

2014

## Vibrational Spectroscopy: Disease Diagnostics and Beyond

Hugh Byrne

*Technological University Dublin, [hugh.byrne@tudublin.ie](mailto:hugh.byrne@tudublin.ie)*

Kamila Ostrowska

*Technological University Dublin*

Haq Nawaz

*Technological University Dublin, [haq.nawaz@tudublin.ie](mailto:haq.nawaz@tudublin.ie)*

*See next page for additional authors*

Follow this and additional works at: <https://arrow.tudublin.ie/biophonbk>



Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#), [Chemicals and Drugs Commons](#), [Diagnosis Commons](#), and the [Physics Commons](#)

---

### Recommended Citation

Byrne, H. J. et al. (2014) "Vibrational Spectroscopy: Disease Diagnostics and Beyond", in M. Baranska (ed.) *Optical Spectroscopy and Computational Methods in Biology and Medicine*, Springer Series, Challenges and Advances in Computational Chemistry and Physics, (2014) ISBN 978-94-007-7831-3

This Book Chapter is brought to you for free and open access by the DIT Biophotonics and Imaging at ARROW@TU Dublin. It has been accepted for inclusion in Books/Book Chapters by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie), [vera.kilshaw@tudublin.ie](mailto:vera.kilshaw@tudublin.ie).

---

## Authors

Hugh Byrne, Kamila Ostrowska, Haq Nawaz, Jennifer Dorney, Aidan Meade, Franck Bonnier, and Fiona Lyng

## **Vibrational Spectroscopy: Disease Diagnostics and Beyond**

Hugh J. Byrne, Kamila M. Ostrowska, Haq Nawaz, Jennifer Dorney, Aidan D. Meade, Franck Bonnier, Fiona M. Lyng  
Focas Research Institute,  
Dublin Institute of Technology,  
Camden Row,  
Dublin 8,  
Ireland

### **Summary**

This chapter outlines some developments in the applications of vibrational spectroscopy for disease diagnostics and demonstrates how the applications of the spectroscopic techniques can be extended to the analysis and evaluation of disease aetiology and the mechanisms of interaction with and the cellular and subcellular responses to, for example, chemotherapeutic agents and nanoparticles. The primary emphasis is on Raman spectroscopy, although some examples are based on infrared absorption spectroscopy. The studies presented are chosen to illustrate how a range of multivariate analytical techniques can be employed to maximize the potential benefits of the complex spectral information obtained from tissue or cells.

### **Introduction**

Vibrational Spectroscopy is a subset of spectroscopy which analyses vibrations within a molecule (or material). The vibrations are characteristic of the molecular structure and, in polyatomic molecules, give rise to a spectroscopic “fingerprint”. The spectrum of vibrational energies can thus be employed to characterise a molecular structure, or changes to it due to the local environment or external factors (eg radiation, chemical agents). Vibrational energies fall within the mid Infrared (IR) region of the electromagnetic spectrum and are commonly probed through IR absorption spectroscopy. Following the discovery of IR radiation by Herschel in 1800 [1], initial applications of Infrared absorption spectroscopy were limited to astronomy and astrophysics [2]. In material sciences, significant advances were made by 1900 when Abney and Festing recorded spectra for 52 compounds, correlating absorption bands with molecular structures [3]. Coblentz helped establish IR spectroscopy as a routine analytical tool, cataloguing the spectra of hundreds of substances, both organic and inorganic [3]. Technological developments post world war II aided considerably in establishing IR spectroscopy as a routine laboratory characterisation technique, but none more so than the development of commercial Fourier Transform IR (FTIR) spectrometers in the 1960s and 70s [4,5] and FTIR microscopes in the late 1980s [6].

IR spectroscopy is now a routine technique for materials characterisation and has found numerous applications in forensics [7], environmental science [8] and pharmacology [9]. Raman spectroscopy is a complementary technique with origin in the discovery of the Raman effect in 1928 [10], for which C.V. Raman was

awarded the Nobel prize in physics in 1930. In 1998 the Raman Effect was designated an ACS National Historic Chemical Landmark, in recognition of its importance in materials and process analysis. Raman spectroscopy remained largely a curiosity until the advent of the laser in the 1960s, and the revolution in Charged Coupled Detector (CCD) arrays in the 1980s and 1990's added to the benefits of high laser source intensities. In addition, the development of narrow band laser line rejection filters meant that the huge losses in signal from traditional triple monochromator systems could be overcome with the combination of a filter set and a single spectroscopic grating. Furthermore, the significant reductions in acquisition time with multichannel signal detection enabled significant improvements in signal to noise ratios [11]. The combination of technology developments led to a new range of Raman spectroscopic microscopes in the 1990s, establishing Raman spectroscopic microscopy as a relatively inexpensive benchtop laboratory tool to rival conventional infrared spectroscopy.

Similar to IR spectroscopy, Raman spectroscopy entails the coupling of incident radiation with molecular vibrations and the resultant Raman spectrum is characteristic of the compound or material. However, whereas IR spectroscopy involves the absorption of radiation, Raman spectroscopy is a scattering technique, whereby the incident radiation couples with the vibrating polarisation of the molecule and thus generates or annihilates a vibration. The differing underlying mechanisms gives rise to a complementarity of the two techniques. For a vibration to be active in IR spectroscopy, a change in dipole is required, whereas to be Raman active, a change in polarisability is required. Vibrations of asymmetric, polar bonds thus tend to be strong in IR spectra, whereas Raman is particularly suitable as a probe of symmetric, nonpolar groups. Notably, OH vibrations of water are very strong in IR spectra, whereas they are extremely weak in Raman spectra, rendering Raman a potentially more suitable technique for bioapplications.

A further implication of the differing physical origins of the techniques is that whereas IR monitors the absorption of IR radiation, Raman scattering can be employed in the UV, visible or near IR regions of the spectrum. Raman scattering thus offers intrinsically higher spatial resolution for mapping or profiling, the diffraction limit being determined by the wavelength ( $<1\mu\text{m}$  for Raman,  $\sim 5\text{-}10\mu\text{m}$  for IR). For many applications, however, near IR is favoured as a source for Raman analysis, to minimise interference from scattering, fluorescence, or photodegradation of the sample [12].

FTIR has become an accepted tool in biophysics for analysis of the structure and interactions of proteins [13], lipids [14], carbohydrates [15] and nucleic acids [16]. Applications to tissue samples for (cancer) diagnostic applications were first reported in the early 90s [17], and since this time a range of pathologies has been investigated [18 – 20]. The application of Raman spectroscopy to biomolecules and even tissues was first demonstrated as early as the 1960s [21-23] and by the mid 1970s biomedical applications were explored [24]. Whole cell and tissue studies have been carried out on a range of pathologies [25-29] and *in vivo* studies [30-32] have demonstrated the prospective for diagnostic applications. The potential

of vibrational spectroscopy in conjunction with multivariate analysis techniques as a diagnostic tool has thus been well demonstrated and the concept of Spectral Cytopathology has been coined [33].

This chapter is based on the experience of the research activities within the Focas Research Institute, Dublin Institute of Technology, and outlines efforts and progress in the applications of both Raman and infrared spectroscopy for the routine screening of disease as well as the understanding disease origin. It further describes studies demonstrating the potential of the techniques for *in vitro* diagnostics and biochemical analyses, for example of the efficacy of chemotherapeutic agents and the interaction of nanoparticles in cells.

### **Vibrational Spectroscopy for Cervical Cancer Pathology [34]**

Cervical cancer is the second most common cancer among women worldwide and generally is more common in developing countries. However, substantial decreases in cervical cancer mortality have been observed in Western countries and some developing countries with well developed screening programmes [35]. A Papanicolaou test (Pap smear test) is used to screen for Cervical Intraepithelial Neoplasia (CIN) and cervical cancer in the general female population. Such cervical cytology can have a high specificity of 95-98% but a sensitivity of lower than 50% [36]. Other methods such as automated cytology and human papilloma virus (HPV) testing [37-39] have been introduced to reduce the false negative rates. An abnormal Pap smear is followed by colposcopic examination, biopsy and histological confirmation of the clinical diagnosis. This involves the visual examination of histological sections. The grading characteristics are quite subjective and pre-malignancy may not be visually perceptible at all.

Optical technologies such as fluorescence spectroscopy [40], polarised light scattering spectroscopy [41], optical coherence tomography [42, 43] and confocal reflectance microscopy [44] have emerged in recent years as promising tools for diagnosis of cervical and other cancers. The potential of vibrational spectroscopy for cervical cancer diagnosis has also been recognised. A number of studies have shown that Fourier Transform Infra Red (FTIR) spectroscopy [45-47] or Raman spectroscopy [32, 48] can be used for detecting spectral changes in malignant and pre-malignant cells. Raman spectroscopy is, however, more suited to *in vivo* studies because of minimal interference from water. Coupled with adapted multivariate analysis techniques, Raman spectroscopic analysis of cervical biopsies and cytological samples can provide high specificity and sensitivity for the detection of cervical cancer and early pre-malignant states. The following provides an illustration of such applications combining Raman analysis and Principal Component Analysis/Linear Discriminate Analysis [34].

Formalin fixed paraffin preserved (FFPP) cervical tissue samples were obtained from the National Maternity Hospital, Holles St., Dublin. Two parallel 10 $\mu$ m FFPP sections were cut from each block using a microtome, mounted on glass slides and dried. Samples were dewaxed in-house prior to investigation by immersion in baths of Xylene (BDH), Ethanol Absolut (Merck) and Industrial Methylat-

ed Spirits 95% (Lennox) and air-dried. Cervical tissue was dewaxed using the same procedure outlined above, but xylene was substituted with hexane, due to the improved level of dewaxing [49]. One section from each sample (the reference section) was stained with Haematoxylin and Eosin and the other kept unstained for spectroscopic examination.

FFPP cervical tissue sections were characterised by consultant pathologist Dr. P. Kelehan, National Maternity Hospital, Holles St., Dublin, and the samples consisted of 20 normal and 20 invasive carcinoma sections from 40 individuals. Of the 20 carcinoma samples, 10 samples were identified as having various grades of CIN, which were also marked for examination.

An Instruments S.A. Labram Raman spectroscopic confocal microscope was used, with an Argon Ion laser operating at an excitation wavelength of 514.5nm. A x50 objective lens was used. The laser power at the sample was measured and found to be  $6.5 \pm 0.05$  mW, focused to a spot size of  $\approx 2 \mu\text{m}$  at the tissue surface. Raman spectra of the various tissues were accumulated for 150 seconds. A total of 10 spectra were recorded from different spots on each sample.

Selected spectra were baseline corrected, dark current subtracted and were smoothed using a 10 point moving average before being normalised to the spectral maximum, from 0 to 1. Basic matrix manipulations and data reduction was carried out in Microsoft Excel Professional 2003 (v. 11.0), before being exported into Minitab to perform Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). Minitab Release 14.1 Statistical Software Analysis Programme was used to produce PCA scores and LDA plots, as well as to carry out leave-one-out cross validation.

Figure 1 shows the layer structure of normal cervical tissue in an unstained FFPP tissue section together with the Raman spectra recorded from basal cells, epithelial cells and connective tissue. The spectra of the three different regions do have a degree of similarity, but notably there are many differentiating spectral features. The spectra of basal cells show strong bands at  $724 \text{ cm}^{-1}$ ,  $779 \text{ cm}^{-1}$  and  $1578 \text{ cm}^{-1}$  which are characteristic of nucleic acids. The morphology of basal cells consists of a single line of tightly packed cells, with large nuclei in relation to the compacted surrounding cytoplasm. In addition, these cells are constantly dividing, providing cells to the parabasal layer. For both of these reasons, a high concentration of DNA would be expected in the basal cells. Spectra of epithelial cells have characteristic glycogen bands at  $482 \text{ cm}^{-1}$ ,  $849 \text{ cm}^{-1}$ ,  $938 \text{ cm}^{-1}$ ,  $1082 \text{ cm}^{-1}$  and  $1336 \text{ cm}^{-1}$ . Collagen contributions can be clearly seen in the spectra of connective tissue at  $850 \text{ cm}^{-1}$ ,  $940 \text{ cm}^{-1}$  and  $1245 \text{ cm}^{-1}$ .

Figure 2a compares the Raman spectra collected from normal epithelial cells and invasive carcinoma from a selection of different patients. Glycogen contributions are clearly visible in the spectra from the normal epithelial tissue. The most obvious bands arise at  $482 \text{ cm}^{-1}$ ,  $849 \text{ cm}^{-1}$  and  $938 \text{ cm}^{-1}$  and are due to glycogen skeletal deformation, CCH aromatic deformation and CCH deformation respectively. However, there are also other glycogen contributions not as apparent, including a CC stretching band at  $1082 \text{ cm}^{-1}$  and  $\text{CH}_3\text{CH}_2$  wagging at  $1336 \text{ cm}^{-1}$ . These glycogen bands ( $482 \text{ cm}^{-1}$ ,  $849 \text{ cm}^{-1}$  and  $938 \text{ cm}^{-1}$ ) are absent in the spectra from inva-

sive carcinoma, and a reduction in the intensity of the CC stretching mode ( $1082\text{ cm}^{-1}$ ) is also apparent. Glycogen, a polysaccharide, is the principal storage form of glucose. Cervical cells are unusual among other epithelial cells in that they accumulate large amounts of glycogen during the maturation process. Glycogen is known to be linked with cellular maturation and disappears with the loss of differentiation during neoplasia. The spectra of invasive carcinoma also show characteristic nucleic acid bands. These include prominent bands at  $724\text{ cm}^{-1}$ ,  $779\text{ cm}^{-1}$  and  $1578\text{ cm}^{-1}$ , but also at  $829\text{ cm}^{-1}$ ,  $852\text{ cm}^{-1}$ ,  $1002\text{ cm}^{-1}$ ,  $1098\text{ cm}^{-1}$ ,  $1240\text{ cm}^{-1}$ . Distinct bands are also seen at  $1366\text{ cm}^{-1}$ , a shoulder at  $1484\text{ cm}^{-1}$  and a band at  $1578\text{ cm}^{-1}$ . An increase in the intensity of the Amide I band ( $1655\text{ cm}^{-1}$ ) was observed in the spectra of carcinoma samples compared to the normal tissue samples. The increased nucleic acid and protein bands are a result of the increased proliferation of these tumour cells.

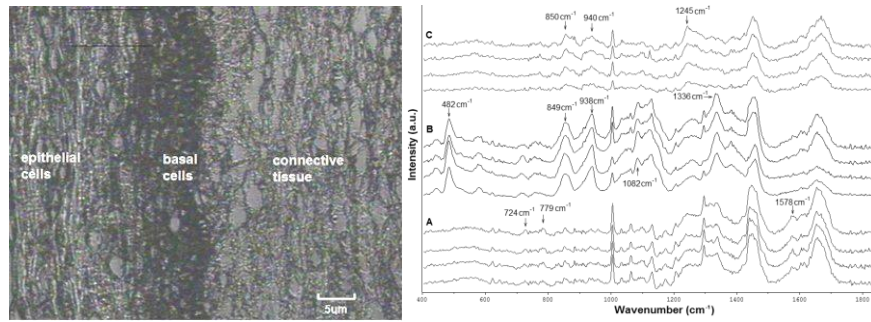


Figure 1: (a) Photomicrograph of unstained tissue section, with different cell types identified and (b) Raman spectra recorded from basal cells (A), epithelial cells (B) and connective tissue (C) in cervical tissue sections. The main spectral features associated with each cell type are highlighted. (Reproduced from [34])

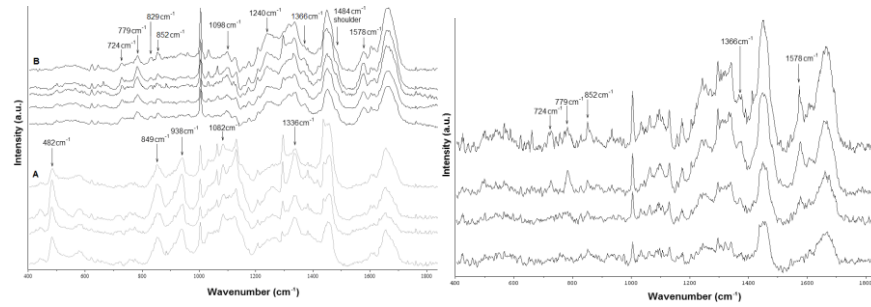


Figure 2: Raman spectra of (a) normal cervical epithelial cells (A) and invasive carcinoma cells (B) and (b) Raman spectra of CIN tissue. Assignments of the main Raman vibrational modes are detailed in table 1. (Reproduced from [34])

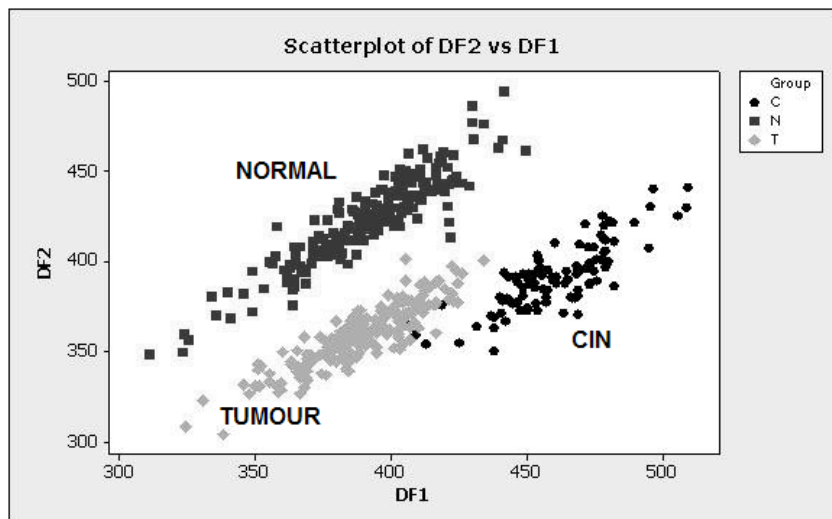


Figure 3: Linear discriminant analysis of the principal components of the first derivative, normalised, 10 point averaged spectra, over the entire spectral range, C = CIN, N = normal and T = invasive carcinoma. (Reproduced from [34])

To investigate whether pre-malignant changes can be elucidated using Raman spectroscopy, 10 areas of CIN from 10 different patients were marked by the pathologist and a selection of the resulting Raman spectra are shown in Figure 2b. A number of the spectral features identified in the invasive carcinoma samples were also observed in the CIN samples such as the nucleic acid bands at  $724\text{ cm}^{-1}$ ,  $779\text{ cm}^{-1}$ ,  $852\text{ cm}^{-1}$ ,  $1366\text{ cm}^{-1}$  and  $1578\text{ cm}^{-1}$ . This indicates that biochemical changes associated with CIN can be identified using Raman spectroscopy.

