Quantifying the Concentration of Glucose, Urea, and Lactic Acid in Mixture by Confocal Raman Microscopy

Zhengyuan Tang
National University of Ireland, Maynooth

Sinead Barton
National University of Ireland Maynooth

Thomas Ward
Dublin City University

See next page for additional authors

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

Part of the Medicine and Health Sciences Commons

Recommended Citation

This Article is brought to you for free and open access by the NanoLab 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, vera.kilshaw@tudublin.ie.

This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 4.0 International License.
Quantifying the concentration of glucose, urea, and lactic acid in aqueous mixtures by confocal Raman microscopy

Zhengyuan Tang\textsuperscript{a}, Sinead J. Barton\textsuperscript{a}, Tomas Ward\textsuperscript{b}, John P. Lowry\textsuperscript{c}, Michelle M. Doran\textsuperscript{c}, Hugh J. Byrne\textsuperscript{d}, and Bryan M. Hennelly\textsuperscript{a,e}

\textsuperscript{a}Department of Electronic Engineering, Maynooth University, Maynooth, Co. Kildare, Ireland.  
\textsuperscript{b}School of Computing, Dublin City University, Dublin, Ireland  
\textsuperscript{c}Department of Chemistry, Maynooth University, Maynooth, Co. Kildare, Ireland.  
\textsuperscript{d}FOCAS Research Institute, Dublin Institute of Technology, Kevin St, Dublin, Ireland  
\textsuperscript{e}Department of Computer Science, Maynooth University, Maynooth, Co. Kildare, Ireland.

ABSTRACT

Raman spectroscopy has numerous applications in the field of biology. One such application is the simultaneously measurement of the concentration of multiple biochemical components in low volume aqueous mixtures, for example, a single drop of blood serum. Over twenty years ago, it was shown for the first time that it was possible to estimate the concentration of glucose, urea, and lactic acid in mixture by combining Raman Spectroscopy with Partial Least Squares Regression analysis. This was followed by numerous contributions in the literature designed to increase the number of components and reduce the limits of concentration that could be simultaneously measured using Raman spectroscopy, by developing various optical architectures to maximise the signal to noise ratio. The aim of this paper is to demonstrate the potential of a confocal Raman microscopy system for multicomponent analysis for the case of physiologically relevant mixtures of glucose, urea, and lactic acid.

Keywords: Raman Spectroscopy, Partial Least Squares Regression, Confocal microscopy, Multiple component analysis

1. INTRODUCTION

The quantification of multiple-components in liquid mixture is of interest in the field of biology, due to the abundant physiological information that can be found in body fluids such as blood and urine. Numerous diseases result in abnormal concentrations of various chemicals in body fluids. The most obvious example is the high blood glucose concentration for diabetes patients. Kidney disease is another example, which is often characterized by high serum creatinine levels and high urine albumin levels.\textsuperscript{1–3}

An accurate multicomponent measurement is important for disease diagnosis. Well established approaches require relatively large sample volumes and require lengthy delays, as well as the need for chemical reagents.\textsuperscript{4,5} In contrast, Raman spectroscopy has been proposed to be a qualitative and quantitative multicomponent predictive technique with the advantages of short processing duration, the need for small sample volumes, and the potential for non-invasive application.\textsuperscript{6}

Raman spectroscopy is based on detecting the vibrations of molecules via the scattering of photons associated with these vibrations. A monochromatic laser source is used to probe the vibrational and rotational states of a molecule, and the scattered photons are collected by a spectrometer using a low noise camera. The Raman spectrum often contains peaks that can be used to identify the presence of a specific chemical as well as the concentration. Following on from the initial experiment of Gotez\textsuperscript{7} to measure the concentrations of multicomponents aqueous mixtures, many researchers have designed experiments to reduce the predictive error of the technique, which allows for increasing in the number of chemical components that can be detected in

Further author information: (Send correspondence to Bryan Hennelly)  
Bryan Hennelly: E-mail: bryanh@cs.nuim.ie, Telephone: +353 1 708 3338
the mixture as well as reducing the smallest measurable concentrations of each component; this is achieved by increasing the signal to noise ratio of the system. A description of Raman spectroscopy, and a brief review of the background research in multicomponent analysis is provided in Section 2.

