Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1758 (2006) 883–891
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.elsevier.com/locate/bbamem

Identification of primary tumors of brain metastases by SIMCA classification of IR spectroscopic images

Christoph Krafft [a,*], Larysa Shapoval [a], Stephan B. Sobottka [b], Kathrin D. Geiger [c], Gabriele Schackert [b], Reiner Salzer [a]

[a] *Institute for Analytical Chemistry, Dresden University of Technology, 01062 Dresden, Germany*
[b] *Clinic for Neurosurgery, University Hospital Dresden, Dresden University of Technology, Fetscherstr. 74, 01307 Dresden, Germany*
[c] *Department of Neuropathology, Institute for Pathology, University Hospital Dresden, Dresden University of Technology, Fetscherstr. 74, 01307 Dresden, Germany*

Received 1 December 2005; received in revised form 24 March 2006; accepted 1 May 2006
Available online 19 May 2006

Abstract

Brain metastases are secondary intracranial lesions which occur more frequently than primary brain tumors. The four most abundant types of brain metastasis originate from primary tumors of lung cancer, colorectal cancer, breast cancer and renal cell carcinoma. As metastatic cells contain the molecular information of the primary tissue cells and IR spectroscopy probes the molecular fingerprint of cells, IR spectroscopy based methods constitute a new approach to determine the origin of brain metastases. IR spectroscopic images of 4 by 4 mm^2 tissue areas were recorded in transmission mode by a FTIR imaging spectrometer coupled to a focal plane array detector. Unsupervised cluster analysis revealed variances within each cryosection. Selected clusters of five IR images with known diagnoses trained a supervised classification model based on the algorithm soft independent modeling of class analogies (SIMCA). This model was applied to distinguish normal brain tissue from brain metastases and to identify the primary tumor of brain metastases in 15 independent IR images. All specimens were assigned to the correct tissue class. This proof-of-concept study demonstrates that IR spectroscopy can complement established methods such as histopathology or immunohistochemistry for diagnosis.

Keywords: Vibrational imaging; Biomedical spectroscopy; Secondary brain tumors; Chemometric methods; Molecular pathology

1. Introduction

Cerebral metastasis occur in about 15–40% of all cancer patients representing about 12 patients per 100000 of the U.S. population per year (review in [1]). This proportion has increased in recent years partly owing to the increased sensitivity of MRI for detecting these tumors and also the longer survival of these patients. Secondary intracranial lesions are estimated to become symptomatic in about 80,000 patients per year compared to 18000 with primary brain tumors which originate from cerebral tissue. The secondary brain tumor is responsible for the first symptom in almost one-half of patients operated on for a cerebral metastasis. The clinical presentation with brain symptoms may be due to the fact that brain metastases cause pronounced surrounding edema and thus give symptoms at a much earlier stage while other manifestations such as thoracic symptoms occur rather late in lung cancer which is the most frequent primary site (51%) causing brain metastasis. Other frequent primary sites include breast cancer, colorectal cancer and renal cell carcinoma.

The diagnosis of brain metastasis in a patient with a known systemic cancer is a simple matter. However, the diagnosis in a patient with a brain metastasis of unknown primary (BMUP) is difficult. The majority of cases are adenocarcinomas with poorly differentiated tumors which show even less differentiation in brain metastases. Unfortunately, adenocarcinomas metastatic from different locations have similar microscopic appearance, impeding histological diagnosis, and almost invariably demanding the application of additional immunohistochemical methods. Consequently, in up to 15% of the patients

* Corresponding author. Tel.: +49 351 46332507; fax: +49 351 46337188.
E-mail address: christoph.krafft@tu-dresden.de (C. Krafft).

0005-2736/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2006.05.001

with brain metastasis the primary tumor will not be identified despite thorough investigation by standard techniques. Thus, BMUP are estimated to be almost equal in numbers to primary brain tumors. There is a strong need to develop new diagnostic methods to determine the origin of brain metastasis and to treat the primary tumor effectively and early as the majority of patients die from the extracranial disease [2].

This report presents an approach combining infrared (IR) spectroscopic imaging and multivariate classification to identify the primary tumor of brain metastases, as metastatic cells contain the molecular information of the primary tissue cells and IR spectroscopy probes chemical and structural properties of tissues at the molecular level. The approach utilizes the concept that biological molecules such as proteins, nucleic acids and lipids possess specific IR spectra which can be considered as their fingerprints, and IR spectra of tissues represent sensitive fingerprint-like signatures with spectral contributions of all constituents. Advantages of IR spectroscopy for diagnostic applications include that the method is rapid and label-free as spectral data can be collected and interpreted within minutes by automated techniques, and the excited molecular vibrations monitor molecular structures and cellular chemistry of samples without external markers and with minimal or no preparation. Biomedical applications have recently been reviewed [3–5].

Usually the spectral information which characterizes tissue classes is distributed throughout the IR spectra. Simple univariate evaluation techniques (such as band intensities, positions or ratios) are often unable to find a correlation between IR spectra and tissue classes. Therefore, more complicated multivariate algorithms are required to recognize diagnostic patterns in IR spectra and to utilize them for classification such as linear discriminant analysis (LDA), artificial neural networks or support vector machines [6,7]. In general, classification models based on these supervised classification algorithms are trained by data of known class assignment. It is of the utmost importance that training data are carefully selected. Then these models are validated by identification of independent test data.

Besides the wide distribution of spectral descriptors, another difficulty is the fact that tissue specimens usually encompass a certain degree of inherent heterogeneity. Therefore, spatially resolved IR data are required for precise tissue studies. In mapping mode, IR spectroscopic images are recorded sequentially in a raster pattern by a FTIR spectrometer with a single channel detector and a movable sample stage. In imaging mode, IR spectroscopic images of globally illuminated samples are recorded by a FTIR spectrometer with a multichannel detector and stepwise acquisition of spectral data. The imaging mode offers a significant time advantage as typical acquisition of IR spectroscopic images using a 64×64 focal plane array (FPA) detector comprising 4096 IR spectra just takes 4 min while the mapping mode requires ca. 800 min to acquire the corresponding number of IR spectra. After introduction of IR imaging spectrometers [8], first studies have been reported on breast tissue [9,10], cartilage [11,12], bone [13], colon tissue [14], prostatic tissue [15], spleen [16], brain [17,18], skin [19] and cervical tissue [20].

