UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

ICFO
The Institute of Photonic Sciences

Statistical analysis and Plasmonic effects to extend the use of Raman Spectroscopy in Biochemistry

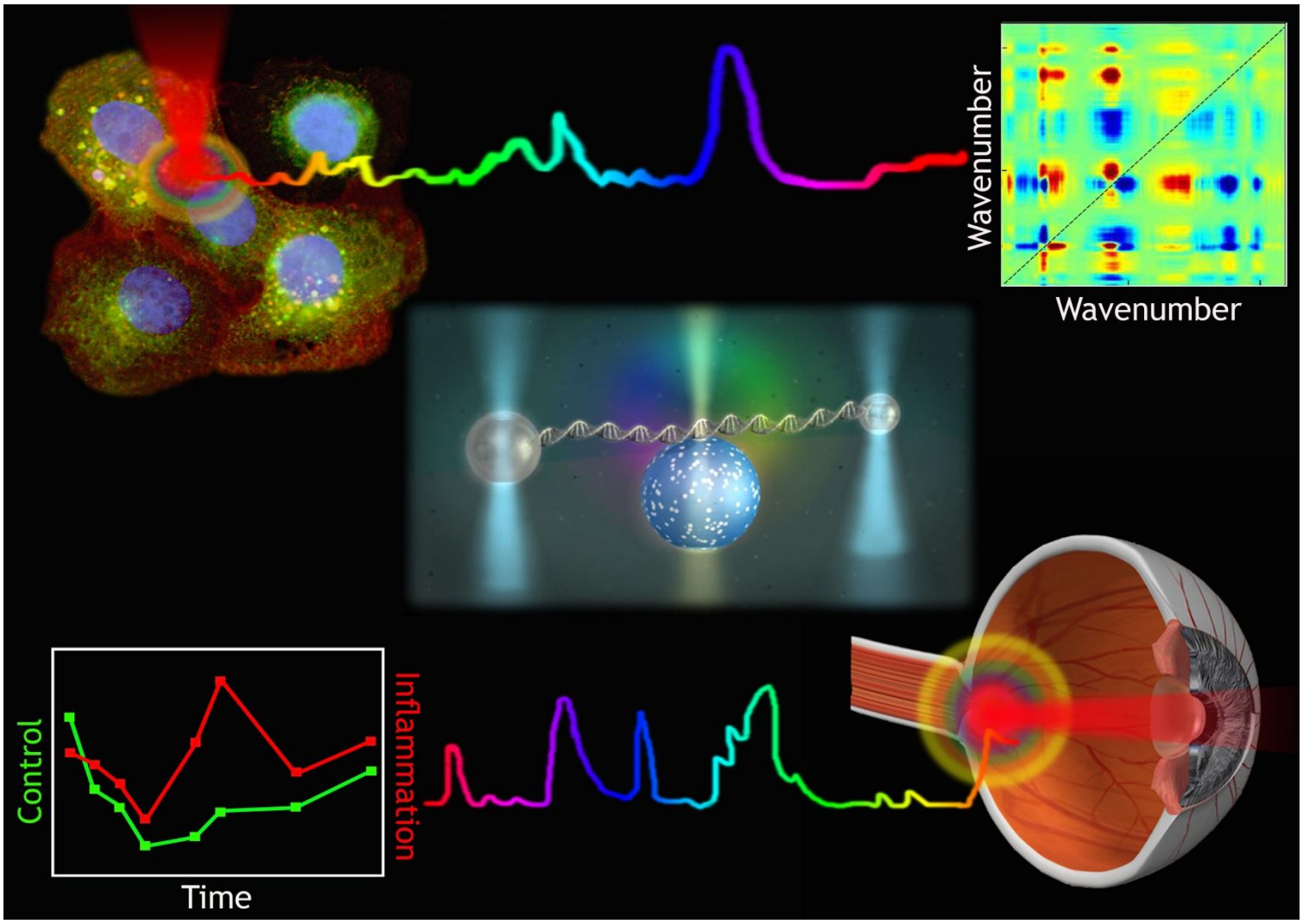
Mónica Marro

Advisor: Prof. Dmitri Petrov

ICFO - The Institute of Photonic Sciences

Universitat Politècnica de Catalunya

Barcelona 2013



Statistical analysis and plasmonic effects to extend the use of Raman spectroscopy in biochemistry

Mónica Marro

Under the supervision of

Prof. Dmitri Petrov

submitted this thesis in partial fulfillment
of the requirements for the degree of

Doctor

by the

Universitat Politècnica de Catalunya

Barcelona, February, 2013

To my family

Men love to wonder, and that is the seed of science
Ralph Waldo Emerson

Acknowledgements

This thesis could never have been finished without the support of lots of people that in many different ways contributed to the completion of this work.

First of all, I would like to express my sincere gratitude to Prof. Dr. Dmitri Petrov for giving me the opportunity to pursue this doctoral thesis under his supervision. He always provided me with all necessary to conclude this work and guided me to present good scientific research. I am in debt to him for letting me work in the exciting topic of my thesis: photonics to improve people's life and health.

I also want to express my gratitude to the director of ICFO, Prof. Lluis Torner. He knew how to construct ICFO to become a distinct place ideal to do research at the highest international level.

I would like to acknowledge the economic support of Fundació privada Cellex. The altruism of his founder Dr. Pere Mir has permitted many research projects to be founded in Catalunya.

The work of this thesis could not be possible without the collaboration with Biomedical Research Institutes and Hospitals in Barcelona. First, I am thankful to Prof. Pablo Villoslada, from Institut d'Investigacions Biomèdiques Agustí Pi Sunyer (IDIBAPS) and Hospital Clinic, and the members in his group: Alice Taubes and Alice Abernathy. We work together in a passionate study about how to diagnose with light Multiple sclerosis non-invasively and in early stages through the retina. They provided their vast expertise in retina inflammation and tissue culture. Second, I also express my gratitude to Prof. Angels Sierra and Dr. Claudia Nieva, from Intitut d'Investigacions Biomèdiques de Bellvitge (IDIBELL) and Hospital de Bellvitge. We collaborate in a project to study the stratification of breast cancer cells and provide a prognosis. They shared with us their knowledge in cell models and cancer.

I would like to thank also the interesting discussions that researchers related with the topic of my thesis had with me in conferences or workshops: Anna de Juan, Jürgen Popp and Thomas Bocklitz. This conversations were always very encouraging and brought me new ideas.

The environment during this almost four years would have been very different without the helpful and cheerful company of my group members: Iñaki, Stefan, Saurabh, Filip, Satish, Pau, Raul and Michal. We interchanged knowledge in an open environment during our lunch and coffe-break discussions as well as share good and difficult moments trying to be always supportive.

But not only scientific knowledge is necessary to fulfill a doctoral thesis with its pleasant surprises and hard work experiments... I deeply need to express my gratitude to a long list of friends with which I had very moving moments.

Finally, I would like to thank my family: Francesc, Maria, Gemma, Iñaki, Irene, Jesús. You were in grey times the only strength to continue and finish my thesis. Thanks you to share together joyful and also difficult moments that made our links tighter. Thank dad and mum to be an example that "difficulties are advantages" and can serve to increase your love. Thank you to teach us to see life as a challenge that we need to face with optimism, joy and giving the best of ourselves.

Abstract

Since its discovery in 1928, Raman spectroscopy (RS) has produced a revolution in the fields of analytical chemistry and molecular detection. Thanks to the latest technical advances, the expectations of the applicability of RS in biology have increased. Most recently, RS emerged as an important candidate technology to detect and monitor the evolution of the biochemical content in biomedical samples non-invasively and with high specificity. However, the inherent properties of Raman scattering have limited its full exploitation for biomedical applications. In the past decade, Surface Enhanced Raman Scattering (SERS) and multivariate analysis have emerged as possible solutions for overcoming the low efficiency and the complexity of the Raman signals obtained from biological material. Until 2009, only a few studies had been reported using multivariate approaches, and these techniques were only employed to group different types of samples. Moreover, although the SERS effect was demonstrated for cells, SERS probes were not used in their full capacity to study complex biological processes inside cells.

This thesis is a step towards combining and using statistical analysis and SERS to expand the applicability of RS in biochemistry: from single molecule to cell and tissue level. This mehtodology could reveal novel insights, otherwise inaccessible using previous techniques.

Specifically, we began studying the changes in Raman spectra of a single DNA molecule and a RBC under stretching employing optical tweezers. SERS and statistical techniques such as 2D correlation and PCA were used to reveal important structural properties of those biological materials.

An experiment to study intracellular pH changes in glioma cells after Photodynamic Treatment (PDT) was performed by using SERS probes embedded in the cells. The evolution in the SERS spectra was analyzed using 2D correlation. To the best of our knowledge, this study represents the first use of the 2D correlation technique to study cellular SERS spectra.

Furthermore, more complex systems were investigated, to reveal the molecular evolution of cells or tissues undergoing a biochemical process. PCA was used to study how lipid metabolism varied in different breast cancer cell lines depending on the degree of malignancy. However, PCA does not provide meaningful components that could be assigned directly to molecular Raman spectra. Consequently, Multivariate Curve Resolution (MCR) was proposed and applied to extract physically and chemically meaningful molecular components that changed in cancer cells during the Epithelial to Mesenchymal transition (EMT).

We monitored the retina composition *ex-vivo* when neuroinflammation was induced. Our study was the first application of MCR to decompose and monitor the molecular content of biological tissue with RS. Biomarkers for the early detection of neuroinflammation processes were identified and monitored. This is the first step in establising of a non-invasive and rapid screening technique for the early detection of multiple sclerosis or other neurodegenerative diseases in patients.

Finally, the flexibility of MCR-ALS algorithm was exploited to remove the presence of background signals in Raman spectra of cytological studies that

mask and degrade the results of a statistical analysis. Application of MCR-ALS enabled identification of molecular components that play an important role in the progression of breast cancer cells towards bone metastasis.

This research demonstrated a powerful method that adds a new dimension to the field of analytical chemistry. Sensitive and highly specific information can be extracted non-invasively, rapidly, and without sample preparation. The samples can be monitored in vivo, quantifying molecular components difficult or impossible to obtain with current technology.

Resumen

Desde que se descubrió la espectroscopia de Raman (ER) en 1928, esta técnica ha producido una revolución en el área de química analítica. Gracias a los últimos avances técnicos, las expectativas de aplicar ER en biomedicina han aumentado. La posibilidad de detectar y monitorizar la evolución del contenido bioquímico en muestras biomédicas de forma no invasiva con alta especificidad se ha convertido en una visión a perseguir. Sin embargo, las propiedades inherentes a la dispersión Raman han inhibido su completa explotación para aplicaciones biomédicas. En la última década, Surface Enhanced Raman spectroscopy (SERS) y el análisis multivariante se han erigido como posibles soluciones para superar la baja eficiencia y la complejidad de las señales Raman de material biológico. Hasta 2009, estas técnicas no habían sido explotadas para su uso en aplicaciones biomédicas. La presente tesis representa un paso hacia la combinación y el uso de análisis estadísticos y SERS para expandir la aplicabilidad de la ER en bioquímica: desde moléculas individuales hasta células y tejidos. Nuevos descubrimientos inaccesibles a técnicas bioquímicas usadas habitualmente, se han podido revelar con esta metodología.

En concreto, empezamos estudiando cambios en el espectro Raman de una molécula individual de ADN y un glóbulo rojo (RBC) sometidos a diferentes estiramientos por medio de pinzas ópticas. Se han utilizado SERS y técnicas estadísticas como correlación 2D (2DC) y Análisis de Componentes Principales (PCA) para revelar importantes propiedades estructurales de esos materiales biológicos.

Se realizó un experimento para estudiar los cambios de pH intracelulares en células gliales después del tratamiento fotodinámico (PDT) utilizando sondas SERS implantadas en el interior de las células. La evolución del espectro SERS fue analizado utilizando 2DC. Hasta donde sabemos, este estudio representa el primer uso de la técnica 2DC para estudiar espectros SERS celulares.

Además, se han investigado sistemas más complejos para revelar la evolución molecular de células y tejidos a lo largo de un proceso bioquímico. Con este objetivo, se utilizó el PCA para estudiar el metabolismo lipídico en diferentes líneas celulares de cáncer de mama relacionándolo con su grado de malignidad. No obstante, el PCA no proporciona componentes significativos que podrían ser asignados directamente a espectros moleculares Raman. En consecuencia, se propuso la Multivariate Curve Resolution (MCR) para extraer componentes moleculares con significado físico y químico que cambiaban en las células cancerígenas durante la transición epitelio-mesenquima (EMT).

En otra aplicación, se monitorizó la composición de la retina ex-vivo cuando se inducía una neuroinflamación. Nuestro estudio representaba la primera aplicación de MCR para descomponer y monitorizar el contenido molecular de un tejido biológico con ER. Se identificaron biomarcadores para la detección precoz de procesos neuroinflamatorios. Esto representa el primer paso hacia el establecimiento de una técnica no invasiva y de diagnóstico temprano de esclerosis múltiple u otras enfermedades neurodegenerativas en pacientes.

Finalmente, se explotó la flexibilidad del algoritmo MCR-ALS para eliminar la presencia de ruido de fondo en el espectro de Raman para estudios citológicos que enmascaran y degradan los resultados del análisis estadístico. Gracias a

eso, se pudieron identificar nuevos componentes moleculares que ejercían un papel muy importante en la progresión de células de cáncer de mama hacia la metástasis ósea.

Esta investigación ha revelado un potente método que añade una nueva dimensión al campo de la química analítica. Se ha podido extraer información con alta especificidad y sensibilidad de forma no invasiva, rápida y sin preparación especial de la muestra. Las muestras pueden ser monitorizadas in vivo, cuantificando sus componentes moleculares difíciles o imposibles de extraer con la tecnología actual.

Contents

Introduction

Methods to quantify the presence of specific molecules in unknown samples are required in a vast number of applications that extend from the most basic scientific research areas, to engineering, food safety monitoring, and the most challenging medical applications. In this thesis we focus on the area of biology and medicine.

Non-invasive monitoring with high specificity of the molecular content of diverse samples such as tissues and cells is crucial when diagnosing or investigating the mechanisms that govern the progress of a disease or the effects of drugs. For example, in cancers, profiling tools for metabolite analysis are necessary to provide a more comprehensive picture of tumor development and progression [1]. These "metabolomic" approaches can provide important information about tumorigenesis, revealing new therapeutic targets and therefore helping in the diagnosis. Multiple sclerosis (MS) is an inflammatory demyelinating disease of the central nervous system (CNS) that lead to neurological and autoimmune manifestations [2]. In this field a non-invasive tool to access the CNS and to detect the metabolites or biomarkers related to neuro-inflammation as well as to study the effect of new drugs and to find new, more accurate biomarkers is required. Usually the symptoms start when about 70% of the CNS is already affected. Thus, the screening of healthy patients is necessary even if the symptoms are still not noticeable. It is important to establish routine tests for the detection of this disease in its early stages with non-invasive, inexpensive, and rapid tools.

The study of a biomolecule such as DeoxyriboNucleic Acid (DNA) is also very important because the molecule contains key information related to various diseases. Specifically, the study of the mechano-chemistry of living cells and DNA can help in understanding mechanisms such as the twisting of the DNA during the DNA–protein interaction, or the effect of external forces on cell motility or metabolic reactions. In all the above-mentioned health-related fields, a good understanding of the biochemical content of the studied samples is required along with the changes in an external parameter that could be: the adhesion of a drug, the differences in malignancy between cells, the time evolution, the progression of the disease, the force applied, etc.

In the past decades, several non-optical profiling tools have been developed for a more comprehensive picture of the biochemical composition and progression of a biological sample. Specifically, biochemical techniques such as High-Performance Liquid Chromatography (HPLC) [3], Gas Chromatography-Mass Spectrometry (GC-MS) [4], Liquid Chromatography-Mass Spectrometry (LC-MS) [5], Metabolite arrays [6], and Nuclear Magnetic Resonance (NMR) spectroscopy [7] have been used routinely in laboratories. A brief comparison of the above mentioned techniques is given in Table 1. However, most of these methods are invasive, costly, complex, and time consuming. Furthermore most of them require previous information on the molecule that needs to be quantified or identified. Usually, to provide that information on the molecules constituting the sample is not very easy.

Table 1: Different methods used in metabolomic analysis. Table adapted from [1]

Technique	Description	Advantages	Disadvantages
Gas chromatography-mass spectrometry (GC-MS)	The method uses the gas chromatography to separate the metabolite mixtures prior to mass spectrometry to identify different metabolites	A relatively cheap and reproducible method that also has a high degree of sensitivity	Sample preparation can be time consuming; not all compounds are suitable for gas chromatography
Liquid chromatography-mass spectrometry (LC-MS)	A similar approach to GC-MS, except separation occurs during liquid chromatography	This method is increasingly being used in place of GC-MS as sample preparation is not as time consuming; similar in sensitivity to GC-MS	More costly than GC-MS and depends on the reproducibility of the liquid chromatography (potentially more difficult to control than gas chromatography); also can suffer from ion suppression, where metabolites are poorly ionized when in the presence of cations and anions
Metabolite arrays	These devices use a 96-well plate assay system for phenotyping	Good as a screening tool when produced for a given situation	The number of metabolites that can be measured is limited by the number placed on the chip; difficult to screen for unknowns and follow metabolism of xenobiotics
Nuclear magnetic resonance (NMR) spectroscopy	Relies on the phenomenon of nuclear magnetic resonance and can provide detailed information about the structure, dynamics, reaction state, and chemical environment of molecules. This approach has been widely used by the pharmaceutical industry and in the screening of patient urine and blood plasma samples	A non-invasive technique - the use of NMR spectroscopy demonstrates that metabolomic analysis of tissues in humans is possible; it can be fully automated and has a high degree of reproducibility; relatively easy to identify metabolites from simple one-dimensional spectra	Lower sensitivity than mass spectrometry; co-resonant metabolites can be difficult to quantify; drug metabolites can be co-resonant with metabolites of interest

On the contrary, techniques based on the interaction of light with matter are very promising. Optical techniques can help the diagnosis be less invasive and more informative. In order to introduce these techniques, we briefly consider how incident light interacts with a molecule in a sample (Fig. 1). After the interaction, the scattered photons can have the same energy (Rayleigh scattering) or a different energy as the incident one had. To understand the spectrum of the scattered light, we need to review the basics of molecular physics [8, 9]. Briefly, as is shown in Figs. 3 and 2, the energy of the molecule is quantized, and depends on its electronic, vibrational, and rotational states: $E = E_e + E_v + E_r$, where E is the total energy of the molecule, E_e is the energy of the electronic state, E_v is the vibrational energy, and E_r the rotational energy. When light interacts with molecules, the photons may be absorbed or scattered. If the energy of the incident photons is higher than the energy difference between two electronic states of the molecule, the photons are absorbed, and the molecule is promoted to excited states with higher energy. After this event, non-radiative processes can occur, and the photons can be emitted from the lower energy of the excited state, leaving the molecules in a vibrational state of the fundamental electronic state. Processes of this type are called photoluminescence, and as a result, the emitted photons have lower energies than the incident ones.

Other absorption processes can occur when the light energy matches the energy difference between two different vibrational states. Then, the photons may be absorbed, leaving the molecule in a higher vibrational state. This is the basis of infrared (IR) absorption spectroscopy, since the vibrational energies are of the order of $10^4 - 10^2$ cm^{-1}, i.e., the same as IR photons [10, 11].

On the other hand, the light is scattered also by the molecule. In this case, the energy of the incident photon can have a mismatch with the difference between two molecular energy levels. When the scattered photons have the same energy as the incident photons, the process is called Rayleigh scattering. On the contrary, photons can undergo inelastic scattering due to the interaction with the molecular vibrations and, therefore, the energy of the scattered photons can be different from the incident ones. In this case, the process is called Raman scattering.

Three processes, luminescence, IR absorption, and Raman spectroscopy (RS), have been used for the identification of the molecular content of a sample, however only vibrational spectroscopy (Raman scattering and IR absorption) possesses the highest specificity due to its inherently sharp bands (typically on the order of $10-20$ cm^{-1}). Photoluminescense, on the contrary, possesses broad bands and therefore the molecular information is difficult to access.

To understand the capabilities of RS, we briefly discuss its basic fundamental principles [11–14]. According to the classical theory, Raman scattering can be described as follows. The electric field strength of an electromagnetic wave (incident light) fluctuates with time (t) according to

$$\vec{E} = \vec{E_0} \cos 2\pi\nu_0 t, \tag{1}$$

where $\vec{E_0}$ is the amplitude of the electric field, and ν_0 is the frequency of the incident light. When a diatomic molecule is irradiated with this electromagnetic wave, an electric dipole moment $\vec{P}$ is induced:

$$\vec{P} = \alpha\vec{E_0} \cos 2\pi\nu_0 t, \tag{2}$$

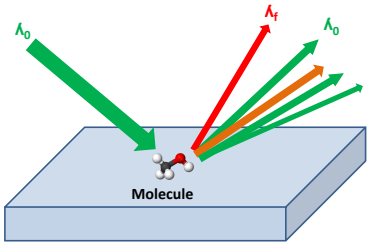


Figure 1: Scheme of the interaction of light with a molecule. Photons can be scattered with the same frequency (Rayleigh scattering) or with different frequencies or they can be absorbed.

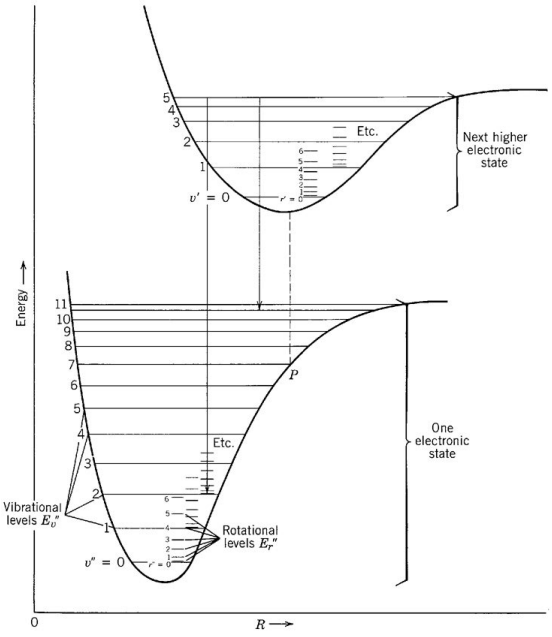


Figure 2: Molecular energy diagram showing electronic, vibrational and rotational levels. Adapted from [8].

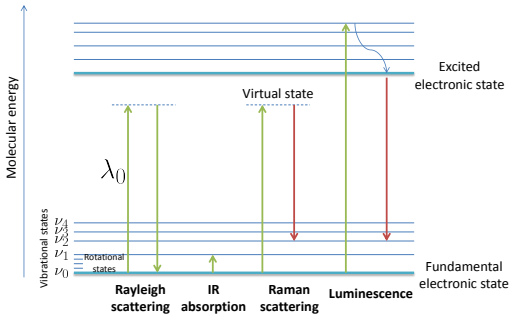


Figure 3: Molecular energy diagram including the main possible interactions of light with molecules.

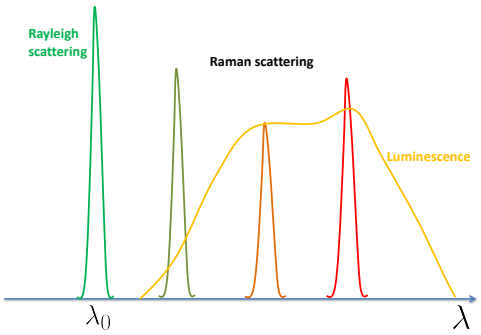


Figure 4: Representation of a typical spectrum obtained when a sample is illuminated with a monochromatic beam. Rayleigh scattering will appear at the same wavelength as the incident photons (λ_0), discrete Raman bands at different wavelengths and luminescence is represented by broad and intense bands.

where α is a constant called the *polarizability*. In the case of isotropic molecules, α is a scalar, and for non-isotropic molecules, α is a tensor.

If the molecule oscillates with a frequency ν_m, the nuclear displacement can be written as [12]:

$$q = q_0 \cos(2\pi\nu_m t), \tag{3}$$

where q_0 is the vibrational amplitude. For small amplitudes of vibration, α is a linear function of the nuclear displacement:

$$\alpha = \alpha_0 + \left(\frac{\partial\alpha}{\partial q}\right)_0 q + ..., \tag{4}$$

where α_0 is the polarizability at the equilibrium position, and $\left(\frac{\partial\alpha}{\partial q}\right)_0$ is the rate of change of α_0 with respect to the change in q, evaluated at the equilibrium position. This holds for every component in the polarizability tensor. Thus, for a given component of the induced dipole moment,

$$P_x = \left(\alpha_{xx}E_x^0 + \alpha_{xy}E_y^0 + \alpha_{xz}E_z^0\right)\cos\omega_0 t, \tag{5}$$

$$\begin{aligned} P_x &= \left(\alpha_{xx}^0 E_x^0 + \alpha_{xy}^0 E_y^0 + \alpha_{xz}^0 E_z^0\right)\cos\omega_0 t \\ &+ \left\{\left(\frac{\partial\alpha_{xx}}{\partial q}\right)_0 E_x^0 + \left(\frac{\partial\alpha_{xy}}{\partial q}\right)_0 E_y^0 + \left(\frac{\partial\alpha_{xz}}{\partial q}\right)_0 E_z^0\right\} \\ &\{q_0\cos\omega_0 t\cos\omega_m t\}, \end{aligned} \tag{6}$$

where $\omega_0 = 2\pi\nu_0$ and $\omega_m = 2\pi\nu_m$.

$$\begin{aligned} P_x &= \left(\alpha_{xx}^0 E_x^0 + \alpha_{xy}^0 E_y^0 + \alpha_{xz}^0 E_z^0\right)\cos\omega_0 t \\ &+ \frac{q_0}{2}\left\{\left(\frac{\partial\alpha_{xx}}{\partial q}\right)_0 E_x^0 + \left(\frac{\partial\alpha_{xy}}{\partial q}\right)_0 E_y^0 + \left(\frac{\partial\alpha_{xz}}{\partial q}\right)_0 E_z^0\right\} \\ &\{\cos(\omega_0+\omega_m)t + \cos(\omega_0-\omega_m)t\}. \end{aligned} \tag{7}$$

The first term represents an oscillating dipole that radiates light of frequency ν_0 (Rayleigh scattering), while the second term corresponds to the Raman scattering of frequency $\nu_0 + \nu_m$ (anti-Stokes) and $\nu_0 - \nu_m$ (Stokes). If $\left(\frac{\partial\alpha_{xi}}{\partial q}\right)_0$ for i=x,y and z in 7 is zero, the vibration is not Raman active. In other words, the selection rules for a molecular vibration to be Raman active are that the rate of change of the polarizability (α) with the vibration, i.e. $\left(\frac{\partial\alpha_{xi}}{\partial q}\right)_0$ for i=x,y and z, must not be zero.

According to quantum mechanics, the spectrum of molecular vibrations is discrete and unique for each molecule, thus, Raman spectrum gives a fingerprint of the molecule investigated.

Another vibrational spectroscopic technique is IR absorption. Unlike RS, a molecular vibration is IR active if the dipole moment is changed during the vibration. There are some rules for the IR and Raman activities [11, 13, 15]. The *mutual exclusion principle* states that vibrations that are symmetric with respect to the center of symmetry are Raman active but not IR active, whereas vibrations that are antisymmetric with respect to the center of symmetry are IR active but not Raman active. Some vibrations are inherently weak in IR and

strong in Raman spectra. In general, covalent bonds are stronger in Raman and ionic bonds stronger in IR.

Therefore, IR and RS could be considered as complementary techniques, but some drawbacks inherent to IR absorption's being applied to biological samples make RS more promising in this field.

First, constraints on the sample thickness need to be taken into account for IR spectroscopy. Second, vibrational modes with low frequencies are generally not observable in IR spectroscopy. Third, water is a weak Raman scatterer but IR spectroscopy suffers from the strong absorption of water. Therefore, aqueous samples can be studied with RS without major interference from the water, which is a major advantage in the case of biological specimens. Also, there are no great difficulties due to background signals from sample container materials such as glass windows [11].

An additional advantage of RS is its simple technical implementation [11]. In RS, the observation of the vibrational modes is usually independent of the excitation frequency. Since RS measures the shift in frequency from the excitation frequency, it can be performed using any excitation frequency range from UV to NIR. The IR absorption technique, on the contrary, needs a continuum light source in the IR spectral region. This makes necessary a lengthy and continual intensity calibration for each wavenumber. Also, being an absorption technique, implementation is more complex because it depends on the sample thickness and material properties, and its applicability to in vivo study of tissues, for example, becomes difficult. Consequently, the technical implementation of Raman spectroscopy is simpler than that of IR absorption spectroscopy and provides potential benefits for its application for the investigation of biomedical samples. The minimal components that a RS system must have are an excitation source (generally a continuous-wave (CW) laser), a sample illumination source, a collection system for the scattered light, a wavelength selector (spectrometer), and a detection and computer control system.

Another advantage of RS is its specificity [12, 14]. Each molecule has a different Raman spectrum, allowing the identification of molecules from among a mixture [16]. Also, the intensity of the bands is proportional to the molecular concentration in the sample being studied [17].

However, there are some drawbacks that need to be solved to fully exploit RS in the biomedical field. First, as was mentioned before, each molecule has its characteristic Raman spectrum with its associated bands. In the case of a biological sample containing a large number of different biomolecules, the Raman spectrum become increasingly complex. Furthermore, some Raman bands of different molecules can overlap. Consequently, the analysis of Raman spectra becomes difficult. For this reason, we use mathematical processing such as multivariate analysis and mathematical methods such as 2D correlation analysis to interpret and extract the information encoded in the spectra.

The requirements of these mathematical methods vary, depending on the application, but the main objectives of the statistical analysis are: firstly, a high ability to deconvolve the correlated signal of different molecules in the Raman spectra and to study the band's behavior with the changes of one parameter in the system. The band's evolution, such as shifts, broadenings, etc., also provide molecular information. Second, Raman spectra are inherently weak ($\approx 1/10^7$ of the scattered photons). Hence, some work is necessary to enhance the intensity or overcome this issue. There are techniques to enhance Raman signals, such

as Surface Enhanced Raman scattering (SERS). SERS employs rough metallic surfaces to increase the Raman output [18]. The first use of SERS for cellular studies was in 2005 [19] by using gold and silver colloid nanoparticles. After this publication, the number of studies based on SERS probes for biological applications increased [20–22], but until 2009 these probes were not fully exploited to study complex biological processes inside cells. One reason could be the incompatibiliy of the probes with the cell environment, or the difficulty of analyzing the SERS spectra. Depending on the application, the use of metallic structures in biosamples might not be feasible. Again, it would be necessary to use mathematical methods to overcome the low efficiency of RS.

At the beginning of my PhD (September 2009), a few reports had been known on using RS in biomedical applications combining it with statistical analysis or SERS (see Fig. 5). A significant increase in publications about RS applied to biomedical problems has occurred, mainly due to improvements in the methodology from the application of mathematical methods and SERS.

Before 2009, the applications of RS to the study of biological specimens such as cells or tissues were limited to single band analysis or basic multivariate analysis. There are some examples of the use of multivariate analysis that exploit Principal Components (PCA) to separate different groups of samples [23–25]. Other discriminant techniques such as Linear Discriminant Analysis have helped to identify different group of samples with Raman spectroscopy [26–28]. Also, Multiple Least Squares (MLS) has been used to asses the contribution of a previously measured Raman spectrum of a specific molecule to the Raman spectra of the sample measured [29–31]. However, no studies were perform to monitor the biochemical evolution of biological processes in cells or tissue.

Also, only a limited number of publications existed for the study of the Raman spectra obtained by changing one of the parameters in the system. Specifically, there was the need of an objective technique to deconvolve the Raman spectra in a number of meaningful components that could be assigned to different biomolecules or metabolites. In this way, new and otherwise inaccessible insights could be revealed about the processes studied.

The group "Optical Tweezers" at ICFO had performed several studies (including yeast cell measurements) [32–34] before 2010 that showed that the dynamics of some well-known biochemical changes could be monitored at the single cell level. After this, the group changed its strategy, concentrating on applications of RS to real demands originating from biological laboratories associated with medical organizations in Barcelona. Compared with the previous experiments, in which a simple hypothesis needed to be corroborated with RS unambiguously, now our collaborators have several hypotheses for understanding biological processes, and by the use of RS we provide additional information that permits one to evaluate the probabilities of each hypothesis. As we mentioned above, the inherently low level of the Raman scattering signal and the complexity of those signals for biomedical samples are the main difficulty in such estimations.

Therefore, for all the above mentioned reasons, **the main goal of the present thesis is to provide a systematical study of several advanced mathematical approaches for processing Raman data and to compare their ability to get otherwise inaccessible biochemical information. Together with applications of SERS as a mean to increase the Raman efficiency, this combination is used in a broad range of ap-**

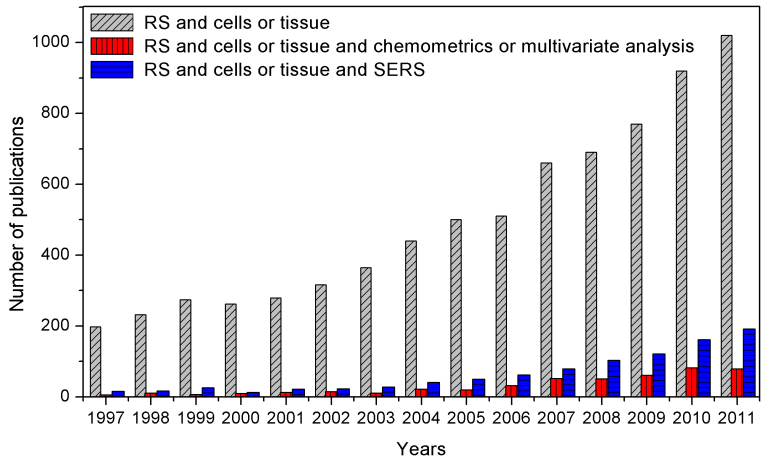


Figure 5: Number of publications during the past 15 years with the topics: RS and cells or tissue or DNA (gray color, slanted lines); Raman spectroscopy RS and cells or tissue or DNA and multivariate analysis or chemometrics (red color, vertical lines); RS and cells or tissue or DNA and Surface Enhanced Raman spectroscopy (SERS). Source: ISI web of knowledge.

plications. This will start from simple cases, but with a view to fundamental interests such as the study of Raman signals from single extended DNA molecules or the stretching of a red blood cell (RBC). Then, we will show the application of these techniques to study the pH changes in cancer cells after photodynamic therapy, and drug diffusion inside cancer cells. Finally, we used our approaches for real demands originating from biological laboratories associated with hospitals such as the study of neuroinflammation processes that occur in the retina (Institut d'Investigacions Bio mediques Agusti Pi i Sunyer (IDIBAPS) and Hospital Clinic) and the degrees of malignancy of cancer cells associated with metabolic changes (Institut d'Investigacions Biomediques de Bellvitge (IDIBELL) and Hospital de Bellvitge).

We cover a wide range of biochemical applications in which previously the use of RS has been restricted due to the drawbacks inherent in the technique when applied to biological samples. For each case, we studied how to fulfill the objectives by comparing, selecting, and developing the best suitable statistical technique and using SERS if possible.

This thesis is organized as follows: in **chapter 1**, the principles of RS and SERS are discussed briefly. **Chapter 2** describes the mathematical methods used and developed in this thesis. In **chapter 3**, the setups used for the experiments, the protocols to prepare the SERS substrates, and the software used are described. In **chapter 4**, the SERS spectra of a single DNA molecule under stretching is studied by using 2D correlation spectroscopy. Some important features in the spectra were revealed thanks to this technique, such as a shift in the O-P-O Raman band. In this study, optical trapping and surface-enhanced Raman scattering (SERS) are combined to establish a direct relationship between the DNA's extension and structure in the low force, entropic regime. A DNA molecule is trapped close to a SERS substrate to facilitate a detectable

Raman signal. DNA Raman modes shift in response to the applied force, indicating phosphodiester mechanical alterations. In **chapter 5**, a increasingly more complex experiment was performed to study intracellular pH changes in glioma cells after Photodynamic Treatment (PDT) by using SERS probes embedded in the cells. The evolution in the SERS spectra caused by the changes in the intracellular pH were studied with 2D correlation. A complete picture of the Raman band's behavior was achieved, that permits selecting the best Raman band capable of tracking the pH changes. To the best of our knowledge, this study represents the first use of the 2D correlation technique in the study of SERS spectra. In **chapter 6**, another study of the effect of a mechanical force on a living cell is discussed, the case of RBC stretching. The band dynamics along different force loads are investigated with principal component analysis (PCA) and 2D correlation. The results revealed important structural proper ties of the RBC cells related with their function in the body.

In the previous chapters, the use of 2D correlation gave important insights on the changes in the band shape or position correlated with the evolution of the parameter being perturbed (pH or force loads) and with other Raman bands in the spectra. This can give information about important and detailed molecular conformation changes in the system studied. However, in some biochemical studies, one needs a technique to extract from the spectra information about the molecular components that evolve during the process analyzed. 2D correlation is limited in that respect since the system starts to be complex (like cells or tissues when the number of biomolecules in the samples become large).

In the next chapters, more complex systems are going to be investigated, such as the molecular evolution of cells or tissues during a biochemical process. A complete understanding of the behavior of the biochemical content is pursued when having very limited or no a priori information about the molecules involved and their concentration changes during the process. Pursuing this goal, in **chapter 7**, PCA is used to study the lipid metabolism in different breast cancer cell lines, depending on the degree of malignancy in the CH stretching region of the Raman spectra. Also a Partial Least Squares-Discriminant analysis (PLS-DA) was perform achieving a high discrimination between metastatic and benign cell lines. However, PCA does not provides meaningful components that could be assigned directly to molecular Raman spectra spectra. Consequently, in **chapter 8** we use a multivariate technique, Multivariate Curve Resolution (MCR), to extract physically and chemically meaningfull molecular components that changed in cancer cell composition during Epithelial to Mesenchymal transition (EMT). MCR permitted to conclude that the EMT process affects the lipid profile of cells, increasing tryptophan but maintaining a low phosphatidylserine content in comparison with highly metastatic cells. Specifically, the use of MCR enabled to deconvolve and track the molecular content of cancer cells during a biochemical process, being a powerful non-invasive tool for identifying metabolic features of breast cancer cell aggressiveness. The application of MCR to Raman spectra, then, opens up the possibility to dynamically monitor molecular components in living tissue. For this reason, in **chapter 9**, we use MCR, to monitor the retina composition when a neuroinflammation is induced. Meaningful molecular Raman spectra could be deconvolved from the spectra acquired at different time points. Assigning each spectrum to a molecular component related with the neuroinflammation, new, complete, and otherwise inaccessible information about the biochemistry behind the process of neuro-inflammation

could be extracted. MCR has rarely been used in the RS of biological specimens, but nowadays it is increasingly being used for Raman images [35]. Our study represents the first application of MCR to decompose and monitor the molecular content of a biological tissue with RS. The MCR-ALS algorithm is very flexible [36] and exploiting the use of initial constraints or estimates can provide a powerful method to solve crucial problems still not overcome in the use of RS in biology. For instance, in **chapter 10**, we exploit the use of the initial constraints in the MCR-ALS algorithm to remove from the spectra the background signal not intrinsic to the cells in the study of cytological preparations of different cancer cell lines. Thanks to the removal of these contributions to the spectra, important molecular components could be revealed which are key in the progression of breast cancer cells towards bone metastasis. Finally, the **last chapter** summarizes the main conclusions of the research performed during this thesis.

Briefly, the main achievements made during this thesis consist in:

- Combining SERS with statistical techniques such as 2D correlation to study the dynamics and correlation of Raman bands of single cells and DNA when a parameter is changed in the system (force, pH, time, ...).
- Extracting from the experimental Raman spectra the individual signatures of the different molecular components in a sample and monitoring their concentration changes during a biochemical process.
- Exploiting the MCR-ALS algorithm to solve a serious issue in the application of RS to biological samples, viz., the background signals that mask the Raman spectral signatures.
- Revealing biochemical information otherwise inaccessible to other biochemical techniques, such as the evolution of the biomolecular content during retina inflammation and the Epithelial to mesenquimal transition of cancer cells.

A list of publications and conference contributions is placed at the end of thesis.

1

Raman spectroscopy

In this chapter we describe different spectroscopic techniques that provide information about molecular vibrations and thus, molecular composition of samples being analyzed. We will show that Raman spectroscopy represents the best suitable technique for the biomedical applications subject to our study: inspect the molecular composition of cancer cells, tissues, biological fluids and changes in conformation of single molecules when a dynamic external parameter is changing: (pH, force, concentration...). SERS, a phenomena to improve Raman efficiency will be introduced at the end of the chapter.

Raman scattering phenomena was discovered in 1928 by Sir Chandrasekhra Ventaka Raman with only crude instrumentation. The source used was the sunlight and a telescope was the colector; as a detector he used his eyes. That such feble phenomena as the Raman scattering was detected in these conditions is very remarkable. From that time improvements in the components of Raman instrumentation have been performing. First, excitation sources were developed to achieve higher ligh intensities until the laser invention in 1960. Second, progress was conducted in the detection systems for Raman measurements. At the beginning photographic plates were used, following the photomultipliers in the 50's and the Cary Model 81 Raman spectrometer. Finally, developments in the optical train of Raman instrumentation took place in the 1960s. All developments in the detection system allow the expansion of the applicability of Raman spectroscopy in other fields like biology. Several books have been published to explain Raman scattering phenomena and its last applications [11–15].

Raman spectroscopy is based in the following phenomena. When a monochromatic light of frequency ν_0 is incident on a molecule, the scattered light consists of two contributions: one, called Rayleigh scattering, which is strong and has the same frequency as the incident beam (ν_0), and the other, called Raman scattering, which is very weak ($\approx 1/10^7$ of the scattered photons) and has frequencies $\nu_0 \pm \nu_m$ where ν_m is a vibrational frequency of the molecule in the sample. In Raman scattering photons interact inelastically with the sample and

as a result, scattered light have wavelengths different than the incident light fig(1.1).

Raman spectroscopy presents several advantages with respect to other analytical techniques [13]. First, it is a scattering technique and a sample only needs to be placed under the excitation light and collect the scattered light, making it more flexible in terms of sample thickness. Second, water and glass have weak Raman spectra, extending the applicability of the technique to fields like Biology where the use of aqueous samples is basic. For Raman spectroscopy, therefore, only minimal sample preparation is required.

Biological processes like diseases implies changes in molecular concentration in cells and tissues. Therefore, vibrational spectroscopy is ideal for sensitive detection of those changes and diagnosis. From the fact that no sample preparation is required, sample measurements are simple and collection times range from seconds to minutes, Raman spectroscopy presents a rapid, objective, non-invasive and cost-effective tool for early diagnosis of disease processes in patients. IR spectroscopy, however, being also a vibrational spectroscopy, requires a tunable source of light, which is a much complex system than the monochromatic source used in Raman spectroscopy.

When analyzing biomedical samples with Raman spectroscopy, all active vibrational modes present in the mixture of the sample are observed in a single spectra. Therefore it leads to a very complex spectra with superimposed spectral features through all the spectral range. Statistical analysis is then required to extract the high amount of information stored in one single spectra. Additional chapters in this thesis are devoted to explain this techniques. On the contrary other routinary techniques used in biochemistry to asses the molecular composition of samples like fluorescence, gas chromatography, UV-VIS spectroscopy, HPLC analysis, can analyze only a portion of the biological content. In addition, most of the techniques needs a very complex sample preparation and pretreatment, arriving to be destructive while Raman spectroscopy no damage of the biomedical sample is required [13]. This allows Raman spectroscopy to study biomolecules in physiological conditions avoiding problems like sample preparation, invasiveness and cross-contamination.

Advantages of Raman versus IR spectroscopy exist from the fact that they are based in a scattering and absorption process respectively [11, 12]. First, Raman spectra is composed by narrower bands allowing to extract more information in an simplified manner. Second, constraints about sample thickness needs to be taken into account for IR spectroscopy. Third, lower vibrational modes are generally not observable in IR spectroscopy. Fourth, in Raman spectroscopy no major difficulties about water interference or background signal from samples containers like glass exist. In the case of Raman spectroscopy, spatial resolution is dependent on the excitation wavelength used and it can be optimized. In the case of IR spectroscopy several micrometers are obtained for spatial resolution while Raman spectroscopy is able to analyze a sample at sub micrometer resolution by selecting an appropriate laser wavelength. Therefore cellular components can be studied being not possible with IR spectroscopy.

Additional advantage of Raman spectroscopy is that the observation of vibrational modes is usually independent on the excitation wavelength [14]. Since Raman spectroscopy measures the shift in frequency from the excitation laser, it can be performed using any operating range from UV to NIR. It permits the study of vibrational information with wavelengths ranging from $2 - 200\mu m$

Other advantage of Raman spectroscopy is its specificity. Each molecule has different Raman spectrum, allowing the identification of molecules among a mixture. Also, the intensity of the bands is proportional to the molecular concentration in the sample being studied. Thus, Raman spectroscopy combined with statistical techniques is a promising technique in biomedicine due to its specificity, sensitivity and quantitative information.

1.1 Classical theory of Raman scattering

According to the classical theory Raman scattering can be described as follows [12]. The electric field strength of an electromagnetic wave (incident light) fluctuates with time (t) as:

$$\vec{E} = \vec{E_0} \cos 2\pi\nu_0 t \tag{1.1}$$

where $\vec{E_0}$ is the amplitude of the electric field and ν_0 the frequency of the incident light. When a diatomic molecule is irradiated with this electromagnetic wave, a electric dipole moment $\vec{P}$ is induced :

$$\vec{P} = \alpha\vec{E_0} \cos 2\pi\nu_0 t \tag{1.2}$$

where α is a constant called *Polarizability*.

For an isotropic molecule, $\vec{P}$ and $\vec{E}$ are both in the same direction and thus, polarizability (α) of the molecule is a scalar. However, for non-isotropic molecules the application of an electric field in a given direction induces a dipole moment that in general can have three components in the space, and α becomes a tensor. For non-isotropic molecules, the polarizability along different principal axes of the molecule can be writen as:

$$P_x = \alpha_{xx}E_x + \alpha_{xy}E_y + \alpha_{xz}E_z, \tag{1.3}$$

$$P_y = \alpha_{yx}E_x + \alpha_{yy}E_y + \alpha_{yz}E_z, \tag{1.4}$$

$$P_z = \alpha_{zx}E_x + \alpha_{zy}E_y + \alpha_{zz}E_z, \tag{1.5}$$

where P_x, P_y and P_z are the induced electric dipole moment in x, y and z directions respectively. The tensor α is defined by these nine coefficients $\alpha_{xx}, \alpha_{xy} ... \alpha_{zz}$. However, $\alpha_{xy} = \alpha_{yx}$, $\alpha_{yz} = \alpha_{zy}$ and $\alpha_{xz} = \alpha_{zx}$, hence the tensor α is really defined by six coefficients.

If the molecule is vibrating with a frequency ν_m, the nuclear displacement can be writen like:

$$q = q_0 \cos(2\pi\nu_m t), \tag{1.6}$$

where q_0 is the vibrational amplitude. For a small amplitude of vibration, α is a linear function of the nuclear displacement and we can write:

$$\alpha = \alpha_0 + \left(\frac{\partial\alpha}{\partial q}\right)_0 q + ..., \tag{1.7}$$

being α_0 the polarizability at the equilibrium position, and $\left(\frac{\partial\alpha}{\partial q}\right)_0$ the change of rate of α_0 with respect to the change in q, evaluated at the equilibrium

position. This polaritzability holds for every component in the polarizability tensor. Thus, for a given component of the dipole moment induced:

$$P_x = \left(\alpha_{xx}E_x^0 + \alpha_{xy}E_y^0 + \alpha_{xz}E_z^0\right)\cos\omega_0 t, \tag{1.8}$$

$$\begin{aligned} P_x &= \left(\alpha_{xx}^0 E_x^0 + \alpha_{xy}^0 E_y^0 + \alpha_{xz}^0 E_z^0\right)\cos\omega_0 t \\ &+ \left\{\left(\frac{\partial\alpha_{xx}}{\partial q}\right)_0 E_x^0 + \left(\frac{\partial\alpha_{xy}}{\partial q}\right)_0 E_y^0 + \left(\frac{\partial\alpha_{xz}}{\partial q}\right)_0 E_z^0\right\} \\ &\{q_0\cos\omega_0 t\cos\omega_m t\}, \end{aligned} \tag{1.9}$$

where $\omega_0 = 2\pi\nu_0$ and $\omega_m = 2\pi\nu_m$.

$$\begin{aligned} P_x &= \left(\alpha_{xx}^0 E_x^0 + \alpha_{xy}^0 E_y^0 + \alpha_{xz}^0 E_z^0\right)\cos\omega_0 t \\ &+ \frac{q_0}{2}\left\{\left(\frac{\partial\alpha_{xx}}{\partial q}\right)_0 E_x^0 + \left(\frac{\partial\alpha_{xy}}{\partial q}\right)_0 E_y^0 + \left(\frac{\partial\alpha_{xz}}{\partial q}\right)_0 E_z^0\right\} \\ &\{\cos(\omega_0+\omega_m)t + \cos(\omega_0-\omega_m)t\} \end{aligned} \tag{1.10}$$

The first term represents an oscillating dipole that radiates light of frequency ν_0 (Rayleigh scattering), while the second term corresponds to the Raman scattering of frequency $\nu_0+\nu_m$ (anti-Stokes) and $\nu_0-\nu_m$ (Stokes). From the equation (1.10), if $\left(\frac{\partial\alpha_{xz}}{\partial q}\right)_0$ is zero, the vibration is not Raman active. In other words, the selection rules for a molecular vibration to be Raman active are that the rate of change of polaritzability (α) with respect to the change in q, evaluated at the equilibrium position, i.e. $\left(\frac{\partial\alpha_{xi}}{\partial q}\right)_0$ for i=x,y and z, must not be zero [11, 12]. From quantum mechanics a vibration is IR active if the dipole moment is changed during the vibration. There are some rules for the IR and Raman activities. The *mutual exclusion principle* states that vibrations that are symmetric with respect to the center of symmetry are Raman active but not IR-active, whereas vibrations that are antisymmetric with respect to the center of symmetry are IR-active but not Raman-active. Some vibrations are inherently weak in IR and strong in Raman spectra. In general, covalent bonds are stronger in Raman and ionic bonds stronger in IR. Also water is a weak Raman scatterer but IR spectroscopy suffers from the strong absorption of water. Therefore, aqueus samples can be studied without major interference from water with Raman spectroscopy, signifying a big adventage in the case of biological specimens.

1.2 Quantum theory of Raman scattering

Although classical theory is able to predict Stokes and anti-Stokes scattering, experimentally it is observed that Stokes scattering is stronger than anti-Stokes scattering and classical theory predicts equal intensities for both terms (fig 1.3). Thus, to complete the explanation of Raman scattering it is necessary to use quantum theory [8, 9].

According to quantum mechanics, molecular vibrations are quantized. The scattering process is explained as the creation and annihilation of vibrational excitations (phonons) when a photon interacts with the molecule. In Raman

scattering incident light raises the molecule to a more energetic electronic state (virtual state) from which it returns to a different vibrational level. In the case of Stokes scattering the molecule initially is in the lower vibrational ground state and because of the incident photon raises a virtual state to return to a higher vibrational level. As a result of this process, scattered photons have less energy. On the contrary, in anti-Stokes Raman scattering, molecule initially is in an excited vibrational level and after the interaction with incident photons it returns to the ground vibrational state. As a result, scattered photons will have a higher energy (fig. 1.2)

Since each atom can move in three directions (x, y, z), an N-atom molecule has 3N degrees of freedom of motion [12]. However, the 3N includes six degrees of freedom originating from translational motions of the whole molecule in the three directions and rotational motions of the whole molecule about the three principal axes of rotation, which go through the center of gravity. Thus, the net vibrational degrees of freedom (number of normal vibrations) is Q=3N-6. In the case of linear molecules, it becomes Q=3N-5 since the rotation about the molecular axis does not exist.

In quantum theory the expectation value of the component α_{ij} of the polarizability tensor is given by:

$$\langle \alpha_{ij} \rangle = \int u_b^*(q) \alpha_{ij} u_a(q) dq, \tag{1.11}$$

where the function $u(q)$ represent the molecular eigenfunctions in the initial level a and the final level b. The integration extends over all nuclear coordinates. Therefore the computation of Raman intensity lines requires the knowledge of the molecular wave function of the initial and final states.

In the case of small displacements q_n the molecular potential can be approximated by a harmonic potential and then, the coupling between the different vibrational modes can be neglected [8,9]. The functions $u(q)$ are then separable into a product:

$$u(q) = \prod_{n=1}^{Q} w_n(q_n, v_n) \tag{1.12}$$

of vibrational eigenfunction of the n^{th} normal mode with v_n vibrational quanta, where Q=3N-6 (or 3N-5 for linear molecules) is the number of normal vibrational mode for N nuclei. Using the orthogonality relation:

$$\int w_n w_m dq = \delta_{nm}, \tag{1.13}$$

of the functions $w_n(q_n)$, one obtains from eq. (1.11) and eq. (4):

$$\langle \alpha_{ij} \rangle_{ab} = \alpha_{ij}^0 + \sum_{n=1}^{Q} \left(\frac{\partial \alpha_{ij}}{\partial q_n} \right)_0 \int w_n(q_n, v_a) q_n w_n(q_n, v_b) dq_n, \tag{1.14}$$

where the first term is constant and responsible of the Rayleigh scattering. For non-degenerate vibrations the integrals in the second term vanish unless $v_a = v_b \pm 1$. In this cases its value is $[\frac{1}{2}(v_a+1)]^{1/2}$. The parameter for intensity of the vibrational Raman bands is the derivative $(\partial \alpha_{ij}/\partial q)$, which can be determined

from the Raman spectra.

Quantum theory is able to explain why Stokes Raman scattering is stronger than anti-Stokes [12]. The intensity of a Raman line at a Stokes or anti-Stokes frequency ($\omega_0 \pm \omega_n$) is determined by the population intensity $N_i(E_i)$ in the initial level E_i, by the intensity I_L of the incident pump wave, and by the Raman scattering cross section $\sigma_R(i \to f)$ for the Raman transition $E_i \to E_f$:

$$I_{Raman} = N_i(E_i)\sigma_R(i \to f)I_L. \tag{1.15}$$

The population density, $N_i(E_i)$, follows Maxwell-Boltzmann distribution at thermal equilibrium:

$$N_i(E_i) = \frac{N}{Z} g_i e^{-E_i/kT}, \tag{1.16}$$

where $N = \sum N_i$, the partition function $Z = \sum g_i e^{-E_i/kT}$, and g_i is the statistical weight.

From this equation the proportion of intensity of Stokes line versus anti-Stokes line for a vibrational frequency ω_n will be given by the factor:

$$\frac{I_{Stokes}}{I_{anti-Stokes}} = e^{(-\hbar\omega_n/kT)}. \tag{1.17}$$

The other factor responsible of the intensity of a given Raman band is the scattering cross section. From quantum mechanics it depends on the matrix element of the polarizability tensor [eq. (1.14)] and contains furthermore the ω^4 frequency dependence. It can be written as follow [80]:

$$\sigma_R(i \to f) = \frac{8\pi\omega_s^4}{9\hbar c^4} \left| \sum_j \frac{\langle \alpha_{ij} \rangle \hat{e}_L \langle \alpha_{jf} \rangle \hat{e}_S}{(\omega_{ij} - \omega_L - i\gamma_j)} + \frac{\langle \alpha_{ji} \rangle \hat{e}_L \langle \alpha_{jf} \rangle \hat{e}_S}{(\omega_{jf} - \omega_L - i\gamma_j)} \right|^2, \tag{1.18}$$

where $\hat{e}_L$ and $\hat{e}_S$ are the unit vectors representing the polarization of the incident and the scattered beam respectively. The sum extends over all molecular levels j with homogeneous width γ_j accessible by single-photon transitions from the initial state i. We see from eq. (1.18) that the initial and the final states are connected by two-photon transitions which implies that both states have the same parity.

Raman scattering is a relatively weak process with typical cross section of $10^{-30} - 10^{-25}$ $cm^2/molecule$). Thus, the number of Raman scattered photons is quite small. To overcome this issue and enhance the sensitivity of Raman spectroscopy, some techniques can be used and will be discussed in the following chapters like Resonance Raman scattering and SERS (Surface Enhanced Raman Spectroscopy).

1.3 Resonance Raman spectroscopy

From eq. (1.18) follows that when $\omega_{ij} - \omega_L \to 0$ the scattering cross section grows. This phenomena has been used in the so called Resonance Raman (RR) scattering and it occurs when the exciting line is chosen so that its energy intercepts the manifold of an electronic excited state [12]. As a result RR spectra show extremely strong enhancement of Raman bands originating in this

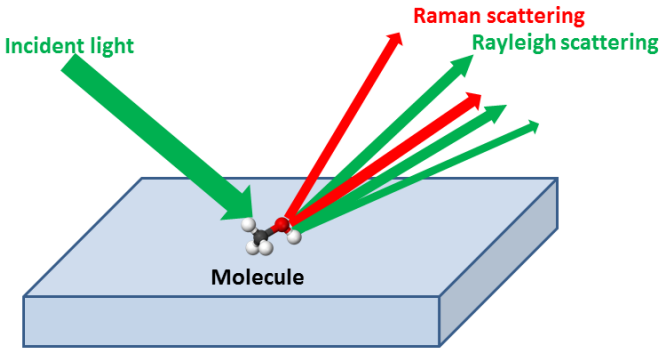


Figure 1.1: Scattered light from a molecule illuminated with a monochromatic light can be Rayleigh (with the same incident frequency) or Raman (with different frequency)

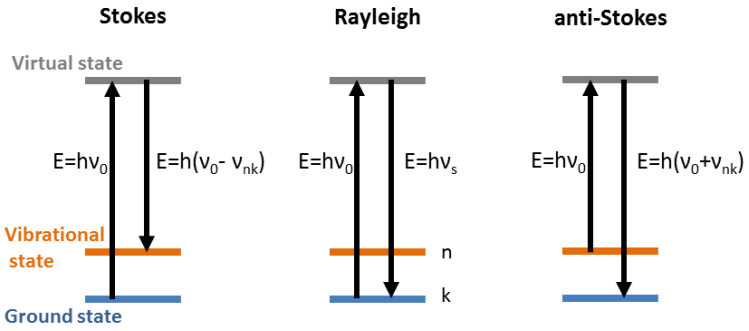


Figure 1.2: Schematic diagram of Raman scattering by molecules.

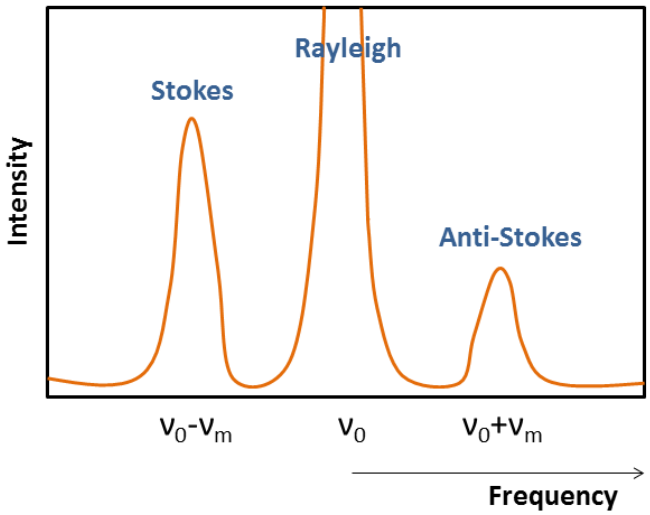


Figure 1.3: Raman spectra with Stokes and anti-Stokes parts.

particular electronic transition (by as much as a factor of 10^6). The lifetime of the excited state in RR is very short ($\approx 10^{-14}$s) compared with the Resonant Fluorescence ($\approx 10^{-8}$ to $\approx 10^{-5}$s).

But RR spectroscopy presents some disadvantages. First, a tunable laser sources is necessary to study different molecules in the sample from the fact that with one excitation line only a molecule can be studied. Second, due to the enhancement of bands of the molecule under study, other molecular signatures become neglectible and can not be investigated in the same experiment. Third, the enhancement factor is still not enough for obtaining single molecule spectra.

1.4 Surface Enhanced Raman Spectroscopy

Although Raman spectroscopy is a very sensitive technique, it is weak. Consequently, biomolecules, that have low Raman cross-section and may exhibit in the spectra high background or fluorescence signal, are difficult to study from the poor visibility of Raman bands. SERS presents a method to enhance the Raman signal [18]. In the course of this work we will use SERS for single molecule studies and to monitor pH changes inside a living cell.

SERS was discovered, but not recognized as such, by Fleischmann et al. [37] in 1974. They observed an intense Raman scattering from the pyridine adsorbed onto a roughened silver electrode surface from aqueous solution. Later, Albrecht and Creighton [38] pointed out that the large intensities observed could not be accounted for simply by the increase in the number of scatterers. This started a research competition to find the mechanisms responsible for the SERS effect. By 1985, the experiments agreed with the essential features proposed as the mechanisms.

SERS has been observed for a very large number of molecules adsorbed on

the surfaces of some metals with different morphologies and physical environments. The largest enhancements achieved were produced in rough surfaces at the nanoscale ($10 - 100nm$). Some examples are electrode surfaces, films deposited by evaporation or sputtering and colloids.

SERS differs from ordinary Raman spectra in some aspects. For instance intensities of the bands observed generally fall off with increasing vibrational frequency. Mechanisms proposed to explain SERS effect accounts for the experimental facts observed. The two mechanisms that are still considered as valid to explain SERS are the electromagnetic and chemical [39]. From their names, the former focus on the enhanced electromagnetic fields which can be produced on metal surfaces with appropriate morphologies and the latter on changes in the electronic structure of molecules which occur upon adsorption and which can lead to resonance Raman scattering.

1.4.1 Electromagnetic enhancement

A plasmon is the collective excitation of the electron gas of a conductor [18]. When the excitation is confined in the near surface region it is called a surface plasmon. Surface roughness or curvature is necessary for the excitation of surface plasmons by light. An electromagnetic field of light at the surface can be enhanced under conditions of the surface plasmon excitation. The amplification of both the incident optical field and the scattered Raman field through their interaction with the surface constitutes the electromagnetic SERS mechanism.

The physics that governs the electromagnetic mechanism can be easily explained for a metal sphere in an external electric field. For a spherical particle whose radius is much smaller than the wavelength of light, the electric field is uniform across the particle and the electrostatic approximation is valid [39]. The field induced at the surface of the sphere is related to the applied, external (laser) field by eq. (1.19)

$$E_{induced} = [\epsilon_1(\omega) - \epsilon_2]/[\epsilon_1(\omega) + 2\epsilon_2]E_{laser}, \tag{1.19}$$

where $\epsilon_1(\omega)$ is the complex, frequency dependent dielectric function of the metal and ϵ_2 is the relative permitivity of the ambient phase. From the equation the function is resonant in the frequency that $Re(\epsilon_1) = -2\epsilon_2$. The excitation of the surface plasmon highly increase the local field experienced by a molecule adsorbed on the surface of the particle.

1.4.2 Chemical enhancement

Some evidences suggested that there is a second enhancement mechanism independent from the electromagnetic mechanism which produces multiplicative effects when is operative [18]. Chemical enhancement can be explained by a resonance Raman mechanism in which new electronic states arised from the chemisorption serve as resonant intermediate states in Raman scattering. Sometimes the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of the adsorbate are symmetrically disposed in energy with respect to the Fermi level of the metal. In this situation, charge-transfer excitations (from the metal to the molecule or viceversa) can occur at about half of the energy of the intrinsic intra-molecular excitations of the adsor-

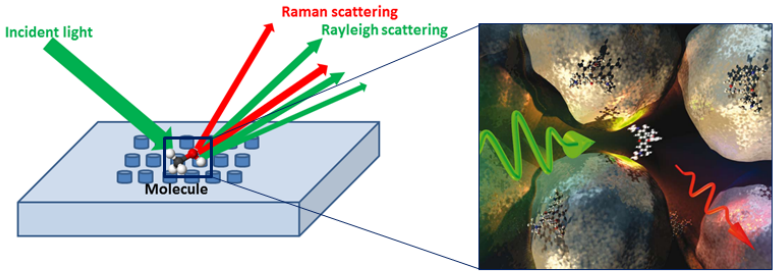


Figure 1.4: Surface Enhanced Raman scattering. Adapted image from [40]

bate. Molecules commonly studied by SERS, that normally have their lowest electronic excitations in the near ultraviolet would experience charge transfer excitations according to this simple model in the visible region of the spectrum.

