The Development of Enhanced Raman Scattering for the Trace Analysis of Biomolecules

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2013

David Paul Cowcher

Supervised by Professor Royston Goodacre

School of Chemistry
## Contents

CONTENTS ........................................................................................................................................... 2

LIST OF FIGURES .............................................................................................................................. 7

LIST OF TABLES ............................................................................................................................. 14

LIST OF EQUATIONS ...................................................................................................................... 15

LIST OF SCHEMES ........................................................................................................................ 15

LIST OF ABBREVIATIONS .............................................................................................................. 16

ABSTRACT .......................................................................................................................................... 18

ACKNOWLEDGEMENTS .................................................................................................................. 19

DECLARATION ................................................................................................................................. 20

COPYRIGHT STATEMENT ............................................................................................................. 21

PREFACE ............................................................................................................................................ 22

1. INTRODUCTION ........................................................................................................................ 24

1.1. ELECTROMAGNETIC RADIATION AND WAVE-PARTICLE DUALITY ....................................... 25

1.2. MOLECULAR ENERGY LEVELS ................................................................................................ 28

1.3. RAMAN SPECTROSCOPY ....................................................................................................... 28

1.3.1. Comparing Raman with Infrared (IR) Spectroscopy ............................................................ 32

1.3.2. Fluorescence interference ..................................................................................................... 33

1.3.3. Raman Instrumentation ....................................................................................................... 33

1.3.4. Raman microscopy ................................................................................................................ 35

1.3.5. Portable Raman and Liquid Sample Analysis ....................................................................... 36

1.3.6. A Typical Raman Spectrum .................................................................................................. 37

1.3.7. Enhancing the Raman Signal ............................................................................................... 38

1.3.7.1. Resonance Raman scattering (RRS) ................................................................................ 38
### Contents

1.3.7.2. Surface-enhanced (resonance) Raman scattering (SER(R)S) .................. 39

1.4. SURFACE-ENHANCED AND TIP-ENHANCED RAMAN SCATTERING (SERS & TERS) ...... 40

1.4.1. Electromagnetic Theory ........................................................................ 40

1.4.2. Chemical Enhancement Theory .............................................................. 41

1.4.3. SERS Surfaces ......................................................................................... 43

1.4.3.1. Nanoparticles .................................................................................... 44

1.4.3.2. Patterned surfaces .............................................................................. 46

1.4.3.3. Non-metal SERS surfaces .................................................................. 47

1.4.4. Tip-enhanced Raman scattering (TERS) ................................................ 47

1.5. ANALYSIS OF RAMAN DATA .................................................................. 49

1.5.1. Preprocessing ......................................................................................... 50

1.5.2. Principal Components Analysis (PCA) .................................................. 51

1.5.3. Multivariate Regression Analysis ............................................................ 52

1.6. AIM OF THE THESIS .............................................................................. 52

1.7. REFERENCES ........................................................................................... 55

2. PORTABLE, QUANTITATIVE DETECTION OF *BACILLUS* BACTERIAL SPORES

USING SURFACE-ENHANCED RAMAN SCATTERING ........................................ 60

2.1. DECLARATION ............................................................................................ 61

2.2. ABSTRACT .................................................................................................. 62

2.3. INTRODUCTION .......................................................................................... 63

2.4. EXPERIMENTAL ........................................................................................ 66

2.4.1. Colloid ..................................................................................................... 66

2.4.2. Instrumentation ....................................................................................... 66

2.4.3. SERS Analysis of Pure DPA .................................................................. 66

2.4.4. Culture of Bacterial Spores ................................................................... 67

2.4.5. DPA Extraction and Analysis ................................................................. 68

2.4.6. Data Processing ...................................................................................... 68
2.5. RESULTS AND DISCUSSION ................................................................................. 69

2.5.1. Univariate Analysis of Pure DPA ...................................................................... 72

2.5.2. Univariate Analysis of DPA Extracted from Bacillus Spores ......................... 75

2.5.3. Multivariate Data Analysis for Estimating DPA Concentrations and Spore Counts. 75

2.6. CONCLUSION ..................................................................................................... 78

2.7. REFERENCES ..................................................................................................... 79

2.8. SUPPLEMENTARY INFORMATION .................................................................... 81

3. AN INVESTIGATION OF BORON NITRIDE AS A SUBSTRATE FOR SURFACE-ENHANCED RAMAN SCATTERING AT 244 NM EXCITATION ............................................. 82

3.1. DECLARATION .................................................................................................... 83

3.2. ABSTRACT ......................................................................................................... 84

3.3. INTRODUCTION ................................................................................................. 85

3.4. EXPERIMENTAL ............................................................................................... 87

3.4.1. Reagents ........................................................................................................ 87

3.4.2. Preparation of surfaces and samples: .............................................................. 88

3.4.3. Instrumentation and measurement: ............................................................... 88

3.4.4. Data analysis: ............................................................................................... 89

3.5. RESULTS AND DISCUSSION ......................................................................... 90

3.6. CONCLUSION ................................................................................................... 99

3.7. REFERENCES .................................................................................................... 100

4. THE DETECTION OF PROTEIN GLYCOSYLATION USING TERS ...................... 102

4.1. DECLARATION .................................................................................................. 103

4.2. ABSTRACT ....................................................................................................... 104

4.3. INTRODUCTION ............................................................................................... 105

4.4. EXPERIMENTAL .............................................................................................. 108

4.4.1. Reagents ...................................................................................................... 108

4.4.2. Protein sample preparation ......................................................................... 108
4.4.3. Instrumentation and TERS measurement ........................................ 108
4.4.4. Data analysis .................................................................................. 110
4.5. RESULTS AND DISCUSSION .............................................................. 111
4.6. CONCLUSION ................................................................................... 127
4.7. ACKNOWLEDGEMENTS .................................................................. 128
4.8. REFERENCES ................................................................................... 129
4.9. SUPPLEMENTARY INFORMATION .................................................... 132

5. QUANTITATIVE ON-LINE LC-SERS OF PURINE BASES ...................... 135
5.1. DECLARATION .................................................................................. 136
5.3. ABSTRACT ....................................................................................... 137
5.4. INTRODUCTION ............................................................................... 138
5.5. EXPERIMENTAL ............................................................................... 142
  5.5.1. Reagents and materials ................................................................. 142
  5.5.2. Nanoparticle synthesis ................................................................. 142
  5.5.3. Sample preparation ..................................................................... 143
  5.5.4. Instrumentation .......................................................................... 143
  5.5.5. Sample analysis .......................................................................... 145
  5.5.6. Data analysis .............................................................................. 146
5.6. RESULTS AND DISCUSSION .............................................................. 147
5.7. CONCLUSION .................................................................................. 158
5.8. ACKNOWLEDGEMENTS .................................................................. 158
5.1. REFERENCES ................................................................................... 159
5.2. SUPPLEMENTARY INFORMATION .................................................... 162

6. DISCUSSION ......................................................................................... 171
6.1. DISCUSSION AND FUTURE PERSPECTIVES .................................... 172
6.2. REFERENCES ................................................................................... 179
APPENDIX I: PUBLISHED WORK IN ORIGINAL FORMAT ............................................. 180

APPENDIX II: ADDITIONAL PUBLICATIONS .......................................................... 182

Word count: 40152
List of Figures

**Figure 1.1.** The oscillating electric field (E, red) and magnetic field (B, blue) with propagation through space (λ) that constitutes electromagnetic radiation. ........................................25

**Figure 1.2.** A cartoon representation of the electromagnetic spectrum (Atkins and De Paula, 2006) ..................................................................................................................26

**Figure 1.3.** General molecular energy diagram for the ground (E₀) and first (E₁) electronic energy levels, showing the approximate relative energies of the related vibrational (V) and rotational (R) energy levels. .................................................................29

**Figure 1.4.** Interaction of electromagnetic radiation with a molecule in its electronic ground state that cause a change in vibrational energy level. V₀ represents the molecular vibrational ground state and V₁ represents the first molecular vibrational excited state. Adapted from (Brown, 1998, Lewis and Edwards, 2001) ..........................................................30

**Figure 1.5.** Normal vibrational modes of carbon dioxide: (A) symmetric stretching, (B) asymmetric stretching, (C) bending and (D) bending (perpendicular to (C)). Adapted from (Atkins and De Paula, 2006) .................................................................31

**Figure 1.6.** Schematic representation of the fundamental components of a Raman spectrometer. Green is used in this figure to depict the laser excitation light and red the inelastically scattered light: either Stokes or Anti-Stokes Raman. .................................................................34

