2015-06-01

Optimal Choice of Sample Substrate and Laser Wavelength for Raman Spectroscopic Analysis of Biological Specimen

Hugh Byrne
Technological University Dublin, hugh.byrne@tudublin.ie

Laura Kerr
NUI Maynooth

Bryan M. Hennelly
NUI Maynooth

Recommended Citation

"Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen", Laura T. Kerr, Hugh J. Byrne, Bryan M. Hennelly, *Analytical Methods*, 7, 5041-5952 (2015) doi:10.1039/b000000x

Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen

Laura T. Kerr,[*a] **Hugh J. Byrne,**[b] **and Bryan M. Hennelly**[ac]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

Raman spectroscopy is an optical technique based on the inelastic scattering of monochromatic light that can be used to identify the biomolecular composition of biological cells and tissues. It can be used as both an aid for understanding the etiology of disease and for accurate clinical diagnostics when combined with multivariate statistical algorithms. This method is non-destructive, potentially non-invasive and can be applied *in vitro* or *in vivo* directly or via a fiber optic probe. However, there exists a high degree of variability across experimental protocols, some of which result in large background signals that can often overpower the weak Raman signals being emitted. These protocols need to be standardised before the technique can provide reliable and reproducible experimental results in an everyday clinical environment. The objective of this study is to investigate the impact of different experimental parameters involved in the analysis of biological specimen. We investigate the Raman signals generated from healthy human cheek cells using different source laser wavelengths; 473 nm, 532 nm, 660 nm, 785 nm and 830 nm, and different sample substrates; Raman-grade calcium fluoride, IR polished calcium fluoride, magnesium fluoride, aluminium (100 nm and 1500 nm thin films on glass), glass, fused silica, potassium bromide, sodium chloride and zinc selenide, whilst maintaining all other experimental parameters constant throughout the study insofar as possible.

1 INTRODUCTION

Conventional Raman microspectroscopy is the most commonly used Raman technique for the analysis and diagnosis of biological specimen. This method can be applied to cells and tissues *in vitro* and *ex vivo* by mounting them onto a substrate. It involves the use of a microscope and a confocal aperture in order to isolate the Raman spectrum from a specific microscopic location within the sample (e.g. the nucleus of a cell). Further information on the experimental set-up of a conventional Raman microspectroscopy system can be found elsewhere.[1,2]

The weak Raman signals associated with biological samples are often obscured by a broad slowly-varying background signal caused by fluorescent signals or stray light due to Mie scattering.[3] These signals can originate from a number of sources including the sample itself, the sample substrate and the optical elements in the system that are common to both the delivery path and the collection path, especially the microscope objective.[4] The presence of this background can compromise the ability to extract reliable and reproducible compositional information from biological Raman spectra.[5] It is therefore essential to obtain spectra that are free from background signals inasmuch as possible. The most common strategies to reduce the background in spectra are (i) using an appropriate source wavelength/substrate combination, (ii) designing a confocal system, and (iii) post-processing whereby baseline correction algorithms are applied to the recorded spectra. The use of an immersion microscope objective has also been reported by Bonnier et al. (2011)[3] to substantially reduce the background signal whilst increasing the Raman signal intensity due to a reduction in Mie scattering.

The objective of this paper is to investigate the impact of the source laser wavelength and sample substrate on the quality of Raman spectra obtained from biological samples with the aim of finding the optimum combination of laser wavelength and substrate for biological measurements. This is achieved by measuring the Raman signals obtained from fresh human cheek cells on a range of sample substrates for various source laser wavelengths ranging from the visible region through to the NIR. Results are presented both graphically and numerically based on the analysis of the data obtained using three different statistical metrics.

The breakdown of this paper is as follows; firstly the importance of standardising the experimental parameters associated with Raman spectroscopy is discussed, including sample substrates, source wavelengths and background subtraction methods. Section 3 provides information about the experimental materials and methods used in our study. Results from all experiments are shown in Section 4, followed by a discussion of

[a] *Department of Electronic Engineering, National University of Ireland Maynooth, Ireland. E-mail: lauratheresekerr@gmail.com*
[b] *Focas Research Institute, Dublin Institute of Technology, Ireland*
[c] *Callan Institute, National University of Ireland Maynooth, Ireland*

these results in Section 5.

2 STANDARDISATION OF RAMAN SPECTROSCOPY

While Raman spectroscopy has been demonstrated to produce accurate diagnostic results, further development is necessary to ensure that Raman spectroscopic systems are sufficiently robust for everyday clinical usage.[6] The lack of standardisation, in terms of equipment, consumables and measurement protocols has resulted in the recording of significantly differing spectra across studies to date and we believe that this has hindered further advancement of this technique. Even within the application of Raman spectroscopy to the diagnosis of one particular pathological disease (e.g. bladder cancer), a wide range of sample substrates, source laser wavelengths and integration times have been applied to date, all resulting in moderately varying spectra for the same disease.[2] If Raman spectroscopy is ever to become a commonly used clinical tool, we believe that it is important that a standardised procedure is established in order to overcome these inconsistencies.[7] The aim of this paper is to identify the optimum source wavelength and sample substrate for Raman spectroscopic measurements of biological samples, and also to identify the similarities, or lack thereof, between spectra recorded using these different experimental parameters.

2.1 SAMPLE SUBSTRATES

Good sample substrates for Raman spectroscopy should produce low background signals, be biocompatible and non-toxic for the cells and tissues placed on them, and be as cost effective as possible. Previous studies have reported the use of a wide range of substrates including calcium fluoride[8–10], aluminium[11,12], quartz[13–15] and 3D collagen gels[16]. However, substrates that produce low background signals for NIR sources are often expensive, and are available in different levels of purity/film thicknesses, which can produce variable results with Raman spectroscopy. Therefore, by analysing the results obtained from a wide range of both substrates and source wavelengths in this study, we hope to identify which substrates produce the lowest background signals at each wavelength while monitoring the cost effectiveness associated with each substrate. Our core motivation is to establish the optimal substrate/wavelength in terms of cost and performance for applications in biomedicine. In order to gauge the performance of each substrate in a controlled manner, we record spectra from fresh cheek cells on each substrate using each wavelength and then we apply a number of different quality related metrics to this data. These metrics are described later in Section 4.

An important consideration that is not included in our study is the biocompatibility of the various substrates. This is particularly important for those cases where living cells are under investigation which often involve the incubation or growth of cells directly on to the substrate. However, this has been previously explored by Meade et al. (2006), whereby substrates were coated with fibronectin, laminin and gelatin, resulting in improved cell proliferation and similar Raman spectra to those achieved without the use of any coatings.[17]

2.2 SOURCE WAVELENGTH

One of the most important elements in any Raman spectroscopy experiment is the choice of source laser. There are advantages and disadvantages to each laser wavelength that need to be considered in order to obtain the optimum Raman signals. Many materials, including the sample itself, the substrate and the microscope objective, produce a background signal in a particular wavelength region, which may swamp any weak Raman signals present. In particular some substrates (especially glass) emit a large fluorescent background signal as the laser moves from the blue region up to the red or NIR region, making a laser in the lower end of the visible spectrum more desirable. On the other hand, it has been shown that a large background can result from scattering from the sample itself, which enters the spectrometer as stray light.[3] This scattering is minimised at longer wavelengths, as is photodegradation of tissue samples.[18] It should be noted that when using a laser emitting in the NIR region, it may be preferable to use a fluorite microscope objective in order to reduce the fluorescent background signal emanating from the objective. These objectives, which are produced by the large optics companies, often have various names, e.g. fluor, fluorite, fluortar, neofluar, and are manufactured using a mineral form of calcium fluoride. The background signal from the microscope objective can be reduced by using a suitable confocal aperture.[4]

