2017

Improved Protocols for Pre-Processing Raman Spectra of Formalin Fixed Paraffin Preserved Tissue Sections

Hugh Byrne

Technological University Dublin, hugh.byrne@tudublin.ie

Fiona Lyng

Technological University Dublin, fiona.lyng@tudublin.ie

Ola Ibrahim

Technological University Dublin, ola.ibrahim@mydit.ie

See next page for additional authors

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

Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Byrne, H.J., Lyng, F.M. & Maguire, A. (2017). Improved Protocols for Pre-Processing Raman Spectra of Formalin Fixed Paraffin Preserved Tissue Sections. *Analytical Methods*, Advance Article 24 Jul 2017. doi:org/10.1039/C6AY03308C

This Article is brought to you for free and open access by the Radiation and Environmental Science Centre at 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.

Authors

Hugh Byrne, Fiona Lyng, Ola Ibrahim, Adrian Maguire, Aidan Meade, Steven Flint, and M. Toner

Analytical Methods

Paper

Improved protocols for pre-processing Raman spectra of formalin fixed paraffin preserved tissue sections

O. Ibrahim^{a*}, A. Maguire^a, A. D. Meade^a, S. Flint^b, M. Toner^c, H. J. Byrne^a, and F. M. Lyng^a

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Although formalin fixed paraffin preserved (FFPP) tissues are a major resource for retrospective studies of disease progression, their use in vibrational spectroscopy studies has been undermined by issues of contributions of substrate and paraffin wax which persist in the spectra and can compromise spectral analysis. Recognising the microcrystalline nature of the wax in the tissue, which are inhomogeneously oriented with respect to the polarisation of the Raman source laser, in this study, we have developed a novel method for removing the paraffin wax contributions to the spectra using matrices of multiple wax spectra. FFPP tissue sections from the oral mucosca were obtained and, with no further chemical processing, the Raman spectral analysis of two regions, epithelium and connective tissue were compared. Matrices of multiple wax spectra were collected from different regions and subtracted from the epithelial and connective tissue spectra using a least squares analysis with non-negative constraints. Spectra of multiple cell components such as DNA and RNA were used in fitting the least squares model to reduce the residual error. The use of a data matrix of multiple wax spectra, as opposed to a single spectrum, results in a more accurate removal of the wax, hence reducing its contribution to spectral analysis. In unprocessed FFPP tissue sections, the contribution of the glass substrate is seen to be minimised in comparison to chemically dewaxed FFPP tissue sections. Contributions of the glass substrate were also successfully removed digitally using the same methodology. The combined results indicate that direct analysis of FFPP tissue sections is feasible using Raman spectroscopy, avoiding the need for chemical dewaxing. Additionally, the ability to use glass slides is very important in translation to the clinic.

Introduction

The standard method used in histopathological tissue processing is formalin fixation followed by paraffin wax embedding. The tissue is first dehydrated through different grades of ethanol, then infiltrated with paraffin wax. In this way, the samples are stabilized and can be stored for years. These formalin fixed paraffin preserved (FFPP) archival tissue libraries can potentially be a vast resource for retrospective studies of patient history and disease progression. Vibrational spectroscopy, including Raman scattering and infra-red absorption, has emerged as a promising candidate for rapid, label free automated screening of tissue pathologies, and the term spectral histopathology has been coined¹. However, paraffin wax poses a difficulty in vibrational spectroscopic studies of biological tissues, as the wax peaks are prominent and tend to interfere with either FTIR or Raman spectroscopic analysis^{2, 3}. A number of different approaches to remove the paraffin wax have been tested. Chemical dewaxing agents such as xylene, hexane, and histoclear were found to reduce but not completely eliminate the presence of