2

Statistical Methods for Processing of Raman spectra

2.1 Principal Component Analysis

Principal Component Analysis constitutes in many ways the basis for multivariate data analysis. PCA provides an approximation of a data matrix Y in terms of the product of two smaller matrices R and L' which capture the essential data patterns of Y. Plotting the colums of R (named *scores*) gives a picture of the dominant *object patterns* of Y and, analogously, plotting the rows of L' (*loadings*) shows the complementary *variable patterns*. In the case of an spectroscopic data matrix the object will be samples measured and the variables the Raman shift or wavenumber.

Pearson [41] was the fist formulating PCA. Later, Fisher and McKenzie [42] mention it as more suitable than analysis of variance for the modelling of response data. They also outlined the NIPALS algorithm. Hotelling [43] further developed PCA to its present stage. Some good books and tutorials on PCA are [44], [45]. In chemistry, PCA was introduced by Malinowski around 1960 with the name Factor Analysis and after 1970 a large number of chemical applications were explored.

In multivariate data analysis we have a Y matrix with n rows termed *objects* and p colums called *variables*. The main goals of PCA are concerned with finding relationships between objects. For example in finding classes of similar objects. Related with this is the detection of outliers in the experimental data set. Another goal could be the data reduction, useful for large amounts of data with a complex model structure. PCA can be used to build a model of how a physical or chemical system behaves, and this model can be used for prediction when new data are measured for the same system. PCA can also be used to unmix the effects of a mixture of components. PCA, furthermore, estimates the correlation structure of the variables. The importance of a variable in a PC is

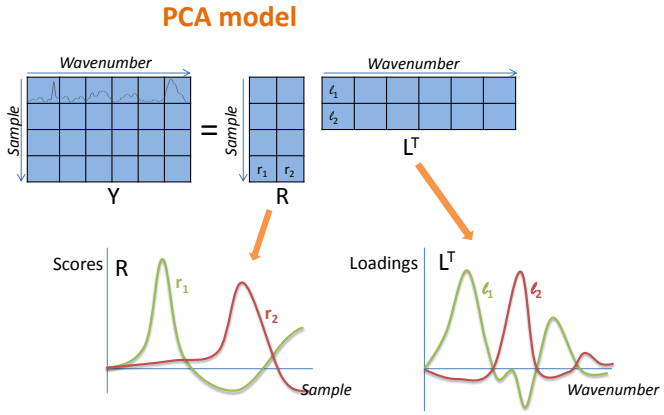


Figure 2.1: PCA provides an approximation of a data matrix Y in terms of the product of two smaller matrices R and L' which capture the essential data patterns of Y

indicated by the size of its variance.

A data matrix Y with n objects and p variables can be represented as an ensemble of n points in a p dimensional space. A p-space is difficult to visualize when p is bigger than 3 and the inner data structure is therefore difficult to reveal. For this reason, we need to apply PCA to this type of complex data structure as an spectroscopic experimental data matrix.

First, let's consider the spectra as a column vector $\vec{y_i}$:

$$\vec{y_i} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ij} \\ \vdots \\ y_{ip} \end{pmatrix}$$

where p is the last wavenumber pixel recorded from the CCD camara of the spectrometer and the intensity for each wavenumber is stored in rows. The transform of this vector is:

$$\vec{y_i}' = \begin{pmatrix} y_{i1} & y_{i2} & \cdots & y_{ij} & \cdots & y_{ip} \end{pmatrix}$$

Then, it is possible to construct a matrix with all spectra adquired in an experiment stored in rows:

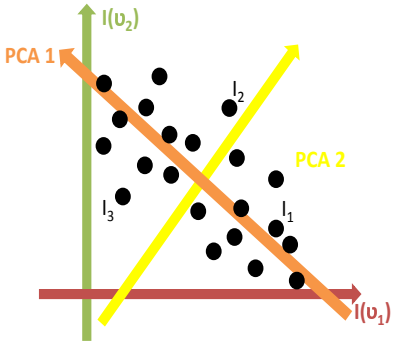


Figure 2.2: Simple diagram showing the Principal Components of a data set representation. The dimensionality of the data could be reduced by selecting the PC1 axis

$$Y = \begin{pmatrix} \vec{y_1}' \\ \vec{y_2}' \\ \vdots \\ \vec{y_i}' \\ \vdots \\ \vec{y_n}' \end{pmatrix} = \begin{pmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,j} & \cdots & y_{1,p} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,j} & \cdots & y_{y,p} \\ \vdots & \vdots & & & & \vdots \\ y_{i,1} & y_{i,2} & \cdots & y_{i,j} & \cdots & y_{i,p} \\ \vdots & \vdots & & & & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,j} & \cdots & y_{n,p} \end{pmatrix} \tag{2.1}$$

where n is the number of spectra adquired. Then, $y_{i,j}$ represents the Raman intensity for the ith spectra acquired at the jth wavenumber.Thus, the data points can be represented in a set of coordinates were each axes are the intensity for a given wavenumber and, by the inherent wavenumber relationships encoded in the Raman spectra of a mixture of molecules, the representation will have priviledge directions. By finding this priviledge directions of the swarm of points we could get insight into the relationships between wavenumbers that are the responsible for the major variance in the data. By selecting those axes as principals we could simplify the representation by selecting only the priviledge directions that retain the maximum variance but without loosing relevant information.

The aim of Principal Component Analysis is to identify the most meaningful basis to re-express the dataset. With the new set of coordinates the hidden structure of the data (relationships between wavenumbers) will be revealed and retaining only the components that represent the major variance in the data, the noise will be neglected. From algebra, the axes can be rotated as:

$$\vec{z_i} = A\vec{y_i}, \tag{2.2}$$

where A is an orthogonal matrix and $\vec{z_i}$ is the spectra $\vec{y_i}$ represented in the new set of coordinates.

Hence, Principal Component Analysis consists in calculating the orthogonal matrix A that rotates the axis in the way that they line up with the natural extensions of the group of data points. But to calculate this matrix, first, let's present the covariance matrix S_Y of matrix Y:

$$S_Y = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1k} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2k} & \cdots & s_{2p} \\ \vdots & \vdots & & & & \vdots \\ s_{j1} & s_{j2} & \cdots & s_{jk} & \cdots & s_{jp} \\ \vdots & \vdots & & & & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pk} & \cdots & s_{pp} \end{pmatrix} \tag{2.3}$$

where $s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(y_{ij} - \overline{y}_j)(y_{ik} - \overline{y}_k)$ and $\overline{y}_j = \frac{1}{p}\sum_{i=1}^{p}(y_{ij})$ and $\overline{y}_k = \frac{1}{p}\sum_{i=1}^{p}(y_{ik})$.

The eigenvectors and eigenvalues of S_Y can be calculated from the following condition:

$$S_Y\vec{y_j} = \lambda\vec{y_j}; \qquad (S_Y - \lambda I)\vec{y_j} = 0 \tag{2.4}$$

and the only non trivial solutions are for

$$\det(S_Y - \lambda I) = 0 \tag{2.5}$$

Thus, the rotation of the initial axis to line up with the natural extensions of the experimental data points will be translated in the fact that the new variables $z_{i1}, z_{i2}..., z_{ip}$ (Principal Components) need to be uncorrelated and thus, the sample covariance matrix of Z must be diagonal:

$$S_Z = \begin{pmatrix} (s_z)_{11} & 0 & \cdots & 0 \\ 0 & (s_z)_{22} & \ddots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & (s_z)_{pp} \end{pmatrix}$$

Furthermore, from algebra,

$$C'S_YC = D = diag(\lambda_1, \lambda_2, ..., \lambda_p),$$

where the λ_i are the eigenvalues of S and C is an orthogonal matrix whose columns are the normalized eigenvectors of S_Y. Thus, from

$$S_Z = AS_YA' \Longrightarrow A = C'$$

where the A rows are the normalized eigenvectors of S_Y and the diagonal elements of $S_Z = AS_YA'$ are the eigenvalues of S_Y,

$$(s_z)_{ii} = \lambda_i \tag{2.6}$$

Hence, as a result, the steps for calculating the Principal Components z_i, can be reduced to first calculate S_Y and its eigenvalues and eigenvectors, and finally construct A, which is formed by the eigenvectors of S_y displayed in rows. The new $\vec{z_j}$ are calculated as $\vec{z_j} = A\vec{y_j}$

From 2.6, the eigenvalues of S_Y are the sample variances of the principal components. And, since the rotation follows the natural extension of the group of points, $\vec{z_1}$ retains the largest variance and $\vec{z_p}$ the smallest.

In summary, PCA model could be writen as:

$$Y = RL' + E \tag{2.7}$$

PCA projects the Y matrix into a lower dimensional subspace by means of the projection matrix L' and the matrix R. The later gives the sample coordinates in this new subspace and L' rows comprise the direction coefficients of the Principal Components hyperplane. The colums of R are the *scores* and the rows of L' are called the *loadings*. L' are the eigenvectors displayed in rows ($L' = A$) and the scores $R = YA'$. The deviations between projections and the original coordinates are termed the residuals and are collected in matrix E.

2.2 Partial Least squares - discriminant analysis (PLS-DA)

PLS-DA is a supervised classification method in which knowledge of the sample is included. A matrix X is constructed with the information about the class membership of samples. Like in PCA, matrix Y contains each sample spectra. PLS-DA employs the fundamental principle of PCA but further rotates the components (latent variables, LVs) by maximizing the covariance between the spectral variation and group affinity so that the LVs explain the diagnostically relevant variations rather than the most prominent variations in the spectral dataset (Fig. 2.3).

The performance of the PLS-DA diagnostic algorithm can be validated using a cross validation method. Thus, the PLS-DA analysis includes two steps: calibration and validation. In the calibration step, the variations in the spectra correlated to the variations in the class values are extracted and then used to calculate the predicted concentration. In the validation step, the validity of the model is verified by adding new data and comparing the predicted values with the reference values. Although the best validation of the model is performed with new data, an independent and representative validation set is rare. In the absence of a real validation set, a commonly used solution is cross-validation (CV), which simulates the accuracy in detection of new data. CV divides the data into a number of groups and develops a number of parallel models obtained by deleting one group from the data. Having a number of groups similar to the number of samples is not recommended [46]. For this reason, we used a venetian blinds methodology for the CV, consisting in the following: the number of data splits s is selected, and then each test set is determined by selecting every s^{th} object in the data set, starting at objects numbered 1 through s.

During the calibration, it is necessary to determine the correct complexity of the model. In selecting the number of PLS components, there is the risk of "over-fitting" by using too many components and getting a well fitting model with small predictive power. Thus, it is necessary to test the predictive significance of each PLS component and stop when the components start to be non-significant. CV is a practical and reliable tool to check the predictive significance. In this study, we selected the number of components for each model using this methodology.

The model accuracy was checked by calculating the following statistical parameters. First, the RMSE (root mean square error), which characterizes how

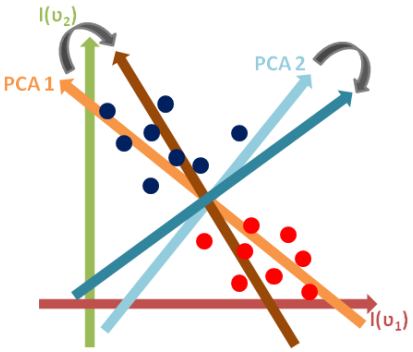


Figure 2.3: Simple diagram showing the Principal Components of a data set representation and the rotation of the PCs axis to achieve the goals of PLS-DA

well the model fits the data used for the calibration, was calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_{ical} - y_{iref})^2}{N}}, \tag{2.8}$$

where y_{ical} is the calculated (predicted) value for a sample included in the model, y_{iref} is the reference value, and N is the total number of samples used in the calibration. Another parameter is the root mean square error of the cross validation (RMSECV), which measures the ability of the model to predict the parameters of the samples that were not used in the calibration. The RMSECV is also defined using equation 2.8, where the y_{ical} are the predictions for the samples not included in the calibration model. To evaluate the model, multiple correlation coefficients (R^2) [46] between the actual and predicted values were calculated. A plot of the RMSECV and RMSE vs. the number of PLS components helps to select the number of components. When the RMSECV was close to the RMSE, the corresponding number of PLS components was chosen.

Finally, parameters as sensitivity and specificity can be extracted to asses the ability of Raman spectroscopy to discriminate between two different types of samples.

2.3 Multivariate Curve Resolution

Multivariate resolution methods represent a set of mathematical tools that can be applied to the analysis and interpretation of spectroscopic data recorded in the monitorization of a chemical or physical process with multichannel detectors. The goal of resolution methods is the extraction of the underlying chemical and physical information from the experimental data set. Nowadays, by means of the technological advances, it is possible to acquire large data sets in short periods of time. In the case of Raman spectroscopy the availability of rapid and reliable detectors allows the monitoring of biochemical reactions or biophysical processes by obtaining a complete spectrum in each point of the process. In the past, only single or few spectroscopic signals (for instance, absorbance) could

be read at each step.

Some techniques like PCA have been used for the analysis of these multidimensional data sets. PCA reduces the dimensions of the experimental data matrix without a significative loss of information. PCA is a useful tool to estimate the number of the most significative components and then neglect the ones that provides noise or non-informative portions. However, the optimization criterium is to find the maximum variance and orthogonal constrains are applied. As a result, PCA in general do not provide physical or chemical meaningful signatures (spectra) and relative concentration profiles of the molecular constituents in the sample. PCA loadings can have negative portions (fig.2.1) therefore not representing a pure molecular Raman spectrum.

Multivariate Curve Resolution (MCR) has been proposed and applied to resolve multiple pure responses and concentrations of the components present in unknown mixtures [47]. It has been used to analyze multicomponent chemical systems like chemical reactions [48], spectroscopic mixtures [49], environmental monitoring data [50] and it can be applied to many other mixture analysis. In MCR, the measured analytical signals are assumed to follow a generalized bilinear additive model (like the extension of Beer- Lambert law in absorption spectroscopy). The contribution of each component to the measured signal depends on its concentration and on its own spectral sensitivity response (pure spectrum). MCR, then, can be applied to obtain quantitative information and provides physical and chemical meaningful solutions. All this is achieved by means of the flexibility of the algorithm to include initial constrains. Altough the possibility of the introduction of the constrains give more chemically undestandable results, MCR is a data-driven method which means that no prior knowdlege (chemical or mathematical) is needed, compounds have a measurable signal and a different (uncorrelated) concentration/response profiles among them. Spectroscopic data is bilinear by nature and therefore MCR provides a very powerful methodology to deconvolve from the spectral data set acquired real chemical profiles (spectra).

MCR has been used in Raman spectroscopy to monitor inorganic chemical reactions and images as [35,51]. In all previous Raman spectroscopy applications pure molecular components in the sample had good and distinct Raman spectra. However, biological Raman spectra are complex and lot of different molecules have common spectral features. Therefore, monitoring biological processes in living tissue or cells is challenging and MCR is a good candidate technique to deconvolve the pure molecular structures that change during the biochemical process studied. Recently, MCR-ALS has been applied to image biological samples with Raman spectroscopy [52] but there are not applications of MCR to monitor or study biochemical process in tissue or cells using Raman spectroscopy. However this field opens a new way to improve the analyis of biological specimens with Raman spectroscopy.

There are several algorithms to perform MCR method: the non iterative algorithms in which the output matrices C and S (explained below) are obtained in a single calculation and rely on mathematical information (Heuristic Evolving Latent Projections (HELP) [53] and SFA) and the iterative algorithms in which C and/or S are obtained from an iterative optimization procedure and rely on chemical and mathematical knowledge (Iterative Target Transformation Factor Analysis (ITTFA) [54], RFA and MCR-Alternating Least Squares (MCR-ALS)). We will focus on MCR-ALS algorithm [36] because of its flexibility to include

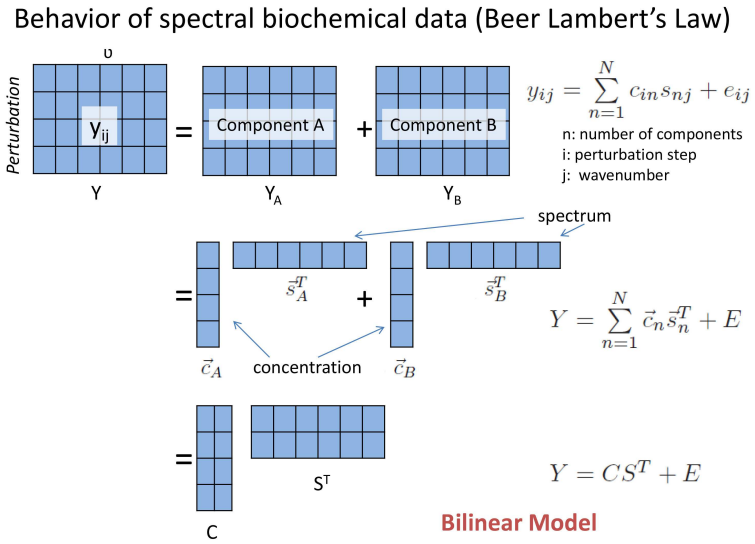


Figure 2.4: Graphical description of the construction of the data matrix used for MCR analysis from the experimental spectroscopic data. Spectra adcquired at different perturbations of the system are displayed in rows and form the Y matrix. This Y matrix then, can be decomposed into several components following a linear combination. A bilinear model is assumed from the nature of the spectroscopic data (Beer Lambert's law).

initial estimates and constrains in the iterative resolution procedure.

2.3.1 MCR-ALS method

Data obtained from the spectroscopic monitoring of a process when some external parameter (perturbation) is applied in each acquisition, can be grouped in a data matrix Y, whose n rows contain the spectra recorded during the process and whose p columns are the wavenumber (Raman shift) (fig. 2.4). The MCR-ALS method is used to decompose the hyperspectral dataset into the signatures or pure spectra of the molecular sample constituents and their concentration (relative amounts) on each sample or acquisition. MCR-ALS is based on a bilinear model which assumes that the observed spectra are a linear combination of the spectra of the pure components in the system. This model can be written in matrix form as:

$$Y = CS^T + E \tag{2.9}$$

where C is a $n*m$ matrix where m is the number of pure components deconvolved and represents the relative amounts or concentrations and ; S^T is a $m * p$ matrix and contains the pure spectra; and E is the matrix associated to noise or experimental error. Fig.2.5 shows a graphical representation of the MCR decomposition of matrix Y into pure components S^T and its contributions C in each acquired spectrum. An element of matrix Y can be represented by

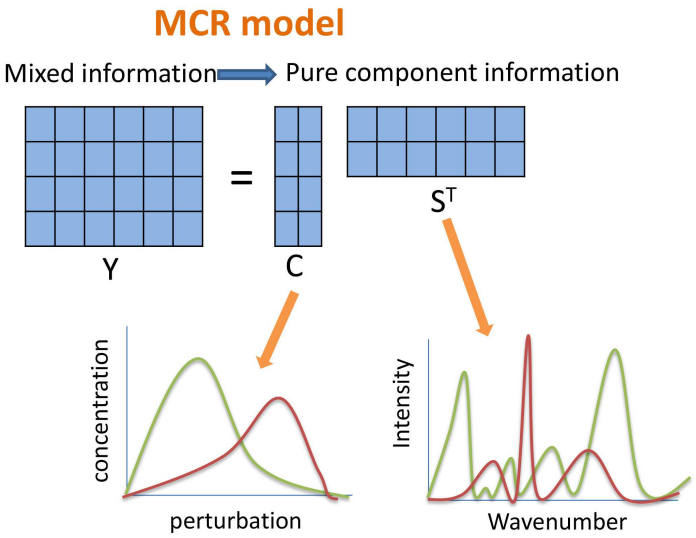


Figure 2.5: MCR provides a decomposition of the spectral data matrix Y in terms of the product of two smaller matrices C and S^T with information about the underlying molecular pure components that are changing during the perturbation

$y_{ij} = \sum_{n=1}^{N} c_{in}s_{nj} + e_{ij}$ where y_{ij} is the Raman intensity for the ith spectrum at the jth wavenumber; c_{in} is the concentration at the ith sample measured of the nth component deconvolved; s_{nj} is the intensity of the nth component at the wavenumber jth and e_{ij} is the residual error at the ith spectrum and jth wavenumber.

The decomposition of a matrix, however, is subjected to ambiguities [55]. Many sets of paired C and S^T type matrices can reproduce the original data matrix with the same fit quality. Specifically, the response profiles that reproduce the original matrix can differ in shape (*rotational ambiguity*) or in magnitude (*intensity ambiguity*) from the real one. This two ambiguities come from the fact that equation 2.9 can be transformed as:

$$D = C(PP^{-1})S^T; D = (CP)(P^{-1}S^T); D = C^*S^{*T}$$

where $C^* = CT$ and $S^{*T} = (T^{-1}S^T)$ describe the D matrix with the same fit quality as the true C and S^T matrices. This ambiguity is called *rotational ambiguity* and indicates that resolution methods will provide as many C and S matrices as P matrices exist. Thus, if C and S are not forced to obey certain conditions, solutions for those matrices are infinite. Appart from *rotational ambiguity*, eq.2.9 can be rewritten as:

$$Y = \sum_{i=1}^{N} \frac{c_i}{k_i}(k_i s_i^T); Y = C^*S^{*T}$$

where k_i are scalars and the concentrations profiles of the new C^* matrix would have the same shape as the real ones, but being k_i time smaller, whereas new S^* matrix would be shaped like the S spectra, but k_i times more intense. This ambiguity is named *intensity ambiguity*. All these ambiguities will be minimized by introducing constrains to the MCR algorithm.

MCR-ALS solves iteratively eq.2.9 using an Alternating Least Squares algorithm which calculates C and S^T matrices by fitting optimally the experimental data matrix Y. The optimization is performed for a given number of components previously fixed and using initial estimates for C or S^T.

For obtaining the number of components necessary to explain the variance in the data, it is possible to use PCA. In PCA decomposition the variance in the data is explained with a set of Principal components each of which having an associated eigenvalue. The magnitude of each eigenvalue is related with the variance of the overall data that explain the associated particular PC. Thus, by ploting the eigenvalues of each PC it will be possible to distinguish which components describe most of the variance in the data and which ones can be neglected. The number of relevant Principal components will be equivalent to the number of components needed for MCR decomposition.

Some methods can be used to obtain the initial estimates: Evolving Factor Analysis [56] or SIMPLISMA [57].

During the ALS optimization, several constraints can be introduced for C and S^T profiles such as non-negativity, unimodality or closure. Some of them are clearly useful for Raman spectroscopy like the non-negativity of spectra and concentrations. Convergence is achieved when in two consecutive iterative cycles, relative differences in standard deviation of the residuals between experimental and calculated ALS data values are less than a previous selected number, normally chosen as 0.1%.

To asses the quality of the results some figures of merit of the optimization procedure can be calculated. The *lack of fit* is defined as the difference between the experimental data matrix Y and the data reproduced by the CS^T product calculated with MCR-ALS. The mathematical expression is:

$$LOF = 100 \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} e_{i,j}^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} y_{i,j}^2}} \tag{2.10}$$

where $y_{i,j}$ designs an element of the experimental data matrix Y and $e_{i,j} = y_{i,j} - \hat{y}_{i,j}$ is the related residual obtained from the difference between the original element $y_{i,j}$ and the MCR-ALS reproduction $\hat{y}_{i,j}$. Other figure of merit is the percent of variance explained:

$$R^2 = 100 \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} y_{i,j}^2 - \sum_{i=1}^{n}\sum_{j=1}^{m} e_{i,j}^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} y_{i,j}^2} \tag{2.11}$$

As a result, MCR algorithm is constructed in four main steps: the determination of the number of components, the construction and introduction of the

initial estimates (for C and S), the iterative Least Squares calculation of C and S subject to constrains and finally check for satisfactory CS^T data reproduction.

2.4 2D correlation Raman spectroscopy

The idea of 2D correlation spectroscopy emerged in the field of NMR to simplify the visualization of complex spectra consisting in many overlapped bands by spreading spectral peaks over the second dimension [58]. Although traditionally and in the present 2D correlation spectroscopy has been dominated by NMR and other resonance spectroscopy methods, nowadays the applicability in other types of spectroscopic techniques is increasing [59]. The introduction of 2D spectroscopy to IR and Raman spectroscopy occurred much later than NMR. In 1986 Noda presented the concept of *perturbation-based* two dimensional spectroscopy applicable to infrared (2D IR) [60]. From that time the technique is establishing itself as a powerful general tool for the analysis of spectroscopic data.

2D correlation technique is based in the fact that when a certain perturbation is applied to a sample, the chemical composition of the sample under study is altered and varies depending on the perturbation exerted. This perturbation-induced changes are translated then into spectral changes when sample is monitored with an electromagnetic probe and therefore a *dynamic spectra* is obtained (fig. 2.6). 2D correlation spectra consists in two orthogonal components, the synchronous and asynchronous correlation spectrum that provide useful and complementary information to analyze the intensity changes, band shifts, band shapes changes..., typical spectral variations observed under external perturbation.

Among the main advantages of 2D correlation spectroscopy, four stand out: (i) simplification of complex spectra consisting in many overlapped bands and enhancement of the spectral resolution by spreading peaks over the second dimension; (ii) identification of ambiguous assignments thanks to the correlation of bands in the 2D spectra and study relationships between molecules during the dynamic process; (iii) probe the order in which spectral changes occur through the asynchronous spectra; (iv) the *heterospectral* correlation (2D correlation analysis between two different spectroscopic techniques) permits the identification of correlation among bands from different types of spectroscopy. Furthermore the applicability of generalized 2D correlation is not limited to any type of spectroscopy with a specific electromagnetic probe. It has been used in IR, NIR, Raman UV-vis, fluorescence and circular dichroism (CD). Also, 2D correlation is not restricted to the type of perturbation that induces spectral changes.

2.4.1 Principle of Two-dimensional Correlation Spectroscopy

In this section the fundamental concept of 2D correlation spectroscopy will be presented. The formal mathematical procedure to generate 2D correlation spectra, the basic properties of synchronous and asynchronous spectra and some conclusions about characteristic patterns in 2D maps will be discussed.

The schematic description of the basic planning for a 2D correlation experiment is shown in fig. 2.6. In spectroscopic techniques like Raman spectroscopy,

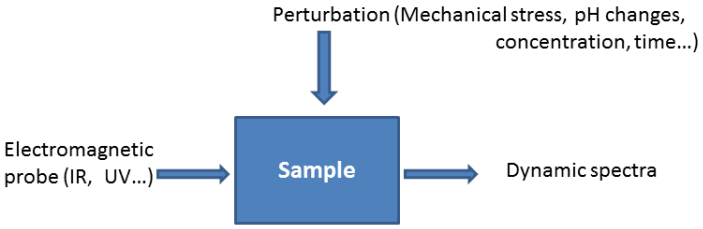


Figure 2.6: Experiment flow used for 2D correlation spectroscopy.

the interaction between an electromagnetic probe and the sample is represented in the form of a spectrum that contains information about the system being analyzed. In 2D correlation spectroscopy the effect of an additional external perturbation exerted on the system is studied from the changes in the spectrum. The external perturbation may cause some physical changes in the sample and it will we translated in distinctive modifications in the measured spectrum. This spectral variations induced by an applied perturbation is referred to as a *dynamic spectrum*.

In the generalized 2D correlation spectroscopy scheme, series of perturbation-induced dynamic spectra are acquired by modifying the external parameter in a systematic way. After, this set is transformed into a set of 2D correlation spectra by cross-correlation analysis. This form of analysis helps to the identification and interpretation of spectral variations not available from or at least not apparent in the original set of conventional 1D spectra. This is the major motivation behind 2D correlation analysis.

The scheme presented in fig. 2.6 is very general and the forms of perturbations that affects the system can be very broad. It can be physical: mechanical, electrical, thermal, magnetic, chemical, acoustic, etc. and the *waveform* of the perturbation can be selected freely. In fact, any spectroscopic experiment, which uses an external perturbation to generate some form of dynamic spectra, is a good potential candidate for benefiting from 2D correlation analysis. In this context a time-evolution study of spectral signatures is also a good candidate for 2D correlation analysis. Because historically 2D correlation analysis comes from statistical series analysis, some traditional terminology like dynamic spectra will be used to describe the perturbation-induced spectral changes, even if the other types of perturbations are considered.

2.4.2 Generalized two-dimensional correlation

Dynamic spectrum

Let us consider a spectral intensity $y(\nu, p)$ which is varying with a perturbation induced in the system and observed during a fixed interval of the external variable p from P_{min} to P_{max}. As it was mention before, p can be any reasonable measure of a physical quantity such as temperature, pressure, concentration, voltage, ph, etc. that is inducing a change in the spectral intensity. It is also

possible to include time as external variable p. The variable ν is any spectral index used in spectroscopy: Raman shift, wavenumber, wavelength, etc.

The dynamic spectrum $\tilde{y}(\nu, p)$ is defined then as [59]:

$$\tilde{y}(\nu, p) = \begin{cases} y(\nu, p) - \overline{y}(\nu) & \text{for } P_{min} < p < P_{max} \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

where $\overline{y}(\nu)$ is the *reference spectrum* of the experiment. Although the selection of the proper reference spectrum is not specified, in most cases the *stationary or averaged* spectrum is selected:

$$\overline{y}(\nu) = \frac{1}{P_{max} - P_{min}} \int_{P_{min}}^{P_{max}} y(\nu, p) dp \tag{2.13}$$

For some applications the reference spectrum can be chosen as some fixed reference point in our experiment like the first spectrum, or the last one. Without any prior knowledge of the system, the selection of the averaged spectrum provides a robust form for correlation analysis.

2.4.3 Two-dimensional correlation concept

The main concept behind 2D correlation spectroscopy is to quantitatively compare the patterns of spectral intensity variations along the external variable p at two different spectral variables, ν_1 and ν_2, over the interval between P_{min} and P_{max} [59]. The 2D correlation spectrum can be expressed as:

$$X(\nu_1, \nu_2) = \langle \tilde{y}(\nu_1, p) \cdot \tilde{y}(\nu_2, p') \rangle \tag{2.14}$$

The symbol $\langle\rangle$ signifies the cross-correlation function that compare the intensity patterns at two different spectral variables ν_1 and ν_2 along p. In other words, the intensity of 2D correlation spectrum $X(\nu_1, \nu_1)$ represents the quantitative measure of a comparative similarity or dissimilarity of spectral intensity variations $\tilde{y}(\nu, p)$ measured at two different spectral variables ν_1 and ν_2 during a fixed interval along p.

$X(\nu_1, \nu_1)$ can be treated as a complex number function to simplify the mathematical manipulation, with two orthogonal (real and imaginary) components, called *synchronous* and *asynchronous* 2D correlation intensities.

$$X(\nu_1, \nu_2) = \Phi(\nu_1, \nu_2) + i\Psi(\nu_1, \nu_2) \tag{2.15}$$

The *synchronous* 2D correlation intensity $\Phi(\nu_1, \nu_2)$ represents the similarity or coincidental intensity changes at two different spectral variables along the value p from P_{min} to P_{max}. Complementarily, the *asynchronous* 2D correlation intensity $\Psi(\nu_1, \nu_2)$ can be understood as a measure of the dissimilarity or *out-of-phase* character of the spectral intensity variations.

It is important to point out that real and imaginary parts of the 2D correlation spectrum $X(\nu_1, \nu_2)$ were termed *synchronous* and *asynchronous* historically because first conceptual developments were applied to time-series analysis. However, as p could be any other physical variable these terms could be called for instance: *synthermal* or *asynbaric*.

Generalized two-dimensional correlation function

The generalized 2D correlation function is defined like [59]:

$$\Phi(\nu_1,\nu_2)+i\Psi(\nu_1,\nu_2)=\frac{1}{\pi(P_{max}-P_{min})}\int_0^\infty \tilde{Y}_1(\tau)\cdot\tilde{Y}_2(\tau)d\tau \tag{2.16}$$

Where $\tilde{Y}_1(\tau)$ is the forward Fourier transform of the spectral intensity variations $\tilde{y}(\nu_1,p)$ observed at ν_1 with respect to the external variable p and can be expressed as:

$$\tilde{Y}_1(\tau)=\int_{-\infty}^{\infty}\tilde{y}(\nu_1,p)e^{-i\tau p}dp=\tilde{Y}_1^{Re}(\tau)+i\tilde{Y}_1^{Im}(\tau) \tag{2.17}$$

being $\tilde{Y}_1^{Re}(\tau)$ and $\tilde{Y}_1^{Im}(\tau)$ the real and imaginary components of the Fourier transform respectively and τ the individual frequency component of the variation of $\tilde{y}(\nu_1,p)$ traced along the external variable p. The conjugate Fourier transform of $\tilde{y}(\nu_2,p)$ at ν_2 is given by:

$$\tilde{Y}_2^{*}(\tau)=\int_{-\infty}^{\infty}\tilde{y}(\nu_2,p)e^{+i\tau p}dp=\tilde{Y}_2^{Re}(\tau)-i\tilde{Y}_2^{Im}(\tau) \tag{2.18}$$

2.4.4 Properties of 2D correlation spectra

Results of 2D correlation analysis are represented by two contour maps plotting the intensity of synchronous and asynchronous 2D correlation spectra respectively. It is worthy to understand some characteristic properties of these maps to extract the potentially useful information encoded otherwise inaccessible.

Synchronous 2D correlation spectrum

Fig. 2.7 shows schematically the intensity of a synchronous 2D correlation spectrum $\Phi(\nu_1,\nu_2)$ which represents the simultaneous or cointidential changes of two separate spectral intensity variations at ν_1 and ν_2 (in the case of Raman spectroscopy at two different wavenumbers) during the interval between P_{min} and P_{max}.

The synchronous spectrum is symmetric with respect to the diagonal line ($\nu_1=\nu_2$) and correlation peaks appear at the diagonal and off-diagonal positions. The intensity of the peaks along the diagonal are mathematically the autocorrelation function of spectral intensity variations occured from P_{min} to P_{max}. These peaks are refered as *autopeaks* and the spectrum along the diagonal that they form is called the *autopower spectrum*. In figure 2.7 four autopeaks are present (A, B, C and D). The intensity of an autopeak represents the overall extent of the changes occured for the spectral intensity at that particular spectral variable ν_1. It is allways positive and then permits to identify which regions in the spectra are more susceptible to change with the external perturbation p.

On the other hand, the *cross peaks* are located off-diagonal and can be positive or negative. They represent the simultaneous or coincidental changes of spectral intensities at two different spectral variables ν_1 and ν_2. The fact that two spectral regions are *synchronously* correlated suggest interesting properties of the molecular content of our sample. For instance two molecules that experience the same changes in concentration during the process observed, will

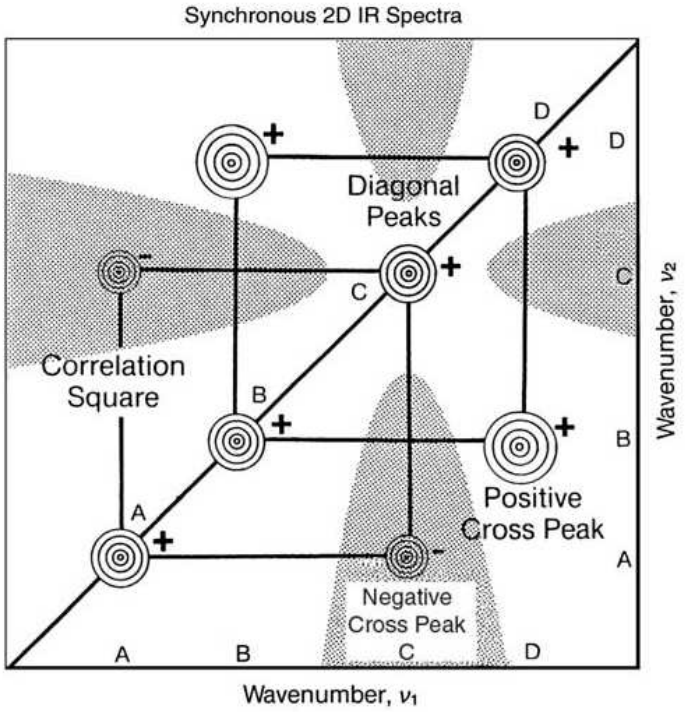


Figure 2.7: Simplified scheme of a synchronous 2D correlation contour map. Shaded areas indicate negative correlation intensity. Adapted image from [59]

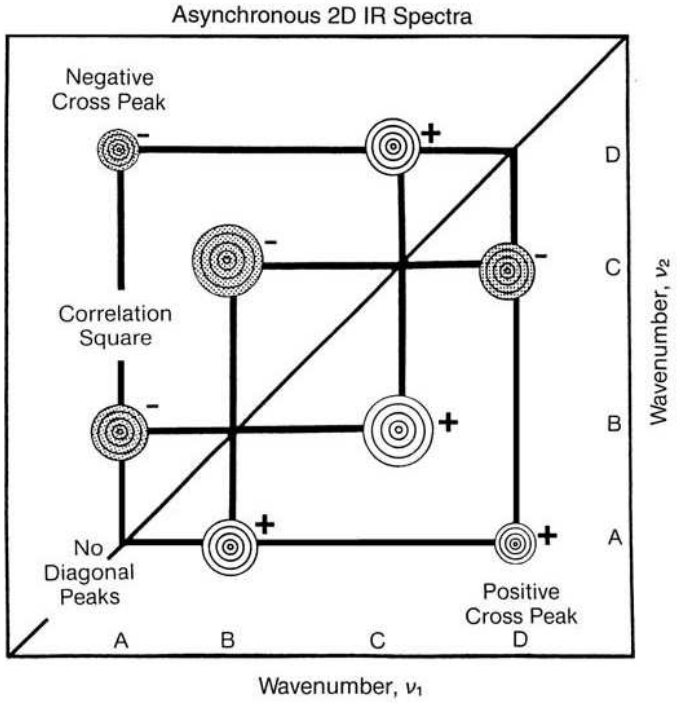


Figure 2.8: Simplified scheme of an asynchronous 2D correlation contour map. Shaded areas indicate negative correlation intensity. Adapted image from [59]

present cross-peaks in synchronous spectra showing their relationship. Also, it permits to identify unknown bands from a pure molecule because all bands will be related synchronously if the concentration for that molecule is changing in the sample. It is useful to identify the *correlation squares* joining the pair of cross peaks located at opposite sides of the diagonal line to extract the bands that are synchronously correlated. In the example A and C from one side and B and D from the other are synchronously correlated.

In the case of cross peaks they can be positive or negative. The sign becomes positive if the spectral intensities at the two spectral variables in the coordinates of the cross peak are either increasing or decreasing together as functions of the external variable p along the observation interval. On the contrary, a negative sign of the cross peak intensity indicates that one of the spectral intensities is increasing while the other is decreasing. In the example from fig. 2.7 cross peak at A and C is negative meaning that one band is increasing while the other is decreasing. Cross peak at coordinates B and D is positive indicating that both bands are increasing or decreasing together.

Asynchronous 2D correlation spectrum

The intensity of an asynchronous spectrum represents sequential or successive, but not coincidental, changes of spectral intensities occurring at two different spectral variables ν_1 and ν_2. Fig. 2.8 shows an example of an asynchronous 2D correlation spectrum. In a descriptive manner, asynchronous maps are anti-symmetric and do not present autopeaks. They consist basically in cross-peaks located at off-diagonal positions. Similarly to synchronous maps, it is possible to construct correlation squares that helps to the interpretation of asynchronous features.

From the theory, having an asynchronous cross peaks signifies that the two spectral features change out of phase (delayed or with a higher speed if we consider time as external variable). This fact is particularly useful to differentiate highly overlapped bands arising from different molecules changing in a different way in our dataset. Asynchronous 2D correlation map will resolve clearly the two bands with a different sign in the cross peaks.

Other very interesting property of asynchronous spectra is that it gives information about the sequential order of the changes happening in the spectral data set. An asynchronous peak will be positive if the intensity change at ν_1 occurs predominantly before that at ν_2 in the sequential order of p . On the contrary, the peak sign becomes negative if the change at ν_1 occurs predominantly after ν_2. However, this sign rule is reversed if the synchronous correlation intensity at the same coordinate becomes negative ($\Phi(\nu_1, \nu_2) < 0$). From the last rules, in the example of fig. 2.8 intensity changes at A and C occur after changes in B and D.

The above rules also known as Noda's rule apply in the majority of cases having spectral intensity variations reasonably monotonic during the observation period.

Special cases and exceptions

However there are few exceptions for the interpretation of 2D correlation spectra from the rules explained before. Basically, they do not apply for some special

kinds of dynamic signals. First, if the intensity variations are very different in waveforms they can not be understood as a simple synchronous correlation based on coordinated increase or decrease of signals. An example of this case is a spectral intensity in which ν_1 increases gradually and the one at ν_1 changes in an erratic manner. Second, the rules can not apply with data sets involving many increasing and decreasing intensity changes with mismatched frequencies. As this kind of signals are clearly uncorrelated, it does not make sense to apply 2D correlation analysis for these cases.

The basic idea behind 2D correlation analysis is that the perturbation exerted to the system by the external variable p is affecting the spectral intensity in a similar waveform for all features in the spectra and not having very different frequencies.

It is very important to point out that other changes in the spectra different than simple intensity variations, like band position shifts or line shape intensities are necessary to be treated separately and the rules before do not apply. In these cases, characteristic features in 2D correlation maps needs to be studied and will be explained below by means of mathematical simulations.

Mathematical simulation of band effects

In Raman spectroscopy band's shape contains important information. From quantum theory, molecules possess well defined energy levels. Therefore, transitions between these levels, caused when molecules absorb or emit energy, occurs in well defined intervals (quanta), which gives rise to the vibrational spectrum. Thus, the absorption of energy by an isolated molecule going from a ground state to a first excited state occurs at a single, well defined, frequency [8]. However, most vibrating molecules exist in a bath of surrounding molecules (the environment), with which they interact. Each molecule interacts with its environment in a slightly different (and dynamic) way, and thus vibrates at a slightly different frequency [9]. The observed line shape is the sum of these individual molecules absorbing or scattering. At equilibrium, the population of vibrational states is controlled by the Boltzmann distribution. The majority of molecules in a normal IR or Raman experiment initially are in the ground state. Some of these molecules are transfered to the excited state when the IR radiation or Raman laser interacts with the sample. The resulting absorption (IR) or change in scattering (Raman) is captured by the spectrometer. Excited state molecules rapidly return to the ground state. For vibrations, this occurs after a few picoseconds (10^{-12} sec). This relaxation is called the lifetime (or amplitude correlation time). However, initially, all of the excited molecules are vibrating together (coherently), but motion and slight differences in vibrational frequencies randomizes this over time. Thus, the effective lifetime is a combination of all this effects. The overall line shape originates from the sum of all the individual vibrations, and the exact vibrational frequency of a particular molecule is controlled by its environment [61]. In the case of solids, the various molecules of the solid experience a statistical distribution of environments, and the line shape takes on the bell curve or Gaussian profile. This profile has the well-known shape from statistics, with a curving (not sharp) center and wings that fall away relatively quickly. On the contrary, in the case of gases where rotation and collisions happen quickly. The resulting line shape is Lorentzian (due to the exponential vibrational population relaxation), which is sharp in

the center, but has long wings. Liquids exist in between these limits, where interactions prevent extremely rapid motion, but the molecules are not locked in place. As a result, the two lifetimes can be close, and the line shape has features of both Gaussian and Lorentzian character. The simplest model for this involves the combination Gaussian-Lorentzian (G-L) profile. In conclusion band width in vibrational spectra can change due to interactions of molecules with its environment and neighbor molecules.

The position of a peak is controlled first by the natural vibrational frequency of the isolated molecule. However, the actual peak location also depends upon interactions with the environment. Depending on the strength of the bonding with the neighbors, the peak can shift to lower or higher energies. For instance, if the molecule is hydrogen bound to neighbors, this effectively lowers the bond energy (spreads the bonding energy over more space), and the peak will red-shift (to lower energy). If the molecule is experiencing repulsions, the peak blue shifts (to higher energy). In proteins, this phenomenon leads to the assignment of slightly variant frequencies of the amide I band to different secondary and tertiary structural elements (helices and sheets).

The height of a peak depends upon the number of molecules present (concentration) and the strength of the absorption. The Beer-Lambert Law uses this for concentration determination. As dilution can shift the peak, or there may be peak broadening, the area of the peak is a better indicator of concentration. In some cases, the peak height can be changed by a broadening mechanism, but the area will remain unchanged as the total number of molecules is constant.

Due to the valuable information that can be extracted from the shape and the behaviour of a Raman band, we used 2D correlation analysis to study different cases of band effects. By means of mathematical simulations, we intend to distinguish from 2D correlation maps: band broadening, shifts or band overlapping. By only observing the spectra is very difficult to differentiate band effects that seems similar at the spectra level. However, characteristic features in 2D correlation maps can discriminate objectively these different cases.

First, the case of a band shift is presented split in two cases: when the shift is bigger than the width of the band (fig.2.10) and when the shift is smaller than the width (fig. 2.9). When the shift is smaller than the band width, the pattern observed in the synchronous map is formed by four lobes and the asynchronous map has the shape of a "sandwich" oriented with respect to the diagonal alternating negative (blue) with positive (red) areas. When the shift is bigger than the band width, these patters are similar but distinct spots of each band can be observed in synchronous and asynchronous map. The simulation of a band shifting in steps bigger than the band width permits to distinguish it from the case in which of different neighbor bands change in intensity in a discorded way (simulation not shown). The characteristic 2D correlation maps on fig. 2.10 allow to detect this pattern and relate it to a band shifting.

Sometimes it is difficult to differentiate a shift from two overlapped bands changing in opposite directions. The doubts can be vanished by comparing the 2D correlation maps. Figure 2.11 shows the 2D correlation maps and spectra of two bands changing in intensity in opposite directions. The synchronous map is very similar to the one obtained with a small shift. However, asynchronous map is very different, being almost zero. Thus, 2D correlation is a powerful tool to differentiate difficult band behavior when analyzing series of spectra.

On the other hand 2D correlation can be used to detect band broadening or

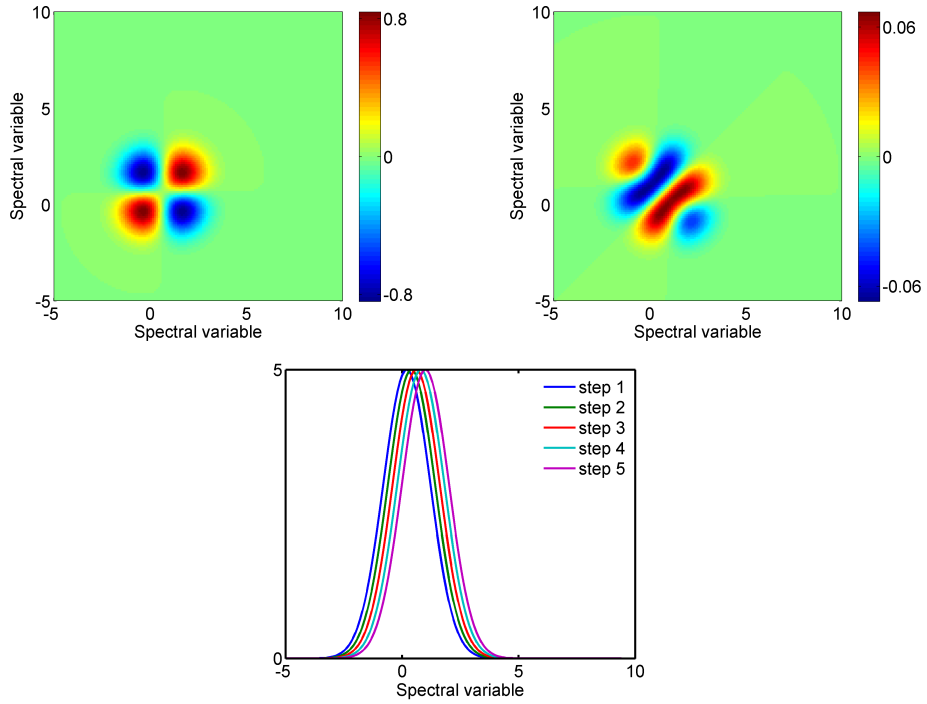


Figure 2.9: **Above:** Synchronous (**left**) and asynchronous (**right**) 2D correlation maps for a band shifting in steps smaller than its width. **Below:** A plot on the simulated band shifting.

narrowing combined with shifts or other effects. First, we present a simulation of a band broadening (fig. 2.12). The synchronous map shows a intense spot in the diagonal at the position of the band. The asynchronous map present a characteristic pattern with four lobes in a cross-like shape. This pattern can be altered and become asymmetric due to a small shift coupled with the broadening. Figures 2.13 and 2.14 presents the simulation of the combination of a band broadening with shifts to higher and lower wavenumber respectively. Furthermore, band can narrow. In figure 2.15, a band narrowing is shown. 2D correlation maps present similar features than broadening maps but with opposite colors (signs). When narrowing is coupled with a shift, patterns become asymmetric (fig. 2.16).

2.5 Comparison of PCA, MCR and 2D correlation for the analysis of band shifts and band overlapping

In the last section we analyzed with 2D correlation the case of a band shift and two overlapped bands changing intensity in opposite directions. We observed that 2D correlation was able to distinguish these situations, difficult to detect

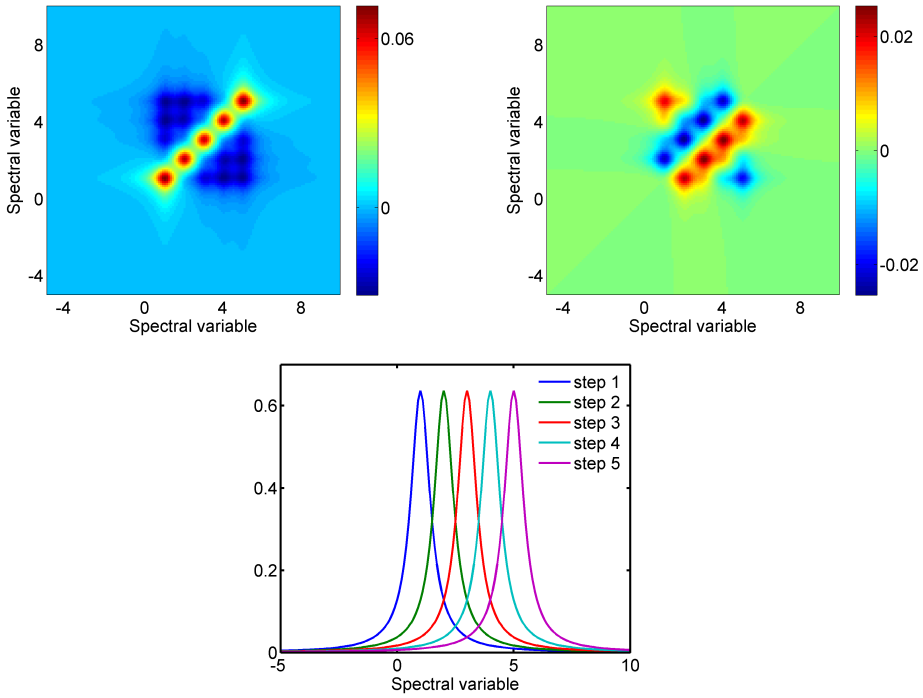


Figure 2.10: **Above:** Synchronous (**left**) and asynchronous (**right**) 2D correlation maps for a band shifting in steps bigger than its width. **Below:** A plot on the simulated band shifting.

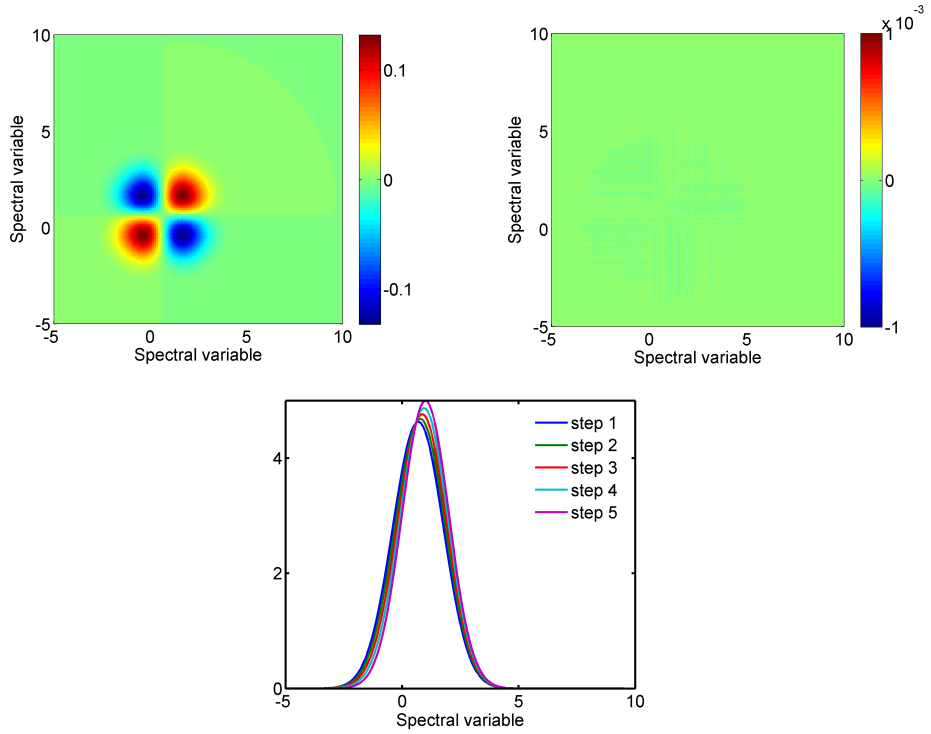


Figure 2.11: **Above:** Synchronous (**left**) and asynchronous (**right**) 2D correlation maps for two highly overlapped bands changing in intensity in opposite directions. **Below:** A plot on the simulated overlapped bands.

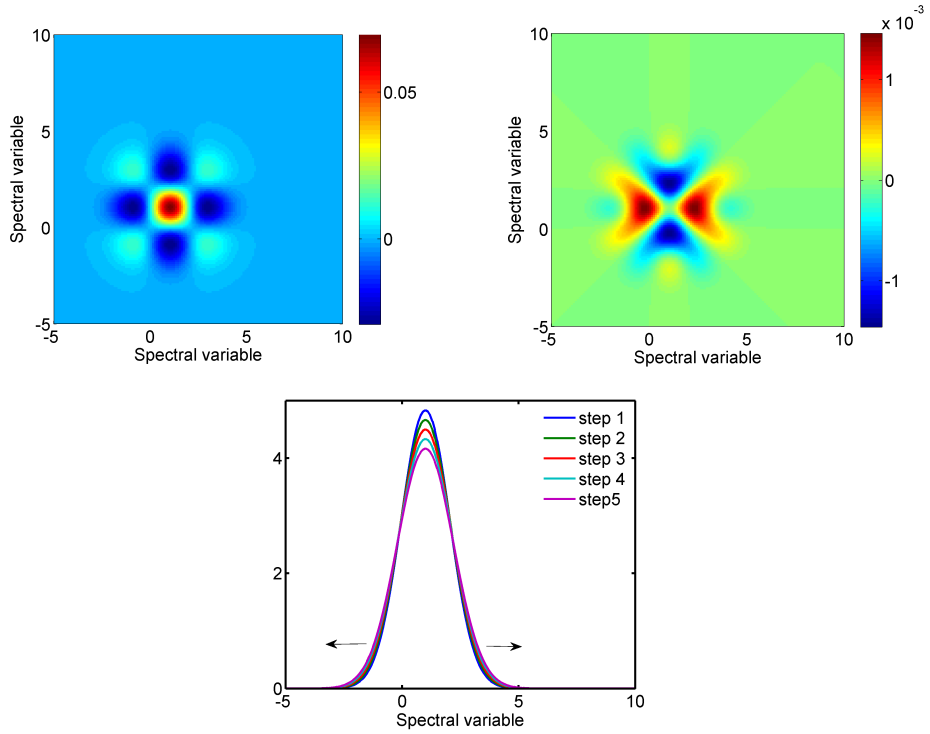


Figure 2.12: **Above:** Synchronous (**left**) and asynchronous (**right**) 2D correlation maps for a band broadening. **Below:** A plot on the simulated band broadening.

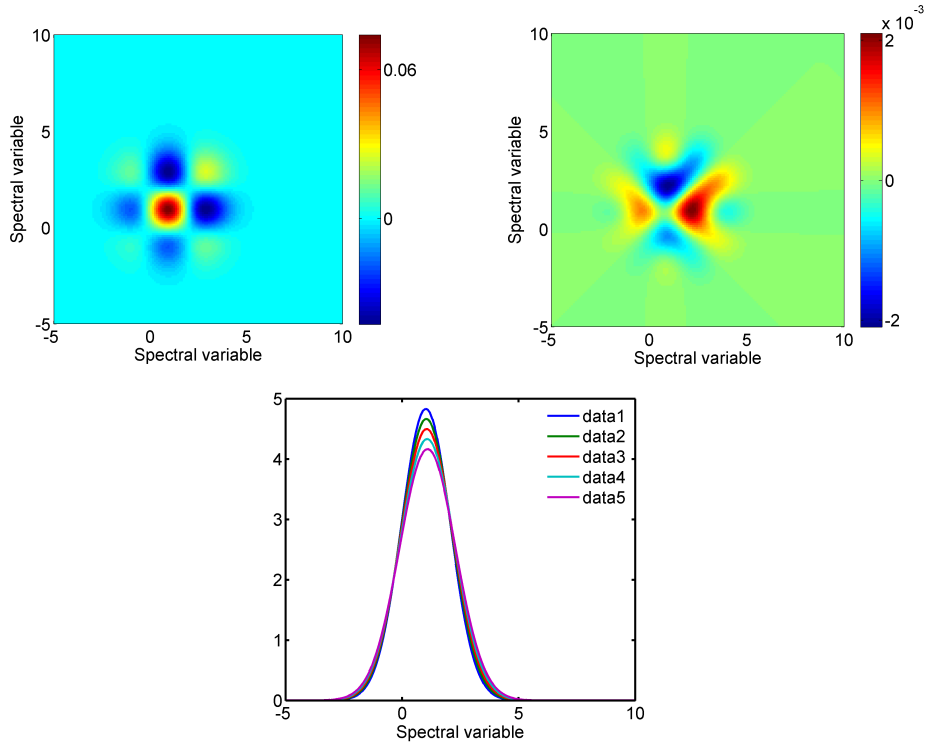


Figure 2.13: **Above:** Synchronous (**left**) and asynchronous (**right**) 2D correlation maps for a band broadening and shifting towards higher wavenumbers. **Below:** A plot on the simulated band broadening and shifting.

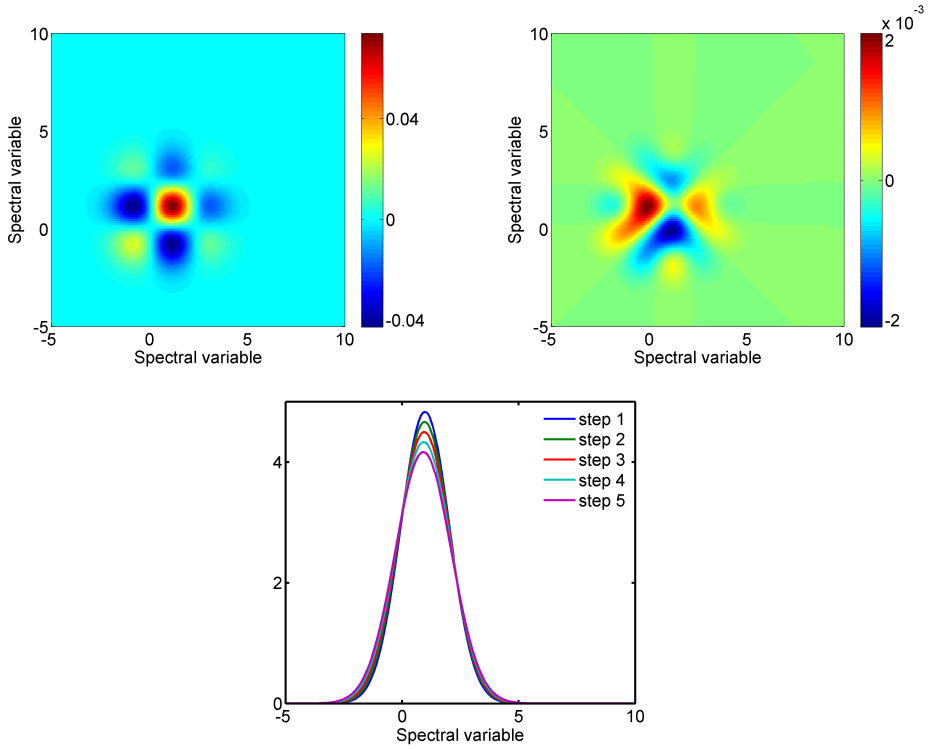


Figure 2.14: **Above:** Synchronous (**left**) and asynchronous (**right**) 2D correlation maps for a band broadening and shifting towards lower wavenumbers. **Below:** A plot on the simulated band broadening and shifting.

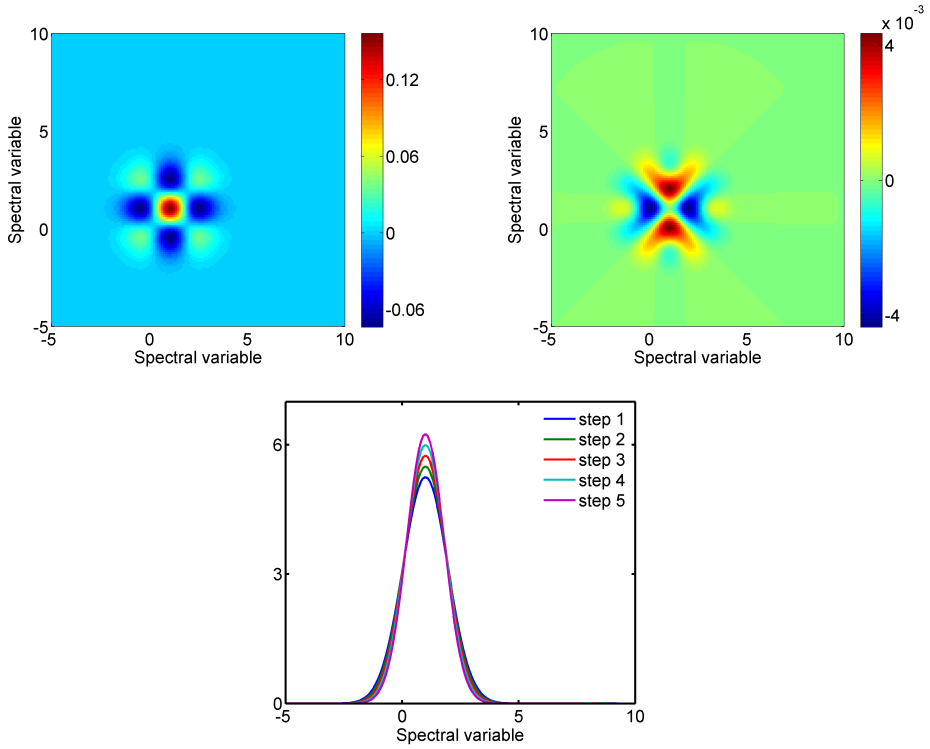


Figure 2.15: **Above:** Synchronous (**left**) and asynchronous (**right**) 2D correlation maps for a band narrowing. **Below:** A plot on the simulated band narrowing.