The number of photons scattered is also indirectly related to the laser wavelength, with the intensity of Raman lines being proportional to the fourth power of the laser frequency[19]:

$$I \propto \nu^4 \tag{1}$$

Therefore, when a comparison is made between a 473 nm laser and an 830 nm laser:

$$\left(\frac{\nu_{473}}{\nu_{830}}\right)^4 = \left(\frac{c}{\lambda_{473}} \times \frac{\lambda_{830}}{c}\right)^4 = 9.481 \tag{2}$$

it can be seen that the 473 nm laser produces Raman lines that are approximately 9.5 times more intense than those produced by an 830 nm laser for the same laser powers, assuming non-resonant conditions.[20] In Figure 1, the relative Raman scattering intensity is shown for a source laser at 532 nm,

a common wavelength used in Raman anaylsis of biological specimen. While a 473 nm laser will produce almost twice as many Raman scattered photons, a 785 nm laser will produce approximately one fifth as many. Therefore, it can be seen that the scattering efficiency is higher at lower wavelengths, resulting in the use of shorter integration times and lower powered lasers in the blue/green regions, although the quantum efficiency of the grating and CCD detector being used must also be taken into consideration. Another important factor is that the optical window for biological tissues exists within the NIR region, where the absorption of light within the tissue sample is minimal.[21] Within this optical window (700–900 nm), biological samples are considered relatively transparent, and measurements in this region result in significantly less tissue damage from the laser.[18]

Another significant factor when considering the laser wavelength is the spectral resolution of the system. Several elements need to be considered for the spectral resolution such as the diffraction grating, detector, and the focal length of the spectrometer. The wavelength bandwidth that can be observed at the output of a spectrometer is a function of the angular dispersion and the focal length of the spectrometer. Since the angular dispersion is itself a function of wavelength, the observable bandwidth differs depending on the centre wavelength that is chosen. Precise calculation of this bandwidth for a particular centre wavelength can be achieved using the grating equation, and requires knowledge of the various spectrograph parameters such as grating angle, focal length, output aperture size and grating period.[22] For gratings with relatively low dispersion, e.g. 300 grooves per mm (gr/mm), the observable wavelength bandwidth (and therefore resolution in nm) is approximately constant for different centre wavelengths ranging from 500 nm to 900 nm, though this approximation fails for highly dispersive gratings. However, in Raman spectroscopy, where units such as Raman shift or wavenumbers are more often used, the spectral resolution (in wavenumbers) of the system increases with the source wavelength. When the wavelength is shifted through Raman spectroscopy from the excitation wavelength λ_e to the scattered wavelength λ_s, the shift in wavelength is given by:

$$\Delta\lambda = \lambda_e - \lambda_s \tag{3}$$

but the corresponding wavenumber shift ($\Delta\tilde{\nu}$) is given by:

$$\Delta\tilde{\nu}\ (\mathrm{cm}^{-1}) = \left[\frac{1}{\lambda_e(\mathrm{nm})} - \frac{1}{\lambda_s(\mathrm{nm})}\right] \times \frac{10^7\ (\mathrm{nm})}{(\mathrm{cm})} \tag{4}$$

$$= \left[\frac{\Delta\lambda}{\lambda_e(\lambda_e - \Delta\lambda)}\right] \times 10^7$$

Therefore, assuming that the wavelength bandwidth, $\Delta\lambda$, remains constant regardless of the centre wavelength chosen, it follows that the ratio of spectral bandwidths for two different source wavelengths, λ_{e1} and λ_{e2}, will be given by:

$$\frac{\Delta\tilde{\nu}_1}{\Delta\tilde{\nu}_2} = \frac{\lambda_{e2}(\lambda_{e2} - \Delta\lambda)}{\lambda_{e1}(\lambda_{e1} - \Delta\lambda)} \tag{5}$$

Thus, for a sample case of $\Delta\lambda$ = 150 nm, we can conclude that 473 nm source wavelength results in approximately 3.7 times more spectral bandwidth than 830 nm. Conversely, the resolution at 830 nm is ~3.7 times smaller than that at 473 nm. The overall efficiency of a grating also depends on the blaze angle of the grating, which we do not consider here. Typically, gratings would be selected to provide the best resolution within a specific acquisition time frame over the desired spectral range for the chosen laser wavelength. For direct comparison in this study, we use two Raman systems (both Horiba Jobin Yvon LabRam 800 HR) and two identical CCD cameras operating with a 300 gr/mm grating for all measurements. It is noted that different results could be obtained by changing any of these parameters (Raman system, CCD or grating line number or blaze), but for the sake of uniformity, these parameters were kept constant in this study. In Figure 1, a relative comparison is shown for the wavenumber resolution (blue) and bandwidth (red) obtained using a 532 nm source laser and those obtained using other lasers from 450 nm up to 1064 nm. This graph is based on Equation 5, and therefore is dependent on the assumptions that were made in its derivation; a wavelength bandwidth of 150 nm is assumed.

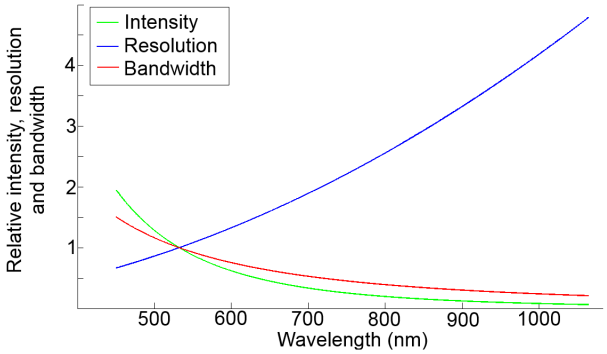


Fig. 1 A relative comparison of Raman signal intensity (green), wavenumber resolution (blue) and bandwidth (red) obtained using source lasers from 450 nm up to 1064 nm relative to a 532 nm laser.

The effect of the source optical mode for Raman spectroscopic measurement of human tissues has previously been analysed by Li et al. (2014)[23], whereby the influence of single-mode (SM) and multi-mode (MM) source lasers at 785 nm and 830 nm were compared in terms of the background signal intensity generated by tissue autofluorescence and the Raman signal intensity measured from human tissue

samples. Overall, a reduction in background, increase in SNR and a reduction in Mie scattering was found for 785 nm SM when compared to 785 nm MM. In terms of source wavelength, 785 nm spectra had a lower background intensity than that at 830 nm, and it was found that the SNR was 1.2–1.6 times higher for 785 nm MM than that recorded from the 830 nm MM.[23] Therefore, taking this into consideration, all measurements in this study are based on a SM source.

2.3 BACKGROUND SUBTRACTION METHODS

Spectra obtained using Raman spectroscopy consist of three main components: the Raman signals, the background signal and noise. The background signal and noise are generally reduced in the recorded spectra using preprocessing techniques before further analysis. Numerous techniques are available to do this, such as polynomial fitting, wavelet transform or the rolling ball technique.[24–26] In this study, we develop an algorithm based on some minor changes to the method proposed by Beier et al. (2009), which is based on the subtraction of a known background signal and a fifth-order modified polynomial fit.[27]

Firstly, three spectra were recorded from the substrate and an average background spectrum $B_0(\lambda)$ was calculated. Since in this section we outline a discrete numerical algorithm, it is more appropriate to describe the signals in terms of their discrete representations; where $\lambda \rightarrow n\delta\lambda$ and n takes integer values from $0 \rightarrow N-1$, where $\delta\lambda$ denotes the sampling interval of the recorded spectrum and $N\delta\lambda$ is equal to the bandwidth. A Gaussian smoothing function was applied to the background spectrum to remove any noise. Equation 6 represents the process of discrete convolution with a Gaussian filter that is sampled at the same rate as the spectrum. We can take the width of the Gaussian function to be $M\delta\lambda$, which corresponds to the region where the function has appreciable values, for example if we choose $M\delta\lambda = 6\sigma$ the region will contain over 99% of the Gaussian signal's energy.

$$B_0(n\delta\lambda) = \sum_{m=0}^{N-1} B_0(m\delta\lambda)\,\exp\left[\frac{\delta\lambda^2(m-n)^2}{2\sigma^2}\right] \qquad (6)$$

where the standard deviation, σ, was set to 2 cm^{-1} and n takes the same range of values as previously mentioned. This equation can be calculated by applying zero padding followed by multiplication of the Discrete Fourier Transform (DFT) of the signal and that of the Gaussian filter followed by an inverse DFT.

