Articles

2014

Processing ThinPrep Cervical Cytological Samples for Raman Spectroscopic Analysis

Franck Bonnier  
*Dublin Institute of Technology*

Damien Traynor  
*Dublin Institute of Technology*

Padraig Kearney  
*The Coombe Women and Infants Hospital*

See next page for additional authors

Follow this and additional works at: https://arrow.tudublin.ie/despart

Part of the Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons

Recommended Citation


This Article is brought to you for free and open access by ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, gerard.connolly@tudublin.ie, vera.kilshaw@tudublin.ie.
Authors
Franck Bonnier, Damien Traynor, Padraig Kearney, Colin Clarke, Peter Knief, Cara Martin, John J. O’Leary, Hugh J. Byrne, and Fiona M. Lyng

This article is available at ARROW@TU Dublin: https://arrow.tudublin.ie/despart/7
Processing ThinPrep Cervical Cytological Samples for Raman Spectroscopic Analysis

Franck Bonnier\textsuperscript{1*}, Damien Traynor\textsuperscript{1}, Padraig Kearney\textsuperscript{3}, Colin Clarke\textsuperscript{2}, Peter Knief\textsuperscript{4}, Cara Martin\textsuperscript{3}, John J. O’Leary\textsuperscript{3}, Hugh J. Byrne\textsuperscript{4}, Fiona Lyng\textsuperscript{1}

1. DIT Centre for Radiation and Environmental Science, Focas Research Institute, Dublin Institute of Technology (DIT), Camden Row, Dublin 8, Ireland
2. National Institute for Cellular Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland
3. Department of Pathology, Coombe Women & Infants University Hospital, Dublin, Ireland
4. Focas Research Institute, Dublin Institute of Technology (DIT), Camden Row, Dublin 8, Ireland

*Corresponding Author: Franck Bonnier,
Focas Research Institute, Dublin Institute of Technology,
Kevin Street, Dublin 8, Ireland.
E mail address: fbonnier@dit.ie
Ph: +353 1 4027917
Fax: +353 1 4027904
Abstract

Raman microspectroscopy has been proven to be a promising technique for diagnosis and early detection of pathologies. The data collected delivers a chemical fingerprint allowing the identification of specific biomarkers indicating the presence of abnormalities. Label free, fast and cost effective, Raman spectroscopy has already been proposed as the new generation of diagnostic tool with a strong potential but has not emerged in the medical field as yet. Notably, it is crucial to improve and adapt the protocols used to reach suitable reproducibility for screening large cohorts of patients. In this study, it is demonstrated that the variability existing in the data sets collected can be limiting. Notably, when working on cervical ThinPrep samples, the presence of blood residue can be detected by Raman spectroscopy swamping the cellular signal. However, combining a washing of the slides using H₂O₂ and alcohols (70% ethanol and 100% Industrial Methylated Spirits), the blood features are removed from the data without altering either the cell morphology or the spectral features. Ultimately, this work demonstrates the improved potential of Raman spectroscopy for ThinPrep analysis based on improved protocols for sample preparation. Therefore, the screening of cervical cells for the detection of abnormalities and identification of patients with Cervical Intraepithelial Neoplasia (CIN) is achievable.

Key words: Raman microspectroscopy, ThinPrep sample, blood, cervical cancer, Cervical Intraepithelial Neoplasia, diagnostic
1. Introduction

Cervical cancer kills approximately 300,000 women each year worldwide, 80% of fatalities occurring in developing countries, and, with about 50,000 new cases diagnosed every year\textsuperscript{1-3}, it is the second most common malignancy among women worldwide. Epithelial abnormalities have the potential to progress into an invasive neoplasm. They have traditionally been categorised either as dysplasia or as carcinoma \textit{in situ}. Dysplastic changes within the epithelium are graded as being of a mild, moderate, or severe degree. Those cervical epithelial abnormalities associated with an increased risk of invasive carcinoma have been classified into a single diagnostic category of cervical intraepithelial neoplasia (CIN). Three grades of abnormality are recognised: CIN I which corresponds to mild dysplasia; CIN II which is equivalent to moderate dysplasia; and CIN III which encompasses both severe dysplasia and carcinoma \textit{in situ}. A screening test is employed to detect these premalignant abnormalities before they can develop into an invasive carcinoma. The main method for cervical cancer screening is the Papanicolaou test (Pap test). The Pap test requires cells to be scraped from the cervix and fixed onto a glass slide, whereupon they are stained and viewed under a microscope. Any cellular abnormalities present are identified by the cell morphology and staining characteristics. Depending on the technology available, most examinations are conducted by eye and the classification of the appropriate grade of dysplasia (CIN 1-3) is subjective. This results in a high rate of patient recall or gynaecological referral.

