This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the author guidelines.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the ethical guidelines, outlined in our author and reviewer resource centre, still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

This article can be cited before page numbers have been issued, to do this please use: B. Brozek-Pluska, M. Kopec and H. Abramczyk, Anal. Methods, 2016, DOI: 10.1039/C6AY02559E.
Development of a new diagnostic Raman method for monitoring epigenetic modifications in the cancer cells of human breast tissue

Beata Brozek-Pluska¹, Monika Kopeć¹, Halina Abramczyk*¹

¹Lodz University of Technology, Institute of Applied Radiation Chemistry, Laboratory of Laser Molecular Spectroscopy, Wroblewskiego 15, 93-590 Lodz, Poland

*To whom correspondence should be addressed: Halina Abramczyk, Lodz University of Technology, Institute of Applied Radiation Chemistry, Laboratory of Laser Molecular Spectroscopy, Wroblewskiego 15, 93-590 Lodz, Poland, telephone number: +48 42 6313188, fax: +48 42 6840043, e-mail: abramczy@mitr.p.lodz.pl.

ABSTRACT

Cancer diagnosis requires better screening of early stages of pathology and monitoring patient responses to treatment. Current technologies in clinical sector are expensive, sophisticated and time consuming. This paper develops a novel Raman based alternative for currently existing epigenetics research approaches. The proposed Raman approach can ‘upgrade’ cancer epigenetics tests and answer many questions by monitoring the biochemistry of cancer cells. We will show that Raman spectroscopy and Raman imaging can detect the relative amounts of acetylated and methylated lysine by monitoring the vibrations of the acetyl and methyl chemical functional groups. In comparison to existing tests and assays used to monitor molecular processes of acetylation and methylation of all proteins, Raman-based methods have potential to be a powerful alternative for conventional methods of cell biology, because they allow non-invasively detect cellular acetylation and methylation processes that are not limited to only those events that are sensitive to a specific antibody. Vibrational Raman signatures of acetylation and methylation processes in epithelial cells of human breast
tissue (ductal and lobular carcinoma) have been used to identify and discriminate structures in normal and cancerous tissues. Our results demonstrate that the stretching vibration of the acetyl group observed at around 2938 -2942 cm\(^{-1}\) and the methyl group at 2970 cm\(^{-1}\) by Raman spectroscopy is useful in monitoring these epigenetic molecular processes in cancer cells. The Raman vibrational marker is markedly shifted blue from 2905 cm\(^{-1}\) (non-acetylated C-H vibrations) for normal cells to 2942 cm\(^{-1}\) (vibrations of the acetylated functional group) in cancer cells. The sensitivity and specificity obtained directly from PLSDA and cross validation for a chosen model gives a sensitivity and specificity of 86.1% and 91.3% for calibration and 85.3% and 91.3% for cross-validation, respectively. The results presented in the paper provide strong evidence that the global acetylation level of histone and non-histone proteins increases in human breast cancer cells.

**Key words:** epigenetic modifications, cancer, Raman spectroscopy, chemometric analysis

**Introduction**

In the paper we will concentrate on development of new diagnostic method for monitoring epigenetic modifications in cancer using Raman spectroscopy and Raman imaging. We have shown that the main metabolic channels that are absolutely required for tumorigenic transformation are the following: (a) rapid ATP generation to maintain energy status; (b) increased biosynthesis of macromolecules; and (c) maintenance of appropriate cellular redox status \(^1\). The link between cancer and altered metabolism is usually described through the first channel comprising glycolysis, also known as the Warburg effect, where ATP generation through oxidative phosphorylation is replaced by ATP generation through glycolysis, even under normal oxygen concentrations. We have shown that there is more and more evidence
that metabolic alterations in tumors extend beyond the Warburg effect and pathways (b) and (c) are equally important. These paths are associated with the production of fatty acids activated via multiple lipogenic enzymes affected at all levels of regulation, including transcription, translation, protein stabilization and protein phosphorylation. Many of these processes are associated with epigenetic modifications in cancer. In recent years it has come to the attention of researchers that many types of cancer are not caused by a modification in the nucleotide sequence, but they are largely due to epigenetic alterations, such as DNA and histone methylation that modify the transcriptional potential of a cell and post translational modifications, such as acetylation, methylation, and phosphorylation. Now is just the beginning of our understanding of epigenetic factors involved in development of cancer diseases, but it has become evident that cancer cells have many epigenetic differences compared to normal cells in the same patient. The list of all possible epigenetic modifications is not yet complete, but it is obvious that in the future the manipulation of the epigenetic landscape will become a key element of cancer therapy. Over the past few years there has been an explosion in studies of epigenetics in breast cancer using various approaches such as mRNA purification, DNA purification, detection of histone modifications by mass spectrometry. We will demonstrate that Raman approach has many advantages over these conventional methods that have been used so far.