In many cases, Raman systems that have been optimized for multicomponent analysis are often no longer useful for many other life science applications; the most obvious example is the liquid core optical fiber technique, in which the spectrum is collected from an aqueous sample that has been pumped into an optical fiber. While this set-up has demonstrated the best results for multicomponent analysis to date, the system is not readily available for the analysis of cell or tissue samples. The most common form of Raman spectrometer found in life science and material science laboratories is based on the confocal architecture. The confocal Raman microscopy system can effectively reduce the unnecessary stray scattering from optical elements and from the sample substrate by inserting a pinhole between the sample and the spectrometer. In Section 3, the set-up of a confocal Raman microscopy system and the concentration estimation methods that we used in this study are discussed. The predictive results are provided in Section 4 and a brief discussion is offered in Section 5.

2. A LITERATURE REVIEW OF RAMAN MULTICOMPONENT ANALYSIS

2.1 Raman spectroscopy

Raman spectroscopy is based on the inelastic scattering of monochromatic light, and is a powerful technique that can be used to identify specific chemical bonds. Scattering of photons occurs when incident photons and chemical molecules interact. Most of scattered photons result from Rayleigh scattering, a form of elastic scattering, and have the same wavelength as the excitation laser. A small percentage of the scattered photons result from an inelastic scattering process known as Raman scattering. Although relatively few in number when compared with Rayleigh scattered photons, the wavelength and intensity of these photons carry qualitative and quantitative information on the presence of a specific chemical bond. A set of peaks appearing at the specific set of wavenumber positions in a given Raman spectrum can be described as a ‘fingerprint’ for identifying a specific chemical, and simultaneously the height of the peaks can be related to the concentration of this chemical.

2.2 A review of Raman multicomponent analysis

Multicomponent analysis is one application of Raman spectroscopy. Over the past two decades, a number of research groups have proposed optical Raman setups that are specifically designed to improve the capability of the technique for measuring multicomponent concentrations. These systems are specifically designed to decrease the error of the overall approach, which in turn allows for increasing the number of analytes in the mixture under investigation, as well as reducing the limit of concentration that can be measured for a given chemical.

In the first experiment of its kind, Gotez et al. quantified the glucose, urea and lactic acid in aqueous solution using a basic Raman setup as shown in Fig 1(a), whereby an Argon-ion laser with a wavelength of 514.5 nm and 10 mW power was directed at the sample container. The scattered light was collected at 90° relative to the incident optical path. A notch filter was used in the collection path in order to remove Rayleigh scattered photons, the Raman scattered photons were transmitted to a spectrometer with a 1200 lines/mm grating. Four spectra of 10s integration time were recorded and averaged together. This basic Raman setup also was used in a number of later experiments.

One year later, Berger et.al. used a NIR Raman spectrometer, which has advantages in terms of sample penetration as well as a reduced background signal, to predict the concentration of glucose, lactic acid and creatinine in saline solution. The authors used an optical fiber bundle to facilitate the delivery of more photons to the entrance slit of the spectrograph from a large area on the sample, without sacrificing spectral resolution. The setup for this experiment was depicted in Fig 1(b); the excitation source used in this experiment was a 830 Argon-ion laser pumped dye laser providing 200 mW. The backscattered photons pass through the dichroic beamsplitter and were collected by the fiber bundle. The exposure time of recording a spectrum in the experiment is 100s.