Vibrational spectroscopic techniques are potential tools for the non-invasive, label free investigation of biological samples at a molecular level\textsuperscript{4-6}. Over the course of the last few years, the range of applications of Infrared and Raman spectroscopy has considerably increased\textsuperscript{7-10}, elucidating different mechanisms involved in pre-cancer
and cancer as well as probing the efficiency of new drugs and the toxicity of nanoparticles at the cellular level. Vibrational spectroscopy is a fast developing discipline showing a strong potential in the field of cervical cancer screening.

Since vibrational spectroscopy was first proposed as a tool for mass screening of cervical cancer in the 1990s by Wong et al., there has been significant progress. Numerous studies have shown the potential of vibrational spectroscopic techniques in the detection of cervical cancer and pre-cancerous lesions. Moreover, spectroscopic data handling and multivariate analysis procedures have developed significantly over the last two decades.

Equipped with high magnification objectives such as x60 or x100, the lateral resolution offered by Raman microspectroscopy (typically better than 1 μm) makes it a perfectly suitable tool for single cell analysis. Therefore, it could potentially be used to complement existing cytology screening techniques, due the high degree of objectivity of the assessment based on molecular composition making spectroscopic techniques extremely sensitive and specific tools, leading to the capability of detecting early abnormal changes that are not morphologically apparent and thus undetectable using standard methods.

Nevertheless, many improvements have still to be realized to increase the specificity of the information which can be gleaned. The instrumental response, background and substrate contribution can add to intrinsic sample variability. Additionally, when studying human samples, a source of variability related to each individual has to be tolerated while interpreting the data. However, in most studies, additional sources of variability, mostly instrumentation related, are added to the problem, decreasing the specificity and reproducibility of the analysis. In the present study, Raman spectroscopy has been used for the analysis of 50 normal cervical cell samples.
prepared using the ThinPrep method, routinely used for screening of cervical cancer. The aim of the study was to visualize, understand and overcome the limitations encountered during the collection of the spectral datasets, the main challenge being to use protocols compatible with clinical practices. Therefore, the use of cost efficient glass slides has to be implemented in the approach, instead of the routinely used spectroscopically invisible CaF$_2$ substrates. In the present work, the presence of blood residue has been highlighted and alternative protocols for sample preparation have been described. Ultimately, the observations are encouraging in assessing the feasibility of using Raman spectroscopy for the analysis of ThinPrep samples and demonstrate that, although the technique is often presented as label free and not requiring special sample preparation, particular attention has to be given to the washing of the ThinPrep samples prior to data collection. Ultimately, it is expected that these protocols will be beneficial in the application of Raman spectroscopy for the screening of abnormal cells by helping to collect more specific and accurate spectral datasets fromThinPrep samples which can be used for diagnosis.

2. Materials and Methods

2.1 ThinPrep

Cervical liquid based cytology (LBC) samples were provided by the Coombe Women and Infants University Hospital (CWIUH), Dublin, Ireland for spectroscopic analysis, after routine cytological screening had been carried out by the medical staff (cytologist and senior cytologist) using Pap staining. In the case where suspicious cells are identified, the slides are referred to a pathologist for clinical diagnosis. The remainder of the (unstained) cell suspensions were then provided to the Focas Research Institute, DIT, accompanied by the final clinical report prepared by the
cytologist and/or pathologist.. Finally, a set of 63 samples (50 identified as cytology negative and 13 prepared from patients with confirmed presence of CIN3 cells) was prepared using the ThinPrep technique. All samples were prepared using a ThinPrep 2000 processor (Hologic Inc., Marlborough, MA 01752).

The ThinPrep system for Pap specimens is FDA approved as a replacement for the conventional method for detecting atypical cells, cervical cancer and its precursor lesions as well as other cytological categories. The ThinPrep process begins with the patient’s gynecological sample being collected by the clinician using either a broom-like device or a brush/plastic spatula combination. The device is then rinsed in the specimen vial containing PreservCyt transport medium (ThinPrep Pap Test; Cytyc Corporation, Boxborough, Mass.). The ThinPrep sample vial is then capped, labeled, and sent to a laboratory equipped with a ThinPrep processor.

The ThinPrep processor homogenizes the sample by spinning either the filter (T2000) or the vial (T3000), creating shear forces in the fluid that are strong enough to disaggregate randomly joined material, break up blood, mucus and non-diagnostic debris while keeping true cell clusters intact. The cells are then collected onto the membrane of the TransCyt filter and transferred onto a glass slide to create a monolayer deposit of cells, ~20 mm in diameter. The slide is then ejected automatically into a fixative bath of 95% ethanol.

2.2 Blood scale evaluation

In cytology laboratories, cell suspensions prepared for ThinPrep or smear samples are annotated according to a blood scale, whereby 0 indicates a clear solution and 3 indicates a bloody sample. Samples presenting a grade 2 or 3 on the blood scale are either rejected as being unsuitable or treated using Cytolyt or various solutions to
wash the cells before preparing the samples. In this study, only samples classified as blood scale 0 or 1 have been employed.