wax and the reduction was seen to be proportional to the time the tissue remained in the dewaxing agent, resulting in long (up to 18 hours) processing times for optimised protocols³. The amount of residual wax was also found to be influenced by the choice of substrate and tissue type. Abnormal tissue, such as metastatic tissue, was found to retain more wax after dewaxing than normal tissue, while calcium fluoride substrates were seen to retain more wax than low E slides⁴. Another disadvantage of chemical dewaxing is that chemical dewaxing agents have been shown to modify the tissue composition and therefore spectra^{5, 6}. The process of formalin fixation gives rise to formalin peaks appearing in the spectra at 1041 cm⁻¹ and 1492 cm⁻¹ which can be removed by washing in PBS⁷. As the molecular structure of paraffin wax is very similar to that of tissue lipids, the Raman spectral profile of the wax overlaps significantly with that of the tissue lipids, especially in the region 1000 cm⁻¹ to 1700 cm⁻¹. Because of the structural similarities, the process of chemical dewaxing can result in modification of the tissue lipidic content⁶. Digital wax removal methods such as independent component analysis and non-negatively constrained least squares analysis (NNLS) have been shown to be a promising alternative approach. However, these methods tend to introduce artefacts in the spectra due to over or underestimation of the wax contribution^{8, 9}.

Another consideration in the spectral analysis of biological tissue is substrate choice, as the optical properties of the substrate can influence the Raman spectrum of biological samples¹⁰. For research purposes, spectroscopists prefer to use spectrally neutral substrates

^a FOCAS Research Institute, Dublin Institute of Technology, Kevin St, Dublin 8, Ireland

Email: ola.ibrahim@mydit.ie

^b School of Dental Science, Trinity College Dublin, Lincoln Place, Dublin 2, Ireland

^c Central Pathology Laboratory, St. James Hospital, James Street, Dublin 8, Ireland

†Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

such as calcium fluoride or magnesium fluoride, as they contribute a much lower background signal compared to glass¹¹. However, they are significantly more expensive and therefore not a feasible option for routine diagnostic use^{11, 12}. Glass is commonly used in histopathological diagnosis as it is the clinical standard, is inexpensive and allows the processing of hundreds of samples a day. Its use in Raman spectroscopy, however, is hampered by the fact that glass exhibits a strong fluorescence emission under 785nm excitation⁴, overlaps with the biological spectra in the fingerprint region¹² and its contribution was found to vary with cellular morphology¹³. Notably, the glass contribution is significantly lower at 532nm than at the commonly employed 785nm, making it the wavelength of choice for such clinically oriented studies. Additionally, a shorter wavelength means greater confocality, higher intensity and a higher scattering efficacy¹¹. A number of different methodologies have been proposed for removal of the glass contribution, including iterative subtraction of a polynomial¹⁴ or glass reference spectrum¹² and Kerr et al. have recently performed a comparison of EMSC based techniques¹⁵ to such iterative techniques, suggesting that the former are more effective¹³.

The aim of this study was to explore improved methodologies of digitally removing the wax contributions to Raman spectra, for the example of oral tissue sections, thereby retaining the clinically established tissue processing protocols. The improved methodology is based on a recognition that the wax in the tissue has a microcrystalline structure. While this is not a significant factor for macro- or unpolarised micro-spectroscopy, as commonly the case for FTIR, in the case of Raman microspectroscopy, the microcrystalline domains are randomly oriented with respect to the polarisation of the source laser and so contribute inhomogeneously to the response, depending on the sampled spot. This inhomogeneity is accounted for by a matrix of multiple wax spectra used to digitally process the spectra. The same methodology was applied to removal of the glass substrate contributions, illustrating that spectroscopic differentiation of tissue regions can be achieved using Raman spectroscopy, free from confounding factors of wax and substrate contributions.

Materials and Methods

Sample preparation

Following ethical approval from the Research Ethics Committee, St James' Hospital, Dublin, Ireland, informed consent was obtained from 3 patients who had undergone an oral biopsy which was diagnosed as benign oral hyperplasia. Tissue blocks from these patients were obtained, for use in the study. 10µm parallel sections were cut from the FFPP tissue blocks and mounted on glass slides for Raman spectroscopy.

For the chemical dewaxing of tissue sections, routine xylene dewaxing was performed, which involves sequentially immersing the FFPP tissue sections in two baths of xylene, two of ethanol and one bath of industrial methylated spirits for 5,4,3,2 and 1 minute respectively³. Routine Hematoxylin and Eosin (H&E) was performed

on parallel tissue sections to aid in the delineation of the different regions in the tissues¹⁹.