Fig. 1 Mechanism of acetylation and deacetylation maintained by antagonistic activity of two groups of enzymes - histone acetyltransferases (HAT) and deacetylases (HDAC).

The impact of acetylation on transcription and post translational modifications of proteins produced in cancer development has been widely reviewed. Briefly, mechanism of acetylation and deacetylation presented in Fig. 1 is maintained by opposing activities of histone acetyltransferases (HAT) and deacetylases (HDAC).
acetylation transfers an acetyl group from acetyl CoA to form ε- N-acetyl lysine. The acetylation mechanism induced by HAT is particularly important in post translational alterations of histone proteins, because the positive charge of the histone tail can be changed to neutral by acetylation. Ultimately, such a modification leads to weaker binding via electrostatic forces between the histone proteins and DNA of the nucleosome components that results in structure alteration of chromatin (less compact euchromatin) and affects the expression of genes making them more active. The opposite effect is due to the HDAC, which removes the acetyl groups from the lysine residues, making the DNA more tightly wrapped around the histone cores (and resulting in a repression of gene activity known as a gene silencing). These opposite effects of HAT/HDAC on chromatin structure are presented in Fig. 2.

Fig 2. Histone acetylation/deacetylation by HAT and HDAC enzymes alters chromatin structure by decreasing/increasing accessibility of exposed sites on DNA to silencing/activating gene transcription.

Histone modifications are considered to be one of the most important mechanisms contributing to cancer development. The opposing actions of histone acetyltransferases and histone deacetylases that regulate the expression and repression of genes have a potential to be used as exciting new agents for the treatment of cancer. There are many anticancer drugs in development targeting histone acetylation mechanisms in cancerous cells. HDAC inhibitors are proven to be effective anticancer drugs that inhibit tumor growth or reactivate tumor suppressor genes.

However, at the moment we do not have enough insights into mechanisms of acetylation and we do not yet fully understand their dual role in controlling cancer development. In the
simplest view, HAT and HDAC inhibitors can both initiate and amplify tumorigenesis targeting oncogenes or inhibit tumor growth and remove silencing in tumor suppressors. Reduced HDAC or increased HAT activities via specific inhibitors or suppressors lead to an increase in histone acetylation; increased HDAC or decreased HAT cause decreased histone acetylation that compacts chromatin and precludes access by the transcription mechanisms, resulting in transcriptional repression \(^8\).

The opposing actions of HAT and HDAC indicate that the new anticancer drugs must have selectivity and remarkable tumor specificity, but specific modifications on the tumorigenic process are still largely unclear \(^9,11,14,16,17\).

Within the paper we will assess the diagnostic potential of the Raman spectroscopy and Raman imaging for monitoring acetylation and methylation processes in cancer cells of human breast tissues. We will show that Raman spectroscopy and Raman imaging can detect the relative amounts of acetylated and methylated lysine by monitoring the vibrations of the acetyl and methyl chemical functional groups. In comparison to existing tests and assays, Raman-based methods represent a powerful alternative that allows non-invasively detect cellular acetylation and methylation processes that are not limited to only those events that are sensitive to a specific antibody.

In this paper we will study samples of normal tissue from the safety margin (known to be cancer negative from the ‘gold standard’ technique of histopathology) and cancerous tissues from the tumor mass of the human breast: invasive ductal cancer (IDC) and invasive lobular cancer (ILC). IDC and ILC are the most common forms of breast cancer, representing over 80 percent of all breast cancer diagnoses. IDC and ILC are cancers that begin growing in the duct (IDC) or the lobule (ILC) and invade the extracellular matrix, dominated by the fatty tissue of the breast outside of the duct or the lobule.
Vibrational signatures of acetylation and methylation processes will be used to identify and discriminate structures in normal and cancerous tissues. Since Raman method is label free, and does not involve any staining or antibody attachment for detecting acetylated or methylated proteins, it offers straightforward sample handling over complex assays that must take into sensitivity and specificity to antibody. Just only a few Raman-based papers on post translational modifications (glycosylation, phosphorylation, acetylation, trimethylation, ubiquitination) have already been published \(^1,5,6,18-20\).

According to our best knowledge this is a first attempt to introduce alternative Raman-based assays to monitor epigenetic modifications in human cancer.

**Experimental**

Lysine (L5501), acetylated lysine (A4021), and methylated lysine (04685) were purchased from Sigma-Aldrich. Lysine, acetylated lysine and methylated lysine were measured as powders without any purification or thin films produced by the evaporation from water solution.