In the present study, IR spectroscopic imaging is applied to distinguish normal brain tissue from brain metastases and to determine the primary tumor of brain metastases from renal cell carcinoma, lung cancer, colorectal carcinoma and breast cancer. Inherent biological variances constitute a severe problem to select training data from IR spectroscopic images for supervised classification algorithms. Intra-sample variances can be observed if tissue specimens encompass several tissue types. Inter-sample variances can occur between specimens from different patients with the same diagnosis. Therefore, training data were extracted from five IR images which consider these variances and accurately represent the diagnostic features of each tissue class. Tissue sections are evaluated by neuropathology in order to confirm the selection of training data. Then, a multivariate classification model based on the algorithm soft independent modeling of class analogies (SIMCA) is developed and used to assign 15 independent IR images. The accuracy of the classification model is discussed and compared to a classification of the same data using LDA [21].

2. Materials and methods

2.1. Tissue samples

All experimental procedures were approved by the Human Ethics Committee of the Dresden University of Technology. Normal brain specimens were obtained from three individuals postmortem during autopsies. Tumor specimens were obtained from 17 patients undergoing neurosurgery in the period between February 2000 and March 2003. Fresh samples were snap frozen in liquid nitrogen immediately after tumor resection without fixatives and antifreeze media. 10 μm tissue sections were obtained from a cryostat, transferred onto a glass slide for histopathologic assessment after hematoxylin and eosin (H&E) staining and onto an IR transparent calcium fluoride slide for IR spectroscopy in transmission mode. All tissue sections were dried on air before further processing. Pathological diagnosis was also obtained independently from additional formalin-fixed and paraffin-embedded tissue samples using standard histological and immunohistological techniques.

2.2. Data collection

IR spectroscopic images were recorded using a Bruker FTIR spectrometer IFS66/S (Bruker Optik, Ettlingen, Germany) equipped with a FPA detector of 64×64 pixels. Coupling to the macro chamber IMAC (Bruker) gave a field of view of 4×4 mm^2 per IR image with spatial resolution of 63×63 μm per pixel. The FTIR spectrometer and the sample chamber were continually purged by dry air from air purifiers. Images of 4096 IR spectra at 4 cm^{-1} spectral resolution were acquired by the OPUS software (Bruker) operating the FPA in continuous scan mode by coadding 21 frames. Recording time took ca. 4 min per IR image. The single beam spectra of each sample IR image were ratioed against the corresponding spectra of a background IR image from a pure calcium fluoride slide and converted to absorbance.

2.3. Image preprocessing

Spectral data sets were imported into MatLab (The Matworks, Natick, MA, USA). Data pretreatment was performed by in-house written routines. The spectral region from 950 to 1800 cm^{-1} was used for analysis. A linear baseline was subtracted to compensate for unspecific differences in spectral offset and background. Raw spectra were filtered to remove low absorbance spectra with poor signal to noise ratios from further analysis because they mainly originate from points without sample, from holes and fissures in tissue sections or from defect pixels of the FPA detector due to delamination. Minimum–maximum

normalization was performed to compensate for different intensities due to thickness variations.

2.4. Cluster analysis

In the present context cluster analysis partitions the ensemble of spectra into groups (clusters) with distinct similarities within each cluster and differences between the clusters. Our implementation distributes the spectra over a fixed, pre-selected number of clusters. The unsupervised classification method k-means cluster analysis under the MatLab environment included an Euclidean distance metric and data normalization by the multiplicative scattering method. The initial spectra for the centers of the clusters were taken randomly from the spectral data set. All spectra were then compared with these cluster centers, a new cluster was calculated by taking the mean of all spectra that were assigned to that cluster. This procedure was repeated until the spectral differences within the clusters were minimized and between clusters were maximized.

2.5. Soft independent modeling of class analogies (SIMCA)

Principal component analysis (PCA), also called singular value decomposition (SVD), projects a data matrix into a new coordinate system in order to describe the variances within the data set by principal components (PCs) which are orthogonal to each other [22]. As the explained variance per PC decreases with increasing number of PCs, the first PCs encompass almost all variance whereas the higher PCs are dominated by noise. PCA is a frequently used multivariate approach to analyze IR images [6,23] or to reduce noise in spectral data [24]. A single PC has been used as input for a LDA classification model to distinguish tumor tissue and necrosis by Raman spectroscopy [25]. The SIMCA algorithm which is based on a concept by Wold [26] is able to utilize a defined number of PCs for supervised classification. First SIMCA classifications of IR spectra have been reported in pharmaceutical research [27], in food research [28], and for identification of bacteria [29]. Here, SIMCA is applied for the first time to classify brain tissue based on IR images.

The SIMCA algorithm was implemented in the Matlab environment using the PLS toolbox (Eigenvector Research, Manson, MA, USA). For each class, training data were selected which represent the spectral properties of the tissue including its inherent variability. A PCA was performed separately for each class of the training set in the interval 950 to 1800 cm^{-1}. The SIMCA model consisted of 10 PCs per class. After performing a PCA from test data, they were projected into the 10-dimensional data space of each model and assigned to the class with the closest distance. The class assignments were color coded for display.