82 component spectra were available from an in house library of spectra of pure biological macromolecules. Matrigel, which is a solubilised basement membrane preparation, was also trialled as a multicomponent reference biological spectrum, as it has previously been employed as such in background correction procedures for FTIR spectroscopy¹⁸.

Cell components (listed in Supplementary Table 1) were purchased from Sigma Aldrich. The components were placed on calcium fluoride disks prior to Raman spectroscopy measurement. All the components were measured in powder form except for DNA, RNA, cytochrome C, stearic acid, and collagen which were measured in an aqueous solution. The spectra of the 20 components (listed in table 1) selected by the Data Analysis procedure are presented in Supplementary Fig 1. Matrigel (Sigma Aldrich) was placed on a calcium fluoride disk prior to Raman recording of a matrix of 300 spectral points.

Instrumentation

A confocal Horiba Jobin Yvon LabRam HR 800 Raman (upright) spectroscopic microscope was used to record the spectra of the tissue and biochemical components. The microscope has an automated xyz stage and is coupled to a Peltier cooled CCD detector. A 50mW diode laser with 532 nm wavelength was used and the grating was set at 600 grooves/mm, while the confocal hole was set at 100µm. For mapping acquisition, the regions to map were selected using a 100X objective (MPLAN N Olympus, Japan), which also collected the backscattered light. The spectra were acquired over two accumulations, totalling 20 seconds per spectrum. The step size was set at 5µm and the spectral range was 400-1800cm⁻¹. The spectra of the cell components were acquired between 10-30 seconds and over 2-3 accumulations depending on the quality of the signal from the specific sample.

Data Analysis

All the data analysis was carried out using Matlab (Mathworks, US) with the PLS-Toolbox (Eigenvector Research Inc.) and in-house algorithms. The spectra were preprocessed by first smoothing using a Savitsky Golay filter (5th order, 13 points). They were then rubberbanded to correct the baseline and vector normalised²⁰. Principal component analysis (PCA), an unsupervised multivariate technique, was used to reduce the dimensionality of the data by calculating principal components which describe the greatest variance in the dataset²¹. PCA was employed to demonstrate the spectral differentiation of regions of epithelium and connective tissue, and the PC loadings were examined to elucidate the biochemical origin of the differentiation, as well as contributions of wax or substrate spectral features.

Classical least squares (CLS) fitting is a supervised technique used to estimate the weighted contributions of a set of input spectra to a

sample spectrum. It assumes that any complex spectrum is the weighted sum of all the base components that contribute to the spectrum as described in Equation 1²².

$$A = a_1 C_1 + a_2 C_2 + a_3 C_3 \dots \quad \text{Equation 1}$$

where A represents a sample spectrum, a1, a2 and a3 are component spectra and C1, C2 and C3 are the weighting coefficients, or concentrations, applied to each component spectrum in the summation. If one then considers the case of Raman spectroscopy in which a large number of variables are recorded and not all contributing spectral components are known, then there is an error (residual) associated with the estimation which will equate to the difference between the sample spectrum and the sum of the weighted spectra:

$$E_j = A_j - (a_{1j} C_1 + a_{2j} C_2 + a_{3j} C_3) \quad \text{Equation 2}$$

where E_j is the error associated with the intensity of the j th term (or j th wavenumber). The least squares method aims to minimise the error in fitting the sample spectrum, A. The minimum error occurs when the differentiation of the sum of squared errors is equal to zero²³. Differentiation of Equation 2 with respect to each of the coefficients results in a set of equations which can be solved using simultaneous equations or matrix mathematics. Non negatively constrained least squares analysis (NNLS) is a technique similar to that of CLS. NNLS is used to estimate the weighted contributions of a set of input spectra in a sample spectrum. However, unlike CLS, NNLS introduces non-negative constraints on the weighting coefficients of the input spectra²⁴.