Details of sample preparation and research methodology have been described in detail in our previous papers \(^1,21,22\). Here we will remind the most essential details. The Raman microscope (an alpha 300 RA (WITec, Ulm, Germany) model) equipped with an Olympus microscope coupled via the fiber of a 50 μm core diameter with a spectrometer UHTS (Ultra High Throughput Spectrometer) and a CCD Camera (Andor Newton DU970N-UVB-353) operating in standard mode with 1600x200 pixels at -60 °C with full vertical binning has been employed to analyze fresh human breast tissues specimens obtained during the surgery. Raman data were pre-processed using the WITec Control/Project Plus 1.6/Project 2.10 package. The sharp spikes attributed to cosmic rays were removed by using frequency
and spatial filtering. The corrected Raman spectra were smoothed (order:3, 7pt) by a Savitzky and Golay procedure. The reduction of the low frequency spectral range to 600-1800 cm\(^{-1}\) were performed to eliminate the Rayleigh scattering and the Raman scattering of the support. The baseline subtraction was performed manually (using Origin) and the Raman spectra were normalized (area=1, using MATLAB). All procedures have been conducted under a protocol approved by the Bioethical Committee at the Medical University of Lodz (RNN/45/14/KE/11/03/2014). All experiments were performed in compliance with relevant laws and guidelines of the Ministry of Health of the Republic of Poland. All experiments were performed in compliance with relevant laws and guidelines of the Bioethical Committee at the Medical University of Lodz (RNN/45/14/KE/11/03/2014).

Briefly, we recorded Raman spectra and Raman images of the samples of normal tissue from the safety margin (known to be normal from the ‘gold standard’ technique of histopathology) and cancerous tissues from the tumor mass of the human breast: invasive ductal cancer (IDC) and lobular invasive cancer (ILC). The samples from 9 patients have been analyzed. The samples of IDC represent G2 (moderately differentiated (intermediate grade)), G3 (poorly differentiated (high grade)) and GX (grade cannot be assessed, after chemotherapy). For each patient the Raman spectra from the safety margin and the tumor mass have been recorded and analyzed. In the paper we will present the results for the following histopathology identifications: invasive ductal carcinoma (G2, G3, GX), invasive lobular carcinoma, mucinous carcinoma with intraductal carcinoma, invasive ductal carcinoma with in situ solid and cribriform type carcinoma, and invasive ductal carcinoma with in situ comedo and cribriform type carcinoma.

To extract important biochemical information and discriminate between cancerous and normal samples we have used statistical chemometric analysis by Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA)\textsuperscript{23,24}. 
For modeling, compression and visualization of the Raman features of the samples we used Principal Component Analysis (PCA). PLSDA was used for building predictive classification models, to validate the classification models and to calculate sensitivity and specificity of Raman biomarkers of acetylation/methylation in human cancers. A ROC curve analysis has been also performed. The Raman spectra of the samples were analyzed with statistical/chemometric software packages MATHLAB and PLS_Toolbox Version 4.0 for use with MATLAB.  

Results and discussion

In this section we will show how specific Raman vibrational features found in breast cancer can elucidate molecular epigenetic mechanisms. We will concentrate on monitoring alterations in cancer cells associated with acetylation and methylation processes.

Figure 3 shows the microscopy image, Raman image, histological image, Raman spectra of the human breast tissue from the negative safety margin (normal tissue) and the tumor mass of the same patient P130. The comparison between the images provides a clear evidence of full morphological correspondence between them. The advantage of Raman images over the conventional images is that they provide also detailed biochemical information from Raman spectra recorded in various areas of the images and presented in Fig. 3D and 3H.

Fig. 3. The microscopy image (A), Raman image (B), histological image (C) and Raman spectra (D) (colors of the spectra correspond to the colors of the Raman image) of the human breast tissue from the safety negative margin (normal tissue) and the microscopy image (E), Raman image (F) histological image (G) and Raman spectra (H) (colors of the spectra
correspond to the colors of the Raman image) of the human breast tissue from the tumor mass for patient P130.

To get insight into acetylation and methylation processes we shall concentrate on the spectral region of 2000-3700 cm\(^{-1}\) where the -C-H,=C-H, C\(\alpha\)-C-H, CH\(_2\), CH\(_3\) stretching vibrations of saturated or unsaturated hydrocarbon chains in lipids and peptide chains in proteins occur. In this spectral region we can record the vibrations of the acetyl and methyl chemical functional groups that allows non-invasively detect cellular acetylation and methylation by Raman spectroscopy. For example, Fig. 4 shows these processes in lysine, which was here chosen since it is well known that among the twenty amino acids, lysine is one of the most heavily acetylated and methylated amino acid [26-29].

Fig. 4 Acetylation and methylation in proteins.

