2010-01-01

Imaging Live Cells Grown on a Three Dimensional Collagen Matrix Using Raman Microspectroscopy

Franck Bonnier
Technological University Dublin, Franck.Bonnier@tudublin.ie

Peter Knief
Technological University Dublin, peter.knief@tudublin.ie

Basil Lim
Technological University Dublin, basil.lim@tudublin.ie

See next page for additional authors

Authors

Franck Bonnier, Peter Knief, Basil Lim, Aidan Meade, Jennifer Dorney, Kunal Bhattacharya, Fiona Lyng, and Hugh Byrne

Imaging live cells grown on a three dimensional collagen matrix using Raman microspectroscopy

F. Bonnier[1*], P.Knief[1], B. Lim[2], A.D.Meade[2], J. Dorney[1], K. Bhattacharya[1], F.M. Lyng[1], H.J. Byrne[1]

1. *Focas Research Institute, Dublin Institute of Technology (DIT), Camden Row, Dublin 8, Ireland*
2. *School of Physics, Dublin Institute of Technology (DIT), Kevin Street, Dublin 8, Ireland*

Abstract

Three dimensional collagen gels have been used as matrices for the imaging of live cells by Raman spectroscopy. The study is conducted on a human lung adenocarcinoma (A549) and a spontaneously immortalized human epithelial keratinocyte (HaCaT) cell line. The lateral resolution of the system has been estimated to be <1.5 μm making it possible to access the subcellular organization. Using K-means clustering analysis, it is shown that the different subcellular compartments of individual cells can be indentified and differentiated. The biochemical specificity of the information contained in the Raman spectra allows the visualization of differences in the molecular signature of the different sub-cellular structures. Furthermore, to enhance the chemical information obtained from the spectra, principal component analysis has been employed, allowing the identification of spectral windows with a high variability. The comparison between the loadings calculated and spectra from pure biochemical compounds enables the correlation of the variations observed with the molecular content of the different cellular compartments.

---

[*] Corresponding author. *Tel.:* +353 1 4027917; *Fax*: +353 1 4027904;
 E-mail address: fbonnier@dit.ie (F. Bonnier)

1. Introduction

The evolution from 2D to 3D *in vitro* tissue models has been an important development in the understanding of the different mechanisms involved in tumour formation. In an environment which better mimics the extracellular conditions encountered by cells *in vivo*, it has been possible to access more accurate information about cell homeostasis, differentiation and migration[1, 2]. It is increasingly apparent that the basic 2D cell culture models commonly used fail to provide the cell a microenvironment for adequate adhesion and proliferation resulting in discrepancies between observations made *in vitro* and *in vivo*, especially for the testing of potential anticancer drugs[3].

Over the last decade, new models such as collagen gels and more complex multicomponent systems such as Matrigel have emerged, providing the cells a matrix which more accurately reproduces the extra cellular matrix (ECM) [4-6]. It has been demonstrated in various studies that the behaviour of cells cultured in such an environment differs considerably compared to those cultured on traditional 2D surfaces[7]. Thus, the interaction between cells and a matrix mirroring the ECM encountered *in vivo* has various effects on their phenotype and metabolism. Wang et al. published a study in 1998 showing that antibodies against B1-integrins exhibit different behaviours when tested in 2D compared to 3D models[8]. More recently, Harisi et al. described the effect on induced doxorubicin-resistance by the extra cellular matrix in human osteosarcoma cells[9]. This observation was confirmed in 2008 by Fourre et al., demonstrating the protective effect of the extracellular matrix on the HT1080 cell line, against the action of doxorobucin[10]. Reduced radiation induced toxicity when cells are grown in a 3D environment has also been recently

demonstrated[11]. Numerous other publications can be found in the literature describing modification of the cell phenotype, metabolism or composition when grown in a 3D matrix, compared to 2-D environments[12, 13].

The use of 3D matrices for cell culture and analysis is still relatively recent, however, and many improvements have to be made before they can be routinely employed for cellular analysis. For instance, the pH used during the polymerisation of collagen gels can influence the fibre diameter which has repercussions on the cellular behaviour[14], showing a need for the optimisation of the biocompatibility. Furthermore, microscopic techniques routinely used for assessing morphological and physiological changes of cells grown on 2D models could be limited in a 3D microenvironment due to the different optical properties (scattering, distortion of the light) of thicker systems. Due to their increased depth resolution, confocal microscopy or multiphoton microscopy would be preferred for such models[15].

Raman microspectroscopy presents several advantages for the study of live cells, combining molecular analysis with optical imaging. In comparison to infrared spectroscopic analysis, the weak contribution from water offers the possibility to study the cells in an aqueous environment and thus to keep them alive for the duration of the measurement[16]. The specific information contained in the Raman spectra provides a signature of the samples studied which can be related to molecular content or changes to the physiology as a result of external stimuli [17, 18]. The spatial resolution is of the order of 1-2 μm, providing access to the subcellular organisation of the cells at a molecular level[19, 20].