For this study, in lieu of using an average spectrum of wax or glass, matrices acquired from maps of the wax and glass were used. The wax maps were taken from regions in the samples with no tissue, while the glass maps were taken from a clean glass slide. The matrices included about 300 spectral points and were processed in a similar fashion to the sample spectra. To account for spectral drift from day to day, the wax spectra were interpolated to every quarter wavenumber. Interpolated spectra were then shifted in both directions in increments of 0.25cm^{-1} to a maximum of $\pm 1\text{cm}^{-1}$, and then added to the original matrix of wax. Using an in-house Matlab algorithm, whereby the Raman spectra of 82 cell components (details in supplementary table 1) were iteratively tested to represent the tissue for NNLS and the combination of components associated with the least residual error was selected for use in the wax and glass subtraction, as listed in Table 1. The wax and glass matrices were then subtracted from the original processed spectra. The steps are summarised in Fig 1. The NNLS procedure weighted the contributions of each in an unsupervised fashion. However, the intention was to optimise the wax removal protocol, rather than to model the tissue analytically, based on the relative contributions of the specific components. Least Squares analysis, either supervised or unsupervised, has previously been used in a number of studies as a spectral analysis protocol^{16,17}.

CLS was used to determine the wax contribution to the first principal component of the unprocessed, chemically processed and digitally processed tissue.

Table 1 Cell components (Sigma Aldrich) used as inputs for NNLS

Cell Components			
Proteins	Lipids	Nucleic acids	Glycosaminoglycans
Actin	Stearic acid	RNA	Heparin
Collagen	P-Choline	DNA	Hyaluronic acid
Cytochrome C	P-Ethanolamine	Thymidine	
L-Histidine		Uracil	
L-Serine		Guanine	
L-Threonine			
Peroxidase			
Ubiquitin			
Taurin			
Apo E2

Fig. 1 Summary of the processing steps for the Raman spectra.

Results

Fig 2 shows a H&E stained FFPP tissue section, highlighting the epithelium and connective tissue regions. Raman maps were taken of the epithelium and connective tissue regions in the unprocessed FFPP tissues, the mean and standard deviation of which can be seen in Fig 3.

Fig. 2 Bright field image of H&E stained FFPP tissue showing the epithelium and connective tissue regions.

Fig. 3 Mean Raman spectra of epithelium and connective tissue in the unprocessed tissue sample. Shading denotes the standard deviation.

Initially, PCA was performed on the tissue spectra without removing the glass and wax contributions (Fig 4 (A)). The first principal component (PC1), which explained 59% of the variance, does not differentiate the two tissue types, and is dominated by wax peaks at 1063 cm^{-1} , 1134 cm^{-1} , 1172 cm^{-1} , 1296 cm^{-1} , 1419 cm^{-1} , 1441 cm^{-1} and 1464 cm^{-1} (Fig 4 (B)), which account for 38% of the loading. The broad streaking of the datasets indicates a significant variability of these contributions across both datasets. The second principal component (PC2), according to which the epithelium and connective tissue spectra are differentiated, explained 5% of the variance (Fig 4 (C)). The datasets are more tightly clustered with respect to PC2 and the loading has strong peaks at 815 cm^{-1} , 857 cm^{-1} , 920 cm^{-1} and 1245 cm^{-1} . These relate to collagen type 1 which, together with fibroblast cells, is the main component of the connective tissue^{25, 26}. The loading of PC2 also contains contributions from wax, but notably, some of them contribute negatively to the loading (1419 cm^{-1} , 1438 cm^{-1}), while some contribute positively (1136 cm^{-1} , 1296 cm^{-1}) indicating that the spectrum of the wax itself is not uniform across the tissue section. Glass peaks do not appear in the first or second principal component (Fig 4 (C)) indicating that the glass contribution to the spectra is not significant. Thus, the differentiation, using Raman microspectroscopy, of the two clearly biochemically distinct regions of tissue is based on their wax content, rather than their biochemical composition, which does not bode well for the potential to distinguish differing pathologies in the same tissue regions.

Fig. 4 (A) PCA scatter plot of unprocessed epithelium and connective tissue shows a clear separation on the second principal component with even distribution along the first principal component. (B) The first principal component is dominated by wax spectral peaks. (C) The second principal component is mainly collagen type 1 associated peaks on the positive side, relating to connective tissue, and protein associated peaks on the negative side, relating to epithelium.