Principal component analysis (PCA) is a method of multivariate analysis widely used with datasets of multiple dimensions. PCA is a chemometric technique that can resolve a complete spectral data set into a few principal components and can thus identify and isolate important trends within the data set [50]. 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,  $\text{var}(PC_1) \geq \text{var}(PC_2) \geq \text{var}(PC_p)$ , where  $\text{var}(PC_i)$  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 (wave-number) 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.



Peak Position (cm <sup>-1</sup> )	Assignment	Peak reference (figure 4)
622	C-C twisting (Proteins)	1
642	C-C twisting mode of Tyr and Phe	2
670	C, T,G (DNA/RNA)	3
720	C-N stretching in A and lipids	4
724	CH <sub>2</sub> deformation	
746	CH <sub>2</sub> rocking	
750	Symmetric ring breathing (protein)	5
782	U, T, C (ring breathing modes in the DNA/RNA)	6
827	PO <sub>2</sub> stretching in DNA, Tyr	7
832	CCH deformation aliphatic	
854	Ring breathing in Tyr and Pro (proteins)	8
873	CC stretch	
922	C-C stretching	
935	C-C stretching mode of Pro and Val	9
1004	CC aromatic ring breathing	10
1031	C-H bending mode in Phenylalanine	11
1034	C-N stretching in proteins	
1065	C-N stretch	
1093	symmetric PO <sub>2</sub> stretching of the DNA backbone; lipids	12
1096	C-C chain stretching	
1098	CC stretch	
1102	CC stretch	
1124	CC skeletal stretch <i>trans</i>	
1128	C-N stretching in proteins; C-O stretching in carbohydrates	13
1155	C-C and C-N stretching of proteins	14
1175	C-H in plane bending mode of Tyr and Phe; C, G	15
1208	C-C <sub>6</sub> H <sub>5</sub> stretching mode in Trp, Phe;	16
1214	CC stretch backbone carbon phenyl ring	
1220-1280	Amide III; A,C,T ring breathing modes of the DNA/RNA	17
1236	CN stretch, NH bending Amide III band	
1240	CN stretch, NH bending Amide III band	
1311	CH <sub>3</sub> /CH <sub>2</sub> twisting mode of collagen and lipid	18
1314	CH deformation	
1337	CH <sub>2</sub> deformation	
1335	CH <sub>2</sub> deformation	
1340	G (DNA/RNA), CH deformation in proteins and carbohydrates; Trp	19
1366	CH <sub>2</sub> bending	
1440	CH <sub>2</sub> scissoring	
1450	CH (CH <sub>2</sub> ) bending mode in proteins and lipids	20
1484	CH <sub>2</sub> deformation	
1548	NH deformation; CN stretch Amide II band	
1578	C=C olefinic stretch	
1583	A,G (DNA/RNA); C=C bending mode of Phe	21
1585	C=C stretching	
1550-1700	H <sub>2</sub> O bending mode	22
1602	CO stretching	
1618	C=C Phe, Tyr and Trp	23
1660 - 1665	C=O stretch Amide I $\alpha$ -helix	24
2930	CH <sub>2</sub> stretching (2930 cm <sup>-1</sup> )	
2932	CH <sub>3</sub> symmetric stretch (2932 cm <sup>-1</sup> ).	

Table 1: Peak position and assignments of main Raman vibrational modes

<b>Predicted (Raman)</b>	<b>Diagnosis (Histopathology)</b>		
	<i>Normal</i>	<i>CIN</i>	<i>Invasive carcinoma</i>
<i>Normal</i>	198	0	0
<i>CIN</i>	0	99	3
<i>Invasive carcinoma</i>	2	1	195
<i>Total</i>	200	100	198
<i>Number correct</i>	198	99	195
<i>Proportion</i>	0.990	0.990	0.985
<b>Sensitivity</b>	99.5%	99.0%	98.5%
<b>Specificity</b>	100%	99.2%	99.0%

Table 2: Summary of cross-validation classification results and sensitivity and specificity values

Principal component analysis was used to reduce the number of parameters needed to represent the variance in the spectral data set. The principal components were then used to generate a linear discriminant model. LDA applies a linear discriminant function that maximises the variance in the data between groups and minimises the variance between members of the same group [51]. All three tissue classes were successfully discriminated as shown in figure 3. The classification model was tested using a leave one out cross validation in which all but one spectrum was used to build the model. This model was then used to predict the remaining spectrum. This was repeated for all 498 spectra and the results are shown in table 2. Of 498 tissue spectra, 492 were correctly classified as normal, invasive carcinoma or CIN. The cross validation mis-classified 6 spectra, 2 of which were normal samples assigned as invasive carcinoma. The other 4 were either invasive carcinoma or CIN mis-classified as either CIN or invasive carcinoma respectively. Importantly, no abnormal samples were classified as normal. Based on the cross validation results, sensitivity and specificity values were calculated as 99.5% and 100% respectively for normal tissue, 99% and 99.2% respectively for CIN and 98.5% and 99% respectively for invasive carcinoma. It is possible that these values may be slightly over optimistic because of the LDA method used with leave one out cross validation. However, similar methods have been used previously to classify bladder and prostate cancer [52], breast cancer [53] and basal cell carcinoma [54]. Ideally, the use of an unknown test set would provide a more robust analysis of the sensitivity and specificity of the technique and this is planned for future work.

The results show the ability of Raman spectroscopy to classify cervical cancer and pre-cancer with high sensitivity and specificity. These classifications are based on biochemical changes known to accompany cervical cancer such as loss of differentiation and increased proliferation. The study presented details of the application to cervical biopsies, but they can equally be applied to cytological samples. Studies in a range of different pathologies have emerged and technologi-

cal development is progressing beyond the stage of “proof of concept” [55, 56]. Technologically, *ex-vivo* screening of histological or cytological samples is less challenging, but as the technologies are optically based, the emergence of *in-vivo* probes is inevitable [57]. The development of such an automated technique measuring biochemical changes with improved diagnostic capability could allow faster, more effective patient management and inevitably would increase survival rates. Raman spectroscopy therefore offers enormous potential as a ‘next generation’ technology to assist pathologists and cytologists with cervical cancer, and other pathological screening and diagnosis.

In the development of the technologies towards clinical and other applications, it is important to ensure that the preprocessing of the spectral data is appropriate and in this context it is important to understand the physical and instrumental origin of any measurement artefacts. Significant progress has been made in FTIR spectroscopy to identify and correct for so-called dispersion artefacts [58-60]. In Raman microspectroscopy, it has been demonstrated that measurement in immersion significantly reduces the spectral background [61] and photodamage [62], minimizing the need for spectral preprocessing.

Notably, the analytical capabilities of the spectral techniques rely heavily on multivariate analytical techniques to classify the complex spectral signatures, and changes to them, associated with disease. A broad range of data analytical techniques are available including Genetic algorithms, Support Vector Machines, Neural Networks etc. To date, no direct comparison of the efficacy or suitability of such techniques has been made and such a comparison on experimental data may not be conclusive, as the true result is not known. Ultimately, the field would benefit from an exhaustive comparison of the range of data preprocessing and analysis techniques on appropriate simulated datasets [59].

### **Exploring the aetiology of disease – HPV infection [63, 64]**

The probability of developing cervical dysplasia is much higher for women infected with the human papillomavirus (HPV) than those free of infection [65]. Infection with oncogenic or high-risk HPV (HPV-16, HPV-18) is the main risk factor for cervical cancer [66], 99.7% of invasive cervical neoplasia being associated with HPV infection [67]. Identifying the presence of HPV is a very important factor in the prevention of cervical cancer and thus, HPV testing has been added to the range of clinical options for cervical cancer screening. HPV testing is based on detecting the DNA of the virus. It has been found that the DNA of high risk strains of HPV is assimilated within the cellular DNA of the host [68]. A drawback to HPV testing is that it is more expensive and time-consuming than other screening tests, and it requires a sophisticated laboratory infrastructure.

Raman spectroscopy has also been used to study cell lines derived from the cervix, revealing spectral variations, mostly in peaks originating from DNA and proteins, in cell lines expressing the E7 gene of HPV-16 compared to cells not affected by HPV [69]. Furthermore, it has been demonstrated that in low-grade cytology infected with high-risk oncogenic HPV16 or HPV18, it was possible to dif-

ferentiate samples from women based on whether they were aged 20–29 years vs. 30–39 years using attenuated total reflection infrared spectroscopy [70].

To further demonstrate the potential of vibrational spectroscopy for clinical screening, four cervical cancer cell lines were investigated using both Raman and FTIR spectroscopy; HPV negative C33A, HPV-18 positive HeLa with 20-50 integrated HPV copies per cell, HPV-16 positive SiHa with 1-2 integrated HPV strands per cell and HPV-16 positive CaSki containing 60-600 integrated HPV copies per cell. The main aim of the experiment was to investigate biochemical changes in cells caused by high risk HPV strands (HPV-16 and HPV-18) and to examine whether it is possible to differentiate between the cells with no HPV copies and cells infected by low number, medium number and high number of integrated HPV copies, using vibrational spectroscopic techniques. Both common forms of vibrational spectroscopy, Raman and Fourier Transform Infrared absorption, were used to gain complementary information [63]. The Raman spectroscopic analysis is discussed here.

A Horiba Jobin Yvon LabRam 800HR was employed for the measurement, operating at a wavelength of 532nm and utilising a x100 objective. After incubation, cells were centrifuged to form pellets and washed thoroughly before measurement. The mean of 30 Raman spectra recorded from two pellets (approximately 15 spectra per pellet) for C33A, SiHa, HeLa and CaSki cell lines were derived from the raw spectra via the pre-processing procedures and are presented in figure 4. Assignments of the main peaks (numbered in figure 4) are shown in table 1. In the mean Raman spectra, almost no significant differences between HPV negative C33A and SiHa cells, containing 1-2 HPV copies per cells (low HPV concentration) are discernible. Only the peak at  $1583\text{cm}^{-1}$  assigned to DNA/RNA and Phenylalanine vibrations was found to vary between C33A and SiHa cells. For cells containing a higher number of integrated HPV copies per cell, HeLa cells (medium HPV concentration) and CaSki cells (high HPV concentration), the observed differences are more prominent for protein, nucleic acid and lipid vibrations. CaSki cells exhibited a significant increase in Amide I (proteins) intensity at  $1660\text{cm}^{-1}$  and a decrease in lipid levels at  $1093\text{cm}^{-1}$ ,  $1311\text{cm}^{-1}$  and  $1450\text{cm}^{-1}$  compared to C33A, SiHa and HeLa. HPV presence in a cell appears to change protein expression, which may result in the protein variations observed in the Raman spectra, primarily for the Amide I region. Further differences in protein vibrations between the cells affected by non- and low HPV number compared to medium and high HPV number were observed in the Amide III region ( $1254\text{cm}^{-1}$ ), and protein constituent amino acids Tryptophan (Trp), Proline (Pro), Tyrosine (Tyr) and Phenylalanine (Phe) vibrations at  $854\text{cm}^{-1}$ ,  $935\text{cm}^{-1}$ ,  $1003\text{cm}^{-1}$ ,  $1583\text{cm}^{-1}$  and  $1618\text{cm}^{-1}$ . Variations in vibrations associated with nucleic acid constituents:  $\text{PO}_2$  group, A, C, T, U at  $1093\text{cm}^{-1}$ ,  $720\text{cm}^{-1}$ ,  $782\text{cm}^{-1}$ ,  $1254\text{cm}^{-1}$ ,  $1583\text{cm}^{-1}$  respectively were also observed. The differences in protein and nucleic acid composition support the molecular data which demonstrates integration of HPV DNA at chromosome fragile sites and the subsequent downstream effect of HPV E6 and E7 oncoprotein expression on host cell proteins such as p53 [71].

The principal component (PC) scatterplot for Raman spectroscopic results presented in figure 5A shows a distinctive separation between the cell lines, suggesting a HPV concentration dependence factor. The C33A (non HPV) and SiHa (low HPV concentration, LC) cell lines are grouped together, confirming observed similarities between these two types of cells based on mean spectra analysis. The HeLa (medium HPV concentration, MC) and CaSki (high HPV concentration, HC) cell lines are clearly separated from each other and from the C33A-SiHa group. Although the 3D PCA scatterplot demonstrates the HPV concentration dependent separation, the cell lines are not grouped according to increasing (or decreasing) number of HPV copies. CaSki (HC) cells lie between HeLa (MC) and SiHa (LC). This suggests the influence on the data point distribution of another factor – HPV strand type. HeLa are cells infected by HPV-18, while CaSki and SiHa are infected by HPV-16. Principal component loadings are presented in figure 5B. Peak numbers correspond to assignments presented in table 1. The percent of variance explained by PC1, PC2 and PC3 is 71.9%, 10.8% and 6.2%, respectively. The first principal component is dominated by the contribution of the H-O-H bending of water, a broad peak in the range of  $1550\text{--}1700\text{cm}^{-1}$  with maximum at  $\sim 1644\text{cm}^{-1}$  [72]. Water is not, however, the only potential contribution within this range, as Amide I vibrations ( $\beta$ -sheet) exhibit a maximum at  $1640\text{cm}^{-1}$ . Nucleic acid contributions to PC1 may also be present at  $782\text{cm}^{-1}$ ,  $1093\text{cm}^{-1}$  and  $1583\text{cm}^{-1}$ , although it is very difficult to judge what is the percentage of contribution of each of these components. The second principal component shows predominantly protein and lipid contributions. Lipids are exhibited at  $1095\text{cm}^{-1}$ ,  $1310\text{cm}^{-1}$  and  $1450\text{cm}^{-1}$ , while proteins are at  $1660\text{cm}^{-1}$  – Amide I ( $\alpha$ -helix),  $1270\text{cm}^{-1}$  – Amide III ( $\alpha$ -helix),  $1450\text{cm}^{-1}$ , C=C Phe at  $1582\text{cm}^{-1}$  and Trp at  $750\text{cm}^{-1}$ . The third principal component contains a lot of biological information with mainly nucleic acid and protein contributions but only accounts for 6.2% of variance. There are distinctive peaks corresponding to nucleic acids and protein vibrations. DNA and RNA contributions are observed at  $670\text{cm}^{-1}$ ,  $720\text{cm}^{-1}$ ,  $782\text{cm}^{-1}$ ,  $1092\text{cm}^{-1}$ ,  $1250\text{cm}^{-1}$ ,  $1340\text{cm}^{-1}$  and  $1583\text{cm}^{-1}$ . PC3 shows a strong influence of variation arising from protein vibrations, Phe ( $642\text{cm}^{-1}$ ,  $1002\text{cm}^{-1}$ ,  $1031\text{cm}^{-1}$ ,  $1583\text{cm}^{-1}$  and  $1618\text{cm}^{-1}$ ), Trp ( $1208\text{cm}^{-1}$ ,  $1340\text{cm}^{-1}$  and  $1618\text{cm}^{-1}$ ), Tyr ( $642\text{cm}^{-1}$ ,  $854\text{cm}^{-1}$  and  $1618\text{cm}^{-1}$ ), Pro ( $854\text{cm}^{-1}$ ,  $935\text{cm}^{-1}$ ), Val ( $935\text{cm}^{-1}$ ), Amide I at  $1680\text{cm}^{-1}$  (disordered) and Amide III at  $1260\text{cm}^{-1}$  (disordered). The source of variations shown in the loadings is in very good agreement with those observed in the analysis of the mean spectra for the investigated cell lines. The main peaks in the loadings arise from protein, nucleic acid and lipid component vibrations. Each principal component is dominated by different cellular components; water, and possibly nucleic acids and proteins in PC1, lipids and proteins in PC2 and nucleic acids and proteins in PC3. Amide I and Amide III contributions are present in all principal component loadings indicative of differences in structural conformation of the protein  $\beta$ -sheet in PC1,  $\alpha$ -helix in PC2 and disordered conformation in PC3. As mentioned previously, variations in protein, nucleic acid and lipid levels arise from the fact that the viral DNA interacts with the cellular DNA, causing repro-

gramming of protein expression in the cell. It was noted that the multivariate analysis could be substantially influenced by water contributions in the region of  $1640\text{cm}^{-1}$ . Repeated over the spectral range  $400\text{--}1550\text{cm}^{-1}$ , the PCA gave a similar separation of the cellular types and yielded similar principal components.

