vibrational spectroscopic imaging: integrating microscopy and molecular recognition, Annu. Rev. Phys. Chem. 56 (2005) 429–474.
[6] J. Kneipp, M. Beekes, P. Lasch, D. Naumann, Molecular changes of preclinical scrapie can be detected by infrared spectroscopy, J. Neurosci. 22 (2002) 2989–2997.
[7] H. Fabian, M. Jackson, L. Murphy, P.H. Watson, I. Fichtner, H.H. Mantsch, A comparative infrared spectroscopic study of human breast tumors and breast tumor cell xenografts, Biospectroscopy 1 (1995) 37–45.
[8] B.R. Wood, M.A. Quinn, B. Tait, M. Ashdown, T. Hislop, M. Romeo, D. McNaughton, FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies, Biospectroscopy 4 (1998) 75–91.
[9] L.M. McIntosh, J.R. Mansfield, A.N. Crowson, H.H. Mantsch, M. Jackson, Analysis and interpretation of infrared microscopic maps: visualization and classification of skin components by digital staining and multivariate analysis, Biospectroscopy 5 (1999) 265–275.
[10] M. Diem, S. Boydston-White, L. Chiriboga, Infrared spectroscopy of cells and tissues: shining light onto a novel subject, Appl. Spectrosc. 53 (1999) 148A–161A.
[11] P. Lasch, W. Haensch, D. Naumann, M. Diem, Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis, Biochim. Biophys. Acta 1688 (2004) 176–186.
[12] H. Fabian, R. Wessel, M. Jackson, A. Schwartz, P. Lasch, I. Fichtner, H.H. Mantsch, D. Naumann, IR-spectroscopy and IR-microscopy of human breast tumors, xenografted breast tumors, and breast tumor cell lines, Proc. SPIE 3257 (1998) 13–23.
[13] H. Fabian, P. Lasch, M. Boese, W. Haensch, Mid-IR microspectroscopic imaging of breast tumor tissue sections, Biopolymers 67 (2002) 354–357.
[14] H. Fabian, P. Lasch, M. Boese, W. Haensch, Infrared microspectroscopic imaging of benign breast tumor tissue sections, J. Mol. Struct. 661–662 (2003) 411–417.
[15] J. Schmitt, T. Udelhoven, Use of artifical neural networks in biomedical diagnosis, in: H.-U. Gremlich, B. Yan (Eds.), Infrared and Raman Spectroscopy of Biological Materials, Marcel Dekker, New York, 2000, pp. 379–419.
[16] H. Fabian, W. Mäntele, Infrared spectroscopy of proteins, in: J.M. Chalmers, P.R. Griffiths (Eds.), Handbook of Vibrational Spectroscopy, vol. 5, Wiley, Chichester, 2002, pp. 3399–3425.
[17] A. Barth, C. Zscherp, What vibrations tell us about proteins, Q. Rev. Biophys. 35 (2002) 369–430.
[18] E. Taillandier, F. Liquier, Infrared spectroscopy of DNA, Methods Enzymol. 211 (1992) 307–335.
[19] M. Jackson, H.H. Mantsch, Biomembrane structure from FTIR spectroscopy, Spectrochim. Acta 15 (1993) 53–96.
[20] D.C. Fernandez, R. Bhargava, S.M. Hewitt, I.W. Levin, Infrared spectroscopic imaging for histopathologic recognition, Nat. Biotechnol. 23 (2005) 469–474.
[21] Y. Ci, T. Gao, J. Feng, Z. Guo, Fourier transform infrared spectroscopic characterization of human breast tissue: Implications for breast cancer diagnosis, Appl. Spectrosc. 53 (1999) 312–315.