To monitor properly the acetylation and methylation processes by Raman spectroscopy we must understand vibrational features of the C-H stretching modes. This spectral region exhibits many complexities that are largely due to Fermi resonances between the stretching C-H vibrations and double excitations of the CH\(_2\) bending modes at lower frequencies and strong overlapping of the bands of C-H vibrations of lipids and proteins. The bands at around 2852 cm\(^{-1}\) and 2880 cm\(^{-1}\) are typical for long hydrocarbon chains, particularly for lipids, and corresponds to the C-H symmetric and antisymmetric stretching vibrations of the methylene CH\(_2\) groups\(^{1,30-41}\).

However, the band of CH\(_2\) groups at around 2880 cm\(^{-1}\) overlaps with symmetric stretching vibrational modes of the methyl groups CH\(_3\) and R\(_3\)C-H\(^{30,37,42-44}\).

By contrast, the region at higher frequencies, particularly the band at around 2940 cm\(^{-1}\), exhibits more inconsistencies in assignment of vibrations. The band at around 2935-2940 cm\(^{-1}\)
of proteins and around 2930 cm$^{-1}$ in lipids was attributed to the symmetric stretching vibrational modes of the methyl groups CH$_3$\textsuperscript{1,37,45-48} or antisymmetric stretching vibrational mode of the methyl groups CH$_3$\textsuperscript{40,41}, while the band at around 2960-2970 cm$^{-1}$ was attributed to the antisymmetric stretching vibrational mode of the methyl groups CH$_3$ of proteins and lipids\textsuperscript{37,38,46,49-53}.

The other papers consider the assignment of the band at around 2935-2940 cm$^{-1}$ to be incorrect since it is not compatible with IR assignments and the band was attributed to the antisymmetric stretching vibrational modes of the methylene groups CH$_2$\textsuperscript{30,32,37,54}.

However, it has been shown recently\textsuperscript{42} that the spectral features in this region are more complex. The band at 2882 cm$^{-1}$ was assigned to the overlapping symmetric stretching vibrational modes of both CH$_2$ and CH$_3$; the band at 2938 cm$^{-1}$ was assigned to the two symmetric-CH$_3$ Fermi resonance modes and the weak-CH$_2$ antisymmetric stretching mode; the band at 2983 cm$^{-1}$ was assigned to the symmetric-CH$_2$ Fermi resonance mode and the weak-CH$_3$ antisymmetric stretching mode.

The peak at around 3009 cm$^{-1}$ corresponds to the =C-H stretching vibration\textsuperscript{34-36,46,55} or antisymmetric stretching mode of -C-H\textsuperscript{32}. At the same time the peak at 3067 cm$^{-1}$ corresponds to vibrations of the stretching modes of (C-H) - aromatic residues observed in aromatic amino acids (phenylalanine, tryptophan)\textsuperscript{34-37,46}.

The bands at around 2874-2897 cm$^{-1}$ were attributed to R$_3$C-H stretching bands in proteins\textsuperscript{37,39}, C$\alpha$-C-H stretching to bands of proteins that occur at 2980-2989 cm$^{-1}$\textsuperscript{37,39}. The band at around 2000-2017 cm$^{-1}$ was assigned to CH$_2$ symmetric vibrations\textsuperscript{53} in lipids and proteins\textsuperscript{31,47,54} or CH$_3$ symmetric stretching\textsuperscript{40,47,56}.

To get more information on C-H vibrations in acetylated, and methylated proteins we concentrated on lysine, because protein acetylation and methylation on lysine residue is one of the main epigenetic modifications that impact on gene expression and transcriptional
activity \(^{37,38}\). Furthermore, lysine acetylation and methylation on histone is one of the major post translational regulatory mechanisms of protein activity, such as differentiation and proliferation, signal transduction and metabolism, apoptosis and cytoskeleton dynamics \(^{26,59}\).

Fig. 5 shows the Raman spectrum of lysine, methylated and acetylated lysine in the region of C-H stretching modes in the crystalline and amorphous phases. The comparison can easily identify the vibrations of the methyl and acetyl groups. The bands at 2927 cm\(^{-1}\), 2963 cm\(^{-1}\) in Fig.5 A were attributed to the symmetric and the antisymmetric stretching C-H vibrations in the acetyl groups O=C-CH\(_3\) in crystalline phase acetylated lysine, respectively. The symmetric stretching C-H vibrations in the acetyl groups O=C-CH\(_3\) in the amorphous phase of acetylated lysine is shifted from 2927 cm\(^{-1}\) to 2932 cm\(^{-1}\) as one can see from Fig.5B.

The bands at 2953 cm\(^{-1}\) and 2970 cm\(^{-1}\) correspond to the symmetric and antisymmetric stretching C-H vibrations in the methyl groups CH\(_3\) in methylated lysine (Fig.5A). As this region is susceptible to Fermi resonances, it is possible that the bands at 2927 cm\(^{-1}\) and 2953 cm\(^{-1}\) correspond to the symmetric CH\(_3\) Fermi resonance mode and the -CH\(_2\) antisymmetric stretching mode; while the bands at 2970 cm\(^{-1}\) and 2983 cm\(^{-1}\) correspond to the symmetric -CH\(_2\) Fermi resonance mode and the -CH\(_3\) antisymmetric stretching mode \(^{42}\).