The chemically dewaxed tissue sections were visually different from the unprocessed tissue sections, as they appeared whiter, indicating increased scatter which may be due to increased porosity as the wax and some of the lipidic content is removed in the dehydration process. Raman analysis of the chemically dewaxed tissue sections showed that, although the wax peaks at 1063cm^{-1} , 1134cm^{-1} , 1172cm^{-1} , 1296cm^{-1} , 1419cm^{-1} , 1441cm^{-1} and 1464cm^{-1} were reduced in comparison to the unprocessed tissue, they were not eliminated. Furthermore, the glass contribution in the chemically processed FFPP tissues was higher compared to the unprocessed tissues (Fig 5 (A)). Acquiring the spectra for connective tissue was problematic, as many of the spectra were saturated. This may be due to the increased scattering of the tissue after dewaxing, a phenomenon which can be reduced in a water immersion measurement protocol²⁷. The epithelium appears to retain more wax and has a higher glass contribution in comparison to the connective tissue (Fig 5 (B)).

Fig. 5 (A) Mean Raman spectra of epithelium in the unprocessed (black) and chemically dewaxed (blue) tissue sample and mean Raman spectrum of glass (red). Compared to the unprocessed epithelial tissue, the chemically dewaxed tissue has less wax contribution and more glass. (B) Mean Raman spectra of

epithelium and connective tissue in chemically dewaxed FFPP tissue section. Shading denotes the standard deviation. The black arrows highlight peaks from wax. The epithelium has more wax and glass contribution than the connective tissue.

PCA of the epithelium and connective tissue showed that the collagen and protein related peaks 815 cm^{-1} , 857 cm^{-1} , 875 cm^{-1} , 920 cm^{-1} and 1245 cm^{-1} had been promoted to the first principal component which explains 75% of the variance (Fig 6). However, although the contribution has been significantly reduced, there was still 32% wax contribution in the loading of the first PC, which positively discriminates the epithelium. Additionally, the two broad peaks of glass can be discerned in the loading of the first PC. This is to be expected, as there was a higher contribution of wax and glass in the epithelium.

Fig. 6 (A) PCA scatter plot of chemically dewaxed epithelium and connective tissue shows a separation on the first principal component. (B) On the positive side of the first principal component, collagen type 1 related peaks (red arrows) can be distinguished. While some wax (black arrows) and protein peaks can be distinguished on the negative side.

The spectral dataset for the FFPP sections, in which the wax contribution was removed digitally (using NNLS), were analysed.

Wax removal with a single wax spectrum was found to be insufficient, as most of the wax peaks were either over or underestimated. PCA of the matrix of wax spectra clearly illustrates the variability of the spectral signature of the nominally single chemical constituent, as shown in the scatterplot in Fig 7 (A). Notably, rather than exhibiting a widely scattered cluster centred about zero, the curved distribution is suggestive of an interdependence of PC1 (Fig 7(B)) and PC2 (Fig 7(C)), each indicating the degree of variability of the wax spectrum on a microscopic level. Although, macroscopically, paraffin is an amorphous wax, on the microscopic scale of the Raman laser spot, it presents a microcrystalline structure²⁸, and, as the laser is polarised, the registered spectrum is significantly variable, depending on the relative orientation of the laser polarisation and the microcrystalline structures. A single or averaged spectrum does not represent the specific spectral contributions at any point on the tissue, and therefore a subtraction protocol based on such results in incomplete removal of the spectral contributions of the paraffin.

Fig. 7 (A) PCA scatter plot of Wax matrix. (B) The first Principal component which explains 85% of the variance (C) The second principal component (D) Comparison of digital wax removal in epithelial tissue using a matrix of wax vs a single wax spectrum. Note, although the NNLS protocol constrains the weightings co-efficients to be non-negative, the component spectra can have negative contributions

Rather than averaging; a matrix of 300 individual spectra retains the intrinsic variability of the wax contributions, each of which is individually weighted in NNLS subtraction. The protocol resulted in much more efficient wax removal, as shown in Fig 7 (D). The resultant spectra of epithelium and connective tissue appear more uniform and have reduced wax contribution (Fig 8). This was also reflected in the PCA of epithelium and connective tissue, as the wax contribution to the loading of the first PC, which accounted for 90% of the variance, was reduced to 10%. The collagen and protein related peaks at 815 cm^{-1} , 857 cm^{-1} , 875 cm^{-1} , 920 cm^{-1} and 1245 cm^{-1} can be clearly distinguished in the first PC. Additionally, the spectra were tightly grouped according to PC1 in the PCA scatter plot, suggesting reduced variability within the groups and greater uniformity (Fig 9), although considerable point to point sample variability is evident by the spread according to PC2, particularly for the case of the connective tissue.