**Figure 1.7.** Examples of typical Raman instruments: microscope (left) and portable (right). ....36

**Figure 1.8.** 785 nm Raman spectrum of N-acetylglucosamine (structure inset). Important regions for peak assignments are highlighted (Socrates, 2004) .........................................................37

**Figure 1.9.** Energy level diagram of an analyte molecule adsorbed onto a metal surface. The minimum energy required for an intramolecular electronic excitation of the analyte, as occurs in RRS (A) is larger than the energy required for a charge transfer from analyte to metal (B) or from metal to analyte (C), due to the Fermi level of the metal residing between the HOMO and LUMO of the analyte. This allows the incident light to be resonant with electronic excitation at higher wavelength. Adapted from (Campion and Kambhampati, 1998) ........................................................................................................43
List of Figures

**Figure 1.10.** Schematic representation of a typical TERS setup using an AFM as the scanning probe microscope. .................................................................49

**Figure 2.1.** Baseline-corrected SERS spectra of glutaric acid internal standard, pure DPA, and DPA extracted from *B. cereus* and *B. subtilis* spores using nitric acid. Highlighted are the DPA ring breathing vibration at 1006 cm\(^{-1}\) and the glutaric acid C–H stretch at 2934 cm\(^{-1}\) that are used for quantification. Chemical structures of DPA and glutaric acid are shown.................................................................70

**Figure 2.2.** Time-related plot showing the effect of two aggregating agents on 633 nm SERS signal from DPA. A constant concentration of DPA was used (5 ppm, 29.9 mM) and 30 s SERS signals were collected sequentially after the initiation of aggregation by 20 mM nitric acid and 0.100 M sodium sulphate. The points are the mean areas under the ring-breathing mode at 1006 cm\(^{-1}\) from baseline-corrected spectra, \(n = 3\). Error bars show one standard deviation.................................................................71

**Figure 2.3.** (A) Calibration curve showing the variation of the mean area under the ................................................................73

**Figure 2.4.** Representative example plots of predicted DPA concentration or spore count against known DPA concentrations or spore counts. Three different analysis methods are shown based on univariate assessment (area under 1006 cm\(^{-1}\)), partial least-squares (PLS) and kernel PLS. Solid circles show the training data set (60%) and outline circles shown the test data set (40%).................................................................76

**Figure 2.5.** Area of the 2.0 mM glutaric acid C–H stretch peak that was used as an internal standard. The high standard deviations shown by the error bars (\(n=3\)) and lack of a general trend in the data from 0 - 100 ppb DPA shows that the peak area is not dependent on the concentration of DPA and is therefore a suitable internal standard. Above 200 ppb DPA, however, a decreasing trend in the C–H stretch peak area is observed, along with much lower standard deviations, indicating that the amount of glutaric acid bound to the colloid is decreasing as it becomes saturated with DPA. .................................................................81

**Figure 3.1.** Structures of the hydrochloride salts of (a) rhodamine 6G and (b) crystal violet. 90

**Figure 3.2.** Mean (\(n=3\)) UV Raman spectra of R6G powder (red) and 1.00 x 10\(^{-3}\) M aqueous solution of R6G spotted onto a BN surface (blue) showing a shift in the peak at ~1200 cm\(^{-1}\) as
List of Figures

Figure 3.3. Mean (n=3) UV Raman spectra of (a) BN (blue), chalk (red) and graphite (black) surfaces, and (b) 1.00 x 10^{-4} M R6G solution spotted on the same surfaces. ........................................91

Figure 3.4. Mean Raman spectra (n=3) of crystal violet (a) 1.00 mM on BN, (b) 1.00 mM on chalk, (c) 1.00 mM on graphite and (d) powder. Dashed lines show the spectra of the bare surfaces. The tops of the peak from the BN background signal is cut off to make the crystal violet peaks easier to see. .................................................................94

Figure 3.5. (a) Mean (n=3) Raman spectra of 50.0 mM solutions of the amino acids spotted onto BN surfaces: phenylalanine (solid blue line), tryptophan (dashed red) and tyrosine (dotted-dashed black). The spectrum of the bare BN surface is shown in yellow. The top of the BN peak is cut off in each case to make the amino acid features clearer. (b) Mean (n=3) Raman spectra of the crystalline amino acids on a glass slide: phenylalanine (solid green line), tryptophan (dashed purple) and tyrosine (dot-dashed grey). .................................................................96

Figure 3.6. Repeat UV-Raman analysis of R6G (a) mean spectra (n=3) of 25 mL spots of 1.00 mM solution on BN (solid blue), chalk (dashed red) and graphite (dotted-dashed black); (b) scaled mean spectra (n=3) of R6G powder (solid blue), and 25 μL spots of 1.00 mM R6G solution on BN: original data set (dotted-dashed black) and repeat dataset (dashed red). ...97

Figure 4.1. (a) Raw TERS spectra of RNase A from one grid with examples of cosmic rays highlighted by asterisks. (b) A typical background TERS spectrum with peaks from the silicon tip highlighted with arrows. Also shown are examples of the (c) TERS and (d) Raman spectra of mannose (solid line) and N-acetylglucosamine (dashed line), the sugars that make up the glycan of RNase B. The TERS spectra (c) have had the silicon background removed. Inset shows a cartoon representation of the native form of bovine RNase drawn from atomic coordinates in the PDB (5RSA) using PyMOL. Also shown is the glycosylation point (Asn34 residue) and the RNase B glycan. Optional mannose refers to the variation in number and possible arrangements of mannose between the different glycoforms that occur in RNase B. ...........................................................................................................112

Figure 4.2. Following cosmic ray removal and baseline correction by the asymmetric least squares method, (a) a typical RNase A spectrum, (b) PCA scores plots of all of the RNase A
List of Figures

(Figure 4.3) PCA plots of all data following normalisation against total signal. (A) and (C) scores plots with RNase A in blue circles and RNase B in red squares, with % explained variance in brackets. (B) and (D) loadings plots of PC1 (solid purple) and PC2 (dashed green). (A) and (B) following the normalisation against blank Si and (C) and (D) following the ALS Si-removal route.

(Figure 4.4) Box and whisker plot of the area under the preprocessed TERS spectra in the wavenumber regions indicated that showed differences between the two protein glycoforms using normal Raman spectroscopy. For each region, the shaded box and whisker represents RNase B. In each case the blue box represents the interquartile range (IQR), the red line the median, and the black dashed whiskers the remaining data, with the exception of any data that lie outside the IQR by more than 1.5 x IQR (±2.72), which are shown by red crosses. Vibration assignments for the regions are given in Table 4.1. In this example, the data were preprocessed using the normalisation against blank Si.

(Figure 4.5) Box and whisker plot of the area under the preprocessed TERS spectra in the wavenumber regions did not show differences between the two protein glycoforms using normal Raman spectroscopy. For each region, the shaded box and whisker represents RNase B. In each case the blue box represents the interquartile range (IQR), the red line the median, the black dashed whiskers the remaining data, and any outliers are indicated by the red crosses. In this example, the data were preprocessed using the normalisation against the blank Si method.
List of Figures

**Figure 4.6.** Following all of the preprocessing, (A) shows PCA scores of the area under the curve of 950 – 1000 cm\(^{-1}\) and (B) shows PCA scores of the interquartile range of the same data, with RNase A shown by blue circles, RNase B shown by red squares and % explained variance in brackets. (C) Shows representative spectra from each group in (B) with RNase A in solid blue, RNase B in dashed red and the selected region highlighted in yellow. (D)-(F) show the same information as (A)-(C) but following the ALS Si-removal preprocessing route and the area under the curve of 1700 – 1800 cm\(^{-1}\). (A) and (D) comprise 643 RNase A spectra and 1116 RNase B spectra, while after selection of the spectra within the IQR (B, E) 321 and 558 spectra are retained for the RNase A and B respectively. ..................124

**Figure 4.7.** AFM topography of a typical gold nanoplate. The area used to collect a grid of 62 TERS spectra from this example is shown by the white square.................................132

**Figure 4.8.** PCA scores plot of all data following cosmic ray removal and baseline correction. This is the same plot that is shown in Figure 4.2b but this time coloured to show how the PCA is grouping the data by measurement grid, based mainly on the spectral features of the silicon background, as shown by the loadings plot (Figure 4.2c). ..................133

**Figure 4.9.** The fitting of asymmetric least-squares baselines under the Si overtone peak. The dot-dashed black line shows the fitted baseline as used to correct the whole spectrum (\(\Sigma = 1000, P = 0.001\)) and the dashed red line shows the fitted baseline used to remove the Si overtone from this region (\(\Sigma = 100, P = 0.01\)). The original TERS spectrum is shown in solid blue.................................................................134

**Figure 5.1.** A schematic diagram of the LC-SERS instrumentation........................................145

**Figure 5.2.** SERS spectra of the four nucleobases at their peak retention times in the range 400-2000 cm\(^{-1}\). (A) Xanthine, (B) Hypoxanthine, (C) Guanine and (D) Adenine. The spectra here are preprocessed as they would be for analysis: normalised against total signal and against the mean blank spectrum. To provide a better visual comparison, these examples have also been scaled on the y-axis between 0 and 1.................................................................150

**Figure 5.3.** Example LC-SERS data from a single injection of a sample containing............153

**Figure 5.4.** Plots of mean \(\bar{y} = 3\) predicted concentration against actual concentration for each of the four nucleobases using leave-one-out calibration. All data are from
injections of nucleobase mixtures using SERS detection. Error bars show one standard deviation. Perfect predictions would lie on the red lines.