A range of different weights or concentrations C of this background signal $B_0(n\delta\lambda)$ were subtracted from the recorded cell spectrum $X_0(n\delta\lambda)$, followed by the subtraction of a fifth order polynomial $P(n\delta\lambda)$, which was generated using Matlab's *polyfit* function, resulting in a spectrum, $R(n\delta\lambda)$, consisting of residual values:

$$R(n\delta\lambda) = [X_0(n\delta\lambda) - (C \times B_0(n\delta\lambda))] - P(n\delta\lambda) \qquad (7)$$

The sum of the square of these residual values was obtained for each C value, until a minimum $R_1(n\delta\lambda)$ value was obtained, resulting in a new estimate of the background spectrum $B_1(n\delta\lambda)$:

$$B_1(n\delta\lambda) = X_0(n\delta\lambda) - R_1(n\delta\lambda) \qquad (8)$$

$$B_1(n\delta\lambda) = [C_1 \times B_0(n\delta\lambda)] + P_1(n\delta\lambda) \qquad (9)$$

A new spectrum $X_1(n\delta\lambda)$ is now defined that is made up of the values of both $X_0(n\delta\lambda)$ and $B_1(n\delta\lambda)$, where for each individual spectral component we take the minimum value from the two corresponding components of $X_0(n\delta\lambda)$ and $B_1(n\delta\lambda)$:

$$X_1(n\delta\lambda) = \min[X_0(n\delta\lambda), B_1(n\delta\lambda)] \qquad (10)$$

This entire process is then repeated by replacing $B_1(n\delta\lambda)$ with $B_i(n\delta\lambda)$ until the optimal fit of $B_i(n\delta\lambda)$ to $X_i(n\delta\lambda)$ is found such that the peaks of the original cell spectrum, $X_0(n\delta\lambda)$, all lie above the modeled background $B_i(n\delta\lambda)$:

$$R_i(n\delta\lambda) = [X_{i-1}(n\delta\lambda) - (B_0(n\delta\lambda) \times C_i)] - P_i(n\delta\lambda) \qquad (11)$$

$$B_i(n\delta\lambda) = [C_i \times B_0(n\delta\lambda)] - P_i(n\delta\lambda) \qquad (12)$$

$$X_i(n\delta\lambda) = \min[X_0(n\delta\lambda), B_i(n\delta\lambda)] \qquad (13)$$

The Raman peaks can then be isolated from the original cell spectrum by:

$$\text{Raman peaks} = X_0(n\delta\lambda) - X_{final}(n\delta\lambda) \qquad (14)$$

Figure 2 demonstrates the application of this algorithm to one of the spectra recorded in this study, whereby different values of C are applied to $B_0(n\delta\lambda)$ (and $B_i(n\delta\lambda)$ in proceeding iterations) in order to obtain a residual spectrum containing only Raman peaks without any background contributions.

The background subtraction method algorithm that we have just described is based on the method developed by Beier et al., which the reader may consult for further information. Some notable changes to this algorithm have been made; firstly, we have applied a smoothing filter to the initially recorded background signal to reduce noise; secondly for the estimate of C, a uniformly sampled range of C is searched in each iteration as opposed to the use of an empirically chosen factor and the *fminsearch* function in Matlab; and finally, we model our background on a combination of C times the recorded background, $B_0(n\delta\lambda)$, instead of C times the recorded cell data, $X_0(n\delta\lambda)$ as suggested by Beier and colleagues.[27] A simple implentation of the background subtraction algorithm has been written in Matlab that takes in the background and cell spectra as text files and produces a spectrum consisting only of Raman peaks; this code is freely available upon request.

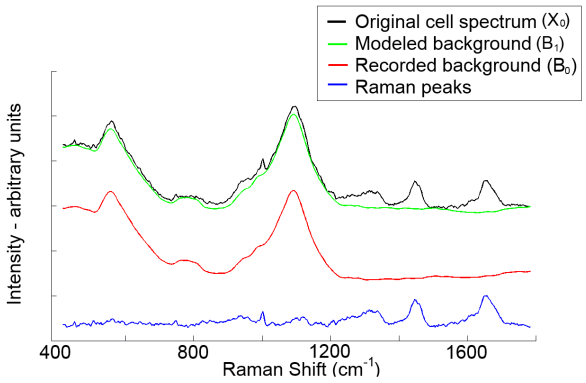


Fig. 2 An example of how the background subtraction algorithm works - whereby the recorded background signal is combined with a fifth order polynomial until a value of C is found such that the modeled background fits directly under the Raman peaks of the original cell spectrum.

3 MATERIALS & METHODS

Fresh cheek cells were swabbed from a single healthy human volunteer and placed directly on to each of the substrates before each set of experiments, and were allowed to air dry for 5 minutes before measurement. No additional preserving/fixing agents or washing steps were applied to the samples, and therefore some spectral signals relating to dried saliva, or other oral contaminants, may be present across all spectra. We note that no debris or contamination was visible in the images of the cells that were used in our experiments and there was no obvious contamination observed in any of the spectra recorded. While we acknowledge the presence of saliva and other possible contaminates, we do not expect their Raman signals to significantly affect the results presented in Section 4.

Raman spectroscopy measurements were performed across two commercial Raman systems (both Horiba Jobin Yvon LabRam 800 HR) with Synapse cooled CCD detectors, a 50x microscope objective (Olympus MPlanN 50x/0.75 ∞/0/FN22), 50 μm confocal aperture, 300 gr/mm grating and an acquisition time of 30 s averaged over 2 iterations, with the subsequent removal of cosmic rays.[28] However, the confocal aperture was opened to 100 μm for measurements with the 830 nm source due to the low photon count and lower quantum efficiency of the CCD in this region. Both systems were calibrated using a silicon wafer prior to measurements.

Background spectra were recorded from each substrate at every wavelength, and were used to isolate the Raman signals from the recorded biological spectra using the background subtraction method that was previously explained in Section 2.3.

The following substrates were used for this study:

- Calcium fluoride – Raman-grade (Crystran Ltd., UK)
- Calcium fluoride – IR grade polished (Crystran Ltd., UK)
- Magnesium fluoride (Crystran Ltd., UK)
- 100 nm aluminium thin film on glass (Deposition Research Laboratory Inc., USA)
- 1500 nm aluminium thin film on glass (Deposition Research Laboratory Inc., USA)
- Glass
- Fused silica – IR grade polished (Crystran Ltd., UK)
- Potassium bromide (Edmund Optics, UK)
- Sodium chloride (Edmund Optics, UK)
- Zinc selenide (Crystran Ltd., UK).

The following lasers were used for this study:

- 473 nm solid state diode laser (50 mW)
- 532 nm solid state diode laser (50 mW)
- 660 nm solid state diode laser (100 mW)
- 785 nm CLDS point mode diode laser (300 mW)
- 830 nm CLDS point mode diode laser (200 mW)

where measurements using the 473 nm, 660 nm and 830 nm lasers were performed on one Raman system, and measurements at 532 nm and 785 nm on the other system, with all other parameters maintained as constant insofar as possible; variances however are present in the output power of each source laser (see Section 3). We have used lasers (and their corresponding powers) which are regularly employed in commercial Raman systems for biological measurements.

4 RESULTS

The background signals recorded are a combination of intrinsic Raman and/or fluorescence from the sample substrate and from optical elements within the Raman system, and in particular from the microscope objective. In order to gauge the contribution of the optical system alone, we begin this section by presenting the spectra obtained from the system without the presence of any samples or substrates. The results are shown in Figure 3 for each source wavelength. The majority of this signal contribution was associated with the microscope objective used.

Figure 4 represents a single baseline-corrected spectrum recorded from fresh human cheek cells on Raman-grade calcium fluoride using a 532 nm source laser, recorded using the parameters outlined in Section 3; this spectrum proved to be the optimum spectrum recorded. Here we have highlighted the key Raman peaks associated with cheek cells, and we have

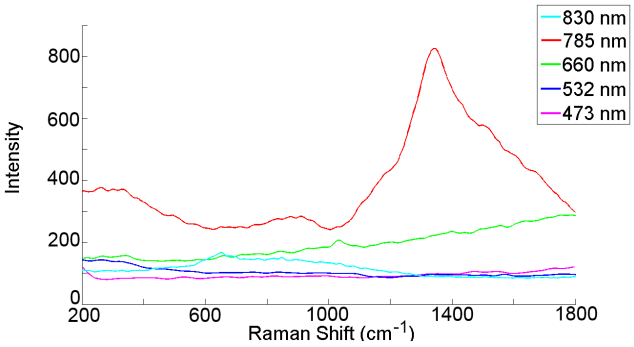
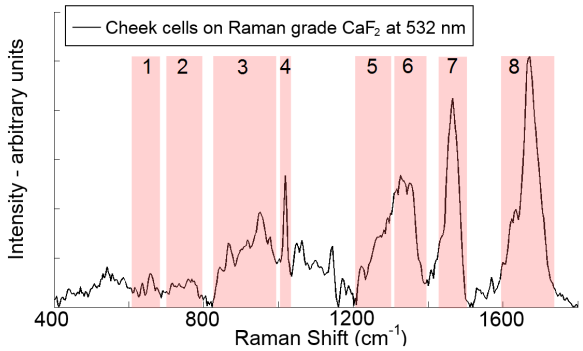


Fig. 3 Background signals generated by the optical elements within the Raman system for each of the above wavelengths - a Gaussian smoothing function was applied to all spectra.