The band that is present both in the lysine, acetylated and methylated lysine at 2905 cm\(^{-1}\) has been attributed the vibrations of the symmetric stretching C-H the methylene group CH\(_2\). It is clear that the previous assignments of the bands at 2905 cm\(^{-1}\) to the C-H symmetric or antisymmetric stretching in the methyl group CH\(_3\) as proposed in \(^{40,47,49,56}\) must be incorrect since the methyl group is absent in lysine. The bands at 2856 cm\(^{-1}\), 2875 cm\(^{-1}\), and 2860 cm\(^{-1}\) that are present in the lysine, acetylated and methylated lysine, respectively have been attributed the vibrations of the antisymmetric stretching C-H the methylene group CH\(_2\).
It is worth emphasizing that methylation of lysine leads to blue shift (higher frequencies) at around 2970 cm\(^{-1}\) of the C-H vibrations of the methyl groups in contrast to acetylation where the C-H vibrations of the acetyl group at usually observed at 2927 cm\(^{-1}\) (crystal) -2932 cm\(^{-1}\) (amorphous phase).

Fig. 5 Raman spectrum of lysine, methylated lysine and acetylated lysine, solid crystalline powder (A), Raman spectrum of lysine, methylated lysine and acetylated lysine amorphous phase evaporated from aqueous solution c=10\(^{-2}\) M (B).

We have concentrated on interpretation of Raman vibrations lysine, methylated and acetylated lysine since it is well known that among the twenty amino acids, lysine is one of the most heavily modified\(^ {26-29}\). Lysine acetylation and methylation are involved in diverse cellular processes, such as chromatin remodeling, cell cycle, splicing, nuclear transport, and actin nucleation\(^ {26}\).

In order to visualize chemical similarities and differences in level of acetylation and methylation in cancerous and normal tissues we have evaluated the predictive validity and robustness of Raman spectroscopy using multivariate statistical methods for data interpretation. Due to high spatial resolution of Raman microspectroscopy that reveals heterogeneity of the sample, it is risky to use the Raman pattern of vibrational features of a single cell in the tissue. It is much more reliable to perform statistical analysis for a large number of cells in tissue samples. To simplify the task of analyzing large number of samples and multidimensional Raman vectors (intensities vs. wavenumbers), data were subjected to dimension reduction by means of Principal Components Analysis (PCA). The PCA method was applied to the 698 Raman spectra of 9 cancer tissue samples from the tumor mass and 9
normal tissue samples from the safety margin, for each of the ten patients both types of samples: normal and cancerous were analyzed.

By plotting the principal components scores, similarities between the samples can be revealed.

Fig. 6 shows PCA score plot for the Raman spectra of the human breast tissue samples from the tumor mass and from the safety margin. The results of the PCA reveal separation of the cancer and normal tissues. The similarities and differences are clearly visible by grouping the Raman spectra in two separate clusters, where the spectra with similar or identical vibrational properties for the samples from the tumor mass (red points) are grouped in the upper area while the samples from the safety margin (blue points) are grouped in the lower area of the plot.

Fig. 6. PCA score plot (model: area normalized to 1 ) for the Raman spectra (preprocessing: background subtraction) from the region (2000 cm\(^{-1}\) -3700 cm\(^{-1}\)) of the human breast tissue samples from the tumor mass (red points) and the safety margin (blue points), integration time 0.5 s, laser power 10 mW; ● invasive ductal carcinoma G3 patient P125, ■ invasive ductal carcinoma GX patient P126, ▲ invasive ductal carcinoma G2 patient P127, ▼ mucinous carcinoma with intraductal carcinoma G2 patient P128, ♦ invasive ductal carcinoma with in situ solid and cribriform type carcinoma patient P129★ invasive ductal carcinoma G3 patient P130, + lobular carcinoma patient P131, ★ invasive ductal carcinoma G3 patient P133, + invasive ductal carcinoma G3 patient P134, ● normal tissue patient P125, ■ normal tissue patient P126, ▲ normal tissue patient P127, ▼ normal tissue patient P128, ♦ normal tissue patient P129, ★ normal tissue patient P130, + normal tissue patient P131, ★ normal tissue patient P133, + normal tissue patient P134.
One can see from Fig. 6 that the first principal component PC1 has the contribution of 94.52% and component PC2 the contribution of 3.34% to variance, respectively. The component PC3 (not presented in Fig. 6) has the contribution of 0.40%. PC1 and PC2 give the dominant account for the maximum variance in the data.