**Figure 5.5.** Photographs of the SERS detector during analysis with the enclosure removed. Side view of the sample droplet forming at the end of the LC tubing in front of the laser aperture (top), front-on view (lower left) and a new droplet forming in the laser path as soon as the previous droplet has fallen (lower right).

**Figure 5.6.** Quartz capillary used for LC-SERS analysis, showing the inside of the tube coated with silver.

**Figure 5.7.** Extracting useful information from the LC-SERS data. (A) shows a 3D plot of the raw data obtained from a LC-SERS injection. In this case the sample is $5 \times 10^{-5} \text{M}$ guanine. The spectra are clearly dominated by the vibrations of the Ag nanoparticles’ citrate capping species, shown in (B). This becomes more apparent when the spectra are normalised against the total signal (C). A loss of signal intensity is observed when the analyte displaces the citrate on the surface. Normalisation against the mean blank spectrum (entirely citrate peaks) reveals the relatively low-intensity guanine spectrum (D).

**Figure 5.8.** Calibration plots for injections of the individual nucleobases using UV detection. The mean UV peak area ($N = 3$) is shown with error bars equal to one standard deviation. Insets show a close-up of the main plot in the low concentration region.

**Figure 5.9.** Plots of mean ($N = 3$) predicted concentration against actual concentration for each of the four nucleobases using leave-one-out calibration. All data are from injections of individual nucleobase samples using UV detection. Error bars show one standard deviation. Insets show a close-up of the main plot in the low concentration region.

**Figure 5.10.** Calibration plots for injections of the individual nucleobases using SERS detection. The mean SERS peak area ($N = 3$) at the relevant wavelength (see Table 5.2) is shown with error bars equal to one standard deviation. Insets show a close-up of the main plot in the low concentration region.

**Figure 5.11.** Plots of mean ($N = 3$) predicted concentration against actual concentration for each of the four nucleobases using leave-one-out calibration. All data are from injections of individual nucleobase samples using SERS detection. Error bars show one
STANDARD DEVIATION. Insets show a close-up of the main plot in the low concentration region.

**Figure 5.12. Calibration plots for injections of mixtures of nucleobases using SERS detection.**

The mean SERS peak area ($N = 3$) at the relevant wavelength (see Table 5.4) is shown with error bars equal to one standard deviation. Insets show a close-up of the main plot in the low concentration region.
List of Tables

**Table 2.1.** Comparison between predictions from three different data analysis methods, aimed at predicting pure DPA concentration or spore counts from *B. cereus* or *B. subtilis*. A

... .......................................................................................................................... 77

**Table 3.1.** Peak shifts in the spectra of R6G spots on the BN substrate between the two repeat measurements and R6G powder. A

... .......................................................................................................................... 98

**Table 4.1.** Tentative Raman band assignments for the regions of the TERS spectra that indicate differences between RNase A and RNase B (Brewster et al., 2011).

... .......................................................................................................................... 122

**Table 5.1.** Final concentration of individual analytes in the mixed solutions, obtained by mixing the appropriate stock solutions.

... .......................................................................................................................... 143

**Table 5.2.** Metrics for HPLC analysis of individual nucleobase samples using simultaneous UV and SERS detection.

... .......................................................................................................................... 151

**Table 5.3.** UV metrics for HPLC analysis of mixtures of nucleobase samples using simultaneous UV and SERS detection. Overlap of peaks in the UV chromatograms meant that a zero measurement and confident integration of the peaks from certain low-concentration samples was not possible. Therefore, the LOD quoted in brackets is the lowest concentration sample from which a peak could be measured, rather than the statistical LOD. None of the other metrics reported for the SERS analysis (Table 5.4) were able to be determined.

... .......................................................................................................................... 155

**Table 5.4.** SERS metrics for HPLC analysis of mixtures of nucleobase samples using simultaneous UV and SERS detection.

... .......................................................................................................................... 156
List of Equations

**Equation 1.1.** \( C = \Lambda N \) .................................................................25

**Equation 1.2.** \( E = H N \) .................................................................27

**Equation 1.3.** \( \Lambda = H P \) .................................................................27

**Equation 1.4.** \( M^* = A E \) .................................................................30

**Equation 1.5.** \( I_R \propto E_{\text{incident}}^2 \cdot E_{\text{scattered}}^2 \) .................................................................41

List of Schemes

**Scheme 5.1.** Tautomerisation of adenine between the lactam and lactim forms. ..................170
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM</td>
<td>atomic force microscopy</td>
</tr>
<tr>
<td>ALS</td>
<td>asymmetric least-squares</td>
</tr>
<tr>
<td>AUC</td>
<td>area under the curve</td>
</tr>
<tr>
<td>CCD</td>
<td>charge-coupled device</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DPA</td>
<td>dipicolinic acid</td>
</tr>
<tr>
<td>EMSC</td>
<td>extended multiplicative scatter correction</td>
</tr>
<tr>
<td>FEP</td>
<td>poly (tetrafluoroethylene-co-hexafluoropropylene)</td>
</tr>
<tr>
<td>FON</td>
<td>film-over-nanosphere</td>
</tr>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetylglucosamine</td>
</tr>
<tr>
<td>HOMO</td>
<td>highest occupied molecular orbital</td>
</tr>
<tr>
<td>HPLC</td>
<td>high-performance liquid chromatography</td>
</tr>
<tr>
<td>ID</td>
<td>internal diameter</td>
</tr>
<tr>
<td>IMS</td>
<td>ion mobility spectrometry</td>
</tr>
<tr>
<td>IQR</td>
<td>interquartile range</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>KPLS</td>
<td>kernel partial least-squares</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>LCMS</td>
<td>liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>LOD</td>
<td>limit of detection</td>
</tr>
<tr>
<td>LOO</td>
<td>leave-one-out</td>
</tr>
<tr>
<td>LOQ</td>
<td>limit of quantification</td>
</tr>
<tr>
<td>LUMO</td>
<td>lowest unoccupied molecular orbital</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>NA</td>
<td>numerical aperture</td>
</tr>
<tr>
<td>NIR</td>
<td>near infrared</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PFA</td>
<td>poly (tetrafluoroethylene-co-perfluoropropylvinyl ether)</td>
</tr>
<tr>
<td>PC</td>
<td>principal component</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PLS</td>
<td>partial least-squares</td>
</tr>
<tr>
<td>PTFE</td>
<td>poly (tetrafluoroethylene)</td>
</tr>
<tr>
<td>PTM</td>
<td>post-translational modification</td>
</tr>
<tr>
<td>R6G</td>
<td>rhodamine 6G</td>
</tr>
<tr>
<td>RMSEC</td>
<td>root-mean-squared error of calibration data</td>
</tr>
<tr>
<td>RMSEP</td>
<td>root-mean-squared error of predicted data</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>ROA</td>
<td>Raman optical activity</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>reversed-phase high-performance liquid chromatography</td>
</tr>
<tr>
<td>RRS</td>
<td>resonance Raman scattering</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>SEM</td>
<td>scanning electron microscopy</td>
</tr>
<tr>
<td>SERRS</td>
<td>surface-enhanced resonance Raman scattering</td>
</tr>
<tr>
<td>SERS</td>
<td>surface-enhanced Raman scattering</td>
</tr>
<tr>
<td>STM</td>
<td>scanning tunnelling microscopy</td>
</tr>
<tr>
<td>TEM</td>
<td>transmission electron microscopy</td>
</tr>
<tr>
<td>TERS</td>
<td>tip-enhanced Raman scattering</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>UV-vis</td>
<td>ultraviolet-visible</td>
</tr>
</tbody>
</table>
Abstract