The applications of spectroscopic analysis, coupled with multivariate statistic techniques can therefore be extended beyond diagnosis based on classification to biochemical analysis of the underlying physiology based on changes to the molecular composition on a cellular and subcellular level. This new technology for cervical screening not only can distinguish cervical cells based on biochemical content but also can elucidate the effect of HPV on the cells. As a further demonstration of the potential of the techniques, FTIR spectroscopy was employed to further explore the known manifestations of HPV infection in cervical cell lines [64].

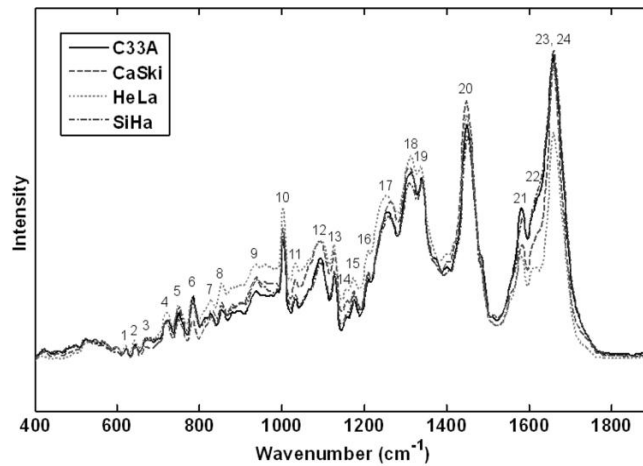


Figure 4: Mean Raman spectra of cervical cancer cell lines. (Reproduced from [63])

Differentially expressed proteins in cancer have potential utility as biomarkers. As the cell cycle is often disrupted in a cancerous cell, proteins associated with it are often candidate biomarkers. Putative biomarkers of cervical cancer that are currently under study include proteins such as CDC 6 (DNA licensing protein), minichromosome maintenance proteins (MCM 2, MCM 5), p53 or p16<sup>INK4A</sup> [73]. These biomarkers have been used to detect the presence of abnormal cells, based upon immunocytochemical methods. p16<sup>INK4A</sup> regulates the levels of active cyclin D/CDK in the cell, providing a feedback mechanism that regulates the levels of MCM (minichromosome maintenance proteins), PCNA (proliferating-cell nuclear antigen) and cyclin E. Overexpression of p16<sup>INK4A</sup>, which is considered a marker of elevated E7 expression, can be detected in some CIN1 lesions, as well as in

CIN2 and CIN3 lesions that show evidence of integration [74,75]. Integration of the HPV genome into the host cell chromosome is a critical event in the development of most cervical cancers [76]. Overexpression of p16<sup>INK4A</sup> has been demonstrated in cervical cancers as a result of functional inactivation of pRb by the HPV E7 protein [77]. This overexpression highlights the potential of p16<sup>INK4A</sup> as a marker for CIN and cervical cancer. HPV positivity and p16<sup>INK4A</sup> positivity have shown a correlation, even though p16<sup>INK4A</sup> expression was also seen in a limited number of HPV negative biopsy samples [78]. Also, a correlation between p16<sup>INK4A</sup> expression and cervical lesion grade and HR-HPV positivity has been documented [79]. p16<sup>INK4A</sup> has been proven to be the most reliable marker of cervical dysplasia and was found to mark all grades of squamous and glandular lesions of the cervix. The use of p16<sup>INK4A</sup> immunocytochemical analysis as a complement to conventional screening programmes could potentially aid in the reduction of false positive and false negative results [80].

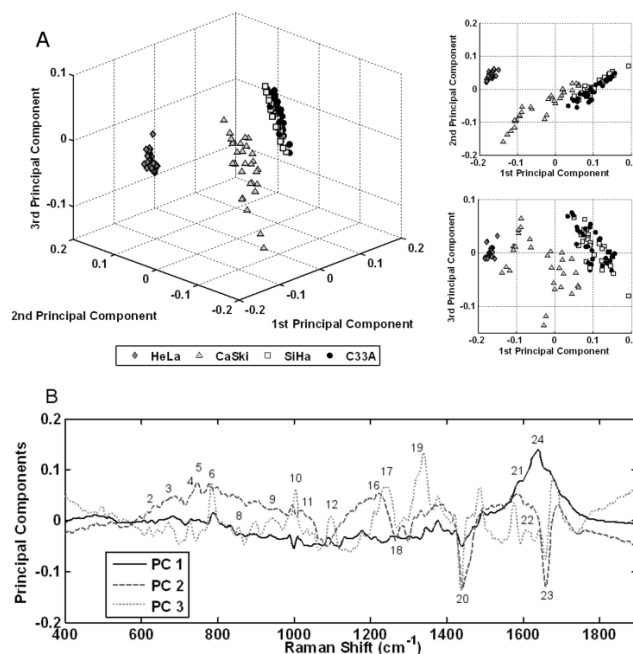


Figure 5: Principal component analysis for Raman spectroscopy results for cervical cancer cell lines. (A) - Three dimensional PCA scatterplot. Graphs show separation of the cervical cancer cell lines suggesting HPV concentration dependence. (B) - Principal component loadings. Peak assignments correspond to those listed in table 1. (Reproduced from [63])

In the first part of this study, expression of p16<sup>INK4A</sup> was analysed in cervical cancer cell lines using immunocytochemical staining and both confocal fluores-

cence microscopy and flow cytometry. [64] Fluorescence images of cells were collected using a confocal fluorescence microscopy system. Representative images of cells are presented in figure 6, where A, C, E and G show only green fluorescence staining (FITC), while B, D, F and H show the overlay of green (FITC) and red (PI) fluorescence signals. The FITC fluorescence intensity is associated with the  $p16^{INK4A}$  expression level, while the PI dye was used to visualise the position of nuclei within the cells.

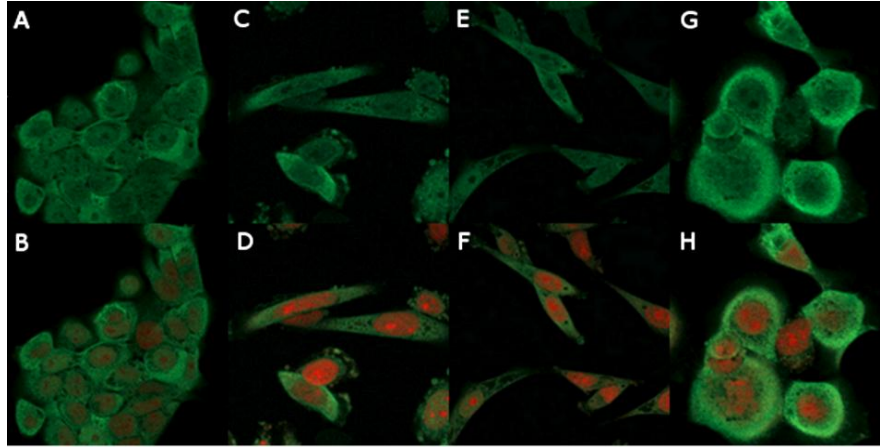


Figure 6: Confocal microscopy images of cervical cell lines: A-B C33A cells, C-D SiHa cells, E-F HeLa cells, G-H CaSki cells. Images A,C, E and G show FITC fluorescence staining (related to  $p16^{INK4A}$  expression), while images B, D, F and H present FITC and PI staining. (Reproduced from [64])

A degree of correlation between the number of HPV copies per cell line and the fluorescence intensity related to  $p16^{INK4A}$  expression was observed for the flow cytometry data (figure 7). The highest level of  $p16^{INK4A}$  expression (intensity of staining) was observed for CaSki cells, followed by HeLa, SiHa and the lowest for C33A. The relationship between  $p16^{INK4A}$  expression and HPV copy number illustrated by figure 7 is supported by a study conducted by Klaes et al. (2001), wherein a correlation between increasing grade of cervical lesion and staining intensity of  $p16^{INK4A}$  was observed [81]. Similarly, Agoff et al. showed that  $p16^{INK4A}$  expression correlates with an increasing severity of cervical disease [82]. Murphy et al. showed a strong correlation between HR-HPV positivity and  $p16^{INK4A}$  staining pattern [78, 83, 84]. In the study conducted by Wang et al. a correlation between  $p16^{INK4A}$  immunostaining and cervical disease severity stratified by HPV status was observed [85].

The results demonstrate a correlation between  $p16^{INK4A}$  immunostaining and the presence of HPV in cervical cancer cell lines, supporting the previous studies and implies that  $p16^{INK4A}$  is a highly sensitive marker of HPV in cervical cancer



cell lines. The hyperbolic-like relationship between the levels of p16<sup>INK4A</sup> and HPV copy number is typical of the response of a cell to the action of an agonist [86] and in figure 7, is approximated by a fit with the following function  $y=a+b\sqrt{x}+cx$ , where  $a=62.4$ ,  $b=3.5$  and  $c=-0.1$ . It should be noted that, in figure 7, the HPV copy number is represented by the average of the range quoted in literature [87, 88] and so error margins in the horizontal axis are potentially very large. However, the sublinear nature of the plot indicates that p16<sup>INK4A</sup> expression levels are particularly sensitive for low HPV copy number.

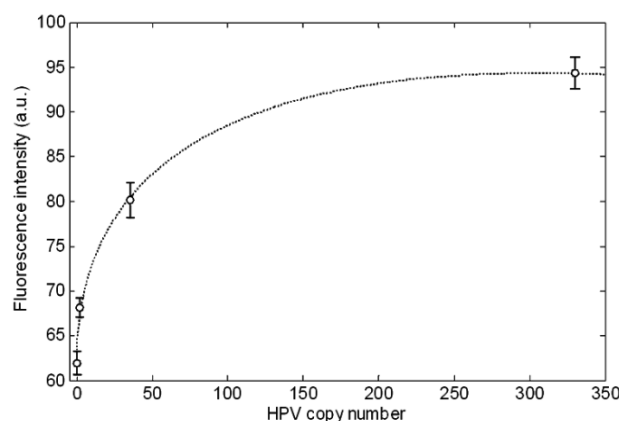


Figure 7: Fluorescence intensity related to the p16INK4A expression level in cervical cancer cell lines plotted against the average HPV copy number present in a cell with the fitted function. The dotted line is a fit of the data with the equation  $y=a+b\sqrt{x}+cx$ , where  $a=62.4$ ,  $b=3.5$  and  $c=-0.1$ . (Reproduced from [64])

FTIR maps were recorded for C33A, SiHa, HeLa and CaSki cells (2-4 maps for each cell line) using the Perkin Elmer Spotlight 400N system. As in Raman spectroscopy, three main biochemical cellular components feature strongly: proteins, lipids and nucleic acids. A detailed examination of the relative intensities of peaks related to vibrations of the cellular components (nucleus and cytoplasm) of the C33A, SiHa, HeLa and CaSki infrared spectra revealed the following tendencies:

- Increase in nucleic acid levels with increasing number of HR-HPV copies in the cell or increasing p16INK4A expression (figure 8A, 8D),
- Decrease in lipid levels with increasing number of HR-HPV copies in the cell or increasing p16INK4A expression (figure 8B, 8E),
- No tendency for changes in protein levels related to HR-HPV copy number or increasing p16INK4A expression (figure 8C, 8F).

The observed increase in nucleic acid levels may be related to an increased number of chromosomes present in HPV infected nuclei. It was reported by Mehés et al. that HPV presence facilitates polyploidisation (increase in chromosome

number in a cell nucleus) in cervical squamous cells [89]. Additionally, it is known that binding of HPV DNA to host DNA disrupts the normal function of the cellular proteins and as a consequence, the host cell accumulates more and more damaged DNA that cannot be repaired [90].

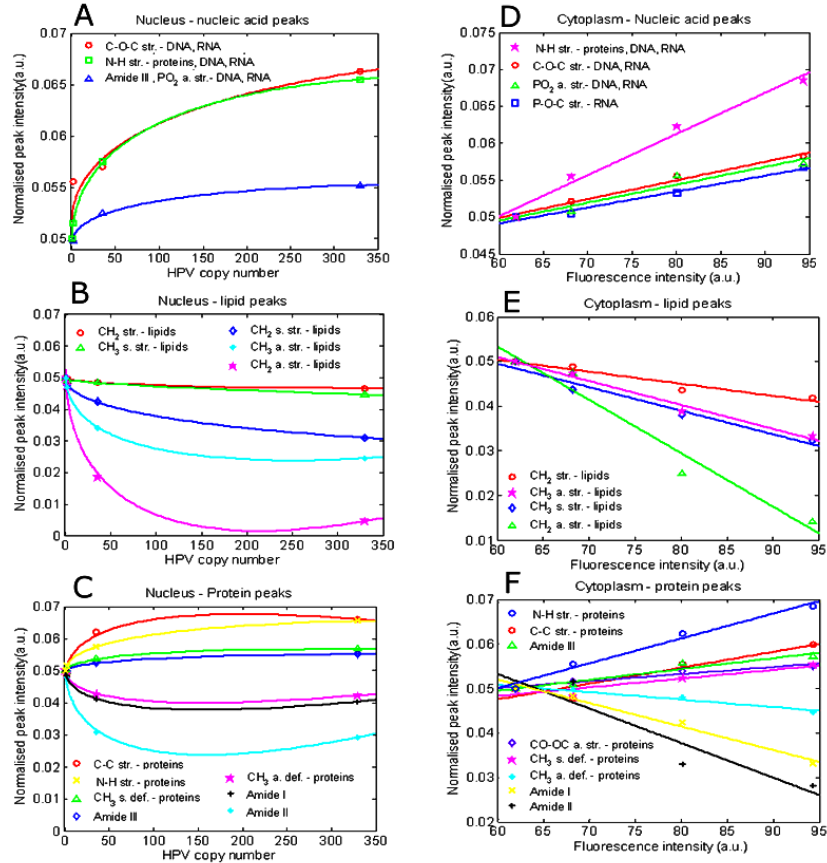


Figure 8: Peak intensity analysis for FTIR spectra of nuclear and cytoplasmic regions of cervical cancer cells. Dependence of peak intensities vs. HPV copy number was fit with  $y=a+b\sqrt{x}+cx$  function, while peak intensities vs. fluorescence intensities (p16INK4A expression level) was fit with a linear function,  $y=a+bx$ . (Reproduced from [64])

CaSki cells are known to be the most malignant and in the spectra of the cytoplasm and nucleus representing this cell line the highest level of nucleic acids was observed. A similar increase of the nucleic acid related peaks in cancerous cervical cells was noticed in previous studies [91, 92] and is confirmed by these observations. Changes in lipid levels are possibly associated with the disruption of the

membrane functionality caused by the virus which influences lipid rafts [93]. A similar behaviour was noted and reported in the previous study on cervical cancer cell lines, wherein both Raman and FTIR (Figure 4) techniques were utilised [63]. Moreover, it was reported previously by Wong et al. that the degree of disorder of methylene chains in membrane lipids increases in cervical neoplastic cells [94]. The absence of any clear relationship between protein levels and HPV copy number is expected, as the presence of HPV in the cell results in functional over- and under- expression of only select biomarker proteins [73], while the level of the other proteins would not be affected. The variation of the spectroscopic features associated with the nuclear lipids, proteins and nucleic acids as a function of the HPV copy number was found to be sublinear and to be well fitted with a similar function to that employed to describe the variation of the p16<sup>INK4A</sup> levels with HPV copy number in figure 7. Notably, the dependences of the cytoplasmic lipids, proteins and nucleic acids on the p16<sup>INK4A</sup> levels, as identified by fluorescence were found to be linear.

In order to further elucidate the multivariate spectral signatures which are specifically related to HPV infection, multivariate regression models were constructed using Partial Least Squares Regression (PLSR) and PLS Jack-knifing was employed as a multivariate feature selection technique [94-96]. The PLSR algorithm seeks to develop a model that relates the spectral data (*X-matrix*) to a series of targets (*Y-matrix*, e.g. concentration of reaction product or analyte) according to the equation  $Y=XB+E$ , where  $B$  is a matrix of regression coefficients and  $E$  is the regression residual. The PLS *Jack-knifing* method developed by Westad and Martens [96] was then used to determine the spectral features that were statistically significant at a particular level of confidence using t-testing of the regression coefficients,  $B$ .