1 - proteins; 2 - nucleic acids; 3 - proteins;
4 - phenylalanine; 5 - amide III; 6 - proteins and nucleic acids;
7 - CH_2 mode of proteins and lipids; 8 - amide I

Fig. 4 A single Raman spectrum obtained from fresh human cheek cells on Raman-grade calcium fluoride with a 532 nm source laser, with key biomolecular regions highlighted.

identified these important biomolecular regions which are often used in the analysis of cells in Raman based cytology studies.

The spectra shown in Figure 5 to Figure 14 represent the background signals and Raman cell peaks associated with each substrate for every wavelength, using the parameters outlined in Section 3 (i.e. all spectra are recorded with an acquisition time of 30 s, averaged over 2 iterations). On the left-hand side, the background signals obtained from each substrate at each wavelength are shown at their recorded intensity values after convolution with a Gaussian smoothing function was applied to remove any additional noise contributions. This noise reduction is necessitated by the low photon count, and therefore high shot noise, brought about by the presence of a 50 μm confocal aperture. The corresponding spectra on the right were recorded from fresh cheek cells on each substrate for each wavelength; these spectra have been baseline corrected using the background subtraction algorithm discussed in Section 2.3. Following this, the spectra were area normalised by dividing each spectrum by the sum of its intensity.

Background spectra were recorded between 200–1800 cm^{-1} for all substrates. Cell spectra are generally shown between 400–1800 cm^{-1}. The reason for this is two fold; firstly this is the fingerprint region for biological specimen and is the most commonly observed spectral region for Raman related biochemical investigations, and secondly, this removes any issues with baseline correction of large background contributions in the lower wavenumber region; problems can occur when applying the background subtraction algorithm in regions where the background has relatively strong intensities, e.g. the CaF_2 peak at approximately 321 cm^{-1}, which results in a breakdown of the algorithm. This is easily avoided by discarding the lower part of the spectrum for both the cell signal and the background. Additionally, some substrates produce large background signals between 400–600 cm^{-1} which were not entirely removed by the background subtraction algorithm, such as zinc selenide (for all source wavelengths) and fused silica at 532 nm. For this reason, these spectra (and their corresponding metric values in Tables 1–3) relate only to the 600–1800 cm^{-1} region.

In order to perform a quantitative evaluation of the quality of each of the recorded spectra, the background subtracted spectra were all directly compared to that recorded from cells on Raman-grade calcium fluoride using a 532 nm laser, as seen in Figure 4. For the sake of completeness, three separate metrics are employed in this comparison; (i) discrete correlation, (ii) normalised covariance, and (iii) mean square error (MSE). These three metrics compare overlapping regions of spectra, and thus, produce a value representing how similar the spectra are to that recorded from Raman-grade calcium fluoride at 532 nm. In this way, we get a good estimate of how similar the Raman peaks and signal-to-noise ratios are when recorded across a range of substrates and source wavelengths.

Discrete correlation is defined by the equation below:

$$(X \star Y)(n\delta\lambda) = \sum_{m=0}^{2N-1} X^*[n\delta\lambda]Y[(n+m)\delta\lambda] \qquad (15)$$

where $\delta\lambda$ is the sampling interval of the two signals X and Y, which both have length N, $\star$ represents correlation, and X^* is the complex conjugate of X. In Section 2.2 we discussed the relationship between spectral resolution and laser wavelength for an identical spectrograph grating. In order to employ the discrete correlation of spectra recorded using different source wavelengths, we must first ensure that they have identical sampling intervals. This was achieved by applying an interpolation function in Matlab. Correlation was then implemented using Matlab's *xcorr* function and the corresponding coefficient values that were recorded can be seen in Table 1, where

1.000 is the optimum result defining a perfect match between spectra, values above 0.900 are a good match, and lower values represent less correlation between spectra.

The second metric, normalised covariance, has a similar range of values between 0 and 1. We define normalised covariance as follows:

$$\text{Covariance} = \frac{(X \cdot Y)^2}{(X \cdot X)(Y \cdot Y)} \tag{16}$$

where $(X \cdot Y)$ represents the dot product of X and Y. This metric has previously been applied to compare the similarity between Raman spectra for cosmic ray removal.[29] The corresponding values obtained using this method are available in Table 2, where 1.000 is the optimum result, and lower values represent less covariance between spectra.

The third metric, MSE, does not have ranges between 0 and 1, but rather between 0, indicating an identical likeness, and some arbitrary maximum value that is dependent on the differences in the values of the two signals. MSE is commonly used in signal processing to compare the likeness of two signals and is defined as follows:

$$\text{MSE} = \text{mean}((X - Y)^2) \tag{17}$$

where mean() denotes the process of taking the average value. For comparative purposes, the MSE values obtained have been presented relative to the maximum value (i.e. all values are compared to the worst spectrum):

$$1 - \left(\frac{MSE}{MSE_{max}} \right) \tag{18}$$

This resulted in values ranging between 0 and 1, which are presented in Table 3, where similar to the other metrics, 1.000 represents the optimum result, and lower values represent a greater mean square error between both spectra.

5 DISCUSSION

Since the overarching goal of this study was to identify the optimum sample substrate and source wavelength for Raman spectroscopic analysis of biological samples, we will begin by discussing the results obtained here in terms of sample substrates, and the advantages and disadvantages associated with each. Then we will discuss the influence of different source lasers on the quality of Raman spectra generated. And finally, we will combine both sample substrates and source lasers to identify which combination should be used to produce the optimum Raman spectra from biological materials.

The second goal of this study was to also be able to identify the similarities, or lack thereof, between spectra recorded using these different experimental parameters. As seen in Figure 5 to Figure 14 in the results section, largely differing background spectra were recorded across the range of substrates and wavelengths, with large background signals appearing within specific spectral regions. However, despite these often large background signals, the standard Raman peaks associated with biological specimen have been isolated, and are comparable in most circumstances.

We note that some of the substrates presented in this paper can provide good cell spectra for 785 nm and 830 nm despite the relatively poor results shown here in Figure 5 to Figure 14, particularly for Raman-grade calcium fluoride (Figure 5). For example, previous studies published by Stone et al.[30], Grimbergen et al.[8] and de Jong et al.[10] have shown good spectra for samples recorded at 830 nm/845 nm using calcium fluoride substrates, which could be achieved here by simply increasing the laser power, enlarging the confocal aperture or increasing the exposure time. The results shown here may also be accounted for by considering the significantly lower number of scattered photons for higher laser wavelengths (see Section 2.2 on spectral resolution) and the quantum efficiency of the CCD detector used for recording spectra. We acknowledge that the use of a CCD detector with a higher quantum efficiency in the NIR region would improve the spectral quality in this region. However, the purpose of this paper was to compare the different substrates under identical experimental conditions and to discover the relative performances of these parameters in a controlled experiment. We accept that the use of two different experimental systems reduces this level of control but identical conditions across all experiments were applied insofar as possible.

From a sample substrate perspective, the most consistent Raman spectra, regardless of source wavelength, were obtained from Raman-grade calcium fluoride and both aluminium coated substrates. This has been verified both visually from the recorded spectra (in the ability to identify key biomolecular peaks and from the signal-to-noise ratio seen in the spectra, see Figure 5, Figure 8 and Figure 9), and from all three of the metrics employed in this study, as shown in Tables 1–3. Good results were also obtained from magnesium fluoride (Figure 11), potassium bromide (Figure 12) and sodium chloride (Figure 13).