The University of Manchester
David Paul Cowcher
Doctor of Philosophy
The Development of Enhanced Raman Scattering for the Trace Analysis of Biomolecules
27th September 2013

Raman spectroscopy is an established analytical technique for determining molecular structure, whose major drawback is lack of sensitivity. Enhanced Raman scattering techniques, such as surface-enhanced Raman scattering (SERS) and tip-enhanced Raman scattering (TERS), utilise nanoscale substrates to enhance the Raman signal through the interaction of surface charges with the incident electromagnetic radiation. Here, nanoparticle-based SERS was used to detect dipicolinic acid (DPA), a biomarker for bacterial spores. Whilst this has been demonstrated previously, the use of a different nanoparticle aggregation mechanism and the inclusion of an internal standard has enabled a SERS detection method to be developed that is quantitative to almost an order of magnitude lower than previously reported. Moreover, for the first time, a nanoparticle-based SERS method was applied to the detection of viable Bacillus spores. Investigations were made into the possibility of SERS enhancement using deep UV laser excitation at 244 nm using a novel boron nitride surface material. This semiconductor has a band gap of comparable magnitude to the laser excitation wavelength and therefore had the potential to impart a SERS enhancement via a chemical enhancement mechanism. Whilst initial results looked promising using Rhodamine 6G as a test analyte, it was not possible to demonstrate reproducibly and no enhancement was observed on other analytes that were tested. TERS was shown to be able to discriminate between glycosylated and non-glycosylated forms of protein molecules, based on the measurement of just a few molecules at a time. This was achieved even without control of the protein interaction with the TERS substrate. The vibrational peak positions in TERS experiments were shown to be highly dependent on the analyte’s orientation relative to the TERS tip, giving variable and complex spectral data. As such, the data processing and analysis methods had to be carefully considered in order to eliminate bias. Lastly, a novel SERS detector for high-performance liquid chromatography (HPLC) was built and tested. It was shown to be able to quantify purine bases from mixtures in tandem with, and in lower amounts than the conventionally used UV absorbance detection, even when the analyte peaks were co-eluting. This quantitative analysis is conducted on-line and in real-time, making it applicable to high throughput applications.

Together the four research projects presented in this thesis make a significant contribution to the field of enhanced Raman scattering and promote its sensitivity and reproducibility as a quantitative analytical technique for the trace analysis of biomolecules.
Acknowledgements

Primarily my thanks go to my supervisor Roy, first of all for his guidance with my MChem project and then for taking me on to study for my Ph. D. Without his support and advice I wouldn’t have achieved the results and positive outcomes from my research that I have presented here. I would also like to convey my gratitude for his encouragement to present my work at conferences and seminars, from which I have learnt a great deal as well as making new friends and contacts in the scientific community.

I would also like to thank the members of the Goodacre group past and present, as well as many others in the MIB for their collective wealth of knowledge, help and advice. A particular mention must go to Sam, Toria and Lorna for their Raman expertise, Nik for his level-headedness and general scientific wisdom, Cate for help with the biological aspects of my work, Yun and Elon for their statistics and Matlab knowledge, David and Steve for their general help and support, and finally all of the above, along with Andrew, Drupad, Ross and Nav, for making the MIB a fun place to work.

My upmost thanks go to Volker Deckert and Tanja Deckert-Gaudig for welcoming me to their lab, their city and their home in Germany. Without their tireless work, help, support and generosity we would not have achieved the positive outcomes to our collaboration. You both have my deepest admiration for the complexity and quality of the research you conduct. I must also thank Regina for being my guide in Jena, and her group of friends for making me feel at home, as well as the other members of the Deckert group for their help and advice.

Thank you to my close friends in Manchester, in particular my housemate Ross, Roxy, Andy, Sam, the Whitby Wanderers and Wednesday football guys for making me laugh, keeping me sane and helping me through difficult patches, even if you didn’t know you were doing it. Thanks also to Nik for getting me through the Three Peaks and for our many adventures. Long may they continue!

Without my girlfriend Lorna I wouldn’t be where I am today and to her, for her, I am extremely grateful. With her love and support I have been able to concentrate on my work without distraction from things at home. I also commend her insistence on numerous sunny holidays to allow us both much needed escape and relaxation.

My final thanks go to my family, in particular my parents Paul and Beverley for always being there for me whatever I wanted to do, and for bringing me up to be the person I am today. I hope I have made you proud.

I would like to finish with a quote from my A Level physics teacher, Mr. Findlay, in my Sixth Form Yearbook that has often been a source of inspiration to make sure I do things properly.

“David, remember what Archimedes said, ‘give me a firm spot on which to stand, and I will move the Earth!’”
Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and the Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regualtions) and in The University’s policy on presentation of Theses.
Preface

The research presented in this thesis is dedicated to furthering the development of enhanced Raman scattering techniques, in particular surface-enhanced Raman scattering (SERS) and tip-enhanced Raman scattering (TERS). These techniques are relatively new within analytical science but are rapidly gaining recognition in fields as diverse as environmental science, materials science, forensic analysis, and medicine. This is due to the rich structural information they can provide, as well as their ever-increasing sensitivity, selectivity and reproducibility. Whilst the work presented here is focussed on biochemical analyses, the underlying science is applicable in a wide range of contexts and it is my hope that it will provide a meaningful contribution to many areas of future tangential research.

This thesis is comprised of six chapters, including an introduction to the prior evolution of the research field, four results chapters and finally a discussion and conclusion of the thesis as a whole, with future perspectives. One of the results chapters is already published and two are pending submission to scientific journals at the time of writing. Thus the results are presented in an alternative format, in which they are written as they would be found in a scientific journal. Fully published papers are included in Appendix I. The outcomes of this work have only been possible with the contribution of my colleagues and collaborators, and therefore their specific individual contributions are credited at the start of each chapter in order that only my own work is credited to me. I was also able to contribute to the research of other colleagues during the course of this degree and although my contribution was not enough to warrant presentation of the results in the thesis proper, the full publications are also included in Appendix II.
Achieving the results presented herein has been a significant but enjoyable challenge and I am thoroughly pleased with the outcomes. I believe this thesis provides valuable support for the idea that enhanced Raman techniques can be sensitive and reproducible when applied to the analysis of trace materials and I hope that other researchers will be able to build upon these findings in the future.

David Cowcher.
1. Introduction
1.1. Electromagnetic Radiation and Wave-Particle Duality

Electromagnetic radiation is a propagation of energy through space via oscillating electric and magnetic fields. These fields are in phase, but perpendicular to one another and to the direction of travel, as shown in Figure 1.1.

![Figure 1.1](image_url)

**Figure 1.1.** The oscillating electric field ($E$, red) and magnetic field ($B$, blue) with propagation through space ($x$) that constitutes electromagnetic radiation.

The electromagnetic radiation propagates through space with a constant speed, $c$, determined only by the medium through which it is travelling. Both its electronic and magnetic components display the properties of a sine wave, with a frequency, $ν$, and wavelength, $λ$, varying according to Equation 1.1 (Atkins and De Paula, 2006).

**Equation 1.1.**

\[ c = \lambda ν \]

Electromagnetic radiation can be emitted over a wide spectrum of energies from very high energy gamma to very low energy radio waves. Whilst the frequency and wavelength are variable, the speed at which the radiation travels is constant (Equation 1.1). The different
regions of the electromagnetic spectrum and examples of their uses are illustrated in Figure 1.2.

**Figure 1.2.** A cartoon representation of the electromagnetic spectrum (Atkins and De Paula, 2006).

Until the late 19th Century, classical physics – treating electromagnetic radiation purely as a wave that can have any amount of energy – accurately accounted for its observed properties, for example reflection and refraction by a prism; however, new observations started to appear that could not be explained by a classical model, such as the photoelectric effect (Fayer, 2010).