Figure 8 demonstrates that p16<sup>INK4A</sup> levels as measured by fluorescence intensity are approximately linearly correlated with the univariate spectral features and thus, as PLS is a linear model, these values are employed as targets for the PLS analysis. The model can therefore be applied to the FTIR data to elucidate multivariate signatures which are correlated to p16<sup>INK4A</sup> expression level, and therefore, in accordance with figure 7, to HPV infection levels. Once established, these variation patterns can then be applied to unknown samples to screen for the biomarker levels. As p16<sup>INK4A</sup> was found to be predominantly expressed in cytoplasmic regions of the cells, signals recorded from the cytoplasm were utilised in the analysis. Calibration (based on 40% of the data) and test set performance (based on 60% of the data) are presented in figure 9A and demonstrate a very good fit to the model. As shown in figure 9B, the PLS loading exhibits variations originating from proteins, primarily as sub-bands of Amide A (3000-3700cm<sup>-1</sup>) and Amide III (1230-1250cm<sup>-1</sup>), and are associated with the chain of biochemical disruption in protein regulation caused by HPV presence [97]. Another prominent spectral feature differentiating the cell lines, and present in PC loadings and the PLS loading, is the lipid contribution. Again, the presence of HPV is seen to significantly influence the lipid balance within the cell [93], and thus it can be expected that the vi-

brations of lipid component groups differentiate cells with various HPV copy number. This differentiation was distinctly manifest in the mean spectra of the cytoplasm and nucleus (figure 4 – enlarged lipid region of 2800-3050 $\text{cm}^{-1}$ ). Furthermore, the DNA and RNA related peaks are exhibited in the range of 1050-1250 $\text{cm}^{-1}$ , (sym. and asym.  $\text{PO}_2$  str. CO str.) and 3200-3400 $\text{cm}^{-1}$ .

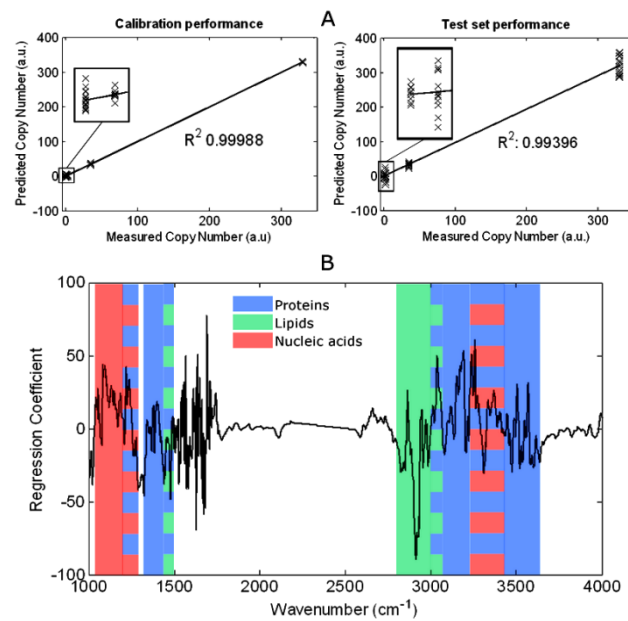


Figure 9: Partial Least Squares analysis results targeting the p16INK4A expression level prediction based on the FTIR spectral features. (Reproduced from [64])

A potential application of vibrational spectroscopy to cervical cancer screening and diagnosis requires a full understanding of the spectral information and its correlation with existing screening and diagnostic methods such as Pap testing or colposcopy. However, the field of the cervical cancer recognition and detection is still developing with new biomarkers [73] being identified that may be considered as adjuncts to existing cervical cytology and pathology methods. Biomarkers such as p16<sup>INK4A</sup> have been found to be useful for low-grade lesions (Benevolo et al., 2006, Passamonti et al., 2010, Negri et al., 2008, Focchi et al., 2007) making them particularly attractive. Thus, a correlation and explanation of the spectral information with these new biomarkers is a new challenge for biospectroscopists working on future cancer diagnostic systems. The study outlined above demonstrates how multivariate regression techniques can help to associate specific spectral signatures with specific biomarkers, potentially paving the way for “spectro-markers” of disease.

### Analysis of the Effects of Anticancer Agents [98, 99]

The previous sections demonstrate how vibrational spectroscopy, coupled with multivariate statistical techniques can be utilized for high sensitivity and specificity disease screening, but also to analyse the underlying biochemical changes at a cellular level. This analytical capability opens up the possibility of a range of *in vitro* applications.

For the design of new anticancer drugs, an in-depth understanding of the mechanisms underlying their biological effects is required. The *in vitro* study of the interaction of anticancer drugs with mammalian cells is therefore important to elucidate the mechanisms of action of the drug on its biological targets and thereby maximise efficacy and guide development strategies. The prevalence of such studies is likely to increase following the EU Directive (2010/63/EU) on the replacement, reduction and refinement (RRR) of animal experimentation in scientific studies. Vibrational spectroscopy has great potential to explore sub-cellular biochemical structure as exemplified by its use in studies investigating the action of various agents on biological macromolecules as well as their interaction with cancer cells [100-106]. It can provide high content information as it can examine spectral changes in the cell membrane, cytoplasm and nucleus of the target cells simultaneously and highlight the biochemical interaction of external agents with the cell or its physiological response [98, 99]. Validation of the technique of Raman spectroscopy in the quantitative measurement of the biochemical and physiological effects of known chemotherapeutic agents is required, however, before it may be used for studies of novel chemical treatments. The study of the spectroscopic signatures of the biochemical effects of commercially available anticancer drugs may be helpful before application for the evaluation of the novel anticancer agents.

Cisplatin (cis-Diamminedichloroplatinum (II)) is a well established chemotherapeutic agent with a known mode of action. In the nucleus, it binds with DNA forming inter-strand and intra-strand crosslinks which lead to cell cycle arrest and apoptosis [107]. The formation of inter-strand and intra-strand crosslinks between cisplatin and DNA leads to conformational changes of the DNA. To understand the changes caused in DNA conformation due to the adducts formed by cisplatin, the formation of 1,2-GG or 1,3-GTG intra-strand cross-links and their characterization by Raman spectroscopy has been carried out [108] and it was found that 1,3 GTG cross linkages induce more distortion in the B-form DNA as compared to the 1,2-GG complex.

Cancer cell lines, such as the A549 human lung adenocarcinoma cell line provide a good *in vitro* model for lung cancer. The high spatial resolution of Raman microscopy can be employed to identify localized spectroscopic changes in the nucleus, cytoplasm and cell membrane and the responses can be correlated with drug dose, as well as physiological response, as determined by the cytotoxicological assay MTT. Feature selection techniques can then be used to identify regions

of the spectrum that are associated with the biochemical effect of exposure to the agent, and with the subsequent cytotoxicological response of the cells.

The MTT assay is a measure of the mitochondrial metabolic activity and can be used initially as a standard against which to map the spectroscopic response of the cells. Maximum cytotoxicity was observed after 96hrs exposure and thus this time period was employed for all measurements. The cytotoxicity induced in A549 cells due to the exposure to cisplatin for 96hrs, as determined by the MTT assay, is shown in Figure 10, where the level of viability in each sample was normalised to that in the control sample. Due to the action of the drug, the mitochondrial activity decreases monotonically which in turn leads to a decrease in cell viability. The Inhibitory Concentration ( $IC_{50}$ ) value was derived from the data by a fit of  $f(x) = \min + (\max - \min) / (1 + (x/IC_{50})^n)$  and found to be  $1.2 \pm 0.2 \mu M$ , which is consistent with the literature. For A549 cells exposed to test drug concentration (TDC) for 72 hrs,  $IC_{50}$  values of  $3.59 \mu M$  and  $2.2 \mu M$  are reported elsewhere with MTT and ATP assays respectively [109]. Also, Cordes et al. have determined an  $IC_{50}$  value of  $2.0 \mu M$  for cisplatin ( $0.1$ - $50 \mu M$ ) exposed A549 cells using the colony formation assay [110].

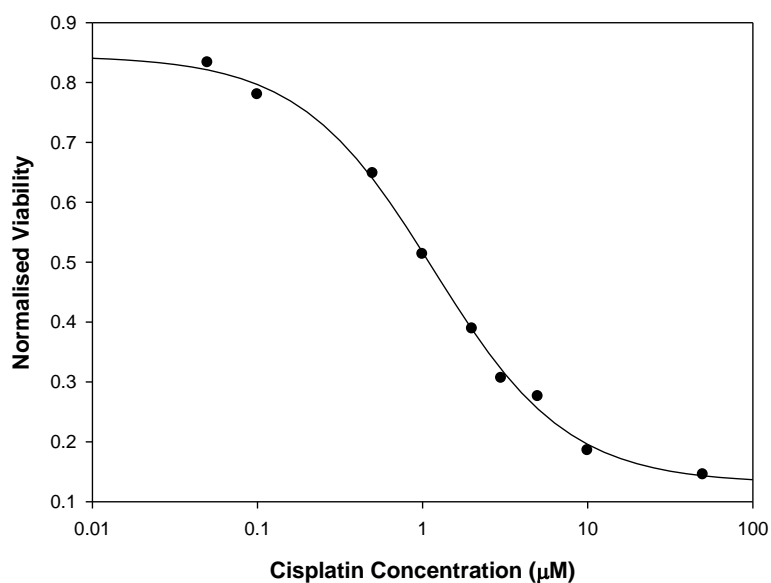


Figure 10. A549 cell viability (measured by MTT absorption) at 96 hours after exposure to Cisplatin. Error bars denote the standard error on the mean at each concentration. (Reproduced from [98])

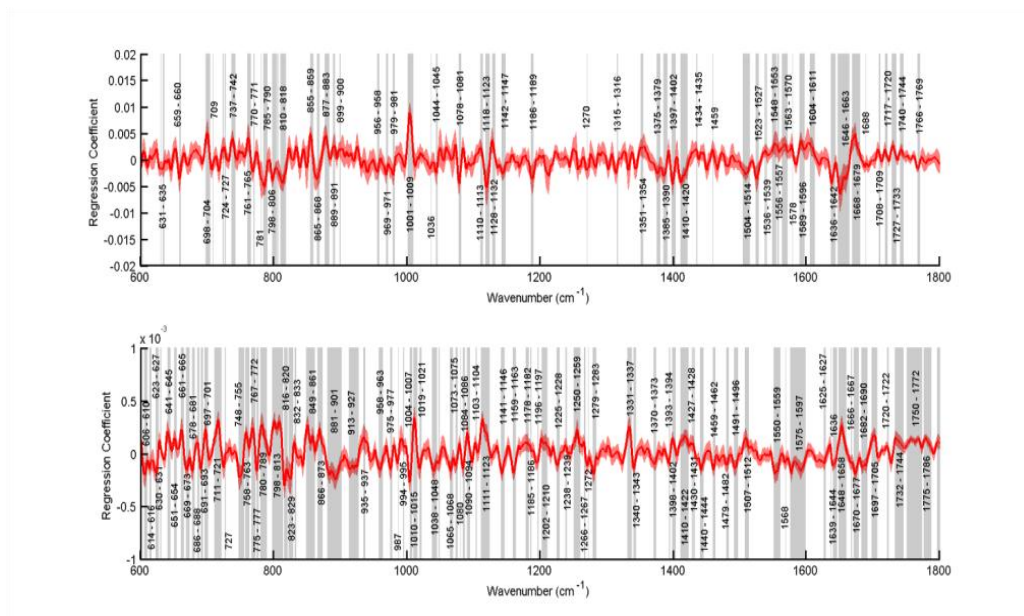


Figure 11: Feature selection in PLSR of Raman spectra of the A549 nucleus against (A) cisplatin concentrations and (B) against cellular viability. The shaded areas denote statistically significant features selected using PLS Jack-knifing with a t-test ( $p < 0.001$ ) of regression coefficients (shown in red with associated standard error). (Reproduced from [98])

While some indication of the localised biochemical impact of the external agent can be gleaned from the mean spectra of the cellular regions [98], the spectral changes are multivariate and may have complex exposure dependence. They include the fingerprint of the interaction of the cisplatin with the nuclear DNA, but also evident are a multitude of other changes associated with exposure and changes in cellular physiology. In an effort to identify the most important spectral features, feature selection was performed using the PLS – Jack knifing analysis described earlier. The regression co-efficients were obtained for separate regressions against cisplatin concentrations and against cell viability with the nuclear spectral data. To elucidate spectral changes that are most statistically significant with respect to regression against either endpoint, a t-test ( $p < 0.001$ ) was applied on the PLS-Jack knifing results. The spectral features identified by this process are highlighted by vertical bands in Figure 11.

The PLS regression of the nuclear data against cisplatin concentration (Figure 11A), identifies many features but, for the purpose of this discussion, only the changes which are related to the known action of cisplatin are highlighted. Features identified include the thymine ( $737\text{--}742\text{ cm}^{-1}$ ,  $1589\text{--}1595\text{ cm}^{-1}$ ), guanine ( $1548\text{--}1553\text{ cm}^{-1}$ ) and adenine ( $1578\text{ cm}^{-1}$ ,  $1604\text{--}1611\text{ cm}^{-1}$ ) bands. The regression co-efficients of guanine and thymine are stronger than that of adenine, indicating

that cisplatin binds preferably with guanine as compared to adenine [111]. Another interesting observation is that the bands at ( $877\text{-}883\text{ cm}^{-1}$ ) and at ( $899\text{-}900\text{ cm}^{-1}$ ), both assigned to deoxyribose ring breathing, are significantly changed, indicating a change in the conformation of B-form DNA [112-114]. Changes in the Raman bands of the proteins are also highlighted, as indicated by the regression coefficients which can be assigned to tryptophan ring breathing ( $761\text{-}765\text{ cm}^{-1}$ ), ( $889\text{-}891\text{ cm}^{-1}$ ), C-N stretching ( $1128\text{-}1132\text{ cm}^{-1}$ ), C-H deformation ( $1315\text{-}1316\text{ cm}^{-1}$ ,  $1351\text{-}1354\text{ cm}^{-1}$ ), amide-I ( $1668\text{-}1679\text{ cm}^{-1}$ ). These changes may indicate some structural changes in the nuclear proteins. The spectroscopic features selected by the PLS regression against drug concentration are consistent with the known direct interaction of the drug with the DNA and thus are identified as spectroscopic markers of the direct chemical interaction of the drug.

The regression against cell viability (Figure 11B), shows the features which are correlated with the change in cell physiology. These include the conformational changes of the DNA B-form ( $832\text{-}833\text{ cm}^{-1}$ ) and DNA A-form ( $798\text{-}813\text{ cm}^{-1}$ ). These changes in the Raman bands of the DNA can be taken as markers for the conformational changes in the DNA caused by cisplatin binding to the guanine and thymine DNA bases (Vrana et al., 2007) and also confirm the observations from the mean difference spectra. Changes related to protein secondary and tertiary structures are indicated by the selection of amide-III random coil ( $1225\text{-}1228\text{ cm}^{-1}$ ), amide-III  $\beta$ -sheet ( $1250\text{-}1259$ ), amide-III  $\alpha$  helix, tryptophan ring breathing ( $711\text{-}721\text{ cm}^{-1}$ ,  $758\text{-}763\text{ cm}^{-1}$ ,  $767\text{-}772\text{ cm}^{-1}$ ) and phenyl ring stretching modes ( $1202\text{-}1210\text{ cm}^{-1}$ ) together with the C-C skeletal vibrations ( $935\text{-}937\text{ cm}^{-1}$ ,  $1159\text{-}1163\text{ cm}^{-1}$ ), C-N stretching vibrations ( $1065\text{-}1068$ ,  $1073\text{-}1075$ ) and C-H bending ( $1178\text{-}1182\text{ cm}^{-1}$ ) ( $1185\text{-}1186\text{ cm}^{-1}$ ). These features suggest changes in the Raman bands of proteins and are indicative of structural alterations associated with protein activation and denaturation during cellular apoptosis and nucleic acid repair responses.

Comparing the regression co-efficients obtained via spectral regression against cisplatin concentration (which highlight the primary biochemical effect of the action of the agent) with those obtained via regression against cell viability (which highlight biochemical markers of the change in the viability of the cell), very different spectral features are selected as being statistically significant and thus allow the differentiation of the chemotherapeutic response. The spectral features of thymine ( $737\text{-}742\text{ cm}^{-1}$ ,  $1589\text{-}1595\text{ cm}^{-1}$ ), guanine ( $1548\text{-}1553\text{ cm}^{-1}$ ) and adenine ( $1578\text{ cm}^{-1}$ ,  $1604\text{-}1611\text{ cm}^{-1}$ ) emerge as prominent as a result of regression against the concentration of the chemotherapeutic agent and are attributable to direct binding of the cisplatin molecule with these DNA bases. In addition, changes related to proteins feature in the regression against the concentration of the chemotherapeutic agent are significantly different than those changes of protein associated bands which appear as a result of regression against the cellular viability. The changes in the features of the proteins due to the chemical effect are attributable to structural alterations in nuclear histone proteins due to cisplatin interactions with the cellular DNA. These features are not selected as being statistically significant by the anal-



ysis for regression against the viability endpoint, suggesting that they are markers of the biochemical interaction between cisplatin and DNA. Contrastingly, the conformational changes to the DNA A-form ( $798\text{--}813\text{ cm}^{-1}$ ) and DNA B-form ( $832\text{--}833\text{ cm}^{-1}$ ), among the changes associated with the protein features, especially the amide-III random coil ( $1225\text{--}1228\text{ cm}^{-1}$ ), amide-III  $\beta$ -sheet ( $1250\text{--}1259$ ), amide-III  $\alpha$ -helix, among others, are selected as being statistically significant in the regression against cellular viability, suggesting that they are markers of this physiological endpoint in the chemotherapeutic response. Overall, Figure 11 demonstrates that the Raman spectrum of the cell may be resolved into distinct groups of statistically significant spectral markers for distinct types of effect. A similar analysis of the spectral changes as a result of the action of the chemotherapeutic agent in the cytoplasm and cell membrane can be performed [99]. Raman microspectroscopy can therefore clearly fingerprint the effects of the interaction of the chemotherapeutic agent with DNA, and identify that fingerprint within the sub cellular spectra of exposed cells.