Focussing on the NIR source wavelengths used here (785 nm and 830 nm); while we found that a number of substrates provided decent results, the qualitatively best spectra (under similar conditions of exposure time and CCD quantum efficiency) were obtained using the aluminium coated substrates. We believe the reason for this is twofold. Firstly, the aluminium coating effectively blocks the glass and has the lowest background signal of all the substrates measured. Secondly, we believe that the reflecive substrates provide approximately four times more Raman scattered

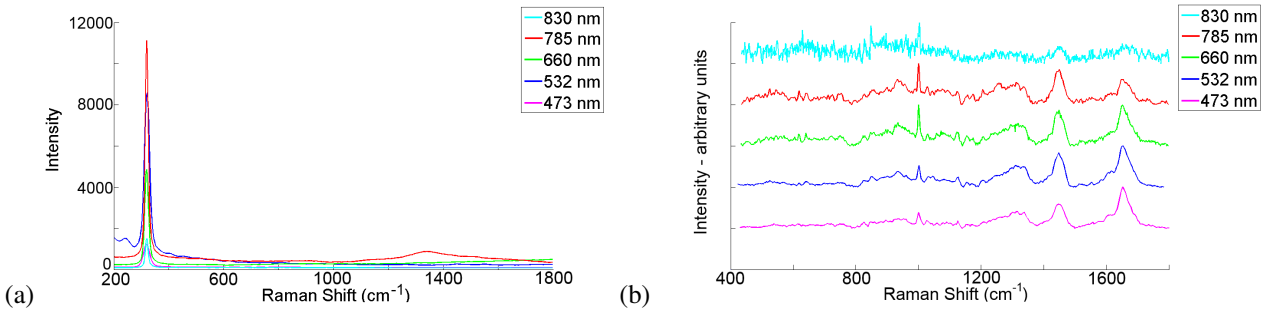


Fig. 5 (a) Background signal of **Raman-grade calcium fluoride** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on Raman-grade calcium fluoride for each of the above wavelengths, with an acquisition time of 30 s (x2). Note, improved spectra at 830 nm can be obtained using a longer acquisition time.

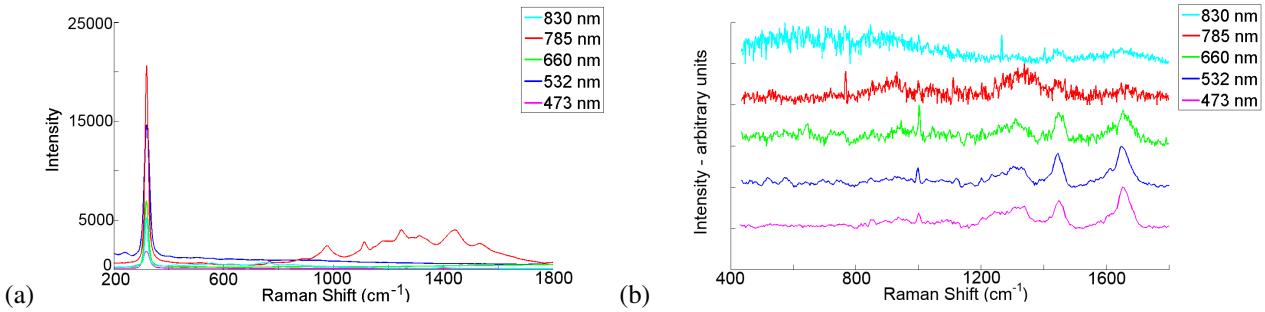


Fig. 6 (a) Background signal of **IR polished calcium fluoride** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on IR polished calcium fluoride for each of the above wavelengths, with an acquisition time of 30 s (x2).

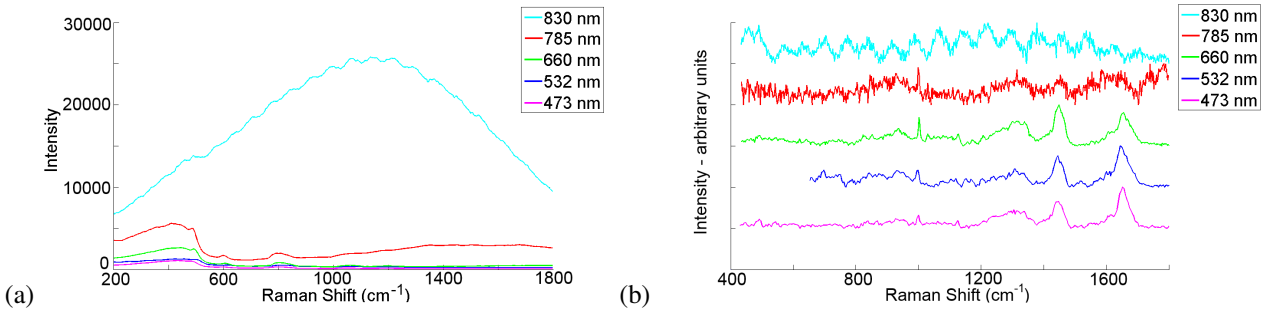


Fig. 7 (a) Background signal of **fused silica** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on fused silica for each of the above wavelengths, with an acquisition time of 30 s (x2). Note, the cell spectrum at 532 nm is only shown between 600-1800 cm^{-1} due to the large background contribution in the lower wavenumber region.

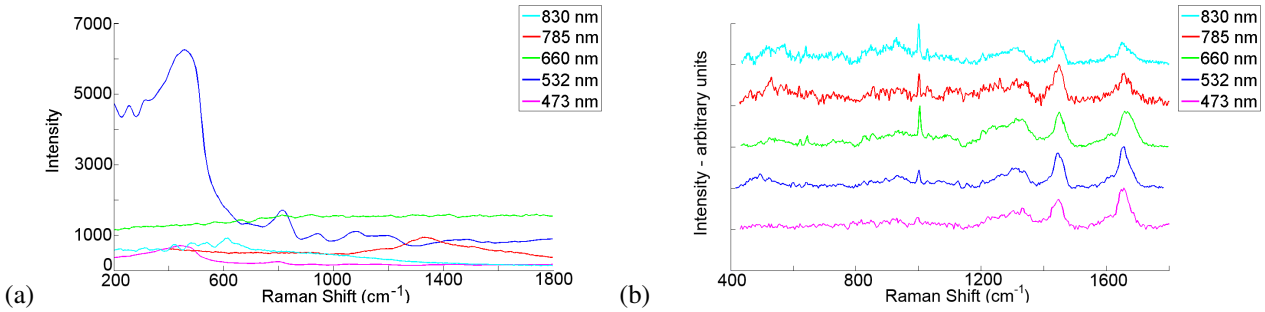


Fig. 8 (a) Background signal of **100 nm aluminium thin film on glass** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on 100 nm aluminium thin film for each of the above wavelengths, with an acquisition time of 30 s (x2).

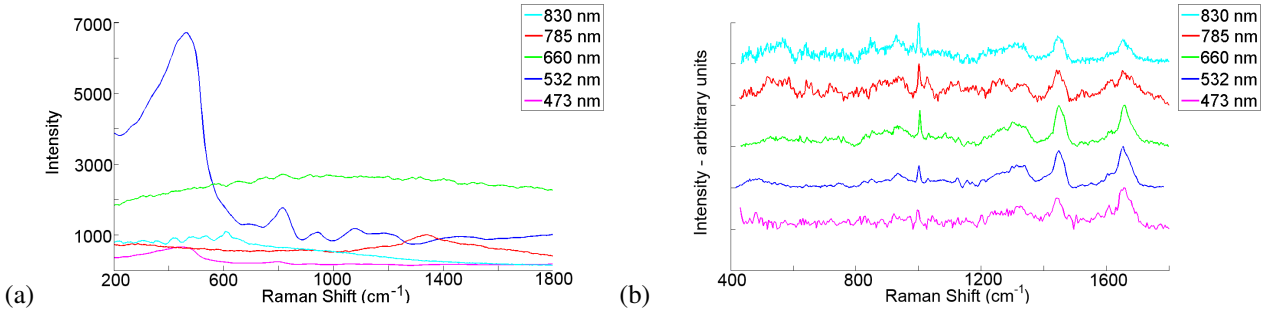


Fig. 9 (a) Background signal of **1500 nm aluminium thin film on glass** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on 1500 nm aluminium thin film for each of the above wavelengths, with an acquisition time of 30 s (x2).