2.3 Pure chemical preparation

Human hemoglobin, hemin and proto-porphyrin (lyophilized powder) were purchased from Sigma Aldrich (Ireland). Each of the chemical compounds was deposited on a CaF₂ window in the lyophilized form before recording.

An aliquot of blood (5 mL) was mixed with PreservCyt solution in a ThinPrep vial containing no cervical cells, and 50 µL were transferred onto a CaF₂ substrate and air dried before recording.

2.4 Hydrogen Peroxide (H₂O₂) treatment of the Thinprep slides

In section 3.3, the presence of blood residues on the ThinPrep slides was found to be the main limiting factor to overcome in order to collect data with minimal variability. Therefore, for subsequent Raman measurements, the slides were treated with a solution of 30% hydrogen peroxide (H₂O₂). Following heating to 37°C in a water bath, the cells were exposed to H₂O₂ on the ThinPrep slide for 5 mins. The slide was then placed into a solution of 70% ethanol for 5 minutes, and was then dipped multiple times in 100% Industrial Methylated Spirits to remove any remaining cellular debris and H₂O₂. The slide was air dried.

2.5 Raman microspectroscopy

Raman studies were performed using a HORIBA Jobin Yvon XploRA™ system (Villeneuve d’Ascq, France), which incorporates an Olympus microscope BX41 equipped with a x100 objective (MPlanN, Olympus, NA = 0.9). A 532nm diode laser source was used throughout this work. In order to avoid any photo damage to the
sample, the power of the laser was set at 50%, resulting in ~8 mW at the objective. The confocal hole coupled to a slit of aperture 100µm, was set at 100µm, for all measurements, the specified setting for confocal operation. The system was pre-calibrated to the 520.7 cm\(^{-1}\) spectral line of Silicon. In the following experiments, the 1200 lines/mm grating was used. The backscattered light was measured using an air-cooled CCD detector (Andor, 1024 x 256 pixels). The spectrometer was controlled by Labspec V5.0 software.

2.6 Data collection

For each cell, 3 spectra were recorded from the cell nucleus, each of them corresponding to the average of 3 accumulations of 10s. After treatment with H\(_2\)O\(_2\), (Section 3.3) a 30s pre-exposure of the cells to the laser has also been applied in order to reduce any residual background present in the data collected \cite{43,44}. Although the full range of 100cm\(^{-1}\) to >4000cm\(^{-1}\) is accessible by the instrument, for expediency, only the spectral range 400-1800 cm\(^{-1}\), was defined by the acquisition software.

For each sample from patients identified as negative (absence of cellular abnormalities), 10 cells were selected randomly for recording. Also, for each slide, a spectrum of the substrate (glass slide used for cell deposition) was recorded for correction of the data. Finally, the 10 spectra recorded from each slide were averaged to obtain only one representative spectrum per patient. Due to difficulties in obtaining samples from patients diagnosed with severe dysplasia (CIN3), a total of 50 abnormal cells have been recorded from 13 different patients and these were treated individually in order to match the number of spectra for statistical analysis. The identification of CIN3 cells being a difficult task based on unstained samples, the confirmation of their abnormality has been made retrospectively. Thus, Raman spectra have been recorded
from suspicious cells identified in the CIN3 ThinPrep samples and their coordinates on the slides were recorded. The pathologist performed further examination of the identified cells after Pap staining and confirmed where the CIN3 cells were, based on coordinates. This approach has been found the most accurate to ensure only spectra recorded from CIN 3 cells are retained in the dataset (figure 7A).

Concerning the recording of the spectra from pure chemicals (mentioned in section 2.3), the laser power was set to 1% in order to avoid photo-damage of the samples. Moreover, the acquisition time was increased to 3 x 100s, providing a final signal to noise ratio comparable with the data collected from the cervical cells.

2.7 Bright field imaging

Bright field images were collected using an Olympus BX51 microscope equipped with a 40x objective under white light. In order to improve the visualization of the different cellular compartments such as nucleus and cytoplasm, the cells were stained with Haematoxylin/Eosin.

2.8 Data preprocessing and Analysis

All the data recorded throughout this work were corrected and analyzed using Matlab software (Mathworks). The scripts have been developed and adapted for the uploading of the spectra and their pre-processing, including smoothing (Savitzky-Golay, k=5; w=13), baseline correction (rubber band) and vector normalization. The spectral features of glass are usually broad and can overlap significantly with the region of interest of the cellular spectra and can therefore be another source of variability. However, using the 532 nm wavelength as a source and a Raman microscope equipped with a 100x objective, the confocality of the Raman can be
greatly improved and the contribution from the substrate significantly decreased. Furthermore, any residual contribution from the glass has been removed by subtraction using a reference spectrum of the slides used to prepare the samples.

Principal Component Analysis (PCA) is an unsupervised method allowing evaluation of the variability existing in the data sets. For more details on the applications of PCA to the analysis of biological samples using Raman spectroscopy, recent work by Bonnier et al. can be found in the literature \(^{16,24,30}\).