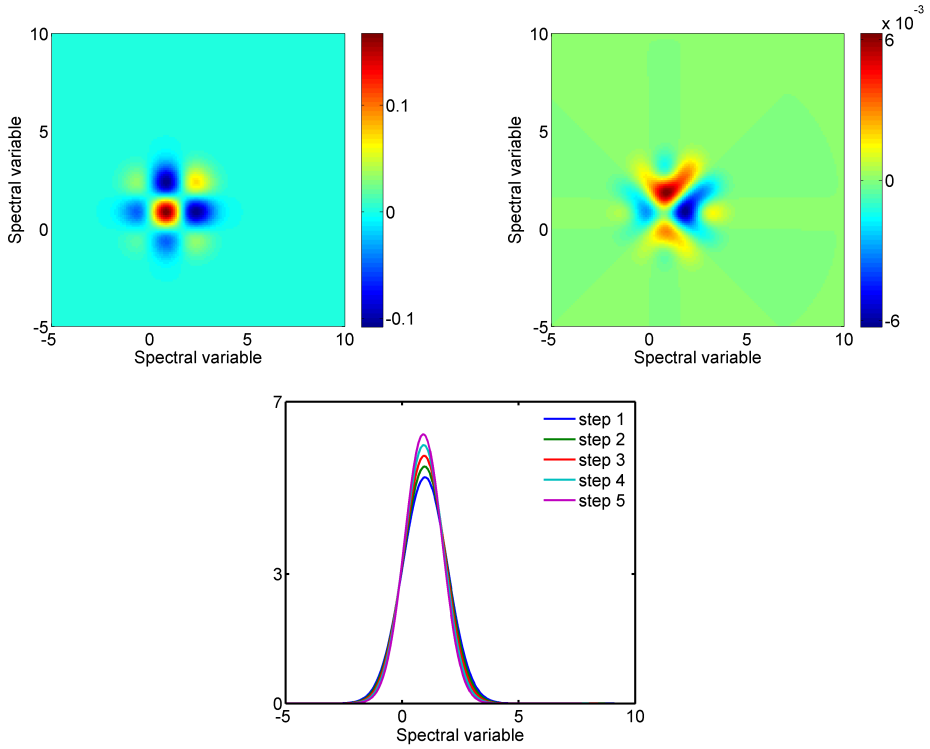


Figure 2.16: **Above:** Synchronous (**left**) and asynchronous (**right**) 2D correlation maps for a band narrowing and shifting towards higher wavenumbers. **Below:** A plot on the simulated band narrowing and shifting.

otherwise. Now, we show the analysis of the same data with PCA analysis and MCR analysis. Figure 2.17 contains the multivariate analysis for the case shown at fig. 2.9. PCA analysis presents a band centered in the mean position of the shift (first loading) and the second and third loading have symmetric and negative portions. On the contrary, MCR analysis give clear and physically understandable results. Five components are necessary to explain the five steps of the shift and loadings are composed by bands at each step position. Figure 2.18 presents the multivariate analysis for two overlapped bands changing in opposite directions (the same case as fig. 2.11). PCA loading give similar results as the case of the shift. However, MCR reproduce what is happening physically in the spectra. The two overlapped bands are deconvolved and the intensity behavior is monitored by the MCR scores.

Comparing PCA and MCR results, we can conclude that PCA do not present significant differences in the analysis of both cases (shift and band overlapping). On the contrary, MCR is able to give a physically meaningful representation of the band behavior, being able to discriminate the two different situations. In conclusion 2D correlation and MCR analysis are able to study band dynamics as shifts and overlapped bands. However, for complicated situations and to have a complete picture, a combination of both methods is recommended. Among all three methods MCR is the best in deconvolving meaningful components from the spectra. Therefore, it can be used to monitor the molecular content in a sample. 2D correlation, has the ability to resume in two maps the dynamic evolution of correlated bands and its shape dynamics. Finally, PCA is an exploratory technique that can be use to have an initial idea of different groups in the sample set or detect outliers. However, to understand chemically and physically which molecular components are evolving in the spectra acquired, MCR is more powerful.

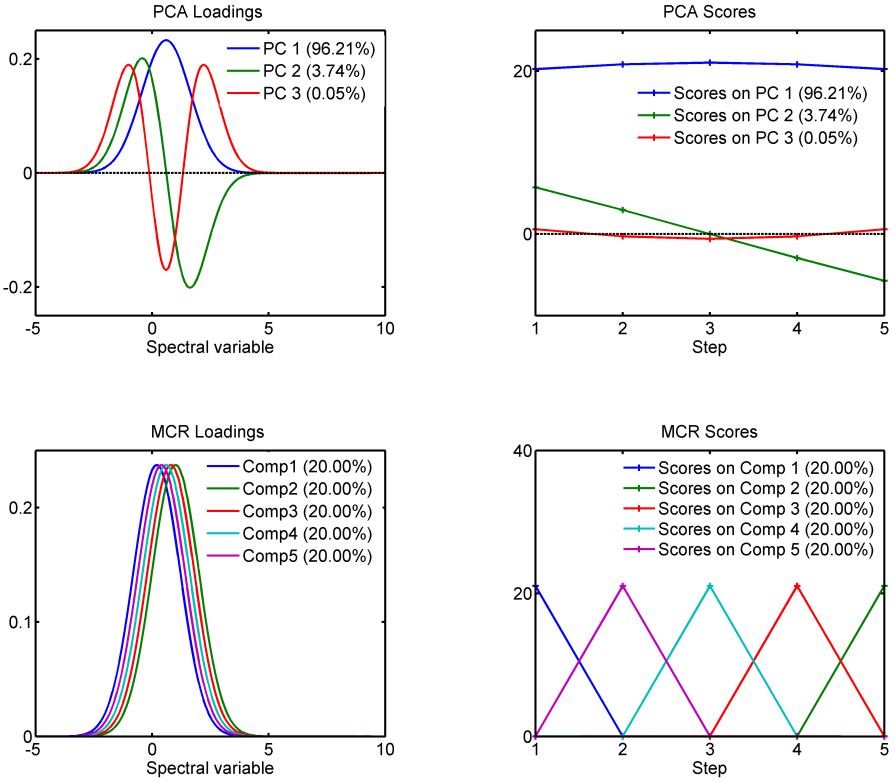


Figure 2.17: Multivariate analysis of a band shift. **Above:** PCA analysis: Loadings (**left**) and scores (**right**). **Below:** MCR analysis: Loadings (**left**) and scores (**right**)

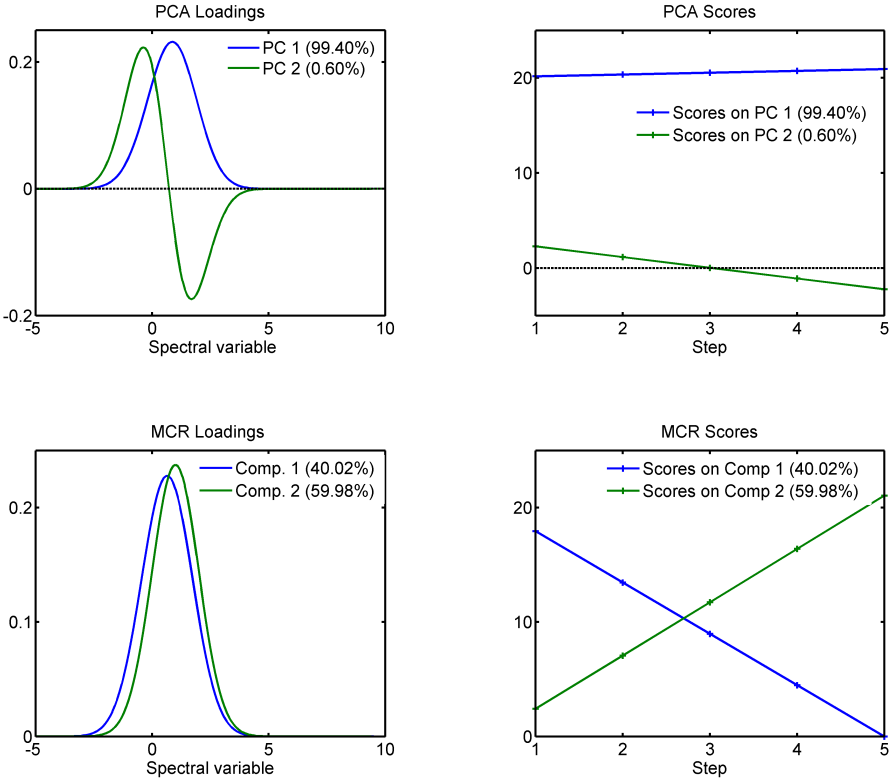


Figure 2.18: Multivariate analysis of two bands shifting in opposite directions. **Above:** PCA analysis: Loadings (**left**) and scores (**right**). **Below:** MCR analysis: Loadings (**left**) and scores (**right**)

3
Materials and Methods

For each experiment, different set-ups and software were utilized depending on the specific requirements. This chapter contains a brief description of the systems and the components used. First, the main features of a system combining optical tweezers with Raman spectroscopy are summarized for experiments using force loads to study single molecules (DNA) or RBC. Second, the Renishaw inVia microscope used for Raman experiments without the need of optical tweezers is described. In a different section a brief overview of the software used or developed to process Raman spectra with mathematical methods is explained.

3.1 Experimental set-up

3.1.1 Raman tweezers setup

A scheme of Raman system in combination with optical tweezers used for experiments is shown in fig. 3.1. The system is composed by two parts: the optical tweezers and the Raman setup. The tweezers part is made using a $1064nm$ diode laser beam (Laser Quantum Limited, Manchester, England) coaxially aligned with a $633nm$ detection beam (He-Ne laser, Research Electro-optics Inc.; Model no. LHRP-0501). With L_1 and L_2 (with focal length of 5 and 25cm respectively) the $1064nm$ laser beam is expanded and after it is divided with the help of a polarizing beam spliter BS_1. Both beams are again combined using other beam splitter BS_2 once they get reflected from mirrors M_1 and M_2. The expansion of the beam is done to overfill the objective (Nikon, oil immersion, 100X, 1.3NA). These mirrors are conjugated with the back focal plane of the objective. This conjugation is achieved by means of lenses L_3 and L_4 (with focal length 20 and 40 cm respectively) which form a collimator [62]. The movement of these lenses control the trap position in the direction of beam propagation. The control on the x-y movement of the optical trap is achieved with the movement in either mirrors M_1 or M_2 without any change in its intensity and shape, keeping the

trapping potential of the traps the same. One of the mirrors is connected to a motorized optical mount (Picomotor, New Focus, 8807).

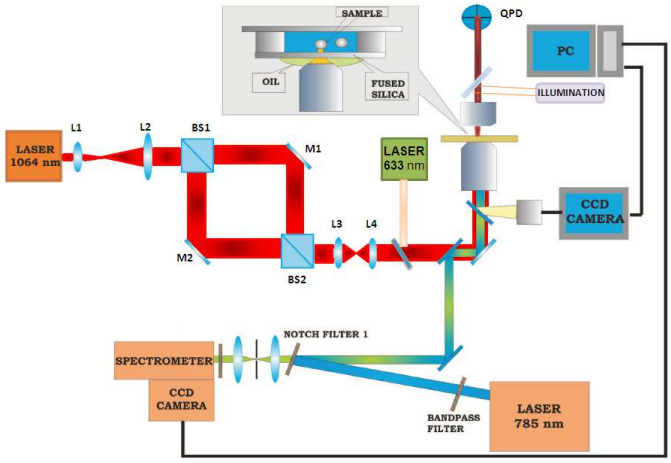


Figure 3.1: Scheme of the Raman tweezers set-up used with 785nm excitation wavelength.

BS_1 and BS_2 are also polarizing beam splitters, thus, after passing BS_2, the beams are orthogonally polarized. Consequently, the two beams do not have any crosstalk at the focus of the microscope objective, giving a stable dual trap. The 633nm beam is coaxially displaced to one of the trapping beams and is used for position detection and force measurement. The forward scattered 633nm beam is incident on a Quadrant Photodetector (QPD) (New Focus, model 2991) and gives a high precision information regarding the displacement of the trapped object from the focus.

In the same system it is necessary to implement a confocal Raman microscope coupled with the optical tweezers. A scheme of this part of the system is shown in fig. 3.2. For excitation of Raman scattering a 785 nm diode laser (Micro Laser systems Inc.) is used. The power is adjusted depending on the experiment in order to prevent a sample damage. The beam is filtered with a 785 nm bandpass filter (BF) to exclude the broad fluorescence from the diode laser and narrow down its bandwidth. After BF the beam is collimated and falls on a holographic notch filter NF (Kaiser Optical Systems) at a small incident angle. The reflected light from the NF goes inside an Olympus-IX51 inverse microscope where the object under study is trapped. The objective also collects the back scattered light. This light contains Rayleigh and Raman photons and is passed through NF and a confocal pinhole of $100\mu m$ dimension. The confocal system discards the most part of the out-of-focus light. Finally, the beam is focused on to the slit of the spectrometer by using a lens L. The spectrometer is a Spectra Pro 2500i, 500mm focal length monochromator, with a resolution of 0.05nm at 435.8nm, containing a 600lines/mm grating (blazed to 750nm) and fitted with a Spec-10:100B back-illuminated CCD, cooled down to $-80\,^{\circ}C$. The data acquired from the CCD is processed with the WinSpec software (Princeton Instruments). The objective used is an oil-immersion objective and we used the oil, that does not give fluorescence. For that reason cargille immersion oil (type 37) is used. In order to visualize and take images of the samples an additional

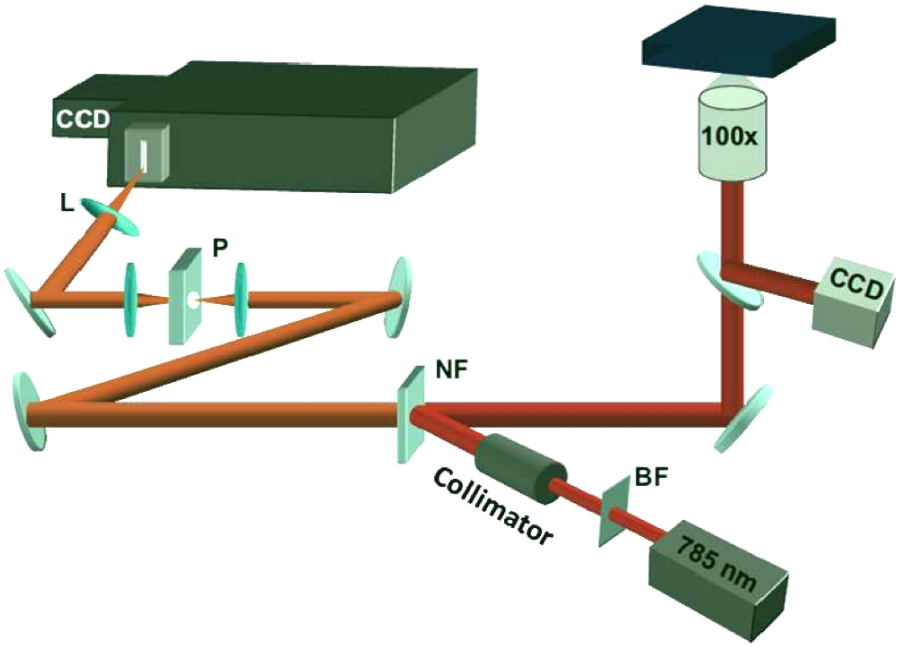


Figure 3.2: Scheme of the confocal Raman microspectroscopy part of the setup.

CCD camara (JAI) is attached to the microscope.

3.1.2 Renishaw Raman microscope

For measurements on cells or tissue without the need of optical tweezers, an inVia Renishaw Raman microscope was used. This system supports multiple excitation wavelength lasers, with an automatic software switching the excitation wavelength. In our case we used two lasers: a 785nm HPNIR785 Renishaw diode laser with a maximum output power 300mW and a 514nm Argon Ion laser of 25mW output power. A scheme on the system is shown in fig. 3.3. The laser light is injected through a hole onto the laser turning mirror. Then, it is expanded while passing thought the beam expander and light is directed to the holografic notch filter where the light is reflected and guided to the objective. A sample is illuminated from the top, what is called an upright microscope. Light scattered from the sample is collected again by the objective and filtered with the notch filter which removes the Rayleigh scattered light. Finally, the Raman emission passes through the entrance slit of the spectrometer were the spectrum is acquired by means of a CCD detector. An additional video camera permits to visualize the sample and take snapshots or videos. The software controlling the system is called Wire and a view on the interface is shown in fig. 3.4. This interface permits the automatic change of the excitation wavelength, grating, laser power and other acquisition parameters like exposure time or wavenumber region. In the left-hand part of the screen is also possible to visualize the sample focal plane in real time. On the right hand, different spectral acquisitions can

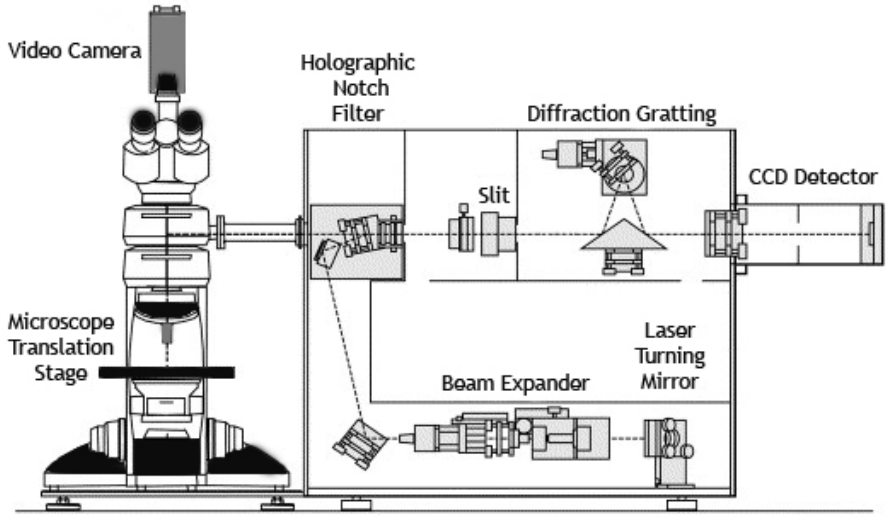


Figure 3.3: Schematic view of the Renishaw Raman microscope system

be obtained.

3.2 SERS probes

3.2.1 Silver colloids

In experiments, we used SERS probes based on silver colloids [63]. First, silver colloids need to be prepared. The method used was the Lee and Meisel method also known as citrate reduction method. Silver colloids are prepared by reducing silver nitrate ($AgNO_3$, Sigma-Aldrich, S8157) with the help of trisodium citrate ($C_6H_5O_7Na_3$, Sigma Aldrich, C3434), according to protocol described by Lee and Meisel [64], with some modifications. 100 ml of 10^{-3}M silver nitrate aqueous solution is heated to boiling ($\approx 90\,^{\circ}$C) and then 2ml of a 1% (w/v) trisodium citrate solution is added slowly and steadily, while continuously stirring the mixture. The mixture is kept boiling for 1 hour (with stirring on) and then is allowed to cool down. The resultant mixture is of dark grey color.

3.2.2 Silver coated silica bead

The SERS probes were obtained by chemically covering a 5 or $2\mu m$ silica bead (Micro Particles, GmbH) with silver colloids obtained from citrate reduction method. To start with, $1000\mu l$ of silica beads is taken from stock and all the liquid in which it is suspended is removed by centrifugation at 3500 rpm for 4 minutes. The beads are then dried overnight at $60\,^{\circ}$C. After, a solution of 1% APTMS (3-aminopropyl trimethoxysilane) is made in pure ethanol. Then $100\mu l$ of APTMS is added to the dried beads in $900\mu l$ of absolute ethanol and allowed to react for 24 h at room temperature with continuous movement. At this stage, the solution was purified by centrifuging at 3500rpm for 4 minutes.

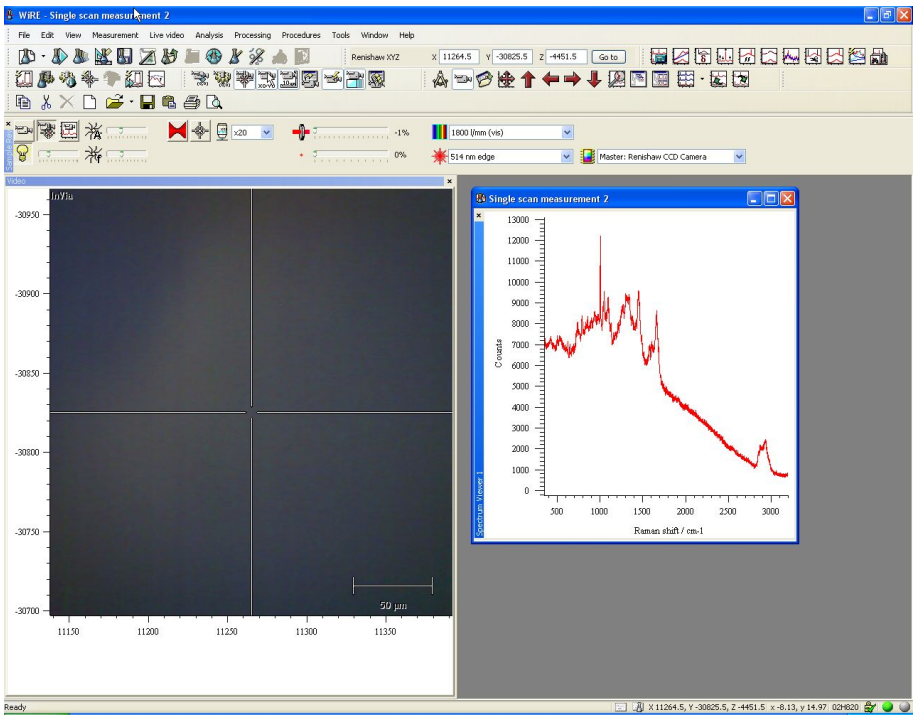


Figure 3.4: The screenshot of the interface "Wire" used to control the Renishaw microscope.

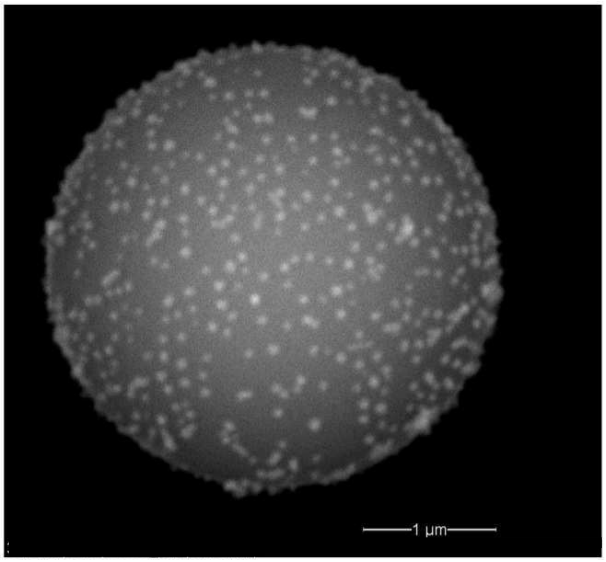


Figure 3.5: SERS images showing the resulting packing density and homogeneity of silver colloid attached to silane monolayer (APTMS) coated silica bead

The supernant from all three washes is discarded and the remaining beads are dried again at 60 °C.

One-tenth of the beads from the dried sample obtained above is redispersed in $1000\mu l$ double-distiled water. Rest of the dried beads can be kept at a cool and dry place for future use. From the redispersed sample, $10\mu l$ beads are taken and are mixed in $940\mu l$ of double-distilled water along with $50\mu l$ of silver colloids obtained from above methodologies. The whole mixture is then incubated at a speed of 300 rpm for 30 minutes. What we obtain as final product is silica beads with a layer of silver nanoparticles over it. A view of the silver coated bead is shown in fig. 3.5. Amount of beads and amount of silver coating can be varied according to the need of experiment. But it should be remembered not to put too much beads at once for the silver coating, as it will hinder the uniform metal coating due to high bead density. To get rid of free colloids in suspension or to change the suspension from DI water to the buffer in which experiment have to be carried out, we can centrifuge the final sample at 3500 rpm for 4 minutes and discard the supernatant. This will remove most of the freely suspended silver colloids. Finally, a solution in which we want the metal coated beads to be suspended is added.

3.3 Software analysis

3.3.1 PLS toolbox

The multivariate analysis (PCA, PLS-DA, PLS regression, MCR) performed during the thesis was performed in most of the cases using the PLS toolbox (Eigenvector Research). PLS-Toolbox software contains an extensive suite of

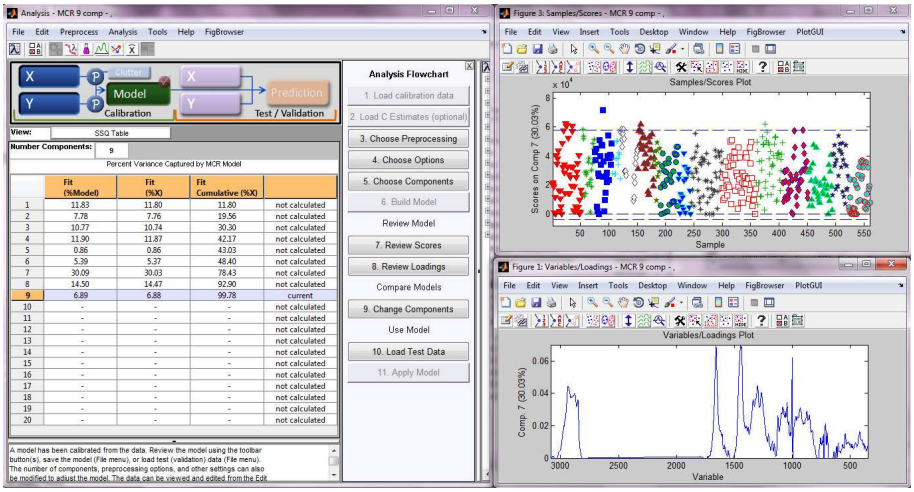


Figure 3.6: PLS-Toolbox interface

essential and advanced chemometric multivariate analysis tools for use within the MATLAB computational environment. Some tools are available also in a stand-alone version called Solo. But for our applications we used the programm in the MATLAB environment due to the flexibility and broader possibilities that it offers. PLS-Toolbox provides a unified graphical interface and for use in a wide variety of technical areas. The tools that can be used for chemical engineers, analytical chemists and other analysis-driven scientists to analyze their data and build predictive models. A view of the interface is shown in 3.6.

3.3.2 MCR-ALS

The application of Multivariate curve resolution (MCR) to Raman spectroscopic data has not been broadly investigated yet. In this thesis we used MCR-ALS algorithm for its flexibility and advantages. MCR-ALS is an algorithm that solves the MCR basic bilinear model using a constrained Alternating Least Squares algorithm. The constraints used to improve the interpretation of the profiles of the concentration C and S^T (Pure component spectra) may respond to chemical properties of these profiles (e.g., non-negativity, unimodality, closure, ...) or have a mathematical origin. The flexibility in applying the constraints and the capability to treat the most diverse multiset structures are the main advantages of this algorithm. The benefits of using MCR-ALS stems from the proper selection and application of the constraints that are really fulfilled by the data set and from the ability to envision how to design and to deal with the most informative multiset structures.

In the last chapters of the thesis we used MCR constrains to deconvolve meaningful component spectra with the minum *a-priori* information. For that, we used a software developed by Romà Tauler, Anna de Juan and Joaquim Jaumot from the department of Analytical Chemistry at the university of Barcelona. MCR-ALS is implemented under MATLAB environment. There are two avail-

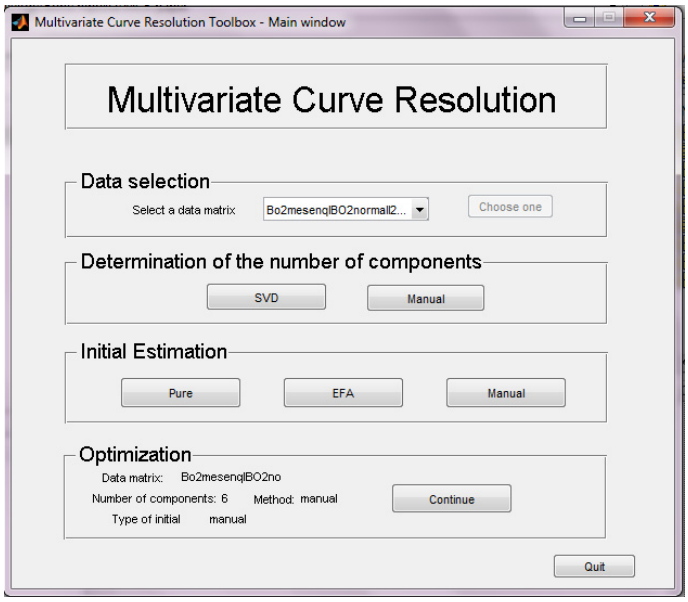


Figure 3.7: MCR-ALS interface of the software developed by the group of Analytical Chemistry at the University of Barcelona [36]. In this step, the initial estimates and the number of components to deconvolve are selected.

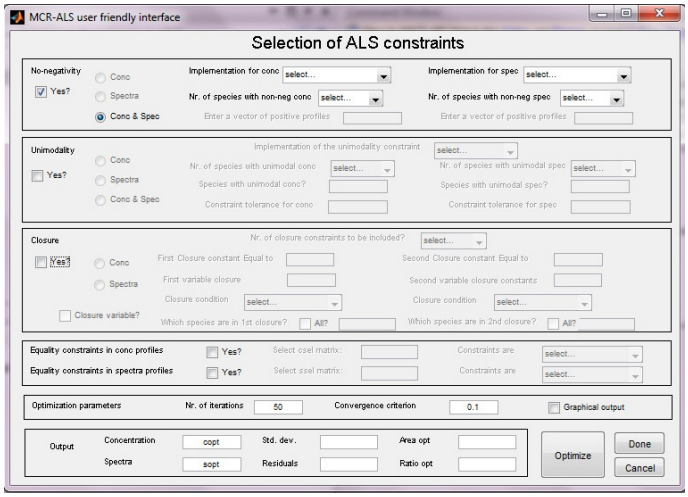


Figure 3.8: MCR-ALS interface of the software developed at the University of Barcelona [36]. In this step of the algorithm, the constraints are selected.

able versions of the algorithm, a command line program and the more popular graphical interface GUI MCR-ALS program, both downloadable in the web page: *http* : *//www.ub.edu/mcr/web_mcr*. Some views on the graphical interface of the software are shown in figures 3.7, 3.8, 3.9 where a user can apply the different constrains and initial estimates in a very flexible way.

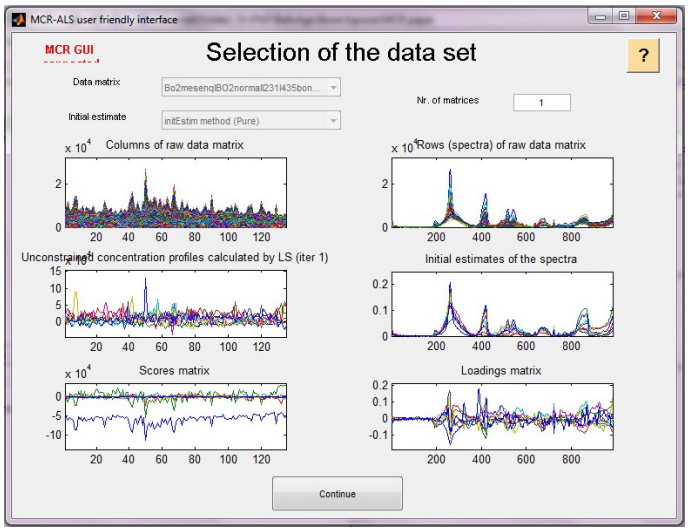


Figure 3.9: MCR-ALS interface of the software developed at the University of Barcelona [36]. Here, a first presentation of the data is shown.

3.3.3 2D correlation

In order to study the Raman band's behavior and the correlation between Raman bands when parameter in the system changes (ph, force, time...), 2D correlation analysis was used. I wrote a Matlab code for the calculation and visualization of 2D correlation maps. The code is shown in the appendix A

3.3.4 Labview background subtraction

Preprocessing of the large amount of spectra acquired after a standard experiment can be difficult and time consuming. For that reason, I buit a Labview interface with the possibility of smoothing, spike removal and background subtraction of the Raman spectra in an automatic way. The background subtraction was implemented with a code that I worte in Matlab and based in the method described at [65]. Once the user enters the specific parameters to preprocess the spectra, the smoothing, spike removal and background subtraction is automatically performed in a selected number of spectra. A view of the interface can be seen in fig. 3.10.

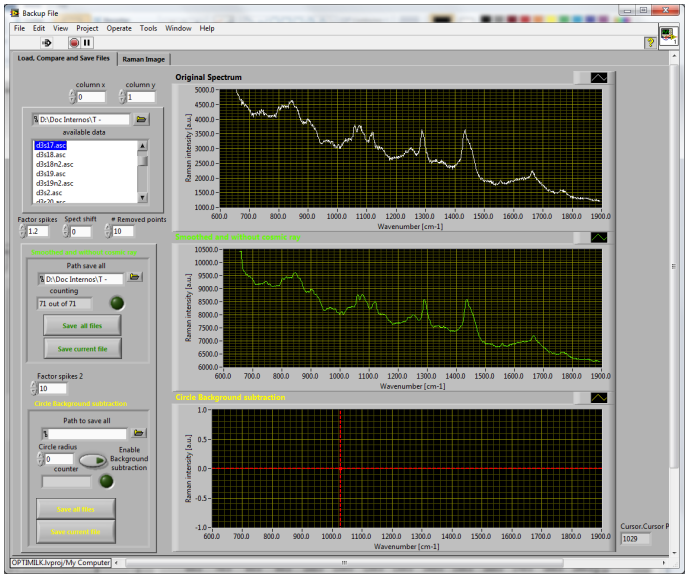


Figure 3.10: A view on the Labview interface developed to subtract the background and remove cosmic rays from the Raman spectra

4

Direct observation of single DNA structural alterations with surface-enhanced Raman scattering

DNA contains genetic information necessary for the functioning and development of all living organism. The DNA molecule acts as a platform for a host of critical functions such as transcription, replication, and other molecular motor driven processes, where the DNA strand undergoes numerous mechanical events that are primarily supported by the polymer-like phosphate backbone. To get insight into DNA mechanics, single molecule force spectroscopy techniques are necessary. Inherent Brownian motion limits current force spectroscopy methods from observing possible bond level structural changes. We combine optical trapping and surface-enhanced Raman scattering (SERS) to establish a direct relationship between DNA's extension and structure in the low force, entropic regime. A DNA molecule is trapped close to a SERS substrate to facilitate detectable Raman signal. Statistical analysis of Raman signal is performed with 2D correlation and some important features are revealed as a shift in DNA Raman modes in response to applied force, indicating phosphodiester mechanical alterations. Molecular dynamic simulations confirm the local structural alterations and the Raman sensitive band identified experimentally. The combined Raman and force spectroscopy technique is a novel methodology that can be generalized to all single molecule studies.

4.1 Introduction

As a passive substrate, DNA maintains mechanical compliance to allow interactions with proteins. Some of these interactions occur at low forces where the mechanical load on DNA is balanced by lowered entropy via unfolding. Thus, understanding DNA structural responses in this entropic regime is essential to

elucidating overall DNA function.

The advancement of single molecule force spectroscopy (SMFS) has connected physiological function to molecular level processes, for example, with DNA [66, 67]. A typical SMFS experiment provides force measurements which are related to models that idealize the molecule mechanically [68]. Advancements in instrumentation have pushed length scale resolution limits leading to dynamical studies of active processes [69]. However, these experiments are either limited to measuring a few averaged parameters or require advanced particle tracking with fluorescence probes [70]. The use of fluorescence imaging is generally applicable to strong topological changes [71] occurring in high force regimes. More importantly, light interacting directly with chemical bonds would provide a more ideal measure of the DNA structure.

Raman spectroscopy is well-suited to provide important insights to SMFS, since it outputs the highest level of chemical structure information with minimal external interference due to its fundamental scattering process at all optical wavelengths. Raman studies of DNA have been present for two decades and have produced a database of Raman peaks that characterize the various components of the DNA structure [72, 73]. Raman signal can be amplified by metal nanostructures through plasmonic effects [74] which makes possible single DNA molecule detection [75–77]. The drawback is that the DNA or its constituent samples are measured as an ensemble or anchored to hard surfaces; far from an ideal physiological state. In my group at ICFO, it have been previously demonstrated a methodology to overcome some of these issues by optically trapping single DNA molecules [34] with silver nanoparticles nonspecifically bound to the phosphate backbone, which made possible the Raman detection of single molecule DNA in its natural aqueous environment.

In this work, we demonstrate a novel combination of optical tweezers and SERS to study DNA structural responses from an applied load in the low force, entropic regime. A single DNA molecule is optically stretched close to an external SERS substrate while its Raman spectrum is simultaneously measured. We identify a correlation between the phosphate backbone structure and molecular extension. The results of this novel experimental technique are confirmed with state of the art theoretical modeling, which combines molecular dynamics (MD) with mixed quantum mechanics/molecular mechanics (QM/MM), to compute Raman modes for DNA structures modeled at different extensions. The modeling was performed thanks to the collaboration with the Supercomputing center in Universitat Politecnica de Catalunya. The overall result demonstrates a structural response in a regime where mechanical load is thought to only be countered by entropy changes. We also highlight an innovative methodology for directly observing single molecule chemical structure changes in response to controllable forces.

4.2 Methods

4.2.1 Dual optical trap and Raman spectroscopy

The experimental setup (Fig. 4.1) is a platform that combines optical trapping with Raman spectroscopy and is similar to the one described in section 3.1.1. Briefly, a specimen held by a dual trap is optically excited by a third

beam and its Raman spectrum is recorded in the backscattered direction. The dual optical trap was constructed with a 985 nm single-mode fiber laser beam (Avanex) passed through an interferometer (100 mW power for each beam) with one mirror motor driven (Newport motorized optical mount, 8807) and controlled by a computer. This mirror was optically conjugated to the back focal plane of the bottom objective (Nikon, CFI PL FL, 100×, NA 1.30). An additional 635 nm low-noise optical beam (0.5 mW) (Coherent, ultralow-noise diode laser LabLaser635) was passed to the sample, coaxial to the non-motor driven trapping beam for position detection of the trapped microsphere. The forward scattered light of the detection beam was collected by a top objective (Edmund Optics, 40×, NA 0.65) and analyzed by a quadrant photodiode (QPD) (Newport, model 2921). The Brownian motion of the bead was measured to calibrate the trap using established procedures [78] and to give the bead position for the force-extension curves. The resulting signals were then transferred through an analog-to-digital conversion card (National Instruments) and recorded with custom software in the Labview environment. A 532 nm beam was used for the Raman excitation (OZ Optics, OZ-2000, 20 mW) with a power density of 1.5×10^6 W/cm^2 at the focus. The dependence of Raman spectra on excitation polarization direction was not tested. The backscattered light was collected by the bottom objective and passed through a holographic notch filter (Semrock, 532 nm RazorEdge Dichroic laser-flat beamsplitter) before entering the confocal system with a 150 μm pinhole (Thorlabs, P150S). Raman spectra were recorded with a spectrometer (Andor, SR163, 1200 lines/mm) equipped with a charged-coupled device (CCD) camera (Andor, DV401A-BV) at a spectral resolution of 3 cm^{-1}. For the measurement, the DNA-bead construct was aligned such that the Raman excitation beam passed between the beads.

None of the auxiliary beams (i.e. the trapping and position detection beams) added background signals to the Raman spectra. These beams are shifted by half the molecule length (2 μm) relative to the Raman excitation focus (see Fig. 4.1) thus propagating outside of the confocal volume of the spectroscopic system. In addition, the power of the detection beam is significantly weaker than the Raman excitation while the trapping beam wavelength is outside the spectral window of the spectrometer that is tuned to wavelengths close to 532 nm. No Raman peaks were observed from two trapped beads without DNA attached between them.

4.2.2 DNA Raman spectrum versus extension measurement

The experiment (see Fig.4.1(top left inset)) consisted of DNA-bead constructs, double-stranded λ-phage 12 kbp DNA (4.25 μm contour length) molecules anchored to silica microspheres via established methods [79], manipulated with the dual optical trap in phosphate buffered saline (pH 7.4) within a custom-built fluid chamber. The Raman signal of the molecule was enhanced by bringing it close to a SERS substrate, 5 μm silica beads with silver nanoparticles (70 nm average diameter) attached to their surface [63], that was previously deposited on to the glass coverslip surface of the fluid chamber. Briefly, the probes were prepared by attaching citrate-reduced silver colloidal nanoparticles [64] to the surface of the silica beads. The particles were anchored to the surface via a self-assembled alkyloxy silane monolayer. Adjusting parameters, such as silane

concentration and the agitation speed of the reaction vial that contained the beads and chemicals, produced a stable silane monolayer with silver nanoparticles attached at less than 20 % packing density on the bead surface, confirmed by scanning electron microscopy (SEM) imaging (Fig. 3.5(bottom right inset)).

For the measurements, a full force versus extension curve was first measured to confirm the presence of a single DNA molecule by fitting to a worm-like chain (WLC) model and confirming the single molecule persistence length [34]. The molecule was then extended to approximately 85% of its contour length before approached in the z direction towards the upper surface of the SERS probe by lowering the bottom objective. This was the initial extension for the measurements and was essential in order to bring the molecule close to the SERS probe without having it contact the trapped beads. The Raman signal was continuously detected at 1 s acquisition times in "live" mode and the DNA z displacement was stopped upon appearance of Raman peaks and adjusted in the x-y directions by slightly shifting the movable trapped bead to maximize peak intensities. The DNA molecule was then extended through two additional 100 nm increments with a single Raman spectrum at 1 s acquisition time recorded at each step. The position of the bead in the fixed trap was continuously recorded to monitor the applied force. The distance between the molecule and SERS probe could not be directly measured or held perfectly constant due to the inherent Brownian motion of the molecule. However, this motion was minimized by the taut state of the molecule throughout the experiment.

4.2.3 Molecular dynamics and QM/MM modeling

Molecular dynamics (MD) modeling was performed with density functional theory (DFT) normal mode frequencies to confirm the experimentally observed Raman modes and their sensitivity to molecular extension. The modeled DNA 3D structures [80] were 30 base pairs long with only cytosine-guanine base pairs, which minimized the number of theoretical modes that are not measurable in the experiment. Four lengths were tested:97Å (unextended), 120 Å, 135 Å and 155 Å. The theoretical structure extensions were higher than in experiment due to the long computational time necessary to resolve smaller extensions (forces). For each modeled structure, normal mode frequencies were calculated for three adjacent segments on each chain in order to observe any variations along the strand. Each segment contained two base-pairs with hydrogen-caps inserted directly above and below surrounding sugars for a total of 134 atoms (Fig. 4.2).

For the initial structure, the Raman active modes in this list of normal mode frequencies were determined by computing Raman activities in a reduced structure containing only two bases with one joining phosphate. Obtaining Raman activities for the full modeled structures was too computationally intensive for this study. The result was a list of frequencies with corresponding Raman activity values. Raman active frequencies in the region of experimental interest were tracked in the full QM/MM systems to see which modes changed with DNA extension. All computed Raman frequencies were lower than in the experimental measurements. For example, the O-P-O stretch mode with the highest Raman activity was at 794 cm^{-1} whereas this mode always appeared above 800 cm^{-1} in the experiment.

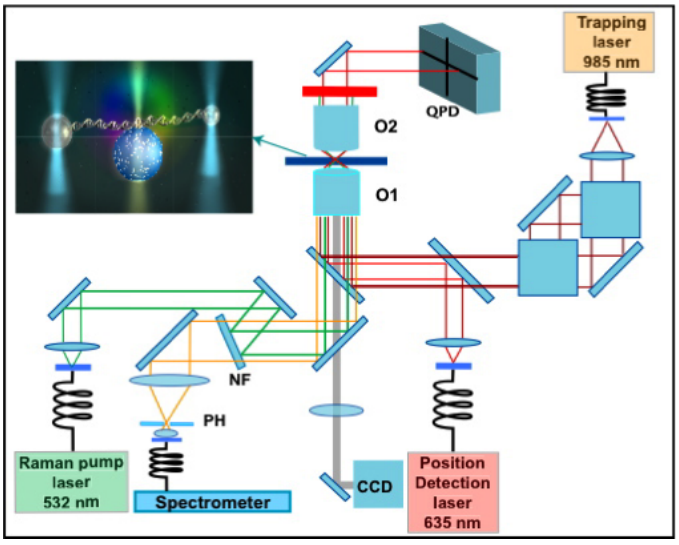


Figure 4.1: Schematic of the combined dual optical trap and Raman spectroscopy setup where a single DNA molecule is stretched by moving one of the attached beads with the optical trap while a laser excites the SERS substrate and DNA molecule (top left inset). All lasers used are coupled through single mode fibers before entering the system. The 985 nm trapping beams are split by an interferometer made with two prism beam splitters. Both beams are passed by a dichroic mirror to the bottom objective (O1). A 635 nm beam also passes through O1 and is used for position detection of the trapped microsphere. The forward scattered light of the 635 nm beam is collected by a top objective (O2) and analyzed by a quadrant photodiode (QPD) with a band-pass filter (Thorlabs, FL635-10) that blocks the trapping and Raman excitation beams. The 532 nm Raman excitation beam is reflected by a 45^o Notch filter (NF) and propagates colinear with the trapping and detection beams through O1. The backscattered light is collected by O1 and passed through NF to a confocal spectroscopy system that consists of two lenses and a pinhole (PH) in front of a spectrophotometer with a CCD camera.

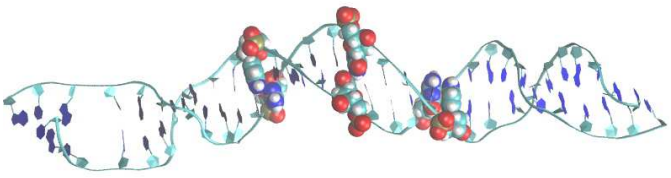


Figure 4.2: Image of DNA molecular model, highlighting the individual segments that were included in the MD normal mode frequency computation. Each segment contained two base pairs with an adjoining phosphodiester chain and three pairs of segments (6 total) between the two chains were calculated to observe any variations along the strand.

4.3 Results and Discussion

4.3.1 Raman signatures of a single DNA molecule

Figure 4.3(a) is a representative series of Raman spectra from a single DNA molecule at three different extensions where a difference in some peak positions between spectra is immediately clear. Each trace was background subtracted using multiplicative scatter correction (MSC) and smoothed with a spline technique [81]. The plotted region contains stretch vibrational modes of the phosphodiester network (C-O-P-O-C) and C-N bonds in the cytosine base [73], with bands at lower and higher frequencies containing bending and scissor modes, respectively, that are typically inconsistent with SERS. The peak positions measured here are similar to what was observed in the previous work of our group [34] where silver nanoparticles were attached to the DNA molecule, demonstrating consistency between two different silver nanoparticle-based SERS systems. The band assignments used here are based on literature that utilized bond symmetry arguments to determine normal mode frequency regions for the DNA components [73]; a direct derivation from the bond orientations of the molecule that is independent of any electromagnetic field enhancement. Moving the DNA molecule away from the SERS bead and Raman excitation focus led to disappearance of signal, thus eliminating band assignment to spurious peaks that could be due to the surface chemistry of the SERS bead. As seen in Fig.4.3(a), the Raman peak position shifts only occur for the phosphodiester stretch mode, and only in an increasing direction with extension, while all other peaks stay in the same position or experience slight random movement. Plotting the Raman peak position versus extension for the phosphodiester and cytosine base mode for this set of traces further exemplifies the difference in sensitivity to molecular stretching between the two bands (Fig. 4.3b inset). The persistence length (53 nm) from the WLC fit to the force extension curve in Fig. 4.3b confirms the presence of a single DNA molecule and indicates the applied force and extension range where this Raman band sensitivity is observed.

Raman spectra from measurements of five DNA molecules at all extensions are grouped together and cross-correlated via a 2D correlation analysis [59] to identify peak movement in the sample set (Fig. 4.4). 2D correlation analysis is an objective method for identifying trends in a set of spectra that respond to a single effect, in this case the extension of the molecule. The outputs are in-phase (synchronous) and out-of-phase (asynchronous) maps where the former represents Raman peak position and intensity variations and the latter the presence of a sequential trend of these variations relative to the molecule extension. Peak position trends with molecule extension will emerge as hotspots in both maps. Features between 800 - 900 cm^{-1} band, where stretch modes of the O-P-O unit lie, are present in both maps, signifying a displacement of this band with extension. The doublet spot is due to a shift greater than the width of the peaks (see simulation on fig. 2.10). The C-N stretch mode of the cytosine ring, 1280 cm^{-1}, only has an in-phase feature due to an intensity change with no position shift in the peak (fig. 4.4). Generally, intensity changes of SERS spectra are difficult to interpret because the enhancement level can vary due to differences in the relative positions of the SERS substrate, molecule, and incoming excitation light [82]. Debris in the solution that could come from the SERS

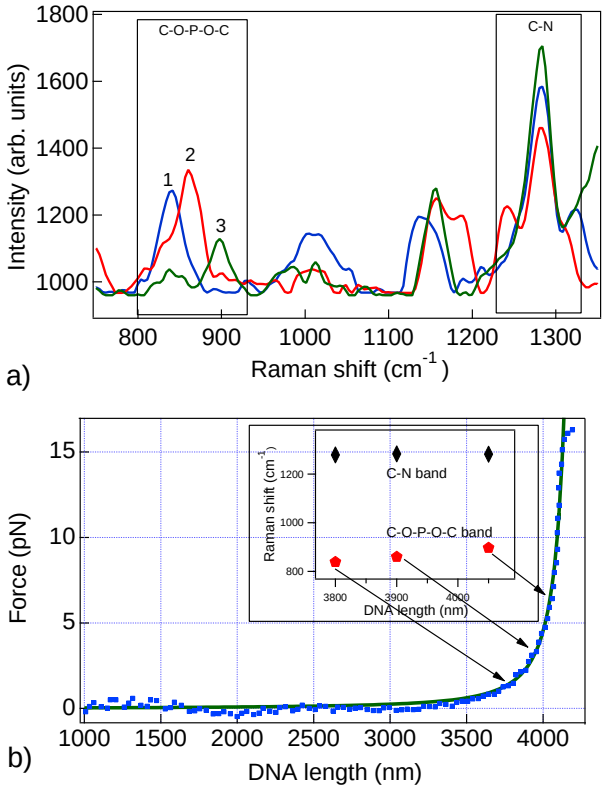


Figure 4.3: A representative set of Raman spectra **(a)** from a single DNA molecule at three different extensions: 3700 nm (1), 3800 nm (2), and 3900 nm (3). The O-P-O Raman band in the 800 - 900 cm^{-1} range undergoes an upward shift in position while the C-N vibration (1280 cm^{-1}) remains constant as the molecule is extended. A single molecule between the trapped beads is confirmed by measuring a force vs. extension curve **(b)** and fitting experimental data (solid squares) to a WLC model [68] (solid line) to validate the single DNA persistence length (53 nm). The peak position versus extension for the two highlighted Raman bands from the three traces in (a) are plotted with their respective location in the WLC curve indicated (inset (b)).

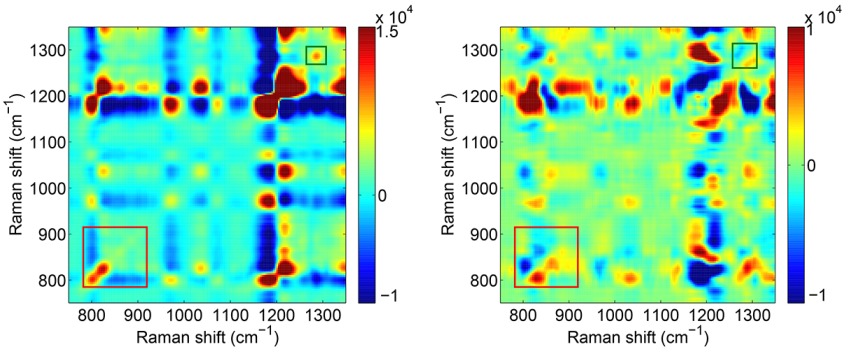


Figure 4.4: 2D correlation maps (synchronous **(left)** and asynchronous **(right)**) from Raman spectra of five different molecules that were extended. The signatures in both maps (boxes) together indicate a shift in peak position of the O-P-O Raman band (800 - 900 cm^{-1}) while the C-N cytosine base vibration (1280 cm^{-1}) remains constant during the molecule extensions. The large spots along the 1200 cm^{-1} axes are due to a large random peak that appeared at this position in two of the spectra which is common with SERS.

beads or DNA sample, could also float in to the excitation path causing brief "flashes" of an intense peak during the acquisition. For these reasons, we chose to ignore a few occurrences of a large peak at 1200 cm^{-1} which caused lines of spots in the maps at this position, because it does not match a particular DNA mode, its intensity was orders of magnitude greater than the next most intense peak, and it appeared in only a few of the spectra.

While intensities are solely dependent on the interaction of the chemical bond with light, band positions also depend on bond symmetry and polarizability which is a direct consequence of the conformation of the bond. Thus, Raman band position changes are expected to correlate with mechanical alteration of the local structure [83]. We conclude that the 800 cm^{-1} phosphodiester stretch mode is sensitive to molecule extension while the cytosine mode (1280 cm^{-1}) is unaffected. A similar mechanical sensitivity of the O-P-O stretch mode was previously observed, where an upward shift of the band is present between looser wound A-form DNA and tighter wound B-form structures [73].

The unique feature of this SERS arrangement is the ability to maintain the molecule free of attachments allowing for a more true measure of the structure. Based on geometrical optics, the Raman excitation focus diameter is approximately 400 nm, however, the SERS effect occurs only within the decay length of the evanescent field from the silver nanoparticles which has been reported to be in the range of 50-100 nm [84], equivalent to about 100 bp. While it is difficult to estimate the exact distance between molecule and SERS substrate, the DNA molecule must be within this decay length in order to observe Raman signal which is a product of multiple base pairs segments simultaneously optically excited. Maintaining the molecule in extended states minimized Brownian motion fluctuations that could cause the DNA to come in contact with the SERS bead, which would alter enhancement levels and potentially hinder extension based effects. As a result, Raman signals remained stable during periods of minutes without observing degradation of the molecule or SERS bead and spectra from

the set of DNA molecules was consistent.

One differing characteristic in the Raman spectra was the adenine peak, previously observed to be dominant in single stranded DNA [85] and from single adenine molecules in silver colloid solutions [76], but detected at smaller intensities (730 cm^{-1} and 1325 cm^{-1}) here. One reason for this discrepancy could be the less understood interaction between silver and the double strand which may cause shielding of the other bases from the silver nanoparticle plasmonic fields. Indeed, the adenine peak is the most responsive to metal nanoparticles adsorbed to uncoiled segments of DNA [86]. The molecule is not in contact with the metal surface as it is in other studies, thus, a proper measure of the SERS cross section or intensity of each component is not possible in this configuration, where the 730 cm^{-1} peak of adenine is traditionally dominant. Adenine signatures, and other bases, are most likely present in the Raman spectra, but the same intensity ratios previously observed with molecules adsorbed to metal surfaces cannot be repeated. The lack of adsorption produces an averaging of the base peaks instead, where the C-N stretch modes of the bases (1100-1700 cm^{-1}) seem to be stronger than the base ring breathing modes (< 800 cm^{-1}). With adenine having the highest SERS cross section, this peak intensity stands to experience the most dramatic reduction in the absence of adsorption to the metal, as is observed here.

4.3.2 Bond orientations alter with molecule extension

Fig. 4.5a is a plot of Raman peak position versus extension for all measurements. The plots contain the phosphodiester and cytosine base stretch modes that had distinct spots in the correlation plots with points from all of the tested molecules plotted together, owing to the known high mechanical uniformity between DNA molecules. Thus, the majority of error will lie in the Raman peak positions. A small spread in the extension points is due to uncertainty in the starting extension, because each molecule must be stretched to a finite initial length in order to be approached to the micron sized SERS substrate. The effect is immediately clear: the phosphodiester mode shifts while the cytosine mode remains unaffected during the DNA extension. Assuming a linear relationship, the most striking parameter is the sensitivity of the phosphodiester Raman band to the force-extension, 0.11 cm^{-1}/nm.

The MD modeling confirmed the measured Raman modes and allowed visualization of the modes' atomic motions. The Raman active modes were determined by computing the frequencies and Raman activities of a reduced structure, which consisted of a single phosphodiester network between a guanine and cytosine base. The phosphate vibration with the highest Raman activity in the vicinity of 800 cm^{-1} prominently involves a single O-P vibration but includes motion throughout the phosphodiester chain (C-O-P-O-C), while the 1280 cm^{-1} cytosine peak comes from a C-N vibration in the cytosine ring. Fig. 4.6a depicts the identified motion of the phosphate mode. Identifying the motion in the reduced system allowed us to locate the corresponding band in the full DNA structure computation and track it through the different extensions. The computed modes for each region (Fig. 4.2) of the full modeled DNA structure are plotted versus extension in Fig. 4.5b. Once again, the phosphate stretch mode shows sensitivity to DNA extension while the cytosine base remains unaffected.

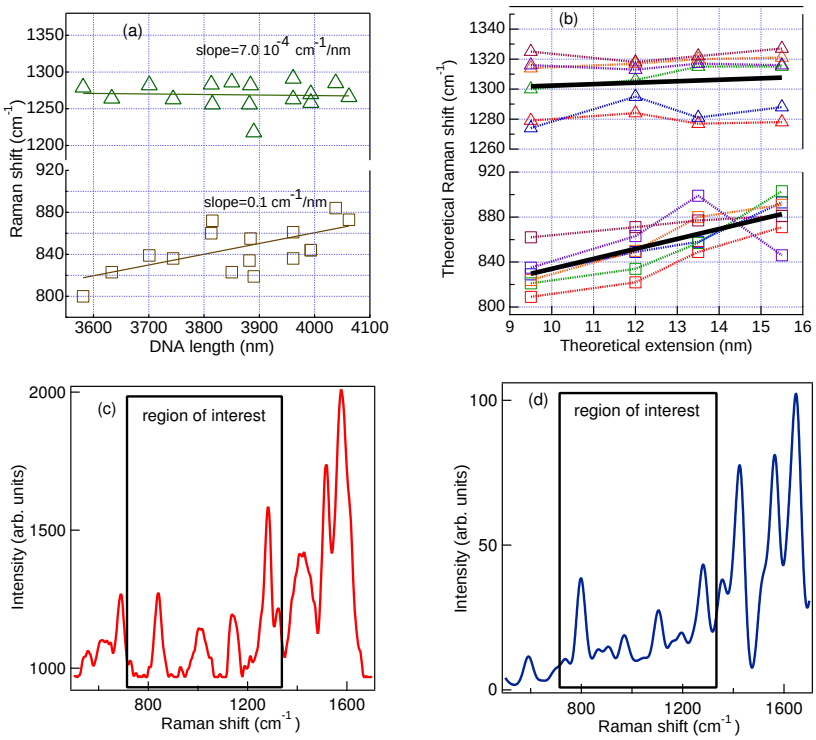


Figure 4.5: Experimental (**a**) and theoretical (**b**) plots of Raman peak position versus DNA extension for the O-P-O phosphate backbone and C-N cytosine base stretch modes. The points from all five molecules and six calculated regions are plotted together for the experimental and theoretical plots, respectively. The data sets are fit to a straight line and the slopes are included for the experimental data. The representative experimental Raman trace for the initial extension from Fig. 4.3a is plotted in full range **(c)** and shares the same x-axis with the theoretical plot of the initial modeled structure **(d)** where frequencies and intensities are taken from the full QM/MM and reduced structure calculation, respectively. All plots show agreement between experiment and theory.

Further agreement between experiment and theory is evident when examining the singular traces. Fig. 4.5c is the representative trace for the initial extension from Fig. 4.3a plotted in full range and compared to the calculated modes of the initial structure in Fig. 4.5d where the peak positions are taken from the full QM/MM calculation and Raman intensities from the reduced structure. The measured Raman bands can be confidently assigned to DNA components based on the consistency between experiment and theory and agreement with previous studies in the literature.

The Raman frequency of a vibration can change due to an alteration of the mechanical and electromagnetic components of the bond. The increase in the phosphate mode frequency is a result of an alteration in geometry and energetics of the system as the molecule is extended. Fig. 4.6b is an overlay of the initial and extended modeled structures with a single calculated region of the DNA chain isolated from the full modeled structure and the phosphorus atom as the common origin. The orientation of the O-P-O unit remains constant and the O-P-O bond angle shows no correlation with extension. The orientations of the carbons and adjacent sugars relative to the O-P-O unit change drastically, however. Nearest neighbors reconfigurations can affect the bond's polarizability and its interaction with light via Raman scattering by changing the symmetry of the local region. From an energetics standpoint, the region experiences a mechanical perturbation as the structure is extended. The total end effect of the above is a shift in the bond's Raman peak position.

The DNA structural alterations observed here at small forces should aide in understanding DNA-protein interactions in the entropic regime. Although DNA can be modeled as a semi-flexible polymer, the result shows that chemical bond orientations contribute to mechanical loading at all scales.

The technique is a novel combination of known methodologies that provides a new addition to the SMFS field. By directly measuring at the bond level with Raman scattering, these structural shifts can be observed at low forces without the need for complex particle tracking. Most importantly, with the constant advancement of SERS substrates, the stability and amplification level will continue to improve while maintaining the DNA strands free of optical probes and leaving them in a more ideal physiological condition. The combination of SERS substrates with advanced mathematical processing of the Raman signal, provides a powerful method to extract otherwise inaccessible information about the mechanochemical properties of single molecules in its natural environment.

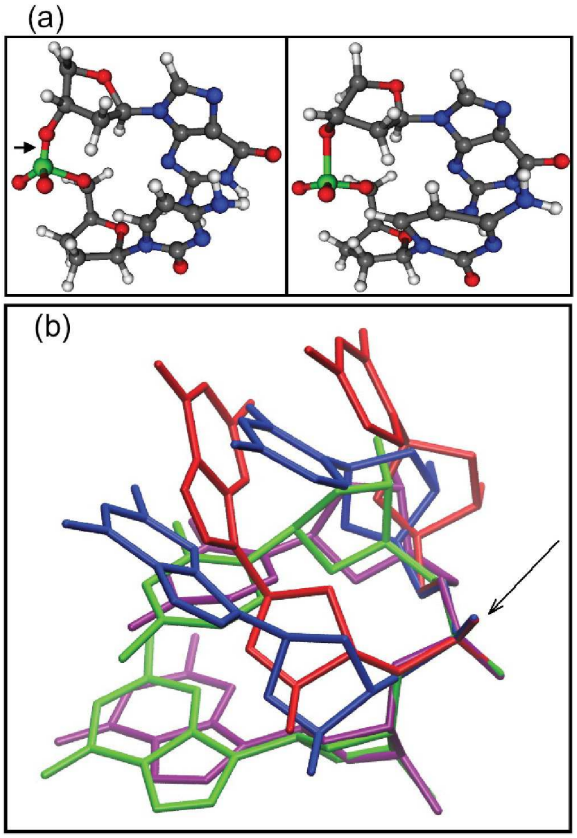


Figure 4.6: Theoretical snapshots (**a**) of the O-P-O vibrational mode responsible for the 800 cm^{-1} Raman peak taken from the reduced structure calculation used to determine the Raman activities. An overlay (**b**) of the calculated DNA structures focused at a single region, with the phosphorus atom (arrow) as the common origin: 97 Å (blue), 120 Å (red), 135 Å (green), 155 Å (purple).