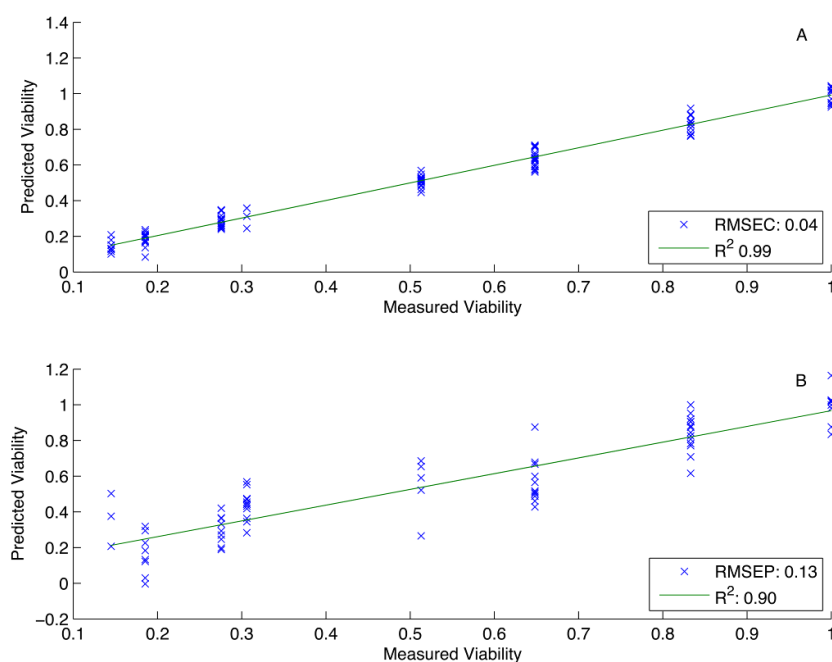


Figure 12: Calibration (A) and test set performance (B) in prediction with PLSR of the cell viability using Raman spectra of the nucleus of A549 cells. Root mean squared errors of calibration (RMSEC) and prediction (RMSEP) are in units of cell viability, which themselves are normalized to the level in the control. (Reproduced from [98])

PLSR models of A549 nuclear spectra versus concentration and reference viability measurements from the MTT assay can then be constructed to determine the ability of the spectral data to predict the level of exposure to cisplatin and the associated level of physiological effect from Raman spectra. All spectra were compiled into a matrix, and were randomly sorted. A total of 60% of the spectra were used to train the PLSR model and 40% of the total was retained as an independent test set to assess the performance of the model in predicting the level of exposure, and the cellular viability, with unseen data. Leave-one out cross validation with the calibration set was used to determine the optimal model complexity for use in testing [115]. This process was performed on fifty separate occasions, with randomization of the data matrix and splitting of the data on each occasion to prevent data bias [116]. The results of prediction of levels of cellular viability is shown in Figure 12, for the example of the nuclear region, where the values of the root mean square error of calibration (RMSEC) and the root mean square error of prediction (RMSEP) denote the prediction uncertainty. The mean values of the RMSEC and RMSEP for the PLSR for prediction of the level of exposure to cisplatin were found to be 1.67  $\mu\text{M}$  (0.45) and 3.41  $\mu\text{M}$  (0.41) respectively, where the figures in brackets denote the standard deviations of these estimates. The associated values of RMSEC and RMSEP for PLSR against normalized cellular viability were 0.05 (0.008) and 0.11 (0.011) respectively. The RMSEP for prediction of cisplatin concentration is therefore 6.8% over the full scale range (0 to 50  $\mu\text{M}$ ), and 11% over the full scale range of viability (from 0 to 1). The proportionally higher observed errors of prediction of cell viability may be due to the development of cellular resistance to cisplatin over the exposure period in the low dose groups [117, 118], which contributes to a non-linearity in the variation in viability with concentration that is difficult for the PLSR algorithm to model.

The study demonstrates the potential of commercially available benchtop Raman microscopy for the examination and prediction of chemotherapeutic responses. In general, for subcellular resolution, Raman spectroscopy is chosen over infrared spectroscopy due to the intrinsically higher spatial resolution. Furthermore, the relative insensitivity to water renders live cell imaging of real time exposures a possibility [119]. Nonlinear optical imaging techniques [120] promise even higher spatial resolution in the future, but at present such short pulse systems cannot be considered benchtop techniques.

Cisplatin, a DNA major groove binder [121], was chosen as a model as its mechanism of action is well characterised. To fully validate the technique, the study should be extended to other chemotherapeutic agents of known interaction mechanisms, such as tallimustine, a known DNA minor groove binder, and ethidium bromide, a known DNA intercalator. In this way the spectroscopic signatures of their interaction can be established and subsequently can act as a database against which novel chemotherapeutic agents can be evaluated. Ultimately, the techniques should be validated in tumour tissue.

### **Live cell imaging [122]**

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 [101]. 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 [123, 124]. The spatial resolution is of the order of 1-2  $\mu\text{m}$ , providing access to the sub-cellular organisation of the cells at a molecular level [125, 126].

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 [127] and chemical maps can be constructed giving information about the distribution of differentiating molecules such as proteins [128]. 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 [129, 130]. Ultimately, if subtle changes in metabolism due to for example chemotherapeutic agents, toxicants, or radiation are to be discerned, live cells should be studied.

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) [131-133]. 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 [134]. In terms of cellular viability and spectral quality, 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 a 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 [119]. 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.

After growing A549 and HaCaT (immortalised human keratinocyte) cells for 24 hours on collagen gels, the medium was substituted by NaCl before mapping, as the phenol red of the cell growth medium limits the visibility of the cells [119]. Examples of photographs obtained using a Horiba Jobin - Yvon Labram HR800 equipped with a x100 immersion objective are presented in figure 13A and 13C. The different structures of the cells can be identified; the cell membrane is clearly

defined for both cell lines and the nucleus is located in the centre part of the cells. The cytoplasm appears smaller in the case of the HaCaT cell. 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. 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 [135, 136], endoplasmic reticulum[136, 137], the mitochondria [138] or the Golgi apparatus[139, 140].

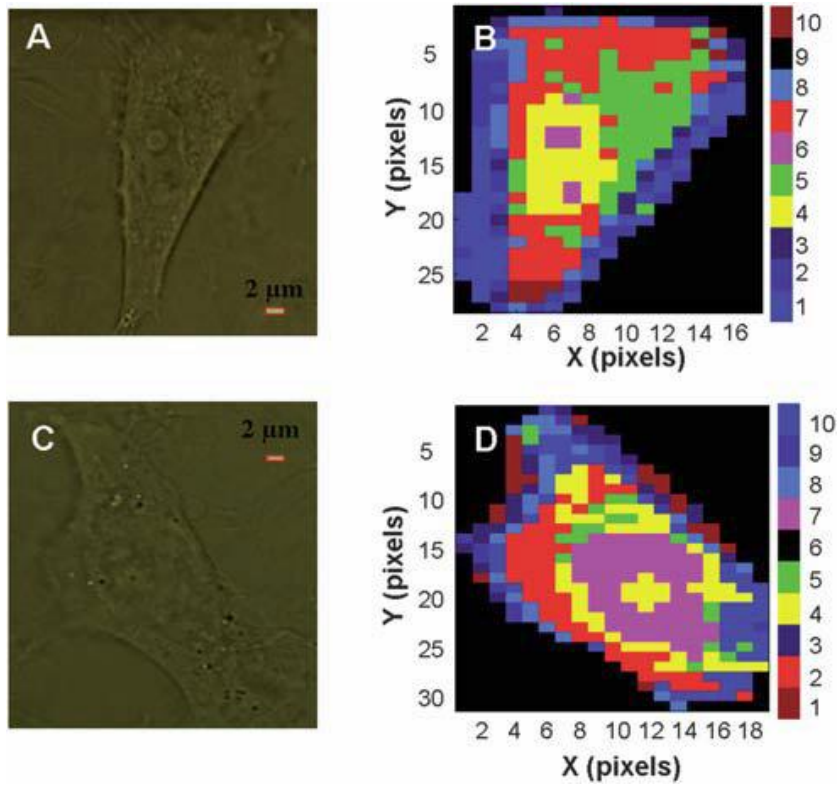


Figure 13: A and C are examples of images taken from the Labram HR 800 equipped with x100 immersion objective respectively from a A549 cell and a HaCaT cell. B and D are the corresponding false colors images obtained after K-means cluster analysis of Raman maps . Each pixel of the Raman map has been attributed to a specific cluster. The cluster numbers are indicated in the colour code to the right of the figure. (Reproduced from [122])

Raman spectral maps were performed in immersion using a Horiba Jobin Yvon LabRam HR800 using a 785nm laser as source. The step between two successive

measurements was set to 1.5  $\mu\text{m}$  and the backscattered Raman signal was integrated for 10 seconds over the spectral ranges from 400 to 1800  $\text{cm}^{-1}$  and accumulated twice to improve the signal to noise ratio. In order to visualise the subcellular structures in the spectral maps recorded, K-means clustering analysis was employed.

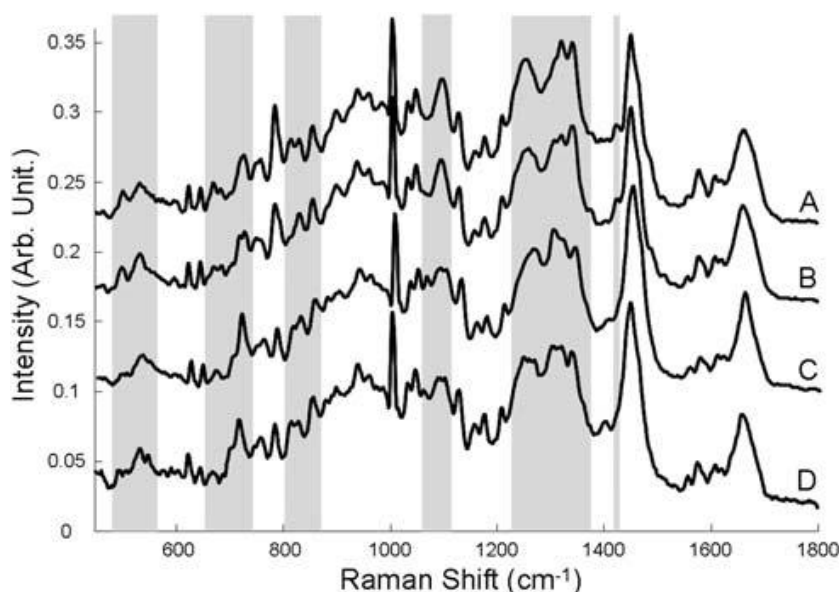


Figure 14: Average spectra obtained from different cluster after K-means clustering analysis. A: nucleus; B: nucleolus; C: cytoplasm cluster 1 and D: cytoplasm cluster 2. (Reproduced from [122])

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 [141]. 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 colour maps can be constructed to visualise the organisation of the clusters in the original image.

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 [142-144]. 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 [127, 145, 145]. The effects of cell fixation on the cellular composition has been already investigated [129, 130] 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 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 13B and 13D. A correspondence between the organisation of the clusters and the structure present in the cells can be seen. In figure 13B, 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 varia-

tions between the different clusters can be visualised through the mean spectra of the clusters. Figure 14 presents the average spectra obtained from selected clusters of the K-means analysis performed on an A549 cell (Figure 13B). (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 14).

The characterisation of live cells on 2D substrates is well documented [123, 124, 147-149] and, based on literature; assignments of the different peaks can be performed. For example, the variations in the spectral range  $718-725\text{ cm}^{-1}$  are related to the lipids and nucleic acids, between  $815$  and  $830\text{ cm}^{-1}$  to RNA and proteins, from  $1088\text{ cm}^{-1}$  to  $1102\text{ cm}^{-1}$  to the nucleic acids, the lipids and carbohydrates, from  $1243-1266\text{ cm}^{-1}$  to the nucleic acids, the proteins and lipids, from  $1296-1322\text{ cm}^{-1}$  to the lipids and from  $1315-1348\text{ cm}^{-1}$  to the nucleic acids and proteins. In analysing the cluster average spectra, it is important to note that, with a  $\sim 1.5\text{ }\mu\text{m}$  spot size, 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 mitochondria, 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  $1423\text{ cm}^{-1}$ , which is completely absent in the cytoplasmic spectra.

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 [150]. Similar results are obtainable for the HaCaT cell line. Figure 15 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 15A 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 15C) 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 (figure 15B) 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 15A. 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 15C. 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.

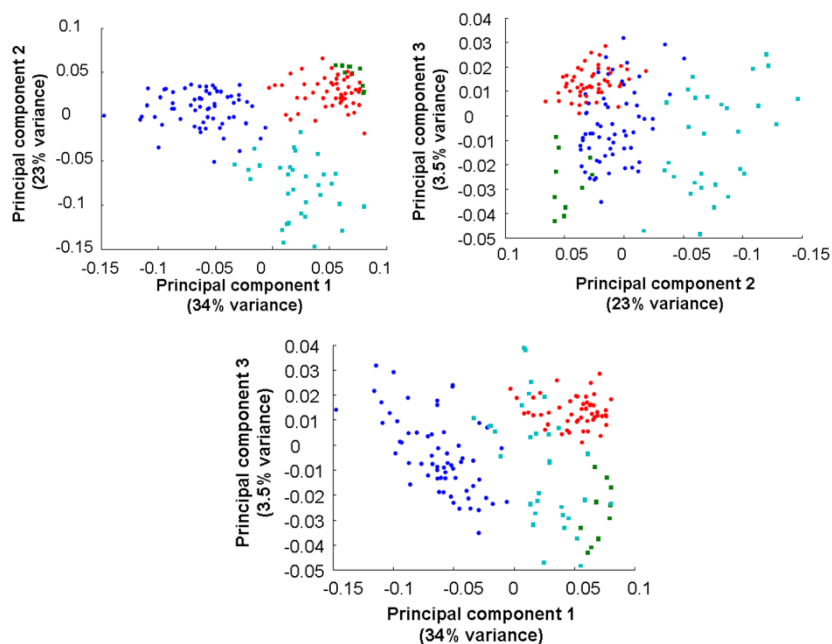


Figure 15: Plots of the first 3 principal components after PCA performed on Raman spectra recorded from A549 cells grown on collagen gels. The spectra have been selected for the cluster corresponding to the nucleus (red), nucleolus (green) and cytoplasm (cluster 1: dark blue; cluster 2: light blue). (Reproduced from [122])

The loading is a representation of the variability described by a principal component as a function of wavenumber of the spectra. Figure 16A 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 15. 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 16 B,C,D and E, offset for clarity, are from DNA, RNA, L- $\alpha$ -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 16A. Each of them has a correspondence with one or more of the four spectra from compounds tested. For example, the peak at 791  $\text{cm}^{-1}$  is prominent in the spectra of DNA and RNA. In figure 14, 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  $\text{cm}^{-1}$ , which is prominent in the spectra of the L- $\alpha$ -Phosphatidylcholine and Sphingomyelin indicating significant variations in lipidic content. The loading of PC2, plotted in figure 17A is the basis of the discrimination between compartment number 2 of the cytoplasm and the nucleus and nucleoli (figure 15A). Once again the comparison with the spectra of pure compounds highlights peaks matching with the loading. The highest variability can be seen at 699  $\text{cm}^{-1}$  and 1436  $\text{cm}^{-1}$ , both attributed to the lipid content, and at 1342  $\text{cm}^{-1}$ , related to the DNA and RNA. The loading of PC3 (figure 18A) represents the variability between the nucleus and the nucleolus (figure 15B and 15C). Many peaks can be identified as related to the DNA and RNA, especially at 651, 813, 1091, 1482 and 1572  $\text{cm}^{-1}$ , showing that the main difference between the spectroscopic signatures of the nucleus and nucleolus is the DNA/RNA content.