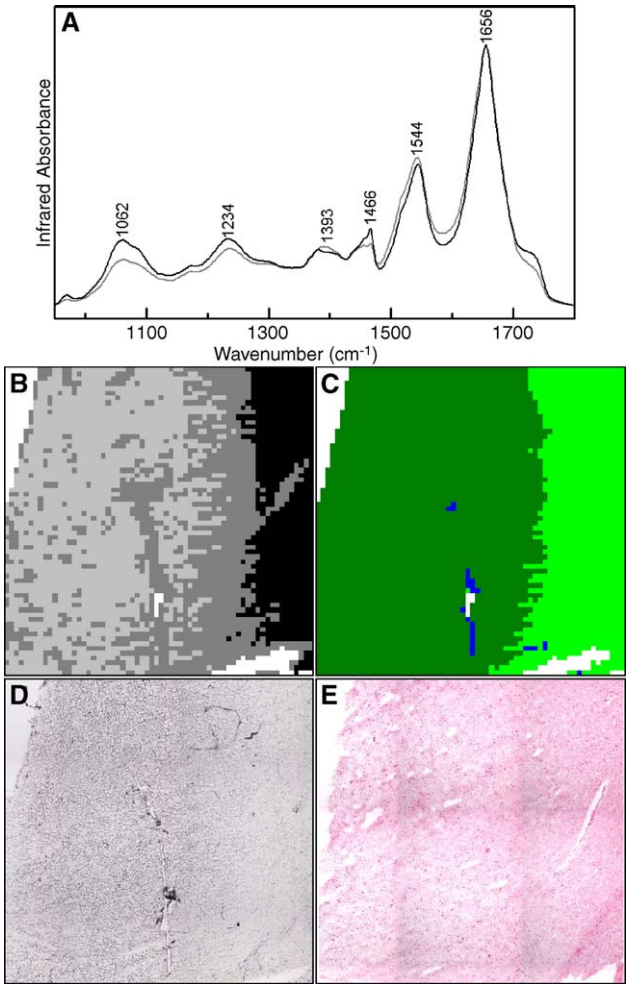

Fig. 1. IR spectroscopic imaging of normal brain tissue N1. Averaged IR spectra from 950 to 1800 cm^{-1} (A) of the black and light gray clusters (B). Color-coded classification of IR spectra by the SIMCA model (C). Gray matter (dark green), white matter (light green), LC (blue). Photomicrographs of the unstained cryosection (D) and the consecutive H&E stained tissue section (E) have been included for comparison.

3. Results

The first part of this chapter describes the spectral properties of the tissue specimens, the selection of training data for developing of the SIMCA model, and histological evaluation of H&E-stained consecutive tissue sections of the training specimens. The second part presents the SIMCA model and the reclassification results of the training data. In the third part, the SIMCA model is applied to classify independent test data.

3.1. Normal brain tissue

Fig. 1A compares cluster-averaged IR spectra of white matter and of gray matter of normal brain tissue N1. The numbering refers to Table 1 which also indicates the sex and age of the patients if known. These spectra represent the light gray and black cluster of the cluster analysis in Fig. 1B which segmented the spectra of the IR image in three clusters. The spectral properties of the dark gray cluster lie between the presented IR spectra (Fig. 1A). It is required to use two classes of normal brain tissue for training of a classification model as gray matter and white matter differ significantly in their chemical composition. Increased band intensities at 1062, 1234, 1466 and 1735 cm^{-1} indicate a higher lipid content in white matter after normalization to the protein amide I band at 1656 cm^{-1} [30]. There is a tendency that bands near 1393 and 1544 cm^{-1} which are associated to proteins show higher relative intensities in gray matter. Here, the light gray and black cluster were selected as training data for white and gray matter in the SIMCA algorithm. The microscopic image of the unstained tissue section from which the IR image was recorded (Fig. 1D) and that of the H&E stained consecutive tissue section (Fig. 1E) confirm the transition from gray matter on the left to white matter on the right.

3.2. Brain metastasis of renal cell carcinoma (RCC)

The overlay of the IR spectra in Fig. 2A representing the dark and light gray clusters of RCC1 reveals spectral differences at

Table 1
Listed values show the assignments of spectra in 20 IR images to the tissue type obtained by the SIMCA model (in %)

Specimens	Sex	Age	Gray	White	RCC	LC	CC	MC
N1 (T)			**70.1**	29.3	0.1	0.5	0	0
N2			**17.1**	77.0	0.2	3.8	1.9	0
N3			**4.6**	94.0	0.1	1.3	0	0
RCC1 (T)	M	73	0	0	**100**	0	0	0
RCC2	M	39	0	0	**99.0**	0.9	0.1	0
RCC3	M	62	0	0	**95.7**	1.8	2.0	0.5
RCC4	M	63	3.6	0.3	**68.1**	15.5	4.5	8.0
RCC5	W	86	0	0	**100**	0	0	0
RCC6	W	55	3.2	0	**39.8**	19.4	33.8	3.8
RCC7	W	66	3.7	3.2	**65.7**	10.9	15.5	1.0
LC1 (T)	M	68	0	0	0	**99.8**	0	0.2
LC2	M	61	0	0	1.9	**90.8**	6.4	0.9
LC3	W	58	0	0	0.6	**82.8**	13.3	3.3
CC1 (T)	W	69	0	0	0	2.1	**97.8**	0.1
CC2	M	67	0	0	0.4	3.4	**90.2**	6.0
CC3	M	61	2.9	0	1.9	13.3	**72.7**	9.2
CC4	M	72	0	0	0.2	48.1	**51.6**	0.1
BC1 (T)	W	40	0	0	0	6.3	2.3	**91.4**
BC2	W	48	0	0	0.3	15.2	4.1	**80.4**
BC3	W	57	0	0	0.1	15.2	34.9	**49.8**

Bold entries represent correct classifications. Reclassification of training data is indicated by (T). Sex and age of the patients are indicated if available. Abbreviations: brain metastasis of renal cell carcinoma (RCC), of lung cancer (LC), of colorectal cancer (CC) and of breast cancer (BC).

1028, 1080 and 1154 cm^{-1}. The appearance of these bands together with reduced intensities of lipid related bands constitute the main difference compared to the IR spectra of normal brain tissue (Fig. 1A). Other notable differences are evident as band shifts to 1238, 1387, 1453, and 1541 cm^{-1}. The signatures in Fig. 2A point to accumulation of glycogen in RCC which is a typical property of renal cells. An IR spectrum of pure glycogen is included as reference. Consequently, spectral contributions of glycogen serve as a molecular marker for RCC. As IR spectra throughout the IR image of RCC1 contain glycogen bands of variable intensities, both clusters were used as training data for RCC in the SIMCA algorithm. Inhomogeneities in the unstained tissue section (Fig. 2D) and in the H&E stained tissue section (Fig. 2E) are consistent with variations in the glycogen content.

3.3. Brain metastasis of breast cancer (BC)

Three IR spectra (Fig. 3A) represent the light gray, dark gray and black clusters of BC1 (Fig. 3B). Compared to the previous IR spectra (Figs. 1A and 2A), the IR spectra differ mainly in the spectral ranges 1000 to 1150, 1400 to 1480 and 1700 to 1750 cm^{-1} due to the lack of glycogen and the reduced lipid content. Besides these tissue-specific differences, intra-sample variances are present. The increase of background in the dark gray and black clusters is accompanied by downshift of the amide II band to 1537 cm^{-1} and broadening of the amide I band. The black cluster corresponds to locations near tissue borders and holes. Therefore, it is assumed that this phenomenon is correlated with scattering and thickness properties of the cryosection and not with chemical or molecular properties as demonstrated for RCC and normal brain tissue before. A similar observation was made for IR spectra of cracks in spleen sections [16]. While the IR spectra of both gray clusters were used for training, the IR spectra of the black cluster were excluded. The cluster segmentation also correlates with properties of the unstained (Fig. 3D) and H&E stained tissue section (Fig. 3E). Dark gray and black clusters coincide with lighter and darker structures which point to necrotic tissue.