Following these experiments, it was proposed that a paraboloidal mirror could be used in the collection path to further increase the number of Raman scattered photons that could be recorded. A Raman system equipped
Figure 1. Raman spectroscopy systems for Raman multicomponent analysis that have previously been proposed: (a) A basic Raman system similar: LPF, line pass filter; NF, notch filter; M, mirror; (b) Raman system with optical fiber bundle and NIR excitation: BPF, bandpass filter; L, lens; OF1, optical fiber; OF2, fiber bundle; DB, dichroic beamsplitter filter; (c) Raman setup using paraboloidal mirror; PM, paraboloidal mirror; (d) A typical LCOF Raman system; PM, power meter; EF, edge filter; FB, fiber bundle.

with a paraboloidal mirror with high numerical aperture and f-number is shown in Fig 1(c). This type of Raman system has been shown to effectively measure the blood analytes concentration by a number of different research groups.\textsuperscript{12, 20, 21}

Another powerful approach for Raman multicomponent analysis is to use a liquid core optical fiber (LCOF). This approach is capable of significantly enhancing the signal-to-noise ratio (SNR) of the collected spectrum, resulting from a significantly larger collection volume, by injecting the sample into a LCOF instead of traditional sample container. A typical LCOF Raman setup is shown in Fig 1(d). When using the LCOF technique, it is important to take into account the attenuation and absorption of the collected spectrum based on the Beer-Lambert law. This is achieved by recording a white light reference spectrum, from which the wavelength-dependent coefficient of light loss in the LCOF can be calculated. Several publications from Qi and Berger have demonstrated the feasibility and advantages of LCOF Raman system, which can obtain a high quality prediction result within a few seconds.\textsuperscript{22–25}

3. EXPERIMENT DESIGN

3.1 Confocal Raman microscopy

Compared with all of the set-ups mentioned in Section 2 for which the application range is essentially limited to recording Raman spectra from bulk (aqueous) samples, confocal Raman microscopy provides for a diversity in application including neurophysiology, cytology, material science, etc.\textsuperscript{26–28} The simple design includes a pinhole between the sample and the detector, the benefit of which is to reduce the unnecessary background scattering.
from optical elements and from the substrate, as well as to improve the three dimensional spatial resolution to approximately one micrometer in all three dimensions. In this paper, we repeat Goetz’s experiment to measure the concentrations of glucose, urea, and lactic acid in aqueous solution using a confocal Raman microscopy system, and we evaluate the performance of this approach for multicomponent analysis.

The confocal Raman microscopy system used in this research is shown in Fig 2. A laser with 532nm wavelength and 100 mW is first passed through a line pass filter and neutral density filter, in order to reduce unwanted laser emissions and adjust the light power. Two lenses are used to control the size of the laser beam, which is directed into the back aperture of a microscope objective (MO) with 10x magnification and 0.3 numerical aperture. The MO focuses the laser onto the sample container containing the mixture, which has a calcium fluoride substrate base that produces a negligible background spectrum. The backscattered photons are collected by the MO and pass through a dichroic beamsplitter and a notch filter in order to reduce the Rayleigh scattered photons and are imaged through a confocal pinhole of size 200 µm. The photons are directed into the spectrograph, using a grating of 600 lines/mm of grating and a liquid nitrogen cooled CCD (Andor Idus) camera operating at -80°C.

3.2 Sample preparation

Glucose, urea and lactic acid were obtained from Sigma Aldrich and mixed together with selected weight ratios and dissolved in deionized water. Twenty two sets of mixtures with different concentrations are used in the experiment that follows. The concentration range of glucose, lactic acid and urea are 150-800mg/dL, 54-226mg/dL, and 51-253mg/dL respectively, and the concentration of each chemical was varied randomly in all of the mixtures. These concentrations are similar to those used in the original Raman multicomponent experiment. It must be noted that lactic acid can be dissociated in different pH environments, resulting in a change in concentration in the mixture. The formula for calculating the actual concentration of lactic acid is shown below,

$$C_{la} = (1 - 1.38 \times 10^{(-4+PH)}) \times C_{ex}$$

where $C_{ex}$ is the concentration of lactic acid used to create the mixture and $C_{la}$ is the measurable concentration in the mixture following dissociation. Three Raman spectra with 20s integration time are recorded for each mixture solution, and then averaged together to produce the final Raman spectrum for each mixture.
3.3 Experimental method

3.3.1 Calibration

Calibration of the Raman spectrometer is a necessary first step in order to correct any unwanted shift in wavenumber. Wavelength miscalibration can result from the minute displacement of various optical elements as well as from small deviation in the rotation angle of the grating. Wavelength calibration is achieved by using a Neon lamp, which produces a set of sharp peaks at well defined wavelengths. The wavenumber axis can also be miscalibrated by variation in temperature. Wavenumber calibration is performed by recording the Raman spectrum from a reference silicon sample, which contains a strong sharp peak at 520 cm\(^{-1}\).