To understand the molecular information contained in the first two PCs we used the loading plots presented in Fig. 7 that reveal the most important characteristic features in the Raman spectra. Fig. 7 shows the loading plot of PC1 as a function of the wavenumber. We can see that the loading plot shows the most pronounced changes around the characteristic Raman peaks of lipids and proteins.

Fig. 7 PCA loading plot for PC1 (A) and PC2 (B) (model: area normalized to 1) for the Raman spectra of the human breast tissue samples from the tumor mass and the safety margin, integration time 0.5 s, laser power 10 mW, 2000–3700 cm$^{-1}$.

The first principal component PC1, which gives the highest contribution to variance, reaches its maximum at 2927 cm$^{-1}$ in Fig. 7A and was attributed to the symmetric stretching vibration of the methyl group of lipids and proteins$^{1,34,36,40,60}$ and acetylated groups of proteins as we showed in Fig. 5.

The second principal component PC2 provides more subtle pattern of differences between the samples. The characteristic maxima and minima in the loading plot in Figure 7B are at 2856, 2905, 2953, 2970, 3056 cm$^{-1}$. The negative peaks at 2856 and 2905 cm$^{-1}$ corresponding to the C-H symmetric stretching vibrations of the methylene CH$_2$ groups$^{1,30,33,37,38}$ suggest decreased lipid contents. The positive peaks at 2953, 2970, 3056 cm$^{-1}$ corresponding to the symmetric stretching vibrational modes of the methyl groups CH$_3$$^{1,37,38,46,48}$, the antisymmetric stretching vibrational mode of the methyl groups CH$_3$ of proteins and lipids$^{40,41,49,54,56}$, and the stretching modes of (C-H) - aromatic residues observed...
in aromatic amino acids (phenylalanine, tryptophan) suggest enhanced acetylated/methylated protein profile corresponding to increased number of methylated and acetylated proteins.

Fig. 8 The characteristic spectra corresponding to the various coordinates of the PC1-PC2 score plot.

As it is evident from Fig. 7 all positive peaks indicate elevation in the expression of methylated/acetylated and aromatic amino acids rich proteins in the tissue.

Fig. 8 shows also the characteristic spectra corresponding to the various coordinates of the PC1-PC2 score plot. The most prominent bands that determine the coordinates on the PC1-PC2 plot is the band at around 2923 cm\(^{-1}\) -2942 cm\(^{-1}\) corresponding to acetylated proteins (CH\(_3\) symmetric stretching vibration of the acetyl group) and the band at 2005-2010 cm\(^{-1}\) of non-methylated/non-acetylated proteins (CH\(_2\) symmetric stretching vibration). For the positive PC2 corresponding to the samples of the cancerous tissue from the tumor mass the Raman peak frequency is observed at 2923-2942 cm\(^{-1}\) corresponding to acetylated proteins. (Fig 8). For the negative PC2 corresponding to the samples of normal tissue from the safety margin the Raman peak frequency is observed at 2904-2910 cm\(^{-1}\) (right lower corner) corresponding to non- methylated/non-acetylated proteins (CH\(_2\) symmetric stretching vibration) (Fig. 8). It indicates that the maximum of the Raman peaks is shifted from 2932-2942 cm\(^{-1}\) for the cancerous cells from the upper left corner to 2904-2910 cm\(^{-1}\) for the normal cells in the lower right corner of the PCA score plot as presented in Fig. 9.

In Fig. 8 one can see also the peak at around 3311 cm\(^{-1}\) characteristic for water confined in the breast tissue. We have proved that a single band with a maximum at 3311 cm\(^{-1}\) which does
not reveal any additional bands when analyzed by deconvolution is typical for cancerous tissue \(^{61}\).

In the view of the results presented so far one can provide strong evidence that the global acetylation level of histone and non-histone proteins increases in human breast cancer cells. The interplay between hypermethylation and hypomethylation or hyperacetylation and hypoacetylation is one of the most constant features of the cancer \(^{62}\). CpG island hypermethylation and global genomic hypomethylation of DNA are common epigenetic features of cancer cells \(^{63}-74\). Less attention has been focused on histone modifications in cancer cells. The global loss of monoacetylation and trimethylation of histone H4 has been suggested to be a common hallmark of human tumor cells \(^{59}\) which is apparently inconsistent with our findings presented in this paper.

Furthermore, HDACs expression and activity in tumor cells is very different from normal cells. Overexpression of HDACs is observed in the majority of cancers \(^{75}-80\).