3.4. Brain metastasis of lung cancer (LC)

Specimen LC1 (Fig. 4B) could be completely imaged as its dimension was smaller than 4×4 mm^2. Two clusters constitute the main IR spectral signatures of LC1. Analogous to Fig. 3A, the averaged IR spectrum from the cluster near the tissue borders (black cluster) is characterized by a higher background, downshift of the amide II band and

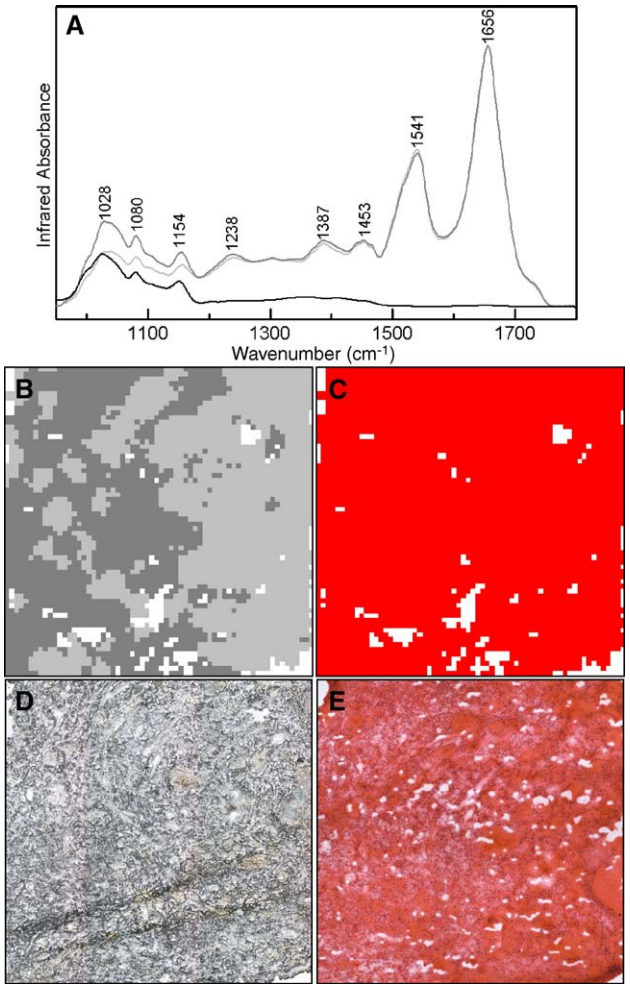

Fig. 2. IR spectroscopic imaging of brain metastasis of renal cell carcinoma RCC1. Averaged IR spectra from 950 to 1800 cm^{-1} (A) of the dark and light gray clusters (B). Color-coded classification of IR spectra by the SIMCA model (C). All spectra were assigned to RCC (red). Photomicrographs of the unstained cryosection (D) and the consecutive H&E stained tissue section (E) have been included for comparison.

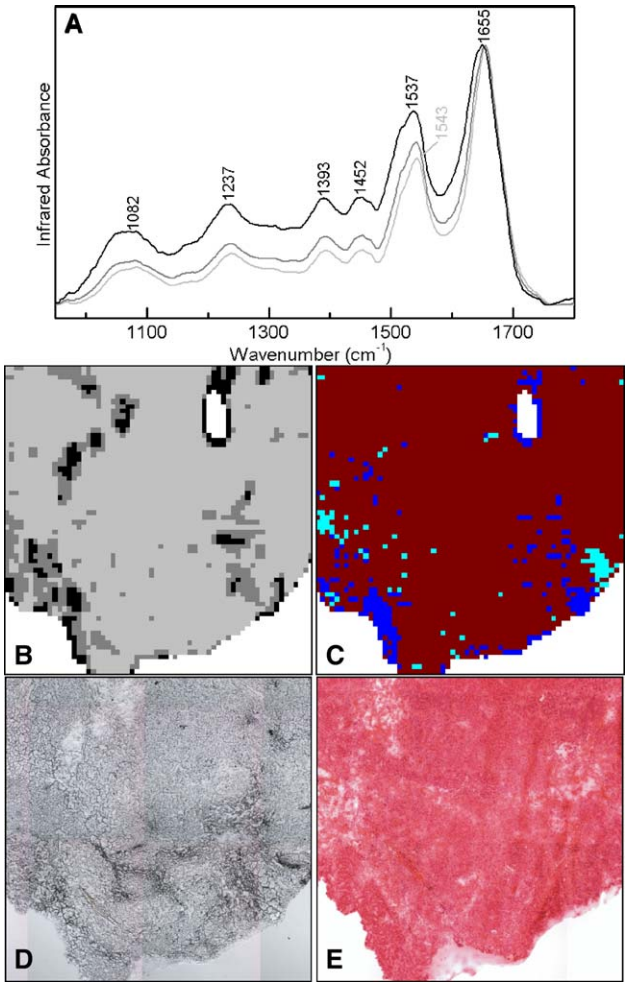

Fig. 3. IR spectroscopic imaging of brain metastasis of breast cancer BC1. Averaged IR spectra from 950 to 1800 cm^{-1} (A) of the light gray, dark gray and black clusters (B). Color-coded classification of IR spectra by the SIMCA model (C). BC (brown), CC (cyan), LC (blue). Photomicrographs of the unstained cryosection (D) and the consecutive H&E stained tissue section (E) have been included for comparison.

broadening of the amide I band (Fig. 4A). The majority of IR spectra in the inner regions of the cryosection (gray cluster) are characterized by a lower position of the amide I band and a higher intensity ratio of amide II to amide I than the IR spectra of the other tissue types in this study. Consistent with the training data above, only the IR spectra of the gray cluster are considered as training data. The contrast of light and dark structures is more pronounced in unstained tissue section of LC1 (Fig. 4D) than in BC1 (Fig. 3D). According to the H&E stained tissue section (Fig. 4E), the light region is assigned to vital tumor and the dark region to necrotic tissue. Extended necrosis is not unusual in brain metastases.