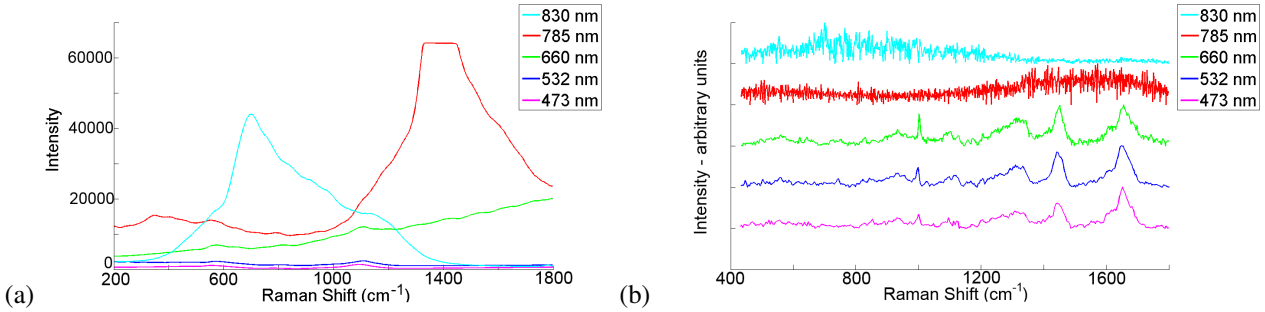


Fig. 10 (a) Background signal of **glass** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on glass for each of the above wavelengths, with an acquisition time of 30 s (x2). Note, the recorded background signal at 785 nm saturated in the 1400 cm^{-1} region.

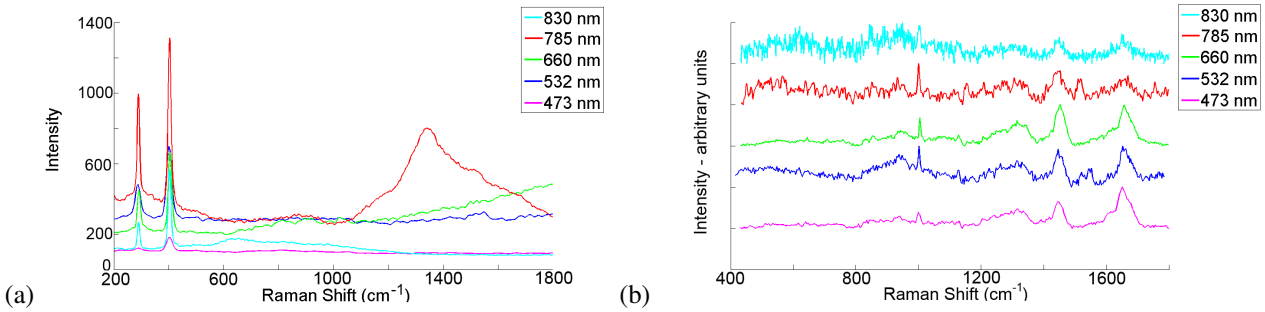


Fig. 11 (a) Background signal of **magnesium fluoride** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on magnesium fluoride for each of the above wavelengths, with an acquisition time of 30 s (x2).

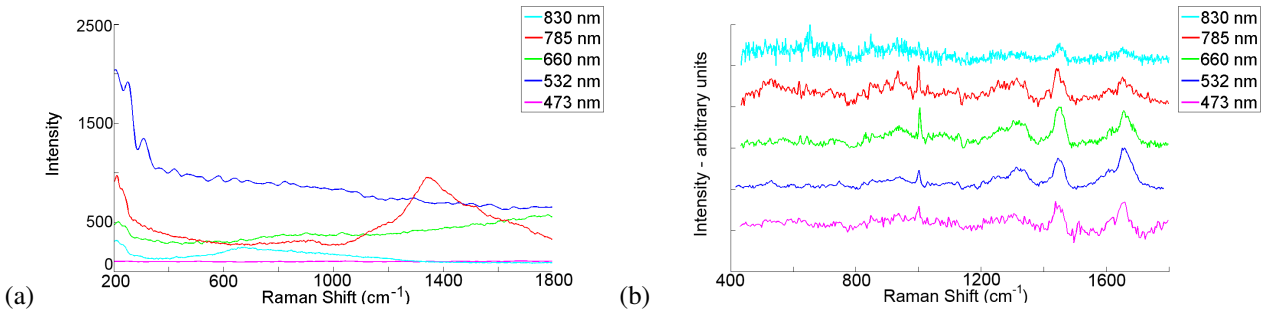


Fig. 12 (a) Background signal of **potassium bromide** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on potassium bromide for each of the above wavelengths, with an acquisition time of 30 s (x2).

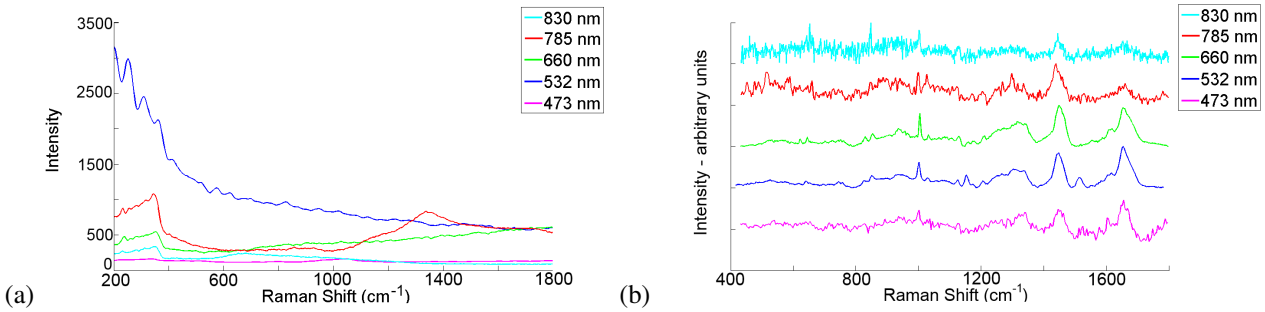


Fig. 13 (a) Background signal of **sodium chloride** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on sodium chloride for each of the above wavelengths, with an acquisition time of 30 s (x2).

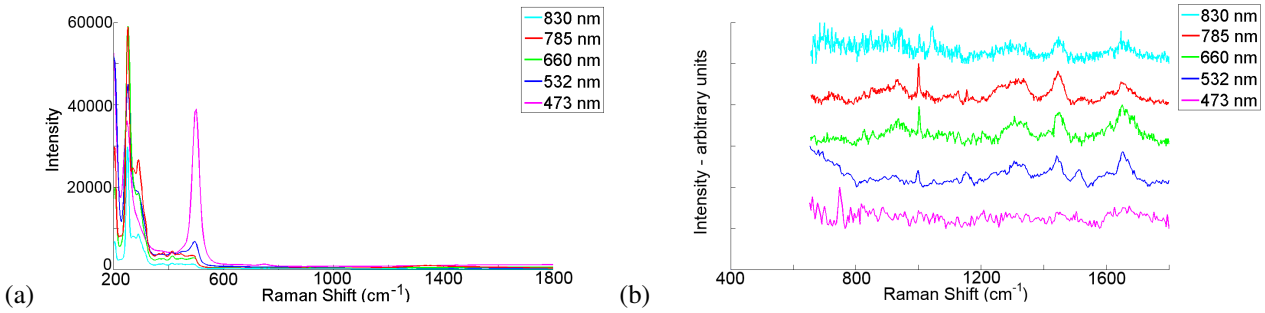


Fig. 14 (a) Background signal of **zinc selenide** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on zinc selenide for each of the above wavelengths, with an acquisition time of 30 s (x2). Note, the cell spectra are only shown between 600–1800 cm^{-1} due to the large background contribution in the lower wavenumber region.

Table 1 Comparison of cross correlation (*xcorr*) results for each substrate at each wavelength when compared to a cell spectrum recorded on a Raman-grade calcium fluoride substrate using a 532 nm laser, where all spectra were recorded with an acquisition time of 30 s (x2).