5

Monitoring of local pH in photodynamic therapy treated live cancer cells using surface-enhanced Raman scattering probes

The local pH inside individual live glioma (U-87 MG) cancer cells were monitored after treatment by the photodynamic therapy drug, emodin. The cellular pH is tracked by the real-time measurement of the surface-enhanced Raman scattering (SERS) from a probe that is embedded in the cell. The probe is a micron-sized silica bead that is covered by nano-sized silver colloids, which enhance Raman signal, and 4-mercaptobenzoic acid (pMBA) whose molecular vibrations and resulting Raman spectrum are sensitive to pH. Visible excitation at different light dosages are used to activate the drug. Evolution in the SERS spectra caused by the intracellular pH changes were studied with 2D correlation. A complete picture of the Raman band's behavior was achieved, that permits to select the best Raman band capable to track the pH changes. The results indicate cell maintenance of internal pH and cell death at low and high light dosage, respectively. We demonstrate that these SERS probes are an effective tool for ex vivo pH monitoring in a live cell due to their high optical sensitivity and non-invasive usage. To the best of our knowledge, this study represents the first use of the 2D correlation technique to analyze SERS cellular spectra.

5.1 Introduction

Photodynamic therapy (PDT) is a viable technique for the treatment of certain types of cancer [87] The main engine of PDT is the production of highly reactive oxygen species (ROS) when activated by visible light [88] This oxidative stress

can lead to apoptosis or necrosis of the cell [89] where the mode of cell death is dependent on several factors such as the type, concentration, and localization of the drug, light activation dose, and the cell genotype.

The PDT drug that we utilize in this work is emodin (6-methyl-1,3,8 - trihydroxyanthraquinone). Emodin is a naturally occurring anthraquinone that was previously reported to have antitumoral activity against certain types of cancers [90, 91] Emodin induces apoptosis in various types of cells [92, 93] through a ROS-dependent mitochondrial signaling pathway.

In general, one of the main PDT targeted cell compartments for inducing apoptosis is the mitochondria [94] Through this pathway, the intracellular pH plays an important role because of the high pH sensitivity of the mitochondria that can lead to cell death [95] Thus, the intracellular pH can be a good measure of the stage of apoptosis and consequently the effect of the PDT drug. Healthy cells already contain variations of pH between the cellular compartments.

One successful method for determining intracellular pH is through the use of fluorescence. Modified green fluorescent proteins were utilized to determine the pH of the cytosol, Golgi apparatus, and mitochondria to be ~ 7.3, ~ 6.5, and ~ 8.0, respectively [96] Fluorescent dyes were also used in determining the acidic pH (~ 4.3) in lysosomes [97] Although these methods have provided good estimations of the intracellular pH, they maintain drawbacks in terms of signal level, toxicity of the dyes, and the relative complexity of expressing fluorescent proteins in the cells. In principle, the optical spectra of the photosensitizer itself could be utilized as a marker where pH sensitivity has been determined previously. However, as is the case with emodin, these optical characteristics are highly affected by aggregation [98] Most importantly, in light of the PDT mechanism, optical spectroscopy that requires visible excitation would further activate the drug during a measurement. An improved method of pH monitoring is needed that is less invasive while carrying the ability of locally measuring pH with good spatial placement and resolution, while not requiring visible wavelength optical excitation.

Based on coulombic interactions, vibrations of certain chemical bonds can be sensitive to the level of protonation in the local environment. For example, it was demonstrated that 4-mercaptobenzoic acid (pMBA), which strongly adsorbs on to silver and gold substrates, [99] can be used as an intracellular pH sensor by calibrating against intensity changes of sensitive peaks of the pMBA Raman spectrum [100] The main issue is the weak cross section of the Raman scattering which can be circumvented with the surface-enhanced Raman scattering (SERS) process where localized surface plasmons from nanoscopically textured metal structures are incorporated with the sample as in [18]. However, the use of nanosized metal colloids alone does not allow good spatial resolution when embedded in a cell. Metal colloids need to be distributed throughout the entire cell in order to ensure a consistent particle distribution that would validate a point measurement or a static pH map.

For example, such a pH map was previously constructed in a living cell where gold nanoagregates conjugated to pH sensitive molecules were accumulated in the studied cell through the process of endocytosis [101] In this case, the locations of the nanoagregates cannot be controlled leaving an inhomogenous spatial distribution of sensors which negatively effects the spatial accuracy of the differing pH regions. Additionally, the pH distribution in the cell nucleus is impossible to measure by this method, because metal nanoparticles cannot

enter this part of the cell [102,103] Other groups have demonstrated probes for pH monitoring using SERS active substrates, along with pH sensitive molecules, anchored to capillaries, however were unable to access the cell in a non-invasive manner [104, 105] Anchoring colloidal material to a substrate of micron scale, which we propose in this work, would allow for safe and accurate placement inside cells using current microinjection methods and in principle, could facilitate the use of combined optical trapping and laser injection techniques [106] for less invasive entry with better spatial control. Moreover, opposed to previous experiments, we conduct an active pH measurement over *time* which is a more accurate representation of the dynamical physiological processes in live cells such as those induced by drug interaction.

In this work, we demonstrate the use of novel pH probes to study the cellular response to the activation of the PDT drug emodin at different light dosages. To accomplish this, we construct a new local pH probe that is based on a micron sized silica bead covered with nanosized silver particles, which we have previously synthesized, characterized [107], and utilized for cancer cell studies [108] The pH sensitivity is added by attaching pMBA molecules to the bead's surface before they are passively embedded in cancer cells through overnight incubation. Emodin is then added to the cells and activated by visible light. After a given light dosage, the SERS of the embedded probe is excited by a near-infrared beam and spectra are continuously recorded. Alterations in the pMBA SERS spectrum are indicative of pH changes at the local environment of the probe, which we choose to be in the mitochondria, far away from the nucleus. The results demonstrate the reaction of the cell to the activated PDT drug, showing recovery to a balanced pH at low activation light dose and apoptotic cell death at high light dose.

The altering of pH is deduced from intensity changes in the SERS spectra which are revealed by 2D correlation spectroscopy analysis [59] This method is a powerful tool for isolating and extracting complex variations in spectral features of a system under external physical perturbations. This objective approach in assessing the data is essential for demonstrating the use of these pH probes as an effective tool in studying the biochemical changes of a live cell and its response to drug interaction.

5.2 Experimental methods

5.2.1 Materials

Emodin (1,3,8-trihydroxy-6-methylantraquinone), dimethylsulfoxide (DMSO), 3-aminopropyl trimethoxysilane (APTMS), silver nitrate ($AgNO_3$), trisodium citrate ($Na_3C_6H_5O_7.2H_2O$), 4-mercaptobenzoic acid (pMBA), hydrochloric acid (HCl), sodium hydroxide (NaOH), methanol (CH_3OH) and ethanol (CH_3CH_2OH) were obtained from Sigma-Aldrich. The pMBA stock solution was prepared in 100 % methanol. Dulbecco's modified eagle medium (DMEM), fetal bovin serum (FBS), trypsin/EDTA, penicillin and streptomycin were purchased from Gibco, Invitrogen (United Kingdom). Silica beads (2 μm diameter) were purchased from Micro Particles GmbH (Germany). The aqueous solutions were prepared with deionized water.

5.2.2 Probe construction

Previously, we completed a study on the silver particle attachment to micron sized beads, with a complete discussion of the critical parameters in the protocol [107] Briefly, the probes were prepared by attaching citrate reduced silver colloidal nanoparticles, [109] with an average diameter of 70 nm, to the surface of 2 μm diameter silica beads. The particles were anchored to the surface via a self-assembled alkyloxy silane monolayer (aminopropyltrimethoxy silane (APTMS)). Through judicious selection of reaction parameters, in particular the silane concentration and the agitation speed of the reaction vial that contained the beads and chemicals, a stable silane monolayer could be formed with a less than 20% packing density of the silver particles on the bead surface. Finally, to add the pH sensitivity, the silver coated silica beads were mixed with 10^{-3} M of pMBA, which creates a monolayer of the molecule on the metal surface through a sulfur atom and carboxyl group [99]

5.2.3 Cells

Human glioma cells (U-87 MG) were cultured as a monolayer and grown in DMEM containing L-glutamine (862 mg/l), sodium pyruvate (110 mg/l) and glucose (4500 mg/l), supplemented with 10 % FBS, penicillin (50 μg/ml) and streptomycin (50 μg/ml). For all measurements, the cells were plated in plastic Petri dishes (35 mm x 10 mm) with #0 cover glass at an optimal density of 5 x 10^5 cells/ml. During the measurements, the cells were enclosed in an incubator system (Live Cell Instrument) mounted on the microscope which maintained at 37 °C with 5 % CO_2 atmosphere.

5.2.4 Optical setup

A full description of the Raman system, which utilizes a 785 nm beam (10 mW) for excitation, has been given previously [110] An additional 532 nm beam was passed through a 40× top objective to facilitate the emodin activation. All spectra were collected with a 1 s acquisition time at a spectral resolution of 3 cm^{-1}. For all spectra, a five-point adjacent average smoothing was applied and the background was removed using an established method [111]

5.2.5 Calibration of pH probes

In order to correlate the SERS peaks to local pH, the pMBA Raman bands that are sensitive to pH were first identified and their intensity changes against known pH values resolved. This was done by measuring the SERS spectra from the probes in aqueous solutions at 8 different pH values between 2 - 7.5, accomplished by HCl/NaOH titration. A total of 15 spectra were recorded and averaged to produce a single Raman spectrum at each pH. The averaged spectra were analyzed with 2D correlation spectroscopy, which is described at the end of this section. Finally, a calibration curve was constructed of the pMBA marker band intensity versus pH.

5.2.6 PDT treated cell measurement

The cells were incubated during the night before the experiment with a fresh culture medium containing the pH probes. The micron sized probes enter the cells through phagocytosis [112] which is a specific form of endocytosis where larger particles are encapsulated by the cell via the membrane that reforms around the particle in order to transfer it inside. We have previously demonstrated the ability to passively embed similar probes in the cell line studied here while maintaining cell viability [108]

For the measurement, the cells were incubated for two hours with media containing 2 μM emodin with less than 1 % v/v DMSO. A dish of cells was then removed from the incubator and rewashed three times with fresh media to remove the excess emodin that had not diffused in to the cell. The samples were placed in the Raman setup where the PDT activation was performed using the 532 nm beam focused to a 3 mm diameter spot around the cell and the irradiation time held at a constant five minutes. By varying the laser power, two dosage levels were applied: low (1.0 J/cm^2) and high (6.0 J/cm^2). These values are below and above the level of energy that seems to assure photoactivation with consequent apoptosis from similar PDT drugs [113] For the control test, cells containing emodin were not photoactivated before the Raman measurement. For the Raman measurement, visual inspection was used to locate cells with a single embedded probe in the cytoplasm, away from the nucleus, where the majority of the PDT process is presumed to occur. Raman spectra from these probes were measured and for all cases, a total of 30 cells were tested and their spectra averaged to produce a single spectrum for each elapsed time after PDT activation. The averaged spectra were then analyzed with 2D correlation spectroscopy before pH versus time curves were constructed using the calibration curve described in the previous section.

5.2.7 2D correlation analysis methodology

The pH calibration and cell experiments produce an array of Raman spectra with two parameters (intensity and wavenumber) that change under a certain perturbation along a third parameter. The 2D correlation analysis identifies and extracts which Raman bands are changing the most with respect to the third parameter and relative to each other. The method has been described and applied previously in section 2.4 and in [59], therefore we describe only briefly here. All of the analysis is performed in the Matlab environment. The spectra were background subtracted, smoothed, and normalized to the out-of-plane vibration of the phenyl ring at 1070 cm^{-1} before entering the 2D correlation analysis.

The output of the analysis are contour graphs where the in-plane axes are the Raman frequencies and the third axis is the level of correlation between two Raman bands. The method produces two results for a given set of spectra: the synchronous and asynchronous graphs which refer to the in- or out-of-phase relationship between bands, respectively, over a third parameter. The third parameter that induce changes in the spectra can be any physical variable, such as pH, concentration or time, where each value matches a measured spectrum of the system. For the synchronous spectra, autopeaks occur along the diagonal and are mathematically equivalent to the autocorrelation of a band, thus,

signifying bands that are susceptible to change under the studied perturbation that is represented by the third parameter. Crosspeaks appear off diagonal and represent bands that increase or decrease together during the perturbation. The picture of the mode behavior is completed by looking at the asynchronous spectrum. Here, peaks on the diagonal do not exist and crosspeaks represent bands that are increasing or decreasing during the perturbation, but out-of-phase with each other.

In this work, we use 2D correlation analysis for the pH calibration of the sensor where the spectra were taken at different pH values for the solution, thus the third parameter is the pH. The 2D maps identify the band that is the most sensitive to the pH range and then a pH versus Raman intensity curve is constructed from the raw data that is a calibration of the probe. Following this, the same analysis is applied to study the pH changes of the living cell after PDT treatment where Raman spectra are acquired at points in time after the drug photoactivation, thus making time the third parameter with the understanding that the cell pH changes by an unknown quantity. The 2D maps reconfirm the marker band in the live cell and then the consequent Raman intensity versus time curve is translated to a pH versus time curve for each case using the calibration from the previous step.

5.3 Results

The synchronous and asynchronous 2D correlation maps are given in Fig. 5.1a and 5.1b, respectively, for the calibration of the pH probe. Autopeaks and crosspeaks are observed indicating Raman bands from the pMBA that are sensitive to the pH. The 1480 cm^{-1} peak is identified to be the best suited pH marker. Figure 5.2(inset) presents a characteristic spectrum of pMBA attached to the SERS probe with this band highlighted with an arrow. The resulting calibration curve, the normalized intensity of the 1480 cm^{-1} plotted against the pH of the solution, is shown in Fig. 5.2.

The synchronous and asynchronous maps are given in Fig. 5.3a, b and c for the control, low, and high dosage cell experiment cases, respectively. The pH sensitive 1480 cm^{-1} peak appears as auto- and crosspeaks again in all of the cases. Using the calibration from Fig5.2, the intensity changes of the marker band are converted to pH values and plotted over time for the control, low, and high dosage cases in Fig5.4a, b and c, respectively. We included the standard error, defined as the standard deviation divided by the square root of the number of tested cells. A series of camera images (Fig.5.4d) are included for the high dosage case to demonstrate the morphological breakdown of the cell over time after the emodin activation.

5.4 Discussion

5.4.1 pH calibration curve

The utility of the 2D correlations maps is evident in the immediate identification of the pMBA pH marker Raman band from Fig.5.1. Four distinct spots (1180, 1280, 1440, 1480 cm^{-1}) appear along the diagonal, the autopeaks, that signify

well resolved bands and sensitivity to the pH of the solutions. Previous experiments have utilized the 1440 cm^{-1} band as the pH marker [100, 101], however, we do not consider this choice to be suitable due to its coincidence with protein Raman bands in the cell [114] The synchronous map does not sufficiently determine the most suitable band, because it lacks information about the band's range of sensitivity. The asynchronous map (Fig.5.1b) completes this picture. The first two bands, at 1180 and 1280 cm^{-1}, have numerous crosspeaks in the asynchronous map indicating that they are changing out-of-phase with the rest of the spectrum. Thus, although these two bands are sensitive to the pH change, determined from the strong autopeaks in the synchronous spectrum, their intensities only change in a small portion of the entire pH perturbation range. Finally, the last band at 1480 cm^{-1} has only two asynchronous crosspeaks that occur at the positions of the first two chaotic bands (1180 and 1280 cm^{-1}), and so, experiences in-phase alterations of intensity, i.e., intensity changes throughout the full pH perturbation range. Thus, this is the correct choice for the pH marker for the cell experiments. The 1480 cm^{-1} peak is in the region of COO^- vibrations of the pMBA [100,101] where the bands' sensitivity to the number of dissociated carboxylate groups allows them to be responsive to pH alterations.

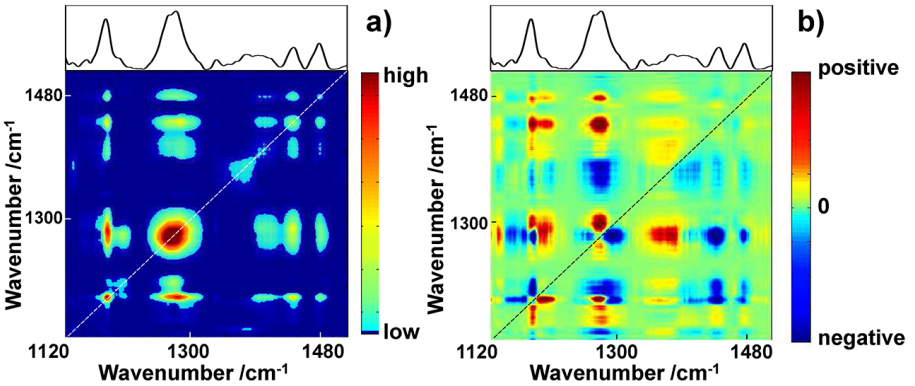


Figure 5.1: (Color online) Synchronous (a) and asynchronous (b) 2D correlation spectra of the pH probes, constructed from measurements of the probes in solutions of various known pH. The averaged spectra are laid along the top of the map. The dotted line runs along the diagonal of the maps, passing through the autopeaks of the spectra. All spectra are normalized to a pMBA phenyl ring vibration at 1070 cm^{-1} which is insensitive to pH, thus accounting for differences in signal enhancement between the probes. The correlation intensity values are not used, thus only a qualitative legend is given for the maps.

The calibration curve (Fig.5.2) fits well to a sigmoidal Boltzmann distribution which is expected in a pH titration calibration [115] and was previously observed in a pMBA SERS band pH calibration [101] For all curves, the normalization to the 1070 cm^{-1} band is essential because of its insensitivity to pH. Thus, any intensity variations observed here must be due to differences in the metal enhancement amongst the probes and should therefore be accounted for through this normalization [101]

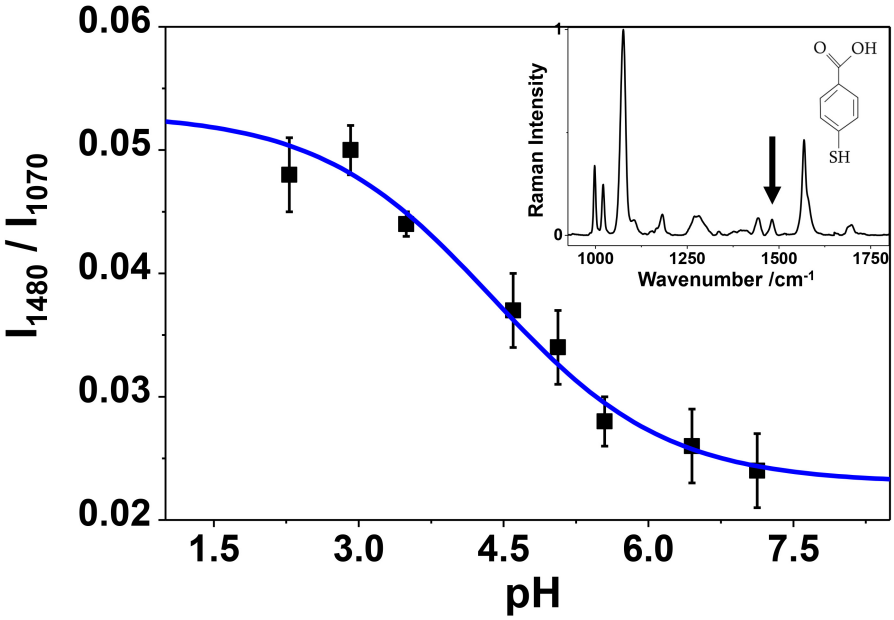


Figure 5.2: (Color online) Calibration curve of the pH probes obtained from the intensity of the carboxyl band at 1480 cm^{-1} normalized to the pMBA phenyl ring at 1070 cm^{-1} for each control solution in the pH range of 2.0 - 8.0. Error bars indicate the standard error from measurements of 15 different probes at each pH value. A typical SERS spectrum of pMBA (inset) measured at acidic pH is included with an arrow showing the pH sensitive Raman band (1480 cm^{-1}).

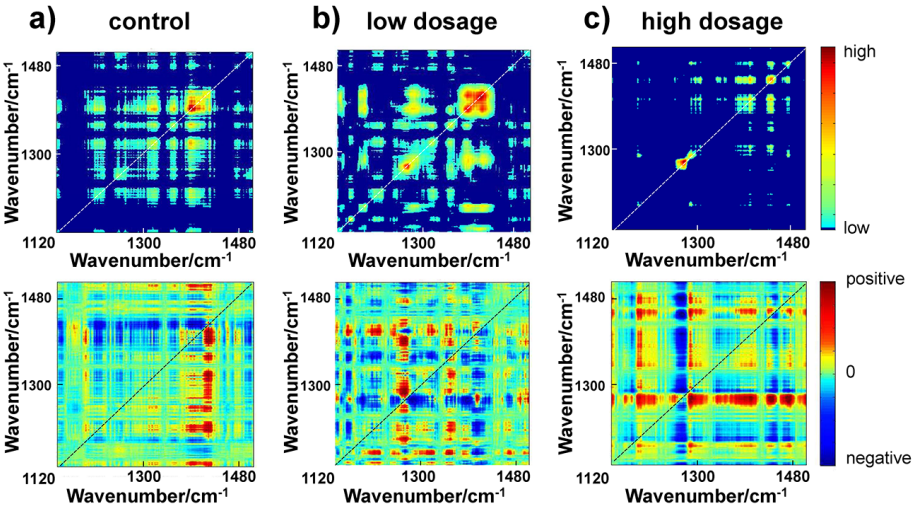


Figure 5.3: (Color online) Synchronous (top) and asynchronous (bottom) 2D correlation maps of the embedded pH probe for the control (a), low (b), and high (c) light dosage cases. The maps are constructed in the same way as in the calibration result of Fig. 5.1, including the normalization method of the Raman spectra. Once again, only a qualitative legend is given.

5.4.2 Identification of pH changes in PDT treated cells

For the control case, the correlation intensities are generally low and there are many crosspeaks in both the synchronous and asynchronous maps (Fig.5.3a). The pH marker band at 1480 cm^{-1} is present, however the intensity of its autopeak is low, indicating the relative pH stability of the cell in the presence of inactivated emodin. The highest autocorrelation intensity comes from the 1400 cm^{-1} band which was related to a vibrational mode of cytochrome C that is sensitive to the spin state of its heme structure [116] Cytochrome C is known to be involved in cell death through a transition from a mitochondrial membrane structure to a free form in the cytosol [117] Thus, intensity fluctuations of the cytochrome C Raman bands could appear depending on the activity of the protein. In this control case, the large autopeak at 1400 cm^{-1} indicates intensity fluctuations for this band over time, however, its almost continuous set of asynchronous crosspeaks means that these fluctuations only occur during a portion of the measuring period, as would be expected for the control.

The pH marker band autopeak intensity increases in the low dosage case (Fig.5.3b). The synchronous autopeak signifies an alteration in acid level and minimal crosspeaks in the asynchronous map show that the pH changes occur during the entire time period. The correlation intensities in both maps are low compared to the rest of the Raman band alterations in the cell. The pMBA peak at 1280 cm^{-1} has a higher autocorrelation, indicating more fluctuations than the 1480 cm^{-1} peak, however, its behavior in the asynchronous map is similar to what was observed in the pH calibration: numerous crosspeaks appear that conclude the alterations are occurring during a part of the total measuring period and therefore, are not synchronized with the fluctuating pH. The spin marker band of cytochrome C (1400 cm^{-1}) has a stronger autopeak in the

synchronous map and less crosspeaks in the asynchronous map than in the control case. The alterations of cytochrome C are seemingly higher and occur through more of the measuring period than before.

In the high dosage case, the pMBA Raman bands dominate the synchronous and asynchronous maps (Fig.5.3c). The bands at 1280 and 1480 cm^{-1} have high synchronous autopeaks. The 1480 cm^{-1} band has only a few asynchronous crosspeaks, indicating in-phase movement, that are shared with the 1180 and 1280 cm^{-1} bands which are again moving out-of-phase with the rest of the spectrum. Thus, pH changes are occurring throughout the measurement period. Additionally, asynchronous crosspeaks appear between 1480 cm^{-1} and 1440 cm^{-1} that were not present in the pH calibration. This substantiates the claim that there are other components in the cell that contribute to the 1440 cm^{-1} Raman band, which causes it to no longer be synchronized with the 1480 cm^{-1} peak and hence, should be ignored as a pH marker as stated previously here. The cytochrome C synchronous autopeak at 1400 cm^{-1} is significant once again with a low number of asynchronous crosspeaks. Therefore, like the pH marker band, this peak is also changing in intensity throughout the entire measurement range.

The 2D correlation spectroscopy is capable of providing more quantitative information because the correlation maps are derived directly from the Raman intensities and there is a temporal component that is present whenever time used as the third parameter. Thus, further information could be objectively extracted, for example reaction steps in a chemical process [118]. However, for this experiment, the objective identification of the pH marker band and confirming its presence in the live cell was the critical result of the 2D maps. Utilizing this approach to extract other quantitative and time-dependent parameters is a consideration for our next studies. From synchronous 2D correlation map interesting quantitative information can be extracted. For instance, the intensity plot of a line in the synchronous map will provide an approximate spectrum about bands that are highly correlated and thus, originating from related molecular structures. The intensity of these plots will provide an estimation about the amount of change this bands have undergo during the process. Selecting the most different "lines" in the synchronous 2D correlation map will be equivalent to the loadings in Principal Components analysis. This, will be analyzed in detail in the next chapter for other application.

5.4.3 Monitoring of pH in PDT treated cells

Following the band identification, the pH versus time curves further illustrate the intracellular response to the drug activation. In the control case (Fig.5.4a), despite the presence of inactivated emodin, the pH of the cytoplasm is steady at a value of 5.5 . The cytosol pH is typically at neutral, however the position of the probe, whose surface is in contact with the mitochondrial membrane (pH = 8) and lysosomes (pH = 4.3), causes a lowering of the initial measured pH in all three cases. The cell pH will vary slightly around the average value due to constant ion exchange during normal physiological processes. In the presence of inactivated emodin, no morphological changes of the cells were observed through camera images (data not shown).

This pH variation grows in the low dosage case (Fig.5.4b) while still keeping the same average value of 5.5. The emodin has now been activated with low

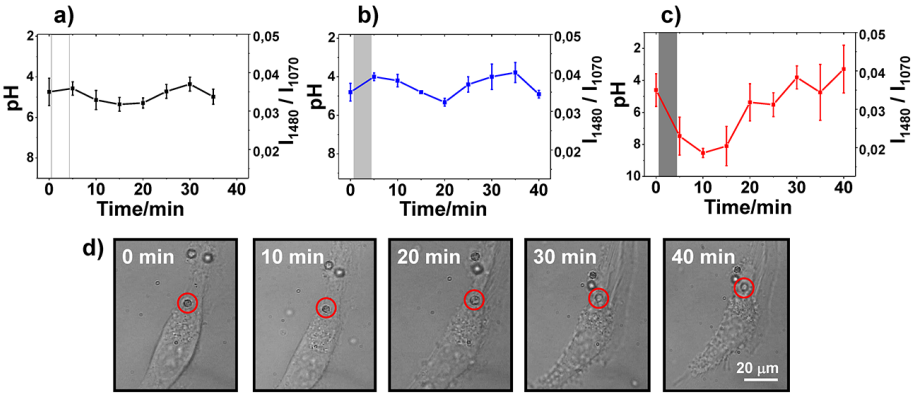


Figure 5.4: The constructed pH curves versus time for the control (a), low (b), and high (c) light dose cases. The shaded regions indicate the time period of light activation (absent for the control case). Camera images (d) of the cell morphological changes after photodynamic treatment at the high dose level (6.0 J/cm^2) are included at the indicated times in order for direct comparison to the measured pH in (c).

energy which causes some ROS production leading to acidification in the cell. However, this disruption is not enough to physiologically unbalance the cell, i.e., the recovery mechanisms are able to keep the pH fluctuating around the natural condition as in the control case. Once again, cell camera images (data not shown) do not reveal any morphological changes.

The conditions change for the cell in the high dosage case. The ROS production is expected to increase substantially at this light activation level [119] which has two major effects in the cell: fast acidification and pore opening in the mitochondrial membrane. The two together cause a cell recovery induced influx of potassium, sodium, and calcium ions that increase the pH of the cell. Second, cytochrome C and apoptosis factors pass from the mitochondria to the cytosol. These components, since they are typically at a pH of 8 in the mitochondria, raise the pH of the intracellular medium as well while triggering a cascade of processes that induce cell death [95, 120] Thus, as is observed in Fig.5.4c, there is an initial increase in the pH. The 2D correlations maps show the band fluctuations of cytochrome C which is reasonable considering its involvement in the cell death process, where apoptotic processes have been invoked and the cell continues through the stages of cell death. Membrane swelling causes more components to spill in to the cytosol and cellular compartments begin to break down. Lysosomes, with an internal pH of 4.3, break down to components and lower the pH of the intracellular medium. Thus, the measured pH lowers after the initial increase, however, it moves beyond the stable initial value in this case. Morphological changes in the cell are observed (Fig.5.4) during this pH lowering period and coupled with the known effects of emodin mentioned here, it can be concluded that this cell death occurs through apoptotic pathways.

The role of pH in the regulation or signaling of cell death processes is still a subject under study. However, what we have shown here is a less invasive approach to monitoring the cell recovery and death process in response to PDT treatment. The pH probe can, in principle, be placed anywhere in the cell with

current microinjection techniques, which avoids the need to distribute metal colloids or fluorescent dyes throughout the entire cell to accomplish the same point measurement, as was the case in previous works [97, 101] We have also demonstrated an approach where the application of Raman probes can be utilized to study dynamic physiological processes in live cells in real time. This key point is critical for future experiments that look to observe cellular responses to external perturbations. In this experiment, the response of the cell in the high PDT dosage case agrees with previous knowledge of the physiological response to PDT treatment, demonstrating the effectiveness of these pH probes here. Unlike previous studies using similar methods, 2D correlation spectroscopy provides an objective approach to assess biological spectra such as those related to the dynamics of the pH of the intracellular medium.

5.5 Conclusion

We have demonstrated the monitoring of pH dynamics after photodynamic treatment in a live cancer cell using SERS probes. Micron sized silica beads were partially covered with nanosize silver colloids that enhance the Raman signal from pH sensitive pMBA molecules which are adsorbed on the bead surface. The probes were passively embedded into the cell which was photodynamicaly treated by photoactivation of emodin at low and high light dosage levels. Peak intensity changes of a pMBA pH Raman marker band were observed in both cases that were not seen in a control test. The consequent pH versus time curves reveal the ability of the cell to recover at the low light dosage and apoptotic cell death to be induced in the high light dosage case. These local pH probes offer a minimally invasive approach for studying the dynamics in a live cell. Due to their stability, in regards to both physical position and optical activity, the probes offer a particular advantage for observing physiological action over time as has been demonstrated here in terms of monitoring the cellular response to the perturbation of drug interaction.

6

Mechanochemistry of single red blood cells monitored using Raman tweezers

The main function of red blood cells (RBCs) is redistribution of oxygen throughout human body. The high deformability of RBC is essential in the gas transport [121]. In particular, it allows RBCs to flow through microcapillaries with inner diameter typically less than half of the cell diameter. This is possible due to characteristic structure, shape, and mechanical properties of the cells [122]. Specifically, altered membrane properties can lead to human diseases such as malaria, spherocytosis, elliptocytosis or sicke cell anemia [123]. Thus, understanding the molecular origin of the extraordinary mechanical properties of RBCs is of critical interest. In this chapter, we combine optical tweezers with Raman spectroscopy to directly extract biochemical information. Two microparticles were biochemically attached to a red blood cell at diametrically opposite parts and held by optical traps allowing to impose deformations. The cell deformation was monitored from the microscopy images. Raman spectra of the cell under tunable deformations were studied. In the past chapters, we used 2D correlation to analize band behaviour and correlation when force loads or pH changes were applied to the sample. Here, the combination of two statistical techniques (PCA and 2D correlation) are exploited to identify, monitor and study band behaviour and correlation in the dynamic Raman spectra of RBC when it is being stretch. Principal Component Analysis distinguishes the most prominent changes in spectra while 2D correlation technique monitors the evolution of Raman bands during stretching. The measurements show significant changes in the cell chemical structure with stretching however the changes saturate above 20 % of cell deformation. Mechanical deformation of the cell mainly affects the bands corresponding to hemoglobin but contributions from spectrin and membrane proteins can not be excluded. The saturation of bands at higher deformations suggests some structural relaxation that RBC has to undergo to bear extra load. The results confirm widely accepted belief that spectrin released from membrane proteins allows for significant shape changes of the

cells. We therefore tentatively suggest that interaction between membrane and cytoskeleton during deformation can be efficiently probed by confocal Raman spectroscopy, in particular via the peak around 1035 cm^{-1}.

6.1 Introduction

Linking spectroscopic techniques with optical tweezers opens unique possibility to directly extract biochemical information at single-cell level under controllable mechanical conditions

Erythrocytes are built of lipid bilayer which contains many transmembrane proteins and underlying cytoskeleton. It forms flexible biconcave disks which, consist mainly of hemoglobin but lacks a cell nucleus and most organelles. Chemical content of RBCs provides information about human health conditions. Therefore blood is one of the most common sample submitted nowadays for medical diagnosis.

Since, in human body, erythrocytes adopt different squeezing/stretching states, mechano-chemical processes associated with deformation of RBCs have been of particular research interest. First studies on deformability of RBC were performed using pipette aspiration [124, 125] and permitted to observe a high degree of deformability. This extraordinary property allows RBC to squeeze through microvasculature and transport oxygen during blood circulation [122]. Micromanipulation techniques such as optical or magnetic tweezers have allowed for more sophisticated microrheological studies. Most of them were focused on measuring the static changes of cell length [4-7] or area expansion [126] in response to a deforming force. More recently two point microrheology measurements of single RBC [127] showed the frequency dependence of complex stiffness confirming strong nonlinear deformation of cells. Optical tweezers provide an excellent method for probing single cells at different stretching states, giving potentially powerful tool for diagnosis of diseases [128], since mechanical properties of RBCs can be affected by malaria, spherocytosis, elliptocytosis or sickle cell anemia [123].

Understanding of molecular origin of extraordinary mechanical flexibility or nonlinear response of the single cell requires more sophisticated experimental approach and extensive theoretical simulations. In this context combination of optical tweezers with single cell spectroscopy seems to be very promising. Linking spectroscopic techniques with optical tweezers opens unique possibility to directly extract biochemical information at single-cell level under controllable mechanical conditions. The simultaneous use of external force and spectroscopic detection can also provide direct insight into molecular changes caused by mechanical deformations. Here, Raman spectroscopy is of particular interest because of its high chemical resolution and relative simplicity. It yields rich, multi-peaks vibrational spectra containing structural information about RBC's basic constituents [12-17]. As previously demonstrated, Raman spectroscopy combined with optical tweezers is extremely valuable tool for reliable estimation of cell's vitality [18-21]. Sensitivity of this method is high enough for identification and monitoring of changes in intra-cellular components induced by cell stretching. In particular, direct evidences for deoxygenation of cells with stretching was shown [129].

Till now most of Raman studies on RBC were performed with cells being di-

rectly trapped with focused optical beam. This is not an ideal approach because of potential light induced cell damage within the trapping volume. Reducing the power of trapping beam leads to significant diminution of Raman signal-to-noise ratio, extensive averaging at the expense of single experiment acquisition time is required. In this case external forces applied to cells by optical tweezers are also strongly reduced. The problem can be avoided by attaching RBCs to chemically functionalized microbeads, which might hold the cell during Raman scattering experiments [130]. Although heating via beads (size dependent process) or effects of scattered light can not be still completely excluded, the risk of direct light-induced damage by trapping beams is significantly reduced. Therefore sensitive experiments at higher forces and lower noise for longer time periods become possible.

We studied changes in molecular structure of a single RBC when it is gradually stretched by optically trapped beads attached to the cell. We extended our previous Raman measurements [129] made in our group to many different RBC's stretching states, which required longer time for experiments. The applied force is intended to simulate step-by-step deformation experienced by cells in normal conditions and induced by blood flow as they squeeze through microvasculature. To improve further the sensitivity of the experiments and facilitate their interpretation, we used also statistical techniques (2D correlation and principal component analysis (PCA)), that permit us to observe previously inaccessible changes in Raman spectra. The purpose of this work is to unravel direct relationships between mechanical deformation of RBC and chemical changes occurring in the cell structure on molecular level.

6.2 Materials and methods

6.2.1 Experimental set-up

The experimental setup is a combined dual-beam optical trapping system with confocal Raman spectrometer and has been described previously in section 3.1.1 of chapter 3 as well as in [34, 129, 131].

Briefly, a 785 nm beam was used for excitation of Raman spectra with average power of 5 mW at the sample. Samples were injected into a custom-made fluid chamber. The chamber was placed on an inverse Olympus IX 51 microscope equipped with a 100× 1.3 NA oil immersion objective and a micrometer controlled stage. The backscattered light was collected by the trapping objective, passed through a holographic notch filter and a confocal system with a 100-μm pinhole. The spectrometer had a 600 lines/mm grating and was fitted with a thermoelectrically controlled charge-coupled device (CCD), cooled to -60°C. A CCD camera attached to the microscope provided optical images used for calculations of cell deformation.

Raman spectra were recorded with a spectral resolution of 3 cm^{-1} for 30 seconds at each acquisition. Measured Raman spectra were background subtracted, normalized using Multiplicative Scatter Correction (MSC), smoothed using smoothing spline technique [81] and median centered before performing 2D correlation and PCA. Analysis was done in Matlab platform.

6.2.2 Sample preparation

About 30 μl of blood was obtained from a healthy donor and washed two times. Final volume was diluted in 600 μl of Alsever's solution and stored at 4°C. The sample was used within no longer than two days. 3μm COOH functionalized polystyrene beads were purchased from Spherotech (CPX-30-10). The beads were washed four times in 50 mM MES buffer (Sigma Aldrich M1317). Their functionalization was obtained with Lectin (Sigma Aldrich L9640) using chemicals and protocol provided with PolyLink Protein Coupling Kit from Bang Labs (PLO1N). After functionalization, beads were again washed 4 times, and finally resuspended and stored in MES buffer. For measurement, 20 μl of resuspended beads were washed in Alsever's solution.

The beads were attached to a single RBC by proceeding as follows: two beads were optically trapped while RBC was held by the optical trap generated by the Raman excitation beam. The two trapped beads were carefully approached towards the diametrically opposite parts of RBC, and the connection of RBC with the beads was obtained in around 5-10 seconds. The position of beads were then adjusted with the help of movable mirrors to assure, that the Raman excitation beam is hitting the edge of RBC. Cells were stretched by moving one of the trap. Measurements of Raman spectra were done on 15 different cells. Raman spectra of most of them behaved by application of mechanical deformation in a similar fashion.

6.2.3 Statistical processing of the Raman spectra

In order to get insight into spectral changes induced by cell deformation, in particular those which are inaccessible by direct visual inspection, we used multivariate analysis [132]. Multivariate methods have been developed to study complex data with many variables analyzed simultaneously. Specifically, Raman spectra from biological samples are a good example of such data, when extraction of important, characteristic variables from some parts of Raman spectra is quite complicated or even impossible without statistical methods.

Principal component analysis (PCA) reduces the dimensionality of the data by finding a set of orthogonal coordinates, principal components (PCs), which accounts for the maximum variance in the Raman spectral dataset, and describes the major trends in the data. In other words, with PCA it is possible to extract the principal contributions in spectral changes due to molecular deformations by RBC stretching. Loading plots represent the characteristic spectra in principal component coordinate system.

Other statistical technique used in this work is 2D correlation analysis [59]. In the previous chapter, we demonstrated this processing by monitoring local pH in photodynamic therapy-treated live cancer cells with surface-enhanced Raman scattering probes [133]. Briefly, the 2D correlation analysis identifies and extracts Raman bands which are changing the most with respect to the external perturbation and relative to each other. The output of the analysis consists of contour graphs, where the in-plane axes are the Raman shifts and the third axis is the level of correlation between Raman bands. The method produces synchronous and asynchronous graphs, which refer to the in-phase and out-of-phase relationship between bands, respectively, as a function of cell deformation. For the synchronous spectra, autopeaks occur along the diagonal of the plots, and

are mathematically equivalent to the autocorrelation of a band, thus signifying bands that are susceptible to change under cell deformation. Cross peaks appear off diagonal and represent bands that increase or decrease together during the perturbation. The picture of the modes behavior is completed by the asynchronous spectrum where autopeaks and cross peaks represent bands that increase or decrease with perturbation but out of phase with one another.

In this chapter we take the advantage of the combination of both mathematical methods. First, PCA and 2D correlation identify the main changes in the spectra. Second, PCA analysis monitor the band intensity changes along the process. Finally, band effects such as broadening or shifts are studied with 2D correlation.

6.3 Results

Figure 6.1 shows a typical set of Raman spectra of single RBC. Raman signals at fifteen different stretches were measured. Raman spectra of RBCs depend strongly on excitation wavelength [134, 135]. At 785 nm excitation the bands associated with proteins can be observed in addition to those associated with the porphyrin macrocycle [136]. Presented Raman spectra are characteristic of single RBC and for unstretched cells have been discussed in details in [135–137].

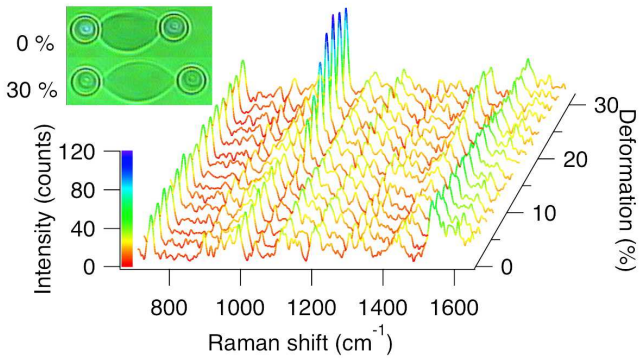


Figure 6.1: 3D representation of Raman spectra of single RBC at 15 different cell deformations. Inset shows microscope image of RBC with the beads attached, at rest and stretched by 30 %.

Deformation was conventionally measured using image analysis by calculating the distance between the beads attached to the diametrically opposite parts of the cell. It was calculated in the following way: $Deformation(\%) = (\Delta L/L_0) \times 100$ where, $\Delta L = L - L_0$ is the difference between diameters of stretched (L) and relaxed (L_0) cell. Even without statistical processing Fig. 6.1 demonstrates effects of the cell stretching on some Raman bands (see also color gradient of the plot), in particular the bands centered at about 991 cm^{-1}, 1035 cm^{-1} and 1442 cm^{-1}. 1035 cm^{-1} peak is associated with in-plane CH_2 asymmetric mode of vinyl substituent in porphyrin [136, 137] and/or Phenylalanine (Phe) [138,139]. The bands at 991 cm^{-1} and 1442 cm^{-1} are assigned to ν_{45}/Phe and CH_2/CH_3 in plane modes of haem, respectively [137]. At low deformations

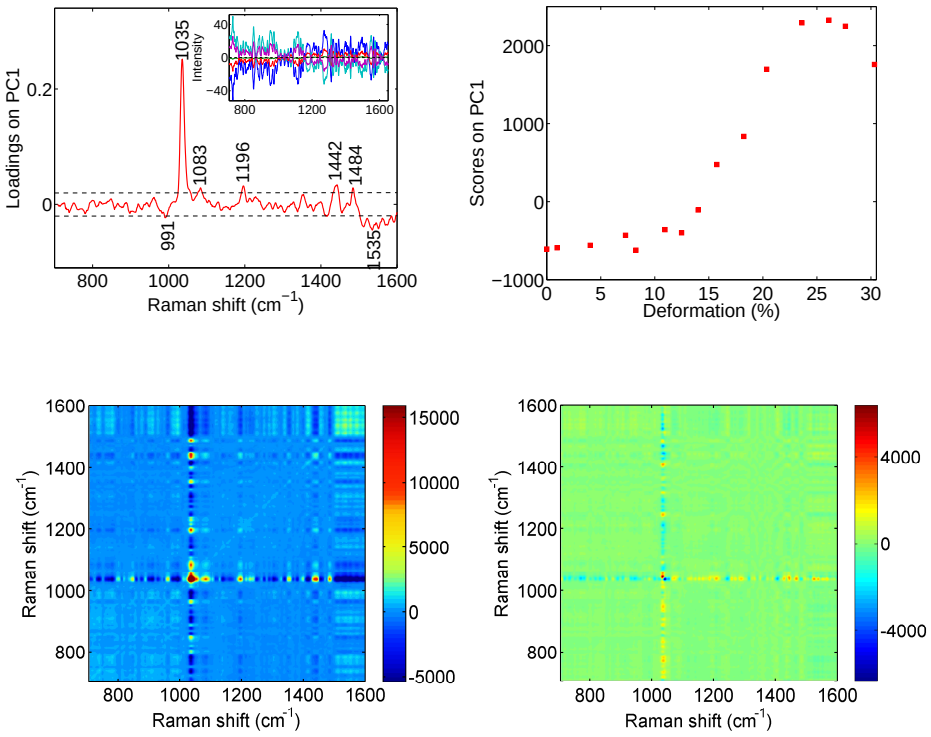


Figure 6.2: PCA and 2D correlation analysis of measured Raman spectra. **Top:**Left: Loading plot with threshold (dashed lines) estimated from experimental noise analysis. The inset demonstrates data used to define the threshold. Right: Scores plot showing overall intensity of all bands above the threshold with increasing cell deformation. **Bottom:** 2D correlation analysis for whole measured spectral window (synchronous map (Left) and asynchronous map (Right)). Cross correlation peaks can be seen in synchronous map indicating bands correlated during stretching.

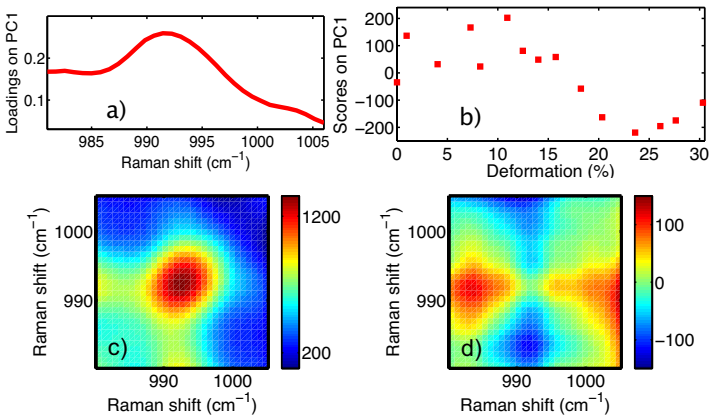


Figure 6.3: Statistical analysis for Raman band at 991 cm^{-1}. **Top:** PCA analysis (loading plot (a) and scores plot (b)). **Bottom:** Expanded view of 2D correlation maps from Fig. 6.2 (synchronous map (c) and asynchronous map (d)).

the 1035 cm^{-1} and 1442 cm^{-1} peaks are unchanged. Their growth starts when the cell deformation is more than 10%, and the growth is saturated at higher cell deformations (above 20 %).

The most remarkable finding of our measurements - the extremely high sensitivity of 1035 cm^{-1} peak to the cell extension - has not been reported yet. It permits us to suggest the monitoring of this peak intensity as a marker to characterize internal deformations in the cell in possible lab-on-a-chip applications. Other Raman bands also change with cell deformation, but their behavior could be revealed only using statistical techniques (see below).

6.4 Discussion

To obtain the dynamics of conformational changes we first performed the PCA analysis using data in the whole spectral range (Fig. 6.2). The score plot Fig. 6.2(b) shows that at small cell deformations (up to 10%), spectral changes are negligible (see also Fig. 6.1). Hence, the maximum intensity variation of spectra obtained at five initial values of deformation may be used to estimate a threshold to discard the peaks generated by the electronic and mechanical noise and errors by numerical calculations. Then we can consider only those Raman bands which are above this threshold.

To establish the numerical value of the threshold in the loadings plot Fig. 6.2(a) we performed PCA analysis of the Raman spectra obtained for the first five values of deformations. Multiplying each data point on the scores plot gives us the intensity of the corresponding spectra with respect to the median spectrum. The maximum intensity was found to be around 40 units (inset of Fig. 6.2 (a)). To translate this threshold (in intensity) to the loadings plot of complete set of spectra we divided it by the highest value in the scores plot of Fig. 6.2 (b). The threshold (± 0.02 in loadings) was then used for further

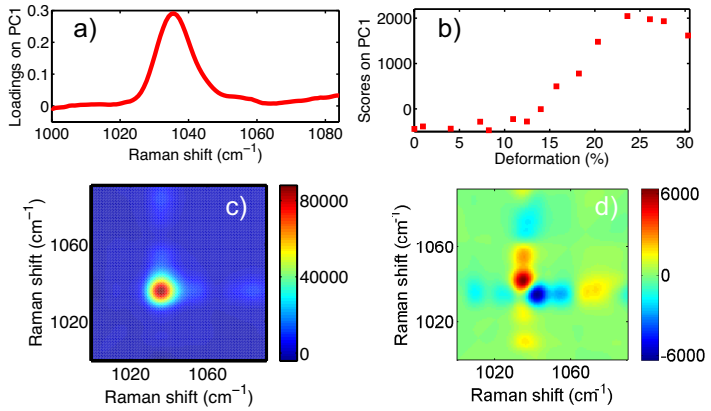


Figure 6.4: Statistical analysis of Raman band at 1035 cm^{-1}. **Top:** PCA analysis (loading plot (a) and scores plot (b)). **Bottom:** Expanded view of 2D correlation maps from Fig. 6.2 (synchronous map (c) and asynchronous map (d)).

analysis.

We conclude that the loadings of the first principal component, which captures almost 80% of the total variance, are dominated by bands centered at 1083, 1196, 1484, and 1535cm^{-1} along with above mentioned bands at 991, 1035, and 1442 cm^{-1}. The former four bands can be assigned to $\delta(= C_bH_2)_{asym}$, $\nu_5 + \nu_{18}$, ν_3 and ν_{38} modes respectively [136, 137, 140].

These selected peaks are also consistent with the most prominent features observed in the complementary technique, 2D correlation analysis, where strong correlation between these bands is observed (Fig. 6.2(a) and Fig. 6.2 (b)). Positions of cross-correlation peaks in synchronous map are consistent with the PCA analysis. This confirms that this set of Raman bands is changing synchronously with the cell stretching. Therefore, two statistical methods, with different mathematical background provide similar results regarding the parts of the Raman spectra sensitive to force loads. Synchronous 2D correlation map give information about the negative or positive correlation among bands. Negative cross peaks signify negative synchronous correlation, and therefore bands change in intensity in different directions at the same time. On the contrary positive cross correlation peaks in synchronous map, means that the two bands are changing in intensity in the same direction synchronously. PCA can give similar picture taking a look on the loadings plot. The first loading plot of figure 6.2 contains positive and negative portions meaning that all bands in the loading are correlated changing synchronously. However, negative bands will change in opposite intensity direction than the positive ones. If we compare 2D synchronous map with the PCA loading we can observe similar results. Bands 991 and 1535 cm^{-1} are inversely correlated with bands 1035, 1083, 1196, 1442 and 1484 cm^{-1}. Further information can be extracted, in principle, from asynchronous 2D correlation maps, like the order in which changes are occuring. But, due to the complexity of the spectra studied, where band witdh and positions are changing, this information can not be directly extracted for this case.

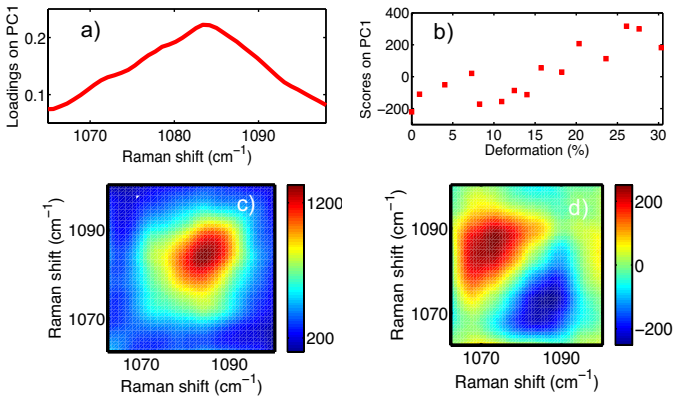


Figure 6.5: Statistical analysis for Raman band at 1083 cm^{-1}. **Top:** PCA analysis (loading plot (a) and scores plot (b)). **Bottom:** Expanded view of 2D correlation maps from Fig. 6.2 (synchronous map (c) and asynchronous map (d)).

However asynchronous 2D correlation maps will be very valuable to extract band behaviour (position and width) along the stretching exerted.

In order to characterize the intensity behavior of correlated bands with deformation, we performed the PCA analysis with reduced spectral windows to include the selected bands only. Resulting scores plots are intended to mimic the band intensity behavior of single band with stretching. In a way, PCA acts as a fitting that permits us to monitor the selected Raman bands even when the spectra are noisy and without distinguishable features. Results for bands at 991, 1035, 1083, and 1196 are shown separately in Fig. 6.3(b), 6.4(b), 6.5(b), and 6.6(b), respectively, where all the bands exhibit a similar tendency with the cell stretching. Raman peak at 1442 cm^{-1} does not give new information and its behavior is very similar to Raman peak at 1196 cm^{-1}. If the cell is deformed less than 10%, these bands exhibit almost constant Raman intensity. Higher mechanical forces lead to almost linear increase (or decrease) dependence with stretching. This behavior is observed up to about 20% of cell deformation. Finally, the bands saturated for deformations exceeding 20%.

However, the bands at 991 and 1535 cm^{-1} showed inverse behavior with increased deformation, i.e. their intensity decreased in the deformation range of 10-20% and saturated above 20% (see, for example, data shown for 991 cm^{-1} band in Fig. 6.3.

To study the dynamics of band's position and their shapes we took advantages of 2D correlation method. Figures 6.3, 6.4, 6.5, and 6.6 show 2D maps for the individual bands. Similar trends for some groups of peaks are clearly visible. In the asynchronous map we observe a characteristic pattern for band broadenings formed by four lobes in a cross-like shape (fig. 2.12). It can also be seen that this pattern is asymmetric. This is caused by the small shift of the Raman band maxima to higher wavenumbers (Fig. 2.13). At 1035 cm^{-1} we observe a similar pattern but with opposite signs in asynchronous maps indicating the band narrowing (Fig. 2.15) and shifting to smaller wavenumbers with stretching. Bands at 991 and 1535 cm^{-1} have also the same cross-like

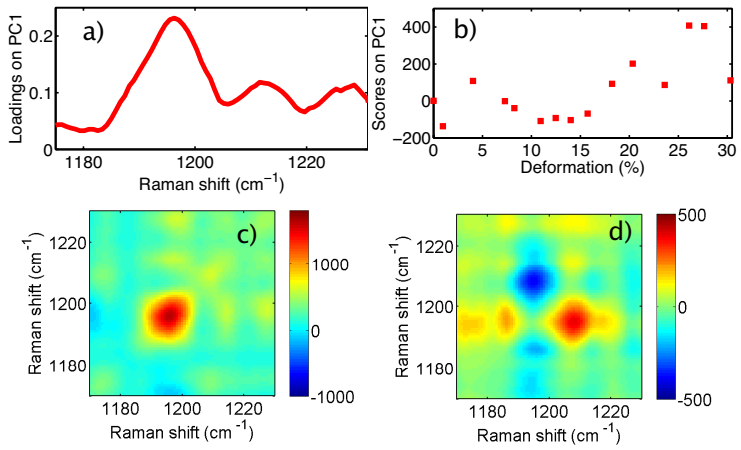


Figure 6.6: Statistical analysis for Raman band at 1196 cm^{-1}. **Top:** PCA analysis (loading plot (a) and scores plot (b)). **Bottom:** Expanded view of 2D correlation maps from Fig. 6.2 (synchronous map (c) and asynchronous map (d)).

shape but their intensity decrease (see, for example, data shown for 991 cm^{-1} band, Fig. 6.3). Comparing these patterns with mathematical simulations of different band behavior (included in section 2.4), we suggest that 991 cm^{-1} band broadens and shifts to lower wavenumbers (Fig. 2.14). 1535 cm^{-1} band broadens and shifts to higher wavenumbers (data not shown). A different pattern was observed for 1083 cm^{-1} band, which exhibits two lobes at both parts of the diagonal in the asynchronous map (Fig. 6.5). This indicates that the width of the band remains the same but a shift to lower wavenumbers occurs. All observed peak frequency shifts are rather small and do not exceed 5 cm^{-1}, close to the spectrometer resolution.

Raman signals collected from RBC are average signals from the confocal volume consisting of large number of molecules that include haemoglobin, spectrin and other biomolecules (fig. 6.7). The stretching might move the cell, but obviously the basic constituents of cell under investigation do not change. In other words, net effect of stretching in Raman signals should not depend on the probed region of the cell. Moreover, the measurements were carried out at many stretching states and for various cells. These cells demonstrated unidirectional changes in their bands with stretching. That is why we can exclude that the cell movement within focal volume has any noticeable effect on Raman spectra.

The main constituents of RBC are hemoglobin and cytoskeleton (including spectrin), all surrounded by thin membrane with many embedded proteins. RBC vibrational spectra, therefore, consist mainly of bands characteristic of porphyrin macrocycle and various proteins. The shape of cells is preserved by spectrin network and ankyrin proteins, which anchor cytoskeleton to membrane. About 10^6 hemoglobins per cell are strongly bound to the inner wall of RBC membrane with possible intercalation [142, 143]. Stretching of the cell is ex-

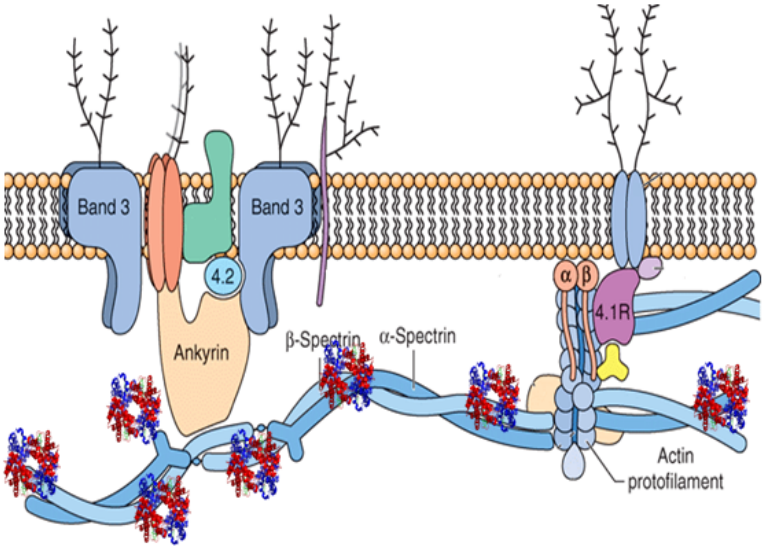


Figure 6.7: Drawing of the RBC inner molecular structure close to membrane. The shape of cells is preserved by spectrin network and ankyrin proteins, which anchor cytoskeleton to membrane. Hemoglobins are strongly bound to the inner wall of RBC membrane with possible intercalation. Adapted from [141].

pected to affect mostly membrane and cytoskeleton, which absorb most of the forces, and also hemoglobins, in particular those which are bounded. Raman excitation close to the edge of RBC allowed us to get signals from the above mentioned cell's regions which undergo maximum deformation.

The statistical analysis, in particular the scores plots, shows the similar behavior of the Raman bands: their intensity remains almost unaffected at low cell deformations, and changes occur at intermediate forces following by band saturation at highest deformations achieved in the experiments. Most of these bands can be associated with different vibrational modes in hemoglobin. It indicates that RBC's chemical structure is strongly perturbed by deformation. Such activation/deactivation of selected Raman transitions can be caused by significant deformation-induced changes in molecular symmetry. Although the role of hemoglobin in maintaining the shape of the cell is rather secondary we believe that via direct binding to membrane it acts as an important probe of membrane/cytoskeleton interaction.

The 785 nm Raman excitation beam is slightly absorbed by hemoglobin, and resonant enhancement of Raman bands may affect measurements. To the best of our knowledge, absorption measurements of single RBC at different stretching states have not been performed yet. Electronic structure of hemoglobin is expected to change with cell deformation, at least via deoxygenation [129,144]. This leads to changes in RBC's absorption spectrum, in particular, the absorption grows at 785 nm [144]. Raman spectra excited at 785 nm show strong enhancement of the B_{1g} modes and the vinyl modes [136]. Therefore expected increase of absorption (at 785 nm) with stretching should further enhance all above mentioned modes almost proportionally. Our data show that intensities

of some modes (in particular at 1035 cm^{-1}) grow much stronger with stretching than the others so this hypothesis appears to break down. Hemoglobin concentration in the cell may also affect proportionally all measured Raman intensities. Elongating the RBC decreases the internal volume of cell and leads to the corresponding increase in hemoglobin concentration [129]. This effect should not only promote Raman intensities at all wavenumber but also neighbor-neighbor interaction between hemoglobins. Such enhanced interaction can be partially responsible for observed broadening of the peak at 1196. Nevertheless hemoglobin concentration effect alone can not fully explain observed behavior of Raman bands. We have to consider significant structural changes caused by mechanical deformation. Exact nature of structural changes in RBC are not straight forward to determine mainly because Phenylalanine (Phe), which is an essential amino acid that can be found not only in hemoglobin but also in various membrane proteins e.g. ankyrin, band3 proteins and spectrin [145]. Although hemoglobin is most likely the main source of Raman signal perturbation, we can not completely exclude contributions from proteins embedded in membrane and cytoskeleton which bears most of the forces during deformation. Direct exposure of membrane to Raman excitation beam is supposed to enhance total scattering probability from it. Interestingly, in Raman studies of RBC ghost [138], strong peak at about 1035 cm^{-1} was also observed which might suggest partial membrane contribution in our data. From many membrane proteins it is ankyrin which anchors cytoskeleton to membrane and that is why this protein together with spectrin presumably undergoes maximum deformation.

Taking into account all the above mentioned aspects, the behavior of Raman bands intensities as a function of applied deformation can be tentatively explained as follows. We suppose that at low deformations, when bands intensities remain almost constant, spectrin bears most of the forces and rearranges itself without significant changes in its primary chemical structure. It is likely that in this range of deformation, structural changes might occur in its higher order structure. At intermediate deformation range (10-20%), the stress is high enough and can lead to significant structural perturbations of linker proteins, spectrin network as well as hemoglobin attached to membrane.Therefore significant changes in Raman bands intensities were observed. At higher deformations (when bands intensity growth saturates), we need to consider mechanical nonlinearity of RBCs. It was proposed that nonlinear response of the cells can originate from the release of spectrin filament from linker proteins (ankyrin) which then re-bond in a configuration of lower stress [127,130]. We believe that observed saturation of the peaks corresponds to filament release from the linkers. This process is followed by creation of new bonds but in a configuration of similar or even lower stress. Behavior of all the bands discussed here are consistent in a way that they remain constant up to 10% cell deformation, increase (or decrease) in intermediate deformation range, then saturates for a small region and finally decrease (or increase) slightly at higher deformation (above$\sim$25%).

6.5 Conclusion

We have presented Raman spectra of RBC at relaxed and various stretched states and discussed the spectral changes induced in RBC by mechanical deformation. Statistical techniques, such as principal component analysis and 2D

correlation spectroscopy were applied to facilitate detailed analysis of spectral changes. Mechanical deformation of the cell mainly affects the bands corresponding to hemoglobin but the contribution from spectrin and membrane proteins can not be excluded. While Raman intensity changes with deformation were provided by both - 2D contour maps as well as PCA, other spectral details such as bands broadening and their shift were extracted from the 2D correlation analysis. We found that most of the affected bands show similar intensity behavior with stretching. The saturation of bands at higher deformations suggests some structural relaxation that RBC has to undergo to bear extra load. The results confirm widely accepted belief that spectrin release from membrane proteins allows for significant shape changes of red blood cells. We therefore tentatively suggest that interaction between membrane and cytoskeleton during deformation can be efficiently probed by confocal Raman spectroscopy, in particular via the peak around 1035 cm^{-1}.