3. Results and discussion

3.1 Raman spectroscopy of ThinPrep samples: sample variability

Based on the numerous documents available in the literature \(^{47-49}\), it is easy to determine a typical spectrum recorded from a single cell. Variations due to, for example, different cell lines or cells in different phases of the cell cycle \(^{50,51}\) are usually minimal and can rarely be discerned by eye, necessitating the use of advanced multivariate analysis methods. However, as highlighted in the data collected from the nuclei of cells of 50 patient samples with negative cytology in figure 1A, large differences in the spectral patterns are observed, and although the variations are continuous, two sub-groups can be qualitatively defined based on (i) the background intensity, particularly influencing the intensity of the spectra in the higher wavenumber region, (ii) the spectral features exhibited across the spectral range 400 – 1800 cm\(^{-1}\). The first subgroup is dominated by features of standard cellular spectra \(^{25,30}\) (referred to as pattern 1 (n = 28)), whereas the red spectra are considered as atypical (pattern 2, n = 22). Table 1 shows a comparison of the different peaks identified in both pattern 1 and pattern 2 spectra.
A more detailed view of the pattern 1 samples can be found in figure 1B, which highlights the presence of a subgroup presenting similar features to pattern 1, but which exhibit a slightly higher background (plotted in green). Baseline correction followed by vector normalization eliminates this variability in pattern 1 spectra (figure 1C), indicating the difference observed in the raw spectra is related to the morphology of the cells rather than biochemical content. Within the two subgroups, whereas pattern 1 spectra are fairly consistent regarding the peak intensities, pattern 2 spectra show a significantly high degree of variability that can be clearly seen for the peaks located at 1642 cm$^{-1}$. Close inspection indicates the presence of an isosbestic-like point in the region of 1646 cm$^{-1}$, also visible at 1481 cm$^{-1}$ (Figure 1D). This indicates that a single spectral component is responsible for this specific Raman signature and it is present in different concentrations. Notably, when a pattern 2 signature is found on a sample, the entire cellular population will display the same profile, indicating that the effect is not related to the cell type recorded but is more likely cell suspension (patient) dependent.

3.2 Understanding the origin of the pattern 2

Observation of the cells using bright field optical microscopy, as per section 2.7, does not give any indication of abnormal morphology of the cells. Figure 2IA and 2IC present a comparison of cells which exhibit respectively pattern 1 and pattern 2 Raman signatures. The cells look perfectly normal in size and the nuclei have a perfectly circular shape and there are no perceptible morphological differences between samples classified as pattern 1 and pattern 2. As collection of the spectra was performed using a 100x objective, giving a spot size of less than a 1 µm at the focal
point, the presence of small molecules at the surface of the cells not discernable by eye could easily contribute to the data collected. Therefore, the presence of biomolecules in cell suspension could contaminate the samples generating abnormal Raman signatures.

Figure 2II presents the mean spectra obtained from the pattern 1 samples (blue) and the pattern 2 samples (red). The spectral windows exhibiting important variations in the Raman signatures have been highlighted by the black squares. A full list of the peaks and assignment can be found in table 1. In order to understand the origin of the pattern 2 signature, the mean spectrum of a typical pattern 2 sample was compared to different spectra recorded from pure compounds (figure 2II). The different spectral regions of interest have been highlighted in grey and the close similarity existing between the spectrum of the blood and the pattern 2 mean spectrum is clearly visible. The comparison with different molecules present in blood such as hemoglobin (figure 2IIIIC), hemin (figure 2IIID) or proto-porphyrin (figure 2IIIE) validate this observation and the closest match is hemoglobin. Notably, hemoglobin is resonant at a wavelength of 532nm, and therefore the Raman signal between 1500-1650 cm$^{-1}$ is resonantly enhanced, such that, even after the blood scale evaluation and treatment described in Section 2.2, residual amounts can dominate the cellular signature. Thus, it may be concluded that, during the different steps of the sample preparation, the lysed blood present in the ThinPrep suspension interacts with the cervical cells and can remain on the surface of the cells after drying. Although not visible by eye, the presence of hemoglobin can be detected by Raman spectroscopy. Therefore, the example presented in figure 2IIA is mainly dominated by the hemoglobin signature and only a few specific features from the cells can be identified. Even the strong signature amide I band at ~1650 cm$^{-1}$ is almost completely obscured and appears only
as a shoulder. The contamination of the samples with blood residue was found to be present in about 50 percent of the samples screened, and because almost half of the spectral range is affected, the specificity of the information contained in the spectra is greatly reduced. The variability generated within the normal samples (see figure 1C) is too high for the data to be used for the purpose of discrimination between normal and abnormal samples.