Fig. 8 Mean Raman spectra of epithelial and connective tissue after digital wax removal. Shading denotes the standard deviation.

blood vessels and oil ducts leading to a greater variability in the latter according to PC2 (B) First PC after digital wax removal. Wax contribution, black arrows, is reduced. The red arrows refer to 815 cm^{-1} , 857 cm^{-1} , 875 cm^{-1} , 920 cm^{-1} and 1245 cm^{-1} which are collagen related peaks.

For the glass removal, NNLS was also used. In a similar fashion to the wax removal, the glass contribution was reduced without affecting the integrity of the spectra (Fig 10). However, in contrast to the case of wax, where the use of a matrix conferred a marked improvement, there was no real difference between the use of a glass matrix or single spectrum in glass removal, indicating that the glass spectrum is more spatially homogenous (Fig 10 (B)). This is confirmed by the spectra being equally dispersed around the PC1 and PC2 axes, and showing no systematic variability (Fig 10 (C)).

Fig. 9 (A) PCA scatter plot of epithelium and connective tissue after digital wax removal using the cell components as a model. Epithelium is made up of tightly packed cells whereas connective tissue contains collagen fibrils, fibroblasts,

Fig. 10 (A) Mean Raman spectra of chemically dewaxed epithelial tissue before and after digital glass removal. (B) Mean Raman spectra of chemically dewaxed epithelial tissue with glass removal using a matrix and a single spectrum. (C) PCA scatter plot of the glass matrix.

Conclusions

One of the advantages of using vibrational spectroscopy techniques, such as Raman spectroscopy, to analyse biological tissues is that they require minimal sample preparation. In this study, chemical dewaxing, which increases the processing time and alters the composition of the tissue, has been substituted by an improved protocol for digital dewaxing. The demand for an improved protocol for spectroscopic analysis of the FFPP tissue sections had been previously highlighted²⁹. As demonstrated in Fig (4) the contributions from the wax significantly influence spectral analysis, which establishes the importance of subtracting it. From the results, it is clear that this method of digital removal greatly reduces the spectral contribution of wax, as the variance from the wax contribution to the first principal component is reduced from 38% and 32% in unprocessed and chemically processed samples respectively to 10% in the digitally dewaxed samples. Furthermore, the chemically dewaxed FFPP tissue samples had more glass background in the epithelium and more scattering background in the connective tissue compared to their unprocessed counterparts. This is most likely due to refractive index matching between the wax and glass resulting in reduced scattering of the incident light and Raman bands themselves in the unprocessed samples. The superiority of digital versus chemical wax removal has also previously been reported in FTIR spectroscopy, whereby modified extended multiplicative signal correction (EMSC) was used for the wax removal³⁰. Because of the variability between, and within, different types of tissue, while using a single mean cellular spectrum as a model or reference, such as has been previously reported for EMSC, is effective for cytology samples, it underestimates the complexity of histopathological tissue samples which can include epithelium and connective tissue, as well as different types of pathology¹³. Similarly, both EMSC and independent component analysis rely on a definitive model of wax (such as a three source model⁸) which does not take into account the variability of the paraffin wax spectrum, accentuated in Raman spectroscopy due to the polarisation dependence of commonly employed instruments. The use of multiple spectral components for the reference and wax contributions provides flexibility to account for the observed point to point variations of both.