7

The lipid phenotype of breast cancer cells characterized by Raman microspectroscopy: towards a stratification of malignancy

Although molecular classification based on gene and protein analysis brings interesting insights into breast cancer taxonomy, its implementation in daily clinical care is questionable because of its expense and the information supplied in a single sample allocation is not sufficiently reliable. New approaches, based on a panel of small molecules derived from the global or targeted analysis of metabolic profiles of cells, have found a correlation between activation of de novo lipogenesis and poorer prognosis and shorter disease-free survival for many tumors. We hypothesized that the lipid content of breast cancer cells might be a useful indirect measure of a variety of functions coupled to breast cancer progression. Raman microspectroscopy can help in the stratification of breast cencer cells being a rapid, reagent-free-tool bringing sensitive molecular information. We used RS to characterize metabolism of breast cancer cells with different degrees of malignancy: MDA-MB-435, MDA-MB-468, MDA-MB-231, SKBR3, MCF7 and MCF10A. We used Principal Component Analysis and Partial Least Squares Discriminant Analyses to assess the different profiling of the lipid composition of breast cancer cells. Characteristic bands related to lipid content were found at 3014, 2935, 2890 and 2845 cm^{-1}, and related to lipid and protein content at 2940 cm^{-1}. A classificatory model was generated which segregated metastatic cells and non-metastatic cells without basal-like phenotype with a sensitivity of 90% and a specificity of 82.1%. Moreover, expression of SREBP-1c and ABCA1 genes validated the assignation of the lipid phenotype of breast cancer cells. Indeed, changes in fatty acid unsaturation were related with the epithelial-to-mesenchymal transition phenotype, a trasition present in

the most aggresive tumors. Raman microspectroscopy is a promising technique for characterizing and classifying the malignant phenotype of breast cancer cells on the basis of their lipid profiling. The algorithm for the discrimination of metastatic ability is a first step towards stratifying breast cancer cells using this rapid and reagent-free tool.

7.1 Introduction

Despite the reduction in mortality in breast cancer patients due to earlier diagnosis and implementation of adjuvant chemo- and hormone therapies, breast cancer is still the commonest cause of cancer death in women worldwide [146]. Many factors and genes are involved in the initiation of breast cancer, but mortality is due to metastatic disease [147]. Patients who go on to develop life-threatening metastases in the visceral tissues have a much higher mortality rate and shortened life expectancy [148, 149]. Although the different biological behaviors and metastatic patterns observed among the distinct breast cancer phenotypes may suggest different mechanisms of invasion and metastasis, the biological features of breast tumors have proven insufficient for a comprehensive description of progression at first diagnosis, due to the heterogeneity of the disease [150]. The datasets available use specific genomic alterations to define subtypes of breast cancer [151]. However, the large number of genetic alterations present in tumor cells complicates the discrimination between genes that are critical for maintaining the disease state and those that are merely coincidental [152]. Thus, although molecular classification based on genetic analysis provides interesting insights into breast cancer taxonomy, its implementation in clinical care is questionable because it is too expensive to be introduced in daily pathological diagnosis, and because the information supplied is of insufficient reliability in single sample allocation [153].

Many observations during the early period of cancer biology research identified metabolic changes as common features of cancerous tissue, such as the Warburg effect [154, 155]. New approaches based on a panel of small molecules derived from the global or targeted analysis of metabolic profiles of cells are being developed to link cancer and altered metabolisms and to characterize cancer cell-specific metabolisms [156, 157]. One of the clearest signals is the de novo production of fatty acids in tumor cells associated with cancer progression, linked to an increased need for membranes during rapid cell proliferation as a part of a more general metabolic transformation, which provides cancer cells with autonomy in terms of their supply of building blocks for growth [158]. This metabolic change occurs as a result of common oncogenic insults and is mediated by the activation of multiple lipogenic enzymes affected at all levels of regulation, including transcription, translation, protein stabilization and protein phosphorylation [159–161]. Activation of de novo lipogenesis correlated with a poorer prognosis and shorter disease-free survival for many tumor types [162, 163]. A low ratio of TUFA/TFA has been proposed as a molecular marker for these aggressive tumors, which is called the lipogenic phenotype. The pathway that regulates synthesis of fatty acid in normal and tumor cells shares identical downstream elements including the SREBP-1c (transcriptional regulator sterol regulatory element-binding protein-1) and LXR (liver X receptor) [160, 164]. We hypothesized that the lipid content of breast cancer cells might be an indirect

measure of a variety of functions coupled to breast cancer progression, and that it could discriminate between different genetic features of breast cancer cells, providing new information on the aggressiveness of their phenotype.

To explore the lipid phenotype associated with breast cancer malignancy we used Raman microspectroscopy (RS). When applied to biological tissue, the technique can distinguish between pathologies based on the differences in their biochemical makeup [165]. RS is a rapid, reagent-free and non-destructive alternative for the analysis of cell biology systems [166]. Recent advances in Raman spectroscopy have given way to a wide range of biomedical applications including cancer. Its ability to detect variance related to DNA/RNA, proteins, and lipids have made it an excellent tool for quantifying changes on the cellular level, as well as differentiating between various cell fingerprints over all the Raman spectral range. The collection of spectra can be performed in vitro, ex vivo or in vivo without disrupting the cellular environment [167]. This is a major advantage of Raman spectroscopy, as most biological assays utilize chemical biomarkers and often require conditions nonnative to the biological environment. Usually, Raman spectra of biological samples are highly complex, and so mathematical processing of the spectroscopic data is required to obtain objective information. Multivariate techniques reduce the dimensionality of the spectral data and allow extraction of useful, objective and less complex information [168, 169]. We used Principal Component Analysis (PCA) [170] and Partial Least Squares Discriminant Analyses (PLS-DA) [171] to assess the different profiling of the lipid composition of breast cancer cells, which permitted differentiation of the lipogenic phenotype according to the proportion of unsaturated fatty acids. Moreover, PCA clearly distinguished cells with the epithelial-to-mesenchymal transition (EMT) phenotype, which is widely linked with breast cancer cell aggressiveness [172]. A discriminative model was generated that segregates metastatic cells and non-metastatic cells without basal-like phenotype with 90% sensitivity and 82.1% specificity.

7.2 Materials and methods

7.2.1 Cell culture and treatments

MDA-MB-435, MDA-MB-468, MDA-MB-321, SKBR3, MCF7 and MCF10A cells were obtained from the American Type Culture Collection. With the exception of MCF10A, all lines were maintained under standard conditions in 1:1 (v/v) mixture of DMEM and Ham F12 medium (DMEM/F12) supplemented with 10% fetal bovine serum (FBS), 1 mM pyruvate and 2 mM L-glutamine in 5% CO2-95% air at 37°C in a humidified incubator. MCF7 medium was supplemented with 0.01 mg/ml bovine insulin. MCF10A was grown in DMEM/F12 medium supplemented with 5% horse serum, 1 mM pyruvate, 2 mM L-glutamine, 0.01 mg/ml bovine insulin, 20 ng/ml EGF, 1 mg/ml hydrocortisone and 100 ng/ml Tetanus toxine, in the same incubator conditions described above. The treatment with the LXR agonist T0901317 (Cayman Chemical Company, Michigan), dissolved in DMSO, was performed at 2 μM final concentration (controls were treated with DMSO at the same concentration).

7.2.2 Immunocitochemistry and labeling of cells

For immunocytochemistry $8x10^4$ cells were seeded in 24 well-plates containing cover slips and were fixed after 24 h using cold methanol for 1 min. MCF10A in sparse conditions was obtained with $8x10^3$ cells/well. Cells were washed three times with PBS1x and treated with PBS1x-5% FBS for 30 min at room temperature. The antibodies used were: Vimentin, mouse anti-human (Dako, Atlanta); E-cadherin, mouse anti-human (BD Biosciences, NJ). Antibodies were diluted 1:50 in PBS1x-1% FBS and used for 1 h at room temperature. After three washes with PBS1x the secondary antibody, Alexa 555 anti mouse IgG (Life technologies, NY) was used diluted 1:1000 in PBS1x-1% FBS for 30 min at room temperature. After three washes with PBS1x the cover slips were mounted on slides using Vectashield (Vector laboratories, Burlingame) with DAPI for nucleus visualization. Preparations were analyzed with an Olympus BX60 fluorescence microscope (Olympus, Japan), using the optimal filters and 40x magnification.

For Nile Red and filipin staining $8x10^4$ cells were seeded in 24 well-plates containing coverslips and 24 h later cells were fixed with 4% cold paraformaldehyde (PFA) in PBS1x for 15 min. After fixing, cells were washed three times with PBS1x and stained with Nile Red at a final concentration of 1 μg/ml for 1 h, or filipin at a final concentration of 50 μg/ml for 2 h. Coverslips were then mounted as described above (filipin staining without DAPI) and analyzed with the confocal microscope (Leica TCS SP5, Wetzlar, Germany) for Nile red and the Olympus BX60 fluorescence microscope for filipin, with 40x magnification.

7.2.3 Raman spectroscopy

For analysis, $3x10^5$ cells were used, and for MCF10A cells in sparse conditions $3x10^4$ cells were used. For measurements in the 2820-3030 cm^{-1} range, cells were seeded in Petri dishes with #0 coverglass (Mattek, Ashland, MA). After 24 h, cells were treated as indicated for Nile red staining. The Raman system Renishaw (Apply Innovation, Gloucestershire, UK) comprises a 514 nm laser that supplies an excitation beam of about 10 mW power, which is focused onto the sample via a microscope with 60x objective (Edmund, York, UK). The same objective collects the scattered light from the sample and directs it to the spectrometer. The spectrometer processes this scattered light, by rejecting the unwanted portion and separating the remainder into its constituent wavelengths. The Raman spectrum is recorded on a deep depletion charge-coupled device (CCD) detector (Renishaw RenCam). We acquired the Raman spectrum of the cytoplasm of 20 cells per each cell line studied. The recorded Raman spectrum is digitized and displayed on a personal computer using Renishaw WiRE software which allows the experimental parameters to be set. The spectra were background subtracted with a custom-written Labview program and the Gaussian fits for total fatty acids (TFA) and total unsaturated fatty acids (TUFA) bands (2845 cm^{-1} and 3015 cm^{-1} respectively) were performed in Matlab allowing the quantification of the two types of fatty acids in the cytoplasm [173].

7.2.4 Statistical analysis

Raman spectroscopy is a promising technique in biomedical studies due to its non-invasive character and high specificity but Micro-Raman spectra of biomedical samples are inherently complex and weak. The use of multivariate analysis can improve their applicability and extract the useful information that Raman spectroscopy can provide to biomedicine.

In this study two multivariate techniques: Principal Component Analysis (PCA) and Partial least square-discriminant analysis (PLS-DA) were performed over the pre-processed Raman spectra in order to evaluate the spectral differences between the cancerous cell lines studied and to develop a model allowing their discrimination and classification.

PCA operates in an unsupervised manner (no previous knowledge of the samples under study is provided) and finds an alternative set of coordinates, the principal components, (PCs) to reduce the dimensionality and complexity of the data set. All the spectra can then be explained in a much simpler fashion through a small number of PCs that accounts for the maximum variance in the data. By plotting the first Principal Components scores, relations between samples (grouping) are revealed. In addition, plotting loadings as a function of the wavenumbers reveal the most important diagnostic variables or regions in the spectra related with the differences found in the data set.

PLS-DA is a supervised classification method in which knowledge of the sample (in our case, malignant or benign phenotype) is included. PLS-DA employs the fundamental principle of PCA but further rotates the component (latent variables, LVs) by maximizing the covariance between the spectral variation and group affinity so that the LVs explain the diagnostically relevant variations rather than the most prominent variations in the spectral dataset. In this study, the performance of the PLS-DA diagnostic algorithm was validated using the venetian blinds cross validation methodology with eight data splits. The number of retained LVs was determined based on the minimal root mean square error of cross validation (RMSECV) curves, and finally six were taken.

Multivariate statistical analysis was performed using the PLS toolbox (Eigenvector Research, Wenatchee, WA) in the Matlab (Mathworks Inc., Natick, MA) programming environment. SPSS (Statistical Package for the Social Sciences) for Windows was used for the statistics of TFA and TUFA quantification. In all the analyses, differences were considered significant when student's "t" was lower than 0.05.

Before including Raman spectra in the multivariate statistical techniques, correct preprocessing must be performed. In this case, background subtraction was achieved with a Matlab and Labview algorithm [65], and then normalization under all Raman spectra was performed to correct for the different amplification in the signal. This normalization can be based on the fact that the spectral region used (the CH stretching region) can be considered as the total biomass present in our confocal volume [174].

7.2.5 Real-time reverse transcription- PCR

Real-time reverse transcription-PCR (qRT-PCR) was performed with gene-specific fluorescent SYBR Green probes (Applied Biosystems, NY, USA) using a 7300 Real time PCR system detection Instrument and the associated software

(Applied Biosystems), following the manufacturer's instructions. Primers were designed using Primer Express software (primer sequences are available on request). We calculated relative changes by the comparative CT method using cyclophilin A as the reference gene. Each reaction was performed in triplicate.

7.3 Results and discussion

7.3.1 The expression of lipid metabolic genes is correlated to the metastatic ability of cells.

The transcription factors SREBP-1c (transcriptional regulator sterol regulatory element-binding protein-1) and LXR (liver X receptor) maintain cholesterol homeostasis through complementary pathways of feedback inhibition and feed-forward activation [160, 175, 176]. To assess their coordinated action in the lipid phenotype of breast cancer cells, we explored the LXR pathways in a set of breast cancer cells according to their malignant phenotype including both non-metastatic and metastatic cells: MCF7, which expressed hormone receptors like luminal A tumors; SKBR3, a phenotype with amplifications of the ErbB2 oncogene; MDA-MB-468, p53 mutated cells with basal-like phenotype; and two different metastatic models: MDA-MB-435, with lung metastasis tropism, and MDA-MB-231 with bone metastasis tropism, both belonging to the basal-like phenotype (also called post-EMT cells) [177]. We analyzed the expression of SREBP-1c, gene target of LXR, and ABCA1, other direct LXR target gene involved in cell cholesterol export [164]. Twenty-four hours after treatment with 2μM LXR agonist T0901317 (Fig. 7.1A), the up regulation of SREBP-1c was evident in the metastatic cells MDA-MB-231 and MDA-MB-435 compared with non-metastatic cells: the transcriptional induction of SREBP-1c was 20 times higher in MDA-MB-231 and 17.4 times higher in MDA-MB-435. In contrast, treatment with the agonist produced only moderate increases in the expression of SREBP-1c in SKBR3 (2.9 times) and MCF7 (3.6 times), and a decrease in MDA-MB-468 cells. Moreover, the cholesterol related gene ABCA1 was increased in MDA-MB-231 (6.8 fold) and MDA-MB-435 (8 fold) and differently induced in non-metastatic cells (SKBR3 cells, 27 fold, MCF-7, 2.4 fold, and MDA-MB-468, 1.3 fold). These results showed the differences in regulation of lipid metabolism pathways in breast cancer cells.

Like nutritional control, neoplastic lipogenesis is controlled through the modulation of the expression and/or maturation status of the transcription factor SREBP-1c, a crucial intermediate of the pro- and anti-lipogenic actions of nutrients and hormones, which stimulates fatty acid synthase transcription in normal and malignant cells [176,178]. In tumor cells, SREBP-1c expression and/or maturation is constitutively driven by the aberrant hyperactivation of these pathways in response to a variety of oncogenic changes, including overproduction of growth factors (GFs), ligand-dependent or independent hyperactivation of GF receptors (GFRs), and loss of function of components of the signalling cascade such as the phosphatase and tensin homologue (PTEN), a potent tumor suppressor [179,180].

SKBR3 cells, a classical ErbB2 amplified model, responded against the LXR agonist with increased ABCA1 expression, different to that of MDA-MB-468 cells which have two populations with different degrees of EGFR ex-

pression [181], displaying low response against the agonist. It is well known that endogenous synthesized fatty acids increase the signal-to-noise ratio in the HER1/HER2-driven progression of human breast epithelial cells towards malignancy [158]. Malignant cells have devised a mechanism to subvert the normal pathways for feedback inhibition via the EGFRvIII and PI3K-dependent activation of SREBP-1c [160]. SKBR3 overexpressed SREBP-1c at basal levels by a factor of five with regard to MDA-MB-468 and metastatic cells; therefore, LXR might respond to excess cellular cholesterol by promoting ABCA1-dependent cholesterol efflux [156]. On the other hand, in normal cells, PI3K activation is tightly controlled by dephosphorylation of PIP3 by the phosphatase PTEN. Activity of the pathway is deregulated in cancer through a variety of mechanisms, including activating mutations in PI3K or PTEN loss [182, 183]. Indeed, the role of cholesterol metabolism in cancer pathogenesis and its association with EGFR/PI3K signaling has recently been described as a potential therapeutic target [160].

The inverse correlation between estrogen receptors in breast tumors and genes involved in lipid storage is well known [184]. Indeed, MCF7 cells had the lowest induction of SREBP-1c and ABCA1. In addition, the increased expression of SREBP-1c and ABCA1 in the estrogen negative metastatic cell lines indicated that both genes are functionally implicated in the most malignant phenotype. Therefore, the pathogenesis of metastasis may include the conjunction of both constitutive metabolic features: fatty acid synthesis and cholesterol cell content.

7.3.2 The lipid phenotype characterized by Raman microspectroscopy

To explore the lipid phenotype associated to breast cancer malignancy we optimized the Raman microspectroscopy (RS) system to acquire Raman spectra in the range of 2820-3030 cm^{-1}, where TFA (2845 cm^{-1}) and TUFA (3015 cm^{-1}) bands were located. In the analysis we included MCF10A cells as benign breast tumor cells, unable to spread outside the basal membrane, despite their basal-like phenotype (Fig. 7.1B).

The cytoplasm lipids were measured by RS in a position near the nucleus and outside the endoplasmic reticulum area, where the Nile red staining showed major lipid concentration (see asterisk in Figure 7.1C). Each spectrum line in Figure 7.1D represents the Raman intensity versus the Raman shift measured in a single cell, and illustrates the cell variability inside each cell line. The bands corresponding to TFA and TUFA for individual cells were used to quantify the TFA and the TUFA average content in each cell line (Figure 7.2A and 7.2B). To obtain the relative quantities of unsaturated fatty acids (% UFA) in each cell line, which indicate the lipogenic phenotype, the values of individual cells followed by the average of cell lines were calculated (Figure 7.2C). Low but significant changes in the TFA bands intensities were found when the cell lines were compared with the MCF10A cells (Student's "t" < 0.0009). The TFA content was clearly highest in MDA-MB-435 cells, followed by the MCF10A cells, and lowest in the SKBR3 cells. These results are in agreement with the increasing evidence that lipid accumulation is a hallmark of aggressive cancer cells, and is involved in the production of membranes for rapid cell proliferation [185].

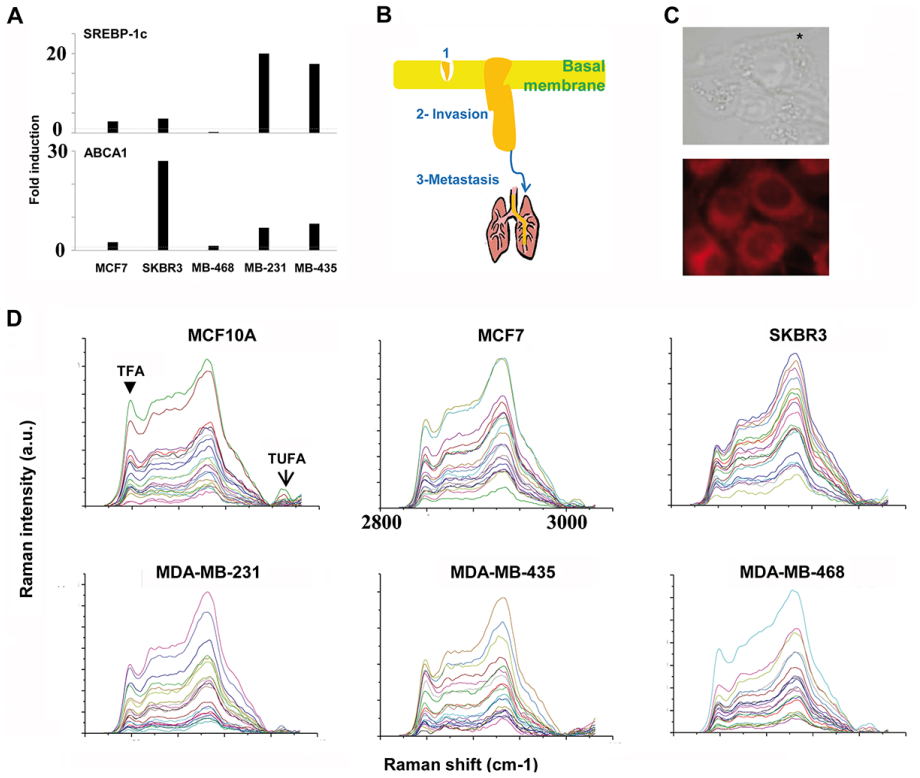


Figure 7.1: Variability in lipid metabolic genes expression and FA composition in MDA-MB-231, MDA-MB-435, MDA-MB-468, MCF7 and SKBR3 cell lines analysed by Raman microspectroscopy. A) The gene expression of SREBP-1c and ABCA1 were examined after 24 h treatment with the LXR agonist T0901317 2 μM compared to the basal conditions by RT and real-time PCR. The fold induction is represented over the pointed line. Cyclophilin A gene was used for normalization. B) Simplistic representation of the progression status of breast cancer cells used in the study: 1) MCF10A cells; 2) MCF7, SKBR3 and MDA-MB-468 and 3) MDA-MB-231 and MDA-MB-345. C) Above, brightfield image of MDA-MB-435 cells, with an asterisk indicating the position of the measurements in the cytoplasm. 60x magnification and 9 mW power were used. Down, fluorescence microscopy image of MDA-MB-435 cells stained with Nile red. 40x magnification was used. D) Measured raw Raman spectra of the cell lines where the axes are intensity (in arbitrary units) versus Raman shift (cm^{-1}). MCF10A cells were also measured. The TFA (2845 cm^{-1}) and TUFA (3015 cm^{-1}) bands are indicated with the arrows in the first spectra.

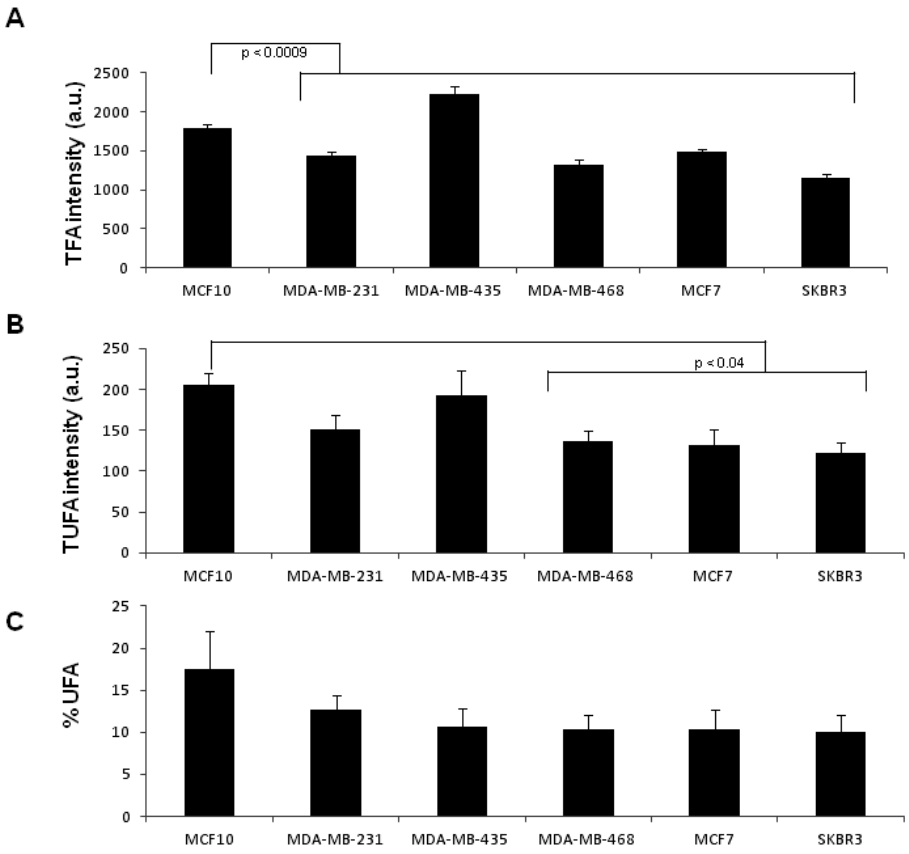


Figure 7.2: Analysis of the lipid content in the breast cancer cell lines using Raman microspectroscopy. A) Total fatty acid (TFA) and B) total unsaturated fatty acid (TUFA) Raman band intensity average in the cell lines is represented in arbitrary units. C) Relative unsaturated fatty acid content is represented as %. The average was calculated with the individual cell ratio values. The lines and the p values (student's "t") indicate the significance between bars compared to the MCF10A values.

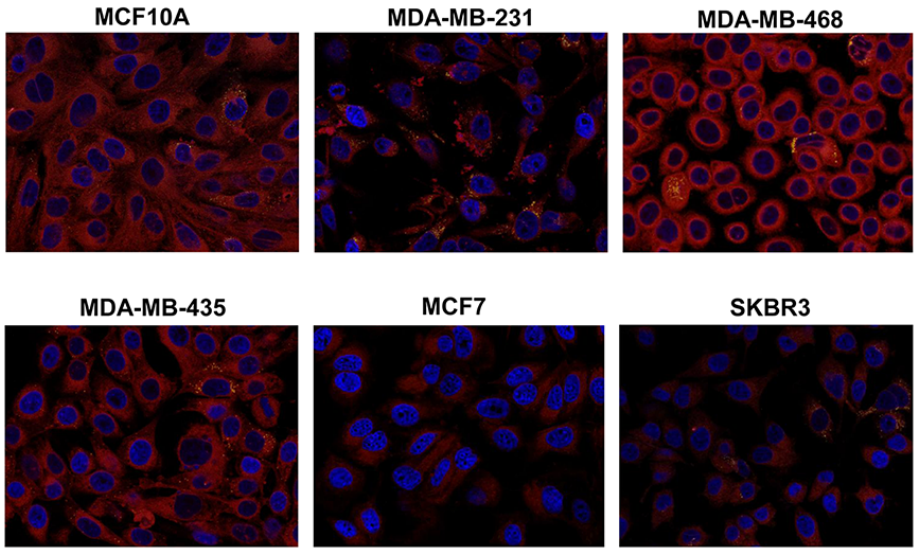


Figure 7.3: Analysis of the lipid content in the breast cancer cell lines using Nile red staining and confocal microscopy. Cells were fixed in 4% PFA and treated with Nile red (1 μg/ml) for 1 h at room temperature and analysed as indicated in material and methods. Hydrophilic fatty acids, mainly phospholipids, are seen in the red channel. Hydrophobic fatty acids, mainly cholesterol esters and triglycerides, are seen in the merge image in yellow. DAPI staining labels the nuclei. 40 x magnification was used.

TUFA bands intensities were only significantly lower in MDA-MB-468, SKBR3 and MCF7 cells when compared to MCF10A cells (Student's "t" < 0.04). Moreover, no significant differences were found between MCF10A and the metastatic cells MDA-MB-231 (Student's "t" $= 0.076$) and MDA-MB-435 (Student's "t" $= 0.661$) with regard to TUFA. The unsaturation ratio, and not the TUFA value, is indicative of de novo lipogenesis and cell malignancy. For this reason, the percentage of unsaturated fatty acids in each cell line was calculated, but we did not observe significant differences due to the high dispersion in the individual cell values. The Student's "t" scores for the cells compared to MCF10A were: MDA-MB-435=0.179; MCF7=0.166; MDA-MB-231=0.323; MDA-MB-468=0.148 and SKBR3=0.139. Thus, the quantification of these two bands was not sensitive enough to differentiate benign from malignant breast cancer cells.

The total amount of lipids was analyzed with an alternative technique using Nile red staining (Figure 7.3). The red channel showed mainly the membrane phospholipids (hydrophilic lipids) and the green channel mainly the hydrophobic lipids (in yellow in the merge image), which accumulated in the typical cytoplasm storage vesicles derived from the endoplasmic reticulum compartment (called lipid droplets). They contain mainly esterified cholesterol and triglycerides [186]. The confocal images of lipids showed similar results to the RS quantifications with the exception of MDA-MB-468 cells, which showed the highest Nile red intensity. As expected, this technique was less informative than Raman for differentiating the cells. The green channel intensities labeling the lipid droplets did not correspond to the TUFA quantification obtained by Raman. MCF7 cells did not show lipid droplets, but their RS quantification was similar to that of SKBR3 and MDA-MB-468 cells. The RS results might also lead us to expect more droplets in MCF10A cells. The overestimation of unsaturated lipid content in MCF7 and MCF10A cells using RS may be due to differences in the lipid composition of the droplets [185].

The idea that exacerbated lipogenesis provides immortalized epithelial cells with a profound neoplastic growth and/or survival advantage over those that maintain physiological levels of endogenous fatty acid biosynthesis strongly suggests that some lipogenic enzymes may work as metabolic intermediates of oncogenesis by linking cellular anabolism and malignant transformation [187, 188]. Indeed, the level of fatty acid saturation indicative of de novo lipogenesis decreased when LNCaP prostate cancer cells were treated with soraphen A (a lipogenesis inhibitor) [162].

To improve the information obtained with the lipid phenotype measurements, we performed a PCA analysis using the 2820-3030 cm^{-1} spectral data to study the grouping and the homogeneity of the sample distribution (Figure 7.4). PC1 and PC2 scores accounted for 47% and 39% respectively of the total variance in the dataset. Raman band regions responsible for the PC1 score discrimination were 3014, 2890 and 2848 cm^{-1} (related to lipid content) and 2940 cm^{-1} (related to lipid and protein content). Raman band regions responsible for the PC2 score discrimination were 2846 cm^{-1} (TFA) and 2935 cm^{-1}, associated with the chain end -CH3 [189].

MCF7 and SKBR3 cells were grouped in the low PC2 region and separated for high and low levels of the PC1 axis respectively. In contrast, other cell lines like MCF10A and MDA-MB-435 appear to be more heterogeneous, spreading through a larger area in the PC axes (Figure 7.4, left panel). Although Raman spectral region 2900 to 3100 cm^{-1} has been labeled the CH stretching

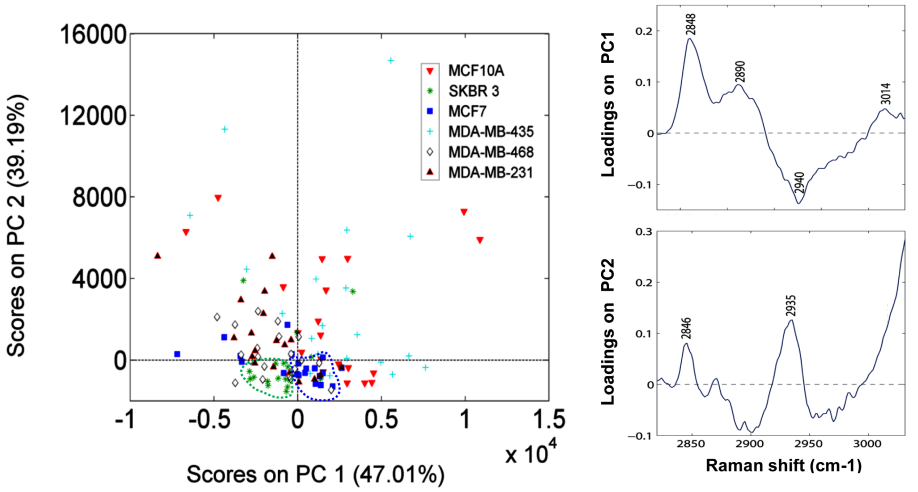


Figure 7.4: PCA scores showing the cell variability present inside each cell line and between the different cell lines. Left: illustration of PCA scores from MCF10A, MDA-MB-231, MDA-MB-435, MDA-MB-468, MCF7 and SKBR3 cell lines RS acquisition. SKBR3 cells are shown in the green circle and MCF7 cells in the blue circle. MDA-MB-435 and MCF10A cells are the most dispersed in the plot. On the right, the loading plots for each Principal Component, both related to fatty acid and protein content (TUFA: 3014 cm^{-1}; protein and lipid: 2940 cm^{-1}; TFA: 2871, 2890, 2846 and 2848 cm^{-1}; -CH3: 2935 cm^{-1}). Percentages in the score plots represent the variance accounted for each PC.

region [189] and therefore, contains bands common for many biomolecules, we attempted to extract a hypothesis from the PCA score plot. As lipids were included in both PC1 and PC2 loadings, we interpret that SKBR3 and MCF7 had the lowest content. MDA-MB-231 and MDA-MB-468 cells had intermediate lipid content and MDA-MB-435 and MCF10A cells differed widely, though it was always high. Our interpretation of the PC1 loading was that it represents mainly TFA, and that lipid and protein cell content were inversely correlated, because we had lipid bands in positive and the 2940 cm^{-1} band (which includes both lipids and proteins) in negative. MCF7 and SKBR3 cells, with similar PC2 scores, had different PC1 values, suggesting different protein content, higher in SKBR3.

The most prominent band included in the PC2 loading was 2935 cm^{-1}. No one substrate is clearly associated with the 2935 cm^{-1} band due to the fact that many biomolecules contain $-CH_3$ side terminal groups. The 2846 cm^{-1} band, also included in PC2, corresponds to total fatty acids and we hypothesized a contribution of cholesterol and cholesterol esters in the 2935 cm-1 band, because it is a lipid with many $-CH_3$ side terminal groups. We also observed differences in the ABCA1 gene expression, and differences in cholesterol have been associated to proliferation and migration of breast cancer cells [184, 185], suggesting the involvement of cholesterol in the malignant phenotype of the cells studied.

We analyzed the cholesterol content of the cell lines using filipin staining, which labels free (unesterified) cholesterol present in the cytosol and membranes

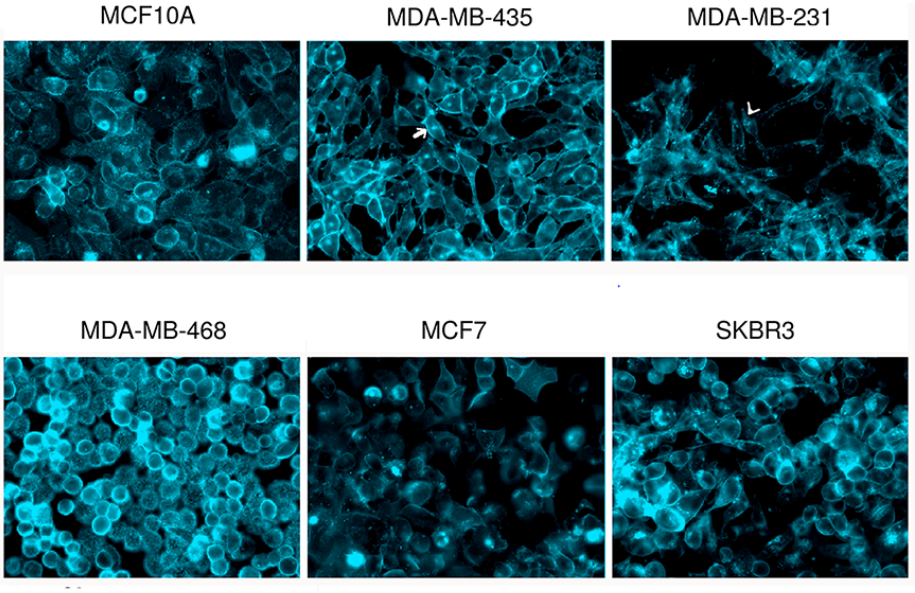


Figure 7.5: Analysis of the cholesterol content in MCF7, SKBR3, MDA-MB-231, MDA-MB-468, MDA-MB-435 and MCF10A cells with filipin staining and fluorescence microscopy. Cells were fixed in 4% PFA and treated with filipin (50 μg/ml) for 2 h at room temperature and analyzed as indicated in material and methods. Filipin labels free cholesterol present in the membranes (arrow) and in the cytosol (arrow head). 40x magnification was used.

of cells. We found differences in the content and distribution of free cholesterol between the cell lines. MDA-MB-231, MDA-MB-468 and MDA-MB-435 cells presented the highest cholesterol intensity, though with different distributions: in MDA-MB-231 and MDA-MB-468 it was accumulated in large cytoplasm spots, whereas MDA-MB-435 cells mainly presented cholesterol in the plasma membrane together with some smaller spots in the cytosol (Figure 7.5). The levels of cholesterol were also high in MCF10A but low in MCF7 cells. SKBR3 cells had higher cholesterol content than expected, given their localization in the PCA, but reinforcing our hypothesis that their high ABCA1 gene expression occurs in response to an excess in cholesterol content (see Figure 7.1A). Taken together, these results suggested that cholesterol might be involved in the lipid differences between metastatic and non-metastatic cells. Moreover, its contribution in the PC2 loading may also include cholesterol esters, not observed with filipin.

7.3.3 The lipid profiling of breast cancer cells distinguishes metastatic ability from malignancy

In the second step, a PLS-DA was used to construct a classification model. This is a supervised method, meaning that prior knowledge of the class membership was included. First we built a classification algorithm to discriminate between non-metastatic non-basal-like (MCF7 and SKBR3) and metastatic basal-like

(MDA-MB-231 and MDA-MB-435) cell lines. Then, in the ideal prediction model, the first group will have class 0 and the second one class 1 (Figure 7.6). The PLS-DA model was carried out over the pre-processed Raman spectra and a cross-validation was performed in order to check the strength of the algorithm to predict new samples. The method for cross-validation was venetian blinds w/10 splits and the errors for the prediction and cross validated model were RMSEC: 0.3 and RMSECV: 0.45 respectively, showing good stability for predicting new samples. A good discrimination between metastatic and non-metastatic cell lines was achieved with sensitivities and specificities of 92.5% and 97.4% for the calibration and 90% and 82.1% for the cross-validation respectively. These results showed good accuracy in discriminating metastatic ability of breast cancer cells, better than those reported for the Raman spectral window (2,800-3,100 cm^{-1}) comparing benign disease and breast cancer tissue in vivo samples, which had specificity and sensitivity of 81.2 and 72.4, respectively [190].

We used the PLS-DA model to test the membership of MCF10A and MDA-MB-468 cells, which were not included in the groups (Figure 7.6). The result indicated that most of the MCF10A cells were very similar to the metastatic group. Seventy-five per cent of the MCF10A cells analyzed were predicted to belong to the metastatic class (above the threshold). The rest of the MCF10A cells were localised below the threshold. Following the same criteria eight of the twenty MDA-MB-468 cells analyzed (40%) were localized in the metastatic group and the rest in the non-metastatic group. The PLS-DA model showed that MCF10A and MDA-MB-468 cells had different basal-like phenotypes. These results suggested that the classification algorithm might discriminate two different basal-like phenotypes: MDA-MB-468 cells, common to a subgroup of MCF10A cells, and the MCF10A cells, common to MDA-MB-231 and MDA-MB-435 metastatic cells.

It has been described that MCF10A cells with basal-like phenotype, which present many features of mesenchymal cancer cell lines in sparse cultures, have intrinsic plasticity for undergoing EMT, transition present in the most aggressive breast tumors with a basal phenotype [172]. Since MCF10A cells were grown at low confluence, we hypothesized that the similarities between MCF10A and MDA-MB-435 cells in the PCA and between MCF10A and the metastatic cells group in the PLS-DA might be related to culture conditions. We performed Raman analysis in sparse and in dense MCF10A cultures, and in cells growing at the edge of dense cultures (Figure 7.7A). In the PCA (Figure 7.7B), PC1 had a prominent TUFA band contribution, and PC2 loading was formed by three lipid bands: 3014, 2890 and 2845 cm^{-1}, and the 2935 cm^{-1} band with inverse correlation. The PCA scores plot clearly separated MCF10A and MDA-MB-435 cell lines: the phenotype of the MDA-MB-435 cells was characterized by higher lipid content (y axis) and lower TUFA (x axis). TUFA band intensity (PC1, x axis) also distinguished between the MCF10A subtypes, being lower in the sparse area and higher in the confluent area.

We also analyzed the EMT phenotype of MCF10A cells grown in sparse and confluent conditions. As expected, like MDA-MB-435, cells lost E-cadherin and expressed more vimentin in sparse conditions (Figure 7.7C) than in confluence. These results confirmed that the spectroscopical differences were secondary to phenotypic changes and correlated well with malignancy; clearly the degree of similarity between MDA-MB-435 and MCF10A cells depends strongly on the culture conditions of MCF10A.

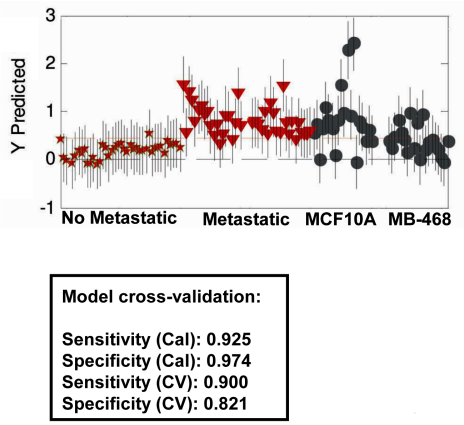


Figure 7.6: PLS-DA discriminative model using Raman microspectroscopy spectra of non-metastatic (SKBR3 and MCF7) and metastatic (MDA-MB-231 and MDA-MB-435) cell lines. PLSDA classification algorithm, in which non-metastatic cells are predicted with class 0 and metastatic cells with class 1. A threshold is assigned (red line) corresponding to the best specificity and sensitivity parameters that separate groups of cells. RMSECV is represented by the error bars. A sensitivity of 90% and a specificity of 82.1% were achieved. Once the model was built, MCF10A and MDA-MB-468 were included to predict their membership. Seventy-five per cent of MCF10A and 40% of MDA-MB-468 cells are related to the metastatic group.

The expression in the set of breast cancer cells of E-cadherin, CK18 and vimentin at the mRNA level, and E-cadherin and vimentin at protein level as well, confirmed the close relationship between the lipid phenotype and the EMT process (Figure 7.8). MCF7 did not express vimentin protein and SKBR3 did so in less than 5% of the cells (Figure 7.8B), similar to MDA-MB-468 cells. These results suggested that in addition to metastatic ability the PLS-DA model discriminated cells with basal-like phenotype that undergo EMT (MCF10A) from basal-like cells with no EMT (MDA-MB-468).

The combination of multivariate statistical techniques applied to the Raman spectral data (PCA and PSL-DA analysis) provided a powerful quantitative method to discriminate cancer phenotypes. These mathematical methods used the whole range of the spectra for the differentiation of the cells. Intensity analysis of a single band does not provide so powerful information, as it was shown in fig. 7.2. Our results suggest that the lipid phenotype of these cells is a signal of the proclivity to mesenchymal transition related to the high aggressiveness and metastatic spread [191]. EMT is an essential developmental process by which cells of epithelial origin lose epithelial characteristics and polarity, and acquire a mesenchymal phenotype with increased migratory behavior. Thus, the characterization of this functional phenotype of cancer cells with RS provides information on intercellular cell adhesion, down-regulation of epithelial markers, up-regulation of mesenchymal markers, acquisition of fibroblast-like (spindle) morphology with cytoskeleton reorganization, increase in motility, invasiveness, and metastatic capabilities [191–193]. The PSL-DA model described discriminates luminal or HER-2 overexpressing cells without EMT and post-EMT cells

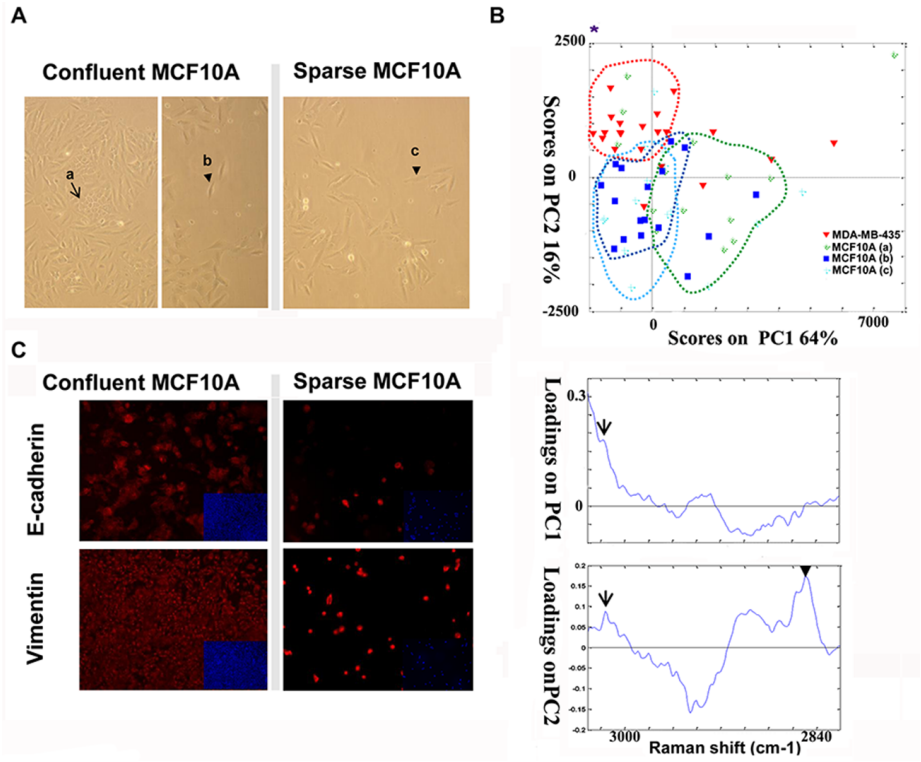


Figure 7.7: Raman microspectroscopy and PCA differentiate the MCF10A cells grown in confluent and sparse conditions. A) MCF10A microscopy images of the cells measured by RS in confluent and sparse conditions. Brightlight images were obtained with an inverted microscope and 10x magnification. Arrows indicate the different areas that were measured by Raman (a: confluent; b: separate cells grown at the edge of confluent cultures; c: sparse). B) PCA representation of MDA-MB-435 cells and MCF10A cells (grown in high confluence and in sparse conditions). PC 1 and 2 separate different groups of cell lines. MDA-MB-435 cells have higher PC2 scores, separated from the MCF10A. MCF10A grown in high confluence are displaced from the ones grown in sparse conditions showing higher PC1 scores. Asterisk indicates the localization of the "lipogenic phenotype" in the axis. PC1 and PC2 loadings are described down with the bands related to TFA (arrow head) and TUFA (arrow) indicated. The percentage means the variance accounted for each PC. C) Immunofluorescence images of E-cadherin and vimentin proteins in both MCF10A culture conditions. DAPI staining appears minimized inside each picture.

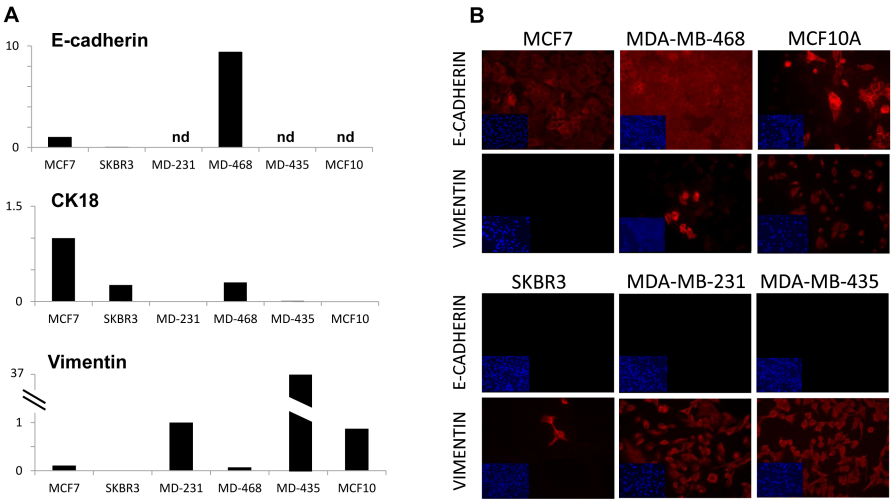


Figure 7.8: Epithelial and EMT marker gene expression in MCF7, SKBR3, MDA-MB-231, MDA-MB-468, MDA-MB-435 and MCF10A cells. A) The gene expression of epithelial cell markers (E-cadherin, cytokeratin 18) and the mesenchymal cell marker vimentin were examined by RT and real-time PCR using 200 ng of RNA. E-cadherin and CK18 are represented compared to the luminal MCF7 cell line expression and vimentin is represented compared to the MDA-MB-231 cell line. Cyclophilin A gene was used to normalize gene expression. B) Immunofluorescence staining of the epithelial E-cadherin and the mesenchymal vimentin markers in MCF7, MDA-MB-468, MCF10A, SKBR3, MDA-MB-231 and MDA-MB-435 cells. DAPI staining appears minimized in each picture. 40x magnification was used.

with a sensitivity of 90% and a specificity of 82.1%. Aggressive cells with basal phenotype (related to EMT plasticity) can also be differentiated, although it may be necessary to include other spectral regions to increase the sensitivity in the differentiation of metastatic and non-metastatic basal-like cell phenotypes. Recently, it has been reported that the analysis of human tumor gene expression profiles identifies triple negative breast cancer subtypes with an overall false-positive rates of 1.7%, 1.7%, and 0.9% for ER, PR, and HER2, respectively [194].

Breast cancer is a heterogeneous disease that includes a wide range of histological subtypes and a diversity of clinical behaviors and patient outcomes [195]. We used representative cell variants, including different phenotypes of breast cancer cells: estrogen receptor expression, ErbB2 amplification, p53 mutation and aggressive metastasic. The molecular and cellular characterization of their associated "lipid signatures" by RS, combined with multivariate statistical analysis, is a promising technique for characterizing the malignant phenotype of breast cancer cells and might provide a helpful adjunct to gene-expression profiling or proteomics in the classification, diagnosis and prognosis of human cancers. Using different spectral ranges of RS, similar results have been obtained regarding the lower lipid content in SKBR3 compared to MDA-MB-231 and -435 cells [196]. These findings support the use of this technology in the study of the lipid phenotype of cells, with possibilities to be used in experimental tumors [190, 197] and in human samples to distinguish between ductal carcinoma in situ and invasive ductal carcinoma of the breast [198]. Serum samples have been used to discriminate between breast cancer patients and healthy individuals; the bands analyzed were statistically accepted as markers corresponding to proteins, polysaccharides and phospholipids [199]. Moreover, the identification of new spectral signatures expanding the RS window may offer more accurate classification of cells for diagnostic purposes, providing rapid, reagent-free and non-destructive alternatives for the analysis of tumor samples.

Raman spectroscopy has shown promise for use as a clinical tool for diagnosis of breast cancer. Optimization of spectral acquisition times and spatial resolution for clinical use is an area which needs further investigation. Studies of larger patient population samples will be needed to establish comparisons between spectral makers for breast cancer cells and pathological indicators that are used for current diagnosis. Moreover, improvements on current data analysis techniques, including the application of advanced data mining methods, along with novel preprocessing techniques will also be critical to introduce RS in the clinical practice.

7.4 Conclusion

Raman spectroscopy is a promising technique in biomedical studies due to its non-invasive character and high specificity. The lipid phenotype associated to breast cancer malignancy belongs to Raman spectra adquired in the range of 2820-3030 7.8, where TFA (2845 7.8) and TUFA (3015 7.8) bands were located. The combination of multivariate statistical techniques, which use the whole range of the spectra, applied to the Raman spectral data (PCA and PSL-DA analysis) provided a powerful quantitative method to discriminate cancer phenotypes. In addition, an algorithm to differentiate metastatic from non metastatic

and non basal phenotype breast cancer cells was design using PLS-DA, with 90% sensitivity and 82.1% specificity. Our results suggest that the lipid phenotype of these cells is a signal of the proclivity to mesenchymal transition related to the high aggressiveness and metastatic spread, then the identification of new spectral signatures expanding the Raman spectral window may offer more accurate classification of cells for diagnostic purposes.

8

Raman spectroscopy supported by Multivariate Curve Resolution improves biochemical analyses of cancer cells

In breast cancer the presence of cells undergoing the epithelial-to-mesenchymal transition (EMT) is indicative of metastasis progression. In chapter 7, PCA has been used to study the lipid metabolism in different breast cancer cell lines depending on the degree of malignancy in the CH stretching region of Raman spectra. However, PCA does not provides meaningful components that could be assigned directly to molecular Raman spectra. Consequently, in this chapter Multivariate Curve Resolution (MCR) was applied to Raman spectroscopy to assess the metabolic composition of breast cancer cells undergoing EMT deconvolving the individual spectra of molecular components. The fingerprint region of the Raman spectra was used which permits a better metabolite discrimination. Benign breast cancer cell line MCF10A, which has a basal-like phenotype and shows phenotypic and genetic changes associated with EMT when cells are cultured in sparse conditions, was used. MCR led to the conclusion that the EMT process affects the lipid profile of cells, increasing tryptophan but maintaining a low phosphatidylserine content in comparison with highly metastatic cells. A Partial Least Squares-Discriminant Analysis (PLS-DA) model classified MCF10A post-EMT cells with 94% sensitivity and 100% specificity. In conclusion, Raman microspectroscopy coupled with MCR enables deconvolution and tracking of the molecular content of cancer cells during a biochemical process, being a powerful non-invasive tool for identifying metabolic features of breast cancer cell aggressiveness.

8.1 Introduction

Breast cancer metastasis is a complex multistep process that consists of a series of dynamic interactions between tumour cells and host cells that allow tumour cells to leave the primary site and establish a distant lesion [200]. Tumour cells at the primary site are influenced by cross-talk with the stroma, leading to breakdown of the basement membrane. This allows migration and invasion [201], possibly through molecular mechanisms akin to the epithelial-to-mesenchymal transition (EMT), which ultimately drive epithelial cells to acquire mesenchymal characteristics: loss of cell-cell adhesion and apical-basal polarity, downregulation of epithelial cytokeratins, upregulation of vimentin, cytoskeleton reorganization, and increased motility and invasiveness [172, 202, 203]. The preferential expression of EMT-related genes has been found in basal-like breast tumours, the most invasive breast carcinomas [204] with the worst prognosis and greatest resistance to chemotherapy [205]. Basal-like tumour cells express markers characteristic of the normal mammary gland myoepithelium, such as epidermal growth factor receptor (EGFR), p63 and basal cytokeratins CK14, CK5/6 and CK178. Cellular remodeling occurring as a consequence of EMT, whereby cells have altered responses to agents in the circulatory system or secondary tumour site, could be advantageous for the process of metastasis [172, 203].

EMT can be induced during in vitro cell culture under the influence of extracellular matrix components and growth factors, such as transforming growth factor beta (TGFbeta), scatter factor/hepatocyte growth factor, fibroblast growth factors, epithelial growth factor family members and insulin-like growth factors 1 and 29. Signal transduction pathways such as Wnt, Hedgehog, Notch and integrin signaling can also coordinate EMT programmes. A number of transcription factors induce EMT through transcriptional control of E-cadherin, including SNAI1 (zinc finger protein snail 1), SNAI2, ZEB1 (zinc finger E-box-binding homeobox 1), ZEB2, TWIST, FOXC1 (forkhead box protein 1), FOXC2, TCF3 (transcription factor 3 - also known as E47) and GSC (homeobox protein goosecoid) [206]. EMT confers mesenchymal properties on epithelial cells and has been closely associated with the acquisition of aggressive traits by carcinoma cells [207]. Moreover, the dynamic interactions among epithelial, self-renewal and mesenchymal gene programmes determine the plasticity of epithelial tumour-initiating cells [208]. The constitutive activation of signaling cascades that stimulate cell growth has a profound impact on anabolic metabolism [208]. One of the principal mechanisms of aerobic glycolysis resides in the activation of hypoxia-inducible factor (HIF), a transcription factor activated by hypoxic, oncogenic, metabolic and oxidative stress, and also involved in EMT, the initial signal of the lethal metastatic phenotype of breast cancer cells [209]. Since metabolic features of tumour cells are critical in cancer progression and drug resistance [210], there exists a need to develop alternative/adjunct, user-freed, cost effective, rapid, objective and unambiguous methods for the early detection and diagnosis of metastatic breast cancer cells.

We have previously shown [211] that RS and PCA is able to characterize the metabolic phenotype of breast cancer cells, based on a panel of small molecules derived from the global or targeted analysis of metabolic profiles of cells. This emerged as a useful method for stratifying malignancy, being well correlated with the EMT phenotype and the expression of SREBP-1c and ABCA1 genes [211]. A useful model for characterizing the metabolic features associ-

ated with EMT is based on in vitro studies on benign breast cancer cell line MCF10A, which has a basal-like phenotype, showing phenotypic and genetic changes associated with the mesenchymal transition when cells are cultured in sparse conditions [172], thus increasing the fatty acid saturation ratio [211]. This might be due to an increase in lipid biosynthesis, leading to the "lipogenic phenotype", which is common in aggressive cancer cells [205, 212, 213].

In the present research we used a Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) algorithm to study cellular RS, which allowed the deconvolution of meaningful molecular RS of biomolecules that are related to metabolites that change concentration in the cells under study. Using RS and a first PCA on the dataset of spectra, the number of components necessary to explain the variance in the data was selected, accounting for 95% of the total variance. The output of the MCR-ALS analysis gives more chemically and physically understandable results than classical techniques such as PCA, where only a mathematical exploration of the data can be performed [23, 24, 28, 214]. This approach led to the discovery of tryptophan as a metabolite that fits with the hypothesis that the EMT phenotype initiates metastatic cells, with the phospholipid level controlled as indicated by the phosphatidylserine content. MCR algorithm has been applied in a broad range of chemical analysis [47,48,50] and recently to Raman images [35,189]. To our knowledge this study reports the first application of MCR to RS to deconvolve and track the molecular content of cancer cells during a biochemical process, in our case the EMT process.

8.2 Material and Methods

8.2.1 Cell culture

MCF10A cells were obtained from the American Type Culture Collection and were grown in DMEM/F12 medium supplemented with 5% horse serum, 1 mM pyruvate, 2 mM L-glutamine, 0.01 mg/ml bovine insulin, 20 ng/ml EGF, 1 mg/ml hydrocortisone and 100 ng/ml Tetanus toxin in 5% CO2-95% air at 37 °C in a humidified incubator. MDA-MB-435 cells maintained under standard conditions in a 1:1 (v/v) mixture of DMEM and Ham F12 medium (DMEM/F12) supplemented with 10% fetal bovine serum (FBS), 1 mM pyruvate and 2 mM L-glutamine, in the same incubator conditions as described above were used in some experiments.

8.2.2 Raman spectroscopy

Each measurement required 3x105 cells, or 3x104 cells for MCF10A cells in sparse conditions. Cells were seeded in six-well plates (Becton Dickinson, NJ) over a quartz crystal (ESCO products, Oak Ridge, NJ), which was used to reduce the background signal. After 24 h, the cells were fixed with 4% cold paraformaldehyde (PFA) in PBS 1x for 15 min, washed with PBS 1x and maintained in the same solution at 4 °C until measurements. The Renishaw Raman system (Apply Innovation, Gloucestershire, UK) comprises a 514 nm laser that supplies an excitation beam of about 5 mW power, which is focused onto the sample via a microscope using a 60x objective (Edmund, York, UK). The same objective collects the scattered light from the sample and directs it to

the spectrometer. The spectrometer processes this scattered light, by rejecting the unwanted portion and separating the remainder into its constituent wavelengths. The Raman spectrum is recorded on a deep depletion charge-coupled device (CCD) detector (Renishaw RenCam). The recorded Raman spectrum is digitalized and displayed on a personal computer using Renishaw WiRE software, which allows the experimental parameters to be set. The spectra were background subtracted using a custom-written Labview program.

8.2.3 Statistical analysis

Two multivariate techniques were performed on cellular Raman spectra, in order to evaluate the spectral differences among the cancerous cell lines studied and to develop a model allowing their discrimination and classification: MCR and PLS-DA.

In Multivariate Curve Resolution (MCR), a bilinear relation of the spectrum is assumed from the matrix, taking into account the Beer Lambert law, X=CST+E, where X is the Raman spectra acquired matrix, C is the concentration matrix transposed and E is the matrix of spectral errors. The algorithm used was Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS). Using a first PCA on the dataset of spectra, the number of components necessary to explain the variance in the data was selected, accounting for 95% of the total variance. For ALS optimization, constraints of non-negativity in the spectra and concentrations were applied. A limit of 300 iterations was applied to reach the 0.95 confidence limits. MCR-ALS allowed visualization of the bands (biomolecular vibrations) responsible for the changes in metabolite concentrations in the different cell lines. Finally, five molecular components were deconvolved from the spectra.

PLS-DA is a supervised classification method in which knowledge of the sample is included (in our case metastatic, malignant and benign MCF10A phenotypes). PLS-DA employs the fundamental principle of PCA but further rotates the component (latent variables, LVs) by maximizing the covariance between the spectral variation and group affinity so that the LVs explain the diagnostically relevant variations rather than the most prominent variations in the spectral dataset. The performance of the PLS-DA diagnostic algorithm was validated using leave-one-out cross validation methodology. The number of retained LVs was determined based on the minimal root mean square error of cross validation (RMSECV) curves.

Multivariate statistical analysis was performed using the PLS toolbox (Eigenvector Research, Wenatchee, WA) in the Matlab (Mathworks Inc., Natick, MA) programming environment. Before including Raman spectra in the multivariate statistical techniques, correct preprocessing must be performed. In this case, multiple scattering correction and background subtraction was achieved using a Matlab and Labview algorithm54. For Raman spectral analysis, the region between 1015 and 1110cm-1 was removed because it contained a background-related signal that reduced the quality and interpretability of our statistical models.

8.2.4 Gene expression analysis

Transcriptomic data from MCF10A cells cultured in sparse or confluent conditions with accession number GSE84305 were downloaded from the Gene Expression Omnibus database [215]. The series matrix contained a normalized log ratio (z-score) for the differences in expression between sparse and confluent cells. The median z-score between the six replicates included in the experiment was calculated for each probe and then used to rank all genes by their level of expression.

In addition, a series matrix including normalized data from GSE18070 that contains different strands of k-ras transfected MCF10A cells was downloaded from the GEO repository [215]. To assess gene expression differences, a t-test between all probe sets from three hybridized MIII samples (cells with ability to metastasize in vivo) and three MII samples (cells without metastasis ability) was conducted. Then the 33853 genes included in the array (Affymetrix U133 Plus 2.0) were ranked based on their statistic expression level..

A pre-ranked GSEA (Gene Set Enrichment Analysis) using both lists of ranked genes was run. Only 81 genesets related to "fatty acid", "lipid", "glucose", "cholesterol", "EMT" and "tryptophan"; and those genesets described in Charafe et al. [216] including differentially expressed genes between different kinds of breast cancer were included in the analysis. The statistical significance of the enrichment score was calculated by permuting the genes 1,000 times as implemented in the GSEA software. Functional terms were considered to be significant at FDR q-value 25%. The conventional statistical methods for microarray gene expression analysis choose a list of differentially expressed genes based on a p-value corrected by multi-testing. Using this astringent criterion, a large number of genes that do in fact contribute to the studied phenotype could be missed. The GSEA (gene set enrichment analysis) approach was designed to address the limitation of single gene analysis since it uses all transcriptomic information included in the array [217].

All computations were performed using R statistical software and Bioconductor [218].

8.3 Results

8.3.1 Multivariate Curve Resolution (MCR) algorithm is able to decompose cell Raman spectra in meaningful Raman spectra of metabolites

RS of biomedical samples is inherently complex and weak. Multivariate analysis improves its ability to obtain useful information from biomedical profiles. Our innovative application consisted of analysing the RS of cells in different EMT states using MCR, an unsupervised computational analysis. This algorithm iteratively derives the pure component spectra from a spectral data set and the contributions of each pure component in each spectrum acquired. This statistical technique efficiently disentangle the information encoded in complex spectra such as Raman.