Optically transparent substrates such as CaF_2, Quartz or ZnSe are commonly used for vibrational spectroscopy but none are adapted for cell culture and each can have toxic effects resulting in death or modification of the metabolism of cells. Improved

biocompatibility for biospectroscopic applications can be achieved by the deposition of a thin layer of gelatine and a thin enough layer does not contribute significantly to the spectral signature[21]. In terms of cellular viability and spectral quality, however, collagen gels have been recently demonstrated to be a substantially improved substrate for live cells analysis using Raman spectroscopy. The gels provide a 3-D matrix for cell growth, although it should be noted that the cells are cultured on the surface of the matrix. As measured using the Alamar Blue cytotoxicity assay, in comparison to uncoated quartz, a human lung cell line exhibited a 39% increase in viability in collagen gels whereas as dermal cell line seemed to be more affected by the different substrates and a 285% increase in viability was observed. Furthermore, after polymerization, the density of the collagen is too low to give any contribution in the spectra recorded[22]. Also no contribution from the medium used for the cell culture is observable in the spectra recorded which allows the cells to be kept alive for prolonged periods and spectroscopically monitored, effectively in real time.

Previous studies have demonstrated the capacity of Raman spectroscopy to map single cells and their subcellular structures. The identification of the sub-cellular structures such as the nucleus, cytoplasm and mitochondria can be achieved[23] and chemical maps can be constructed giving information about the distribution of differentiating molecules such as proteins[24]. However, these have been performed on chemically fixed cells and it has since been demonstrated that chemical fixation induces significant changes to the biochemistry of the cell[25, 26]. Ultimately, if subtle changes in metabolism due to for example chemotherapeutic agents, toxicants, or radiation are to be discerned, live cells should be studied.

In this study, the potential of Raman spectroscopy as an imaging tool for single live cells grown on collagen gel substrates is evaluated. As Raman spectroscopy has been

demonstrated to be a powerful technique for the analysis of tissue and cells [18, 27-29], and because of the high spatial resolution, can potentially be employed to monitor and analyse cellular physiology on a molecular level, as well as changes to it due to external stimuli [30-33] with subcellular resolution, it is critical to demonstrate its performance on cells in a 3-D matrix. This work demonstrates spectral imaging of live cells in collagen matrices using Raman microspectroscopy. Prolonged imaging times are afforded due to the increased viability of the cells in the matrices. Data processing is significantly reduced due to reduced substrate contributions. The technique is demonstrated to provide molecular fingerprints of the subcellular structure.

2. Materials and methods

2.1. A549 and HaCaT Cell lines

A549 cells from a human lung adenocarcinoma with the alveolar type II phenotype were obtained from ATTC (Manassas, VA, USA). Cells were cultured in RPMI 1640 (Gibco, Karlsruhe) supplemented with L-glutamine, penicillin and streptomycin (Gibco) and 10% foetal calf serum (FCS, Biochrom, Berlin) in a humidified atmosphere containing 5% CO_2 at 37ºC.

HaCaT cells are a spontaneously immortalized human epithelial keratinocyte cell line derived from adult skin [34]. They are aneuploid, non-tumourigenic and p53 mutated adult keratinocytes. HaCaT cells were cultured in Dulbecco's MEM: F12 (1:1) medium (Sigma, Dorset, U.K.) containing 10% fetal calf serum (Gibco, Irvine, U.K.) 1% penicillin-streptomycin solution 1000 IU (Gibco, Irvine, U.K.), 2mM L-glutamine

(Gibco, Irvine, UK) and 1 μg/mL hydrocortisone (Sigma, dorset,UK) in a humidified atmosphere containing 5% CO_2 at 37 °C.

2.2. Collagen gels

Solutions of collagen I from rat tail tendons (Gibco) were used for preparation of the collagen gels. The 5mg/mL solution was mixed with sterile 10X phosphate buffered saline (PBS), sterile distilled water (dH_2O) and 1M NaOH. The appropriate relative quantities of these components is determined by the final concentration (3 mg/mL) and volume needed. All the steps were carried out on ice to slow the gelation process. After mixing, the solution was placed in a Petri-dish for Raman spectroscopy before incubation at 37 °C degrees in a 95% humidity incubator until a solid gel is formed (about 30mins).

2.3. Sample preparation

Cell samples: A549 and HaCaT cells were loaded at a concentration of 4×10^4 cells, onto collagen gels (prepared as described above). To facilitate the manipulation of the samples, 30 mm Petri dishes (Stardest, USA) have been preferred for Raman spectroscopic measurements. All samples were incubated for 24H at 37°C, 5% CO_2 before measurements.

Iron Oxide particle samples have been prepared to estimate the best lateral resolution to be used for the mapping of living cells: First, glass slides were coated using a poly L lysine solution (Sigma Aldrich). Then, a drop of Iron oxide suspension was placed on the coated glass slides. After 30 min incubation at 37°C and 100% humidity, the

particles were firmly attached. The substrate was rinsed in water and kept in a Petri dish, immersed in distilled water until measurement.

Chemical samples: Four different compounds were prepared for Raman spectroscopy analysis: Raman spectra have been recorded from lyophilised DeoxyriboNucleic Acid from calf thymus deposited on CaF_2 substrates. Ribonucleic acid from baker's yeast (S. cerevisiae) was first suspended in water and deposited onto CaF_2 substrates for drying before recording. Sphingomyelin from bovine brain and L-α-Phosphatidylcholine from egg yolk were first dispersed in into chloroform and then deposited onto CaF_2 substrates for drying before spectral acquisition.

2.3 Observation of live cells

In order to compare the cell morphology, samples were imaged using a confocal microscope (LSM 510 META, Zeiss, Germany) equipped with a x20 objective.