3.3 Dealing with the pattern 2: reversion to pattern 1

The origin of pattern 2 type spectra having being identified, the challenge is to remove the variability existing in the spectra to generate an acceptable control dataset. In order to solve the contamination of the samples with blood residues, three different approaches can be considered; (i) modify the protocol before the preparation of the samples while the cells are still in suspension, (ii) treat the slides and try to wash the cells after deposition and drying or (iii) manipulate the data and digitally remove the signal of the blood.

Signal decomposition and correction with approaches such as Independent Component Analysis (ICA) coupled with Non-negatively Constrained Least Squares Analysis (NCLSA) has been documented and proven promising for the application to Raman data. The first step consists of identifying the independent components of pattern 2 by analyzing the co-variance existing in the data set. In the second step, the unwanted hemoglobin component is subtracted from the data set using NCLSA. This method has been successfully used to remove the contribution of wax in embedded skin sample for Raman and Infrared analysis. However, the samples exhibiting a pattern 2 signature have been also found to be more susceptible to photodegradation,
causing some difficulties to collect the data from ThinPrep slides heavily contaminated with blood residues which could represent up to 5% of the samples screened. Thus, although the digital removal of the hemoglobin would be a valid approach, the aim of this work was to develop an alternative approach allowing the collection of Raman data from all the different samples, regardless of the quantity of hemoglobin initially present on the cell surface. It has been observed in this study that some samples with a grade 1 or 0 on the blood scale that would routinely be used for cytology present a pattern 2 signature. As described in Section 2.2, during ThinPrep sample screening in the clinical environment, Cytolyt would be used to treat bloody samples identified as level 3 on the blood scale. The observations of the previous section indicate however that, for spectroscopic screening, all samples would have to be washed with Cytolyt, which is unrealistic in terms of time and expense. It is important to bear in mind that, in order to successfully integrate Raman microspectroscopy in the clinical environment, the sample preparation protocols should be compatible with existing methods. The present work has been conducted in collaboration with the Coombe Women and Infants University Hospital (Dublin, Ireland), specialising and nationally recognized in cervical screening. Due to the constant increase in patient numbers and the intensification of the National Cervical Screening Programme, CervicalCheck Ireland (http://www.cervicalcheck.ie/), in 2013 approximately 300 to 400 ThinPrep samples were routinely prepared per week bringing the total number of ThinPrep slides prepared for that year to over 15,000 in this hospital alone. For this reason, improved sample preparation and handling protocols, before the collection of the datasets, were sought. The only approach that can be realistically applicable in a medical environment and for a large number of samples is the treatment of the slides after cell deposition and
drying. To be present at this stage, the blood residues must have interacted with the cellular membrane. Thus, by destabilizing this interaction the residues may be washed away. Different solutions such as Cytolyt, eposti solution and acetic acid have been tested, but none delivered a reproducible result and features relating to the blood residue could still be seen in the spectra recorded (data not shown). Previous work published by Romeo et al. 56 describes the feasibility to use red cell lysis buffer (RCLB) to reduce the presence of erythrocytes in cervical ThinPrep samples for IR spectroscopy. However, this solution has been found to be inefficient after the cells have been deposited and dried on glass slides.

The action of \( \text{H}_2\text{O}_2 \) on the structure and conformation of hemoglobin is well documented 57, 58. Hemoglobin is particularly sensitive to oxidation, which irreversibly changes its structure. A solution of 30% \( \text{H}_2\text{O}_2 \) was therefore employed to treat the samples for 5 mins, after which the samples were washed in 70% ethanol and 100% Industrial Methylated Spirits (see material and methods for full description of the protocol). The first concern was to preserve the cell morphology after such treatment. Figures 2IB and 2ID present cells stained using Haematoxylin/Eosin after treatment with \( \text{H}_2\text{O}_2 \). The cellular shape and size is not affected by the solution processing, probably due to the fact cells are fixed and dried on the glass slides. The use of staining is still achievable and the membrane and nucleus of the cells remain clearly identifiable. However, although the appearance of the cells appears unchanged, the molecular composition could be affected by such approach. Figure 3 presents the spectra collected from cells after \( \text{H}_2\text{O}_2 \) treatment. For consistency, all the slides have been submitted to the same protocols regardless of their initial spectral pattern. The red spectra are those from cells originally presenting a pattern 2, exhibiting strong contributions of blood. These features are completely gone after
washing and the only peaks visible are related to cellular components. Thus, the transformation from pattern 2 to pattern 1 is achievable using a short exposure to H₂O₂ and subsequent drying of the samples. However, comparison with the blue spectra, corresponding to the samples originally exhibiting a pattern 1, highlights the presence of a stronger background still remaining in the spectra recorded from samples originally presenting a pattern 2. Although the two datasets seem to be chemically comparable, they remain spectroscopically different. The problem of background in the data collected using Raman spectroscopy has been well documented in the literature 59-61. Although the background observed in Raman spectroscopy is commonly attributed to fluorescence 60, 62, and its reduction upon illumination described as a photobleaching 44, recent work on proteins and human skin sections has demonstrated that similar effects can be observed on samples devoid of fluorophores or dyes 40. Therefore, the evolution of the background may also be attributed to photothermal “annealing” like effects rather than photochemical bleaching, which, although not visible microscopically, result in local changes to sample morphology and a reduction in scattering 43. Similar effects can be seen in the present study and figure 3II presents an example of the background decay over time. A set of 6 spectra has been recorded on the same spot, without any movement of the sample, using an accumulation time of 10s. Spectrum A represents the signal originally recorded. The background creates a distortion of the baseline that can be seen with the red line. The last spectrum (B) corresponds to the 6th accumulation of 10s acquired from this spot of the sample. The background is completely gone and the spectrum displays a profile similar to those recorded for cells grown in vitro 25, 30, 47. The grey spectra represent the intermediate states for each accumulation of 10s. After 30s exposure to the laser, the background becomes relatively constant and the
decrease is minimal. Therefore, pre-exposure to the laser of 30s has been employed for any further measurement carried out on the ThinPrep samples.