Substrate	473 nm	532 nm	660 nm	785 nm	830 nm
Raman-grade CaF_2	0.8741	1.000	0.9189	0.8286	0.6939
IR polished CaF_2	0.8726	0.9817	0.8671	0.7882	0.5420
Fused silica	0.8391	0.9422	0.9168	0.7128	0.6218
Aluminium (100 nm)	0.8488	0.9766	0.9211	0.8532	0.7674
Aluminium (1500 nm)	0.8435	0.9854	0.9233	0.8302	0.7837
Glass	0.8473	0.9822	0.9221	0.7469	0.4754
Magnesium fluoride	0.8648	0.5968	0.9403	0.7561	0.7120
Potassium bromide	0.6067	0.9819	0.9112	0.7860	0.6668
Sodium chloride	0.6393	0.9839	0.9354	0.7324	0.7210
Zinc selenide	0.7044	0.3035	0.7643	0.8305	0.4284

Table 2 Comparison of normalised covariance coefficient results for each substrate at each wavelength when compared to a cell spectrum recorded on a Raman-grade calcium fluoride substrate using a 532 nm laser, where all spectra were recorded with an acquisition time of 30 s (x2).

Substrate	473 nm	532 nm	660 nm	785 nm	830 nm
Raman-grade CaF_2	0.9429	1.000	0.8814	0.6767	0.1641
IR polished CaF_2	0.9191	0.9059	0.7117	0.3558	0.0066
Fused silica	0.9004	0.7682	0.9119	0.2423	0.0318
Aluminium (100 nm)	0.8668	0.8868	0.8457	0.6694	0.4999
Aluminium (1500 nm)	0.7390	0.9245	0.9179	0.5582	0.5167
Glass	0.8817	0.9125	0.8856	0.1905	0.0020
Magnesium fluoride	0.9317	0.5953	0.9343	0.3487	0.2117
Potassium bromide	0.5901	0.9235	0.8927	0.4848	0.1249
Sodium chloride	0.5216	0.9148	0.9512	0.2973	0.2646
Zinc selenide	0.2005	0.5840	0.7944	0.6929	0.2173

Table 3 Comparison of mean square error results for each substrate at each wavelength when compared to a cell spectrum recorded on a Raman-grade calcium fluoride substrate using a 532 nm laser, and then compared to the value from glass at 830 nm (i.e. the cell spectrum least similar to that recorded on CaF_2 at 532 nm); where all spectra were recorded with an acquisition time of 30 s (x2).

Substrate	473 nm	532 nm	660 nm	785 nm	830 nm
Raman-grade CaF_2	0.9634	1.0000	0.9817	0.7973	0.4800
IR polished CaF_2	0.9417	0.9434	0.7986	0.6139	0.1820
Fused silica	0.9360	0.8517	0.9437	0.5428	0.3187
Aluminium (100 nm)	0.9185	0.9273	0.9073	0.6726	0.7001
Aluminium (1500 nm)	0.8399	0.9530	0.9443	0.6151	0.7089
Glass	0.9269	0.9473	0.9293	0.5079	0.0000
Magnesium fluoride	0.9584	0.6758	0.9605	0.6083	0.5197
Potassium bromide	0.6067	0.9536	0.9221	0.6756	0.4489
Sodium chloride	0.7313	0.9540	0.9703	0.5790	0.5595
Zinc selenide	0.3527	0.6398	0.8457	0.7849	0.4377

photons than transmissive substrates; this results in the source laser passing through the sample twice, effectively doubling the source power, as well as reflecting all forward scattered photons back towards the microscope objective which would otherwise be lost for a transmissive substrate. This effect is true for a transmissive sample, such as an epithelial cell, but may not be valid for a thicker tissue sample that appears to be opaque. In the NIR, the Raman scattering efficiency is significantly lower than in the visible region (see Equation 2) resulting in up to ten times less scattered photons. In addition, the quantum efficiency of most cameras is significantly reduced in the NIR and is often less than half that for visible wavelengths. The overall effect is that a detected Raman photon is significantly scarcer in the NIR than in the visible range. The quadrupling of this number achieved using an aluminium substrate is therefore a significant advantage over calcium fluoride and other transmissive substrates in this region. We note that the same effect is true for other source wavelengths with the caveat that the effect of doubling the source power may have the effect of burning the sample.

For specific source wavelengths, other substrates performed as well as Raman-grade calcium fluoride and aluminium. In the lower wavelength regions (473 nm, 532 nm and 660 nm), IR polished calcium fluoride, fused silica and glass produced good biological Raman spectra as shown in Figure 6, Figure 7 and Figure 10. However, the background signals produced by these substrates in the higher wavelength regions swamped the Raman peaks generated from the cells. On the other hand, zinc selenide performed poorly in the low wavelength region, but generally improved with increasing wavelength and even proved to be the best substrate for measurements with a 785 nm laser, as seen in Figure 14 and Tables 1–3, thus making it a good choice of substrate for the red and NIR region.

Other factors must also be considered when choosing a sample substrate such as cost, reusability and biocompatibility. Calcium fluoride, and in particular Raman-grade calcium fluoride, is the most expensive of all of the substrates in this study at approximately 150-200 euro for each slide, but it is biocompatible and produces good biological Raman spectra throughout the full range of laser wavelengths applied in this study. Potassium bromide and sodium chloride are a much more cost effective solution to calcium fluoride (with associated spectra shown in Figure 12 and Figure 13), however both of these substrates are soluble in water, making it more difficult to clean and re-use these substrates for multiple studies as is desired in biological research. Magnesium fluoride (Figure 11) is more expensive than potassium bromide and sodium chloride, yet it is cheaper than calcium fluoride, it has the ability to produce good Raman spectra throughout the range of source wavelengths, and it is biocompatible and reusable. Glass is the most widely available and cheapest substrate but unfortunately it is not ideal for the recording of spectra in the higher wavelength region where the biological window exists. However, for studies in the lower source wavelength region, glass has the ability to produce results comparable to that from calcium fluoride, as shown in Tables 1–3. Similarly, IR grade calcium fluoride and fused silica work better in the lower wavelength region, as shown in Figure 6 and Figure 7, but they are more expensive than glass. Finally, zinc selenide is a significantly cheaper alternative to the other substrates for Raman spectroscopy in the NIR region. However it is not biocompatible, although there may be potential to functionalise this substrate with a 'Raman-friendly' biofunctional layer that would improve biocompatibility without compromising the spectral quality. [16,17]

A key point of interest is that while the Raman signals generated from cells on a glass substrate in the NIR region were completely swamped by background fluorescence, by simply depositing a thin film of aluminium onto the glass substrate, it is possible to effectively eliminate the background signal emitted by the glass. We believe that this is due to both the scattering properties of aluminium and the reflective surface created by the aluminium which prevents photons from reaching and interacting with the glass substrate. It is also possible that the reflective nature of the aluminium effectively quadruples the number of scattered photons collected, whereby the source laser takes two passes through the sample and both forward and backward scattered photons are propagated towards the confocal aperture. This may account for the clearly higher quality spectra recorded at 830 nm when compared to other substrates that are suitable for the NIR region, see Figure 8 and Figure 9. This feature will be valid so long as the biological sample is thin and is placed directly on the aluminium surface. For thicker samples, we can expect that the collection of forward scattered photons by the confocal aperture will not occur. Whilst the cost of a single substrate consisting of a thin film of aluminium on glass (the slides used in this study were approximately 7 USD each) is higher than glass alone, mass production of these substrates could reduce this cost substantially.

In terms of source wavelength, there was a greater signal-to-noise ratio in the lower wavelength regions, with spectra in the NIR being noisy for the integration time of 30 s used in this study, as shown in Figure 5 to Figure 14. However, this noise can be reduced by using a longer integration time or using a CCD detector with a higher quantum efficiency in this region. Spectra in the NIR region have a higher resolution than those from lower wavelength lasers, but resolution for the lower laser wavelengths can easily be increased using a grating with higher line numbers than that used here. NIR lasers are also better suited for biological analysis due to the optical window in tissues, but larger background signals, from the sample substrate and the optical elements in the system,

are present in this region. Therefore, there must be a compromise when deciding upon a laser for Raman spectroscopic analysis of biological specimen; one must choose between a laser that results in lower background signals but is more likely to cause biomolecular damage to the sample, or a laser that is biologically-friendly but which is more likely to produce large background signals.