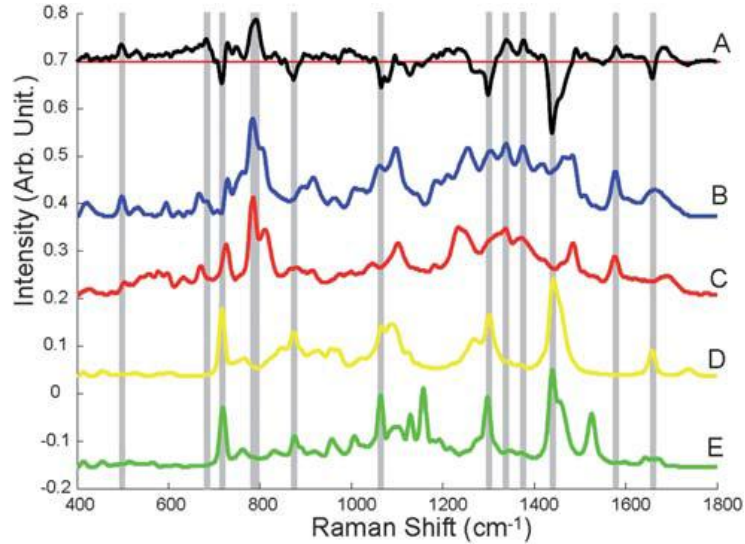


Figure 16: Plot of the loadings number 1 (A) corresponding to the first principal component resulting of the PCA analysis. This loading has been compared with spectra recorded from different compounds: DNA (B) RNA (C) L- $\alpha$ -Phosphatidylcholine (D) and Sphingomyelin (E). (Reproduced from [122])

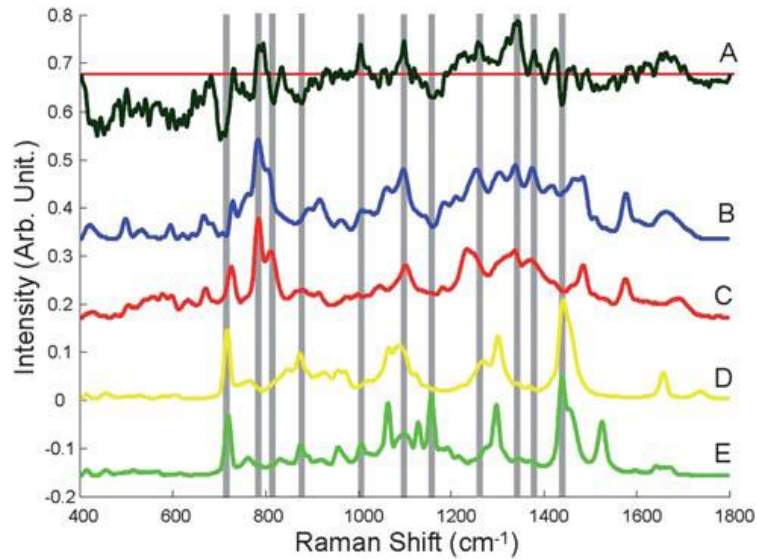


Figure 17: Plot of the loadings number 2 (A) corresponding to the second principal component resulting of the PCA analysis. This loading has been compared with spectra recorded from different compounds: DNA (B), RNA (C), L- $\alpha$ -Phosphatidylcholine (D) and Sphingomyelin (E). (Reproduced from [122])

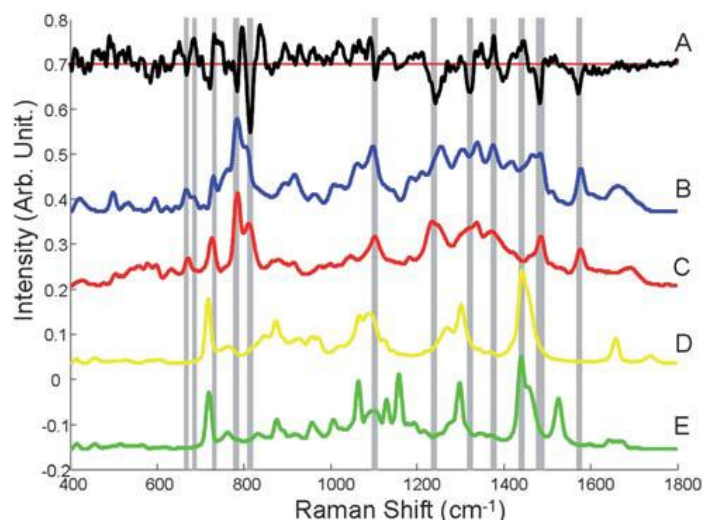


Figure 18: Plot of the loadings number 3 (A) corresponding to the third principal component resulting of the PCA analysis. This loading has been compared with spectra recorded from different compounds: DNA (B) RNA (C) L- $\alpha$ -Phosphatidylcholine (D) and Sphingomyelin (E). (Reproduced from [122])

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 other 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.

### NanoParticles in Cells [151]

The potential of vibrational, particularly Raman, spectroscopy as a realtime label free subcellular analytical technique opens up a range of applications, as for example in the field of nanotoxicology and nanomedicine. One of the major concerns regarding the possible toxic effects of nanoparticles is the capacity of these materials to penetrate cells and possibly translocate to other cells around the body. Conventional *in vitro* cytotoxicity assays, such as Alamar Blue, Neutral Red, MTT, etc, provide indications of impact on cell proliferation, viability, metabolic activity, lysosomal and mitochondrial activity. However, they fail to provide a verification of nanoparticle internalisation or elucidate intracellular trafficking mechanisms and subcellular distribution within cells. Indeed, false positive results have been demonstrated due to the extracellular interaction of the nanoparticles with the cell culture medium *in vitro*, and with the cytotoxic assays themselves [152, 152]. The underlying principles of internalisation and distribution of nano-

particles within cells and any associated toxicity to humans are still relatively poorly understood. However, there are numerous studies that have demonstrated the ability of nanoparticles to cross membranes and internalise within many different cell types [153-155]. In order to assess potential risks, but also benefits in terms of nanomedicine, a greater understanding of migration of nanoparticles intra- and inter-cellularly and essentially from one part of the body to another, along with information on the translocation of nanoparticles within organelles, is urgently required.

Among the main challenges faced by nanotoxicologists are the detection and identification of nanoparticles that have crossed the cell membrane and monitoring of their intracellular trafficking. Imaging cells exposed to fluorescently labelled nanoparticles using confocal fluorescence microscopy is one of the most common ways in which to ensure nanoparticles can be tracked and monitored as they enter and localise within cells [156-158]. However, not all nanoparticles can be easily fluorescently labelled. Furthermore, there have been reports that labelled nanoparticles can release the dye into the surrounding biological environment, and so the distribution of fluorescence within the cell does not necessarily represent the presence or subcellular distribution of the nanoparticles [159-161]. Furthermore, it is not clear that the transport mechanisms of smaller nanoparticles, fluorescently labelled with anionic moieties, are the same as their unlabelled counterparts. Transmission Electron Microscopy can also be employed to visualise nanoparticles within cells, but significant sample processing (fixing and ultramicrotoming) is required and only particles with sufficient electronic contrast to the cellular environment can be visualized [162, 163]. Thus, there is a requirement for a sensitive technique to localise and identify nanoparticles internalised in cells, ideally based on their chemical composition, rather than fluorescence labels or electronic contrast. Identification of their local environment (e.g. endosomes, lysosomes, mitochondria) could aid in elucidating their intracellular trafficking and interaction mechanisms, and the resulting changes in cellular metabolism.

Raman spectroscopy potentially offers a label free probe of nanoparticles within cells, which can potentially analyse their local environment, and ultimately changes in the cellular metabolism which can be correlated with cytotoxic responses [164], oxidative stress, or inflammation. Kneipp et al. have previously demonstrated the use of Surface Enhanced Raman Scattering from gold nanoparticles and nanoaggregates to probe the environment of the subcellular compartments through which they are trafficked [165, 166]. They have also demonstrated the use of molecular labelled nanoparticles as more specific probes of the local environment [167-169]. However, the uptake rates and mechanisms as well as the subsequent trafficking may be specific to the nanoparticle type, size and surface chemistry. Furthermore, the molecular specificity of the surface enhancement process is not well understood. Therefore, a truly label free method for monitoring and characterising the cellular uptake and subcellular localisation of nanoparticles in general is still required.

It has been demonstrated that, based on the spectroscopic fingerprint of polystyrene, Raman spectroscopy, coupled with routine multivariate analyses such as K-means clustering, is a more specific probe to detect and localise nanoparticles within cells [151]. Furthermore, Principal Component Analysis can be employed to differentiate the local environment of the nanoparticles from the cytoplasm. For convenience, formalin fixed (A549) cells were employed. Raman mapping was performed in immersion using a x100 Olympus LUMPlanF1 objective, a wavelength of 785nm, and a step size of 0.75 $\mu$ m. As seen in figure 13, using a 100x objective, the visible image obtained from the cells is rather detailed (Figure 19 IA). The nucleus is clearly identifiable and the nucleoli can also be seen. The membrane of the cells is well defined and the cytoplasm presents different structures and the region containing the endoplasmic reticulum can be easily recognised. As in the previous section, K-means cluster analysis of the Raman spectral map using 10 clusters clearly differentiates the sub-cellular regions of interest (Figure 19 IB).

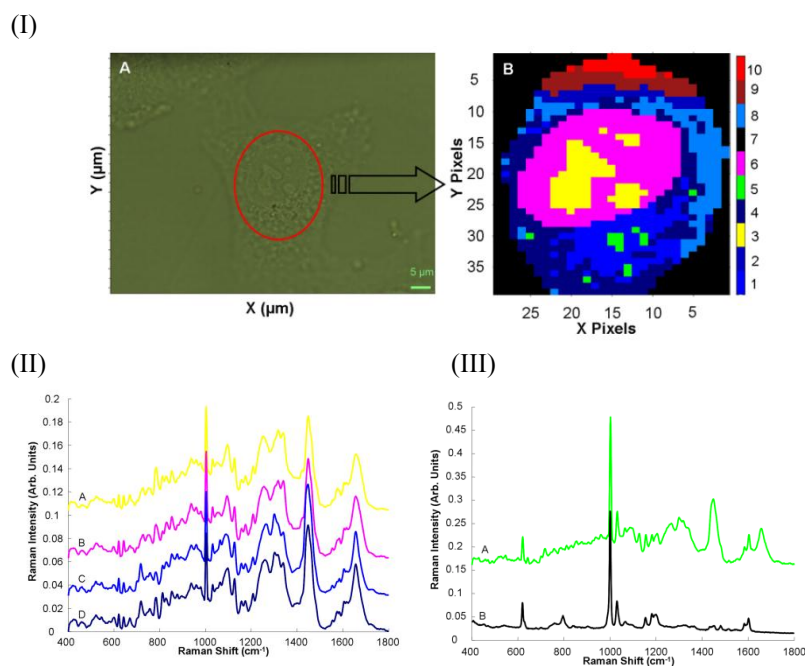


Figure 19: I; (A) Microscopic image of an A549 cell, showing the reduced area identified for spectral mapping. (B) K-means cluster map of the Raman profile of the same reduced area. II; K-means spectra of clusters 3 (A – representing nucleoli), 6 (B – representing nucleus), 1 and 4 (C and D, both from the cytoplasm). Spectra are offset for clarity. III; K-means spectrum of Cluster 5 (A), compared to the Raman spectrum polystyrene nanoparticles (B). Spectra are offset for clarity. (Reproduced from [151])

The nucleoli spectra form a specific group (cluster 3), distinct from the nucleus (cluster 6). The corresponding mean spectra are displayed in figure 19 II. The cytoplasm is represented by a number of different clusters (1, 2, 4, 8, 9, 10). Notably, nanopolyesterene can now be localised in the cytoplasm according to cluster 5, the corresponding mean spectrum of which is presented in figure 19 IIIA. Although some cellular features clearly contribute to the cluster spectrum, it is dominated by specific features originating from polystyrene (figure 19 IIIB). K-means cluster analysis yields the average spectra for the identified clusters. These average spectra can contain contributions from points on the boundaries between the cellular regions and a direct comparison does not always easily facilitate identification of the differences in the biochemical composition. Cluster 5 clearly identifies the presence of polystyrene nanoparticles localised in the cytoplasm, but the averaged spectrum also exhibits clear contributions of the biological environment. PCA can be employed to highlight the biochemical differences between the subcellular regions [150]. In figure 19 IB, cluster 5 is predominantly surrounded by cytoplasmic regions grouped within cluster 1. Figure 20 (top) is a scatter plot of the PCA of spectra associated with K-means clusters 1 and 5. For visualisation purposes, spectra associated with cluster 1 are coloured blue, while those from cluster 5 are coloured green, and although the data is rather dispersed, some differentiation between the two clusters according to PC1 is apparent. PC1 accounts for 61% of the variance and, as shown in the loadings plot of figure 20 (bottom), it is dominated by positive contributions of polystyrene (figure 20 B). Notably, the negative contributions of PC1 match well with the Raman features of lipids such as L- $\alpha$ -Phosphatidylcholine (figure 20 C) and also sphingomyelin (figure 20 D). For comparison, the Raman spectra of DNA (figure 20 E) and RNA (figure 20 F) are also shown, and it is clear that they exhibit few or no similarities with PC1. High loadings of PC1 for cluster 5 are therefore an indication of the dominance of polystyrene. Positive loadings of PC1 for some of Cluster 1 may indicate regions of overlap between the localised nanoparticles and the surrounding cytoplasm, which, although they have significant polystyrene contributions, are represented in the K-means cluster analysis by the average spectrum of cluster 1.

An established mechanism for the transport of nanoparticles in cells is via endosomes and later lysosomes [170]. A number of studies have also demonstrated that, in the later stages of trafficking, individually endocytosed nanoparticles accumulate in larger multivesicular bodies [165] and they have also been seen to be localized in the endoplasmic reticulum and Golgi apparatus [171]. Notably, in figure 19 IA, the majority of the nanoparticles, identified in green as cluster 5, are surrounded by cytoplasmic regions of cluster 1, which is relatively spatially extensive, particularly compared to typical sizes of lysosomes (1-2 $\mu$ m). In the white light image of figure 19 IB, this region can be identified as the Endoplasmic Reticulum (ER), which may be expected to be rich in lipidic compounds, and a PCA of the different regions of the cytoplasm confirms this. Cluster 4, on the other hand is relatively rich in nucleic acids, which may be due to the presence of mitochondria. Co-staining the cells to highlight the ER and examination using confocal

fluorescence microscopy indicates a strong correlation between the ER stain and the fluorescence of the nanoparticle label, and any locally concentrated fluorescence which may be indicative of encapsulated nanoparticles is contained within the ER [151]. The analysis supports the attribution of the subcellular environment of K-means cluster 1 as endoplasmic reticulum.

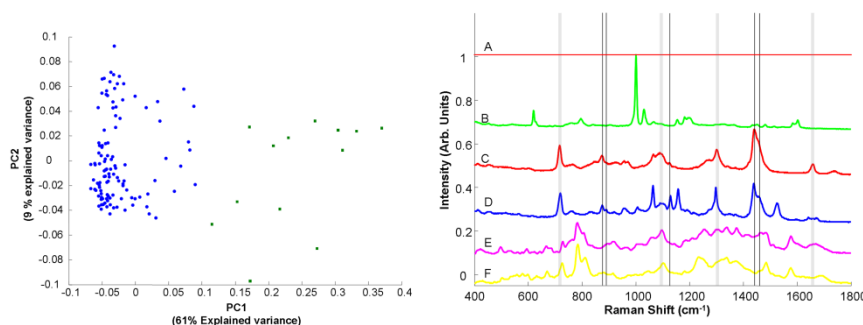


Figure 20: Top; Scatter plot of the PCA of spectra associated with clusters 1 (Blue) and 5 (Green). Bottom; (A) Loading of PC1 (B) Raman Spectrum of polystyrene nanoparticles (C) of L- $\alpha$ -Phosphatidylcholine, (D) of Sphingomyelin, (E) of DNA and (F) of RNA. (Reproduced from [151])

The study represents an initial proof of concept, to demonstrate the potential use of Raman spectroscopy to identify and locate nanoparticles in cells and to probe their subcellular environment. No attempt has, as yet, been made to quantify the amount of nanoparticles located or the sensitivity and detection limit of the technique. Instrumentally, the spatial resolution can be improved by moving to lower wavelength sources, although the excitation of any fluorescence from the nanoparticles and the cell constituents should be avoided. However, as it scales linearly with wavelength, a move from 785 nm to 473 nm should result in an improvement in lateral resolution by a factor of  $\sim 1.7$ . Ultimately, the confocal mode of operation can provide 3D localisation of the nanoparticles at submicron resolution.

Other multivariate analytical approaches have also been applied in the field of Raman microspectroscopy of cells. In addition to KMCA, other clustering methods such as Fuzzy C means clustering (FCM) and hierarchical cluster analysis (HCA) have been used to separate the cellular Raman data into clusters and subsequently reshape the data into images [172, 173]. However, as highlighted by Headegaad *et al.*, these approaches have their own limitations. In particular, boundaries between sub-cellular features can often result in the addition of extra clusters with mixed spectral signatures. This can be overcome by increasing the number of clusters; however, this in turn can result in added complexity to interpretation and inaccuracies in regional separation. Additionally, the reproducibility

of these methods can also be questioned as the starting point for the centroid based KMCA and FCM is subjective [172].

PCA and vertex component analysis (VCA) have also been used to separate out distinct regions of the cell. With regards to PCA, separation is based on the variances between the spectra in the data set, the majority of the variance being described by the first three principal components [172]. Thus, the score values can be used to construct a composite image of the cell in which the biochemical contributions of each component are described by the corresponding loadings plot. Unlike KMCA and FCM, PCA identifies quite accurately the boundaries between each feature. However, the images generated suffer from inferior contrast and in some instances interpretation may be difficult as biochemical features may be spread across different loadings.

VCA is another method which has been used for similar analytical purposes. In brief, VCA computes a linear combination of supposed pure component spectra which are termed endmember spectra. As described in Miljkovic *et al.*, the endmember spectra are acquired under the assumption that the most extreme data points in the dataset are representative of pure component spectra [173]. However, it has been pointed out that the endmembers generated are not truly representative of the pure component they describe in the data set and can often contain a mixture of biochemical constituents i.e. DNA and proteins. While this is representative of the true nature of nucleic acids *in-situ*, it could lead to inaccuracies in interpretation.

A more recent study introduced the technique of spectral cross correlation to screen spectral maps for the spatial distribution of specific spectral signatures, including those of polystyrene nanoparticles, RNA and lipids, and argued that it was more effective than Classical least squared analysis, both supervised and unsupervised, which is commonly available on instrument software [174].