3.5. Brain metastasis of colorectal cancer (CC)

Two clusters (Fig. 5B) with distinct spectral signatures (Fig. 5A) are identified in the brain metastasis of colorectal cancer CC1. Following the previous selections, the gray cluster with the low background IR spectra was selected as training data. Although generally similar, the black cluster with the higher background IR spectra was not selected for training of the supervised classification model. The IR spectra differ from the highly similar IR spectra of BC1 (Fig. 3A) by small changes in wavenumber positions. The classification results will show if these tissue-specific variances are large enough to identify the primary tumor of this brain metastasis. Similarly as observed before, the unstained tissue section consists of light and dark regions (Fig. 5D) which correlate with the cluster segmentation. More details are resolved in the H&E stained tissue section (Fig. 5E) which will be discussed below in a photomicrograph at higher magnification. Comparison between the unstained and the top part of the H&E tissue section is impaired because some tissue regions are missing.

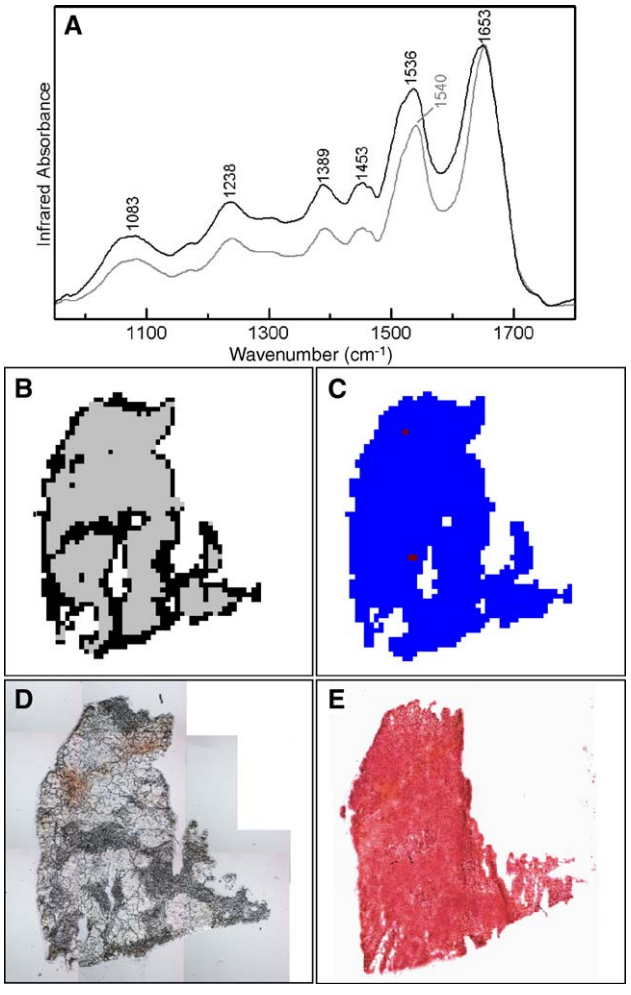

Fig. 4. IR spectroscopic imaging of brain metastasis of lung cancer LC1. Averaged IR spectra from 950 to 1800 cm^{-1} (A) of the light gray and black clusters (B). Color-coded classification of IR spectra by the SIMCA model (C). LC (blue), BC (brown). Photomicrographs of the unstained cryosection (D) and the consecutive H&E stained tissue section (E) have been included for comparison.

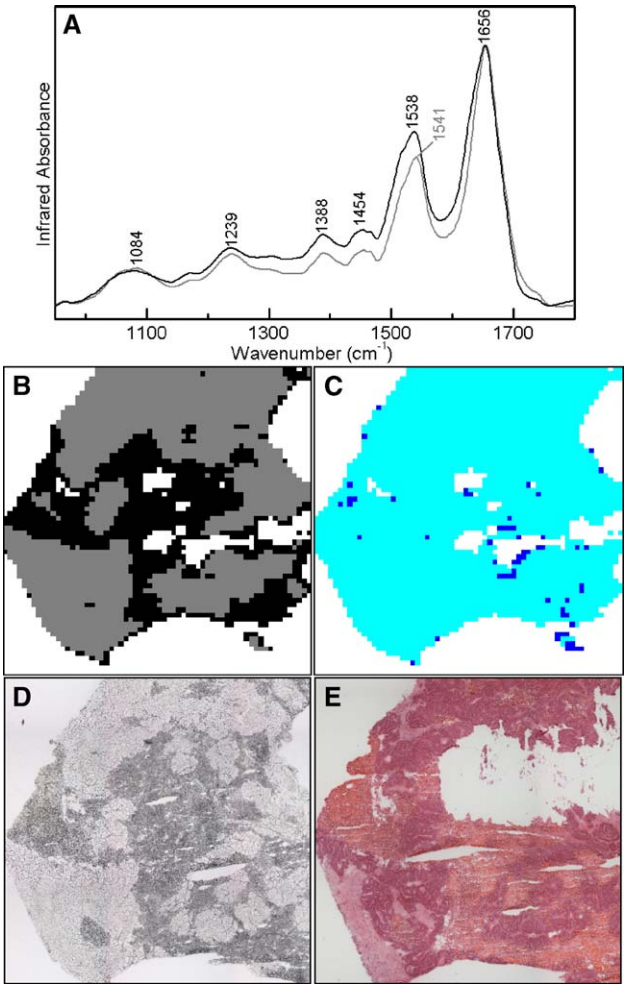

Fig. 5. IR spectroscopic imaging of brain metastasis of colorectal cancer CC1. Averaged IR spectra from 950 to 1800 cm^{-1} (A) of gray and black clusters (B). Color-coded classification of IR spectra by the SIMCA model (C). CC (cyan), LC (blue). Photomicrographs of the unstained cryosection (D) and the consecutive H&E stained tissue section (E) have been included for comparison.

3.6. Neuropathologic evaluation of brain metastases

Fig. 6 shows details of the H&E stained tissue sections of Figs. 2E, 3E, 4E and 5E at higher magnification representing each group of tumors. Their descriptions are given as examples for the respective groups. RCC1 (Fig. 6A) is a poorly differentiated clear cell renal carcinoma with high glycogen content and either solid growth or leaf-like organization of tumor cells along a collageneous matrix. BC1 (Fig. 6B) is a poorly differentiated adenocarcinoma with small glandular-like structures, focal solid growth and “Indian file” patterns. There is also a partial acidic mucopolysaccharide matrix in the center of glandular-like structures and focally surrounding islets growing in minimally altered brain tissue. The observed tumor strongly resembles the structures of the primary tumor with low expression of estrogen and progesterone receptors. LC1 (Fig. 6C) is a large cell anaplastic carcinoma of the lung with histologically confirmed primary lung carcinoma. The tumor cells show exceptionally large cell bodies with bizarre nuclei and a solid growth pattern interspersed with small necrotic areas. CC1 (Fig. 6D) is a moderately differentiated colon carcinoma with tubulopapillary growth pattern reminiscent of intestinal epithelium and polypoid adenomas of the colon. However, nuclear polarization is lost and there are considerable alterations of nuclear shape and chromatin density as well as an increased number of mitosis. In addition, partial mucin production is observed.