The upregulation of HDACs that repress an important growth suppressive gene is an important mechanism to promote cancer cell proliferation. As such, HDAC inhibition can be used to stop the growth of cancer cells. Many HDAC inhibitors are being explored and even used for cancer treatment \(^{81,82}\). The addition of HDAC inhibitors (HDACi) has a similar net effect to increasing the amount of HAT activity, resulting in enhanced transcription levels \(^{16}\). However, reduced HDAC or increased HAT activities via specific inhibitors or suppressors lead to an increase in histone acetylation. Our results show evidently that the potential use of HDAC inhibitors in the clinic practice to treat cancer is exciting, but should be approached with caution, because only the balance between histone acetylation and deacetylation determines the level at which cancer cannot develop. The low ratio of HDAC/HAT leads to enhanced global acetylation that has roles in the development and progression of human cancer as it has been envisaged by our results. Our data suggest that the global increase of
acetylation of histone and non-histone proteins is a common hallmark of human breast tumor cells.

Fig. 9. The characteristic Raman spectra corresponding to cancerous (A), and normal (B) human breast tissue.

Fig. 9 shows the characteristic Raman spectra corresponding to cancerous, and normal human breast tissue. One can see that the samples of the cancerous tissue from the tumor mass have the Raman peak frequency at around 2938 cm\(^{-1}\) correlating quite well with the Raman peak position of acetylated lysine at 2932 cm\(^{-1}\) (Fig.5B). The samples of normal tissue from the safety margin have the Raman peak frequency at 2905 cm\(^{-1}\) corresponding to the vibrations of non-methylated/non-acetylated proteins (CH\(_2\) symmetric stretching vibration).

In contrast to the results presented in ref. [1] where we have concentrated on the average Raman spectra of human breast tissue comprising relatively large areas of the tissue that contain both epithelial cells of ducts, connective tissue and adipose cells of stroma, here we recorded the single Raman spectra with high spatial resolution of 1 \(\mu\)m from the epithelial cell of the tumor mass excluding adipose cells. The average Raman spectra and the single microRaman spectra from the tumor mass obtained in ref. [1] are identical as those in Fig. 9A, while they differ significantly in the region of the normal tissue from the safety margin. The reason is that the average Raman spectra measured in ref. [1] contain contribution from adipose tissue consisting of glycerol monooleate derivatives with Raman spectra at 2854, 2888, and 2926 cm\(^{-1}\), which was not measured in the single Raman spectra in Fig. 9B. One can see from Fig. 9B that the single Raman spectra are dominated by a peak at around 2905 cm\(^{-1}\) that corresponds to vibrations of non-methylated/non-acetylated proteins (CH\(_2\) symmetric stretching vibration).
Although PCA method is useful for visualization of data trends, it cannot be used as Supervised Learning classification method, where the previous knowledge of the samples under study must be known. In our case the class membership of every sample is a priori known from a ‘gold standard’ method of histopathology performed by the certified doctors. One group of samples was identified to be normal (negative) without breast cancer cells (safety margin), and the second group of samples from the tumor mass were identified as having cancer cells (positive). We used supervised learning algorithm PLSDA to calculate the sensitivity (the ratio of cancerous tissue found by the Raman method to the total number of abnormal samples known to be cancerous from the histopathology examination) and the specificity (the ratio of normal tissue found by the Raman method to the total number of normal samples known from the histopathology examination).

Equations 1 and 2 describe how to calculate the sensitivity and the specificity.

\[
\text{sensitivity} = \frac{TP}{TP+FN} \quad (1)
\]

\[
\text{specificity} = \frac{TN}{TN+FP} \quad (2)
\]

where:

TP- true positive result
FN- false negative result
TN- true negative result
FP- false positive result.

The sensitivity and specificity obtained directly from PLSDA and cross validation gives a sensitivity and specificity of 86.1% and 91.3% for calibration and 85.3% and 91.3% for cross-
validation, respectively. The method for cross-validation was venetian blinds w/10 splits and
the root mean square errors for the calibration (RMSEC) and cross validated (RMSECV)
models were 0.307 and 0.311, respectively, showing a good stability for predicting new
samples.

To further evaluate and compare the performance of the PCA and PLSDA based
diagnostic algorithm for breast cancer diagnosis, the receiver operating characteristic (ROC)
analysis was performed.

Fig. 10 shows a bimodal ROC (Receiver operating characteristic) curves graph for the
normal and cancerous human samples. The sensitivity and specificity obtained directly from
the ROC curve are 0.861 and 0.913, respectively.

The ROC curve in Fig. 10 illustrates the ability of Raman spectroscopy to separate enhanced
methylation/acetylation of proteins occurring in cancerous breast tissue. A simulated ROC
curve of two indistinguishable populations is included for comparison.

Fig. 10. The ROC curves graph for plotting specificity versus 1-sensitivity.

The area under the ROC curve (AUC) supports the discriminatory quality of the Raman
analysis. For the random test an AUC=0.5, and represents the area under the diagonal from
the origin of the plot (dashed line). The ROC curve in Fig.10 is shifted towards the upper left
and top indicating the discriminatory performance of the Raman method. The perfect
performance has an AUC=1, while the ROC curve for distinguishing between cancer and
normal tissue has an AUC =0.96.