Figure 3III presents the mean spectra obtained from the 50 samples after the H$_2$O$_2$ treatment and the 30s pre-exposure to the laser. The variability existing between the data sets corresponding to the samples originally displaying pattern 1 (blue) and pattern 2 (red) has been greatly reduced. Although the intensity is slightly different, the background remains similar. Thus, all the spectra have a profile corresponding to those commonly observed on cells grown in vitro.

3.4 Evaluation of the sample variability after pre-processing

H$_2$O$_2$ is a highly reactive species and has an oxidizing effect on proteins. Thus, it is necessary to ensure the cells were not adversely affected by the treatment. Although the cell morphology is intact after treatment and no visual alteration of the composition can be seen (figure 2I) the molecular organization could be slightly different and have an impact on the spectra recorded. As an internal control, the samples presenting a pattern 1 signature before the H$_2$O$_2$ washing have been used. Figure 4IA presents the mean spectrum of the pattern 1 samples before treatment compared to the mean spectrum of the pattern 1 samples after treatment (figure 4IB). The two spectra display strong similarities and only two small bands located at 536 cm$^{-1}$ and 904 cm$^{-1}$ have their intensities changed. Apart from those peaks, no noticeable variability is present in the spectra and an overall mean and standard deviation of 0.0207 +/- 0.012 is observed. Indeed, the overlap between the two spectra is not perfect, but as the cells are picked randomly, some variability has to be expected. The analysis of the data sets before and after washing with H$_2$O$_2$ using PCA gives additional understanding of the origin of any spectral variability. Figure 5.I
displays the PCA scatter plot based on the pattern 2 samples before and after H$_2$O$_2$ treatment, highlighting a clear discrimination between the two datasets which, according to the loading of PC1, plotted in figure 5.III (red), is due to the removal of the hemoglobin from the samples. However, after performing a PCA on the datasets described as pattern 1, the scatter plot also displays a strong discrimination between the spectra recorded before and after washing using H$_2$O$_2$. For comparison, the loading corresponding to PC1 has been plotted in figure 5.III (blue). The similarities existing between the loadings calculated from the samples classified as pattern 1 and 2 are notable. This demonstrates that the classification of individual sample spectral patterns was not exact but, although some samples did not exhibit strong hemoglobin features and were classified as pattern 1, the spectra were influenced by underlying contributions which mixed with the complex signal from the cellular features. However, this can be detected using a statistical analysis approach such as PCA. Thus, the variation seen in the pattern 1 mean spectra plotted in figure 4.I.A and 4.I.B can be more likely assigned to the washing of small hemoglobin residues present on the cell surface.