When a metal is illuminated with electromagnetic radiation above a certain energy, electrons are emitted from the surface. The minimum amount of energy required for a particular metal is known as its threshold frequency. If electromagnetic radiation behaved solely as a wave, increasing the intensity of the incident radiation (the amplitude of the wave) would eventually provide enough energy for the metal to emit electrons; however this is not observed. Instead, only increasing the frequency of the incident radiation can
breach the threshold limit and allow emission of electrons, regardless of how intense the radiation is. Einstein hypothesised (Einstein, 1905) that rather than being a continuous wave, the radiation is arriving at the metal surface in ‘packets’ of energy, known as quanta: in effect, behaving as a stream of particles. The energy, $E$, of these particles, later named ‘photons’, is related to the frequency, $\lambda$, of the electromagnetic radiation by Planck’s constant, $h$, equal to $6.626 \times 10^{-34}$ Js (Atkins and De Paula, 2006) in Equation 1.2.

**Equation 1.2.**

$$E = h\nu$$

The ability of electromagnetic radiation to behave as both a particle and a wave is known as wave-particle duality.

The wave-particle duality does in fact hold for all particles, including those with mass, as proposed by Louis de Broglie in 1924 (de Broglie, 1924), who related the ‘wavelength’ of a particle to its linear momentum, $\rho$, in Equation 1.3.

**Equation 1.3.**

$$\lambda = \frac{h}{\rho}$$

This states that all matter should have wave-like properties as well as particle-like properties; however, the incredibly small magnitude of Plank’s constant means that for macro-scale objects, the de Broglie wavelength is so minutely small that the effects are not observable. de Broglie’s theory states that the wavelength of a particle depends only upon its mass and velocity, therefore a particle with a very low mass would have a much higher wavelength, for example an electron, whose mass is $9.109 \times 10^{-31}$ kg (Atkins and De Paula, 2006). The theory was proved correct in 1927 by Davisson and Germer (Davisson and Germer, 1927), who fired slow-moving electrons at a nickel target and observed diffraction patterns analogous to those of X-rays. This phenomenon could only be explained if the electrons were behaving as a wave.
1.2. **Molecular Energy Levels**

A free quantum particle, such as a single, unbound electron, can possess any energy; however, one that is bound, for example as part of an atom or molecule, can only exist in discrete energy levels (Fayer, 2010). When the appropriate amount of energy is given to or removed from a molecule, it can transition between these electronic energy levels. Energy absorbed by a molecule is not exclusively limited to electronic excitations however, and can also cause the molecule to vibrate, rotate, or move. Due to the quantum mechanical nature of objects at this scale, these excitations are also quantised and therefore each molecule has discrete vibrational, rotational and translational energy levels, as illustrated in Figure 1.3. These different types of excitation are promoted by different magnitudes of energy, as illustrated in Figure 1.2.

1.3. **Raman Spectroscopy**

When electromagnetic radiation interacts with a molecule and cannot be absorbed, it is scattered. The interacting photon promotes the molecule to a virtual excited energy state (not corresponding to any formal molecular energy level), from which it relaxes again, re-emitting the photon. In the majority of cases the molecule relaxes to the same energy level from which it was excited and the emitted photon has the same energy at the incident photon: an effect known as elastic, or Rayleigh scattering. In approximately $10^7$ cases, the molecule relaxes to a different energy level than it was excited from, emitting a photon with a different energy to the incident photon, which is known as inelastic, or Raman scattering (Atkins and De Paula, 2006).
Figure 1.3. General molecular energy diagram for the ground \((E_0)\) and first \((E_1)\) electronic energy levels, showing the approximate relative energies of the related vibrational \((V)\) and rotational \((R)\) energy levels.

Stokes scattering occurs when the molecule relaxes to a higher vibrational energy level than it started in, meaning some of the incident photon energy has been converted into vibrational energy of the molecule, with the remainder emitted as a photon of reduced energy. By contrast, anti-Stokes scattering occurs when the molecule relaxes to a lower energy level, emitting a photon of higher energy than the incident photon, as it has gained extra energy from the change in molecular vibration. It is most probable that a molecule will be in its ground vibrational state when the incident photon arrives, due to the Boltzmann distribution of energies (Atkins and De Paula, 2006) and so Stokes scattering is
much more common than anti-Stokes and is typically collected in a Raman experiment. A diagrammatic representation of light scattering is shown in Figure 1.4.

**Figure 1.4.** Interactions of electromagnetic radiation with a molecule in its electronic ground state that cause a change in vibrational energy level. \( V_0 \) represents the molecular vibrational ground state and \( V_1 \) represents the first molecular vibrational excited state. Adapted from (Brown, 1998, Lewis and Edwards, 2001).

In order for Raman scattering to occur, it is necessary that the molecular vibration results in a change in polarisability of the molecule. Polarisability is a measure of the electronic dipole induced when the electron cloud of a molecule interacts with an electric field, such as that of electromagnetic radiation. The induced dipole, \( \mu^* \), is related to the electric field strength, \( E \), by Equation 1.4, where \( \alpha \) is the polarisability.

**Equation 1.4.**

\[
\mu^* = \alpha E
\]
Larger changes in polarisability with molecular vibration lead to more intense Raman scattering (Atkins and De Paula, 2006). The Raman spectrum therefore shows peaks corresponding to the energies of vibrational levels where this selection rule is met.

For a molecule made up of \( N \) atoms, there are \( 3N-5 \) normal modes of vibration if the molecule is linear and \( 3N-6 \) if it is non-linear (Atkins and De Paula, 2006). For example, carbon dioxide, as shown in Figure 1.5, is a linear, tri-atomic molecule and therefore has 4 normal modes of vibration, which can be described as symmetric stretching, asymmetric stretching and bending.

![Normal vibrational modes of carbon dioxide: (a) symmetric stretching, (b) asymmetric stretching, (c) bending and (d) bending (perpendicular to (c)). Adapted from Atkins and De Paula, 2006.](image)

These different vibrational modes occur at different energies and would appear as different bands on the Raman spectrum, providing they obey the selection rule. Raman spectroscopy is therefore able to provide detailed structural information about a molecule based on the different vibrational normal modes it can detect.

The fact that Raman spectroscopy takes place in a static location, and measures a reversible change in the excitation of a molecule means that it is non-destructive.
contact with the sample is required, meaning it is potentially useful in situations such as forensic analysis, where the sample must not be disturbed, or in sterile manufacturing, where analysis can take place through a window into a vessel. These are significant advantages that Raman spectroscopy has compared with destructive techniques, or those where a sample has to be taken for analysis, such as mass spectrometry (MS) or nuclear magnetic resonance (NMR) spectroscopy.

1.3.1. Comparing Raman with Infrared (IR) Spectroscopy

IR is also a vibrational spectroscopy; however in contrast to Raman, molecular vibrations are initiated by absorbance of incident radiation rather than scattering (Figure 1.4). The same vibrational normal modes apply, although the selection rule is different. Molecular vibrations are IR active if the vibration results in a changing permanent electric dipole in the molecule. This gives IR some limitations, as it cannot detect homonuclear diatomic species due to the one allowed vibrational mode not resulting in a change in the dipole. As such, Raman and IR spectroscopies are often used in complementary experiments because in general, vibrational modes that are not Raman active are IR active and vice versa. In fact for a molecule with a centre of symmetry, no vibrational mode can be both IR and Raman active, although some are neither (Atkins and De Paula, 2006).

Raman scattering is a much weaker effect than IR absorbance and therefore difficult to detect, particularly for molecules with only weakly active vibrational modes (low polarisabilities). However, its main advantage over IR is that water, a strong IR absorber, is very weakly Raman active, which makes Raman more amenable to use with aqueous samples and as such, is a potentially useful tool for the analysis of biological samples. However, there are still considerable challenges for the analysis of biological materials with Raman spectroscopy.
1.3.2. **Fluorescence interference**

Fluorescent molecules can be difficult to analyse by Raman spectroscopy if the emission wavelength of the fluorescence falls within the same region of the spectrum as the Raman scattering. Fluorescence occurs when the energy of the incident light photon corresponds to a real electronic excitation. This excitation will be to a higher vibrational energy level in the electronic excited state than in the electronic ground state, as the timescale of the electronic excitation happens more quickly than the bond can adjust its internuclear distance: a so-called vertical transition. The molecule then reverts to the ground vibrational energy level of the excited state by non-radiative decay, before re-emitting a photon and returning to the ground electronic energy level (Atkins and De Paula, 2006). As a consequence of the non-radiative decay phase the timescale of the fluorescence event, at \(~10^{-8}\) s (Nair, 2006), is much longer than a scattering event, but the emitted photon is still of a lower energy than the incident photon and is therefore allowed to pass through the Rayleigh rejection filter of the Raman spectrometer. The fluorescence process is shown diagrammatically in Figure 1.4.