3.7. SIMCA classification model

A SIMCA classification model was trained to distinguish the following six classes: gray matter and white matter of normal brain tissue, brain metastases of renal cell carcinoma (RCC), of lung cancer (LC), of colorectal cancer (CC) and of breast cancer (BC). Clusters of IR spectra were selected for each class which represent the inherent inter-sample variances as shown above. The SIMCA model was developed with 10 PCs in the interval 950 to 1800 cm^{-1}. If the PCs were increased to 12 or decreased to 8, the accuracy of the classification would not change significantly. The results of the reclassification of the full data sets N1, RCC1, BC1, LC1 and CC1 are displayed graphically in Figs. 1C, 2C, 3C, 4C and 5C, respectively (i.e., including IR spectra which were not used in the training set). The numerical values are summarized in Table 1 with reclassifications of training images indicated by (T). The assignments for 15 independent IR images have been included as test data in Table 1.

Normal brain tissue N1 (Fig. 1C) was assigned to gray matter (70.1%) and to white matter (29.3%). N2 and N3 are dominated by white matter (77% and 94%, respectively) with smaller fractions of gray matter (17.1% and 4.6%, respectively). False positive classification are generally small (0.6% for N1, 5.9% for N2, 1.4% for N3). They occur mainly near fissures or hemorrhages. Tissue with hemorrhage shows an increased amide II band [31] which has also been identified as a marker for LC. Therefore, it seems plausible that IR spectra near hemorrhage were incorrectly assigned to LC in each specimen.

RCC1 (Fig. 2C) was assigned to RCC with perfect 100% accuracy. Similar correct classifications were obtained for RCC2 (99%) RCC3 (95.7%) and RCC5 (100%). The classification rates are smaller for RCC4 (68.1%), RCC6 (39.8) and RCC7 (65.7) because the glycogen content as the main IR spectroscopic marker for RCC is reduced and these specimens also contain significant fractions of normal tissue.

RCC7 deserves special attention as the primary tumor could not be identified unambiguously by standard methods. It was tentatively specified as ”suspicious to renal cell carcinoma or lung cancer”. The SIMCA classification of the IR image segmented the specimen into three main parts (Fig. 7): normal brain tissue consisting of white (3.2%) and gray matter (3.7%), RCC (65.7%) and others (CC 15.5%, LC 10.9%, MC 1%) at the transition between normal tissue and RCC. The detection of glycogen and the majority classification rate for RCC strongly supports that the primary tumor of the brain metastasis is renal cell carcinoma. The intermediate zone between RCC and normal brain tissue seems to be a mixture of necrosis, cerebral

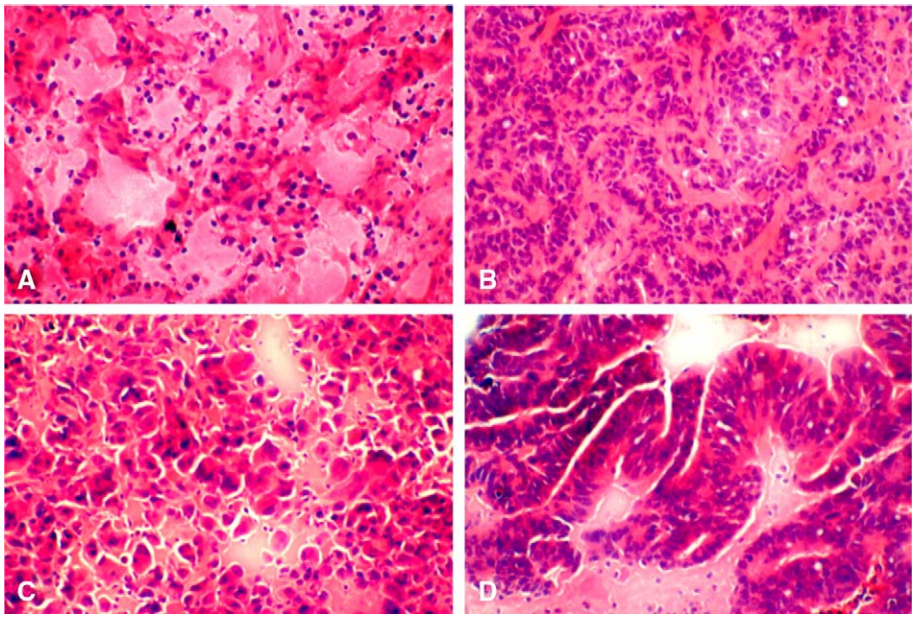

Fig. 6. Photomicrographs highlighting details in intracerebral metastases of different carcinoma types after H&E staining of cryosections. Magnification: original ×20. (A) Clear cell renal carcinoma (RCC1, detail of Fig. 2E). (B) Poorly differentiated breast cancer, mucinous adenocarcinoma (BC1, detail of Fig. 3E). (C) Large cell anaplastic carcinoma of the lung (LC1, detail of Fig. 4E). (D) Moderately differentiated partially mucinous tubulopapillary carcinoma of the colon (CC1, detail of Fig. 5E).

gliosis and tumor tissue from RCC without accumulation of glycogen. As such tissue classes have not been included in the classification model yet, this might explain the incorrect classifications to MC, CC and LC.

99.8% of spectra from IR image LC1 were assigned to LC. Inspection of Fig. 4C reveals that all IR spectra from the tissue border which constitute the black cluster in Fig. 4B were correctly classified although they did not belong to the training data. The SIMCA algorithm seems to tolerate the increase of the background, and the spectral pattern appears to be conserved upon the transition of tumor tissue to necrotic tissue. The simple linear baseline from a minimum near 950 cm^{-1} and a minimum near 1800 cm^{-1} which was subtracted during the course of preprocessing turned out to be sufficient for this approach. The correct classifications decrease slightly to 90.8% for LC2 and 82.8% for LC3. No false negative classification to normal brain tissue is observed.