More importantly, the variability existing between the samples originally exhibiting a pattern 1 signature and a pattern 2 signature has to be investigated after treatment with H$_2$O$_2$. Figure 4II presents the mean spectrum of the samples with a pattern 1 signature before washing (red), compared to the mean spectrum of the samples with a pattern 2 signature after washing (blue). The spectra have been baseline corrected and vector normalized for comparison. The spectral features are perfectly comparable and no alteration of the molecular composition of the cells can be seen. Figure 4III is an attempt to represent the overall quality of the data set after treatment with H$_2$O$_2$. The mean spectrum after baseline correction and vector normalisation has been plotted in
The mean and standard deviation of the full dataset over the full spectral range before and after treatment using H$_2$O$_2$ was found to be respectively 0.0209 +/- 0.0029 and 0.0207 +/- 0.0013. However, the blood residues do not affect the whole spectral range and only few features are strongly altered. Therefore such values do not accurately represent the beneficial effect of such treatment of the slides. A better appreciation of the reduced data variability can be achieved by selecting hemoglobin specific spectral ranges as discussed in section 3.2. For instance, the spectral ranges 1270-1435 cm$^{-1}$ and 1495 – 1651 cm$^{-1}$ are of particular interest as shown in figure 1. The mean and standard deviations for those regions have been found to be 0.043 +/- 0.027 and 0.027 +/- 0.010 before treatment, clearly highlighting the huge amount of variability representing up to 50% of the mean value. However, after treatment with H$_2$O$_2$, the standard deviation for the same spectral windows has been substantially reduced to 0.0426 +/- 0.0012 and 0.0173 +/- 0.0011. Thus, starting with two distinct signatures, obviously different, the adaptation of the sample preparation steps resulted in a significantly reduced within group variability which will give the possibility to perform accurate and more specific analysis of the ThinPrep samples using Raman spectroscopy. However, the cellular variability remains one of the most limiting factors. Thus, when collecting spectra from different cells, differences in the cell morphology, size, but also metabolism or phenotype at the stage of sample preparation can have a significant impact on the features present in the data. The use of a 100x objective gives access to subcellular analysis of single cells but using a wavelength such as 532 nm, the spot being reduced to less than a micron, the sub-nuclear analysis of the cells is also possible$^{25, 30}$. While the identification of the
nucleoli can be achieved when working on cancerous cell lines grown on substrates \(^9, \) \(^16 \), in the case of cervical cells prepared as ThinPrep slides, the nucleus does not present any evidence of sub-nuclear organization, although not being visibly discernible does not necessarily mean the molecular composition of the nucleus is homogeneous. Therefore, although 3 spectra have been recorded for each cell in order to reduce the variability due to random selection of the collection spots; most of the nucleus is not being recorded due to the high magnification offered by the use of a 100x objective. Thus, the remaining standard deviation observed in the data sets is likely to be cell to cell variability due the presence of sub-nuclear structures inducing subtle modifications of the spectral features present in the data sets. This remaining variability can only be removed by adapting the approach used for the spectral data collection and improvement of the nuclear area covered during the spectra collection.

3.5 Discrimination between normal and cancerous cells

The washing of the ThinPrep samples with \(\text{H}_2\text{O}_2\) seems to be necessary for the recording of Raman spectra useful for diagnostic purposes. Although the variability in the data collected from normal samples is clearly reduced, it is also essential to demonstrate that such a process doesn’t take away any specific features in the spectra that can be used to discriminate between normal and abnormal samples. For this reason, the recording of samples identified as CIN3 has been performed. In order to avoid any interference in the analysis, only the range 1150 – 1800 cm\(^{-1}\) has been used for the PCA (figure 6B). As highlighted in figure 6C, a clear separation between the normal and CIN3 spectra can be achieved using this spectral window (only one normal sample mis-classified). The loading in figure 6D gives a representation of the different wavenumbers involved in this discrimination. It is interesting to point out the
presence of features at 1381, 1426 or 1581 cm\(^{-1}\) related to the DNA/RNA content of the cells (Figure 6D and also see table 1). This clearly demonstrates that the washing of the slides with H\(_2\)O\(_2\) does not alter the composition of the cells and also does not interfere with the possibility to identify the abnormal cells on the ThinPrep samples. These observations are really encouraging for the possibility to use Raman spectroscopy for clinical application, and for instance cervical cytology.

**Conclusion:**

The variability present in patient samples can always be an issue for the potential discrimination between normal samples and pathological or abnormal samples. Although it is commonly accepted that more specific analytical methods are needed to extract the relevant information, the sample preparation protocols can have a significant impact on variability. Concerning the study of ThinPrep samples using Raman spectroscopy, the presence of blood can be the limiting parameter as its spectral signatures overlap the cellular features therefore making the use of the raw data impossible for diagnostic purposes. However, this study clearly demonstrates that rather than trying to develop advanced algorithms of multivariate analysis methods to correct the data, a simple treatment of the samples can wash off the blood residue present on the cell membrane, thus the collection of highly reproducible data has been achieved. The protocols developed throughout this work should greatly improve the relevancy of the results collected from ThinPrep samples and pave the way towards, for example, analysis of cytologically normal cells in abnormal samples, and will contribute to realisation of the application of vibrational spectroscopy as a diagnostic tool for screening of large sample sets.
Acknowledgement

This work was funded by Enterprise Ireland co-funded by the European Regional Development Fund (ERDF) and Ireland’s EU Structural Funds Programme 2007-2013. FB was also supported by Science Foundation Ireland, 11/PI/08.
Bibliography

56. Romeo MJ, et al., Biopolymers (Biospectroscopy), 2003, 72, 69-76.
59. Mazet V, et al., Chemometrics and Intelligent Laboratory Systems, 2005, 76, 121.
Figures:

Figure 1: A: Mean spectrum calculated for the raw spectra recorded from 50 normal samples. The blue spectra exhibit regular cellular features (Pattern 1). The red spectra represent the samples with a signature corresponding to abnormal features (Pattern 2); B: Mean spectra of the samples corresponding to the pattern 1. The spectra highlighted in green have typical cellular features but can be discriminated due to a slight offset in their baseline; C: mean spectra of the 50 samples after pre-processing and D: isobestic point in the data set