This study is the first to utilize cell components and a wax matrix in the digital removal of wax from FFPP tissue sections. As the wax spectrum is composed of a number of peaks, and the ratio between these peaks is variable from one wax spectrum to another; using a matrix of wax captures this variability which therefore reduces the residual error in the NNLS model and results in improved estimated spectra. Paraffin wax is made up of linear hydrocarbon molecules which can form crystallites on a microscopic scale²¹. As these can be randomly oriented with respect to the laser polarisation, their contribution to the Raman spectrum of the tissue can be highly spatially variable. Polarisation dependent

studies are frequently employed to study the orientation of, for example, molecules in liquid crystals³¹ and have been employed to study the constituents of bone³². A study conducted, exploiting the polarisation dependence in biological tissues, has shown that in probing the deeper layers of the tissues the backscattered light has a greater chance of being depolarised by diffusion compared to the more superficial layers³³.

In contrast, glass has two broad peaks in the fingerprint region. Almost no difference was observed in the subtraction of glass using a single glass spectrum and using a glass matrix. However, this may be due to the fact that the glass contribution was not high in these samples and more or less uniform across the sample, so the matrix subtraction may be more effective in samples which show greater variability.

The cell components were selected from a pool of 82 components based on their fit of the NNLS model and, similar to the wax matrix, they confer a significant improvement on the wax subtraction. Choosing whether to use a single, commercially available component such as Matrigel or collections of biomolecules as representation of tissue constituents can depend on the application and the expected variability of the samples under investigation. For the application demonstrated here, the cell components gave superior performance in the removal of wax contributions to the spectra as the residual error was reduced from 18% when Matrigel was used in the NNLS correction to 8% when the cell components were used. A further improvement in the efficacy of removal could potentially be achieved by a more representative choice of cellular constituent spectra, targeted at specific tissue types.

By interpolating the wax spectra to every quarter wavenumber and shifting in both directions, the small spectral drifts from day to day calibration were accounted for resulting in a more accurate wax removal.

For a diagnostic technique to be translatable into routine clinical use it has to be affordable, and reliable. Effective digital subtraction of glass contribution from spectra means that glass could be used for FFPP or cytological samples, making vibrational spectroscopy techniques more affordable. Although thicker tissue section may overcome the issue of glass spectral contributions, they would have to be significantly thicker than what is in routinely employed in clinical practice. Furthermore, thicker sections would not resolve the issue of the paraffin wax contributions which are independent of tissue thickness. The digital subtraction of wax using a matrix, decreases the processing time and involves less modification of the samples resulting in more reliable results.

Acknowledgements

This research was funded by Science Foundation Ireland (12/IP/1494) and supported by the National Biophotonics and Imaging Platform (NBIP) Ireland funded under the Higher Education Authority PRTL (Programme for Research in Third Level Institutions) Cycle 4, co-funded by the Irish Government and the European Union Structural Fund.

References

1. M. Diem, A. Mazur, K. Lenau, J. Schubert, B. Bird, M. Miljkovic, C. Krafft and J. Popp, *Journal of Biophotonics*, 2013, **6**, 855-886.
2. M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sule-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner and F. L. Martin, *Nature Protocols*, 2014, **9**, 1771-1791.
3. E. O. Faolain, M. B. Hunter, J. M. Byrne, P. Kelehan, H. A. Lambkin, H. J. Byrne and F. M. Lyng, *Journal of Histochemistry & Cytochemistry*, 2005, **53**, 121-129.
4. L. M. Fullwood, D. Griffiths, K. Ashton, T. Dawson, R. W. Lea, C. Davis, F. Bonnier, H. J. Byrne and M. J. Baker, *Analyst*, 2014, **139**, 446-454.
5. E. O. Faolain, M. B. Hunter, J. M. Byrne, P. Kelehan, M. McNamara, H. J. Byrne and F. M. Lyng, *Vibrational Spectroscopy*, 2005, **38**, 121-127.
6. S. M. Ali, F. Bonnier, A. Tfayli, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. C. Lee, F. M. Lyng and H. J. Byrne, *Journal of Biomedical Optics*, 2013, **18**.
7. Z. W. Huang, A. McWilliams, S. Lam, J. English, D. I. McLean, H. Lui and H. Zeng, *International Journal of Oncology*, 2003, **23**, 649-655.
8. V. Vrabie, C. Gobinet, O. Piot, A. Tfayli, P. Bernard, R. Huez and M. Manfait, *Biomedical Signal Processing and Control*, 2007, **2**, 40-50.
9. A. Tfayli, C. Gobinet, V. Vrabie, R. Huez, M. Manfait and O. Piot, *Applied Spectroscopy*, 2009, **63**, 564-570.
10. L. Mikoliunaite, R. D. Rodriguez, E. Sheremet, V. Kolchuzhin, J. Mehner, A. Ramanavicius and D. R. T. Zahn, *Scientific Reports*, 2015, **5**.