1.3.3. **Raman Instrumentation**

A diagram of the main components of a Raman spectrometer is shown in Figure 1.6. Monochromatic light from a laser source is directed onto a Rayleigh rejection filter, which reflects the laser light onto the sample. Two types of filter are commonly used: notch filters and edge filters. The former transmits light of all wavelengths with the exception of the ‘notch’ region that corresponds to the laser wavelength, whilst the latter only transmits light above or below a particular wavelength, typically 100 cm\(^{-1}\) from the laser wavelength.

The light is scattered by the sample, part of which is scattered back towards the filter, where the Rayleigh-scattered light is reflected and the Raman-scattered light transmitted
though to a diffraction grating, which diffracts it onto a charge-coupled device (CCD) detector.

**Figure 1.6.** Schematic representation of the fundamental components of a Raman spectrometer. Green is used in this figure to depict the laser excitation light and red the inelastically scattered light: either Stokes or anti-Stokes Raman.

The resolution of the instrument will be determined by the quality of the components, for example, the number of lines on the diffraction grating or the number of pixels on the CCD. Depending on the sample, different laser wavelengths may be used for the incident light. The intensity of the scattered light is proportional to the fourth power of the incident light frequency (Lewis and Edwards, 2001). This means that more intense scattering will occur as the laser wavelength is decreased; however, shorter wavelength means higher energy, which increases the risk of damage to the sample, particularly in the case of delicate biological samples. To a point, reducing the wavelength of the laser also makes the scattering more susceptible to interference from fluorescence, giving rise to high background signal, or masking of the Raman peaks, as the Rayleigh rejection filter does not filter out the fluorescence emission. Fluorescence is generally reduced as the
Introduction

wavelength is increased into the near infrared; however, a compromise must be reached between the Raman scattering intensity and the background interference (Lewis and Edwards, 2001). As discussed in Section 1.3.7, Raman enhancement methods may also influence the selection of appropriate laser wavelengths.

Although the fundamentals of Raman instrumentation remain the same, different types of instrument are available for analysis of different sample types; these will be discussed in more detail in the subsequent sections.

1.3.4. Raman microscopy

Raman analysis can be conducted using a conventional light microscope to focus the laser onto the sample and collect the (back) scattered radiation. As such, any sample that can be observed using a light microscope can theoretically be analysed by Raman spectroscopy. This allows for very precise measurement resolution, down to the Abbe diffraction limit (~λ/2), and furthermore, focusing the whole laser beam to a small spot drastically increases the number of photons delivered to the precise position on the sample, thereby increasing the Raman scattering intensity and the measured signal strength.

Samples amenable to Raman microscopy are typically flat solids, or substances that can be immobilised onto a flat, solid substrate, for example crystalline powders (Ali et al., 2008) and bacterial cells (Paret et al., 2010), although it is also possible to focus onto, or within liquid samples (Camerlingo et al., 2007), and there have been reports of gas phase Raman measurements for concentrated samples (Kiefer et al., 2008). Sophisticated sample stages, combined with computer-controlled optics and data processing permit Raman images to be created. In such images, or maps, each pixel of the white light microscope image has a corresponding Raman spectrum, allowing the accurate mapping of individual components.
within a sample, for example detecting the location of proteins, lipids, nucleic acids and cholesterol within a cell (Krafft et al., 2003).

1.3.5. Portable Raman and Liquid Sample Analysis

Whilst Raman microscope instruments tend to be relatively large, bench top, portable instruments are also available (Figure 1.7). These tend to have lower power lasers and lower grade optical components as a consequence of the lack of space inside the instrument, meaning that the resolution and intensity of the spectra produced are not as good as from microscope instruments. There is also little or no room for an accurately controlled sample stage, and therefore these instruments are primarily used for analysing bulk samples, particularly liquids, and most have the facility for a sample vial holder to be fitted (Sorak et al., 2011).

What they lack in measurement sensitivity, they make up for in the speed and portability of analysis, and are of course considerably lower cost. There is no requirement for accurate positioning and focussing on the sample and the instrument can be easily transported for analysis in the field. Fibre optic probes are often available to allow measurements in awkward locations (Battaglia et al., 2004).

![Figure 1.7. Examples of typical Raman instruments: microscope (left) and portable (right).](image)
1.3.6. A Typical Raman Spectrum

A Raman spectrum is a plot of intensity (photon count) against Raman shift – the amount the scattered radiation is changed in energy from the laser line, measured in cm$^{-1}$, known as wavenumber, $\nu$. This is simply the reciprocal of the wavelength, quoted in cm$^{-1}$ for convenience, as the typical spectral range is 200 - 4000 cm$^{-1}$. An example Raman spectrum is shown in Figure 1.8 along with annotation of some of the important regions.

**Figure 1.8.** 785 nm Raman spectrum of $N$-acetylglucosamine (structure inset). Important regions for peak assignments are highlighted (Socrates, 2004).
1.3.7. Enhancing the Raman Signal

Due to the low probability of Raman scattering events, as mentioned previously, there is a desire to increase the detectable signal. The first logical step would be to increase the laser power, thus providing more photons per second at the sample. This would work in theory; however, any background interference, for example fluorescence, would also be increased, meaning no real improvement would be achieved if this were the limiting factor. In addition, increasing the laser intensity requires detectors that are able to cope without becoming saturated and sample management so that the sample does not become damaged by the laser.

More sophisticated methods of Raman enhancement have become commonplace that only enhance the signal from the analyte(s). The most important of these are discussed in the following sections.

1.3.7.1. Resonance Raman scattering (RRS)

Raman scattering usually occurs via a virtual excited state. However, if the energy of the incident photon is approaching that of an excited electronic energy level of the molecule (Figure 1.4.), an enhancement of the Raman intensity is observed, known as resonance Raman scattering (RRS). This enhancement is strongest for vibrational energy levels that couple strongly with the excited electronic state and therefore can provide additional information about the excited state itself. RRS signals can be in the region of four to six orders of magnitude stronger than the corresponding normal Raman signals (Kane and Jensen, 2010).

The theory behind RRS is generally based on two approaches: the vibronic theory (Albrecht, 1961), and the time-dependent theory (Heller et al., 1982); an experimental
comparison of which was presented by Kane and Jensen (Kane and Jensen, 2010). Due to the typical energy gap between ground and excited electronic states, lasers that are appropriate for RRS often fall into the ultraviolet (UV) part of the electromagnetic spectrum. For example, an important application of RRS is the analysis of peptides and proteins, with resonance coming from the 190 nm $\pi - \pi^*$ transition of the amide bond (Song and Asher, 1989) (Asher, 1988).

1.3.7.2. Surface-enhanced (resonance) Raman scattering (SER(R)S)

SERS is the term for Raman enhancements observed when the analyte is located in close proximity to a surface, usually metallic, with nanoscale roughness features. The effect was first observed in 1974 by Fleishmann et al. (Fleischmann et al., 1974), who noted an enhancement in the Raman scattering intensity of pyridine when it was adsorbed onto a roughened silver electrode. This was initially believed to be due to increased surface area provided by the surface roughness; however, Jeanmaire and Van Duyne (Jeanmaire and Van Duyne, 1977), and Albrecht and Creighton (Albrecht and Creighton, 1977) both showed that the enhancement factor was significantly larger than would be predicted by the increase in surface area alone. SERRS is a combination of RRS and SERS that provides multiplicative enhancement by both of the methods simultaneously and is usually used to measure target analytes (dyes) in very low concentrations (Kneipp et al., 1995).

As it forms a major part of the work reported in this thesis, SERS, along with the related technique tip-enhanced Raman scattering (TERS), is discussed in greater depth in Section 1.4.
1.4. Surface-Enhanced and Tip-Enhanced Raman Scattering (SERS & TERS)

Even though SERS was first observed nearly 40 years ago its exact mechanism is still not fully understood, although two theories have been proposed based on experimental observations. Which theory is correct has been a matter of fierce debate in the SERS community; however it is likely that both methods are valid and provide their own contribution to the observed enhancement. In the majority of cases where SERS is observed the two types of enhancement occur simultaneously, although in some cases, one type of enhancement dominates the other.