Another important consideration for the source wavelength is the associated power output. In this study, different powers were used for each source wavelength (details given in Section 3). These powers were chosen as they are the typical values employed in commercial Raman systems for biological measurements. While the Raman signal strength is directly proportional to the source power, and higher powers are therefore more desirable, the associated optical and thermal responses of the biological sample need to be taken into consideration. The laser power density is particularly important here, where one needs to maximise the source signal delivered to the sample without actually burning it. Thus, it can be concluded that different source powers could improve the Raman spectra obtained across different sample substrates and source wavelengths, but that has not been investigated here.

Overall, it has been shown that there exists a high degree of variability across sample substrates and laser wavelengths for the analysis of biological specimen using Raman spectroscopy. This high degree of variability means that there is not one single substrate and one source laser to suit all applications of Raman spectroscopy in the analysis of biological specimen. Here we have discussed which substrates provide the optimum Raman spectra for each source wavelength region. However, to find the true optimum combination, we must consider the cost effectiveness and biocompatibility of the substrates in combination with a source laser within the biological window to prevent photodegradation of the samples which also has a low background contribution. In this paper, we chose the 'benchmark' spectrum, against which the quality of all other spectra were compared, to have come from a Raman-grade calcium fluoride substrate with a source wavelength of 532 nm. However these substrates are very expensive and source lasers in this region are known to result in biomolecular damage within the sample. Therefore, with these considerations, it can be concluded that the use of aluminium coated glass substrates with an NIR laser provides the optimum Raman spectra for everyday clinical studies, being significantly cheaper than most substrates, biocompatible, and resulting in good quality spectra.

With this study, we have also shown that for certain sample substrates and source wavelength combinations that result in relatively strong background signals, it is still possible to obtain a high quality cell spectrum that is very similar to the benchmark spectrum obtained from a cell on Raman-grade calcium fluoride using a 532 nm source. Following background subtraction, the spectra show a strong likeness to the benchmark spectrum, as verified by the comparison based metric values close to 1.0 in Tables 1–3. This suggests that it may be possible for a direct comparison of spectra obtained from different research groups operating under different experimental conditions. It is still unclear whether such spectra could be shared for the purpose of precise biochemical comparisons or for studies involving multivariate classification. The level of accuracy required by such studies may be too stringent to establish if this is feasible. At the very least, a calibration of the spectra to take into account the varying quantum efficiencies for the different detectors over the different wavelength bands of interest would be required, as well as interpolation to account for different sampling rates. It is possible that such a standardisation of recording systems could be achieved using a calibration tool that emits a range of wavelengths of known power, such as the calibration accessory currently sold by Kaiser Optical Systems, Inc.

We note that there are clearly limits on how strong the background signal can be, in order for a weaker Raman signal to still be recoverable using background subtraction techniques. The first limit is based on the high dynamic range of measurement of a signal with a very strong background signal, as well as the quantization noise, both of which will be determined by the detector. The fundamental limit on the recovery of a Raman signal from background will be determined by the shot noise associated with the background signal. If the standard deviation of this noise signal is comparable to the amplitude of the Raman scattering, it will be impossible to recover the Raman signal regardless of the properties of the detector.

ACKNOWLEDGEMENTS

This research was conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 11/SIRG/I2140.

References

1 A. Palonpon, J. Ando, H. Yamakoshi, K. Dodo, M. Sodeoka, S. Kawata and K. Fujita, *Nature Protocols*, 2013, **8**, 677–692.

2 L. Kerr, K. Domijan, I. Cullen and B. Hennelly, *Photonics and Lasers in Medicine*, 2014, **3**, 2193–0643.

3 F. Bonnier, A. Mehmood, P. Knief, A. Meade, W. Hornebeck, H. Labkin, K. Flynn, V. McDonagh, C. Healy, T. Lee, F. Lyng and H. Byrne, *Journal of Raman spectroscopy*, 2011, **42**, 888–896.

4 H. Byrne, G. Sockalingum and N. Stone, *Raman Microscopy: Complement or Competitor? In Biomedical Applications of Synchrotron Infrared Microspectroscopy (Moss, D. (ed))*, Royal Society of Chemistry, RCS Analytical Spectroscopy Monographs, No. 11, 2011, pp. 105–142.

5 N. Kourkoumelis, A. Polymeros and M. Tzaphlidou, *Spectroscopy: An International Journal*, 2012, **27**, 441–447.

6 F. Lyng., E. O Faolain, J. Conroy, A. Meade, P. Knief, B. Duffy, M. Hunter, J. Byrne, P. Kelehan and H. Byrne, *Experimental and Molecular Pathology*, 2007, **82**, 121–129.

7 H. Byrne, M. Baranska, G. Puppels, N. Stone, B. Wood, K. Gough, P. Lasch, P. Heraud, J. Sule-Suso and G. Sockalingum, *Analyst*, 2015. DOI: 10.1039/c4an02036g.

8 M. Grimbergen, C. van Swol, R. van Moorselaar, J. Uff, A. Mahadevan-Jansen and N. Stone, *Journal of Photochemistry and Photobiology*, 2009, **95**, 170–176.

9 N. Stone, M. Prieto, P. Crow, J. Uff and A. Ritchie, *Analytical and Bioanalytical Chemistry*, 2007, **387**, 1657–1668.

10 B. De Jong, T. Bakker Schut, K. Maquelin, T. van der Kwast, C. Bangma, D. Kok and G. Puppels, *Analytical Chemistry*, 2006, **78**, 7761–7769.

11 A. Shapiro, O. Gofrit, G. Pizov, J. Cohen and J. Maier, *European Urology*, 2011, **59**, 106–112.

12 M. Grimbergen, C. van Swol, R. Draga, P. van Diest, R. Verdaasdonk, N. Stone and J. Bosch, *Proc. SPIE 7161, Photonic Therapeutics and Diagnostics V*, **7161**, 716114.

13 E. Cancetta, M. Mazilu, A. De Luca, A. Carruthers, K. Dholakia, S. Neilson, H. Sargeant, T. Briscoe, C. Herrington and A. Riches, *Journal of Biomedical Optics*, 2011, **16**, 037002.

14 M. Prieto, P. Matousek, M. Towrie, A. Parker, M. Wright, A. Ritchie and N. Stone, *Journal of Biomedical Optics*, 2005, **10**, 044006.

15 M. Mariani, P. Lampen, J. Popp, B. Wood and V. Deckert, *Analyst*, 2009, **134**, 1154–1161.

16 F. Bonnier, P. Knief, A. Meade, J. Dorney, K. Bhattacharya, F. Lyng and H. Byrne, *Clinical and Biomedical Spectroscopy and Imaging II. Proc. SPIE*, 2011.

17 A. Meade, F. Lyng, P. Knief and H. Byrne, *Analytical and Bioanalytical Chemistry*, 2006, **5**, 1717–1728.

18 F. Bonnier, S. Ali, P. Knief, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. Lee, F. Lyng and H. Byrne, *Vibrational Spectroscopy*, 2012, **61**, 121–132.

19 T. Dieing, O. Hollricher and J. Toporski, *Confocal Raman microscopy*, Springer Science and Business Media, 2011.

20 E. Smith and G. Dent, *Modern Raman Spectoscopy: A Practical Approach*, Wiley, 2004.

21 K. Svoboda and S. Block, *Annual Review of Biophysics and Biomolecular Structure*, 1994, **23**, 247–285.

22 J. Lerner, *Cytometry, Part A*, 2006, **69**, 712–734.

23 R. Li, D. Verreault, A. Payne, C. Hitchcock, S. Povoski, E. Martin and H. Allen, *Journal of Raman spectroscopy*, 2014.

24 N. Afseth, V. Segtnan and J. Wold, *Applied Spectroscopy*, 2006, **60**, 1358–1367.

25 K. Liland, T. Almoy and B. Mevik, *Applied Spectroscopy*, 2010, **64**, 1007–1016.

26 P. Lasch, *Chemometrics and Intelligent Laboratory Systems*, 2012, **117**, 100–114.

27 B. Beier and A. Berger, *Analyst*, 2009, **134**, 1198–1202.

28 H. Takeuchi, S. Hashimoto and I. Harada, *Applied Spectroscopy*, 1993, **47**, 129–131.

29 U. Cappel, I. Bell and L. Pickard, *Applied Spectroscopy*, 2010, **64**, 195–200.

30 N. Stone, C. Kendall, N. Shepherd, P. Crow and H. Barr, *Journal of Raman Spectroscopy*, 2002, **33**, 564–573.