To distinguish the metabolic features of cells with the EMT phenotype [219], as a signal of the initiation of metastasis, the focus was on characterizing the

spectroscopic parameters of a prototype of metastasis-initiating cells in comparison with the highly metastatic MDA-MB-435 breast cancer cells. Confluent and sparse differentiated MCF10A, according to the TUFA band intensity, which is lower in sparse conditions, were previously associated with the EMT phenotype [211]. RS of both phenotypes (confluent and sparse MCF10A) was acquired using an InVia Raman microscope (Renishaw) and the fingerprint region of the RS (600-1800cm-1). The MCR-ALS algorithm was then used to deconvolve from the spectra the metabolites that differed among the cell groups. One problem in resolution analysis methods is that the components deconvolved from the experimental Raman spectra are mixed with background signals and are therefore difficult to interpret. This problem was solved by adding to the data set a small number of spectra obtained from the PBS (buffer, not cells) at the same focal distance at which the cellular spectra were acquired. Including this extra information in the MCR-ALS algorithm allows the deconvolution of a component that contains only background spectra (quartz and water). Therefore, the remaining components only account for the differences in the metabolite content in the cells. Four pure spectra were deconvolved from inherent cellular components and are shown in Figure 8.1. Score plots relating different components are presented in Figure 8.2. The control spectra from PBS were plotted on the scores plot as blue crosses and were always around the lowest levels of the concentration of the different cellular metabolites, indicating the absence of those molecules from the spectra.

With the help of previously published Raman databases [189, 220–226], we identified component 1 as tryptophan [221], component 2 as phosphatidylserine [220], component 4 to polysaccharides [222–226] and component 5 as cytokeratin [189, 221]. Component 3 accounted only for background contribution (Figure 8.1). Table 8.1 lists the assigned metabolites in each component [189, 220–226]. Component 1 included mainly phenylalanine and tryptophan: 1001, 1170, 1204, 1300, 1315, 1336, 1359 and 1587 cm^{-1}. Component 2 included bands assigned to phosphatidylserine: 1123, 1260, 1300 and 1438 cm^{-1}. Raman bands in component 4 can be assigned to polysaccharides: 940, 986, 1125, 1187, 1207, 1337, 1365 and 1452 cm^{-1} . All the bands contained in component 5 are characteristic of keratin (936, 1296, 1335, 1448 and 1650 cm^{-1}) and amide (1235 and 1240 cm^{-1}).

Therefore, MCR resolved meaningful components in the spectra with respect to classical techniques such as Principal Component Analysis (PCA), demonstrating a high capacity to deconvolve the pure cell component spectra of molecules from a set of Raman spectra. Moreover, the outputs of the statistical analysis provided more chemically and physically understandable results than classical techniques such as PCA, via which only a mathematical exploration of the data can be performed and components have mixed molecular contributions and negative loadings.

8.3.2 RS coupled with MCR dissects the metabolic phenotype of EMT in breast cancer cells

We conducted an in-depth analysis of the most significant differences between confluent and sparse MCF10A cells to assess the different profiles of the spectral composition and to build a classification algorithm for the discrimination of EMT cells with the metastatic-initiating phenotype from the highly

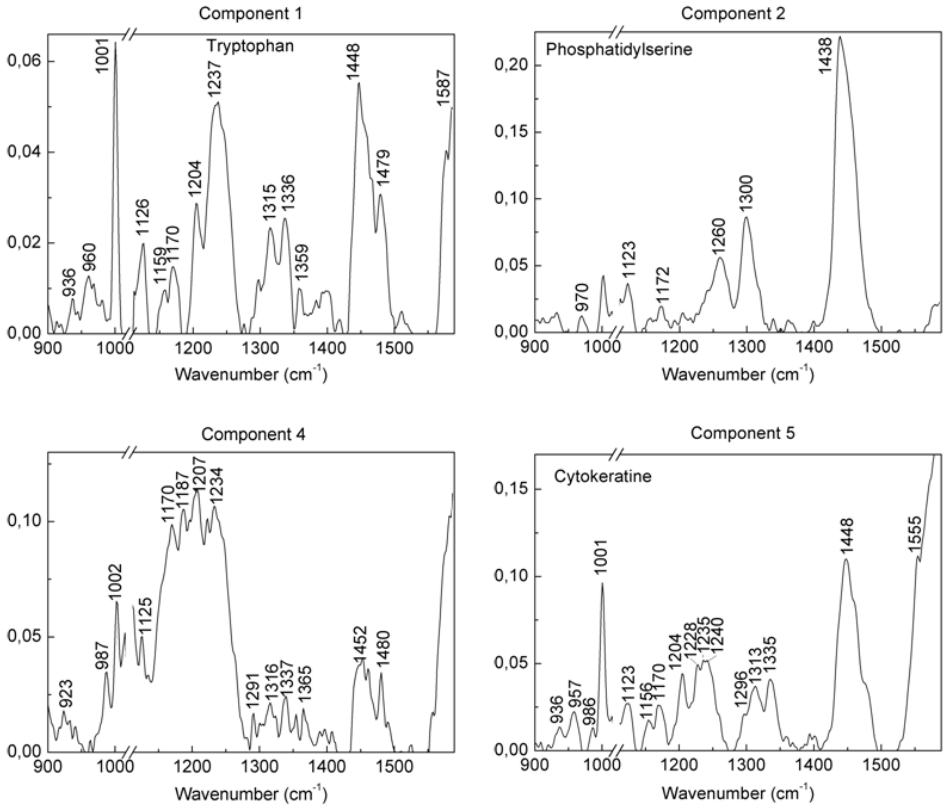


Figure 8.1: Component spectra obtained by Raman microspectroscopy deconvolved from MCR analysis using sparse and confluent MCF10A cells and MDA-BM-435 cells. The main characteristic bands of component 1, 2, 4 and 5 are indicated in each spectrum. Assignments based on references [189, 220, 221]

Table 8.1: Band assignments for the Raman spectra corresponding to components 1, 2, 4 and 5 obtained by MCR analysis of sparse and confluent MCF10A cells and MDA-MB-435 cells. Assignments based on references [189, 220, 221]. PA: Phenylalanine, PhS: Phosphatidylserine, T: Tryptophan, K: Keratin, A: Amide, PS: Polysaccharides

Component 1		Component 2		Component 4		Component 5	
Raman shift (cm^{-1})	Assignm.	Raman shift (cm^{-1})	Assignm.	Raman shift (cm^{-1})	Assignm.	Raman shift (cm^{-1})	Assignm.
1001	T, PA	1123	PhS	940	PS	936	K
1170	T, PA	1260	PhS	986	PS	1001	PA
1204	T	1300	PhS	1125	PS	1235	AIII, K
1237	T, AIII	1438	PhS	1187	PS	1240	AIII, K
1300	T			1207	PS	1296	K
1315	T			1337	PS	1335	K
1336	T			1365	PS	1448	K
1359	T			1452	PS	1650	K
1448	T, CH_2 deform						
1479	T, AII						
1587	T						

metastatic MDA-MB-435 cells. We plotted the contributions of component 2 (phosphatidylserine) with respect to component 5 (keratin) (Fig. 8.2a). EMT appeared inversely correlated with keratin, with MDA-MB-435 cells having low keratin level whereas confluent MCF10A cells had the highest (Figure 8.2a and 8.2c). Component 2 (phosphatidylserine) separated two MDA-MB-435 cell populations: SP435A with the highest phosphatidylserine content, and SP435B with a moderate phosphatidylserine increase. In contrast, sparse MCF10A cells had a low keratin and increased phosphatidylserine content compared with confluent MCF10 cells, more similar to SP435A. Indeed, the most lipid accumulating cells were sparse MCF10A and SP435A. Thus, RS accurately classified EMT based on low keratin and a high lipid and phosphatidylserine content, so called lipid phenotype (LP). In addition, component 1 identified as tryptophan (T) clearly segregated sparse MCF10A from the rest of the cells, and therefore high levels of this metabolite were an idiosyncratic element of the EMT phenotype (Figure 8.2b and 8.2c). Taking into account scores on tryptophan and phosphatidylserine (Figure 8.2b), the aggressive metastatic phenotype of MDA-MB-435 cells was strongly associated with an increase in phospholipids with low tryptophan content (Figure 8.2b). On the other hand, MDA-MB-435 cells were split into two clear groups, one with low lipid content and the other with the highest lipid content, according to its heterogeneity [227]. Moreover, in figure 8.2d it was possible to distinguish the highest phosphatidylserine SP435A cells (with a lipid phenotype) from SP435B cells (with a low lipid concentration and increased levels of Polysaccharides (component 4)), suggesting that this component may act as an alternative fuel source for MDA-MB-435 cells with a lower lipid content [227]. Therefore, component 4 may be related to glucose metabolism. Furthermore, the cloud of sparse MCF10A cells was displaced with respect confluent MCF10A cells in Fig. 8.2d. Sparse MCF10A showed an increase LP and low PS content

whereas confluent MCF10A have higher PS and lower LP what support PS as an alternative fuel source for low lipid accumulating cells.

We performed a PLS-DA model in order to explore the ability of RS to differentiate between sparse and confluent MCF10A and MDA-MB-435 cells. The model (Figure 8.3) gave 100% sensitivity and 100% specificity for discrimination of sparse MCF10A, 94% sensitivity and 89% specificity for confluent MCF10A and 80% sensitivity and 100% specificity for MDA-MB-435 using the cross validation method venetian blinds with 7 splits. Overall there was a clear difference between confluent and sparse MCF10A cells, which reflected the increased EMT of sparse, compared to confluent MCF10A cells; they could thus be stratified by profiling the lipid accumulating phenotype, which might be enriched by the increased tryptophan concentration in sparse MCF10A cells.

8.3.3 Differential expression of genes validates the phenotype of EMT.

We performed functional enrichment analyses to validate the spectral differences between confluent and sparse MCF10A cells. We used public transcriptomic data downloaded from Gene Expression Omnibus database GSE84305 to search for functional differences between sparse and confluent cells. This dataset contains a normalized log ratio (z-score) for the differences in expression of 11500 probe sets representing 9300 genes from the CNIO Homo sapiens Oncochip [201]. Genes were ranked by their median z-score and gene set enrichment analysis (GSEA) algorithm [228] was applied to identify specific functions in the list of pre-ranked genes. A total of 71 gene sets including the keywords "fatty acid", "lipid", "glucose", "cholesterol", "EMT" and "tryptophan", and those gene sets described in Charafe et al. [216], including genes that are differentially expressed in different kinds of breast cancer, were tested. Genes that were downregulated in luminal-like breast cancer cell lines in comparison with mesenchymal-like ones [216] were expressed in sparse MCF10A (FDR$<$0.000), whereas confluent MCF10A contained genes that were upregulated in luminal-like breast cancer cell lines in comparison with the mesenchymal-like ones (FDR$<$0.000). As expected, sparse cells showed a significant enrichment in genes that were upregulated in EMT whereas confluent cells were enriched in EMT downregulated genes, pointing to a mesenchymal phenotype. Surprisingly, both up- and downregulated EMT genes described in Jechlinger et al. [229] were significant gene sets in sparse cells, but downregulated genes were much more statistically significant (FDR=0.01).

Fatty acid and glucose metabolic functions differentiated confluent and sparse MCF10A cells. The metabolic phenotype of sparse MCF10A cells was characterized by genes involved in glucose deprivation (FDR=0.0020) and glucose transport (FDR=0.120), whereas in confluent MCF10A cells the profiling showed an enrichment in genes involved in glucose metabolism (FDR=0.032). Moreover, differences included increased fatty acid metabolism (FDR=0.240) and fatty acid oxidation (FDR=0.150) in confluent MCF10A. Since fatty acid metabolism was increased in these cells their glucose dependence was minimized with regard to sparse cells, suggesting that confluent and sparse MCF10A cells have a different metabolic phenotype in both conditions.

One of the first metabolic alterations observed in cancer cells was the higher dependence upon glucose for their growth, the so-called Warburg effect, indi-

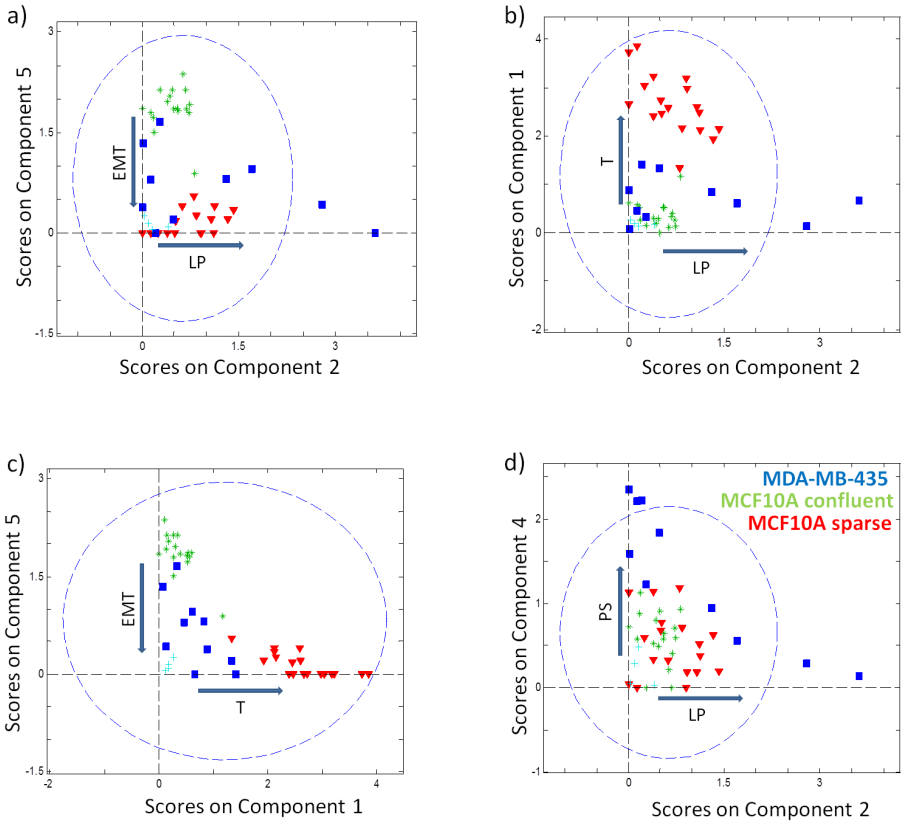


Figure 8.2: MCR scores plot for components represented in Figure 8.1. Confluent MCF10A cells (green asterisks), sparse MCF10A cells (red triangles) and MDA-MB-435 cells (dark blue squares). Light blue crosses are control spectra obtained from the PBS solution at the same focal depth. The arrows indicate an increase in the EMT (epithelial-to-mesenchymal transition), the LP (lipid phenotype), the tryptophan (T) content or the polysaccharide (PS) content. a) Component 5 (keratin), a marker of EMT, is plotted with component 2, which was assigned to phosphatidylserine, an indicator of the LP. b) Component 1, assigned to tryptophan, is plotted with component 2 (phosphatidylserine). c) Keratin is plotted with component 1, assigned to tryptophan. MCF10A cells in both culture conditions are completely separated by differences in the levels of both components. d) Component 4, assigned to polysaccharides, is plotted with the component 2 (phosphatidylserine). MDA-MB-435 cells are clearly split into two subpopulations: SP435A (with high lipid content) and SP435B (with lower lipid content).

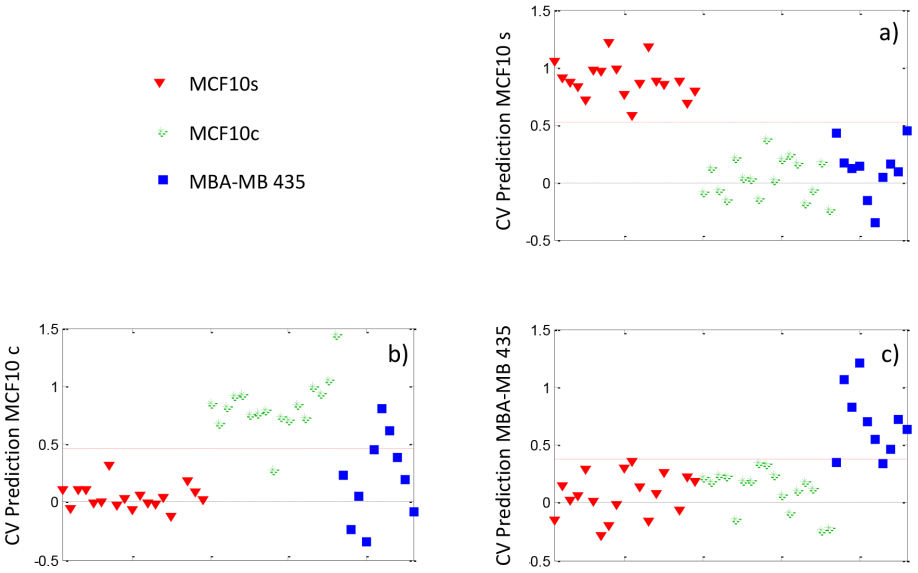


Figure 8.3: Results of PLS-DA analysis to explore the ability of RS to differentiate between sparse and confluent MCF10A cells and MDA-MB-435 cells. The model gives 100% sensitivity and 100% specificity for sparse MCF10A discrimination (a), 94% sensitivity and 89% specificity for confluent MCF10A discrimination (b), and 80% sensitivity and 100% specificity for MDA-MB-435 identification (c) using the cross-validation method venetian blinds with 7 splits.

cating that the EMT process in sparse MCF10A cells altered their metabolism, giving them more oncogenic properties [230]. Moreover, tryptophan metabolism and tryptophan catabolism were statistically significant functions (FDR=0.190 and 0.136 respectively) in confluent cells, which expressed lower levels of tryptophan than sparse cells. Due to the low number of metabolism-related genes represented in the GEO2603 chip that hybridized with sparse and confluent MCF10A cells, an alternative analysis using data from Papageorgis et al. [231] was conducted. To assess differences in gene expression, a t-test between all probe sets from three hybridized MII samples (k-ras transfected MCF10A cells without metastasis ability) and MIII samples (k-ras transfected MCF10A cells with ability to metastasize in vivo) was conducted at the transcriptomic level. The 33853 genes included in the array (Affymetrix U133 Plus 2.0) were ranked by the t-statistic. In this way, positive values corresponded to a MIII phenotype whereas negative values were characteristic of MII. A GSEA using the same gene sets as those in the GSE8430 study was then run. A total of 33853 genes were hopefully hybridized on the chip, so more metabolism-related genes were included in the functional enrichment analysis. Significantly enriched functions at FDR 25% were found such as regulation of insulin secretion by free fatty acids (FDR=0.150) in MIII cells, whereas in MII cells glucose transporters (FDR=0.189), lipid rafts (FRD=0.029), phospholipid binding (FDR=0.124) and cholesterol biosynthesis (FDR=0.002), among others, were enhanced.

Interestingly, upregulated genes from the dataset "Sarrio epithelial mesenchymal transition up" were significantly associated with the MIII phenotype, suggesting a parallel between the MIII vs. MII comparison and the sparse MCF10A vs. confluent MCF10A comparison (FDR = 0.013). Along with this gene set, other sets including genes upregulated in EMT were found to be associated with MIII cells, whereas datasets including downregulated genes in EMT were associated with MII cells. As expected, MIII cells showed a more mesenchymal phenotype, with significant gene sets being those genes described in Charafe et al. [216] as downregulated in basal and luminal breast cancer sub-types. According to these transcriptomic public data, we inferred that fatty acid and glucose metabolic functions differentiate confluent and sparse MCF10A cells, then supporting the RS profile.

8.4 Discussion

To our knowledge this study is the first time MCR has been used to deconvolve and track the molecular content of cancer cells during the EMT process, which involves biochemical changes linked with the early metastatic phenotype of cancer cells. We have demonstrated that the unsupervised computational MCR analysis resolves chemically and physically meaningful component spectra from a set of mixed Raman spectra, allowing the assignment of pure molecular spectra to each component deconvolved and its contribution (concentration). MCR is more powerful than the classical PCA technique, in which components can contain negative regions, impeding the assignment of molecular spectra [214]. Therefore, MCR is a good candidate for tracking the biochemical composition of a biomedical sample during the transition involving molecular changes in the early diagnosis of breast cancer.

During the last two decades MCR has been applied to a wide range of chemi-

cal data [47,48,50] and has been adapted to Raman spectral data from biological tissues for imaging their molecular content [35, 189]. High-contrast images of uterine tissue derived using RS with an empty modelling approach of multivariate curve resolution-alternating least squares [189] have been obtained. We have shown the ability of MCR to track the biochemical composition of a biological sample during a process involving molecular changes. These molecular changes are sometimes masked by background signals that perturb the statistical analysis, an important problem when studying cells using RS. This is because cells need to be grown on a substrate (normally glass) that contains Raman and fluorescence bands, and when acquiring the cellular spectra, the focal point needs to be very close to the window, which in some cases covers part of the confocal volume. The ability of MCR to deconvolve pure component spectra allow these undesired components to be removed and only the inherent changes occurring in cellular RS to be analysed. Our results demonstrate an improvement in the molecular characterization of cells when using RS.

MCR performed on the RS of confluent and sparse MCF10A cells identified the metabolic profile of EMT, which is characterized by a decrease in the cytokeratin content and an increase in the phosphatidylcholine and tryptophan content. These results are consistent with the gene profile attributed to altered lipid metabolism functions like fatty acid metabolism and fatty acid oxidation, which increase in malignant cells [232]. Moreover, the PLS-DA analysis of the cellular RS classified EMT cells in the 600-1800 cm-1 range with 100% sensitivity and 100% specificity, indicating that RS provides powerful information allowing stratification of cells by their metastasis initiation process phenotype.

MCF10A cells in confluent conditions had the highest content of cytokeratin according to the composition of their pre-EMT cytoskeleton. In contrast, MCF10A cells that undergo EMT and MDA-MB-435 cells contain abundant vimentin cytoskeleton [189]. Overall, there was a clear difference between confluent and sparse MCF10A cells, reflecting the increased EMT of the latter. These cells can thus be stratified using the lipid-accumulating phenotype profile, which has been found in the most aggressive cancer cells and is related to their increased migration in vitro [233,234]. Moreover, the increased tryptophan concentration in sparse MCF10A compared to confluent and MDA-MB-435 cells might be secondary to the catabolic activation of L-tryptophan by indoleamine 2,3-dioxygenase (IDO1). The consumption of this amino acid by cancer cells has been related to immune system evasion [235].

The role of IDO1 in tumour cells has been associated with cell cycle regulatory genes, increased proliferation and inhibition of apoptosis [236]. Thus tryptophan metabolism could exert immunological and non-immunological effects, in turn improving cancer progression. Moreover, IDO inhibition with siRNA leads to diminished breast cancer cell proliferation [237]. The activation of genes associated with the tryptophan metabolism has been reported in other types of carcinoma such as the cervical carcinoma, indicating a common metabolic change linked to malignancy [238, 239]. The epithelial-to-mesenchymal transition (EMT) is one of the main mechanisms involved in breast cancer metastasis and most likely contributes to metastasis in all types of carcinoma [229,240,241]. Data obtained by the evaluation of RS using MCR were sensitive enough to detect differences in the breast cancer basal phenotype, since all analysed cells have a basal phenotype and share some similarities. The MDA-MB-435 cell population is basically fed by components 2 (phosphatidylserine) and 4 (polysaccha-

rides). Indeed, MDA-MB-435 cells can be split into two clear groups, one with a low lipid content (phosphatidylserine) and higher levels of polysaccharides and the other with a higher lipid content, indicating the important role of lipids as a fuel source for metastatic cells under nutritional deprivation, which in turn could influence metastasis organ-specificity [227, 242].

Cancer is a very heterogeneous disease with a large number of genetic alterations. The molecular characterization of cancer cells provides interesting insights into breast cancer taxonomy but its implementation in clinical diagnosis is difficult because it is so expensive [243]. RS together with multivariate analysis has shown a high ability to differentiate cells with different degrees of malignancy according to their biochemical composition. This was validated by gene expression analyses, which confirmed that these differences in cell composition were closely associated with alterations in lipid metabolism that are associated with malignancy. Compared with other decomposition techniques used in RS, such as Principal Component Analysis, MCR analysis gives more chemically and physically understandable results.

There is increasing evidence of the association between phospholipids and cancer. The analysis of MCF10A cells showed initial changes in the lipid content of cells associated with their EMT phenotype. In carcinoma cells EMT is related to higher aggressiveness and invasive and metastatic potential [244]. Activation of lipid metabolism is an early event in carcinogenesis and also a hallmark of many cancers [245]. In our attempt to study the metabolic changes involved in the EMT process we analysed sparse MCF10A cells, which show many features of post-EMT cells in in vitro cell culture, such as increased motility5. Our results clearly indicate that an increase in phospholipids is associated with the EMT process in breast cells. Characterization of this initial malignant process, which is closely related to metastasis progression, might be useful for diagnostic purposes. The essential hallmarks of cancer are intertwined with altered cancer cell-intrinsic metabolism, either as a consequence or as a cause [246]. New approaches based on a panel of small molecules derived from the analysis of metabolic profiles of cells are being developed to characterize cancer cell-specific metabolism [247]. Many observations during the early period of cancer biology research identified metabolic alterations such as the Warburg effect [230]. In this scenario RS supported by MCR can provide a useful tool with which to assess the biomolecular composition of tumour cells and metastasis-initiating cells. The applications of spectroscopic methods in cancer detection open new possibilities in the early stage diagnosis of breast cancer.

9

Molecular monitoring of retina inflammation reveals mitochondria stress and phosphatidylcholine decrease using Raman spectroscopy

Retinal tissue sustains substantial damage during the inflammatory processes characteristic of Multiple Sclerosis (MS). The analysis of neuronal and axonal degeneration within the retina represents a unique opportunity to study neurodegeneration in MS because this tissue is accessible to direct imaging by photonics. Furthermore, the analysis of molecular changes at the retinal level will increase our understanding of neuronal loss and axonal transection and would aid the development of new neuroprotective strategies for brain and retina diseases. RS is a promising candidate technique to monitor the molecular changes of the dissease tissue in vivo due to is specificity and non-invasive character. In previous chapter, we demonstrated the ability of MCR to disentangle the molecular components that change in cellular Raman spectra along a biochemical process. Then, the application of MCR to Raman spectra opens up the possibility to dynamically monitor molecular components in living tissue. For this reason, in this study, we propose MCR as a technique able to decompose retina spectra in meaningful molecular components that were tracked over the inflammation process. Retinal inflammation was induced by challenging murine retinal cultures with lypopolysaccharide (LPS), activating microglia and inducing oxidative stress and pro-inflammatory cytokines, which work to promote axonal damage. The evolutions over 48 hours of six different molecular components were deconvolved from the Raman spectra by multivariate curve resolution (MCR), and were assigned based in the Raman spectra from candidate molecules involved in neuroinflammation. We observed the increase in levels of immune mediators such as Lipoxygenase, iNOS and TNFα, peaking at 10-12 h

after LPS challenge, coinciding with maximum production of oxidative stress. We also found significant changes in key molecules involved in energy production by mitochondria, such as the increase of Cytochrome C levels by 24h and the decrease in NADH / NAD^+ levels by 12 h after LPS. Finally, we observed a decrease of the fatty acid Phosphatidylcholine at 12 h after LPS challenge, which parallels tissue degeneration and thinning of the retina. In summary, we present a novel methodology to monitor the evolution of different molecular components in retina due to inflammation by using Raman spectroscopy coupled to MCR, becoming a promising tool that could be used to study the human retina and diseases of the central nervous system in vivo. Our study represents the first application of MCR to decompose and monitor the molecular content of a biological tissue with RS.

9.1 Introduction

The retina is a distinctive component of the central nervous system (CNS) in which retinal ganglion cells from the ganglion cell layer (GCL) projects their axons to the brain through the optic nerve. Axons from the GCL are not wrapped by myelin sheaths until they exit the eye and are frequently damaged in several retinal and brain diseases such as Glaucoma and Multiple Sclerosis (MS). For this reason, the analysis of their axons in the retina will reveal information about mechanisms of axonal degeneration independent of the demyelination that is characteristic of MS. Inflammation of the retina is involved in promoting axonal damage in MS and other diseases of the retina. Retinal inflammation is primarily mediated by microglia activation, including the release of pro-inflammatory cytokines and the creation of oxidative stress that impairs neuronal function and promotes axonal damage [248–253]. MS is an inflammatory disease of the CNS inducing widespread demyelination and axonal loss, involving in almost all cases the anterior visual pathway including the optic nerve and the retina [254]. As such, inflammatory perivascular cuffs and widespread microglia activation is observed in the retinas of patients with MS [255]. Axons damaged by presence of plaques within the optic nerve or by trans-synaptic degeneration due to lesions in optic radiations suffer retrograde degeneration and can be assessed by optic coherence tomography [256, 257]. The analysis of neuronal and axonal degeneration within the retina represents a unique opportunity to study neurodegeneration in MS because this tissue is accessible to direct imaging by photonics. Moreover, axonal degeneration is an active but poorly characterized process that might share commonalities within different brain diseases such as Alzheimer's disease, stroke or brain trauma [258–264]. Therefore, the analysis of molecular changes at the retinal level will increase our understanding of neuronal loss and axonal transection and would aid the development of new neuroprotective strategies for brain and retina diseases.

Raman technology is particularly useful for working in vivo because the near-infrared range of incident wavelengths used is non-damaging to human tissue, and also exhibits a relatively large penetration depth into a sample. Several recent studies have applied this technology to retinal tissue in animal models to quantify advanced glycosilation end products (AGE) levels in Brusch membranes [265], and to differentiate between the molecular content of different retinal layers [266]. Resonance Raman spectroscopy was used in human subjects

in vivo to quantify macular carotenoid pigments [267]. Another Raman related technology is coherent anti-Stokes Raman spectroscopy (CARS), which has been applied for obtaining in vivo microscopy images without staining, such as the elucidation of myelin damage in spinal cords following glutamate excitotoxicity in guinea pigs [268].

Although Raman spectroscopy has the maximum specificity among all optical techniques for detecting molecular changes, the interpretation of Raman spectra is complex. Biomolecules have many Raman bands and some of them have similar molecular structures. Consequently, they share groups of bands, making difficult to deconvolve the contributions of pure molecular components from the Raman spectra. During the past decades applications of Raman spectroscopy have been focused to separate several groups of samples by means of multivariate analysis such as Principal Component Analysis (PCA) or to classify samples by Partial Least squares- Discriminant analysis (PLS-DA) or Neural Networks. However, little information can be extracted from the above mentioned methods to extract meaningful molecular components from the Raman spectra. In this study, we used Multivariate Curve Resolution (MCR) to deconvolve from the set of experimental Raman spectra pure molecular components that changed during retina inflammation and we monitored the evolution of their concentrations in the tissue over a 48 hour period process. MCR is an statistical technique that efficiently extracts the information encoded in complex spectra such as those created by Raman spectroscopy [269]. MCR is an unsupervised computational analysis that iteratively derives the pure component spectra from a spectral data set and the contributions of each pure component in each spectrum acquired [270]. MCR requires minimal a priori knowledge of the system providing objective information. Also, physical characteristics of Raman spectra as non-negativity constrains can be included, giving more meaningful component spectra. In this study, very limited a priori information was known about the molecular components that changed its concentration in the retina along retina inflammation. Furthermore, non alternative chemical technique existed (other than RS) to provide information about the molecular content in a specific region of the retina in real time and in vivo. Thus, the combination of RS with MCR represents an objective and novel methodology with promising applicability to study and monitor the biochemical behavior of diseases in vivo.

Here we aim to assess the molecular changes along time of the GCL during retina inflammation by means of Raman spectroscopy. We made use of an in vitro model of neuroinflammation using murine retinal organotypic cultures, which preserve cellular composition and tissue architecture while allowing direct in vivo imaging analysis [271]. We were interested in identifying the most significant molecular changes associated with axonal damage in retina inflammation and in obtaining a proof of concept that molecular imaging of the retina by Raman spectroscopy combined with MCR can become a useful technology for studying and monitoring for MS and other brain and retina diseases.

9.2 Methods

9.2.1 Chemicals

L-glutamic Acid, cytochrome C purified from Pidgeon Breast Muscle, L-(+)-lactic acid, nicotinamide adenine dinucleotide (NADH), L-+-phosphotidylcholine, and N-acetyl-L-aspartic acid (NAA), were purchased from Sigma Aldrich and used without further modification. N-carboxymethyl lysine (CML), and N-carboxyethyl lysine (CEL), and pentosidine were purchased from Polypeptide Laboratories, CML and CEL were used without further modification. Due to packaging constrains, pentosidine was diluted in PBS for analysis. Flavin adenine nucleotide (FAD) was purchased from Carbon Scientific and used without further modification. For the spectra of molecules in solution, powders were mixed with a small amount (1 to 2 drops) of PBS 1x pH 7.4 and imaged at this high concentration, with the exception of NADH which was measured in Triz pH 7.5 due to its documented interactions with the phosphate content of PBS [272]. Eight different concentration of NAA in PBS pH 7 (4M, 3M, 2M, 1M, 0.5M, 0.1M, 0.05M, and 0.01M) were prepared by dilution methods. Three to five spectra were taken of each concentration, and spectral intensity versus concentration for each spectral peak was displayed graphically.

9.2.2 Retina organotypic cultures

Retinal cultures were prepared in accordance with published protocols (27, 43). Retinal tissue was excised from C57BL/6J p8 mice (Harlan) and whole mounted to tissue culture Millicell-CM culture inserts (pore size 0.4 μm) with the retinal nerve fiber layer (RNFL) facing up, and incubated at 37°C and 5% in culture medium (45% HAMS-F12, 45% DMEM, 10% fetal bovine serum, 10 ml/l HEPES, 10 ml/l Penicillin / Streptomycin, 10 ml/l Anti-Mycotic, 5 μ g/ml insulin). Tissue was cultured for six days, then treated with 15 μ g/ml lipopolysaccharide (LPS; from Sigma Aldrich) in the tissue culture medium for time periods up to 24 hours as described before [271,273–275].

9.2.3 Immunohistochemistry

Retinal cultures were fixed for 40 minutes in 4% paraformaldehyde 24h after treatment with LPS. Retina were blocked and permeabilized in 0.5% Triton and 10% normal goat serum in PBS 1x for 1 hour, and incubated overnight at 4°C with primary antibody in blocking solution. The samples were then washed and incubated with the secondary antibody in blocking solution, washed again, and mounted with GelMount anti-fading medium with DAPI (Sigma). Samples for bio-conjugated Lectin staining were prepared using the same protocol, with the secondary antibody omitted. Pairs of primary and secondary antibodies were as follows: rabbit polyclonal antibody to the 200 kD neurofilament heavy chain (NfH) (Abcam, 1:500) and FluoProbes 488 anti-Rabbit IgG (FluoProbes, 1:200), rat monoclonal to CD11b (anti-OX42) (Abcam, 1:100) and FITC-conjugated Lectin from Lycopersicon esculentum (Sigma Aldrich 1:200). MHC class II staining was omitted because LPS injection in the retina does not cause induction of this particular protein in retinal microglia [253]. Samples were viewed using a Leica confocal microscope Sp5 and Leica viewing software.

9.2.4 qRT-PCR

The mRNA extraction from tissue culture was performed according to the animal tissue protocol from Qiagen RNeasy Mini Kit. The total RNA concentration was measured using an UV spectrometer. 800 ng of total RNA from each sample was subjected to reverse transcription using the Qiagen RT-PCR system. First strand cDNAs were amplified using a real-time PCR thermal cycler. qPCR was performed with Qiagen qPCR kit and iNOS primers from Applied Biosystems. Mouse GADPH (Applied Biosystems) was used as an endogenous control. For relative comparison, the Ct value of each sample was normalized to the equivalent endogenous control and then all values were normalized to a randomly chosen control sample as described before [276].

9.2.5 ELISA

Retinal organotypic culture supernatants were collected at different time points after LPS stimulation (0, 6, 12, 24, and 48 hours) for the measurement of IL-1β, TNF-α and IL-6 levels. ELISAs were performed following the kit instructions from the manufacturer (Peprotech)

9.2.6 ROS measurement

Tissue cultures were incubated for 1 hour with 10 μl of 50 μM H2DCFDA (Peprotech) after LPS stimulation at 0, 6, 12, 18, 24, and 48 h. Tissue was then washed with PBS, and fluorescence was measured using a spectrofluorometer with excitation at 485 nm and emission at 535 nm as previously described [271].

9.2.7 HPLC

We pooled 8 retinas for each time point in order to assess concentration of NAD and NADH, Lactate, Glutamate and NAA. Perchloric acid 0.4 M was added, and the samples were sonicated for 10 s 3 times after centrifugation for 20 min, at 10,000 rpm and 4°C. The supernatant was collected and kept at -80°C until use. For metabolite determination, we applied the HPLC-Uvis (Breeze model; Waters) with a binary pump (model 1525), autosampler (model 2707) and UV detector (model 2489). Sample aliquots of 50 ÎĽl were injected into the Spherisarb ODS column (Waters) with a 5 μm pore size and 4.6 x 250 mm dimension. The mobile phase used was 8 mM NaH2PO4 and 18% acetonitrile adjusted to pH 3 with HCl pure acetonitrile. Quantification of the results was done by using commercial molecules at different concentrations to generate a calibration curve (range: 0.1-1 mg/ml).

9.2.8 Raman spectroscopy

Raman spectra were acquired with an InVia Raman microscope from Renishaw with a backscattered configuration. The Raman excitation was performed with 785 nm laser beam focused through a 60X 0.75 NA objective (Leica). Raman spectra for the candidate molecules were obtained by placing the powder or the liquid samples on a sample holder with a cover glass number zero and illuminated from the top with a power from 5 to 40mW. Retina organotypic cultures were measured using the preparation described above using a power of 100mW to

ensuring no photodamage was induced. The background for all spectra was subtracted using an established method [277]. Background subtraction was implemented in Labview and smoothed using a Butterworth low pass filter and five-point adjacent average. We assigned strong (s), medium (m), and weak (w) labels based on the highest, middle, and lowest third of the peak intensities of each individual spectra. For individual band assignments no shifts bigger than 10 cm-1 were considered. We only assigned a spectrum to a molecule when more than the 90% of the most prominent bands were correlated with the characteristic molecular Raman bands found in the constructed Raman database or in the literature.

9.2.9 Time series and statistical analysis

Spectra were taken randomly from 10 points in the RGC layer of the organotypic retina culture (at a depth of 15-16 um below tissue surface, Fig. 1) in healthy tissue and 2, 4, 6, 8, 10, 12 and 24 hours after LPS challenge. This process was repeated for 6 independent retinas treated with LPS and 6 control retinas. Principal Component Analysis (PCA) was performed to explore the spectra and select the number of the principal components needed to explain the maximum variance in the data [278]. Once the number of components was identified, the spectra were analyzed by Multivariate Curve Resolution (MCR) as described before [269]. In total, nine components of the spectra were deconvolved. Three components were identified as background coming from different parts of our system (the culture membrane or residual fluorescence), these components were not used in further analysis. The remaining six components were identified by their respective molecular Raman spectra and tracked through time after LPS challenge induced retina inflammation. Characteristic statistical parameters such as the p value were calculated to assess the significant difference between molecular content in control and LPS challenged tissues. Partial Least Squares - Discriminant Analysis (PLS-DA) was performed using PLS toolbox from Eigenvector Research in MatLab. The same Raman spectra used for MCR analysis were used for PLS-DA classification. Cross-validation analysis was computed by Venetian blinds (6 bags). The number of retained LVs was chosen to minimize the root mean square error of cross validation (RMSECV) curves. In plots, * signify $p<0.05$ (significant), ** means $p<0.01$ (highly significant) and *** is $p<0.001$ (very highly significant).

9.3 Results

9.3.1 Raman spectra of candidate molecules

We analyzed eight molecules by Raman spectroscopy in powder form and in solution as candidate biomarkers for monitoring retina inflammation. These molecules were selected because they play an important role in neuroinflammation and neurodegeneration and at the same time they show good signal to noise ratio using Raman spectroscopy. The spectrum of each molecule is presented as spectral library (Fig.9.1) with an accompanying chart detailing the strength (strong, medium, or weak) of the Raman activity of each peak (Table9.1). Of the molecules tested, Glutamate, Lactate, NADH, NAA, Cytochrome C and

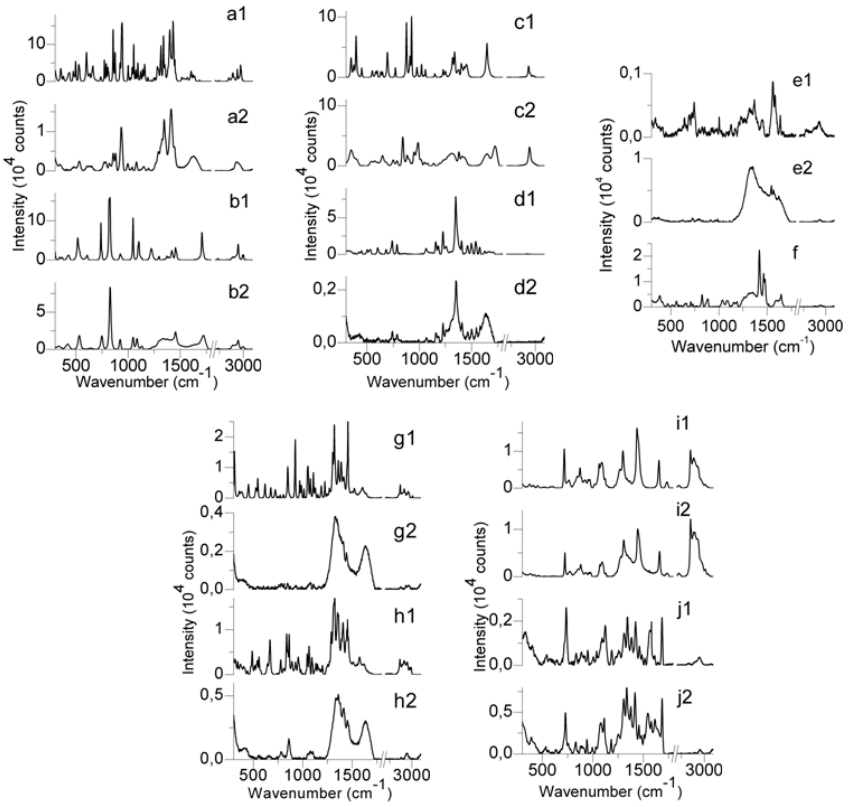


Figure 9.1: Raman spectra of candidate molecules: Glutamate in powder a1) and in PBS a2); Lactate in powder b1) and in PBS b2). NAA in powder c1) and in PBS c2); FAD in powder d1) and in PBS d2); Cytochrome C in powder e1) and in PBS e2); Pentosidine in solution f); CML in powder g1) and in PBS g2); CEL in powder h1) and in PBS h2); Phosphatidilcoline in powder i1) and in PBS i2); NADH in powder j1) and in PBS j2). 125 mW, 785 nm, 30s acquisition time.

Phosphatidylcholine presented a good signal to noise ratio in Raman spectra in solution (which better simulates the signal from living tissue, as compared to the powder form) and were used for further analysis.

9.3.2 Raman spectra from the GCL of cultured retinas

Spectra were taken at a depth of 11-12 μm from the surface of retina cultures, which given the refractive index of the retinal tissue (reported to be at around 1.35 for murine retina), places the focal point at an actual depth of around 15-16 μm below the tissue surface [279]. The GCL has been shown in adult mice to begin from 13 μm below the surface of the retina in healthy tissue to 3 μm below the surface in tissue with RNFL degradation by inflammation. The average depth of GCL is 40 μm. Taking into account the fact that post-natal p8 mice have less developed tissue structure, our probe focal point at a depth of 15-16 μm was within the GCL, with vertical peripheral data including the

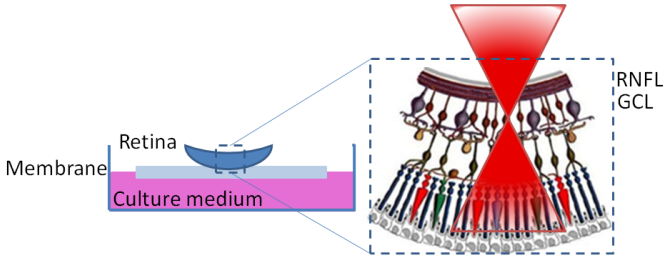


Figure 9.2: Raman spectroscopy analysis of the Retinal Ganglion Cell layer of the retina. Design of the analysis of the Ganglion cell layer (GCL) and Retinal Nerve Fiber Layer (RNFL) of the retina based in the physical properties of laser light and anatomical structure of retinal layers.

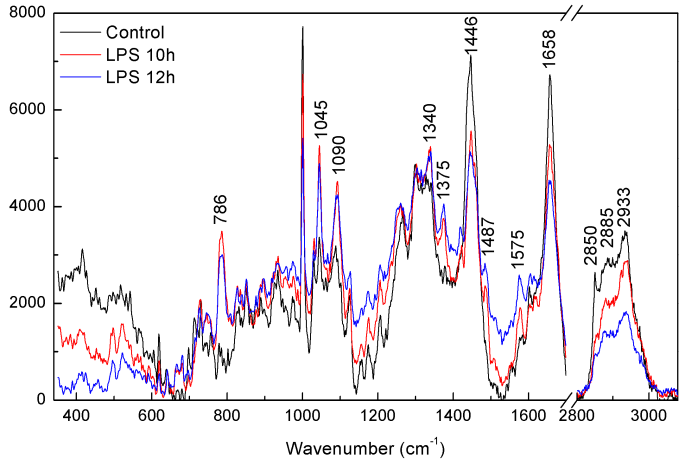


Figure 9.3: Raw Raman spectra from control retina samples after 10 hours incubation time (black) and LPS treated retina samples after 10 hours incubation time (red) and 12 hours incubation time (blue).

entire depth of the RNFL (Fig. 9.2) [280]. Using this approach we obtained the Raman spectra from murine retina cultures, that revealed a complex pattern and contain the peaks of the several candidate molecules under study (Fig. 9.3).

9.3.3 LPS induces microglia activation, oxidative stress and axonal damage in retina cultures

In order to assess the changes in the GCL in response to inflammation, we used a model of retina inflammation in which retina organotypic cultures are challenged with LPS [271, 273–275]. Immunostaining of OX42 reveals activation of retinal microglia 24 h after LPS challenge, as evidenced by their amoeboid shape (Fig. 9.4I a-b). Tomato Lectin staining clearly delineated ramified microglia with typical long branching processes in control cultures, versus the rounder activated amoeboid state found in cultures stimulated with LPS by same time (Fig. 9.4I c-

d). Immunostaining of the neurofilament heavy chain (NfH) clearly showed the formation of axonal spheroids, a feature known to be typical of the occurrence of impaired axonal transport and axonal dysfunction and end-bulbs indicative of axonal transection in cultures treated with LPS (arrows in Fig. 9.4I f) [264,271]. The level of pro-inflammatory cytokine IL-6 was assessed in culture supernatant from baseline to 48 hours after LPS stimulation by ELISA, showing a peak at 18 hours and a subsequent increase to 48 hours after LPS stimulation (Fig. 9.4II). The level of reactive oxygen species was measured from baseline to 24h after LPS stimulation, showing a significant increase 12 hours after LPS challenge and remain elevated for the next 48 hours (Fig. 9.4III). Overall, these results show the time dynamics of microglia activation and axonal damage in retina inflammation over 24 h after LPS challenge. The Raman spectra from murine retina cultures challenged with LPS reveals significant differences compared with control retinas (Fig. 9.3).

9.3.4 Raman spectroscopy combined with Multivariate Curve Resolution reveals significant molecular changes in immune, energy and lipid mediators during retina inflammation

Changes in the Raman spectra during the process of neuroinflammation were analyzed from spectra taken at random locations within the central region of each organotypic retina culture at a constant depth of 15-16 μm, putting the focal point within the GCL of the culture. Two independent experiments were performed with 6 retinas per experiment, three retinas were control and three were stimulated with LPS. Ten spectra were taken for each time point and retina. We characterized molecular changes in the murine retinas during inflammation using MCR analysis. Nine components were deconvolved from the spectra by MCR, capturing 99.7% of total variance. Three components were used for background removal from the culture membrane or media and undesired fluorescence (see methods). The six remaining components were assigned based in Raman databases or the spectra from our candidate molecules known to have a role in neuroinflammation (see methods).

The first MCR component was 5-Lipoxygenase (Fig. 9.5A) [281], an enzyme involved in metabolism of eicosanoids such as prostaglandins and leukotrienes and therefore critical in the innate immune response. From the reconstructed concentration curve a significant increase of 5-Lipoxygenase at 10 to 12 h after LPS challenge was observed. The second component corresponded to iNOS and TNFα (Fig. 9.5B) [282], two well-known proinflammatory molecules involved in CNS inflammation. The temporal evolution of the concentration of iNOS and TNFα demonstrates a significant increase at 10 hours after LPS application with a maximum at 12 h. Component 3 corresponded to Cytochrome C and Phenylalanine (Fig. 9.5C). These molecules experience a significant increase in their concentrations 24 h after the LPS administration. Component 4 was Phosphatidylcholine (Fig 9.6A) [283]. The concentration plot showed a significant decrease at 10 to 12 h after LPS challenge compared with baseline. Component 5 corresponded to NAD+ and NADH (Fig 9.6A) [284]. NAD+/NADH levels fluctuate with an early increase by 6 h and a decrease 12h after LPS challenge. The last component (component 6) corresponded to a mixture of molecules, which

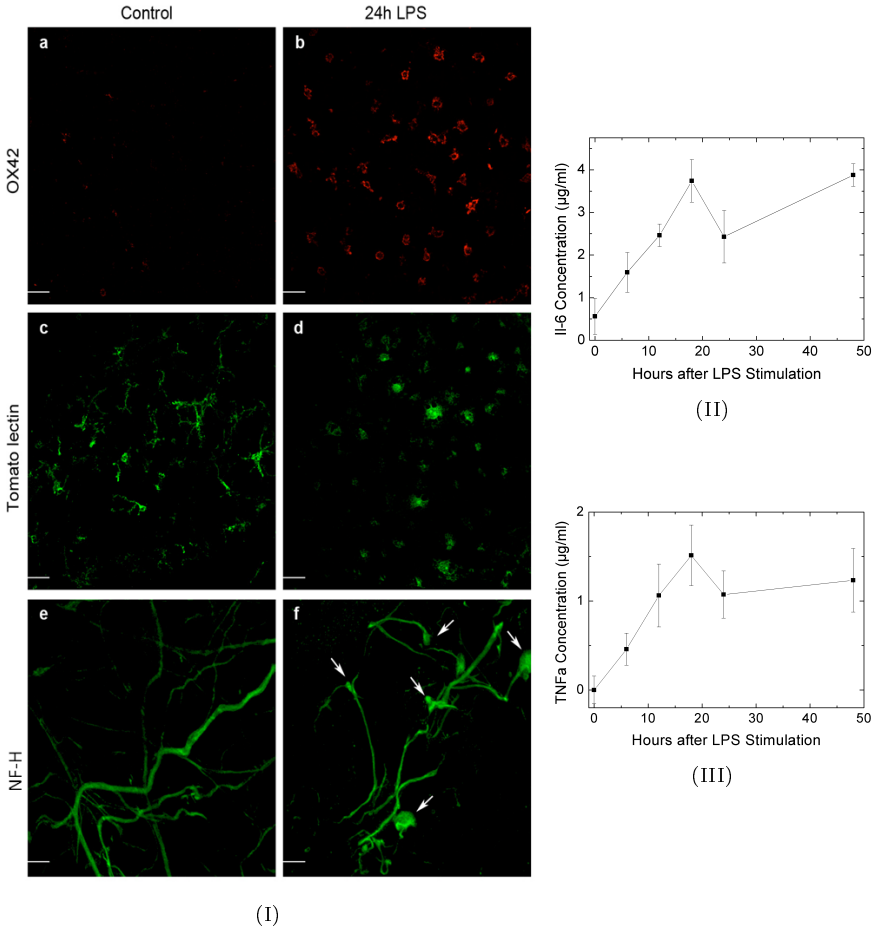


Figure 9.4: **Morphological and molecular characterization of retina inflammation.** Immunohistochemical analysis of organotypic retinal tissue culture 24 hours after LPS challenge as compared to control tissue. I) Representative confocal images of retinal cultures labeled with primary antibodies against OX42 (a,b), Tomato Lectin (c,d), and NfH (e,f). Scale bar 20 μm (a,b,c,d), 50 μm (e,f). Arrows indicate either axonal swellings (spheroids) or axonal end-bulbs (axon transection). Induction of pro-inflammatory cytokines and oxidative stress during retina inflammation: II) time-course analysis via ELISA of cytokine levels for IL6 in culture supernatant at hours 0, 6, 12, 18, 24, and 48 after addition of LPS; III) time-course analysis via spectrofluorometer of reactive oxygen species levels in tissue cultures at 0, 6, 12, 18, 24, and 48 hours after addition of LPS.

include several of our candidates such as glutamate, NAA, lactate or FAD, as well as structural molecules of the retina present in GCL that do not undergo significant changes in concentration during the 24 hours after LPS (9.6C) and for this reason they were not further characterized.

9.3.5 PLS-DA classifier identifies Raman spectra pattern associated with retina inflammation

We made use of Partial Least Squares-Discriminant Analysis (PLS-DA) to asses the ability of Raman spectroscopy to discriminate retina inflammation. The PLS-DA algorithm was able to classify retina tissue in control or LPS challenged with good accuracy (sensitivity: 0.91, 0.87 and 0.95, and specificity 1.0, 0.93 and 0.84 for 10, 12 and 24 h respectively; Fig. 9.7A - 9.7C). Therefore, the fact that the PLS-DA classifier achieved a high accuracy for discriminating between healthy and inflamed tissue support Raman spectroscopy as a tool for monitoring retina inflammation. Also, we identified the best time to detect inflammatory changes was at 10 h, which matches with the maximum changes observed from the MCR results.

9.3.6 Validation of iNOS, TNFα and NAD/NADH changes in retina inflammation

In order to validate several of the changes revealed by Raman spectroscopy during retina inflammation, we quantified level of TNFα in culture supernatant from baseline to 48 hours after LPS stimulation by ELISA. We confirmed that TNFα increased after 12 h and remained elevated for the next 48 hours after LPS stimulation (Fig. 9.8A and 9.4III). RT-PCR analysis of iNOS levels revealed a prominent increase in iNOS expression 24 hours after LPS challenge (Fig. 9.8B). Finally, we assessed the levels of NAD / NADH in retina cultures by HPLC. Retinas challenged with LPS for 12 and 24 hours were compared with control retinas. We observed an initial increase of NAD and NADH after LPS challenge in retinal culture in accordance with Raman spectroscopy analysis (Fig. 9.8C). Moreover, the levels of glutamate, lactate and NAA, which were not significantly changed in the Raman spectroscopy analysis, remained also unaltered by HPLC quantification (data not shown).

9.4 Discussion

In this study we report the use of Raman spectroscopy of organotypic retinal cultures as a molecular imaging tool applicable to study of diseases of the central nervous system. Raman spectroscopy of the retina greatly enhances our ability to study these systems. The accessibility of the retina for laser imaging, and the sensitivity of Raman spectroscopy coupled with MCR-ALS algorithm for detecting and quantifying biomolecules, allowed a more detailed study of the temporal evolution of the molecules implicated in neuroinflammation. By combining biochemical spectra analysis, published Raman databases and multivariate analysis (MCR) we were able to characterize, monitor and quantify several molecules such as Lipoxygenase, iNOS, TNFα, Cytochrome C, Phenyalanine, NADH/NAD^+, and Phosphotidylcholine in the GCL of the retina. Moreover,

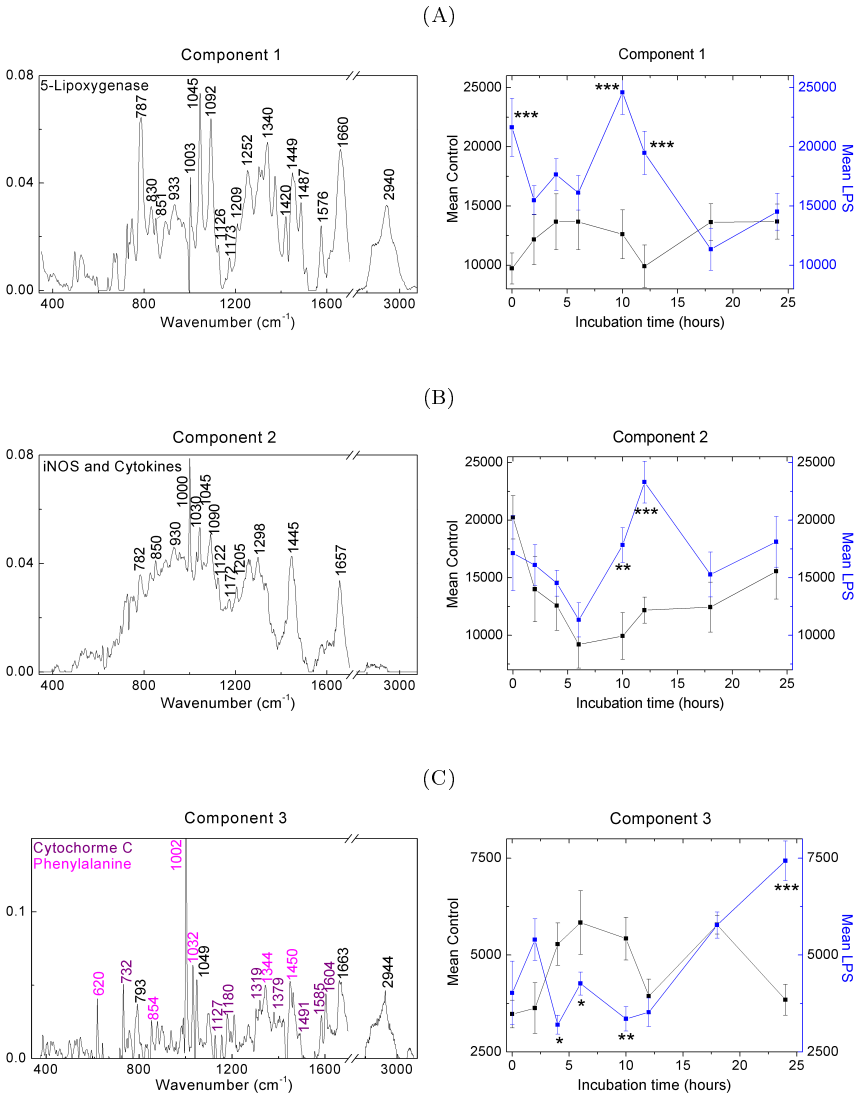


Figure 9.5: **Changes in molecular content in retina inflammation by Raman spectroscopy.** Molecular components that changed during retina inflammation were deconvolved from the Raman spectra using Multivariate Curve Resolution. The 3 first components identified are displayed (A-C): left: spectral profile deconvolved for each molecular component; right: Time course of the concentration profile for each molecular component after the LPS challenge. Curves for the mean of the control samples (black line) and the LPS treated samples (blue line) are compared. *p<0.05, **p<0.01 and ***p<0.001.

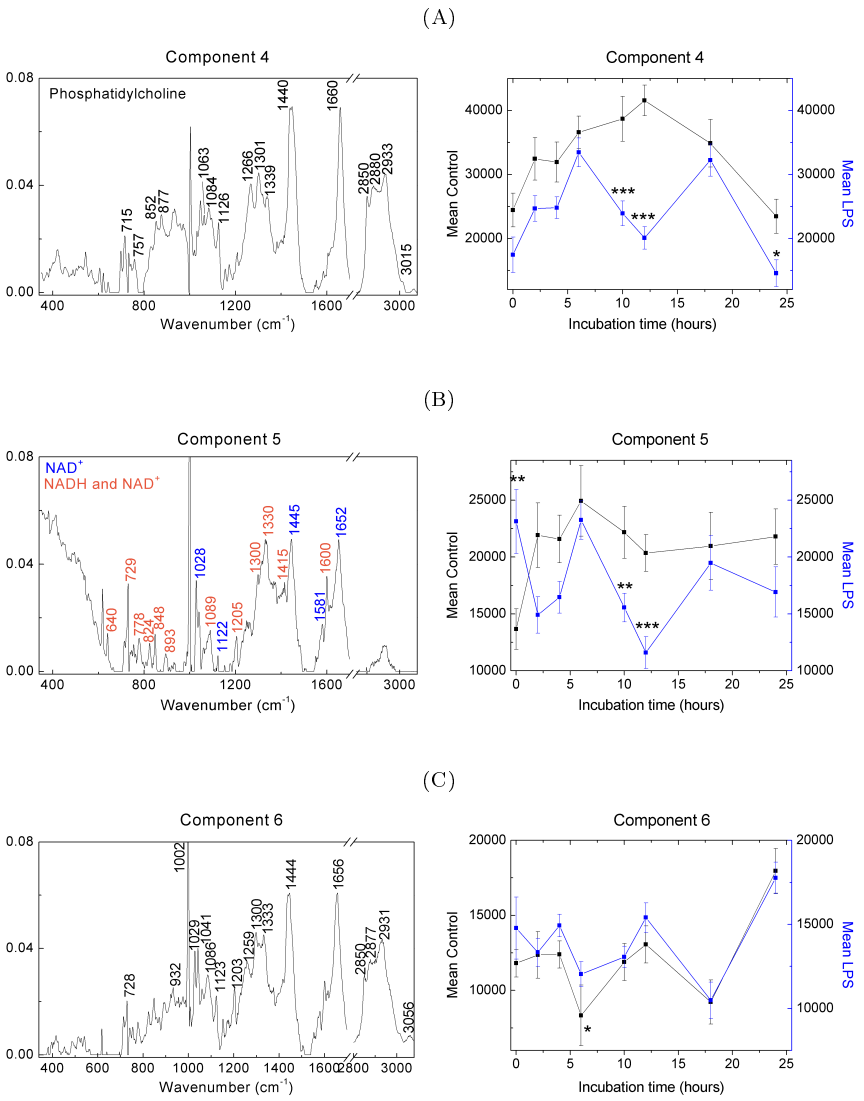


Figure 9.6: Continuation from previous figure. **Changes in molecular content in retina inflammation by Raman spectroscopy.** Molecular components that changed during retina inflammation were deconvolved from the Raman spectra using Multivariate Curve Resolution. The 3 remaining components identified are displayed (A-C): left: spectral profile deconvolved for each molecular component; right: Time course of the concentration profile for each molecular component after the LPS challenge. Curves for the mean of the control samples (black line) and the LPS treated samples (blue line) are compared. *p<0.05, **p<0.01 and ***p<0.001.

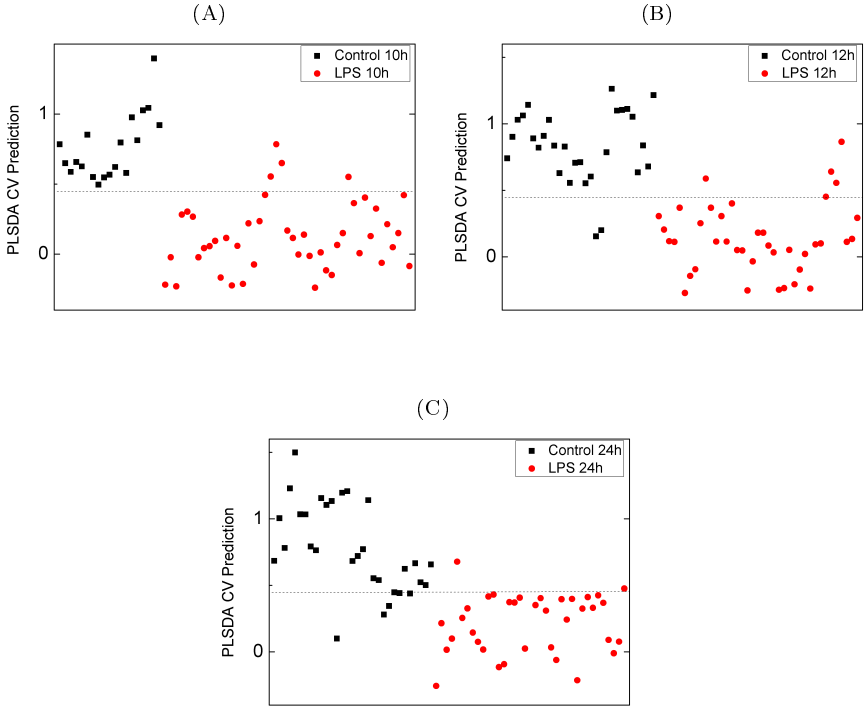


Figure 9.7: **Partial least square - discriminant analysis of Raman spectra in retina cultures in neuroinflammation.** Spectra from 25 control cultures and 25 cultures stimulated with LPS for 24h were analyzed by means of PLS-DA in order to classify samples as control (0) or inflamed (LPS treated) (1). Classification is shown for changes in the Raman spectra at 10 h (A), 12 h (B) and 24 h (C).