2.4. Raman spectroscopic measurements

A Horiba Jobin-Yvon LabRAM HR800 spectrometer with an external 300 mW single mode diode laser operating at 785 nm as source was used throughout this work. For all measurements, a x100 immersion objective (LUMPlanF1, Olympus, N.A. 1.00W) was employed The confocal hole was set at 100μm for all measurements, the specified setting for confocal operation. The system was pre-calibrated to the 520.7 cm^{-1} spectral line of Silicon. The Labram system is a confocal spectrometer that contains two interchangeable gratings (300 and 900 lines/mm respectively). In the following experiments the 300 lines/mm grating was used, which gave a spectral

dispersion of around 1.5 cm^{-1} per pixel. The detector used was a 16-bit dynamic range Peltier cooled CCD detector. Images of the sample were acquired using a video camera within the system.

Lateral resolution testing: Iron oxide particles of 1 μm diameter, deposited on glass slides as described above, were selected for measurement. Linescans through the particles were conducted with a step size of 0.2 μm in order to map out the convolution of the focal profile and particle size.

Live cell mapping: After 24 hours incubation, the cells were well spread out over the collagen gels and firmly attached at the surface of the substrates. Spectral imaging was performed on the surface. An area corresponding to the map to be acquired was defined around the cells in the optical image provide by the video camera. The step between two successive measurements was set to 1.5 μm (in accordance with the results of the lateral resolution determination) and the backscattered Raman signal was integrated for 10 seconds over the spectral ranges from 400 to 1800 cm^{-1} and accumulated twice to improve the signal to noise ratio.

2.5. Data Analysis

The different data analysis steps were performed using Matlab (Mathworks, USA). Before statistical analysis, a Savitsky-Golay filter (5th order, 7 points) was applied to smooth the spectra and the reference spectrum constituting the background signal was subtracted.

Principal component analysis (PCA) is a method of multivariate analysis widely used with datasets of multiple dimensions [35]. It allows the reduction of the number of variables in a multidimensional dataset, although it retains most of the variation

within the dataset. The order of the PCs denotes their importance to the dataset. PC_1 describes the highest amount of variation, PC_2 the second greatest and so on. Therefore, var $(PC_1) \geq$ var $(PC_2) \geq$ var (PC_p), where var (PC_i) represents the variance of PC_i in the considered data set. Generally, the 3 first components represent more than 90% of the variance. This statistical method was preferred for this study to highlight the variability existing in the spectral data set recorded during the different experiments. Another advantage of this method is the observation of loadings which represent the variance for each variable (wavenumber) for a given PC. Analysing the loadings of a PC can give information about the source of the variability inside a data set, derived from variations in the molecular components contributing to the spectra.

K-means clustering analysis is one of the simplest unsupervised learning algorithms used for spectral image analysis. It groups the spectra according to their similarity, forming clusters, each one representing regions of the image with identical molecular properties[36]. The distribution of chemical similarity can then be visualised across the sample image. The number of clusters (k) has to be determined a priori by the operator before initiation of the classification of the data set. K centroids are defined, ideally as far as possible from each other, and then each point belonging to a data set is associated to the nearest centroid. When all the points have been associated with a centroid, the initial grouping is done. The second step consists of the calculation of new centroids as barycentres of the clusters resulting from the previous step. A new grouping is implemented between the same data points and the new centroids. These operations are repeated until convergence is reached and there is no further movement of the centroids. Finally k clusters are determined, each containing the most similar spectra from the image. From here, colours can be attributed to each cluster and false

colours maps can be constructed to visualise the organisation of the clusters in the original image.

3. Results and discussion

3.1 Observation of live cells

Figure 1 presents optical images of both the A549 and HaCaT cell lines using a confocal microscope (LSM 510 META, Zeiss, Germany). The images A,B,D and E were obtained from cells grown in a Petri-dish (plastic substrate). Both cell types exhibit similar size but different morphologies. The A549 cells (A and B) grow individually whereas the HaCaT cells (D and E) remain strongly attached to each other and seem to form colonies of cells. In both cases, the cell membrane is easily identified and different subcellular compartments can be observed. The nucleus appears as a circular unit surrounded by the cytoplasm. Inside the nucleus, dark elements can be observed, corresponding to the nucleoli. The main difference between the cell lines is the size of the nucleus, being slightly bigger in the case of the HaCaT cells line, making the cytoplasm smaller. The organisation of the nucleoli differs between the two cell lines. In A549 cells, usually 2 or 3 nucleoli are clearly visible whereas in the case of the HaCaT cells, the number of nucleoli appears to be higher and more variable (Fig 1 B and E). Similar observations can be made concerning the cells grown on collagen gels (Fig.1 C and F). Although, the thickness of the collagen gels could influence the contrast of the images, the cell membrane, the nucleus and nucleoli are still clearly identifiable. The overall morphology of the cells

on collagen also appears to differ, in both cases, indicating better adhesion to the matrix.