3.3.2 Pre-processing methods

Cosmic ray artifacts are sharp random peaks that can be present in a recorded spectrum due to high energy particles impacting the CCD; this is most likely to occur during long exposure times. A simple solution is to record several spectra with the same exposure time (three spectra are recorded in our experiment), which are then averaged. Any sudden sharp peaks can be easily identified in one of these spectra by comparing with the other spectra; these peaks are omitted from the averaging process.

Dark current noise and read noise are two noise sources that originate in the detector and contribute to noise in the recorded spectrum. Dark current results from random electrons in the detector pixels and can be effectively reduced to small values in low temperatures. Read noise is unavoidable but it has a relatively small standard deviation when compared with shot noise from the Raman spectrum, which is the most significant source of noise in our experiments. Shot noise results from the random arrival of collected photons and is a critical problem when recording weak signals. Raman spectra recorded from low concentration mixtures result in low signal to noise ratios (SNR) due to the relatively large shot noise component. In our experiment, the SNR is increased by using long exposure times. There is a limit to what can be achieved using this approach since long exposure times results in increasing levels of dark current noise. Savizky-Golay smoothing is applied in order to reduce the overall noise in the spectrum; however, it has the adverse affect of smoothing spectral peaks.

Aside from the various noise sources, the raw spectrum recorded from the CCD consists of some other unwanted components such as a slowly varying background signal resulting from scattering and/or photoluminescence from the sample, and the Raman spectrum from optical elements as well as the substrate and the microscope objective. The enhanced multiplicative scatter correction (EMSC)\(^{29}\) algorithm can help to reduce these unwanted Raman signals and recover the Raman spectrum of the mixture in isolation, which can then be used with partial least squares regression (PLSR) for multicomponent analysis. The EMSC algorithm assumes the raw spectrum is constituted by (1) reference spectra from each chemical component in the mixture, (2) the combined background Raman spectrum from the optical system, the substrate as well as from the water in the mixture, which is recorded using a sample that contains no chemicals, (3) a polynomial to describe the unwanted baseline signal. The algorithm uses ordinary least squares to find the optimal weights of each of these different spectral components. Finally, the Raman spectrum of the chemical mixture is calculated by subtracting the weighted copies of (2) and (3).

3.4 Partial least squares regression

PLSR is a multivariate statistical method that is commonly used in Raman multicomponent analysis. All of the experiments discussed in Section 2 exploited PLSR. In PLSR, the dependent variables \(Y\) (the concentrations of the various multicomponents) and the independent variables \(X\) (a matrix of partial Raman spectra, which includes only areas that contain peaks) can be projected to a new space by decomposing \(X\) and \(Y\), respectively, into two matrices of scores \(T\) and \(U\), and two matrices of loadings \(P\) and \(Q\):

\[
X = TP^T \\
Y = UQ^T
\]  

(2)

This decomposition will maximise the covariance between \(T\) and \(U\), thereby enhancing the relationship between \(X\) and \(Y\). The \(X\) loadings matrix \((P)\) represents the features of the Raman spectrum that are used in the
prediction. Only a small number of loadings needs be retained. An optimal choice of the number of loadings and their scores will minimise the effect of noise, and optimize the predictive model between $X$ and $Y$. The Root-Mean-Square-Error (RMSE) is a metric that is commonly used to evaluate the prediction by comparing the difference between the actual concentrations used in the preparing mixtures and predicted concentrations.