Although just the gray cluster with 64% of total spectra from the IR image of CC1 was selected as training data, almost all (97.8%) spectra of CC1 were correctly assigned to CC (Fig. 5C). As noted for LC1, the different background slopes due to necrosis do not significantly influence the classification accuracy. Also CC2, CC3 and CC4 were assigned to CC with accuracies of 90.2%, 72.7% and 51.6%. The decrease of correct classifications is connected with increased overlap to LC ranging from 2.1% to 48.1%. Conversely, most incorrectly assigned IR spectra for LC2 and LC3 belong to CC (6.4% and 13.3%, respectively).

The reclassification rate for BC1 is just 91.4% which is lower than for the other training data. However, the classification rates for BC2 (80.4%) and BC3 (49.8%) are high enough to clearly

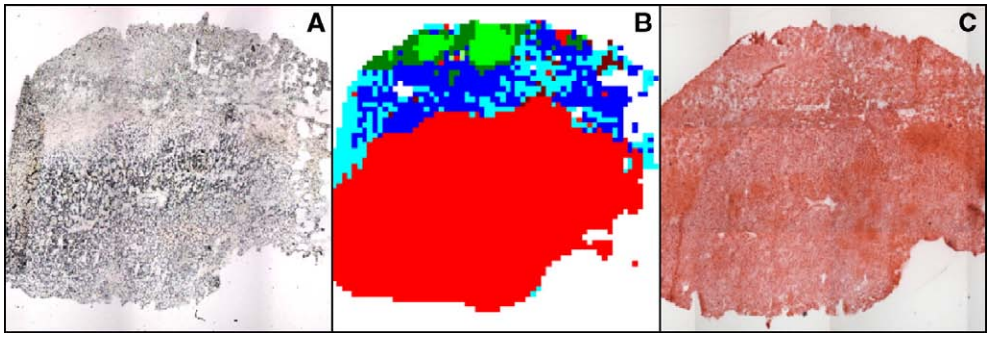

Fig. 7. Unstained cryosection (A), color-coded classification of IR spectroscopic image by the SIMCA model (B) and hematoxylin and eosin-stained tissue section of brain metastasis of renal cell carcinoma RCC7 (C). Color code: white matter (light green), gray matter (dark green), brain metastasis of renal cell carcinoma (red) and intermediate region (cyan, brown, blue).

identify the primary tumor. Misclassifications occurred to LC and CC, but not to RCC or normal tissue. This observation is consistent with the classification results for LC and CC. Obviously, the SIMCA model is better defined for gray matter, white matter and RCC due to well pronounced marker bands. However, the spectral differences between BC, LC and CC are less intense so more overlap between classes was observed.

4. Discussion

A number of studies already reported that IR spectroscopy based methods offer the potential to complement histopathology and immunohistochemistry as an independent method of tissue evaluation. The current proof-of-principle study demonstrates that IR spectroscopic imaging succeeded by multivariate classification algorithms can identify the primary tumor of brain metastases. If the criterion is defined that the majority of spectral classification determines the tissue class, then all normal brain tissue samples and all brain metastases were correctly classified. The results agree qualitatively well with the LDA classification of the IR spectroscopic images which have been reported in an accompanying manuscript [21]. Therefore, we conclude that both independent classification models are able to extract the diagnostic information in the IR spectra and to utilize it for tissue identification. Quantitatively the numbers differ because the training data and the classification algorithms are different. In 12 out of 17 brain metastases, the percentage of IR spectra which were assigned to the correct primary tumor were higher for the SIMCA classification than for the LDA classification. This comparison suggests that the SIMCA model is able to describe the diagnostic spectral features more comprehensive. The conclusions imply to test how accurate other pattern recognition techniques such as support vector machines or artificial neural networks can assign these IR spectroscopic images.

So far, the classification model is restricted as an unknown brain metastasis can only be identified if the primary tumor originated from the sites and varieties of the training set. As just one specimen per class trained the classification model, not all tissue variants could be considered yet. Other tumor varieties might explain the higher rates of misclassification for BC3 and CC4. In general, the fraction of IR spectra that are incorrectly assigned represents the measurable error, e.g., due to impurities of other cell types, physical effects, model deficiencies or an unknown brain metastasis. To improve the latter point, additional frequent primary tumors of brain metastasis (e.g., malignant melanomas) and other varieties of primary tumors (e.g., metastases from breast cancer of glandular vs. lobular adenocarcinoma) have to be included to the SIMCA model. Furthermore, the algorithm can be modified to implement a class called “not assigned”. Refinement of training data might also improve the accuracy of the SIMCA model. Due to a limited number of specimens in the current study, IR images from just one patient per class trained the supervised model. With a larger number of specimens per tumor type, including IR images from several patients to the training set can consider the inherent biological variability better.

This study applied IR spectroscopic imaging in the macroscopic mode. The advantage is that 4×4 mm^2 sample areas can be probed in just few minutes so the heterogeneity of the specimens can be assessed with a lateral resolution of 63 μm which is in most cases sufficient to identify the molecular information of the main tissue classes. The disadvantage is that morphological details which were highlighted in Fig. 6 cannot be resolved. In order to increase the lateral resolution down to the single cell level, the FTIR imaging spectrometer has to be coupled to a microscope [16,17,20]. However, the IR microspectroscopic imaging mode with 15× magnification reduces the field of view to 266×266 μm^2. In this case, assessment of a 4×4 mm^2 area requires to record $15^2=225$ IR images which dramatically increases the total acquisition time and the amount of data to be processed. Therefore, we conclude that IR macrospectroscopic imaging is a better candidate to introduce IR spectroscopy as a rapid, high throughput analytical tool to characterize tissue sections.

The general concept behind this study was to measure spectral profiles from tissue types, to identify relevant spectral markers, to organize them into a predication algorithm and to validate the model by classification of independent specimens. This concept can be transferred to other biological applications making IR spectroscopic imaging a versatile analytical tool. As tissue sections are prepared during standard protocols and remain unperturbed during spectral data recording, integration of IR spectroscopy with conventional laboratory techniques for subsequent analyses is possible.

Acknowledgment

This work is supported by the Volkswagen Foundation within in project “Molecular Endospectroscopy” of the program “Junior research groups at universities”.