The aim of this study was to evaluate the potential of collagen gel substrates for the imaging of live cells using Raman microspectroscopy. Different compartments can be seen in the cells using optical microscopy but the main challenge was to be able to differentiate and categorise them using the biochemical information contained in the Raman spectra recorded. The time needed to obtain a map from the live cells is the main concern. In order to get a signal to noise ratio sufficient to clearly identify the variations between the spectra recorded from the different locations, an acquisition time of twice 10 seconds was required. Thus, to optimise the total time of measurement for the mapping of a single cell, it was necessary to estimate the optimal lateral resolution of the system beforehand. For this purpose, a protocol proposed by Puppels et al. was used[37]. Iron oxide particles were deposited onto poly-L-Lysine coated glass slides and immersed into water for recording. The particles firmly attached to the glass slide were immersed in water to reproduce the conditions of the measurements performed on the live cells. Particles of 1μm diameter were selected and a linescan with a 0.2 μm step size was recorded through the particles. The iron oxide particles exhibit strong peaks in the region of the spectrum from 400-650 cm^{-1}. The peak positioned at 610 cm^{-1} was selected for the estimation of the lateral resolution. The intensity of this peak was monitored for each different position of the linescan. The lateral resolution is determined by the full width at half intensity, which was found to be 1.5 μm with the settings used for the mapping. It should be noted that this is in fact a measurement of the convolution of the particle size and the spotsize, and so is an underestimate of the true spatial resolution achievable.

11

3.2 Mapping of live cells grown on collagen gels

After growing the A549 and HaCaT cells for 24 hours on the collagen gels, the medium was substituted by NaCl before mapping. The feasibility of recording Raman spectra from cells in the presence of the culture medium without interference to the spectra has been reported[22]. However, the main limitations concerning the collagen gels seem to be the luminosity and the contrast when observing the cells by white light in transmission. Although the collagen gels are prepared in Petri dishes with a glass bottom, the light has to travel through the glass, the collagen gel, the cell and culture medium (containing phenol red with a strong absorption at ~560nm) and finally the aqueous solution used to keep the cell immersed, before reaching the camera. Two parameters can be adjusted, the thickness of the collagen and the aqueous solution used. The quantity of collagen used has been reduced as much as possible and 1mL per Petri dish was determined to be the smallest amount of collagen that could possibly be used for the experiments. When using a smaller quantity, different problems were encountered regarding the loss of the stability of the collagen gels which could detached from the Petri dish when changing the medium, or the observation of a contribution from the Petri dish in the spectra due to a too thin collagen matrix. Therefore, to improve the brightness and the contrast it was decided to substitute the culture medium by a 0.9% NaCl solution which is uncoloured and maintains the cells alive for the duration of the measurement. It has been observed that when grown on collagen gels, after 4 to 5 hours in saline solution, the cells don't exhibit any signs of stress due to the lack of medium. Examples of photographs obtained using the Labram HR800 equipped with a x100 immersion objective are presented in figure 3A and 3C. The different structures of the cell can be identified;

12

the cell membrane is clearly defined for both cell lines and the nucleus is located in the centre part of the cells. As observed previously, the cytoplasm appears smaller in the case of the HaCaT cell. Again, the main difference between the two cell lines is the organisation in the nucleus. When observing the A549 cells, the nucleoli are clearly visible inside the nucleus but in the case of the HaCaT cell, no well defined shapes can be observed under these conditions. Another observation is the presence of 2 different organelles in the cytoplasm of the A549 cell, not clearly visible on the image obtained with the confocal microscope (Fig.1B) but now visible with the x100 immersion objective. This observation is not surprising as the cytoplasm of the cells has a complex organisation and different structures can be found such as the actin or microtubules[38, 39], endoplasmic reticulum[39, 40], the mitochondria[41] or the Golgi apparatus[42, 43].

3.3 K-means clustering Analysis

In order to visualise these different structures in the spectral maps recorded, K-means clustering analysis was employed. K-means clustering analysis aims to gather the spectra with the most similarities inside the same clusters, the number of clusters being defined by the operator before the analysis. The data pre-processing was minimised in order not to influence the information contained in the data set. A spectrum corresponding to the background (signal of the collagen gel) was first subtracted from the whole data set. Examples of spectra corresponding to the background and cellular spectrum are given in the figure 2A1 and 2A2. The signal of the collagen doesn't contain any peaks and represents the contribution of the water present in the gel and in the aqueous solution used to keep the cells alive [22]. Before

13

removal of the background signal, the spectra were subjected to a light smoothing (Savitzky–Golay, 13 pts). Thus, spectra of the substrate showed little point to point variation and were not excluded from the analysis. Although, the raw spectra corresponding to the background (the figure 2A1) exhibit a low signal to noise ratio, the subtraction of such spectra from the data set will result in the creation of more noise in the corrected spectra. To avoid any deterioration of the quality of the spectra recorded from the live cells, the signal obtained from the collagen gel was extensively smoothed before subtraction (Savitzky–Golay, 41 pts). Thus, the signal to noise ratio of the data set is unchanged after the pre-processing. After subtraction of the background signal, the baseline of the spectra is already greatly improved (figure2B) and only a slight deviation in the baseline is noticeable between the two extremities of the spectrum. A linear baseline, as illustrated by the dotted line in figure 2B, was subsequently subtracted from all spectra in the data set. Figure 2C represents an example spectrum obtained after the different steps of the pre-processing. After subtraction of background and correction of the baseline, the spectrum exhibits a good signal to noise ration with the presence of well defined peaks all along the spectral range 400-1800 cm^{-1}. Because the signal from the collagen gel is weak and constant, it has been possible to standardise the background removal from the data set. Nevertheless, due to variation in the density related to different structures present in the live cells, slight fluctuation in the intensity of the spectra can be seen. A normalisation has been finally applied to correct for this.