4. RESULTS

Following the removal of cosmic rays, all spectra are processed using Savitzky-Golay smoothing with a fifth order polynomial and a window size of nine samples. In Fig 3(a) a raw sample of one sample mixture is shown before and after smoothing. The same smoothing was applied to all 22 raw spectra, followed by application of the EMSC algorithm. The EMSC algorithm requires the recording of each of the background spectra, including high concentrations of each of the chemicals used in the mixture. The spectrum of 5000 mg/dL urea, 5000 mg/dL lactic acid and 7500 mg/dL glucose were recorded for this purpose and these spectra are shown in Fig 3(b). The combined background Raman spectrum that is used in the EMSC algorithm is obtained by recording the spectrum of a pure water sample in the same container and under the same recording conditions; this spectrum contains all of the unwanted background signals. The baseline that is calculated by the EMSC algorithm is shown in Fig 3(c). Following this, a leave one out cross validation method is exploited and steps are taken to avoid overfitting in the PLSR algorithm. The relationship between the PLS loadings and the RMSE is illustrated in the Fig 3(d). In order to minimise the effect of noise and to avoid overfitting, five loadings are chosen to predict the concentration of each chemical in the 22 mixtures. The RMSE value is 25.1 mg/dL, 19.5 mg/dL, 15.7 mg/dL for glucose, lactic acid, and urea, respectively. The relevant predictive results from the previous studies that are mentioned in Section 2 are tabulated together with our results in Table 1.
Table 1. our predictive result and Partial results from previous researches for comparison

<table>
<thead>
<tr>
<th>Analytes</th>
<th>RMSE (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>22.95</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>17.5</td>
</tr>
<tr>
<td>Urea</td>
<td>15.7</td>
</tr>
</tbody>
</table>

Goetz et. al.’s result from a basic Raman spectroscopy (64s integration time)

<table>
<thead>
<tr>
<th>Analytes</th>
<th>RMSE (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>20.17</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>12.92</td>
</tr>
<tr>
<td>Urea</td>
<td>19.07</td>
</tr>
</tbody>
</table>

Berger et. al.’s result from a fiber bundle Raman spectroscopy (100s integration time)

<table>
<thead>
<tr>
<th>Analytes</th>
<th>RMSE (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>21.62</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>11.71</td>
</tr>
<tr>
<td>Creatinine</td>
<td>13.57</td>
</tr>
</tbody>
</table>

Annika et. al.’s result from a Raman spectroscopy using paraboloidal mirror (5 min integration time)

<table>
<thead>
<tr>
<th>Analytes</th>
<th>RMSE (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>21</td>
</tr>
<tr>
<td>Urea</td>
<td>9</td>
</tr>
</tbody>
</table>

Qi et. al.’s result from a LCOF Raman spectroscopy (64s integration time for creatinine) (150s integration time for glucose)

<table>
<thead>
<tr>
<th>Analytes</th>
<th>RMSE (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>8.8</td>
</tr>
<tr>
<td>Creatinine</td>
<td>4.3</td>
</tr>
</tbody>
</table>

5. DISCUSSION

In this paper, we briefly review the development of Raman multicomponent analysis by discussing the Raman setups that were specifically designed by other researchers. However, optical setups that are designed for multicomponent analysis are usually limited in terms of their application. In contrast, the confocal Raman microscopy system is ubiquitous in terms of application, and is commonly found in life science laboratories. We have validated the predictive ability of this type of system for multicomponent analysis using a simple proof of concept experiment, similar in design to the original work on Raman multicomponent analysis. It has been demonstrated in this paper that it is possible to use a standard confocal Raman microscopy system to perform multicomponent analysis. The limitation of confocal Raman microscopy is that the pinhole not only prevents most of unwanted signal from the optical element and substrate, but also significantly reduces the required sample signal, and therefore reduces the SNR, which can only be increased by significantly longer acquisition times when compared with LCOF.

It must be noted that the LCOF Raman system is optimal for multicomponent analysis when compared with all other system designs and requires the shortest acquisition times. Despite the high performance of the LCOF approach, the overall performance is fundamentally limited by the shot noise of the water in the sample. Despite being a weak scatter, the high concentration of water results is a relatively strong, albeit smooth, spectrum. This can be subtracted using EMSC but the shot noise can never be removed.

In future work, we intend to investigate the role of the microscope objective in the performance of multicomponent
ACKNOWLEDGMENTS

This research was conducted with the financial support of Science Foundation Ireland (SFI) under Grant number 15/CDA/3667.

REFERENCES