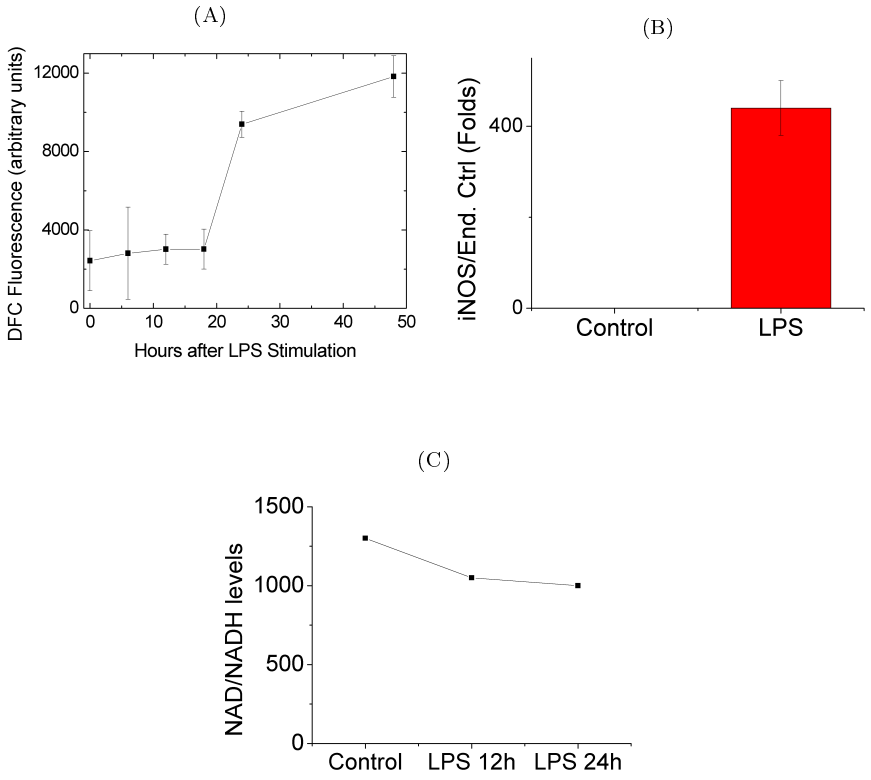


Figure 9.8: **Validation of molecular changes in neuroinflammation.** A) Induction of the pro-inflammatory cytokine TNFα during retina inflammation: Time-course analysis via ELISA of cytokine levels for TNFα in culture supernatant at hours 0, 6, 12, 18, 24, and 48 after addition of LPS. B) Quantitative analysis by qRT-PCR of iNOS expression in tissue cultures after 24h incubation with LPS versus control tissue. Y-axis represents the ratio between the level of iNOS mRNA and the level of endogenous control mRNA. C) Quantification of NAD and NADH by HPLC in retina cultures after LPS challenge: Graphs shows levels of NAD and NADH from a pool of 5 retinas for each time point after LPS stimulation and quantified by HPLC.

we provide evidence that quantification of such molecules reflects tissue damage related to retina inflammation. Finally, we developed mathematical methods that were able to predict the presence of retinal inflammation based on changes in the Raman spectra. Overall, our results suggest that Raman spectroscopy of the retina combined with MCR analysis can become a useful molecular imaging technique for the study and monitoring of retina and brain diseases.

The molecules deconvolved by MCR of Raman spectra of LPS treated retina are important players in the process of neuroinflammation in the CNS. The enzyme Lipoxygenase is associated with the production of leukotrienes, which are associated with inflammation of brain tissue mediated by the innate immune system [285]. In the cycle of neuroinflammation and oxidative stress in the central nervous system, iNOS plays a critical role exacerbating the hypoxic environment through overproduction of nitric oxide (NO). Levels of iNOS increased within the GCL at 12h, as revealed by Raman spectroscopy, and in the entire retina at 24 h as revealed by ELISA. Importantly the production of reactive oxygen species peaks at twelve hours, closely correlated to the peak in iNOS production detected by Raman. The production of the NO by iNOS has deleterious effects on the electron transport chain, and is a major cause of neuronal and axonal damage [271]. In addition to iNOS, the pro-inflammatory cytokine TNF-α is a key component of the signaling cascade associated with the inflammatory response in nervous tissue resulting in cellular damage [286].

In addition, Raman spectroscopic analysis of the inflamed retina revealed changes in critical mediators of mitochondria functioning and energy production. We observed a significant increase in Cytochrome C and Phenylalanine and a decrease in NAD/NADH. Oxidative stress can significantly interfere with the electron transport and mitochondrial function, promoting the export of Cytochrome C to the cytoplasm, which is an important trigger for cell death [287]. Also, the decrease in NAD/NADH levels confirms the failure of the electron transport chain. In addition, NAD seems to play a significant role in the control of the autoimmune attack against the brain in MS through the regulation of the enzyme indoleamine 2,3-dioxygenase (IDO), regulating T cell activation but also the interaction between neurons and microglia [288]. The anoxic cellular environment, in which cells must rely primarily on glycolysis to generate ATP, will increase NADH levels, while decreasing NAD^+ supplies. The increased levels of NO impair the electron transport chain, without which NAD^+ levels are unable to be maintained. Fluctuations of these metabolites over the 24h LPS challenge period reflect metabolic stress experienced by the retina tissue. Also, the role of Phenylalanine is complex, being a precursor of tryptophan and kynurenins and for this reason playing a critical role in the control of the endogenous regulation of neuronal excitability and the initiation of immune tolerance [289], as well as modulating the stability of myelin during neuroinflammation [290].

Raman spectra of the inflamed retina revealed decreased levels of fatty acids that could correlate with cell death and degradation within the GCL. The decrease in levels of Phosphatidylcholine in the GCL as a result of LPS challenge could be caused by cell death or thinning of the tissue and cell membrane disruption [291]. In addition, we observed that 24 to 48h after LPS challenge, the overall signal intensity decreased in the retina tissue, which was not attributable to any single component. Although this effect could be due to thinning of the GCL, changes in Muller cells in response to microglia activation may have had influence in light transmission [292]. Finally, we found no significant changes in

some of the candidate molecules such as lactate, glutamate and NAA neither by Raman spectroscopy or HPLC. The lack of changes of these molecules that are well documented in other settings could be related to the fact that concentration may change between cell compartments or retinal layers, but not in the overall concentration in the tissue.

Our study has several limitations. We have studied the changes induced by LPS in an in vitro model of retinal inflammation in which timing and location was well controlled. However, in humans changes in immune mediators may be more difficult to detect due to heterogeneity between subjects and the changes observed are the result of a chronic process instead of an acute challenge, making it more difficult to identify such patterns. Second, the specificity of the assignment of spectra to molecules (we required a 90% band coincidence) does not completely preclude the possibility that the signal identified may come from other molecules, in addition to the ones identified. For this reason, additional validations in other models of disease are warranted. Also, the identification of the molecules that significantly change in our study were based in the characterization of a set of candidate molecules and the information available in Raman databases, which is limited at present. For this reason, future research would improve the knowledge of the Raman spectra of biomolecules, allowing better identification of molecular changes by this technology. Finally, at present, knowledge of the specificity of Raman spectroscopy for distinguishing isoforms or posttranslational modifications of a given molecule is limited, which may be of great importance for several biological processes.

One of the main benefits of our approach is that Raman spectroscopy combined with MCR analysis is an unbiased screening technique that identifies temporal changes in any molecule present in the spectra, and for this reason may reveal previously uncharacterized molecules involved in a biological process. Furthermore it is a non-invasive technique and monitoring in living subjects is possible as compared with invasive methodologies as HPLC. The use of Raman spectroscopy to analyze findings in inflamed and control tissue culture samples has revealed several peaks of interest within the retinal spectra that can be attributable to changes in inflammatory mediators, components of the mitochondria and fatty acids during neuroinflammation. Further research is required to characterize the Raman band's behavior of these molecular components as a function of a real case of disease progress, and to incorporate their scanning potential into a diagnostic agent. It is clear from these findings that Raman spectroscopy combined with MCR presents the distinct possibility of non-invasively diagnosing and characterizing molecular levels quickly and accurately, and has the potential to drastically change how the process of neuroinflammation can be studied both in culture and in humans. In summary, we present a novel methodology for decomposing experimental Raman spectra from the retina into pure molecular components and monitor the molecular time evolution during retina inflammation, becoming a promising application to study retina diseases progress in vivo.

Table 9.1: Raman peaks and intensity for each molecule in solution. Strong (s), medium (m), and weak (w) designations assigned based on the highest, middle, and lowest third of the peak intensities of each individual spectra (very weak bands not listed). NAA: N-acetyl-aspartate; NADH: nicotinamide adenine dinucleotide; FAD: flavin adenine nucleotide (FAD), CML: N-carboxymethyl lysine; CEL: N-carboxyethyl lysine

Molecule	Buffer	Raman bands (cm^{-1})
Glutamate	PBS	344(w),531(m), 772(m), 787(m), 819(m), 857(s), 881(s), 935(s), 998(w), 1041(w), 1080(m), 1152(w), 1194(w), 1265(m), 1289(s), 1327(s), 1346(s), 1414(s), 1444(s), 1629(m), 2868(w), 2885(w), 2937(m)
Lactate	PBS	335(w), 422(w), 532(m), 642(w), 749(m), 828(s), 924(m), 1049(m), 1085(m), 1133(w), 1245(w), 1298(m), 1334(m), 1377(s), 1456(s), 1722(m), 2731(w), 2890(w), 2946(m), 2995(w)
NAA	PBS	349(s), 396(m), 542(w), 575(w), 649(m), 750(m), 7887(w), 842(s), 885(m), 901(m), 926(w), 949(s), 989(s), 1016(m), 1050(w), 1078(w), 1138(w), 1194(w), 1233(m), 1309(s), 1379(s), 1422(m), 1432(m), 1648(s), 1723(s), 2857(w), 2943(s)
NADH	TRIZ	390(m), 534(w), 635(w), 729(s), 756(m), 779(w), 832(m), 849(w), 882(w), 893(w), 941(m), 995(w), 1031(w), 1075(m), 1081(m), 1114(s), 1183(m), 1222(m), 1254(m), 1306(s), 1337(s), 1379(s), 1419(s), 1458(m), 1483(m), 1511(m), 1544(s), 1580(m) 1617(s), 1688(s), 2900(w), 2956(w)
Phosphatidyl-choline	PBS	374(w), 413(w), 449(w), 459(w), 613(w), 716(s), 766(m), 846(m), 873(m), 924(m), 954(m), 971(m), 1021(w), 1064(m), 1085(m), 1123(w), 1145(w), 1273(s), 1300(s), 1339(s), 1364(s), 1440(s), 1555(m), 1657(s), 1732(w), 2727(w), 2852(w), 2893(s), 2928(s), 3014(m), 3042(w)
FAD	PBS	531(w), 609(w), 680(w), 743(m), 789(w), 1067(w), 1158(w), 1164(w), 1181(w), 1226(s), 1255(m), 1280(m), 1351(s), 1408(s), 1463(m), 1500(m), 1546(m), 1581(s), 1629(s)
Cytochrome C	PBS	645(m), 741(s), 797(m), 811(m), 825(m), 852(w), 929(m), 975(m), 1002(m), 1126(w), 1561(s), 1582(s), 1636(s), 2883(w), 2934(m)
CML	PBS	810(s), 848(s), 929(m), 987(w), 1011(w), 1079(s), 1108(m), 1219(m), 1261(s), 1441(s), 2876(w), 2912(w), 2933(m), 2963(m), 2978(w), 2998(w)
CEL	PBS	332(s), 354(m), 483(w), 539(m), 554(m), 648(m), 733(w), 774(m), 810(m), 852(s), 895(w), 1043(m), 1063(m), 1070(m), 1093(m), 1166(w), 1198(w), 1327(s), 1352(s), 1372(s), 1410(s), 1444(s), 1464(s), 1632(s), 2878(w), 2944(m), 2970(w)
Pentosidine	PBS	406(m), 432(m), 487(w), 534(w), 548(w), 578(m), 600(w), 624(w), 651(w), 663(m), 670(m), 680(m), 727(m), 748(m), 777(w), 786(w), 844(m), 902(m), 1054(m), 1106(m), 1144(w), 1170(m), 1277(m), 1327(m), 1372(m), 1435(s), 1480(s), 1494(s), 1532(w), 1558(w), 1610(m), 1639(m), 1660(m), 2104(m), 2878(w), 2944(w)

10

Exploiting MCR-ALS algorithm to improve cytologic Raman spectroscopy studies. Unraveling the metabolic progression of cancer cells to undergo metastasis.

Raman spectroscopy is a very promising tool for molecular cytologic studies to reveal the underlying biochemical mechanisms that occurs in a cellular process but Raman spectroscopy has some drawbacks that need to be overcome to set this technique as a tool daily used in clinical diagnosis. Specifically, there is a need to improve the methodology to deconvolve from the spectra each pure molecular contribution when looking for sample differences. Also, the contribution of background signal in cell preparations as the glass window or the media can contribute Raman signal masking and degrading the statistical analysis. In this research we compare the strengths of Multivariate Curve Resolution (MCR) vesus PCA to deconvolve meaningful molecular components from a set of mix spectra. We exploit the capability of MCR algorithm to easily include initial estimates and constrains, demonstrating the ability of MCR to deconvolve from the spectra undesired background spectra in cytologic Raman studies and therefore obtaining the inherent biochemical information encoded in the cellular spectra. The use of MDA-MB 231 and MDA-MB 435 breast cancer metastatic cells illustrate the spectroscopic gaining insights which infer into the metabolic changes required to bone metastasis progression, including an increase in amino acids composition and decrease in mitochondria signal. We propose MCR as a very powerful tool for analyzing biological Raman spectra to reveal chemically and physically meaningful components that will provide new and unique insights to the biochemical model encoded in the molecular progression of cell

composition.

10.1 Introduction

Raman spectroscopy has been applied in the past decade to study biological samples with the help of statistical analysis. For instance, some techniques has been used to find groups and clusters in the Raman spectral data set such as PCA [293, 294], K-means clustering [295, 296] or Neural networks [297]. Also, Partial-Least-Squares Discriminant analysis (PLS-DA) allowed to construct discriminatory models to distinguish different types of samples and thus, it has been applied for instance to differentiate metastatic cells from the non-metastatic cells without basal-like phenotype [211]. Studying a dynamic process , 2D correlation was used to get insights into the correlation of Raman bands and the band evolution as broadening, shifts, etc. [298]. However, Multivariate techniques mention above do not provide the sufficient information to extract the molecular picture of the process being analysed and then, do not exploit all the specificity capabilities of Raman spectroscopy. There are some applications as the cytological studies of suspicious cancer cells in which Raman spectroscopy could be very useful to diagnose and study molecular differences among types of cells. But, some drawbacks needs to be overcome to fully exploit all the specificity capabilities of this technique. First, cells are grown on the surface of glass materials and therefore laser focus needs to be close or pass the surface window. Sometimes, part of the confocal volume analyzed covers a portion of the glass material which consequently produces undesired background signals in the Raman spectra. Second, it would be very useful to be able to use already prepared and stored patient samples used for parallel studies such as fluorescence but, sometimes these samples contain stains and are prepared in cover slips that will give a very high signal that will perturb the inherent cellular Raman spectra. This could be palliated by using expensive objectives as high numerical water immersion objectives directly on the liquid preparation, but risk of contamination of samples will exist and no already prepared samples could be used. Furthermore, background signals will be present. Bonnier et al. [299] tried to remove the effect of the background signals in the spectra by growing the cells in collagen matrices on the top of coverslips but this requires sample processing and modification of already established biomedical protocols. In this study a powerful statistical methodology is proposed to remove the undesired background signals and additionally be able to extract meaningful molecular components playing a role from the cellular Raman spectral analysis.

The method proposed is based on the use of the flexible and powerful algorithm Multivariate Curve Resolution- Alternating Least squares (MCR-ALS). MCR-ALS has been applied to several chemical problems [47]. Recently, it has been used in biological Raman spectroscopy as a decomposition algorithm to create Raman maps [52]. However, up to our knowledge it has not yet being used to recover and understand the molecular behavior of a biochemical process from Raman spectra of biological species such as tissue or cells. In this study we propose to exploit the flexibility of MCR-ALS algorithm to enhance and expand the strengths of Raman spectroscopy to be applied in daily clinical studies. Specifically, by making use of initial estimates and introducing physic-chemical constrains as the non-negativity of spectra and concentration,

pure and meaningful molecular components can be deconvolved from biological Raman spectra.

In the present work, MCR-ALS algorithm has been applied to a set of cytological Raman spectra that were adquired from an already prepared sample containing different cancer cell lines. Specifically, first, a comparison on the ability of PCA and MCR analysis applied to the same Raman spectral data has been tested to deconvolve pure components; second, a clever use of MCR-ALS algorithm has been applied to extract the quartz and water background signal from the spectral data and remove it from the Raman spectral data set.

This methodology has been used to get insights into the metabolic changes that breast cancer cells acquire to bone metastasis tropism. Recently, it has been reported the distinct growth and metastatic properties of two cell lines that share a closely-related genetic background with different orga-specificty in lungs or in bone. The different dynamics of hypoxia was secondary to organ-specific metastasis genes suggesting distinct dependence of hypoxia in promoting metastasis [300]. Moreover, we have recently showed that antioxidant proteins such as peroxiredoxins were specifically upregulated in lung metastatic breast cancer cells [242]. In that study, we used Raman microspectroscopy to study the chemical composition of metastatic variants to assess the importance of the metabolic pressure of the target organ to select their metabolic phenotype. These results suggested the glycolytic character of lung metastatic cells subtend by an increased REDOX. Under glucose deprivation, lung and bone metastatic cells were almost completely separated by a component, compatible with NADH.

In the present study, we exploit the use of MCR-ALS to study the Raman spectra of two cell lines having differences concerning their ability to grow in tissues with different oxygen concentration: MDA-MB-231, with tropism to metastasize in bone, and MDA-MB-435, with tropism to metastasize in lung. The study is complemented with the analysis of two variants of these primary cancer cell lines with increased tropism to grow in bone, Bo2 (derived from MDA-MB-231) and MDA-MB-435 bone (derived from MDA-MB-435). In our intention to find the metabolic characteristics necessary to produce bone metastasis and the mechanisms involved in this process, the cell lines set MDA-MB-435, MDA-MB-435 bone, MDA-MB-231 and Bo2 might indicate an increase in these metabolic characteristics. Finally, in a parallel and independent study we proof our approach to test in a separate data set the effect when the PRDX2 protein was over expressed in Bo2 cells. The presence of this protein in high amounts inside the Bo2 cells eliminates their ability to produce bone metastasis [242]. We could show with Raman spectroscopy coupled with MCR analysis that treated PRDX2 Bo2 cells return to a molecular content more similar to the primary 231 cell line.

A new approach has been proposed and tested to push forward the use of Raman spectroscopy in daily used samples as the preparation of cells in hospitals to extract chemically meaningful information about the molecular content of the biological samples and thus be able to build a biochemical model of the processes under study. We propose MCR-ALS algorithm as a powerful method that can be used in a clever and flexible way to reveal the underlying molecular information stored in Raman spectra. Specifically, the use of initial estimates and constrains applied to MCR-ALS algorithm allows to go an step further in the analysis of biological Raman spectra compared with the multivariate techniques already used.

10.2 Materials and Methods

10.2.1 Raman spectroscopy

For analysis $3x10^5$ cells were used. For measurements cells were seeded in 6 wells/plates (Becton Dickinson, NJ) over a quartz crystal (ESCO products, Oak Ridge, NJ). After 24h, cells were fixed with 4% cold paraformaldehyde (PFA) in PBS1x for 15 min, washed with PBS1x and maintained in the same solution at 4 °C until measurements. After the fixation, the quartz substrate containing the cells was mounted to a magnetic holder (Live Cell Instrument) with PBS and covered by other quartz subtrate for Raman analysis.

A total of 135 spectra were taken from the cytoplasm of cancer cells lines MDA-MB-231, Bo2, MDA-MB-435 and MDA-MB-435 Bone. For a second and independent study of PRDX2 role, a total of 90 spectra were obtained from the cytoplasm of MDA-MB-231, Bo2 and Bo2-PRDX2-treated cells.

The Raman system used was an inVia Renishaw (Apply Innovation, Gloucestershire, UK) and comprises a 514 nm laser that supplies an excitation beam of about 10 mW power, which is focused onto the sample via a microscope with 60x objective (Edmund, York, UK) and using a backscattered configuration. The Raman spectrum is recorded on a deep depletion charge-coupled device (CCD) detector (Renishaw RenCam). The recorded Raman spectrum is digitalized and displayed on a personal computer using Renishaw WiRE software which allows the experimental parameters to be set. The spectra were background subtracted with a custom-written Labview program using an stablish method [65].

10.2.2 Cell preparation

MDA-MB-435 cells supplied by Dr Fabra (IDIBELL) and their metastatic variant established from primary cultures of bone (435-B) metastases, maintained under standard conditions have been described elsewhere [301]. Although controversial, it has recently been demonstrated that MDA-MB- 435 cells are a useful breast cancer model and that they express both epithelial and melanocytic markers [302]. We used breast cancer bone metastatic cell lines MDA-MB-231 originally obtained from the European Type Culture Collection (ECACC 92020424. BO2 cell line has been established from bone metastases caused by MDA-MB-231 after six in vivo passages in nude mice using a heart injection model. Cells were cultured under standard conditions: grown in DMEM/F12 medium supplemented with 5% horse serum, 1 mM pyruvate, 2 mM L-glutamine in 5% CO2-95% air at 37 °C in a humidified incubator.

10.2.3 Statistical Analysis

First, an exploration of the spectral data set was performed by using PCA analysis. The number of components selected was the one in which the cumulative variance explained was more than the 99%. Two independent analyses (PCA and MCR) were performed in the same dataset which contained the spectra of the cell lines in rows. Additionally, to improve background component deconvolution five spectra from the PBS solution at the same focal distance than the cell measurements and far from the cells were added. For MCR-ALS algorithm, the protocol followed was: first, the initial estimates were selected

by using SIMPLISMA algorithm which selects the purest rows (spectra) in the data set; second, the iterative least squares calculation started with the addition of constrains: non-negativity of spectra and concentration matrices. Finally, the quality parameters from the MCR model were calculated: Lack of fit $LOF = 100\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} e_{i,j}^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} d_{i,j}^2}}$ and the parameter $R^2 = 100\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} d_{i,j}^2 - \sum_{i=1}^{n}\sum_{j=1}^{m} e_{i,j}^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} d_{i,j}^2}$ where $e_{i,j} = d_{i,j} - \hat{d}_{i,j}$ and $d_{i,j}$ is the original data matrix and $\hat{d}_{i,j}$ the calculated by means of MCR-ALS algorithm . For PCA and MCR analysis two different Matlab toolboxes were used: PLS toolbox (from Eigenvector Research) and MCR-ALS toolbox (from University of Barcelona).

10.3 Results and discussion

10.3.1 MCR is able to deconvolve more meaningful molecular components than Principal Component analysis on cell Raman spectra

A matrix containing all 135 spectra form cytoplasm of MDA-MB-231, MDA-MB-435 cells and their respective bone metastatic variants BO2 and 435B cells was constructed. In addition 5 more spectra of the PBS solution outside the cells were included in the last rows.

First, a PCA was performed on the spectral data matrix and the principal component loadings are plotted in Fig.10.1a. Second, MCR-ALS algorithm subjected to specific constrains (non-negativity of spectra and concentrations) was applied to the same spectral data set and the pure spectra deconvolved are plotted in Fig.10.1b.

From PCA analysis results shown in Fig. 10.1a, we observe that the signals of different substrates are mixed in the loading of principal components. The contributions of Quartz signal [299, 303] (mainly bands around 800 and 1050 cm^{-1}) was found in PC3 mixed with other cellular substrates and water signal. Indeed, the non-negativity of the Principal component loadings made difficult the assignment of each component to a pure molecular Raman spectra. In contrast, MCR decomposed four meaningful components (Fig. 10.1b) that can be assigned to: cellular lipids (component 1), Quartz signal (component 2), water signal [304] (component 3) and proteins and other cellular components (component 4). The variance explained by PCA and MCR models is: 99.74% and 99.66% respectively.

The higher ability of MCR to deconvolve pure signals relies on the flexibility of MCR-ALS algorithm. First, a clever construction of the spectral data matrix was performed by displaying all cell Raman spectra in rows and adding five more rows containing the signal from PBS acquired at the same focal distance than cells but outside of them. Second, the use of SIMPLISMA algorithm to find the purest spectra in the data set allowed giving initial estimates of the MCR-ALS algorithm. As a result the four most dissimilar spectra found were: two signals from PBS solution and two of different cell lines. Once the initial estimates were selected, the iterative least square calculation was initialized by giving some constrains to the algorithm as non-negativity in the output spectra and

concentrations. All this summed gave as a result four chemically and physically meaningful components.

Compared with MCR, PCA finds the directions in the data set with maximum variance and orthogonal between them, but no additional chemical information can be added. Consequently, in the case of complicated biological Raman datasets, PCA could be used for an exploratory analysis of the data set to find for instance outliers or groups but no pure molecular components can be extracted from it. This problem was mentioned in previous works for cellular studies [214] and for image segmentation [35].

10.3.2 Subtraction of the background signals by MCR allowed disentangling the inherent composition differences between cells

In order to clean and remove those signals from the original matrix spectral data set, a direct subtraction of the quartz and water signals (MCR loadings) multiplied by the concentration matrix (MCR scores) of those components for each original spectra was performed (Fig. 10.2). This induced more visibility of some spectral changes inherent to the cellular spectra, making easier to build a statistical model with chemically meaningful outputs. The results are shown in the plots of spectra before (Fig. 10.2A) and after (Fig. 10.2B) the subtraction of Quartz and water signals . Furthermore, by using this method, statistical analysis of cell Raman spectra is improved because is not affected by background contributions that may interfere the results. Thus, based on the clean cell Raman spectra, a new MCR model was calculated to deconvolve the main molecular components that are changing between cells. Recently, some applications of MCR to Raman images, detected the contribution of background signals in the components deconvolved [35, 226]. However, up to our knowledge this is the first time that MCR-ALS algorithm is proposed as a tool to subtract the background signal contributions in the Raman analysis of cells when using standard sample preparation. For clinical relevancy, and acceptance of the Raman spectroscopy by the clinical community, it is important that the sample preparation protocols necessary for Raman spectroscopy are consistent with current practice. Nowadays this is a challenging issue and researchers are working to overcome this limitation. For instance, a recent study compared Raman signals of unprocessed tissue preparations and dewaxed tissue sections, demonstrating that tissue processing has significant impact in the extracellular structure, in particular in the lipidic structure [305]. It will be very convenient if Raman spectroscopy could be used using the conventional sample preparation without the need of further processing to prevent the loose of more cellular information. In this way it would be necessary to deconvolve from the spectra the contribution of signals non-inherent of cells that are present due to chemicals used in the conventional cytological preparation. In this way, parallel and dialy used techniques could be performed in the same samples in hospital laboratories. The use of stored bank samples would be possible permitting studies on prognosis and evolution of patients.

It is important to notice that from a matrix with very similar cell Raman spectra, with MCR-ALS algorithm, three meaningful molecular components were deconvolved giving key information about the biochemical changes occur-

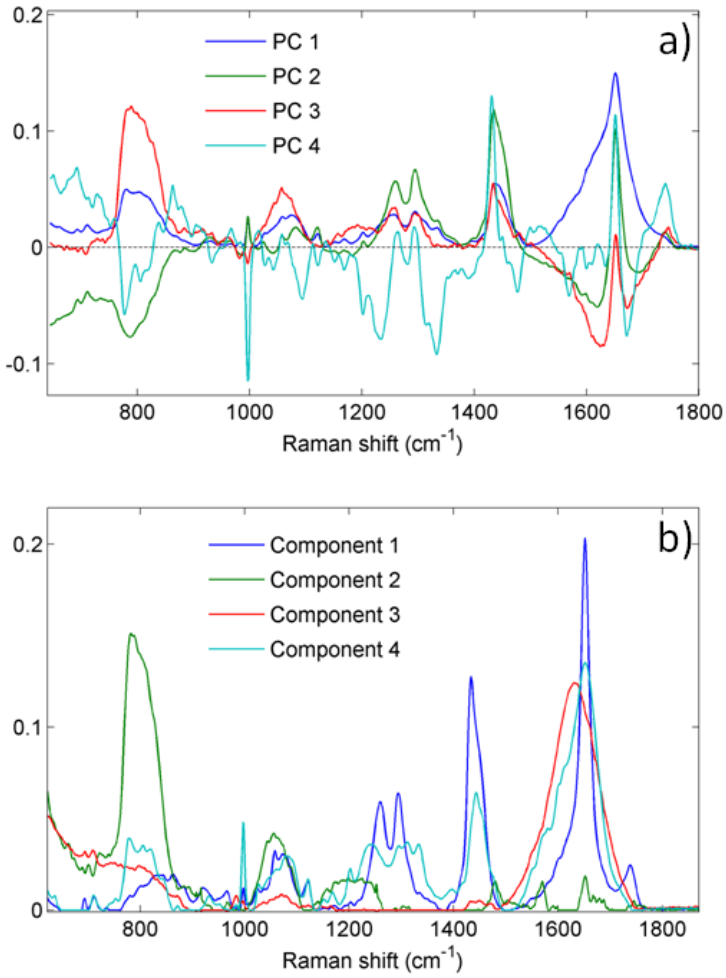


Figure 10.1: a) PCA loadings and b) MCR-ALS loadings of the matrix spectral data set containing the Raman spectra of cytoplasm of cancer cells lines 231, Bo2, 435 and 435Bone and PBS.

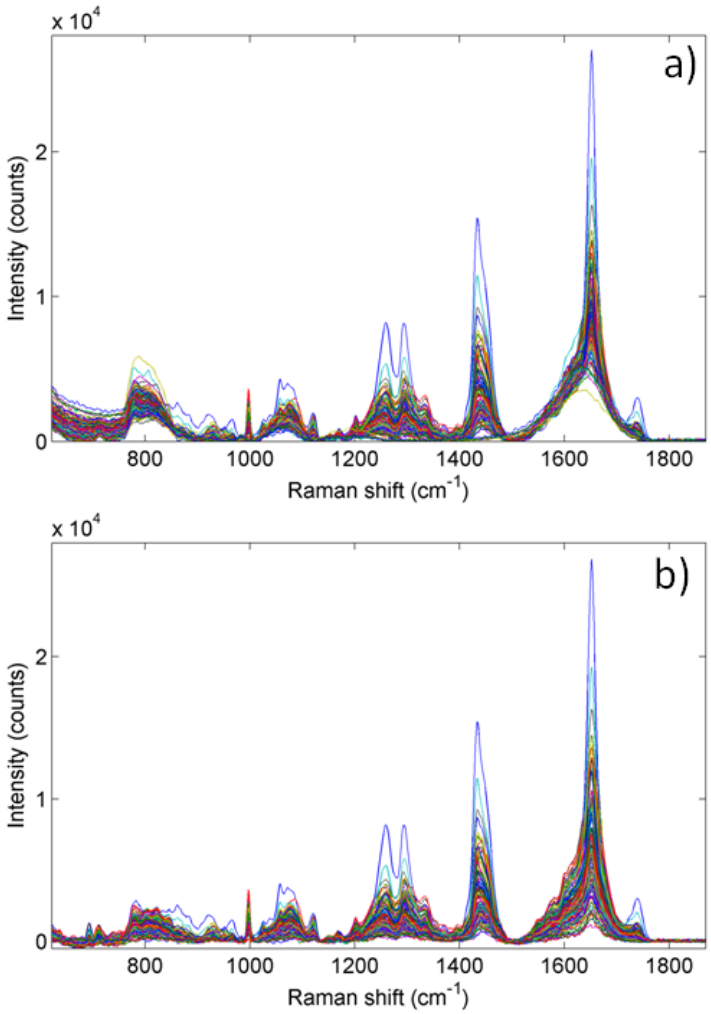


Figure 10.2: Original Raman spectra (a) and Raman spectra after the subtraction of Quartz and water signals (b).

ring in metastatic cancer cells. Indeed, spectra deconvolved with MCR-ALS algorithm was assigned to an almost pure molecule and therefore gave more biochemical information than other commonly used techniques as PCA or cluster analysis could be extracted (Fig. 10.3). In particular component 1 can be assigned to Phosphatidylserine [220]. Raman bands in component 2 have been reported to be related to aminoacids [306, 307]. Finally, component 3 contain bands assigned to mitochondria and cytochrome C [308–316]. This last component also contain bands related with DNA or nucleotides [189]. The fact that mitochondria is the organelle of the cytoplasm that contain the higher concentrations of DNA and Cytochrome C, leads to the conclusion that component 3 can be a measure of the amount of mitochondria present in the cell under study. Table 10.1 shows the main assignments of Raman bands present in each component.

Table 10.1: Band assignments for the Raman spectra corresponding to components 1, 2, and 3 obtained by MCR analysis of MDA-MB-231, Bo2, MDA-MB-435P and 435B cell lines after background signal subtraction. Assignments based on references [189, 220, 306–316]. PA: Phenylalanine, PhS: Phosphatidylserine, T: Tryptophan, AA: aminoacid.

Component 1		Component 2		Component 3	
Raman shift (cm^{-1})	Assignm.	Raman shift (cm^{-1})	Assignm.	Raman shift (cm^{-1})	Assignm.
836	PhS	693	AA, methionine	750	Cytochrome C
1057	PhS	710	AA, methionine	777	Cytosine, Tymine
1075	PhS	822	C-Cstretch	804	Phosphoric acid
1119	PhS	997	PA	996	PA
1260	PhS	1026	PA	1082	Phospholipids, nucleic acid
1295	PhS	1082	Glutamic acid	1120	CytochromeC
1442	PhS	1122	C-N proteins	1203	Nucleic acid
1660	PhS	1201	T, aromatic C-N, AmideIII	1240	RNA
		1263	AmideIII	1312	CytochromeC
		1293	Alanine	1333	Adenine, Guanine
		1336	C-H, AmideIII, T, Glycine	1445	
		1442	CH_2	1570	CytochromeC, mitochondria
		1654	AmideI	1585	CytochromeC, mitochondria

In the past, a gradual increase in the amino acid Raman bands has been observed in bladder cancer at different stages [317]. Also, the increase in 790 and 820cm^{-1} bands were observed. The amount of amino acids like Alanine

is expected to increase during hypoxia, made by transamination of pyruvate to prevent further increases in lactate [1]

10.3.3 Molecular components deconvolved disentangle key metabolic changes to undergo metastasis in bone.

To illustrate the efficient use of MCR in Raman spectroscopy analysis, we obtained the spectra of MDA-MB-231, MDA-MB-435 cells and their respective bone metastatic variants BO2 and 435B cells to assess the differences between them (Fig. 10.3). The concentrations of each MCR component for each cell measuredwere ploted (Fig. 10.3 right) with its corresponding spectra (Fig. 10.3 left). Component 1 referred to phosphatidylserine and was inversely proportional to bone metastasis ability, decreasing its value as the bone tropism increased (had the lower level in BO2 cells). Moreover, we distinguished two populations inside the MDA-MB-435 cell line, one with lower phosphatidylserine content that clearly is enriched in the 435B bone metastatic variant. The clear differences between MDA-MB 231 and MDA-MB 435 cells might be due to the lung metastatic background which characterized MDA-MB 435 cells in contrast to MDA-MB 231 ones. This result suggested that bone metastatic ability is related with low lipid content in the cells (measured as phosphatidylserine).

A decrease in component 2, which referred to aminoacid content, opposite to component 1, is related to a decreased bone metastasis tropism with the minimum level in MDA-MB-435 cells. Component 3 referred to mitochondrial content, which is inversely related to bone metastatic ability (similar to phosphatidylserine content) and was found increased in MDA-MB-435 cells.

It has recently been reported that mitochondrial glucose oxidation may be incompatible with the survival of some cancer cell [318]. Moreover, the balance mechanism of the antioxidant system might prevent collapse by redirecting the glycolytic flux into the pentose phosphate pathway [319]. Indeed, B02 cells displayed low glucose dependence suggesting a metabolic shift from glucose oxidation to fatty acid oxidation to resist metabolic insults in a hypoxic and hypoglycemic tissue like bones. Then, the low metastatic burden of B02/PRDX2 cells in bones suggests that PRDX2 might interfere in the preferential bone metastatic cells metabolism with deleterious consequences to carcinoma cells that try to adapt to the bone microenvironment [242]. Indeed, the mechanism to cope with the cumulative reactive oxygen species is different in lung than in bone tissues, where the lowest oxygen gradient in bone marrow induces fewer metabolic free radical challenges [320].

Mitochondria are organelles that provide the majority of the energy in most cells because of their synthesis of ATP by oxidative phosphorylation. They also have other roles including a contribution to intracellular calcium homeostasis [321], and are critical for many cellular functions including growth, division, energy metabolism and apoptosis in cells. Mitocrondrial dynamics have an important role in breast cancer cell migration and invasion [322]. Fission facilitates and fusion inhibits these processes, Drp1, a protein controlling mitochondrial fission, is an early event in development of metastatic breast cancer [323].

All together our results clearly indicated that a metabolic switch in breast cancer cells might be useful in order to fit with the bone microenvironment increasing their bone metastatic ability. Since no significant increase in lipid

metabolism was found in bone metastatic cells, these results strongly suggested that metabolic changes fitting with bone requirements are related to mitochondrial activity. Moreover, the increase aminoacid levels suggested that they might be relatd with alternative metabolic sources.

To check this hypothesis we used a mitotraker to visualized mitochondrial concentration in breast cancer metastatic variants. In addition we performed an independent experiment that resulted in a matrix containing the Raman spectra of MDA-MB-231 and BO2 cells including PRDX2BO2 cells. PRDX2 influences oxidative and metabolic stress through multiple mechanisms [324]. We have recently reported that PRDX2 knockdown dramatically decreased lung metastasis formation, whereas overexpressed in BO2 cells decreased its bone metastatic ability, suggesting that PRDX2 is a 'metabolic adaptor', which functions stabilizing the redox state required for cell survival in an oxidative atmosphere [242].

A new MCR was built after subtracting Quartz and water signals to confirm the molecular behaviour of the different cell lines and specifically the effect of redox state disruption by PRDX2 in bone metastatic cells.

Similar components than previous MCR analysis (Fig. 10.3) were deconvolved (Fig. 10.3). MCR analysis revealed that PRDX2 increased phosphatidylserine in BO2 cells as was reflected in Component 1 levels (Fig. 10.4A) according to the above observation (Fig. 10.3A) in which the lipid content of cells was inversely related with the bone metastasis potential. Moreover, Component 2 decreased in PRXD2BO2 cells with regard to BO2 cells, then aminoacid content was associated with higher bone tropism. In addition, mitochondrial content represented in Component 3 previously inversely associated to bone tropism, was also increased (Fig 10.4C). These results validated confidently the biological use of MCR applications, confirming our hypothesis that bone metastasis production needs biochemical changes, which are reversed with the PRDX2 over expression in the metastatic Bo2 cells.This is probably because they recover features of their non-hypoxic metabolism and loose the ability to produce metastasis in an hypoxic environment as bone. Mitochondria is well-known for its role in cell respiration [325] and it may play a role in the necessary adaptation of primary tumor cells to proliferate in hypoxic environments as bone tissue and produce metastasis.

As a conclusion making use of the flexible and powerful algorithm MCR-ALS and applying it to cell Raman spectra we were able to recover and extract useful and other-wise inaccessible information about molecular changes occurring in cellular processes. In order to take full advantage of the specificity potential of Raman spectroscopy, it is necessary to expand the Raman spectral database of biological molecules related to biochemical activity in cells and tissue. In this way, Raman spectroscopy could substitute expensive, time consuming and invasive techniques such as HPLC.

In the future other initial estimates could be explore to guide MCR-ALS algorithm to recover contributions of molecules that a priori are known to have a role in the process under study. By introducing the Raman spectra of those pure molecules as initial estimates, the contribution of those molecules in the processes could be explored.

10.4 Conclusion

We propose a new methodology to extract the useful molecular information encoded in Raman spectra of biological samples. By using the flexible MCR-ALS algorithm meaningful molecular components could be extracted thanks to the previously subtraction of background signals that perturbed the inherent cell Raman spectra. This methodology opens up new possibilities for the clinical use of Raman spectroscopy in hospital diagnosis because permit to remove signals from substrates or chemicals used in the cytologic preparation. It also provides a rapid, reliable and label free method to disentangle the biochemical components involved in the differentiation of cells with different ability to metastize in different organs. It provides a new tool to give a prognosis about the probability of a given group of cells to produce metastasis in an organ and prevent it in advance. We apply this methodology to study the molecular differences between primary breast cancer cell lines and their bone metastatic cell lines. We could identify biochemically meaningful molecular components responsible for the changes in their metastatic ability and related to the necessity of bone cancer cells to adapt to a more hypoxic environment. Specifically, mitochondrial levels decrease in bone metastatic cell variants (Bo2) with respect with their correspondent primary cell lines (231) but PRXD2 Bo2 treated cells recover the same levels as 231 cells suggesting that PRXD2 over-expression suppress the bone metastatic ability and thus it gives a good clinical treatment for patients undergoing bone metastasis from a primary breast cancer tumor.

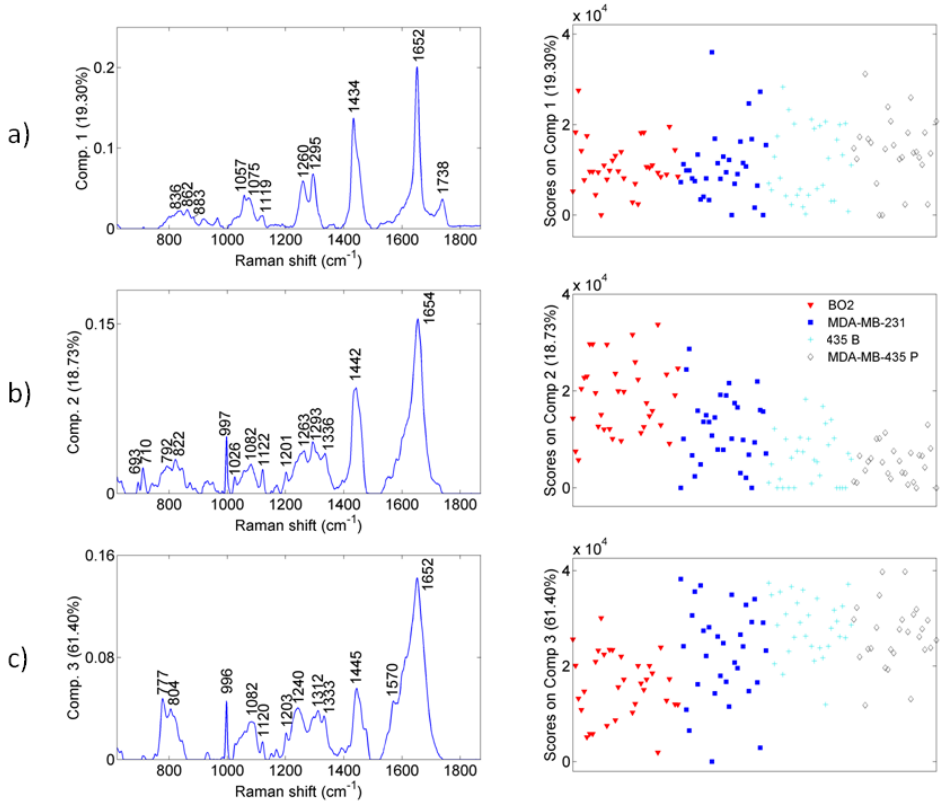


Figure 10.3: MCR spectra analysis from cell Raman spectra of MDA-MB-231, Bo2, MDA-MB-435P and 435B cell lines after the subtraction of Quartz and water signals. Three components were deconvolved (a, b and c) and its respective MCR loadings (left) and scores (right) are plotted. Red triangles: Bo2, blue squares: MDA-MB-231, blue crosses: 435B and white diamonds: MDA-MB-435. The variance explained for each component is written in %.

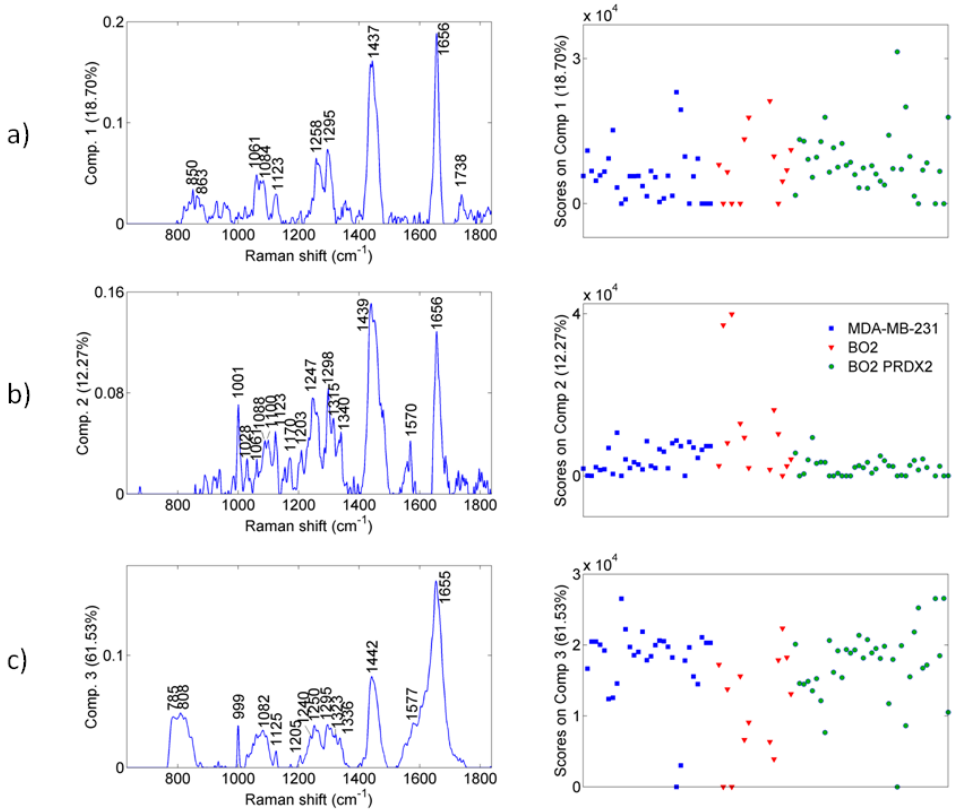


Figure 10.4: MCR analysis of Bo2, MDA-MB-231 and treated Bo2-PRDX2 cell lines. Three components were deconvolved (a, b and c) and its respective MCR loadings (left) and scores (right) are plotted. Red triangles: Bo2, blue squares: MDA-MB-231, green stars: Bo2-PRDX2. The variance explained for each component is written in %.

Conclusion

Since its discovery in 1928, Raman spectroscopy has produced a revolution in the areas of analytical chemistry and molecular detection. Thanks to the latest technical advances, such as sensitive spectrometers and NIR or UV lasers, the expectations of the applicability of Raman spectroscopy in biology have increased. The possibility of detecting and monitoring the evolution of the biochemical content in biomedical samples non-invasively and with high specificity has become a vision to be pursued. However, the inherent properties of Raman scattering have inhibited its full exploitation for biomedical applications. In the past decade, SERS and multivariate analysis have emerged as possible solutions for overcoming the low efficiency and the complexity of the Raman signals obtained from biological material. Until 2009, only a few studies had been reported using multivariate approaches, and these techniques were only employed to group different types of samples. This meant that the rich information contained in Raman spectra was not being fully exploited. Also, although the SERS effect was demonstrated for cells, SERS probes were not used in their full strength to study complex biological processes inside cells. This thesis is a step towards combining and using SERS and multivariate analysis to expand the applicability of RS in biomedicine. Thus, the virtues of RS such as its high specificity can be used to diagnose and study biomedical samples non-invasively, rapidly, and without special sample preparation. New insights otherwise inaccessible could be revealed by this methodology, insights that the routinely employed biochemical techniques in labs or hospitals can not provide.

Specifically, the SERS spectra of a single DNA molecule under stretching and under physiological conditions was studied by using 2D correlation spectroscopy. Some important features in the spectra were revealed, thanks to this technique, such as a shift in the O-P-O Raman band. In this study, optical trapping and surface-enhanced Raman scattering (SERS) were combined to establish a direct relationship between the DNA's extension and its structure in the low force, entropic regime. A DNA molecule was trapped close to a SERS substrate to enhance detectable Raman signals. With the help of 2D correlation, it was possible to observe that the DNA Raman modes shifted in response to an applied force, indicating the occurrence of phosphodiester mechanical alterations.

Also, another study of the effect of a mechanical force on a living cell is discussed, one applied to RBC stretching. The band dynamics along different force loads are investigated with principal component analysis (PCA) and 2D correlation techniques. The results reveal important structural properties of the RBC cells related with their function in the body.

A complex experiment was performed to study the intracellular pH changes in glioma cells after Photodynamic Treatment (PDT) by using SERS probes embedded in the cells. The evolution in the SERS spectra caused by the changes in the intracellular pH was studied using 2D correlation techniques. A complete picture of the Raman band's behavior was achieved, that permits selecting the Raman band most capable of tracking the pH changes. To the best of our knowledge, this study represents the first use of the 2D correlation technique to study biological SERS spectra.

Furthermore, more complex systems were investigated, such as the molecular evolution of cells or tissues during a biochemical process. A complete understanding of the behavior of the biochemical content was achieved, having very limited or no a priori information about the molecules involved or their changes in concentration during the process. Pursuing this goal, PCA was used to study the lipid metabolism in different breast cancer cell lines depending on the degree of malignancy in the CH stretching region of the Raman spectra. A classificatory model was generated which differentiated between the metastatic cells from the non-metastatic cells without basal-like phenotype. This represents a first step towards stratifying breast cancer cells using RS. However, PCA does not provides meaningful components that could be assigned directly to molecular Raman spectra spectra. Consequently, a very powerful multivariate analysis (MCR) was proposed and applied to extract physically and chemically meaningful molecular components that changed in cancer cell composition during Epithelial to Mesenchymal transition (EMT). The analysis of the Raman spectra in the fingerprint region coupled with MCR led to the conclusion that the EMT process affects the lipid profile of the cells, increasing the tryptophan but maintaining a low phosphatidylserine content in comparison with highly metastatic cells. We demonstrated, then, that Raman microspectroscopy coupled with MCR enables deconvolution and tracking of the molecular content of cancer cells during a biochemical process, and is a powerful non-invasive tool for identifying the metabolic features of breast cancer cell aggressiveness.

Also, we monitored the retina composition ex vivo when a neuroinflammation is induced. Meaningful molecular Raman spectra could be deconvolved from the spectra acquired at different time points by using the MCR algorithm. By assigning each spectrum to a molecular component related with the neuroinflammation, new, complete, and otherwise inaccessible information about the biochemistry behind the process of neuroinflammation could be extracted. Our study represents the first application of MCR to decompose and monitor the molecular content of a biological tissue with RS. Biomarkers for the early detection of neuroinflammation processes were identified, and this methodology represents the first step towards the establishment of a non-invasive and rapid screening technique for the early detection of multiple sclerosis or other neurodegenerative diseases in patients.

Finally, the MCR-ALS algorithm is very flexible, and we exploited the use of initial constrains or estimates to provide a powerful method to solve crucial problems not yet overcome in the use of RS in biology, such as the presence of background signals that mask and degrade the results of a statistical analysis. The removal from cellular spectra of the background signals not intrinsic to the cells was achieved by employing the MCR-ALS algorithm. Cytological preparations of different cancer cell lines were studied and, thanks to the removal of the background signals, the molecular components could be identified that have an important role in the progression of breast cancer cells towards bone metastasis.

To summarize, the main results that I have achieved in my thesis are:

- Extracting from Raman spectra information on the molecular components of cells and tissue and monitoring the evolution of their concentrations during a biochemical process.
- Studying the Raman band's behavior and correlations by means of 2D correlation analysis when a single DNA molecule or a living cell is stretched

or experiences a change in one of the parameters in its environment. This permitted revealing important features in the molecular structure of the biosamples studied.

- Combining 2D correlation and SERS of biological samples for the first time to analyze the dynamic changes of the pH in living cells after being treated with PDT.
- Identifying and tracking the concentration of biomarkers in cancer cells during the EMT transition. The proposed method of using Raman spectroscopy supported by MCR provides a powerful non-invasive tool for identifying the metabolic features of breast cancer cell aggressiveness.
- A method exploiting the use of constraints with the MCR-ALS algorithm, representing a new step towards the use of cheap and conventional cytological sample preparation for screening samples with Raman spectroscopy in hospitals. The ability to remove the background signals non inherent to the cells permitted the use of glassware material or the chemicals normally used in cytological preparations that could interfere with the Raman signals.
- Combining and comparing different mathematical methods to study a very broad range of biomedical applications. For each need, a different method was proposed.
- Presenting a novel methodology for monitoring the evolution of different molecular components in the retina during an inflammation by using Raman spectroscopy coupled to MCR. This is a promising tool that could be used to study the human retina and diseases of the central nervous system in vivo.

This research has demonstrated a powerful method that adds a new dimension to the field of analytical chemistry. Sensitive and highly specific information can be extracted non-invasively, rapidly, and without sample preparation. The samples can be monitored in vivo, quantifying molecular components difficult or impossible to obtain with today's technology.

A

2D correlation written code

```
1
function[syn,asyn] = todcosp2(rs,t)
3 % Input vectors:
% rs(raw spectra, the spectra for each time is placed in columns
    where each column represents one time step)
5 % size(rs)=n,m (we have n wavelenght measured points)
% t: vector with each time point, size(t)=m,1
7 %
%Reference spectrum: is the first spectra
9
[n m]=size(rs);
11 t0=2*t(1)-t(2);
tm1=2*t(m)-t(m-1);
13 syn=zeros(n,n);
asyn=zeros(n,n);
15 s=zeros(n,m);
for i=1:m
17 s(:,i)=rs(:,i)-rs(:,1);
end
19 for i = 1:n
    for ii = 1:n
21 asyn(i,ii) = asyn(i,ii) + 0;
        % 1 , 1
23 for k = 2:m-1
            asyn(i,ii) = asyn(i,ii) + s(i,1)*s(ii,k)*(t(k+1)-t(k-1)
                )*(t(2)-t0)/(t(k)-t(1));
25 % 1 , 2:m-1
        end;
27 asyn(i,ii) = asyn(i,ii) + s(i,1)*s(ii,m)*(tm1-t(m-1))*(t(2)
            -t0)/(t(m)-t(1));
        % 1 , m
29 for j = 2:m-1
            syn(i,ii)=syn(i,ii)+s(i,j)*s(ii,j)*(t(j+1)-t(j-1));
31 asyn(i,ii) = asyn(i,ii) + s(i,j)*s(ii,1)*(t(2)-t0)*(t(j
                +1)-t(j-1))/(t(1)-t(j));
            % 2:m-1 , 1
33 for k = 2:m-1
```

```
                    if j==k
35                      asyn(i,ii) = asyn(i,ii) + 0;
                        % 2:m-1 , 2:m-1 (j==k)
37                  else
                        asyn(i,ii) = asyn(i,ii) + s(i,j)*s(ii,k)*(t(k
                            +1)-t(k-1))*(t(j+1)-t(j-1))/(t(k)-t(j));
39                      % 2:m-1 , 2:m-1 (j!=k)
                    end;
41              end;
                asyn(i,ii) = asyn(i,ii) + s(i,j)*s(ii,m)*(tm1-t(m-1))*(
                    t(j+1)-t(j-1))/(t(m)-t(j));
43              % 2:m-1 , m
            end;
45          asyn(i,ii) = asyn(i,ii) + s(i,m)*s(ii,1)*(t(2)-t0)*(tm1-t(m
                -1))/(t(1)-t(m));
            % m , 1
47          for k = 2:m-1
                asyn(i,ii) = asyn(i,ii) + s(i,m)*s(ii,k)*(t(k+1)-t(k-1)
                    )*(tm1-t(m-1))/(t(k)-t(m));
49              % m , 2:m-1
            end;
51          asyn(i,ii) = asyn(i,ii) + 0;
            % m , m
53      syn(i,ii)=syn(i,ii)+s(i,1)*s(ii,1)*(t(2)-t0)+s(i,m)*s(ii,m)*(
            tm1-t(m-1));

55      end;
end;
57  syn=syn/(2*(t(m)-t(1)));
    asyn = asyn/(4*pi*(t(m)-t(1)));
```

B

List of publications and contributions in conferences of Monica Marro

Refereed research papers

1. S. Rao, S. Raj, B. Cossins, M. Marro, V. Guallar, and D. Petrov. Direct observation of single DNA structural alterations at low forces with surface-enhanced Raman scattering *Biophys. J.* 104:156-162, 2013

2. C. Nieva, M. Marro, N. Santana-Codina, S. Rao, D. Petrov, and A. Sierra. The lipid phenotype of breast cancer cells characterized by raman microspectroscopy: towards a stratification of malignancy. *PLoS ONE* 7, e46456, 2012

3. S. Raj, M. Marro, M. Wojdyla, and D. Petrov. Mechanochemistry of single red blood cells monitored using Raman tweezers. *Biomed. Opt. Express* 3: 753-763 2012

4. V. Stresing, E. Baltziskueta, N. Rubio, J. Blanco, MC. Arriba, J. Valls, M. Janier, P. Clezardin, R. Sanz-Pamplona, C. Nieva, M. Marro, D. Petrov and A. Sierra. Peroxiredoxin 2 specifically regulates the oxidative and metabolic stress response of human metastatic breast cancer cells in lungs *Oncogene* online DOI: 10.1038/onc.2012.93, 2012

5. S. Balint, S. Rao, M. Marro, P. Miskovsky, and D. Petrov. Monitoring of local pH in photodynamic therapy treated live cancer cells using surface-enhanced Raman scattering probes. *J. Raman Spectrosc.* 42: 1215-1221, 2011.

6. S. Balint, S. Rao, M. Marro, V. Huntosova, P. Miskovsky, and D. Petrov. Diffusion and cellular uptake of drugs in live cellsstudied with surface-enhanced Raman scattering probes. *J. Biomed. Opt.* 15, 027005, 2010.

Submitted research papers

1. M. Marro, A. Taubes, A. Abermathy, S. Balint, B. Moreno, B. Fernandez-Diez, EH. Martinez-Lapiscina, I. Amat-Roldan, D. Petrov, and P. Villoslada. Raman spectroscopy reveals mitochondria stress and phosphatidylcholine decrease in retina inflammation. *J. Biophot.*

2. M. Marro, C. Nieva, R. Sanz-Pamplona, D. Petrov and A Sierra. Raman spectroscopy supported by Multivariate Curve resolution improves biochemical and metabolic analysis of cancer cells. *Carcinogenesis*

3. M. Marro, C. Nieva, A. Sierra and D. Petrov. Exploiting MCR capabilities for improving metabolic cytological studies using Raman spectroscopy. *J. Biomed. Opt.*

Proceedings papers

1. M. Marro, A. Taubes, P. Villoslada, and D. Petrov. Detection of neuroininfammation through the retina by means of Raman spectroscopy and multivariate analysis. Proc. SPIE 8427, Biophotonics: Photonic Solutions for Better Health Care III, 842715 (June 1, 2012); doi:10.1117/12.921358

2. M. Marro, C. Nieva, R. Sanz-Pamplona, N. Santana-Codina, S. Rao, D. Petrov, A. Sierra. Raman spectroscopy analysis of the MCF10A malignant transformation. Proc. SPIE, Biophotonics: Photonic Solutions for Better Health Care III, (May 2012); doi: 10.1117/12.921356

3. S. Raj, S. Rao, M. Marro, M. Wojdyla, and D. Petrov. Load bearing studies of single DNA molecules and red blood cells using optical tweezers and Raman spectroscopy. Proc. SPIE 8427, Biophotonics: Photonic Solutions for Better Health Care III, 842719 (June 1, 2012); doi:10.1117/12.921357

4. M. Marro, S. Raj, S. Rao, A. Taubes, P. Villoslada, D. Petrov. Using 2D correlation and multivariate analysis combined with plasmonic effects to expand the use of Raman microspectroscopy in biomedical applications. Paper 8087-51 from the proceedings of SPIE/OSA Biomedical Optics Conference in Munich, Germany May 22-26, 2011.

5. M. Marro, S. Raj, S. Rao, D. Petrov. Extending the applicability of Raman microspectroscopy in biomedicine using statistical analysis and plasmonic effects. IEEE 978-1-4244-9837-6/11. 2011

6. N. Santana, C. Nieva, M. Marro, S. Rao, A. Sierra, D. Petrov. Raman microspectroscopy is a tool to identify the metastatic ability of breast tumors. IEEE 978-1-4244-9837-6/11. 2011

Conference talks

1. M. Marro, A. Taubes, P. Villoslada, and D. Petrov. Detection of neuroininfammation through the retina by means of Raman spectroscopy and multivariate analysis. SPIE Photonics Europe conference in Brussels. April, 2012.

2. M. Marro, C. Nieva, R. Sanz-Pamplona, N. Santana-Codina, S. Rao, D. Petrov and A. Sierra. Raman spectroscopy analysis of the MCF10A malignant transformation. SPIE Photonics Europe conference in Brussels. April, 2012.

3. M. Marro. Extending the applicability of Raman microspectroscopy in biomedicine using statistical analysis and plasmonic effects. International Workshop Biophotonics 2011 Parma, Italy- June 8-10, 2011.

4. M. Marro Using 2D correlation and multivariate analysis combined with plasmonic effects to expand the use of Raman microspectroscopy in biomedical applications. Paper 8087-51, SPIE/OSA Biomedical Optics Conference in Munich, Germany May 22-26, 2011.

5. M. Marro. Photodynamic drug induced pH changes in live cancer cells studied with SERS probes. - COST meeting training school: Optical Micro-Manipulation by Nonlinear Nanophotonics. VisegrĂˇd, Hungary . October 5-8, 2010

Other contributions in conferences

1. M. Marro. Extending the applicability of Raman spectroscopy in Biomedicine by using statistical analysis. EPIC workshop on Biophotonics, 25 October, Paris, 2012 (Poster)

2. M. Marro. Extending the applicability of Raman microspectroscopy for biological samples using statistical analysis and plasmonic effects. L4H event at ICFO, 30-31 May 2011. (Poster)

3. M. Marro. Photodynamic drug induced pH changes in live cancer cells studied with SERS probes. - COST meeting training school: Optical Micro-Manipulation by Nonlinear Nanophotonics. Visegrad, Hungary . October 5-8, 2010. (Poster)

Bibliography

[1] J. Griffin and J. Shockcor. Metabolic profiles of cancer cells. *Nature Reviews*, 4:551–561, 2004.

[2] B. Wootla, M. Eriguchi, and M. Rodriguez. Multiple sclerosis an autoimmune disease? *Autoimmune diseases*, page 2012, 2012.

[3] M. Liscovitch, A. Freese, J. Blusztajn, and R. Wurtman. High-performance liquid chromatography of water-soluble choline metabolites. *Anal. Biochem.*, 151:182–187, 1985.