ARTICLE

Journal Name

11. L. T. Kerr, H. J. Byrne and B. M. Hennelly, *Anal. Methods* 2015, **7**, 5041-5052.
12. F. Bonnier, D. Traynor, P. Kearney, C. Clarke, P. Knief, C. Martin, J. J. O'Leary, H. J. Byrne and F. Lyng, *Analytical Methods*, 2014, **6**, 7831-7841.
13. L. T. Kerr and B. M. Hennelly, *Chemometrics and Intelligent Laboratory Systems*, 2016, **158**, 61-68.
14. B. D. Beier and A. J. Berger, *Analyst*, 2009, **134**, 1198-1202.
15. K. H. Liland, A. Kohler and N. K. Afseth, *Journal of Raman Spectroscopy*, 2016, **47**, 643-650.
16. M. E. Keating, F. Bonnier and H. J. Byrne, *Analyst*, 2012, **137**, 5792-5802.
17. L. Notingher, G. Jell, P. L. Notingher, I. Bisson, O. Tsigkou, J. M. Polak, M. M. Stevens and L. L. Hench, *Journal of Molecular Structure*, 2005, **744**, 179-185.
18. P. Bassan, A. Kohler, H. Martens, J. Lee, H. J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke and P. Gardner, *Analyst*, 2010, **135**, 268-277.
19. D. Wittekind, *Biotechnic & Histochemistry*, 2003, **78**, 261-270.
20. J. Trevisan, P. P. Angelov, P. L. Carmichael, A. D. Scott and F. L. Martin, *Analyst*, 2012, **137**, 3202-3215.
21. S. Wold, K. Esbensen and P. Geladi, *Chemometrics and Intelligent Laboratory Systems*, 1987, **2**, 37-52.
22. H. Mark and J. Workman, *Chemometrics in Spectroscopy Classical Least Squares*, Part I : Mathematical Theory).
23. A. C. C. Rencher, William F., *Methods of Multivariate Analysis*, John Wiley and Sons, Third Edition edn., 2012.
24. D. Chen and R. J. Plemmons, in *The Birth of Numerical Analysis*, eds. A. Bultheel and R. Cools, World Scientific, Singapore, 2010.
25. T. T. Nguyen, C. Gobinet, J. Feru, S. Brassart-Pasco, M. Manfait and O. Piot, *Spectroscopy-an International Journal*, 2012, **27**, 421-427.
26. C. Squier and K. Brogden, *Human Oral Mucosa: Development, Structure and Function*, John Wiley & Sons, New Jersey, 2010.
27. F. Bonnier, A. Mehmood, P. Knief, A. D. Meade, W. Hornebeck, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. C. Lee, F. M. Lyng and H. J. Byrne, *Journal of Raman Spectroscopy*, 2011, **42**, 888-896.
28. M. Zbik, R. G. Horn and N. Shaw, *Colloids and Surfaces a-Physicochemical and Engineering Aspects*, 2006, **287**, 139-146.
29. H. J. Byrne, M. Baranska, G. J. Puppels, N. Stone, B. Wood, K. M. Gough, P. Lasch, P. Heraud, J. Sule-Suso and G. D. Sockalingum, *Analyst*, 2015, **140**, 2066-2073.
30. J. Nallala, G. R. Lloyd and N. Stone, *Analyst*, 2015, **140**, 2369-2375.
31. A. P. Davey, R. G. Howard, W. J. Blau and H. J. Byrne, *International Journal of Polymeric Materials and Polymeric Biomaterials*, 1999, **44**, 241-258.
32. A. J. Makowski, C. A. Patil, A. Mahadevan-Jansen and J. S. Nyman, *Journal of Biomedical Optics*, 2013, **18**.
33. Z. J. Smith and A. J. Berger, *Optics Letters*, 2005, **30**, 1363-1365.