K-means clustering analysis is currently commonly used for the analysis of tissue sections by infrared and Raman spectroscopy, especially for the detection of the different structures present in the samples studied and the identification of tumoral regions[44-46]. K-means clustering can also be employed for the study of individual cells

by Raman spectroscopy but only a few studies have elucidated the subcellular organisation and they are usually based on fixed cells grown on a non optimal substrate such a quartz window or equivalent [23, 47, 48]. The effects of cell fixation on the cellular composition has been already investigated[25, 26] and thus it is desirable to be able to obtain the Raman signature from live cells, avoiding any structural and molecular modification which may occur as a result of the chemical processing. Moreover, the use of collagen gels enables extended measurement periods and thus more detailed mapping of live cells. In contrast to quartz substrates which have a toxic effect necessitating fixation of the cells for prolonged measurement times, collagen gels offer a better adhesion of the cells to the substrate, enhancing the cell viability.

Within the cells, it may be predicted that distinct biochemical regions can be identified, corresponding to, for example, nucleus, nucleoli, cytoplasm and cell membrane. However, in analysing the spectral response of the cells, it is observed that the spectra recorded from the edge of the cells contain a high variability due to the transition from the membrane to the substrate, resulting in quite noisy spectra. When the K-means analysis is performed, the variability between these spectra can interfere with the analysis, resulting in the creation of new clusters. Thus, in order to best visualise the different structures existing inside the cells the numbers, the number of clusters has to be increased to 10 groups, which produced the best reproducibility. In this way, the variability in the spectra obtained at the edge of the cells are contained in different clusters and do not interfere with the identifications of the different structures present within the cells. The initiation step was selected randomly for each iteration and it was seen to have no effect on the final results of analysis. After analysis, the spectra were gathered in the 10 different groups according to their

15

similarities and a colour (arbitrarily) attributed to each cluster. False colour maps were then constructed representing the partition of the different clusters in the cells. The resulting images are presented in figure 3B and 3D. A correspondence between the organisation of the clusters and the structure present in the cells can be seen. In figure 3B, specific clusters relating to the cytoplasm, nucleus and nucleoli appear. The specificity of the information contained in the spectra recorded allows discrimination of the different cellular compartments using K-means analysis. The sensitivity of the analysis furthermore allows visualization of the two different compartments present in the cytoplasm of the A549 cell and identifies a specific cluster related to the nucleoli in the HaCaT cell, although they are not clearly seen in the visible image.

K-means clustering analysis appears to be a powerful tool to identify the different structures present in the cells. The false colour maps give a representation of the partition of the different clusters but do not provide any information about the chemical composition of the samples. This biochemical information and the variations between the different clusters can be visualised through the mean spectra of the clusters. Figure 4 presents the average spectra obtained from selected clusters of the K-means analysis performed on an A549 cell (Fig 3B). (A: nucleus from the cluster number 4, B: nucleolus from the cluster number 6, C: cytoplasm from the cluster number 5 and D: cytoplasm from the cluster number 7). The other clusters correspond to the spectra recorded from the collagen gel and those located at the edge between the cell and the substrate (data not shown). Variations can be observed in different spectral windows according to the different clusters (highlighted in grey in figure 4).

The characterisation of live cells on 2D substrates has been already been undertaken by others [17, 18, 49-51] and, based on literature; assignments of the different peaks can be performed. For example, the variations in the spectral range 718 -725 cm^{-1} are related

to the lipids and nucleic acids, between 815 and 830 cm^{-1} to RNA and proteins, from 1088 cm^{-1} to 1102 cm^{-1} to the nucleic acids, the lipids and carbohydrates, from 1243-1266 cm^{-1} to the nucleic acids, the proteins and lipids, from 1296-1322 cm^{-1} to the lipids and from 1315-1348 cm^{-1} to the nucleic acids and proteins. In analysing the cluster average spectra, it is important to note that, with a ~1.5 μm spotsize, while the nucleus can be specifically targeted, the overlying cellular membrane and cytoplasm will also contribute to the spectrum acquired. Similarly, the spectrum of the cytoplasm will contain contributions from the overlying cellular membrane and potentially also from subcellular organelles such as mitocondria, lysosomes, etc.. Across, the spectral range, the variations between the different regions are thus subtle. However, the regions do clearly cluster by K-means analysis, and for example the nucleus and nucleoli are clearly differentiated from the other spectra by for example the peak at 1423 cm^{-1}, which is completely absent in the cytoplasmic spectra.

In figure 3, subcellular resolution is clearly achievable. It is noted, however, that the achievable resolution is underestimated by the technique described in 3.1 which measures the convolution of the particle size and the spot size. Furthermore, it has been reported by Uzunbajakava et al. that optimal lateral sampling is achieved using intervals 2-3 times smaller than the laser spot[24], thus promising higher definition of the subcellular features.