<table>
<thead>
<tr>
<th>Pattern 1</th>
<th>Assignments</th>
<th>Pattern 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>497</td>
<td>C–C twisting mode of Phe (proteins)</td>
<td>499</td>
</tr>
<tr>
<td>625</td>
<td>C–C twisting mode of Tyr and Phe</td>
<td>625</td>
</tr>
<tr>
<td>646</td>
<td>C–N stretching in A and lipids</td>
<td>647</td>
</tr>
<tr>
<td>731</td>
<td>Symmetric breathing of Trp (protein)</td>
<td>732</td>
</tr>
<tr>
<td>761</td>
<td>U, T, C (ring breathing modes in the DNA/RNA)</td>
<td>752</td>
</tr>
<tr>
<td>785</td>
<td>PO2 stretching in DNA, Tyr</td>
<td>786</td>
</tr>
<tr>
<td>832</td>
<td>Ring breathing in Tyr and Pro (proteins)</td>
<td>831</td>
</tr>
<tr>
<td>857</td>
<td>C–C stretching mode of Pro and Val</td>
<td>857</td>
</tr>
<tr>
<td>941</td>
<td>C–C and C–N stretch PO3 2– stretch (DNA)</td>
<td>961</td>
</tr>
</tbody>
</table>
1007  C–C aromatic ring stretching in Phe
              C–H bending mode in phenylalanine, C–N stretching in
              proteins
1035
1095  Symmetric PO2 stretching of the DNA backbone; lipids
              C–N stretching in proteins; C–O stretching in
              carbohydrates
1129
1160  C–C and C–N stretching of proteins
1177  C–H in plane bending mode of Tyr and Phe; C, G
1212  C–C6H5 stretching mode in Trp, Phe
1246 (1220-
1280)  Amide III; A, C, T ring breathing modes of the DNA/RNA

Pattern 2 feature

1316
1340  G (DNA/RNA), CH deformation in proteins and
              carbohydrates

Pattern 2 feature

1366
1402
1424  CH3 asymmetric stretch (lipids, aromatics)
1453  CH (CH2) bending mode in proteins and lipids
1581  A, G (DNA/RNA); C=C bending mode of Phe

Pattern 2 feature

1588
1610  C=C Phe, Tyr
1620  C=C Tyr and Trp

Pattern 2 feature

1642
1673  Amide I
Figure 2: I) Bright field images of Thinprep samples as seen using an Olympus BX51 microscope equipped with a 40x objective. The cells have been stained with Haematoxylin and Eosin. Cells from a Thinprep sample exhibiting pattern 1 before (A) and after (B) treatment compared to cells exhibiting pattern 2 before (C) and after (D) treatment

II) Comparison of the mean spectrum of the pattern 1 sample (blue) and the pattern 2 samples (red)

III) Comparison of the mean spectrum of the pattern 2 (A) with the spectra blood (B), haemoglobin (C), hemin (D) and proto-porphyrin (E).
Figure 3: **I:** Mean spectra from the 50 samples after treatment using H$_2$O$_2$. The blue spectra were already presenting a pattern 1 before treatment. The red spectra were presenting a pattern 2 before treatment. **II:** background decay for extended exposure to the laser. A: spectrum obtained after 10s accumulation. B: spectrum recorded after the 6$^{th}$ accumulation of 10s. **III:** Mean spectra of the 50 samples after background removal using extended exposure to the laser.
Figure 4: I) A: Mean spectrum of the pattern 1 sample before (A) and after (B) washing with H$_2$O$_2$. II) A: Mean spectrum of the pattern 1 (A) and pattern 2 (B) sample after washing with H$_2$O$_2$. III) Mean spectrum of the whole data set after treatment with H$_2$O$_2$ (black curve) and standard deviation calculated on for each spectra compared to the mean spectrum (blue shade).
Figure 5: I) Scatter plot of the PCA analysis performed on the data recorded from Pattern 2 samples before (blue) and after (red) washing using H2O2. II) Scatter plot of the PCA analysis performed on the data recorded from Pattern 1 samples before (blue) and after (red) washing using H2O2. III) Loadings corresponding to the first PC for the pattern 1 and 2 data sets.
Figure 6: A) Bright field micrograph obtained from ThinPrep slides before and after Pap staining. The abnormal cells (CIN3) are indicated by the arrow; B) Comparison of the mean spectra calculated from the normal cells after H₂O₂ washing (blue) compared to the average spectrum obtained from 50 abnormal cells (red) and a typical spectrum of the ThinPrep glass slide used to prepare the samples; C) Scatter plot of the PCA analysis performed on the data recorded from normal (blue) and abnormal samples (red) after washing using H₂O₂ in the spectral range 1150-1800 cm⁻¹ and D) Loading 1 obtained from the PCA.