As in the case of multivariate techniques for disease diagnostics, there is a myriad of potential techniques to explore the spatial variations of the spectral profiles with heterogeneous structures such as biological cells. A truly quantitative comparison of the different techniques is lacking, however.

## Conclusions

Since their origins in the 19<sup>th</sup> and 20<sup>th</sup> centuries, the techniques of Infrared absorption and Raman scattering spectroscopy have come a long way. Technological developments have rendered them routine benchtop techniques for chemical and process analysis. Although their applications for the analysis of complex mixtures have raised some challenges, an understanding of the fundamental physical processes and interactions has helped the scientific community address these challenges. In biomedical applications, the chemical and physical inhomogeneities of cells and tissues have presented many such challenges, not easily understood by a multidisciplinary community. However, concerted efforts through collaborative programmes such as the EU FP6 Special Support Action DASIM [175] have helped understand the origins of spectral artefacts which can mask or obscure the



true information contained within a spectral profile, such that such effects can be minimised by improved sample presentation and/or data preprocessing.

Over the past decades, the potential of vibrational spectroscopy for biomedical applications has been extensively explored and demonstrated. The prospect is of a truly label free, non-invasive, non destructive, rapid, automated and objective technology for disease diagnostics and beyond. Disease diagnostics can be made on the basis of true changes in biochemistry rather than for example cellular and tissue morphology, and disease origin and progression can be better understood. Gerwin Puppels, one of the pioneers of the field and founder of River Diagnostics likens the developing technology to that of Dr. McCoy of Star Trek, whereby a simple optical probe can scan a body for all ailments. While it may seem improbable, the emergence of spectroscopic probes for *in vivo* analyses and the development of applications for the analysis of bodily fluids and identification of spectral markers of disease may mean the realisation of real clinical applications in the near future.

Beyond disease diagnostics, EU legislation is placing increasing emphasis on *in vitro* screening techniques for the replacement, reduction and refinement of animal experimentation in scientific studies. High Content Analysis techniques for automation and higher throughput of screening based on chemical and immuno assays have developed significantly over the past decade. However, the assays are highly specific and the process of assaying is extremely expensive. In the realm of *in vitro* diagnostics and therapeutics, vibrational spectroscopy presents as realistic alternative.

The development of applications of vibrational spectroscopy for biomedical/chemical applications has relied heavily on the development and application of multivariate statistical and data mining techniques. A single spot on a cell or tissue contains a multitude of different biochemical components in itself, and the composition and morphology of this complex melange can vary dramatically over a length scale which is comparable to the sampled spot size. Cell to cell, tissue to tissue, or patient to patient variability must also be considered. The challenge of extracting relevant information is therefore immense, and a myriad of techniques has been applied and adapted as exemplified in a multitude of techniques. Given the impact of the conclusions derived, particularly in medical diagnostics, it is imperative that the combination of spectral analysis and data processing yield the “right answer”. In a real life scenario, however, it is impossible to know for certain what the “right answer” is however. In this context, complex simulated data sets which incorporate instrumental variabilities and measurement artefacts, but also sample to sample variabilities may play a critical role in validating and optimising the data analytical techniques. With a combination of optimised spectroscopic technologies, sample presentation protocols and data processing and analysis techniques, the future for vibrational spectroscopy is bright, for disease diagnostics and beyond.

## References

- [1] [http://coolcosmos.ipac.caltech.edu/cosmic\\_classroom/classroom\\_activities/herschel\\_bio.html](http://coolcosmos.ipac.caltech.edu/cosmic_classroom/classroom_activities/herschel_bio.html)
- [2] G.H. Rieke, "History of infrared telescopes and astronomy". *Experimental Astronomy*, **25**, 125 (2009)
- [3] A.L. Smith, *Applied Infrared Spectroscopy: Fundamentals, Techniques, and Analytical ProblemSolving*. Wiley, New York (1979)
- [4] L. Mertz, *Astron J.*, **70**, 548 (1965)
- [5] P.R. Griffiths, R. Curbelo, C.T. Foskett and S.T. Dunn, *Analytical Instrumentation (Inst. Society of America)*, **8**, II4 (1970)
- [6] R. Messerschmidt and M. Harthcock eds, "Infrared Microscopy, Theory and Applications", Marcel Dekker, New York (1988)
- [7] <http://www.mtholyoke.edu/~mlyount/MySites/ForensicSpectroscopy/ForensicApps.html>
- [8] USA Today, June 23 rd 1993, pg8A
- [9] R.C. Lyon, E.H. Jefferson, C.D. Ellison, L.F. Buhse, J.A. Spencer, M.M. Nasr and A.S. Hussain, "Exploring Pharmaceutical Applications of NearInfrared Technology", *American Pharmaceutical Review*, <http://americanpharmaceuticalreview.com/ViewArticle.aspx?ContentID=225>
- [10] C.V. Raman and K.S. Krishnan, *Nature*, **121**, 501 (1928)
- [11] C. Adjouri, A. Elliasmine, Y. Le Duff, *Spectroscopy*, **44**, 46 (1996)
- [12] G.J. Puppels, J.H.F. Olminkhof, G.M.J. Segers-Nolten, C. Otto, F.F.M. de Mul and J. Greve, *Experimental Cell Research*, **195**, 361 (1991)
- [13] M. Jackson and H. Mantsch, *CRC Crit., Rev. Biochem. Mol. Biol.*, **30**, 95 (1995)
- [14] H.L. Casal and H.H. Mantsch, *Biochim. Biophys. Acta*, **779**, 381 (1984)
- [15] M. Mathlouthi and J.L. Koenig, *Adv. Carbohydr. Chem. Biochem.*, **44**, 7 (1986)
- [16] E. Taillandier, J. Liquier and J.A. Taboury, in R.J.H. Clark and R.E Hester (eds.), Vol. 12, "Advances in Spectroscopy: Advances in Infrared and Raman Spectroscopy", Wiley, New York, (1985) p65.
- [17] P.Wong, S. Lacele, H. Yadzi, *Appl. Spectroscopy*, **47**, 1830 (1993)
- [18] R. Dukor, "Vibrational spectroscopy in the detection of cancer, *Handbook of Vibrational Spectroscopy*", Ed J. M. Chalmers and P.R. Griffiths Vol 5, Wiley, Chichester (2002)
- [19] E. Gazi, M. Baker, J. Dwyer, N. P. Lockyer, P. Gardner, J.H. Shanks, R. S. Reeve, C. Hart, N.W. Clarke M. Brown, *European Urology*, **50**, 750–761 (2006)
- [20] D. C. Fernandez, R. Bhargava, S. M. Hewitt, I. W. Levin, *Nature Biotechnol.*, **23**, 469 (2005)
- [21] R.C. Lord, N.T. Yu, *J Mol Biol*, **20**, 50924 (1970).
- [22] M.C. Tobin, *Science*, **161**, 68, (1998).
- [23] A.G. Walton, M.J. Deveney and J.L. Koenig, *Calcified Tissue International*, **6**, 162, (1970).

- [24] N.T. Yu, B.H. Jo, R.C.C. Chang and J.D. Huber, *Arch. Biochem. Biophys.*, **160**, 614 (1974).
- [25] G.J. Puppels and J. Breve, in “Biomedical Applications of Spectroscopy”, R.H.J. Clark and R.E. Hester eds, *Advances in Spectroscopy*, vol. 25, John Wiley and Sons, New York (1996).
- [26] M. Gniadecka, H.C. Wulf, O.F. Nielsen, D.H. Christensen, J. Hercogova, *Photochemistry and Photobiology*, **66**, 418 (1997).
- [27] C.M. Krishna, G.D. Sockalingum, L. Venteo, R.A. Bhat, P. Kustagi, M. Pluot, M. Manfait, *Biopolymers*, **5**, 269 (2002).
- [28] J. Smith, C. Kendall, A. Sammon, J. ChristieBrown, N. Stone, *Technology in Cancer Research & Treatment*, **2**, 327 (2003)
- [29] A. Molckovsky, L.M.W.K. Song, M.G. Shim, N.E. Marcon, B.C. Wilson, *Gastrointestinal Endoscopy*, **57**, 396 (2003)
- [30] E.B. Hanlon, R. Manoharan, T.W. Koo, K.E. Shafer, J.T. Motz, M. Fitzmaurice, J.R. Kramer, I. Itzkan, R.R. Dasari and M.S. Feld, *Phys. Med. Biol.*, **45**, R1-R59 (2000).
- [31] P. J. Caspers, G. W. Lucassen, R. Wolthuis, H. A. Bruining, G. J. Puppels, *Biospectroscopy*, **4**, S31-39 (1999).
- [32] U. Utzinger, A. MahadevanJansen, D. Hinzelman, M. Follen, R. Richards-Kortum, *Applied Spectroscopy*, **55**, 955, (2001).
- [33] <http://www.chem.neu.edu/web/faculty/Prof.MaxDiem.html>
- [34] F.M. Lyng, E. Ó Faoláin, J. Conroy, A. Meade, P. Knief, B. Duffy, M. Hunter, J. Byrne, P. Kelehan, H.J Byrne, *Experimental and Molecular Pathology*, **82**, 121 (2007)
- [35] D.M. Parkin, F. Bray, J. Ferlay and P. Pisani, *Global Cancer Statistics 2002*, *CA Cancer J. Clin*; **55**,74-108 (2005)
- [36] K. Nanda, D.C. McCrory, E.R. Myers, L.A. Bastian, V. Hasselblad, J.D. Hickey, D.B. Matchar, *Ann Intern Med*, **132**, 810 (2000)
- [37] B.H. Willis, P. Barton, P. Pearmain, S. Bryan, C. Hyde, *Health Technol Assess*, **9**, 1-207 (2005)
- [38] N. Bolger, C. Heffron, I. Regan, M. Sweeney, S. Kinsella, M. McKeown, G. Creighton, J. Russell, J. O'Leary, *Acta Cytol.*, **50**, 483 (2006)
- [39] G. Koliopoulos, M. Arbyn, P. Martin-Hirsch, M. Kyrgiou, W. Prendiville, E. Paraskevaidis, *Gynecologic Oncology*, **104**, 232 (2006)
- [40] S.K. Chang, T. Pavlova, N. Marin, M. Follen, R. Richards-Kortum, *Gynecologic Oncology*, **99**, S61-63 (2005)
- [41] R.S. Gurjar, V. Backman, L.T. Perelman, I. Georgakoudi, K. Badizadegan, I. Itzkan, R.R. Dasari, M.S. Feld, *Nat Med*. **7**, 1245 (2001)
- [42] P.F. Escobar, J.L. Belinson, A. White, N.M. Shakhova, F.I. Feldchtein, M.V. Kareta, N.D. Gladkova, *Int J Gynecol Cancer*. **14**, 470 (2004)
- [43] A.F. Zuluaga, M. Follen, I. Boiko, A. Malpica, R. Richards-Kortum, *Am J Obstet Gynecol.*, **193**, 83 (2005)
- [44] K. Carlson, I. Pavlova, T. Collier, M. Descour, M. Follen, R. Richards-Kortum, *Gynecologic Oncology*, **99**, S84 (2005)

- [45] B.R. Wood, L. Chiriboga, H. Yee, M.A. Quinn, D. McNaughton, M. Diem, *Gynecologic Oncology*, **93**, 59 (2003)
- [46] S. Mordechai, R.K. Sahu, Z. Hammody, S. Mark, K. Kantarovich, H. Guter-  
man, A. Podshyvalov, J. Goldstein, S. Argov, *Journal of Microscopy-Oxford*, **215**,  
86 (2004)
- [47] M. Romeo, C. Matthaus, M. Miljkovic, M. Diem, *Biopolymers*, **74**, 168  
(2004)
- [48] A. Mahadevan-Jansen, N. Ramanujam, A. Malpica, S. Thomsen, U. Utzinger,  
R. Richards-Kortum, *Photochemistry and Photobiology*, **68**, 123 (1998)
- [49] E. Ó Faoláin, M.B. Hunter, J.M. Byrne, P. Kelehan, H.A. Lambkin, H.J. Byr-  
ne, F.M. Lyng, *J. Histochem. Cytochem.*, **53**, 121 (2005)
- [50] J.E. Jackson, "A User's Guide to Principal Components", Wiley, New York  
(1991)
- [51] M. Otto, "Chemometrics: Statistics and Computer Applications in Analytical  
Chemistry", Wiley New York (1999)
- [52] P. Crow, A. Molckovsky, N. Stone, J. Uff, B. Wilson, L.M. Wong-KeeSong,  
*Urology*, **65**, 1126 (2005)
- [53] A.S. Haka, K.E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R.R. Dasari, M.S.  
Feld, *Cancer Res.*, **62**, 5375 (2002)
- [54] A. Nijssen, T.C. Bakker Schut, F. Heule, P.J. Caspers, D.P. Hayes, M.H.  
Neumann, G.J. Puppels, *J Invest Dermatol.*, **119**, 64 (2002)
- [55] UK Patent Application No. GB0823071.6, "A Diagnostic Instrument for Cer-  
vical Cancer Screening", December 2008
- [56] International Patent, F. M. Lyng., H. J. Byrne, E. Ó Faoláin, C. Clarke and K.  
M. Ostrowska (2010), "A Diagnostic Instrument for Cervical Cancer Screening",  
Patent no WO2010070133
- [57] M.L. Almond, J. Hutchings, C. Kendall, N. Stone, H. Barr, *British Journal of  
Surgery*, **98**, 35 (2011)
- [58] P. Bassan, H.J. Byrne, F. Bonnier, J. Lee, P. Dumas, P. Gardner, *Analyst*,  
**134**, 1586 (2009)
- [59] P. Bassan, H.J. Byrne, J. Lee, F. Bonnier, C. Clarke, P. Dumas, E. Gazi, M.  
D. Brown, N.W. Clarke, P. Gardner, *Analyst*, **134**, 1171 (2009)
- [60] P. Bassan, A. Kohler, H. Martens, J. Lee, H.J. Byrne, P. Dumas, E. Gazi, M.  
Brown, N. Clarke and P. Gardner, *Analyst*, **135**, 268 (2010)
- [61] F. Bonnier, A. Mehmood, P. Knief, A. Meade, W. Hornebeck, H.  
Lambkin, K. Flynn, V. McDonagh, C. Healy, T.C. Lee, F.M. Lyng, H.J.  
Byrne, *Journal of Raman Spectroscopy*, **42**, 888 (2011)
- [62] F. Bonnier, S.M. Ali, P. Knief, H. Lambkin, K. Flynn, V. McDonagh,  
C. Healy, T.C. Lee, F.M. Lyng, H.J. Byrne, *Vibrational Spectroscopy*, **61**,  
124 (2012)
- [63] K.M. Ostrowska, A. Malkin, A.D. Meade, J. O'Leary, C. Martin, C.  
Spillane, H.J. Byrne and F.M. Lyng, *Analyst*, **135**, 3087 (2010)

- [64] K.M. Ostrowska, A. Garcia, A.D. Meade, A. Malkin, I. Okewumi, J.J. O’Leary, C. Martin, H.J. Byrne and F.M. Lyng, *Analyst*, **136**, 1365 (2011)
- [65] K.S. Cuschieri, H.A. Cubie, M.W. Whitley, G. Gilkison, M.J. Arends, C. Graham and E. McGoogan, *J. Clin. Pathol.*, **58**, 946 (2005)
- [66] W.A. Tjalma, T.R. Van Waes, L.E. Van den Eeden and J.J. Bogers, *Best Pract. Res. Clin. Obstet. Gynaecol.*, **19**, 469 (2005)
- [67] J.M. Walboomers, M.V. Jacobs, M.M. Manos, F.X. Bosch, J.A. Kummer, K.V. Shah, P.J. Snijders, J. Peto, C.J. Meijer and N. Munoz, *J. Pathol.*, **189**, 12 (1999)
- [68] M.H. Mayrand, E. Duarte-Franco, I. Rodrigues, S.D. Walter, J. Hanley, A. Ferenczy, S. Ratnam, F. Coutlee and E.L. Franco, *N. Engl. J. Med.*, **357**, 1579 (2007)
- [69] P.R. Jess, D.D. Smith, M. Mazilu, K. Dholakia, A.C. Riches and C.S. Herrington, *Int. J. Cancer*, **121**, 2723 (2007)
- [70] J.G. Kelly, K.T. Cheung, C. Martin, J.J. O’Leary, W. Prendiville, P.L. Martin-Hirsch and F.L. Martin, *Clin. Chim. Acta*, **411**, 1027 (2010)
- [71] N. Murphy, M. Ring, C.C. Heffron, B. King, A.G. Killalea, C. Hughes, C.M. Martin, E. McGuinness, O. Sheils and J.J. O’Leary, *J. Clin. Pathol.*, **58**, 525 (2005)
- [72] J.J. Max and C. Chapados, *J. Chem. Phys.*, **131**, 184505, (2009)
- [73] C.M. Martin, L. Kehoe, C.O. Spillane and J.J. O’Leary, *Mol Diagn Ther*, **11**, 277 (2007)
- [74] A. Kalof, M. Evans, L. Simmons-Arnold, B. Beatty and K. Cooper, *Am J Surg Pathol*, **29**, 674 (2005)
- [75] M. Dray, P. Russell, C. Dalrymple, N. Wallman, G. Angus, A. Leong, J. Carter and B. Cheerla, *Pathology*, **37**, 112 (2005)
- [76] T. Yu, M.J. Ferber, T.H. Cheung, T.K. Chung, Y.F. Wong and D.I. Smith, *Cancer Genet Cytogenet*, **158**, 27 (2005)
- [77] N. Ozgul, A. Cil, G. Bozdayi, A. Usubutun, D. Bulbul, S. Rota, M. Kose, A. Biri and A. Haberal, *J Obstet Gynaecol Res*, **34**, 865 (2008)
- [78] N. Murphy, M. Ring, C. Heffron, B. King, A. Killalea, C. Hughes, C. Martin, E. McGuinness, O. Sheils and J. O’Leary, *J Clin Pathol*, **58**, 525 (2005)
- [79] N. Missaoui, S. Hmissa, L. Frappart, A. Trabelsi, A. Ben Abdelkader, C. Traore, M. Mokni, M. T. Yaacoubi and S. Korbi, *Virchows Arch*, **448**, 597 (2006)
- [80] N. Murphy, M. Ring, A.G. Killalea, V. Uhlmann, M. O’Donovan, F. Mulcahy, M. Turner, E. McGuinness, M. Griffin, C. Martin, O. Sheils and J.J. O’Leary, *J Clin Pathol*, **56**, 56 (2003).
- [81] R. Klaes, T. Friedrich, D. Spitkovsky, R. Ridder, W. Rudy, U. Petry, G. Dalenbach-Hellweg, D. Schmidt and M. von Knebel Doeberitz, *Int J Cancer*, **92**, 276 (2001)
- [82] S.N. Agoff, P. Lin, J. Morihara, C. Mao, N.B. Kiviat and L.A. Koutsky, *Mod Pathol*, **16**, 665 (2003)
- [83] N. Murphy, C.C. Heffron, B. King, U.G. Ganuguapati, M. Ring, E. McGuinness, O. Sheils and J.J. O’Leary, *Virchows Arch*, **445**, 610 (2005)