3.4 Principal Component Analysis

The mean spectra obtained from each cluster provide information about the spectral ranges which differentiate the sub-structures of the cells. However, it is difficult to determine precisely the differentiating features and to profile the biochemical

structure of the cellular compartment analysed. Using K-means clustering analysis, the different structures of the cells have been identified and spectra associated with the nucleus, nucleoli and cytoplasm can now be extracted from the map and analysed independently to highlight the modification in the spectral signatures. For this purpose, principal component analysis (PCA) was employed, for the case of the A549 cells for illustration purposes. Similar results are obtainable for the HaCaT cell line. Figure 5 is a representation of the plots obtained from the spectra extracted from the A549 cell map. Each spectrum has been considered as a single unit for the analysis and the difference compared to K-means analysis is that PCA is a totally unsupervised method. The result is a representation of the discrimination between the different spectra. Figure 5A is a plot of PC 1 versus PC 2 for each spectrum. The spectra from the nucleus and nucleoli form one group and are not discriminated by these principal components. Nevertheless, PC 1, which represents 34% of the explained variance, gives a good discrimination between the spectra from compartment 1 of the cytoplasm from the other spectra (this can also be seen in figure 5C) whereas PC 2, which represents 23% of the variance, allows the discrimination between compartment 2 of the cytoplasm and the other spectra. The second plot (fig.5B) represents the loadings of PC 2 versus PC 3. PC 2 again gives a clear discrimination between the spectra from compartment 2 of the cytoplasm and the other spectra, as seen in figure 5A. PC 3, which represents 3.5% of the explained variance, discriminates between the spectra from the nucleus and those from the nucleoli. The same observation can be made in figure 5C. PCA thus provides a clear discrimination between the different groups of spectra analysed. The results are not surprising as the spectra have been extracted from different clusters obtained after the K-means analysis and no additional information about the chemical composition of the samples is given in the plots.

Nevertheless, the observation of the plots is essential to identify the principal component involved in the discrimination between the different groups, as for each principal component a corresponding loadings is given after analysis.

The loading is a representation of the variability described by a principal component as a function of wavenumber of the spectra. Figure 6A shows the loading of PC1 obtained from the PCA analysis based on the A549 cell map which allows visualisation of which part of the spectrum is responsible for the discrimination between compartment 1 of the cytoplasm and the other spectra, observed in the plot of figure 5. The loading is a combination of different peaks with different intensities, the solid red line representing the zero level. It appears that the loading is composed of different sharp peaks, both positive and negative, and correspond to increased or decreased contributions of specific molecular components to the Raman spectra. The spectra in figure 6 B,C,D and E, offset for clarity, are from DNA, RNA, L-α-Phosphatidylcholine and Sphingomyelin respectively. DNA and RNA are abundantly present, especially in the nucleus, phosphatidylcholine is a major component of biological membranes and sphingomyelin is also found in animal cell membranes. To demonstrate further the origin of the spectral variations, and their clustering using PCA, the 12 peaks with highest variability have been highlighted in figure 6A. Each of them has a correspondence with one or more of the four spectra from compounds tested. For example, the peak at 791 cm^{-1} is prominent in the spectra of DNA and RNA. In figure 4, it can be seen that while all average spectra exhibit a peak in this region, it is significantly stronger in the spectra of the nucleus and nucleoli, thus giving rise to a strong positive contribution to PC1, which differentiates between nucleus/nucleoli and cytoplasm 1. The strongest variation, however, is the negative peak at 1439 cm^{-1}, which is prominent in the spectra of the L-α-Phosphatidylcholine

and Sphingomyelin indicating significant variations in lipidic content. The loading of PC2, plotted in figure 7A is the basis of the discrimination between compartment number 2 of the cytoplasm and the nucleus and nucleoli (fig 5A). Once again the comparison with the spectra of pure compounds highlights peaks matching with the loading. The highest variability can be seen at 699 cm^{-1} and 1436 cm^{-1}, both attributed to the lipid content, and at 1342 cm^{-1}, related to the DNA and RNA. The loading of PC3 (fig 8A) represents the variability between the nucleus and the nucleolus (fig 5B and 5C). Many peaks can be identified as related to the DNA and RNA especially at 651, 813, 1091,1482 and 1572 cm^{-1} showing that the main difference between the nucleus and nucleolus is the DNA/RNA content.

The loadings of the principal components provide a multivariate profile of the biomolecular variations within the cell. While a detailed analysis is not undertaken here, it is clear that a wealth of information can be derived from the Raman analysis of single cells. In contrast to previous studies, the work presented here is conducted on live cells rather than chemically preserved, and moreover the details of the spectra are greatly enhanced by the minimal need for data-preprocessing and prolonged measurement periods. These conditions are facilitated by the use of collagen 3-D matrices for cell growth and measurement.

4. Conclusion

This study demonstrates the feasibility of performing imaging of live cells grown on 3D collagen gels. This label free, non invasive study of live cells is really encouraging for future applications and opens up new perspectives. The improvement of the compatibility between the cells and the substrate used clearly enhances the relevancy

of the information obtained from such study. Increased viability means that single studies can now be extended over the cell cycle. The high specificity of the spectra recorded combined with the reduced contribution of the background allows subcellular analysis of physiological processes. Imaging of single live cells using Raman spectroscopy to access information about cellular processes such as cell migration, growth, apoptosis or probe the action of external agents such as toxicants, chemotherapeutic compounds and radiation can now be elevated to a new dimension.