1.4.1. Electromagnetic Theory

Moskovits (Moskovits, 1978) introduced the idea that SERS enhancement is the result of excitation of surface plasmons (oscillations in the cloud of conduction electrons) of the metal by interaction with the electric field vector of the incident laser light. A model for this was developed simultaneously by Gertsen (Gersten, 1980), Gertsen and Nitzan (Gersten and Nitzan, 1980, Gersten and Nitzan, 1981) and McCall et al. (McCall and Platzman, 1980, McCall et al., 1980), known as the electromagnetic theory of SERS. This was advanced further by Kerker et al. (Kerker et al., 1980, Kerker et al., 1984) to include colloidal metal particles as well as roughened metal surfaces.

Incident light can induce many different forms of excitation at a metal surface, but when the surface features (or colloidal particles) are on a smaller scale than the wavelength of the light, all but the dipolar plasmon excitation can be ignored (Kneipp et al., 2006). If an incident laser wavelength is chosen so that it is resonant with the dipolar surface plasmons, the metal particle or surface feature will radiate light. This radiation will be coherent with the exciting field, creating a spatial distribution of field intensities with regions of
enhanced and depleted fields around the particle or surface feature. Analyte molecules located where the field is enhanced near to the surface show the greatest Raman enhancements because of interaction with photons from both the incident radiation and the radiating plasmon. In addition to this, the Raman scattered light can also excite the surface plasmons, causing radiation of light at the Raman-shifted wavelengths as well. The combination of the enhancements of the incident and Raman-scattered light give a Raman intensity, $I_R$, proportional to the square of the local incident field, $E_{\text{incident}}$, multiplied by the square of the Raman scattered light field, $E_{\text{scattered}}$, (Equation 1.5) meaning that for low frequency Raman bands, where $\lambda_{\text{incident}} \approx \lambda_{\text{Raman}}$, the enhancement scales approximately to the fourth power of the incident field intensity (Campion and Kambhampati, 1998).

**Equation 1.5.**

$$I_R \propto E_{\text{incident}}^2 \cdot E_{\text{scattered}}^2$$

For higher frequency bands (i.e. greater Raman shift), the enhancement is defined by a more complex function and will be lower, as the band is further from the resonant frequency (Kneipp et al., 2006). Observed reduction of Raman band intensity for higher frequency Raman shifts provides strong evidence in favour of the electromagnetic theory (Campion and Kambhampati, 1998); however, there is also evidence of deviations from the model.

### 1.4.2. Chemical Enhancement Theory

CO and N$_2$ are isoelectronic, diatomic molecules with similar polarisabilities and therefore electromagnetic theory would predict similar Raman enhancements for both; however, the observed SERS intensities have been shown to differ by a factor of 200 under identical experimental conditions (Campion and Kambhampati, 1998). This indicates that the electromagnetic theory does not fully account for all SERS observations.
Instead, chemical interaction between the analyte molecule and the metal surface is believed to provide an enhancement mechanism, which can be thought of as a lower energy form of RRS. Adsorption of the analyte onto the metal surface is thought to allow resonance at visible wavelengths (compared with UV in normal RRS) by either shifting or broadening the electronic energy states of the analyte, or by creating new electronic states that function as resonant intermediates. For example, if the Fermi level of the metal surface sits between the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) of the analyte, the incident light could be resonant with charge transfer excitations from the HOMO of the analyte to the LUMO of the metal, or from the HOMO of the metal to the LUMO of the analyte, as illustrated in Figure 1.10 (Campion and Kambhampati, 1998). The dominant charge transfer will depend on the individual analyte, for example Morton and Jensen (Morton and Jensen, 2009) showed that for meta- and para- substituted pyridines on small silver clusters, the enhancement factor was more dependent on the energy gap between the metal HOMO and analyte LUMO than the other way around.

Jensen et al. (Jensen et al., 2008) make a useful summary by categorising SERS enhancement into four groups: (1) ground state Raman band enhancement due to adsorption onto the metal surface alone, (2) surface-enhanced resonance Raman scattering (SERRS), which comprises of RRS of the analyte – surface adduct, (3) charge-transfer resonance Raman, involving electronic excitation from metal to analyte or analyte to metal as previously mentioned, or (4) plasmon resonance enhancement, as described by the electromagnetic theory.
**Figure 1.9.** Energy level diagram of an analyte molecule adsorbed onto a metal surface. The minimum energy required for an intramolecular electronic excitation of the analyte, as occurs in RRS (a) is larger than the energy required for a charge transfer from analyte to metal (b) or from metal to analyte (c), due to the Fermi level of the metal residing between the HOMO and LUMO of the analyte. This allows the incident light to be resonant with electronic excitation at higher wavelength. Adapted from (Campion and Kambhampati, 1998).

### 1.4.3. SERS Surfaces

The hierarchy of metals based on their relative Raman-enhancing power was predicted by the electromagnetic theory before its observation, with silver and the alkali metals providing the best enhancements, followed by gold, copper, other good conductors (aluminium, platinum and indium) and finally the remainder of the transition metals (Kneipp et al., 2006). SERS was first observed on roughened silver electrodes (Fleischmann et al., 1974), and it was later deduced that the nanoscale surface features were responsible for the enhancements (Moskovits, 1978). Since then, the design of
surfaces for SERS has taken a variety of different directions based on some common themes, as will be described in the following sections.

1.4.3.1. Nanoparticles

The use of metal nanoparticles for SERS measurements was first demonstrated by Creighton et al. (Creighton et al., 1979) and these have since become the most popular SERS surfaces (Aroca et al., 2005). Nanoparticles are more popular than roughened metal surfaces, such as the electrodes used in the pioneering SERS experiments, due to their relatively simple and inexpensive preparation methods, although nanoparticle syntheses are gradually becoming more and more complex (Xu et al., 2010, Garcia-Leis et al., 2013). Nanoparticles can be suspended in a liquid (known as a sol), or deposited on smooth substrates (Kho et al., 2005), providing an even distribution of nanoscale surface roughness features. Nanoparticle sols have the advantage of being in suspension, meaning that they are in constant (Brownian) motion, thus the Raman spectrum observed is an average over the laser path through the solution for the duration of the analysis, rather than that of a fixed spot on a surface. It is this subject that divides the two currently prevalent fields in nanoparticle SERS investigations: ‘average’ SERS experiments and ‘hot spot’ SERS experiments (Aroca et al., 2005).

‘Average’ SERS experiments utilise sol suspensions and are popular because they allow the analysis of samples in solution. As well as generating an average spectrum and hence more reproducible signal, the natural movement of the analytes also helps to prevent damaging of the sample and substrate by the laser (Mernagh et al., 1984, McGlashen et al., 1991), meaning higher laser powers can be used to generate more intense spectra and/or more delicate samples can be analysed safely, such as biological samples; the scope for which is presented in the review by Jarvis and Goodacre (Jarvis and Goodacre, 2008).
Usually, an aggregating agent is added to the sol-analyte mixture. This is generally an electrolyte solution, or an organic polymer, such as poly (L-lysine) (Aroca et al., 2005) and serves the dual purpose of decreasing the electrostatic potential between individual nanoparticles and between nanoparticle and analyte (Alvarez-Puebla et al., 2005). Reduction in the electrostatic potential will facilitate easier displacement of the colloid capping species, whose vibrations would otherwise dominate the spectra due to their very close proximity to the nanoparticle surface. The analyte itself can act as an aggregating agent if it has the appropriate functionality, for example in the context of negatively surface-charged nanoparticles, R-OH groups on the analyte that are negatively charged in solution will experience high electrostatic potential, whereas R-NH$_2$ groups that become positively charged in solution experience low electrostatic potential, thus inducing aggregation (Faulds et al., 2004). The extent of aggregation will depend on the power of the aggregating agent to reduce the electrostatic potential and its concentration in the sample.

The extent of aggregation will affect the resonance of the laser light with the surface plasmon of the aggregate. If the aggregates formed are too large then resonance will be lost and SERS intensity reduced; however, when a sample is located at the surface of a nanoparticle aggregate of appropriate size to be resonant with the incident laser, more intense SERS will be observed than with individual nanoparticles. It is from this concept that ‘hot spot’ SERS has evolved.

Regions providing extremely intense SERS have been observed between pairs, or small clusters of nanoparticles (Etchegoin et al., 2003) and these so-called hot spots can provide detection of analyte concentrations of $10^{-12}$ to $10^{-14}$ M (Kneipp et al., 2006). The first observed SERS of a single-molecule was reported by Kneipp et al. for crystal violet
Solutions of small nanoparticle clusters were mixed with analyte molecules in a ratio of 10:1 or greater, meaning it was unlikely that more than one analyte molecule would associate with each cluster. Greater aggregation of the individual clusters was not observed. Micro-Raman spectroscopy of these solutions spotted onto a glass slide provided a probed volume of 1fL to 1pL, which combined with the detection limits stated previously, provided an average of less than one analyte molecule in the probed volume (Kneipp et al., 2006). These observations have rapidly increased the research into SERS analysis using hot spots, particularly in the area of engineering improved nanoparticle clusters (Moskovits and Jeong, 2004).