References

[1] K.S. Polyzoidis, G. Miliaras, N. Pavlidis, Brain metastasis of unknown primary: a diagnostic and therapeutic dilemma, Cancer Treat. Rev. 31 (2005) 247–255.
[2] Schackert, A. Steinmetz, U. Meier, S.B. Sobottka, Surgical management of single and multiple brain metastases: results of a retrospective study, Onkologie 24 (2001) 246–255.
[3] R.K. Dukor, Vibrational spectroscopy in the detection of cancer, in: J.M. Chalmers, P.R. Griffiths (Eds.), Handbook of Vibrational Spectroscopy, John Wiley and Sons, New York, 2002, pp. 3335–3361.
[4] M. Diem, M. Romeo, S. Boydston-White, M. Miljkovic, C. Matthäus, A decade of vibrational micro-spectroscopy of human cells and tissue, Analyst 129 (2004) 880–885.
[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] P. Lasch, D. Naumann, FTIR microspectroscopic imaging of human carcinoma in thin sections based on pattern recognition techniques, Cell. Mol. Biol. 44 (1998) 189–202.
[7] T.C. Martin, J. Moecks, A. Belooussov, S. Cawthraw, B. Dolenko, M. Eiden, J. Von Frese, W. Kohler, J. Schmitt, R. Somorjai, T. Udelhoven, S. Verzakov, W. Petrich, Classification of signatures of bovine spongiform encephalopathy in serum using infrared spectroscopy, Analyst 129 (2004) 897–901.

[8] E.N. Lewis, P.J. Treado, R.C. Reeder, G.M. Story, A.E. Dowrey, C. Marcott, I.R. Levin, Fourier transform spectroscopic imaging using an infrared focal plane array detector, Anal. Chem. 67 (1995) 3377–3381.
[9] L.H. Kidder, V.F. Kalasinsky, J.L. Luke, I.W. Levin, E.N. Lewis, Visualization of silicone gel in human breast tissue using new infrared imaging spectroscopy, Nat. Med. 3 (1997) 235–237.
[10] H. Fabian, P. Lasch, M. Boese, W. Haensch, Mid-IR microspectroscopic imaging of breast tumor tissue sections, Biopolymers 67 (2002) 354–357.
[11] K. Potter, L.H. Kidder, I.W. Levin, E.N. Lewis, R.G.S. Spencer, Imaging of collagen and proteoglycan in cartilage sections using Fourier transform infrared spectral imaging, Arthritis Rheum. 44 (2001) 846–855.
[12] N.P. Camacho, P. West, P.A. Torzilli, R. Mendelsohn, FTIR microscopic imaging of collagen and proteoglycan in bovine cartilage, Biopolymers 62 (2001) 1–8.
[13] H. Ou-Yang, E.P. Paschalis, W.E. Mayo, A.L. Boskey, R. Mendelsohn, Infrared microspectroscopic imaging of bone: spatial distribution of CO_3^{2-}, J. Bone Miner. Res. 16 (2001) 893–900.
[14] 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.
[15] D.C. Fernandez, R. Bhargava, S.M. Hewitt, I.R. Levin, Infrared spectroscopic imaging for histopathologic recognition, Nat. Biotech. 23 (2005) 469–474.
[16] C. Krafft, R. Salzer, G. Soff, M. Meyer-Hermann, Identification of B- and T-cells in human spleen samples, Cytometry A 64A (2005) 53–61.
[17] C. Beleites, G. Steiner, M.G. Sowa, R. Baumgartner, S. Sobottka, G. Schackert, R. Salzer, Classification of human gliomas by infrared imaging spectroscopy and chemometric image processing, Vib. Spectrosc. 38 (2005) 143–149.
[18] D.S. Lester, L.H. Kidder, I.W. Levin, E.N. Lewis, Infrared microspectroscopic imaging of the cerebellum of normal and cytarabine treated rats, Cell Mol. Biol. 44 (1998) 29–38.
[19] R. Mendelsohn, H.C. Chen, M.E. Rerek, D.J. Moore, Infrared microspectroscopic imaging maps the spatial distribution of exogenous molecules in skin, J. Biomed. Opt. 8 (2003) 185–190.
[20] W. Steller, J. Einenkel, L.-C. Horn, U.-D. Braumann, H. Binder, R. Salzer, C. Krafft, Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging, Anal. Bioanal. Chem. 384 (2006) 145–154.
[21] C. Krafft, L. Shapoval, S.B. Sobottka, G. Schackert, R. Salzer, Identification of primary tumors of brain metastases by IR spectroscopic imaging and linear discriminant analysis, Technol. Cancer Res. Treat 5 (2006) 291–298.
[22] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chem. Intell. Lab. Syst. 2 (1987) 37–52.
[23] C.P. Schultz, The potential role of Fourier transform infrared spectroscopy and imaging in cancer diagnosis incorporating complex mathematical methods, Technol. Cancer Res. Treat. 1 (2002) 95–104.
[24] N. Uzunbajakava, A. Lenferink, Y. Kraan, E. Volokhina, G. Vrensen, J. Greve, C. Otto, Nonresonant confocal Raman imaging of DNA and protein distribution in apoptotic cells, Biophys. J. 84 (2002) 3968–3981.
[25] S. Koljenovic, L.P. Choo-Smith, T.C. Bakker Schut, J.M. Kros, H.J. van den Berge, G.J. Puppels, Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy, Lab. Invest. 82 (2002) 1265–1277.
[26] S. Wold, Pattern recognition by means of disjoint principal components model, Pattern Recogn. 8 (1976) 127–139.
[27] M. Foot, M. Mulholland, Classification of chondroitin sulfate A, chondroitin sulfate C, glucosami hydrochloride and glucosamine 6 sulfate using chemometric techniques, J. Pharm. Biomed. Anal. 38 (2005) 397–407.
[28] H.H. Nieuwoudt, B.A. Prior, I.S. Pretorius, M. Manley, F.F. Bauer, Principal component analysis applied to Fourier transform infrared spectroscopy for the design of calibration sets for glycerol prediction models in wine and for the detection and classification of outlier samples, J. Agric. Food Chem. 52 (2004) 3726–3735.
[29] A. Oust, T. Moretro, C. Kirschner, J.A. Narvhus, A. Kohler, FT-IR spectroscopy for identification of closely related lactobacilli, J. Microbiol. Methods 59 (2004) 149–162.
[30] C. Krafft, S.B. Sobottka, G. Schackert, R. Salzer, Analysis of human brain tissue, brain tumors and tumor cells by infrared spectroscopic mapping, Analyst 129 (2004) 921–925.
[31] C. Krafft, S.B. Sobottka, G. Schackert, R. Salzer, Raman and infrared spectroscopic mapping of human primary intracranial tumors: a comparative study, J. Raman Spectrosc. 37 (2006) 367–375.