1. J. B. Kim, *Seminars in cancer biology*, 2005, **15**, 365-377.
2. J. B. Kim, R. Stein and M. J. O'Hare, *Breast cancer research and treatment*, 2004, **85**, 281-291.
3. J. L. Horning, S. K. Sahoo, S. Vijayaraghavalu, S. Dimitrijevic, J. K. Vasir, T. K. Jain, A. K. Panda and V. Labhasetwar, *Molecular pharmaceutics*, 2008, **5**, 849-862.
4. V. M. Weaver, A. R. Howlett, B. Langton-Webster, O. W. Petersen and M. J. Bissell, *Seminars in cancer biology*, 1995, **6**, 175-184.
5. O. W. Petersen, L. Ronnov-Jessen, A. R. Howlett and M. J. Bissell, *Proceedings of the National Academy of Sciences of the United States of America*, 1992, **89**, 9064-9068.
6. G. D. Prestwich, *Accounts of chemical research*, 2008, **41**, 139-148.
7. K. S. Smalley, M. Lioni and M. Herlyn, *In vitro cellular & developmental biology*, 2006, **42**, 242-247.
8. F. Wang, V. M. Weaver, O. W. Petersen, C. A. Larabell, S. Dedhar, P. Briand, R. Lupu and M. J. Bissell, *Proceedings of the National Academy of Sciences of the United States of America*, 1998, **95**, 14821-14826.
9. R. Harisi, J. Dudas, J. Nagy-Olah, F. Timar, M. Szendroi and A. Jeney, *Cancer biology & therapy*, 2007, **6**, 1240-1246.
10. N. Fourre, E. Millerot-Serrurot, R. Garnotel, J. M. Zahm, N. Bonnet, J. M. Millot and P. Jeannesson, *Cancer science*, 2008, **99**, 1699-1705.
11. M. B. Sowa, W. B. Chrisler, K. D. Zens, E. J. Ashjian and L. K. Opresko, *Mutation research*, 2010
687, 78-83.
12. T. Lupanova, N. Stefanova, D. Petkova, G. Staneva, A. Jordanova, K. Koumanov, R. Pankov and A. Momchilova, *Molecular and cellular biochemistry*, 2010, **340**, 215-222.
13. Y. M. Wu, J. Tang, P. Zhao, Z. N. Chen and J. L. Jiang, *Experimental and molecular pathology*, 2009, **87**, 133-140.
14. K. E. Sung, G. Su, C. Pehlke, S. M. Trier, K. W. Eliceiri, P. J. Keely, A. Friedl and D. J. Beebe, *Biomaterials*, 2009, **30**, 4833-4841.
15. B. W. Graf and S. A. Boppart, *Methods in molecular biology (Clifton, N.J*, **591**, 211-227.
16. F. Draux, P. Jeannesson, A. Beljebbar, A. Tfayli, N. Fourre, M. Manfait, J. Sule-Suso and G. D. Sockalingum, *The Analyst*, 2009, **134**, 542-548.
17. I. Notingher, S. Verrier, S. Haque, J. M. Polak and L. L. Hench, *Biopolymers*, 2003, **72**, 230-240.
18. I. Notingher and L. L. Hench, *Expert review of medical devices*, 2006, **3**, 215-234.
19. G. J. Puppels, F. F. de Mul, C. Otto, J. Greve, M. Robert-Nicoud, D. J. Arndt-Jovin and T. M. Jovin, *Nature*, 1990, **347**, 301-303.
20. G. J. Puppels, H. S. Garritsen, G. M. Segers-Nolten, F. F. de Mul and J. Greve, *Biophysical journal*, 1991, **60**, 1046-1056.
21. A. D. Meade, F. M. Lyng, P. Knief and H. J. Byrne, *Analytical and bioanalytical chemistry*, 2007, **387**, 1717-1728.
22. F. Bonnier, A. D. Meade, S. Merzha, P. Knief, K. Bhattacharya, F. M. Lyng and H. J. Byrne, *The Analyst*, 2010, **135**, 1697-1703.
23. M. Miljkovic, T. Chernenko, M. J. Romeo, B. Bird, C. Matthaus and M. Diem, *The Analyst*, 2010, **135**, 2002-2013.