One can see that the ROC curve presented in Fig. 10 is useful for a proper decision for a
Raman diagnostic test on acetylation/methylation related to cancer development.
Table 1 presents the Confusion matrix and Confusion table for the model chosen for the analysis.

<table>
<thead>
<tr>
<th>Model results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confusion matrix</td>
</tr>
<tr>
<td>class</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Confusion table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual class</td>
</tr>
<tr>
<td>Predicted as 1</td>
</tr>
<tr>
<td>Predicted as 2</td>
</tr>
</tbody>
</table>

Conclusions

We have demonstrated that the Raman imaging is capable of identification of breast cancer specific alterations and thus it has potential to provide predictive markers and treatment targets. This paper examines the methylation/acetylation epigenetic changes in breast cancer by Raman spectroscopy, and discusses how they contribute to different aspects of tumourigenesis in human breast cancer.

We have found that Raman spectroscopy is capable of monitoring of methylation/acetylation modifications in histone and non-histone proteins which are strongly associated with tumor development. Our results demonstrate that the stretching vibration of the acetyl group observed at around 2927 -2942 cm\(^{-1}\) and the methyl group at 2970 cm\(^{-1}\) by Raman spectroscopy is useful for monitoring these epigenetic molecular processes.
We have demonstrated that Raman vibrations of the acetyl and methyl groups have a potential to be used as Raman markers of cancer development by the assessment of global methylation/acetylation of proteins in the cancer cells of the human breast tissue. The Raman vibrational marker is markedly blue shifted from 2905 cm\(^{-1}\) for normal cells to around 2938-2942 cm\(^{-1}\) for cancer cells. Our results demonstrate that the Raman shift can be used as a hallmark of cancer aggressiveness of epithelial breast cells. The results presented in the paper demonstrate that Raman spectroscopy and Raman imaging are powerful methods that will bring important contribution to understanding of basic epigenetic regulatory mechanisms in biology, medicine and pharmacy. Our Raman results provide strong evidence that the global acetylation level of histone and non-histone proteins increases in human breast cancer cells and is a common hallmark of human breast tumor cells. The sensitivity and specificity obtained directly from PLSDA and cross validation for a chosen model gives a sensitivity and specificity of 86.1% and 91.3% for calibration and 85.3% and 91.3% for cross-validation, respectively.

Acknowledgements

The project was funded through The National Science Centre Poland Grant UMO-2015/19/B/ST4/01878.

Notes and references:


Mechanism of acetylation and deacetylation maintained by antagonistic activity of two groups of enzymes - histone acetyltransferases (HAT) and deacetylases (HDAC).

254x190mm (96 x 96 DPI)
Histone acetylation/deacetylation by HAT and HDAC enzymes alters chromatin structure by decreasing/increasing accessibility of exposed sites on DNA to silencing/activating gene transcription.

254x190mm (96 x 96 DPI)
The microscopy image (A), Raman image (B), histological image (C) and Raman spectra (D) (colors of the spectra correspond to the colors of the Raman image) of the human breast tissue from the safety negative margin (normal tissue) and the microscopy image (E), Raman image (F) histological image (G) and Raman spectra (H) (colors of the spectra correspond to the colors of the Raman image) of the human breast tissue from the tumor mass for patient P130.

161x76mm (300 x 300 DPI)
Acetylation and methylation in proteins.

60x22mm (300 x 300 DPI)
Raman spectrum of lysine, methylated lysine and acetylated lysine, solid crystalline powder (A), Raman spectrum of lysine, methylated lysine and acetylated lysine amorphous phase evaporated from aqueous solution c=10^{-2} M (B).

120x176mm (300 x 300 DPI)
PCA score plot (model: area normalized to 1) for the Raman spectra from the region (2000 cm\(^{-1}\) - 3700 cm\(^{-1}\)) of the human breast tissue samples from the tumor mass (red points) and the safety margin (blue points), integration time 0.5 s, laser power 10 mW

161x96mm (300 x 300 DPI)
PCA loading plot for PC1 (A) and PC2 (B) (model: area normalized to 1) for the Raman spectra of the human breast tissue samples from the tumor mass and the safety margin, integration time 0.5 s, laser power 10 mW, 2000–3700 cm$^{-1}$.
The characteristic spectra corresponding to the various coordinates of the PC1-PC2 score plot.

161x114mm (300 x 300 DPI)
The characteristic Raman spectra corresponding to cancerous (A), and normal (B) human breast tissue.
The ROC curves graph for plotting specificity versus 1-sensitivity

82x109mm (300 x 300 DPI)
Graphical abstract

72x50mm (300 x 300 DPI)