- [84] N. Murphy, M. Ring, C. C. Heffron, B. King, A. G. Killalea, C. Hughes, C. M. Martin, E. McGuinness, O. Sheils and J.J. O'Leary, *J Clin Pathol*, **58**, 525 (2005)
- [85] S.S. Wang, M. Trunk, M. Schiffman, R. Herrero, M.E. Sherman, R.D. Burk, A. Hildesheim, M. C. Bratti, T. Wright, A.C. Rodriguez, S. Chen, A. Reichert, C. von Knebel Doeberitz, R. Ridder and M. von Knebel Doeberitz, *Cancer Epidemiol Biomarkers Prev*, **13**, 1355 (2004)
- [86] Z. Movasaghi, S. Rehman and I. U. Rehman, *Appl. Spectrosc. Rev.*, **42**, 493 (2007)
- [87] J. W. Black and P. Leff, *Proc R Soc Lond B Biol Sci*, **220**, 141 (1983)
- [88] G. Lizard, M.C. Chignol, Y. Chardonnet, C. Souchier, M. Bordes, D. Schmitt and J.P. Revillard, *J Immunol Methods*, **157**, 31 (1993)
- [89] G. Mehés, N. Speich, M. Bollmann and R. Bollmann, *Pathol Oncol Res*, **10**, 142 (2004)
- [90] E.M. Burd, *Clin Microbiol Rev*, **16**, 1 (2003)
- [91] S. Neviliappan, L. Fang Kan, T. Tiang Lee Walter, S. Arulkumaran and P.T. Wong, *Gynecol Oncol*, **85**, 170 (2002)
- [92] H.M. Yazdi, M.A. Bertrand and P.T. Wong, *Acta Cytol*, **40**, 664 (1996)
- [93] F.A. Supryniewicz, G.L. Disbrow, E. Krawczyk, V. Simic, K. Lantzky and R. Schlegel, *Oncogene*, **27**, 1071 (2008)
- [94] P.T. Wong, R.K. Wong, T.A. Caputo, T.A. Godwin and B. Rigas, *Proc Natl Acad Sci USA*, **88**, 10988 (1991)
- [95] H. Martens and M. Martens, *Food Qual. Prefer.*, **11**, 5 (2000)
- [96] F. Westad and H. Martens, *J. Near Infrared Spectrosc.*, **8**, 117 (2000)
- [97] J. Doorbar, *Clin Sci (Lond)*, **110**, 525 (2006)
- [98] H. Nawaz, F. Bonnier, P. Knief, O. Howe, F.M. Lyng, A.D. Meade and H.J. Byrne, *Analyst*, **135**, 3070 (2010)
- [99] H. Nawaz, F. Bonnier, A.D. Meade, F.M. Lyng and H.J. Byrne, *Analyst*, **136**, 2450 (2011)
- [100] C.A. Owen, J. Selvakumaran, I. Notingher, G. Jell, L.L. Hench and M.M. Stevens, *J. Cell. Biochem.*, **99**, 178 (2006)
- [101] F. Draux, P. Jeannesson, A. Beljebbar, A. Tfayli, N. Fourre, M. Manfait, J. Sule-Suso and G.D. Sockalingum, *Analyst*, **134**, 542 (2009)
- [102] J. Ling, S.D. Weitman, M.A. Miller, R.V. Moore and A.C. Bovik, *Appl. Opt.*, **41**, 6006 (2002)
- [103] J.Y. Ling, Q.Z. Yang, S.S. Luo, Y. Li and C.K. Zhang, *Chin. Chem. Lett.*, **16**, 71 (2005)
- [104] I. Notingher, C. Green, C. Dyer, E. Perkins, N. Hopkins, C. Lindsay and L. L. Hench, *J. R. Soc. Interface*, **1**, 79 (2004)
- [105] S. Verrier, I. Notingher, J.M. Polak and L.L. Hench, *Biopolymers*, 157 (2004)
- [106] Y. Yang, J. Sule-Suso, G. D. Sockalingum, G. Kegelaer, M. Manfait, and A. J. El Haj, *Biopolymers*, **78**, 311 (2005)

- [107] B. Kosmider, I. Wojcik, R. Osiecka, J. Bartkowiak, E. Zyner, J. Ochocki and P. Liberski, *Invest. New Drugs*, **23**, 287 (2005)
- [108] O. Vrana, V. Masek, V. Drazan and V. Brabec, *J. Struct. Biol.*, **159**, 1 (2007)
- [109] E. Ulukaya, F. Ozdikicioglu, A.Y. Oral and M. Demirci, *Toxicol. In Vitro*, **22**, 232 (2008).
- [110] N. Cordes, C. Beinke, L. Plasswilm and D. van Beuningen, *Strahlenther. Onkol.*, **180**, 157 (2004)
- [111] M.H. Baik, R.A. Friesner and S.J. Lippard, *J. Am. Chem. Soc.*, **125**, 14082 (2003)
- [112] S.U. Dunham, C. J. Turner and S.J. Lippard, *J. Am. Chem. Soc.*, **B**, 5395 (1998)
- [113] A. Gelasco and S. J. Lippard, *Biochemistry*, **37**, 9230 (1998)
- [114] P.M. Takahara, A.C. Rosenzweig, C.A. Frederick and S.J. Lippard, *Nature*, **377**, 649 (1995)
- [115] A.D. Meade, C. Clarke, H.J. Byrne and F.M. Lyng, *Radiat. Res.*, **173**, 225 (2010)
- [116] K. Varmuza and P. Filzmoser, "Introduction to Multivariate Statistical Analysis in Chemometrics", CRC Press, Taylor and Francis Group, Boca Raton, FL, 2009.
- [117] C.M. Krishna, G. Kegelaer, I. Adt, S. Rubin, V.B. Kartha, M. Manfait and G.D. Sockalingum, *Biopolymers*, **82**, 462 (2006)
- [118] J.M. Le Gal, H. Morjani and M. Manfait, *Cancer Res.*, **53**, 3681 (1993)
- [119] F. Bonnier, A.D. Meade, S. Merzha, P. Knief, K. Bhattacharya, F.M. Lyng and H.J. Byrne, *Analyst*, **135**, 1697 (2010)
- [120] Y. Fu, H. Wang, R. Shi and J.-X. Cheng, *Biophys. J.*, **92**, 3251 (2007)
- [121] G. Colella, S. Marchini, M.D. D'Incalci, R. Brown and M. Broggin, *Br. J. Cancer*, **80**, 338 (1999)
- [122] F. Bonnier, P. Knief, B. Lim, A.D. Meade, J. Dorney, K. Bhattacharya, F.M. Lyng, H.J. Byrne, *Analyst*, **135**, 3169 (2010)
- [123] I. Notingher, S. Verrier, S. Haque, J. M. Polak and L.L. Hench, *Biopolymers*, **72**, 230 (2003).
- [124] I. Notingher and L.L. Hench, *Expert Rev. Med. Devices*, **3**, 215 (2006)
- [125] G.J. Puppels, F.F. de Mul, C. Otto, J. Greve, M. Robert-Nicoud, D.J. Arndt-Jovin and T.M. Jovin, *Nature*, **347**, 301 (1990)
- [126] G.J. Puppels, H.S. Garritsen, G.M. Segers-Nolten, F.F. de Mul and J. Greve, *Biophys. J.*, **60**, 1046 (1991)
- [127] M. Miljkovic, T. Chernenko, M.J. Romeo, B. Bird, C. Matthaus and M. Diem, *Analyst*, **135**, 2002 (2010)
- [128] N. Uzunbajakava, A. Lenferink, Y. Kraan, B. Willekens, G. Vrensen, J. Greve and C. Otto, *Biopolymers*, **72**, 1 (2003)
- [129] F. Draux, C. Gobinet, J. Sule-Suso, A. Trussardi, M. Manfait, P. Jeannesson and G.D. Sockalingum, *Anal. Bioanal. Chem.*, **397**, 2727 (2010)

- [130] A.D. Meade, C. Clarke, F. Draux, G.D. Sockalingum, M. Manfait, F.M. Lyng and H. J. Byrne, *Anal. Bioanal. Chem.*, **396**, 1781 (2010)
- [131] V.M. Weaver, A.R. Howlett, B. Langton-Webster, O.W. Petersen and M.J. Bissell, *Semin. Cancer Biol.*, **6**, 175 (1995)
- [132] O.W. Petersen, L. Ronnov-Jessen, A.R. Howlett and M.J. Bissell, *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 9064 (1992)
- [133] G. D. Prestwich, *Acc. Chem. Res.*, **41**, 139 (2008)
- [134] K.S. Smalley, M. Lioni and M. Herlyn, *In Vitro Cell. Dev. Biol.: Anim.*, **42**, 242 (2006)
- [135] A. Curtis and C. Wilkinson, *Biomaterials*, **18**, 1573 (1997)
- [136] P. Giannakakou, M. Nakano, K.C. Nicolaou, A. O’Brate, J. Yu, M. V. Blagosklonny, U.F. Greber and T. Fojo, *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 10855 (2002)
- [137] A. Kotorashvili, S.J. Russo, S. Mulugeta, S. Guttentag and M.F. Beers, *J. Biol. Chem.*, **284**, 16667 (2009)
- [138] 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*, **29**, 309 (2009)
- [139] 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, *J. Biol. Chem.*, **279**, 11320 (2004)
- [140] S. Grewal, S. Ponnambalam and J.H. Walker, *J. Cell Sci.*, **116**, 2303 (2003)
- [141] J.B. MacQueen, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, (1967)
- [142] 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, *J. Biomed. Opt.*, **9**, 1187 (2004)
- [143] P. Lasch, W. Haensch, D. Naumann and M. Diem, *Biochimica et Biophysica Acta*, **1688**, 176 (2004)
- [144] E. Ly, O. Piot, A. Durlach, P. Bernard and M. Manfait, *Analyst*, **134**, 1208 (2009)
- [145] M. Hedegaard, C. Krafft, H.J. Ditzel, L.E. Johansen, S. Hassing and J. Popp, *Analytical Chemistry*, **82**, 2797 (2010)
- [146] A. Taleb, J. Diamond, J.J. McGarvey, J.R. Beattie, C. Toland and P.W. Hamilton, *J. Phys. Chem.*, **110**, 19625 (2006)
- [147] K.W. Short, S. Carpenter, J.P. Freyer and J.R. Mourant, *Biophys. J.*, **88**, 4274 (2005)
- [148] R. J. Swain, G. Jell and M.M. Stevens, *J. Cell. Biochem.*, **104**, 1427 (2008)
- [149] C.M. Krishna, G.D. Sockalingum, L. Venteo, R.A. Bhat, P. Kushtagi, M. Pluot and M. Manfait, *Biopolymers*, **79**, 269 (2005)
- [150] F. Bonnier, H.J. Byrne, *Analyst*, **137**, 322 (2012)
- [151] J. Dorney, F. Bonnier, A. Garcia, A. Casey, G. Chambers, H.J. Byrne, *Analyst*, **137**, 1111 (2012)
- [152] P. Berntsen, C.Y. Park, B. Rothen-Rutishauser, A. Tsuda, T.M. Sager, R.M. Molina, T.C. Donaghey, A.M. Alencar, D.I. Kasahara, T. Ericsson, E.J. Millet, J.



- Swenson, D.J. Tschumperlin, J.P. Butler, J.D. Brain, J.J. Fredberg, P. Gehr and E.H. Zhou, *J. R. Soc. Interface*, **7**, S331 (2010)
- [153] M. Lundqvist, J. Stigler, G. Elia, I. Lynch, T. Cedervall and K.A. Dawson, *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 14265 (2008)
- [154] M.J.D. Clift, B. Rothen-Rutishauser, D.M. Brown, R. Duffin, K. Donaldson, L. Proudfoot, K. Guy and V. Stone, *Toxicol. Appl. Pharmacol.*, **232**, 418 (2008)
- [155] J. Rejman, V. Oberle, I.S. Zuhorn and D. Hoekstra, *Biochem. J.*, **377**, 159 (2004)
- [156] B. Rothen-Rutishauser, C. Muhlfeld, F. Blank, C. Musso and P. Gehr, *Part. Fibre Toxicol.*, **4**, 9 (2007)
- [157] P.C. Naha, K. Bhattacharya, T. Tenuta, K.A. Dawson, I. Lynch, A. Garcia, F.M. Lyng and H.J. Byrne, *Toxicol. Lett.*, **198**, 134 (2010)
- [157] I. Roy, T.Y. Ohulchanskyy, D.J. Bharali, H.E. Pudavar, R.A. Mistretta, N. Kaur and P.N. Prasad, *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 279 (2005)
- [158] H. Cang, C.S. Xu, D. Montiel and H. Yang, *Opt. Lett.*, **32**, 2729 (2007)
- [159] A. Salvati, C. Aberg, T. Dos Santos, J. Varela, P. Pinto, I. Lynch and K.A. Dawson, *Nanomedicine*, **7**, 818 (2011)
- [160] H. Suh, B. Jeong, F. Liu and S.W. Kim, *Pharm. Res.*, **15**, 1495 (1998)
- [161] K. Yin Win and S.-S. Feng, *Biomaterials*, **26**, 2713 (2005)
- [162] M. Davoren, E. Herzog, A. Casey, B. Cottineau, G. Chambers, H.J. Byrne and F.M. Lyng, *Toxicol. in Vitro*, **21**, 438 (2007)
- [163] K. Shapero, F. Fenaroli, I. Lynch, D.C. Cottell, A. Salvati and K.A. Dawson, *Mol. BioSyst.*, **7**, 371 (2011)
- [164] P. Knief, C. Clarke, E. Herzog, M. Davoren, F.M. Lyng, A.D. Meade and H.J. Byrne, *Analyst*, **134**, 1182 (2009)
- [165] K. Kneipp, A.S. Haka, H. Kneipp, K. Badizadegan, N. Yoshizawa, C. Boone, K.E. Shafer-Peltier, J.T. Motz, R. R. Dasari and M.S. Feld, *Appl. Spectrosc.*, **56**, 150 (2002)
- [166] J. Kneipp, H. Kneipp, M. McLaughlin, D. Brown and K. Kneipp, *Nano Lett.*, **6**, 2225 (2006)
- [167] J. Kneipp, H. Kneipp, A. Rajadurai, R. W. Redmond and K. Kneipp, *J. Raman Spectrosc.*, **40**, 1 (2009)
- [168] J. Kneipp, H. Kneipp, W. L. Rice and K. Kneipp, *Anal. Chem.*, **77**, 2381 (2005)
- [169] J. Kneipp, H. Kneipp, B. Wittig and K. Kneipp, *J. Phys. Chem. C*, **114**, 7421 (2010)
- [170] A.E. Nel, L. Madler, D. Velegol, T. Xia, E.M. V. Hoek, P. Somasundaran, F. Klaessig, V. Castranova and M. Thompson, *Nat. Mater.*, **8**, 543 (2009)
- [171] M.Y. Chang, A.L. Shiau, Y.H. Chen, C.J. Chang, H.H. Chen and C.L. Wu, *Cancer Sci.*, **99**, 1479 (2008)
- [172] M. Hedegaard, C. Mattheus, S. Hassing, C. Krafft, M. Diem and J. Popp, *Theor. Chem. Acc.*, **130**, 1249 (2011)
- [173] M. Miljković, T. Chernenko, M. J. Romeo, B. Bird, C. Matthäus, and M. Diem, *Analyst*, **135**, 2002 (2010)

[174] M.E. Keating, F. Bonnier, H.J. Byrne, *Analyst* (2012), DOI: 10.1039/c2an36169h

[175] <http://www.dasim.eu/website.php>