24.   N. Uzunbajakava, A. Lenferink, Y. Kraan, B. Willekens, G. Vrensen, J. Greve and C. Otto, *Biopolymers*, 2003, **72**, 1-9.
25.   F. Draux, C. Gobinet, J. Sule-Suso, A. Trussardi, M. Manfait, P. Jeannesson and G. D. Sockalingum, *Analytical and bioanalytical chemistry*, 2010, **397**, 2727-2737.
26.   A. D. Meade, C. Clarke, F. Draux, G. D. Sockalingum, M. Manfait, F. M. Lyng and H. J. Byrne, *Analytical and bioanalytical chemistry*, 2010, **396**, 1781-1791.
27.   R. J. Swain and M. M. Stevens, *Biochemical Society transactions*, 2007, **35**, 544-549.
28.   C. M. Krishna, J. Kurien, S. Mathew, L. Rao, K. Maheedhar, K. K. Kumar and M. V. Chowdary, *Expert review of molecular diagnostics*, 2008, **8**, 149-166.
29.   C. A. Lieber, S. K. Majumder, D. Billheimer, D. L. Ellis and A. Mahadevan-Jansen, *Journal of biomedical optics*, 2008, **13**, 024013.
30.   C. Yu, E. Gestl, K. Eckert, D. Allara and J. Irudayaraj, *Cancer detection and prevention*, 2006, **30**, 515-522.
31.   K. Meister, D. A. Schmidt, E. Brundermann and M. Havenith, *The Analyst*, **135**, 1370-1374.
32.   M. M. Mariani, P. J. Day and V. Deckert, *Integr Biol (Camb)*, **2**, 94-101.
33.   C. Matthaus, T. Chernenko, J. A. Newmark, C. M. Warner and M. Diem, *Biophysical journal*, 2007, **93**, 668-673.
34.   P. Boukamp, S. Popp, K. Bleuel, E. Tomakidi, A. Burkle and N. E. Fusenig, *Oncogene*, 1999, **18**, 5638-5645.
35.   K. Varmuza, in *Introduction to multivariate statistical analysis in chemometrics,* ed.Taylor & Francis Group, CRC Press, New york, 2009, ch.3, pp.59-102.
36.   J. B. MacQueen, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, 1967.
37.   G. J. Puppels, W. Colier, J. H. F. Olminkhof, C. Otto, F. F. M. d. Mu and J. Greve, *Journal of Raman Spectroscopy*, 1991, **22**, 217-225.
38.   A. Curtis and C. Wilkinson, *Biomaterials*, 1997, **18**, 1573-1583.
39.   P. Giannakakou, M. Nakano, K. C. Nicolaou, A. O'Brate, J. Yu, M. V. Blagosklonny, U. F. Greber and T. Fojo, *Proceedings of the National Academy of Sciences of the United States of America*, 2002, **99**, 10855-10860.
40.   A. Kotorashvili, S. J. Russo, S. Mulugeta, S. Guttentag and M. F. Beers, *The Journal of biological chemistry*, 2009, **284**, 16667-16678.
41.   T. C. Hsia, J. S. Yang, G. W. Chen, T. H. Chiu, H. F. Lu, M. D. Yang, F. S. Yu, K. C. Liu, K. C. Lai, C. C. Lin and J. G. Chung, *Anticancer research*, 2009, **29**, 309-318.
42.   B. J. Pettus, A. Bielawska, P. Subramanian, D. S. Wijesinghe, M. Maceyka, C. C. Leslie, J. H. Evans, J. Freiberg, P. Roddy, Y. A. Hannun and C. E. Chalfant, *The Journal of biological chemistry*, 2004, **279**, 11320-11326.
43.   S. Grewal, S. Ponnambalam and J. H. Walker, *Journal of cell science*, 2003, **116**, 2303-2310.
44.   S. Koljenovic, T. C. Bakker Schut, J. P. van Meerbeeck, A. P. Maat, S. A. Burgers, P. E. Zondervan, J. M. Kros and G. J. Puppels, *Journal of biomedical optics*, 2004, **9**, 1187-1197.
45.   P. Lasch, W. Haensch, D. Naumann and M. Diem, *Biochimica et biophysica acta*, 2004, **1688**, 176-186.

46.   E. Ly, O. Piot, A. Durlach, P. Bernard and M. Manfait, *The Analyst*, 2009, **134**, 1208-1214.
47.   M. Hedegaard, C. Krafft, H. J. Ditzel, L. E. Johansen, S. Hassing and J. Popp, *Analytical chemistry*, **82**, 2797-2802.
48.   A. Taleb, J. Diamond, J. J. McGarvey, J. R. Beattie, C. Toland and P. W. Hamilton, *The journal of physical chemistry*, 2006, **110**, 19625-19631.
49.   K. W. Short, S. Carpenter, J. P. Freyer and J. R. Mourant, *Biophysical journal*, 2005, **88**, 4274–4288.
50.   R. J. Swain, G. Jell and M. M. Stevens, *Journal of cellular biochemistry*, 2008, **104**, 1427-1438.
51.   C. M. Krishna, G. D. Sockalingum, L. Venteo, R. A. Bhat, P. Kushtagi, M. Pluot and M. Manfait, *Biopolymers*, 2005, **79**, 269-276.

Figure legends

Figure 1: Pictures taken from A549 (A and B) and HaCaT (D and E) cells grown on Petri dish and collagen gels (C and F) using a confocal microscope equipped with a x20 objective.

Figure 2: Pre-processing of the data set. A: Example of spectrum recorded from a live cell (1) compared with the signal of the collagen gel (2). B: Cellular spectrum obtained after background correction. The dotted line represents the baseline correction to be applied to the data set. C: Example of cellular spectrum obtained after the pre-processing of the data set, background contribution has been removed and baseline corrected.

Figure 3: K-means clustering analysis. A and C are examples of images taken from the Labram HR 800 equipped with x100 immersion objective respectively from a A549 cell and a HaCaT cell. B and D are the corresponding false colors images obtained after K-means analysis. Each pixels of the Raman map has been attributed to a specific cluster. The cluster numbers are indicated in the colour code to the right of the figure

Figure 4: Average spectra obtained from different cluster after K- means clustering analysis. A: nucleus; B: nucleolus; C: cytoplasm cluster 1 and D: cytoplasm cluster 2

Figure 5: Plots of the first 3 principal components after PCA performed on Raman spectra recorded from A549 cells grown on collagen gels. The spectra have been selected for the cluster corresponding to the nucleus (red), nucleolus (green) and cytoplasm (cluster 1: dark blue; cluster 2: light blue).

Figure 6: Plot of the loadings number 1 (A) corresponding to the first principal component resulting of the PCA analysis. This loading has been compared with spectra recorded from different compounds: DNA (B) RNA (C) L-α-Phosphatidylcholine (D) and Sphingomyelin (E).

Figure 7: Plot of the loadings number 2 (A) corresponding to the second principal component resulting of the PCA analysis. This loading has been compared with spectra recorded from different compounds: DNA (B), RNA (C), L-α-Phosphatidylcholine (D) and Sphingomyelin (E).

Figure 8: Plot of the loadings number 3 (A) corresponding to the third principal component resulting of the PCA analysis. This loading has been compared with spectra recorded from different compounds: DNA (B) RNA (C) L-α-Phosphatidylcholine (D) and Sphingomyelin (E).