[4] S. Osorio, P. Do, and A. Fernie. Profiling primary metabolites of tomato fruit with gas chromatography-mass spectrometry. *Methods Mol. Biol.*, 860:101–109, 2012.

[5] Y. Liang, G. Wang, L. Xie, and L. Sheng. Recent development in liquid chromatography-mass spectrometry and emerging technologies for metabolite identification. *Curr. Drug. Metab.*, 12(4):329–344, 2011.

[6] M. Salim, P. Wright, and S. Vaidyanathan. A solvation-based screening approach for metabolite arrays. *Analyst*, 137(10):2350–2356, 2012.

[7] C. Zheng, S. Zhang, S. Ragg, D. Raftery, and O. Vitek. Identification and quantification of metabolites in H NMR spectra by Bayesian model selection. *Bioinformatics*, 27(12):1637–1644, 2011.

[8] R. Eisberg and R. Resnick. *Quantum Physics of atoms, molecules, solids, nuclei, and particles.* John Wiley and Sons, New York, USA, 1985.

[9] H. Haken and H. Wolf. *Molecular physics and elements of quantum chemistry. Introduction to experiments and theory.* Springer, Berlin, Germany, 1995.

[10] M. Alonso and E. Finn. *Fundamentos cuánticos y estadísticos.* Fondo educativo interamericano, Barcelona, Spain, 1971.

[11] P. Larkin. *IR and Raman Spectroscopy.* Elsevier, Waltham, USA, 2011.

[12] J. Ferraro, K. Nakamoto, and C. Brown. *Introductory Raman Spectroscopy.* Academic Press, London, England, 2003.

[13] F. Siebert and P. Hildebrandt. *Vibrational spectroscopy in life science.* Wiley-VCH, Weinheim, Germany, 2008.

[14] E. Smith and G. Dent. *Modern Raman spectroscopy - a practical approach.* Wiley and Sons, Sussex, England, 2005.

[15] D. Long. *The Raman Effect*. Willey, West Sussex, England, 2002.

[16] O. Samek, A. Jonas, Z. Pilat, P. Zemanek, L. Nevdbal, J. Triska, P. Kotas, and M. Trtilek. Raman microspectroscopy of individual algal cells: sensing unsaturation of storage lipids in vivo. *Sensors*, 10:8635–8651, 2010.

[17] O. Samek, P. Zemanek, A. Jonas, and H. Telle. Characterization of oil-producing microalgae using Raman spectroscopy. *Laser Phys Lett*, 8(10):701–709, 2011.

[18] E. Le Ru and P. Etchegoin. Principles of Surface-Enhanced Raman Spectroscopy: and related plasmonic effects. *Elsevier, Oxford*, 2009.

[19] J. Kneipp, H. Kneipp, W. Rice, and K. Kneipp. Optical probes for biological applications based on Surface-enhanced Raman scattering from indocyaninice green on gold nanoparticles. *Anal. Chem*, 77:2381–2385, 2005.

[20] R. Gebner, P. Rosch, R. Petry, M. Schmitt, M. Strehle, W. Kiefer, and J. Popp. The application of a sers fiber probe for the investigation of sensitive biological samples. *Analyst*, 129:1193–1199, 2004.

[21] J. Kneipp, H. Kneipp, B. Wittig, and K. Kneipp. Novel optical nanosensors for probing and imaging live cells. *Nanomedicine*, 6:214–226, 2010.

[22] M. Ochsenkuhn, P. Jess, H. Stoquert, K. Dholakia, and C-Campbell. Nanoshells for Surface Enhanced Raman spectroscopy in eukaryotic cells: cellular response and sensor development. *ACS Nano*, 3:3613–3621, 2009.

[23] U. Neugebauer, U. Schmid, K. Baumann, W. Ziebuhr, S. Kozitskaya, V. Deckert, M. Schmitt, and J. Popp. Towards a detailed understanding of bacterial metabolism- spectroscopic characterization of staphylococcus epidermis. *ChemPhysChem*, 8:124–137, 2007.

[24] A. Harris, M. Garg, X. Yang, S. Fisher, J. Kirkham, D. Smith, D. Martin, and A. High. Raman spectroscopy and advanced mathematical modelling in the discrimination of human tyroid cell lines. *Head Neck Oncol*, 1(38), 2009.

[25] J. Chan and D. Lieu. Label-free biochemical characterization of stem cells using vibrational spectroscopy. *J Biophoton*, 2(11):656–628, 2009.

[26] P. Jess, M. Mazilu, K. Dholakia, A. Riches, and C. Herrington. Optical detection and grading of lung neoplasia by Raman microspectroscopy. *Int J Cancer*, 124:376–380, 2009.

[27] U. Schmid, P. Rosch, M. Krause, M. Harz, J. Popp, and K. Baumann. Gaussian mixture discriminant analysis for single-cell differentiation of bacteria using micro-Raman spectroscopy. *Chem Int Lab Syst*, 96(2):159–171, 2009.

[28] J. Chan, D. Lieu, T. Huser, and R. Li. Label-free separation of human embryonic stem cells (hESCs) and their cardiac derivatives using Raman spectroscopy. *Anal Chem*, 81(4):1324–1331, 2009.

[29] A. Beljebbar, O. Bounche, M. Diebold, P. Guillou, J. Palot, D. Eudes, and M. Manfait. Identification of Raman spectroscopic markers for the characterization of normal and adenocarcinomatous colonic tissues. *Onc Hematology*, 72:255–264, 2009.

[30] A. Haka, K. Shafer, M. Fitzmaurice, J. Crowe, R. Dasari, and M. Feld. Diagnosing breast cancer by using Raman spectroscopy. *PNAS*, 102(5):12371–12376, 2005.

[31] P. Caspers, G. Lucassen, E. Carter, H. Bruining, and G. Puppels. In vivo confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles. *Autoimmune diseases*, 116(3):434–442, 2001.

[32] G. Singh, C. Creely, G. Volpe, and D. Petrov. Real time detection of hyperosmotic stress response in optically trapped single yeast cell using Raman microspectroscopy. *Anal. Chem.*, 77:2564–2568, 2005.

[33] G. Singh, G. Volpe, C. Creely, H. Grötsch, I. Geli, , and D. Petrov. The lag phase and g_1 phase of a single yeast cell monitored by Raman microspectroscopy. *J. Raman Spect.*, 37:858–864, 2006.

[34] S. Rao, S. Raj, S. Balint, C. Bardina Fons, S. Campoy, M. Llagostera, and D. Petrov. Single DNA molecule detection in an optical trap using surface-enhanced Raman scattering. *Appl. Phys. Lett.*, 96:213701, 2010.

[35] S. Piqueras, L. Duponchel, R. Tauler, and A. Juan. Resolution and segmentation of hyperspectral biomedical images by multivariate curve resolution-alternating least squares. *Anal Chim Acta*, 705:182–192, 2011.

[36] J. Jaumot, R. Gargallo A. de Juan, and R. Tauler. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom and intel lab systems*, 76(1):101–110, 2005.

[37] M. Fleishmann, P. Hendra, and A. McQuillan. Raman spectra of pyridine adsorbed at a silver electrode. *Chem. Phys. Lett.*, 26:163–166, 1974.

[38] M. Albrecht and J. Creighton. Anomalously Intense Raman Spectra of Pyridine at a Silver Electrode. *J. Am. Chem. Soc.*, 99:5215–5217, 1977.

[39] A. Campion and P. Kambhampati. Surface-enhanced Raman scattering. *Chem society reviews*, 27:241–250, 1998.

[40] Etchegoin and Le Ru. A perspective on single molecule SERS: current status and future challenges. *Phys. Chem. Chem. Phys.*, 10:6069–6200, 2008.

[41] K. Pearson. On lines and planes of closest fit to systems of points in space. *Phylosophycal Magazine*, 6:559–572, 1901.

[42] R. Fisher and W. MacKenzie. Studies in crop variation. The manurial response of different potato varieties. *J. Agric. Sci.*, 13:311–320, 1923.

[43] H. Hotelling. Analysis of complex statistical variables into principal components. *J. Educ. Psycol.*, 24:417–441, 1933.

[44] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemom Intel Lab systems*, 2:37–52, 1987.

[45] J. Joliffe. *Principal Component Analysis*. Springer, Berlin, 1986.

[46] S. Wold, M. Sjostrom, and L. Eriksson. Pls-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab.*, 58:190–130, 2001.

[47] T. Azzouz and R. Tauler. Application of multivariate curve resolution alternating least squares (mcr-als) to the quantitative analysis of pharmaceutical and agricultural samples. *Talanta*, 74:1201–1210, 2008.

[48] M. Garrido, F. Rius, and M. Larrechi. Multivariate curve resolution-alternating least squares (mcr-als) applied to spectroscopic data from monitoring chemical reactions processes. *Anal. Bioanal. Chem.*, 390:2059–2066, 2008.

[49] J. Saurina, S. Hernandez, R. Tauler, and A. Izquierdo. Multivariate resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions. *Chemom.*, 12:183–203, 1998.

[50] R. Tauler, S. Lacorte, M. Guillamon, R. Cespedes, P. Viana, and D. Barcelo. Chemometric modeling of main contamination sources in surface waters of portugal. *Environ. Toxicol. Chem.*, 23:565–575, 2004.

[51] J. Schoonover, R. Marx, and S. Zhang. Multivariate curve resolution in the analysis of vibrational spectroscopy data files. *Appl. spect.*, 57(5):154–170, 2003.

[52] I. Patel, J. Trevisan, G. Evans, V. Llabjani, P. Martin, H. Stringfellow, and F. Martin. High contrast images of uterine tissue derived using Raman microspectroscopy with the empty modelling approach of multivariate curve resolution-alternating least squares. *Analyst*, 136:4950–4959, 2011.

[53] O. Kvalheim and Y. Liang. Heuristic evolving latent projections: resolving two-way multicomponent data. 1. Selectivity, latent-projective graph, datascope, local rank, and unique resolution. *Anal Chem*, 64(8):936–946, 1992.

[54] Z. Zhu, W. Cheng, and Y. Zhao. Iterative target transformation factor analysis for the resolution of kinetic-spectral data with an unknown kinetic model. *Chemometrics Intel Lab Syst*, 64(2):157–167, 2002.

[55] W. Lawton and E. Sylvestre. Self modeling curve resolution. *Technometrics*, 13:617–633, 1971.

[56] M. Maeder. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Anal. Chem.*, 59:527–530, 1987.

[57] W. Windig and J. Guilment. Interactive self-modeling mixture analysis. *Anal. Chem.*, 63:1425–1432, 1991.

[58] I. Bax. *Two Dimensional Nuclear Magnetic Resonance in Liquids*. Kluber, 1982.

[59] I. Noda and Y. Ozaki. *Two-dimensional correlation spectroscopy - applications in vibrational and optical spectroscopy.* John Wiley and Sons, England., 2004.

[60] I. Noda. Two-dimensional infrared (2D IR) spectroscopy: Theory and applications. *Appl. Spectrosc.*, 44:550, 1990.

[61] M. Bradley. Curve fitting in Raman and IR spectroscopy: basic theory of line shapes and applications. *Thermo scientific application note*, page 50733, 2004.

[62] E. Fallman and O. Axner. Design for fully steerable dual-trap optical tweezers. *App Opt*, 36:2107–2113, 1997.

[63] S. Balint, M. P. Kreuzer, S. Rao, G. Badenes, P. Miskovsky, and D. Petrov. Simple route for preparing optically trappable probes for surface-enhanced Raman scattering. *J. Phys. Chem. C.*, 113:17724–17729, 2009.

[64] P. C. Lee and D. Meisel. Adsorption and Surface-Enhanced Raman of dyes on silver and gold sols. *J. Phys. Chem.*, 86:3391 – 3395, 1982.

[65] I. Mikhailyuk and A. Razzhivin. Background subtraction in experimental data arrays illustrated by the example of Raman spectra and fluorescent gel electrophoresis patterns. *Instr. Exp. Tech*, 46:765–769, 2003.

[66] S. R. Quake, H. Babcock, and S. Chu. The dynamics of partially extended single molecules of DNA. *Nature.*, 388:151–154, 1997.

[67] C. Bustamante, Z. Bryant, and S. B. Smith. Ten years of tension: single-molecule DNA mechanics. *Nature.*, 421:423–427, 2003.

[68] J. F. Marko and E. D. Siggia. Stretching DNA. *Macromolecules.*, 28:8759–8770, 1995.

[69] M. Woodside, P. Anthony, W. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science.*, 314:1001 – 1004, 2006.

[70] J. van Mameren, P. Gross, G. Farge, P. Hooijman, M. Modesti, M. Falkenberg, G. J. L. Wuite, and E. J. G. Peterman. Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *P. Natl. Acad. Sci. USA.*, 106:18231–18236, 2009.

[71] H. Fu, H. Chen, J. F. Marko, and J. Yan. Two distinct overstretched DNA states. *Nucleic Acids Res.*, 38:5594 – 5600, 2010.

[72] B. Prescott, W. Steinmetz, and G. J. Thomas Jr. Characterization of DNA structures by laser Raman spectroscopy. *Biopolymers.*, 23:235–256, 1984.

[73] G. J. Thomas, J. M. Benevides, S. A. Overman, T. Ueda, K. Ushizawa, M. Saitoh, and M. Tsuboi. Polarized Raman spectra of oriented fibers of A DNA and B DNA: Anisotropic and isotropic local Raman tensors of base and backbone vibrations. *Biophys. J.*, 68:1073–1088, 1995.

[74] E. Le Ru and P. Etchegoin. *Principles of Surface-Enhanced Raman Spectroscopy and related plasmonic effects.* Elsevier Science, Holland., 2009.

[75] S. Nie and S. R. Emory. Probing single molecules and single nanoparticles by Surface-Enhanced Raman Scattering. *Science.*, 275:1102–1106, 1997.

[76] K. Kneipp, H. Kneipp, V. B. Kartha, R. Manoharan, G. Deinum, I. Itzkan, R. R. Dasari, and M. S. Feld. Detection and identification of a single DNA base molecule using surface-enhanced Raman scattering (SERS). *Phys. Rev. E.*, 57:R6281–R6284, 1998.

[77] E. Bailo and V. Deckert. Tip-enhanced Raman spectroscopy of single RNA strands: towards a novel direct-sequencing method. *Angew. Chem. Int. Ed.*, 47:1658–1661, 2008.

[78] F. Gittes and C. F. Schmidt. Signals and noise in micromechanical measurements. *Method. Cell. Biol.*, 55:129 – 156, 1998.

[79] J.W. Shaevitz, E.A. Abbondanzieri, R. Landick, and S.M. Block. Backtracking by single rna polymerase molecules observed at near-base-pair resolution. *Nature.*, 426:684–687, 2003.

[80] J. Rezac, P. Hobza, and S. A. Harris. Stretched DNA investigated using molecular-dynamics and quantum-mechanical calculations. *Biophys. J.*, 98:101–110, 2010.

[81] C.H. Reinsch. Smoothing by spline functions. *Numerical Mathematics*, 10:177–183, 1967.

[82] X. Qian and S. Nie. Single-molecule and single-nanoparticle SERS: from fundamental mechanisms to biomedical applications. *Chem. Soc. Rev.*, 37:912 – 920, 2008.

[83] T. Yano, P. Verma, Y. Saito, T. Ichimura, and S. Kawata. Pressure-assisted tip-enhanced Raman imaging at a resolution of a few nanometers. *Nat. Photonics.*, 3:473 – 477, 2009.

[84] L. Novotny and B. Hecht. *Principles of Nano-Optics.* Cambridge University Press, England., 2006.

[85] A. MacAskill, D. Crawford, D. Graham, and K. Faulds. DNA sequence detection using surface-enhanced resonance Raman spectroscopy in a homogenous multiplexed assay. *Anal. Chem.*, 81:8134 – 8140, 2009.

[86] A. Barhoumi, D. Zhang, F. Tam, and N.J. Halas. Surface-enhanced Raman spectroscopy of DNA. *J. Am. Chem. Soc.*, 130:5523–5529, 2008.

[87] T. J. Dougherty, Ch. J. Gomer, B. W. Henderson, G. Jori, D. Kessel, M. Korbelik, J. Moan, and Q. Peng. Photodynamic therapy. *Journal of National Cancer Institute*, 90:889–905, 1998.

[88] C. Beghetto, Ch. Renken, O. Eriksson, G. Jori, P. Bernardi, and F. Ricchelli. Implications of the generation of reactive oxygen species by photoactivated calcein for mitochondrial studies. *European Journal of Biochemistry*, 267:5585–5592, 2000.

[89] R. D. Almeida, B. J. Manadas, A. P. Carvalho, and C. B. Duarte. Intracellular signaling mechanisms in photodynamic therapy. *Biochimica et Biophysica Acta*, 1704:59–86, 2004.

[90] Ch. Chang, C. Ashendel, T. Chan, R. Geahlen, J. McLaughlin, and D. Waters. Oncogene signal transduction inhibitors from chinese medical plants. *Pure Appl. Chem.*, 71:1101–1104, 1999.

[91] Y. Yan, X. Su, Y. Liang, J. Zhang, Ch. Shi, Y. Lu, L. Gu, and L. Fu. Emodin azide methyl anthraquinone derivate triggers mitochondrial-dependent cell apoptosis involving in caspase-8-mediateed bid cleavage. *Mol. Canc. Ther.*, 7:1688–1697, 2008.

[92] Y. Su, H. Chang, S. Shyue, and S. Hsu. Emodin induces apoptosis in human lung adenocarcinoma cells through a reactive oxygen species-dependent mitochondrial signaling pathway. *Biochem. Pharm.*, 70:229–241, 2005.

[93] S. Lin, W. Lai, Ch. Ho, F. Yu, G. Chen, J. Yang, K. Liu, M. Lin, P. Wu, M. Fan, and J. Chung. Emodin induces apoptosis of human tongue squamous cancer SCC-4 cells through reactive oxygen species and mitochondria-dependent pathways. *Anticancer Research*, 29:327–336, 2009.

[94] Z. Lu, Y. Tao, Z. Zhou, J. Zhang, C. Li, L. Ou, and B. Zhao. Mitochondrial reactive oxygen species and nitric oxide-mediated cancer cell apoptosis in 2-butylamino-2-demethoxyhypocrellin B photodynamic treatment. *Free Radical Biology and Medicine*, 41:1590–1605, 2006.

[95] S. Matsuyama and J. C. Reed. Mitochondrial-dependent apoptosis and cellular pH regulation. *Cell Death and Differentiation*, 7:1155–1165, 2000.

[96] J. Llopis, J. M. McCaffery, A. Miyawaki, M. G. Farquhar, and R. Y. Tsien. Measurement of cytosolic, mitochondrial, and golgi pH in single living cells with green fluorescent proteins. *Proc. Natl. Acad. Sci.*, 95:6803–6808, 1998.

[97] G. Bach, Ch. Chen, and R. E. Pagano. Elevated lysosomal pH in mucolipidosis type IV cells. *Clinical Chimica Acta*, 280:173–179, 1999.

[98] S. Sanchez-Cortes, D. Jancura, P. Miskovsky, and A. Bertoluzza. Near infrared surface-enhanced Raman spectroscopic study of antiretroviraly drugs hypercin and emodin in aqueous silver colloids. *Spectrochim. Act. A*, 53:769–779, 1997.

[99] A. Michota and J. Bukowska. Surface-enhanced Raman scattering (SERS) of 4-mercaptobenzoic acid on silver and gold substrates. *Journal of Raman Spectroscopy*, 34:21–25, 2003.

[100] Ch. E. Talley, L. Jusinski, Ch. W. Hollars, S. M. Lane, and T. Huser. Intracellular pH sensors based on surface-enhanced Raman scattering. *Analytical Chemistry*, 76:7064–7068, 2004.

[101] J. Kneipp, H. Kneipp, B. Wittig, and K. Kneipp. One and two photon excited optical pH probing for cells using surface enhanced Raman and hyper-Raman nanosensors. *Nano Lett.*, 7:2819–2823, 2007.

[102] A. M. Alkilany and C. J. Murphy. Toxicity and cellular uptake of gold nanoparticles: what we have learned so far? *J. Nanopart. Research*, pages DOI 10.1007/s11051–010–9911–8, 2010.

[103] J. A. Khan, B. Pillai, T. K. Das, Y. Singh, and S. Maiti. Molecular effects of uptake of gold nanoparticles in HeLa cells. *ChemBioChem*, 8:1237–1240, 2007.

[104] S. Al-Hilli, M. Willander, A. Ost, and P. Stralfors. Zno nanorods as an intracellular sensor for ph measurements. *J. Appl. Phys.*, 102:084304Đ1 Đ 084304Đ5, 2007.

[105] J. Scaffidi, M. Gregas, V. Seewaldt, and T. Vo-Dinh. SERS-based plasmonic nanobiosensing in single living cells. *Anal. Bioanal. Chem.*, 393:1135Đ1141, 2009.

[106] Tuchin V., editor. *The Handbook of Photonics and Medical Science*. CRC Press, Taylor and Francis Group, 2010.

[107] S. Balint, M. Kreuzer, S. Rao, G. Badenes, P. Miskovsky, and D. Petrov. A simplistic route for preparing optically trappable probes for surface enhanced Raman scattering. *J. Phys. Chem. C*, 113:17724–17729, 2009.

[108] S. Balint, S. Rao, M. M. Sanchez, V. Huntosova, P. Miskovsky, and D. Petrov. Diffusion and cellular uptake of drugs in live cells studied with surface-enhanced Raman scattering probes. *Journal of Biomedical Optics*, 15:027005–1 – 027005–7, 2010.

[109] P. C. Lee and D. Meisel. Adsorption and surface-enhanced Raman of dyes on silver and gold sols. *J. Phys. Chem.*, 86(17):3391–3395, 1982.

[110] C. Creely, G. Singh, and D. Petrov. Dual-wavelength optical tweezers for confocal Raman spectroscopy. *Opt. Comm.*, 245:465–470, 2005.

[111] I. Mikhailyuk and A. Razzhivin. Background subtraction in experimental data arrays illustrated by the example of Raman spectra and fluorescent gel electrophoresis patterns. *Instr. Exp. Tech.*, 46:765–769, 2003.

[112] E.D. Korn and R.A. Weisman. Phagocytosis of latex beads by acanthamoeba. *J. Cell Biol.*, 34:219–227, 1967.

[113] G. P. Schubart, C.S. Happ, and A. Ruck. Early and late apoptotic events induced in human glioblastoma cells by Hypericin PDT. *Medical Laser Application*, 23:25–30, 2008.

[114] N. Uzunbajakava, A. Lenferink, Y. Kraan, E. Volokhina, G. Vrensen, J. Greve, and C. Otto. Nonresonant confocal Raman imaging of DNA and protein distribution in apoptotic cells. *Biophys. J.*, 84:3968–3981, 2003.

[115] J. M. Dubois, G. Ouanounou, and B. R. Dubois. The Boltzmann equation in molecular biology. *Progress in Biophysics and Molecular Biology*, 99:87–93, 2009.

[116] K. Kelly, B. N. Rospendowski, W. E. Smith, and C. R. Wolf. Surface enhanced resonance Raman scattering as a probe of the spin state of structurally related cytochromes P-450 from rat liver. *FEBS Lett.*, 222:120–124, 1987.

[117] S. Berezhna, H. Wohlrab, and P. M. Champion. Resonance Raman investigations of cytochrome C conformational change upon interaction with the membranes intact and Ca^{2+} - exposed mitochondria. *Biochemistry*, 42:6149–6158, 2003.

[118] I. Noda, W. Allen, and S. Lindberg. Two-dimensional Raman correlation spectroscopy study of an emulsion copolymerization reaction process. *Appl. Spectrosc.*, 63:224–232, 2009.

[119] J. Yi, J. Yang, r. He, F. Gao, H. Sang, X. Tang, and R. D. Ye. Emodin enhances arsenic trioxide-induced apoptosis via generation of reactive oxygen species and inhibition of survival signaling. *Cancer Research*, 64:108–116, 2004.

[120] D. Kessel and Y. Luo. Mitochondrial photodamage and PDT-induced apoptosis. *Journal of Photochemistry and Photobiology B: Biology*, 42:89–95, 1998.

[121] S. Guido and G. Tomaiuolo. Microconfined flow behavior of red blood cells in vitro. *C. R. Physique*, 10:751–763, 2009.

[122] S. Chien. Red cell deformability and its relevance to blood flow. *Annu. Rev. Physiol.*, 49:177–192, 1987.

[123] S. Suresh. Mechanical response of human red blood cells in health and disease: Some structure-property-function relationships. *J. Mater. Res.*, 21:1871–1877, 2006.

[124] E. A. Evans. New membrane concept applied to the analysis of fluid shear- and micropipette-deformed red blood cells. *Biophys. J.*, 13:941–954, 1973.

[125] A. Tozeren, R. Skalak, Kuo-Li P. Sung, and S. Chien. Viscoelastic behavior of erythrocyte membrane. *Biophys. J.*, 39:23–32, 1982.

[126] G. Lenormand, S. Henon, A. Richert, J. Simeon, and F. Gallet. Direct measurement of the area expansion and shear moduli of the human red blood cell membrane skeleton. *Biophys. J.*, 81:43–56, 2001.

[127] Y. Z. Yoon, J. Kotar, A. T. Brown, and P. Cicuta. Red blood cell dynamics: from spontaneous fluctuations to non-linear response. *Soft Matter*, 7:2042–2051, 2011.

[128] G. Bao and S. Suresh. Cell and molecular mechanics of biological materials. *Nature Materials*, 2:715–725, 2003.

[129] S. Rao, S. Balint, B. Cossins, V. Guallar, and D. Petrov. Raman study of mechanically induced oxygenation state transition of red blood cells using optical tweezers. *Biophys. J.*, 96:209–216, 2009.

[130] Y.-Z. Yoon, J. Kotar, G. Yoon, and P. Cicuta. Non-linear mechanical response of the red blood cell. *Phys. Biology*, 5:036007, 2008.

[131] C. M. Creely, G. P. Singh, and D. Petrov. Dual wavelength optical tweezers for confocal Raman spectroscopy. *Optics Commun.*, 245:465–470, 2005.

[132] Alvin C. Rencher. *Methods of Multivariate Analysis*. John Wiley and Sons, USA, 2002.

[133] S. Balint, S. Rao, M. Marro, P. Miskovsky, and D Petrov. Monitoring of local pH in photodynamic therapy-treated live cancer cells using surface-enhanced Raman scattering probes. *J. Raman Spectr.*, 42:1215–1221, 2011.

[134] B. R. Wood, B. Tait, and D. McNaughton. Micro-Raman characterisation of the R to T state transition of haemoglobin within a single living erythrocyte. *Biochimica et Biophysica Acta*, 1539:58–70, 2001.

[135] B. R. Wood and D. McNaughton. Raman excitation wavelength investigation of single red blood cells in vivo. *J. Raman Spectrosc.*, 33:517–523, 2002.

[136] B. R. Wood, P. Caspers, G. J. Puppels, S. Pandiancherri, and D. McNaughton. Resonance Raman spectroscopy of red blood cells using near-infrared laser excitation. *Anal Bioanal Chem*, 387:1691–1703, 2007.

[137] A. Bankapur, E. Zachariah, S. Chidangil, M. Valiathan, and D. Mathur. Raman tweezers spectroscopy of live, single red and white blood cells. *PLoS ONE*, 5:e10427, 2010.

[138] S. C. Goheen, L. J. Lis, O. Kucuk, M. P. Westerman, and J. W. Kaufman. Compositional dependence of spectral features in the Raman spectra of erythrocyte membranes. *J. Raman Spectrosc.*, 24:275–279, 1993.

[139] X.L. Yan, R.X. Dong, L. Zhang, X.J. Zhang, and Z.W. Zhang. Raman spectra of single cell from gastrointestinal cancer patients. *World J Gastroenterol*, 11:3290–3292, 2005.

[140] S.Hu, K.M. Smith, and T.G. Spiro. Assignment of protoheme resonance Raman spectrum by heme labeling in myoglobin. *J. Am. Chem. Soc.*, 118:12638–12646, 1996.

[141] A. Fauci, D. Kasper, E. Braunwald, S. Hauser, D. Longo, J. Jameson, and L. Loscalzo. *Harrison's Principles of Internal medicine*. McGraw-Hill, 2004.

[142] N. Shaklai, J. Yguerabide, and H.M. Ranney. Interaction of hemoglobin with red blood cell membranes as shown by a fluorescent chromophore. *Biochemistry*, 16:5585–5592, 1977.

[143] S. Fischer, R.L. Nagel, R.M. Bookchin, EF Jr. Roth, and I.Tellez-Nagel. The binding of hemoglobin to membranes of normal and sickle erythrocytes. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 375:422–433, 1975.

[144] B. L. Horecker. The absorption spectra of hemoglobin and its derivatives in the visible and near infra-red region. *J. Biol. Chem.*, 148:173–183, 1943.

[145] V. Bennett and P. J. Stenbuck. Human erythrocyte ankyrin. purification and properties. *J Biol Chem.*, 255:2540–2548, 1980.

[146] J. Ferlay, H. Shin, F. Bray, D. Forman, and C. Mathers. Globocan 2008, Cancer incidence and mortality worldwide. *IARC Cancer Base*, 2010.

[147] American cancer society:cancer facts and figures 2010. *Atlanta*, 2010.

[148] J. Sleeman and P. Steeg. Cancer metastasis as a therapeutic target. *Eur J Cancer*, 46:1177–1180, 2010.

[149] R. Weil, D. Palmieri, J. Bronder, A. Stark, and P. Steeg. Breast cencer metastasis to the central nervous system. *Am J Pathol*, 167:913–920, 2010.

[150] D. Rodenhiser, J. Andrews, T. Vandenberg, and A. Chambers. Gene signatures of breast cancer progression and metastasis. *Breast Cancer Res*, 13:201–208, 2011.

[151] M. Guedj, L. Marisa, A. Reynies, B. Orsetti, and R. Schiappa. A refined molecular taxonomy of breast cancer. *Oncogene*, 31:1196–1206, 2012.

[152] R. Brough, J. Frankum, D. Sims, A. Mackay, and A. Mendes. Functional viability profiles of breast cancer. *Cancer discovery*, 1:261–273, 2011.

[153] K. Kasaian and S. Jones. A new frontier in personalized cancer therapy: mapping molecular changes. *Future oncology*, 7:873–894, 2011.

[154] W. Koppenol, P. Bounds, and C. Dang. Otto warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer*, 11:325–337, 2011.

[155] J. Bayley and P. Devilee. The warbug effect in 2012. *Curr Opin Oncol*, 24:62–67, 2012.

[156] J. Bayley and P. Devilee. An in vitro investigation of metabolically sensitive biomarkers in breast cancer progression. *Breast Cancer Res Treat*, 133:959–968, 2012.

[157] H. Hu, C. Deng, T. Yang, Q. Dong, and Y. Chen. Proteomics revisits the cancer metabolome. *Expert Rev Proteomics*, 8:505–533, 2011.

[158] A. Vazquez, R. Colomer, J. Brunet, R. Lupu, and J. Menendez. Overexpression of fatty acid synthase gene activates her1/her2 tyrosine kinase receptors in human breast epithelial cells. *Cell Prolif*, 41:59–85, 2008.

[159] M. Hilvo, C. Denkert, L. Lehtinen, B. Muller, and S. Brockmoller. Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res*, 71:3236–3245, 2011.

[160] D. Guo, F. Reinitz, M. Youssef M, C. Hong, and D. Nathanson. An lxr agonist promotes glioblastoma cell death through inhibition of an EGFRAKT/SREBP-1/LDLR–Dependent pathway. *Cancer Discovery*, 1:442–456, 2011.

[161] K. Bhalla, B. Hwang, R. Dewi, L. Ou, and W. Twaddel. PGC1alpha promotes tumor growth by inducing gene expression programs supporting lipogenesis. *Cancer Res*, 71:6888–6898, 2011.

[162] E. Rysman, K. Brusselmans, K. Scheys, L. Timmermans, and R. Derua. De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Res*, 70:8117–8126, 2010.

[163] D. Calvisi, C. Wang, C. Ho, S. Ladu, and S. Lee. Increased lipogenesis, induced by akt-mtorc1-rps6 signaling, promotes development of human hepatocellular carcinoma. *Gastroenterology*, 140:1071–1083, 2011.

[164] L. Vedin, S. Lewandowski, P. Parini, J. Gustafsson, and K. Steffensen. The oxysterol receptor lxr inhibits proliferation of human breast cancer cells. *Carcinogenesis*, 30:575–579, 2009.

[165] J. Popp, D. Krafft, and T. Mayerhofer. Modern Raman spectroscopy for biomedical applications. a variety of Raman spectroscopical techniques on the threshold of biomedical applications. *Optik and photonik*, 6:24–28, 2011.

[166] W. Min, C. Freudiger, S. Lu, and X. Xie. Coherent nonlinear optical imaging: beyond fluorescence microscopy. *Annu Rev Phys Chem*, 62:507–530, 2011.

[167] M. fen, P. Xanthopoulos, G. Pyrgiotakis, S. Grobmyer, P. Pardalos, and L. Hench. Raman spectroscopy for clinical oncology. *Advances in Optical Technol*, 2011, 2011.

[168] J. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, and J. González-Trujillo. Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. *Lasers med sci*, 22:236–239, 2011.

[169] C. Lieber, H. Nethercott, and M. Kabeer. Cancer field effects in normal tissues revealed by Raman spectroscopy. *Biomed Opt Express*, 1:975–982, 2010.

[170] J. Jackson, editor. *A User's guide to Principal Components*. Wiley, New York, 1991.

[171] J. Westerhuis, H. Hoefsloot, S. Smit, D. Vis, and A. Smilde. Assessment of PLS-DA cross validation. *Metabolomics*, 4:81–89, 2008.

[172] D. Sarrio, S. Rodriguez, D. Hardisson, A. Cano, and G. Moreno. Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer Res*, 68:989–997, 2008.

[173] C. Freudiger, W. Min, B. Saar, S. Lu, and G. Holtom. Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science*, 322:1857–1861, 2008.

[174] T. Bocklitz, A. Walter, K. Hartmann, P. Rosch, and J. Popp. How to pre-process Raman spectra for reliable and stable models? *Anal. Chimica Acta*, 704:47–56, 2011.

[175] N. Zelcer, C. Hong, R. Boyadjian, and P. Tontonoz. LXR regulates cholesterol uptake through idol-dependent ubiquitination of the LDL receptor. *Science*, 325:100–104, 2009.

[176] R. Rawson. The SREBP pathway-insights from insigs and insects. *Nature Rev Mol Cel Biol*, 4:631–640, 2003.

[177] M. Voss, M. Moller, D. Powe, B. Niggemann, and K. Zanker. Luminal and basal-like breast cancer cells show increased migration induced by hypoxia, mediated by an autocrine mechanism. *BMC Cancer*, 11:158–169, 2011.

[178] M. Rudolph, J. Monks, V. Burns, M. Phistry, and R. Marians. Sterol regulatory element binding protein and dietary lipid regulation of fatty acid synthesis in the mammary epithelium. *AJP Endo*, 299:918–927, 2010.

[179] M. Rudolph, J. Monks, V. Burns, M. Phistry, and R. Marians. Role of the phosphatidylinositol 3-kinase/PTEN/Akt kinase pathway in the overexpression of fatty acid synthase in LNCaP prostate cancer cells. *AJP Endo*, 62:642–646, 2002.

[180] T. Porstmannm, B. Griffiths, Y. Chung, O. Delpuech, and J. Griffiths. PKB/AKT induces transcription of enzymes involved in cholesterol and fatty acid biosynthesis via activation of SREBP. *Oncogene*, 24:6465–6481, 2005.

[181] K. Agelopoulos, B. Greve, H. Schmidt, H. Pospisil, and S. Kurtz. Selective regain of egfr gene copies in CD44+/CD24-/low breast cancer cellular model MDA-MB-468. *BMC Cancer*, 10:78–89, 2010.

[182] R. Jones and C. Thompson. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev*, 23:537–548, 2009.

[183] R. Cairns, I. Harris, and T. Mak. Regulation of cancer cell metabolism. *Nature Reviews*, 11:85–95, 2011.

[184] C. Antalis, T. Arnold, T. Rassol, B. Lee, and K. Buhmann. High ACAT1 expression in estrogen receptor negative basal-like breast cancer cells is associated with LDL-induced proliferation. *Breast Cancer Res Treat*, 122:661–670, 2011.

[185] C. Antalis, A. Uchida, K. Buhmann, and R. Siddiqui. Migration of MDA-MB-231 breast cancer cells depends on the availability of exogenous lipids and cholesterol esterification. *Clin Exp Metastases*, 28:733–741, 2011.

[186] G. Diaz, M. Melisa, B. Batetta, F. Angius, and A. Falchi. Hydrophobic characterization of intracellular lipids in situ by Nile Red red/yellow emission ratio. *Micron*, 39:819–824, 2008.

[187] J. Menendez and R. Lupu. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature reviews*, 7:763–777, 2007.

[188] J. Menendez and R. Lupu. Genome-wide changes accompanying knockdown of fatty acid synthase in breast cancer. *BMC Genomics*, 8:168–179, 2007.

[189] Z. Movasaghi, S. Rehman, and I. Rehman. Raman spectroscopy of biological tissues. *Applied Spectroc Rev*, 42:493–541, 2007.

[190] A. Garcia, L. Raniero, R. Canevari, K. Jalkanen, and R. Bitar. High-wavenumber ft-Raman spectroscopy for in vivo and ex vivo measurements of breast cancer. *Theor Chem Acc*, 130:1231–1238, 2011.

[191] G. Moreno-Bueno, F. Portillo, and A. Cano. Transcriptional regulation of cell polarity in emt and cancer. *Oncogene*, 27:6958–6969, 2008.

[192] K. Polyak and R. Weinberg. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer*, 9:265–273, 2009.

[193] E. Tomaskovic-Crook, E. Thompson, and J. Thiery. Epithelial to mesenchymal transition and breast cancer. *Breast Cancer Research*, 11:213–230, 2009.

[194] B. Lehmann, J. Bauer, X. Chen, M. Sanders, A. Chakravarthy, Y. Shyr, and J. Pietenpol. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*, 121:2750–2767, 2011.

[195] F. Tavassoli and P. Devilee, editors. *WHO Classification of Tumors, Pathology and Genetics: Tumors of the breast and female genital organs.* IARC Press, Lyon (France), 2003.

[196] L. Hartsuiker, N. Zeijen, L. Terstappen, and C. Otto. A comparison of breast cancer tumor cells with varying expression of the Her2/neu receptor by Raman microspectroscopic imaging. *Analyst*, 135:3220–3226, 2010.

[197] L. Raniero, R. Canevari, L. Ramalho, F. Ramalho, and E. dos Santos. In and ex vivo breast disease study by Raman spectroscopy. *Theor Chem Acc*, 130:1239–1247, 2011.

[198] S. Rehman, Z. Movasaghi, A. Tucker, S. Joel, and J. Darr. Raman spectroscopic analysis of breast cancer tissues: identifying differences between normal, invasive ductal carcinoma and ductal carcinoma in situ of the breast tissue. *J Raman Spectrosc*, 38:1345–1351, 2007.

[199] J. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-Garcia, R. Huerta-Franco, and G. Gonzalez-Trujillo. Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. *Lasers Med Sci*, 22:229–236, 2007.

[200] B. Eckhardt, P. Francis, B. Parker, and R. Anderson. Strategies for the discovery and development of therapies for metastatic breast cancer. *Nat Rev Drug Discov*, 11:479–497, 2012.

[201] Z. Khamis, Z. Sahab, and Q. Sang. Active roles of tumor stroma in breast cancer metastasis. *Int J Breast Cancer*, page doi: 10.1155/2012/574025, 2012.

[202] E. Tomaskovic-Crook, E. Thompson, and J. Thiery. Epithelial to mesenchymal transition and breast cancer. *Breast Cancer Res*, 11:213, 2009.

[203] F. Davis. Remodeling of purinergic receptor-mediated ca2+ signaling as a consequence of egf-induced epithelial-mesenchymal transition in breast cancer cells. *PLoS One*, 6:doi: 10.1371/journal.pone.0023464, 2011.

[204] T. Sorlie. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*, 11:10869–74, 2001.

[205] E. Rysman. De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Res*, 70:8117–8126, 2010.

[206] G. Moreno-Bueno, F. Portillo, and A. Cano. Transcriptional regulation of cell polarity in emt and cancer. *Oncogene*, 27:6958–6969, 2008.

[207] C. Scheel and R. Weinberg. Cancer stem cells and epithelial-mesenchymal transition: concepts and molecular links. *Semin Cancer Biol*, 22:396–403, 2012.

[208] T. Celia-Terrasa. Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J Clin Invest*, 122:1849–1868, 2012.

[209] W. Luo, R. Chang, J. Zhong, A. Pandey, and G. Semenza. Histone demethylase jmjd2c is a coactivator for hypoxia-inducible factor 1 that is required for breast cancer progression. *Proc Natl Acad Sci USA*, 109:E3367–3376, 2012.

[210] C. Creighton. Development of resistance to targeted therapies transforms the clinically associated molecular profile subtype of breast tumor xenografts. *Cancer Res*, 68:7493–7501, 2008.

[211] C. Nieva, M. Marro, N. Santana, S. Rao, D. Petrov, and A. Sierra. The lipid phenotype of breast cancer cells characterized by Raman microspectroscopy: towards a stratification of malignancy. *PLoS ONE*, 7:e46456, 2012.

[212] A. Vazquez-Martin, R. Colomer, J. Brunet, R. Lupu, and J. Menendez. Overexpression of fatty acid synthase gene activates her1/her2 tyrosine kinase receptors in human breast epithelial cells. *Cell Prolif*, 41:59–85, 2008.

[213] J. Brunet. Brunet, j. et al. brca1 and acetyl-coa carboxylase: the metabolic syndrome of breast cancer. *Mol Carcinog*, 47:157–163, 2008.

[214] F. Bonnier and H. Byrne. Understanding the molecular information contained in principal component analysis of vibrational spectra of biological systems. *Analyst*, 137:322–332, 2012.

[215] R. Edgar, M. Domrachev, and A. Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res*, 1:207–210, 2002.

[216] J. Charafe. Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene*, 25:2273–2284, 2006.

[217] J. Shi and M. Walker. Gene set enrichment analysis (gsea) for interpreting gene expression profiles. *Current Bioinformatics*, 2(2):133–137, 2007.

[218] R: A language and enviroment for statistical computing. R foundation for statistical computing, 2004.

[219] N. Gunasinghe, A. Wells, E. Thompson, and H. Hugo. Mesenchymal-epithelial transition (met) as a mechanism for metastatic colonisation in breast cancer. *Cancer Metast Rev*, 31:469–478, 2012.

[220] C. Krafft, L. Neudert, T. Simat, and R. Salzer. Near infrared Raman spectra of human brain lipids. *Spectrochim Acta A*, 61(7):1529–1535, 2005.

[221] S. Devpura, J. Thakur, S. Sethi, V. Naik, and R. Naik. Diagnosis of head and neck squamous cell carcinoma using Raman spectroscopy: tongue tissues. *J Raman spect*, 43:490–496, 2012.

[222] M. Kacurakova and M. Mathlouthi. Ftir and laser-Raman spectra of oligosaccharides in water: characterization of the glycosid bond. *Carbohydrate Research*, 284:145–157, 1996.

[223] M. Dauchez, P. Lagant, P. Derreumaux, and G. Vergoten. Force field and vibrational spectra of oligosaccharides with different glycosidic linkages-Part II. Maltose monohydrate, cellobiose and gentiobiose. *Spectrochim Acta A*, 50:105–118, 1994.

[224] M. Dauchez, P. Lagant, P. Derreumaux, and G. Vergoten. Force field and vibrational spectra of oligosaccharides with different glycosidic linkages-Part I. Trehalose dihydrate, sophorose monohydrate and laminaribiose. *Spectrochim Acta A*, 50:87–104, 1994.

[225] Y. Huang, T. Karashima, M. Yamamoto, and H. Hamaguchi. Molecular level investigation of the structure, transformation and bioactivaty of single living fission yeast cells by time and space resolved Raman spectroscopy. *Biochemistry*, 44:10009–10019, 2005.

[226] C. Huang, M. Ando, H. Hamaguchi, and S. Shigeto. Disentangling dynamic changes of multiple cellular components during the yeast cell cycle by in vivo multivariate Raman imaging. *Anal Chem*, 84:5661–5668, 2012.

[227] N. Santana, C. Nieva, M. Marro, S. Rao, and D. Petrov. Raman microspectroscopy is a tool to identify the metastatic ability of breast tumors. *IEEE explore*, 2011.

[228] A. Subramanian. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*, 102:15545–15550, 2005.

[229] M. Jechlinger, S. Grunert, and H. Beug. Mechanisms in epithelial plasticity and metastasis: insights from 3d cultures and expression profiling. *J Mammary Gland Biol Neoplasia*, 7:415–432, 2002.

[230] W. Koppenol, P. Bounds, and C. Dang. Otto warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer*, 11:325–337, 2011.

[231] P. Papageorgis. Smad signaling is required to maintain epigenetic silencing during breast cancer progression. *Cancer Res*, 70:968–978, 2010.

[232] P. Chinnaiyan. The metabolomic signature of malignant glioma reflects accelerated anabolic metabolism. *Cancer Res*, 72:5878–5888, 2012.

[233] C. Antalis. High acat1 expression in estrogen receptor negative basal-like breast cancer cells is associated with ldl-induced proliferation. *Breast Cancer Res Treat*, 122:661–670, 2012.

[234] C. Antalis, A. Uchida, K. Buhman, and R. Siddiqui. Migration of mda-mb-231 breast cancer cells depends on the availability of exogenous lipids and cholesterol esterification. *Clin Exp Metastasis*, 28:733–741, 2011.

[235] M. Travers, I. Gow, M. Barber, J. Thomson, and D. Shennan. Indoleamine 2,3-dioxygenase activity and L-tryptophan transport in human breast cancer cells. *Biochim Biophys Acta*, 1661:106–112, 2004.

[236] V. Levina, Y. Su, and E. Gorelik. Immunological and nonimmunological effects of indoleamine 2,3-dioxygenase on breast tumor growth and spontaneous metastasis formation. *Clin Dev Immunol*, doi: 10.1155/2012/173029:106–112, 2012.

[237] L. Kumar and A. Clarke. Gene manipulation through the use of small interfering rna (sirna): from in vitro to in vivo applications. *Adv Drug Deliv Rev*, 59:87–100, 2007.

[238] L. Huang. Identification of a gene-expression signature for predicting lymph node metastasis in patients with early stage cervical carcinoma. *Cancer*, 117:3363–3373, 2011.

[239] M. Platten, W. Wick, and B. Van de Eynde. Tryptophan catabolism in cancer: beyond ido and tryptophan depletion. *Cancer Res*, 72:5435–5440, 2012.

[240] C. Creighton, J. Chang, and J. Rosen. Epithelial-mesenchymal transition (EMT) in tumor-initiating cells and its clinical implications in breast cancer. *J Mammary Gland Biol Neoplasia*, 15:253–260, 2010.

[241] J. Fuxe, T. Vincent, and A. Garcia de Herreros. Transcriptional crosstalk between TGF-? and stem cell pathways in tumor cell invasion: role of EMT promoting Smad complexes. *Cell Cycle*, 9:2363–2674, 2010.

[242] V. Stresing, E. Baltziskueta, N. Rubio, J. Blanco, M. Arriba, J. Valls, M. Janier, P. Clezardin, R. Sanz-Pamplona, C. Nieva, M. Marro, D. Petrov, and A. Sierra. Peroxiredoxin 2 specifically regulates the oxidative and metabolic-stress response of human breast cancer cells. *Oncogene*, page doi:10.1038/onc, 2012.

[243] V. Stresing. Challenges of microarray data and the evaluation of gene expression profile signatures. *Cancer Invest*, 26:327–332, 2008.

[244] D. Drasin, T. Robin, and H. Ford. Breast cancer epithelial-to-mesenchymal transition: examining the functional consequences of plasticity. *Breat Cancer Res*, page doi: 10.1186/bcr3037, 2011.

[245] M. Hilvo. Novel therasnostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res*, 71:3236–3245, 2011.

[246] R. Jones and C. Thompson. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev*, 23:537–548, 2009.

[247] N. Simpson, V. Tryndyak, F. Beland, and I. Pogribny. An in vitro investigation of metabolically sensitive biomarkers in breast cancer progression. *Breast Cancer Res Treat*, 133:959–968, 2012.

[248] P. Yang, A. Vos, and A. Kijlastra. Macrophages in the retina of normal lewis rats and their dynamics after injection of lipopolysaccharide. *Ophthalmol Vis Sci*, 37(1):77–85, 1996.

[249] Detrick B. and Hooks J. Immune regulation in the retina. *Immunol Res*, 47:153–161, 2010.

[250] C. Kaur, G. Rathnasamy, and E. Ling. Roles of activated microglia in hypoxia induced neuroinflammation in the developing brain and the retina. *J Neuroimmune Pharmacol*, 2012.

[251] A. Ebneter, R. Casson, J. Wood, and G. Chidlow. Microglial activation in the visual pathway in experimental glaucoma: spatiotemporal characterization and correlation with axonal injury. *Invest Ophthalmol Vis Sci*, 52(12):6448–6460, 2010.

[252] K. Shindler, E. Ventura, M. Dutt, and A. Rostami. Inflammatory demyelination induces axonal injury and retinal ganglion cell apoptosis in experimental optic neuritis. *Exp Eye Res*, 87(3):208–213, 2008.

[253] P. Yang, L. Chen, R. Zwart, and A. Kijlstra. Immune cells in the porcine retina: distribution, characterization and morphological features. *Ophthalmol Vis Sci*, 43(5):1488–1492, 2002.

[254] E. Frohman. Modeling axonal degeneration within the anterior visual system: implications for demonstrating neuroprotection in multiple sclerosis. *Arch Neurol*, 65(1):26–35, 2008.

[255] A. Green, S. McQuaid, S. Hauser, I. Allen, and R. Lyness. Ocular pathology in multiple sclerosis: retinal atrophy and inflammation irrespective of disease duration. *Brain*, 21, 2010.

[256] J. Sepulcre. Contribution of white matter lesions to grey matter atrophy in multiple sclerosis: Evidence from voxel-based analysis of T1 lesions in the visual pathway. *Arch Neurol*, 66(2):173–179, 2009.

[257] P. Villoslada, A. Cuneo, J. Gelfand, S. Hauser, and A. Green. Color vision is strongly associated with retinal thinning in multiple sclerosis. *Mult Scler*, 18(7):991–999, 2012.

[258] J. Osterloh. sSarm/Sarm l is required for activation of an injury-induced axon death pathway. *Science*, 337:481–484, 2012.

[259] L. Chen, M. Stone, J. Tao, and M. Rolls. Axon injury and stress trigger a microtubule-based neuroprotective pathway. *Proc Nat Acad Sci Am*, 109(29):11842–11847, 2012.

[260] I. Michaelevski, K. Medzihradszky, A. Lynn, A. Burlingame, and M. Fainzilber. Axonal transport proteomics reveals mobilization of translation machinery to the lesion site in injured sciatic nerve. *Mol Cell Proteomics*, 9(5):976–987, 2010.

[261] F. Sun. Sustained axon regeneration induced by co-deletion of PTEN and SOCS3. *Nature*, 480:372–375, 2011.

[262] F. Bareyre. In vivo imaging reveals a phase-specific role of STAT3 during central and peripheral nervous system axon regeneration. *Proc Natl Acad Sci U S A*, 108(15):6282–6287, 2011.

[263] M. Coleman and M. Freeman. Wallerian degeneration, wld(s), and nmnat. *Annu Rev Neurosci*, 33:246–267, 2010.

[264] M. Coleman. Axon degeneration mechanisms: commonality amid diversity. *Nat Rev Neurosci*, 6(11):889–898, 2005.

[265] J. Glenn. Confocal Raman microscopy can quantify advanced glycation end product (AGE) modifications in Bruch's membrane leading to accurate, nondestructive prediction of ocular aging. *FASEB J*, 21(13):3542–3552, 2007.

[266] J. Beattie, S. Brockbank, J. McGarvey, and W. Curry. Raman microscopy of porcine inner retinal layers from the area centralis. *Mol Vis*, 13:1106–1113, 2007.

[267] I. Ermakov, M. Sharifzadeh, M. Ermakova, and W. Gellermann. Resonance Raman detection of carotenoid antioxidants in living human tissue. *J. Biomed Opt*, 10(6):064028, 2005.

[268] Y. Fu, W. Sun, Y. Shi, and J. Cheng. Glutamate excitotoxicity inflicts paranodal myelin splitting and retraction. *PloS ONE*, 4(8):e6705, 2009.

[269] I. Patel. High contrast images of uterine tissue derived using Raman microspectroscopy with the empty modelling approach of multivariate curve resolution-alternating least squares. *Analyst*, 136(23):4950–4959, 2011.

[270] M. Garrido, F. Rius, and M. Larrechi. Multivariate curve resolution-alternating least squares (mcr-als) applied to spectroscopic data from monitoring chemical reactions processes. *Annal Bioanal Chem*, 390(8):2059–2066, 2008.

[271] A. Di Penta. Oxidative stress and proinflammatory cytokines contribute to demyelination and axonal damage in a cerebellar culture model of neuroinflammation. *PloSONE*, pages 2059–2066, 2013.

[272] L. Rover. Study of nadh stability using ultraviolet-visible spectrophotometric analysis and factorial design. *Anal Biochem*, 260:50–55, 1998.

[273] K. Mertsch, U. Hanisch, H. Kettenmann, and J. Schnitzer. Characterization of microglial cells and their response to stimulation in an organotypic retinal culture system. *J Comp Neurol*, 431(2):217–227, 2001.

[274] C. Broderick, L. Duncan, N. Taylor, and A. Dick. IFN-gamma and LPS-mediated IL-10-dependent suppression of retinal microglial activation. *Invest Ophthalmol Vis Sci*, 41(9):2613–2622, 2000.

[275] W. Wang, G. Xu, J. Tian, and A. Sprecher. Inhibitory effect on LPS-induced retinal microglial activation of downregulation of t-PA expression by siRNA interference. *Curr Eye Res*, 34(6):476–484, 2009.

[276] R. Palacios. A network analysis of the human T-cell activation gene network identifies JAGGED1 as a therapeutic target for autoimmune diseases. *PLoS ONE*, 2(11), 2007.

[277] T. Romer. Histopathology of human coronary atherosclerosis by quantifying its chemical composition with Raman spectroscopy. *Circulation*, 97(9):879–885, 1998.

[278] M. Hedegaard. Discriminating isogenic cancer cells and identifying altered unsaturated fatty acid content as associated with metastasis status, using k-means clustering and partial least squares-discriminant analysis of Raman maps. *Anal Chem*, 82(7):2797–2802, 2010.

[279] A. Ajo. On the refractive index of the retina. *Acta Physiol Scand*, 13:130–149, 1947.

[280] I. Galeaa. A web-based tool for personalized prediction of long-term disease course in patients with multiple sclerosis. *Eur J Neurol*, in press:2059–2066, 2012.

[281] P. Guthikonda, J. Baker, and D. Mattson. Interferon-beta-1-b (IFN-B) decreases induced nitric oxide (NO) production by a human astrocytoma cell line. *J Neuroimmunol*, 82(2):133–139, 1998.

[282] L. Hua, J. Liu, C. Brosnan, and S. Lee. Selective inhibition of human glial inducible nitric oxide synthase by interferon-beta: implications for multiple sclerosis. *Ann Neurol*, 43(3):384–387, 1998.

[283] V. Stewart. Pretreatment of astrocytes with interferon-alpha/beta impairs interferon-gamma induction of nitric oxide synthase. *J Neurochem*, 68(6):2547–2551, 1997.

[284] D. Chen. Classical Raman spectroscopic studies of nadh and nad+ bound to liver alcohol dehydrogenase by difference techniques. *Biochemistry*, 26(15):4776–4784, 1987.

[285] H. Vries. Nrf2-induced antioxidant protection: a promising target to counteract ROS-mediated damage in neurodegenerative disease? *Free Radic Biol Med*, 45(10):1375–1383, 2008.

[286] L. Steinman. Inflammatory cytokines at the summits of pathological signal cascades in brain diseases. *Sci Signal*, 6, 2013.

[287] M. Gomez-Lazaro. Reactive oxygen species and p38 mitogen-activated protein kinase activate bax to induce mitochondrial cytochrome c release and apoptosis in response to malonate. *Mol Pharmacol*, 71(3):736–743, 2007.

[288] W. Penberthy and I. Tsunoda. The importance of nad in multiple sclerosis. *Curr Pharm Des*, 15(1):64–99, 2009.

[289] L. Vecsei, L. Szalardy, F. Fulop, and J. Toldi. Kynurenines in the cns: recent advances and new questions. *Curr Pharm Des*, 12(1):64–82, 2012.

[290] J. Mertin and R. Hunt. Hyperphenylalaninaemia and experimental allergic encephalomyelitis. *J Neurol Sci*, 29(2):351–359, 1976.

[291] P. Ho. Identification of naturally occurring fatty acids of the myelin sheath that resolve neuroinflammation. *Sci Transl Med*, 4(137):137–173, 2012.

[292] M. Wang, W. Ma, L. Zhao, R. Fariss, and W. Wong. Adaptive muller cell responses to microglial activation mediate neuroprotection and coordinate inflammation in the retina. *J Neuroinflammation*, 8:173, 2011.

[293] J. Chan, D. Taylor, T. Zwerdling, S. Lane, and K. Ihara. Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophys. J.*, 90:648–656, 2006.

[294] W. Huang, R. Griffiths, I. Thompson, M. Bailey, and A. Whiteley. Raman microscopic analysis of single microbial cells. *Anal. Chem.*, 76:4452–4458, 2004.

[295] M. Hedegaard, C. Krafft, H. Ditzel, L. Johansen, S. Hassing, and J. Popp. Discriminating isogenic cancer cells and identifying altered unsaturated fatty acid content as associated with metastasis status. using k-means clustering and partial least squares-discriminant analysis of Raman maps. *Anal. Chem.*, 82:2797–2802, 2010.

[296] M. Hedegaard, C. Matthaus, S. Hassing, C. Krafft, M. Diem, and J. Popp. Spectral unmixing and clustering algorithms for assessment of single cells by Raman microscopic imaging. *Theor. Chem. Acc.*, 130:1249–1260, 2011.

[297] A. Harris, M. Garg, X. Yang, S. Fisher, J. Kirkham, D. Smith, D. Martin, and A. High. Raman spectroscopy and advanced mathematical modelling in the discrimination of human tyroid cell lines. *Head Neck Onc.*, 38:1–6, 2009.

[298] S. Balint, S. Rao, M. Marro, P. Miskovsky, and D. Petrov. Monitoring of local ph in photodynamic therapy-trated live cancer cells using surface enhanced Raman scattering probes. *J. Raman Spectrosc.*, 2010.

[299] H. Bonnier, P. Knief, A. Meade, J. Dorney, and K. Bhattacharya. Collagen matrices as an improved model for in vitro study of live cells using Raman microspectroscopy. *Proc. SPIE*, 80870F, 2011.

[300] X. Lu and Y. Kang. Hypoxia and hypoxia-inducible factors: master regulators of metastasis. *Clin Cancer Res*, 16(24):5928–5935, 2010.

[301] O. Mendez, Y. Fernandez, M. Peinado, V. Moreno, and A. Sierra. Anti-apoptotic proteins induce non-random genetic alterations that result in selecting breast cancer metastatic cells. *Clin Exp Metastasis*, 22:297–307, 2005.

[302] A. Chambers. Mda-mb-435 and m14 cell lines: identical but not m14 melanoma? *Cancer Res*, 69:5292–5293, 2009.

[303] D. Read. *Handbook of hydrocarbon and lipid microbiology.* Springer, Berlin, 2010.

[304] http://lid.phys.msu.su/englishversion/science/water.html.

[305] S. Ali, F. Bonnier, A. Tfayli, H. Lambkin, and K. Flynn. Raman spectroscopic analysis of human skin tissue sections ex-vivo: evaluation of the effects of tissue processing and dewaxing. *J Biomed Optics*, 18(6):061202, 2012.

[306] G. Zhu, X. Zhu, Q. Fan, and X. Wan. Raman spectra of amino acids and their aqueous solutions. *Spectrochimica Acta Part A*, 78:1187–1195, 2011.

[307] I. Notingher. Raman spectroscopy cell-based biosensors. *Sensors*, 7:1343–1358, 2007.

[308] H. Tang, H. Yao, G. Wang, Y. Li, and M. Feng. Nir Raman spectroscopic investigation of single mitochondria trapped by optical tweezers. *Optics Express*, 15(20):12708–12716, 2007.

[309] M. Miljkovic, T. Chernenko, M. Romeo, B. Bird, C. Matthaus, and M. Diem. Label-free imaging of human cells: algorithms for image reconstruction of Raman hyperspectral datasets. *Analyst*, 135:2002–2013, 2010.

[310] F. Adar and M. Erecinska. Resonance raman spectra of whole mitochondria. *Biochemistry*, 17(25):5484–5488, 1978.

[311] C. Matthaus, T. Chernenko, J. Newmark, C. Warner, and M. Diem. Label-free detection of mitochondrial distribution in cells by non-resonant raman microspectroscopy. *Biophys J*, 93:668–673, 2007.

[312] K. Fujita and I. Smith. Label-free molecular imaging of living cells. *Mol cells*, 26:530–535, 2008.

[313] M. Okada, N. Smith, A. Palonpon, H. Endo, and S. Kawata. Label-free raman observation of cytochrome c dynamics during apoptosis. *PNAS*, 109:28–32, 2012.

[314] K. Hamada, K. Fujita, N. Smith, M. Kobayashi, Y. Inouye, and S. Kawat. Raman microscopy for dynamic molecular imaging of living cells. *J Biom Opt*, 13(4):0440271– 0440274, 2008.

[315] L. Chiu, M. Ando, and H. Hamaguchi. Study of the raman spectroscopic signature of life in mitochondria isolated from budding yeast. *J Raman Spect*, 41:2–3, 2010.

[316] U. Neugebauer, A. Marz, T. Henkel, M. Schmitt, and J. Popp. Spectroscopic detection and quantification of heme and heme degradation products. *Anal Bioanal Chem*, 404:2819–2829, 2012.

[317] R. Draga, M. Grimbergen, P. Vijverberg, C. van Swol, T. Jonges, J. Kummer, and J. Bosch. In vivo bladder cancer diagnosis by high-volume Raman spectroscopy. *Anal Chem*, 82:5993–5999, 2010.

[318] S. Bonnet, SL. Archer, and J. Allalunis-Turner. A mitochondria-k+ channel axis is suppressed in cancer and its normalization promotes apoptosis and inhibits cancer growth. *Cancer Cell*, 11:37–51, 2007.

[319] M. Ralser, M. Wamelink, S. Latkolik, E. Jansen, H. Lehrach, and C. Jakobs. Metabolic reconfiguration precedes transcriptional regulation in the antioxidant response. *Nat Biotechnol*, 27:604–605, 2009.

[320] Y. Jang and S. Sharkis. A low level of reactive oxygen species selects for primitive hematopoietic stem cells that may reside in the low-oxygenic niche. *Blood*, 110:3056–3063, 2007.

[321] G. Szabadkai and M. Duchen. Mitochondria: the hub of cellular ca2b signaling. *Physiology*, 23:84–94, 2008.

[322] Y. Chen, H. Chou, P. Lyu, H. Yin, F. Huang, and W. Chang. Mitochondrial proteomics analysis of tumorigenic and metastatic breast cancer markers. *Funct. Integr Genomics*, 11:228–239, 2011.

[323] J. Zhao, U. Lendahl, and M. Nister. Regulation of mitochondrial dynamics: convergences and divergences between yeast and vertebrates. *Cell Mol Life Sci*, 70(6):951–976, 2013.

[324] M. Olahova, S. Taylor, S. Khazaipoul, J. Wang, B. Morgan, and K. Matsumoto. A redox-sensitive peroxiredoxin that is important for longevity has tissue- and stress- specific roles in stress resistance. *PNAS*, 105:19839–19844, 2008.

[325] http://www.nature.com/scitable/topicpage/mitochondria-14053590.