Research into improved SERS on nanoparticles for both the average and hot spot fields is continually evolving, with different sizes being synthesised (Ojea-Jiménez et al., 2010) as well as different shapes, including rods (Lee et al., 2004), prisms (Jin et al., 2001), plates (Pastoriza-Santos and Liaz-Marzan, 2008), flowers (Xu et al., 2010) and stars (Rodríguez-Lorenzo et al., 2009, Esenturk and Walker, 2009, Vernon et al., 2010, García-Leis et al., 2013)).

1.4.3.2. Patterned surfaces

Nanoscale roughness features can be imparted on a metal surface by other means than deposition of metal nanoparticles. For example, non-metallic nanospheres may be immobilised on a glass substrate and have metal deposited over the top: so called film-over-nanosphere (FON) surfaces, as pioneered by Zhang et al. (Zhang et al., 2005). Dendrite-type structures of silver can be grown on copper surfaces by simple galvanic displacement reactions (Mabbott et al., 2012). Commercially produced SERS substrates are also now available, with precisely controlled surface features, such as Klarite®, manufactured by Renishaw Diagnostics (Perney et al., 2007, Vernon et al., 2010). Due to
their immobile nature, patterned surfaces are thus far restricted to use with Raman microscope instruments.

1.4.3.3. Non-metal SERS surfaces

SERS has also been reported using non-metallic surfaces, for example graphene (Ling et al., 2009) and semiconductor materials, such as silicon and germanium (Wang et al., 2011). The enhancement factors for graphene substrates can be comparable to those of conventional metal nanoparticle SERS (Ling et al., 2009), although the enhancement factors for semiconductor-based SERS are lower. Non-metal SERS is more significantly influenced by the charge transfer processes associated with the chemical enhancement mechanism (see 1.4.2) and in the case of semiconducting substrates, this is the only possible enhancement mechanism due to their structure not supporting surface plasmon resonance (Wang et al., 2011).

1.4.4. Tip-enhanced Raman scattering (TERS)

TERS combines a variation on the enhancement of SERS with the spatial resolution of scanning probe microscopy. Whereas SERS uses ensembles of nanoparticles over the entire sample to enhance the Raman scattering by associated molecules, TERS uses a nanoscale metal tip that is moved over the surface of the sample, initiating enhancement at each location.

The size and shape of a nanoparticle is highly influential on the enhancement it provides at a particular excitation wavelength (Bailo and Deckert, 2008). Even in the most carefully controlled syntheses, nanoparticles exist in a size distribution, meaning there is also a distribution in their individual enhancement capabilities. Although this is averaged out to a certain extent by the fact that many nanoparticles fall within the laser focus, often SERS
experiments tend to suffer from a lack of reproducibility. This is not so much the case with TERS, as the same tip is used for the acquisition of each spectrum, although degradation via surface oxidation or mechanical impact during a scan may attenuate the enhancement, and different tips do of course offer different enhancing capabilities between measurements. If there is more than one nanoparticle in the laser focus, the spatial resolution of SERS microscopy is diffraction limited to $\sim \lambda_{\text{laser}} / 2$, whereas the resolution of TERS can be as small as the enhancement region of the tip which is typically in the order of $<10\, \text{nm}$ (Treffer et al., 2012).

The fundamental equipment required for TERS experiments, in addition to a Raman microscope, is a scanning probe microscope, such as a scanning, tunnelling microscope (STM) or more frequently an atomic force microscope (AFM). This is required not only to scan the nanoparticle over the surface of the sample, and to generate a topological map of the areas to be probed by Raman, but also to control the distance of the nanoparticle from the sample’s surface via some form of feedback (Bailo and Deckert, 2008). A diagram of a typical TERS setup is shown in Figure 1.11.

TERS tips can be prepared by electrochemical etching of metal wires (Dickmann et al., 1996, Ren et al., 2004, Zhu et al., 2007), by depositing metal onto silicon AFM tips or glass optical fibres by evaporation of metal films (Rasmussen and Deckert, 2006), or growth of a colloidal nanoparticle at the tip (Wang et al., 2005).

The near-field resolution of TERS means it can be used to image samples on a much smaller scale than conventional Raman and SER(R)S. When using AFM as the scanning probe, simultaneous AFM topography of the surface and TERS images can be generated, so that the data is effectively 4-dimensional. Applications are numerous, including imaging
of components within living cells and even detecting individual bases along a single strand of DNA (Treffer et al., 2011).

![Schematic representation of a typical TERS setup using an AFM as the scanning probe microscope.](image)

**Figure 1.10.** Schematic representation of a typical TERS setup using an AFM as the scanning probe microscope.

### 1.5. Analysis of Raman Data

A well designed data analysis strategy can draw a lot of useful information from Raman spectra. The most straightforward means of analysing Raman spectroscopic data is with traditional univariate techniques, such as recording the position of peak maxima and calculating the area under peaks in order to quantify relative differences within and between samples.
However, important changes in samples are often only subtly manifested in their spectra, for example with the slight broadening of a peak (Gouadec and Colomban, 2007). Conversely, and particularly when enhanced Raman techniques are used, spectra may change so much that it is near impossible to pick out common trends, as many tens of peaks would need to be integrated individually and compared with one another. As such, more sophisticated, multivariate data analysis techniques are commonly used, where the spectra are analysed as a whole, rather than using selected component peaks. By analysing the whole spectra, these techniques are sensitive to subtle changes in the data that may not be immediately obvious upon visual inspection alone and can also detect patterns that are hidden within complicated data (Feng et al., 2010, de Groot et al., 2001).

1.5.1. **Preprocessing**

In order for the multivariate techniques to work as effectively as possible, some preprocessing of the data is often required. In this case, the term ‘preprocessing’ covers any transformation of the data prior to the eventual analysis method and commonly includes processes such as baseline correction – removing any background signal, such as that caused by fluorescence that might interfere with the analysis of the Raman signal (Eilers and Boelens, 2005), removing any peaks that may be present in the background signal, for example from the sample vessel, smoothing the signal to reduce noise (Savitzky and Golay, 1964), or scaling to correct for differences in scattering intensities between samples.

Different combinations of preprocessing techniques, along with the order in which they are performed, can significantly affect the outcome of multivariate analysis and need to be chosen with care so as not to bias the result (Brewster et al., 2011).
1.5.2. Principal Components Analysis (PCA)

PCA is one of the most widely used multivariate methods and is part of a class of techniques known as unsupervised methods, meaning the results are generated based on the data alone and not on a priori information, such as sample concentration etc. (Daffertshofer et al., 2004). PCA takes a matrix of data – in the case of Raman data with rows representing each sample, made up of $n$ columns (variables) relating to the wavenumber values of each data point. The PCA algorithm then calculates the axis of maximum variance – the first principal component (PC) – when these variables are plotted in $n$-dimensional space, and assigns each sample a score based on its position along this axis. This axis is then fixed and a new axis of maximum variance that is orthogonal to the first is defined as the second PC, and so on up to a pre-defined number of PCs. The relationship between each sample’s score and the original data is defined by another matrix known as the loadings matrix. As a consequence of the way the PCs are calculated, each successive PC accounts for a smaller amount of the overall variance in the data than the previous one, meaning most of the useful information can be gained from the first few PCs. Samples with similar PC scores can generally be defined as similar and therefore plotting PCs against one another leads to clustering of closely related samples. The origins of the similarities or differences can be related back to the original variables via the loadings matrix (Otto, 1998, Brereton, 2007). The applications of PCA are incredibly diverse and a vast wealth of research has been published using the technique to analyse Raman data, so just a few relevant examples are cited here (Webster et al., 2011, Mabbott et al., 2013, Feng et al., 2010).
1.5.3. Multivariate Regression Analysis

Unlike PCA and other unsupervised methods, supervised multivariate methods use prior information about the original data to look for patterns in the data set that can then be used to classify data where this information is not known. Common examples include regression techniques, such as partial least-squares (PLS) regression (Martens and Naes, 1989), where this time, the algorithm aims to maximise the covariance between an axis through the n-dimensional space and a vector containing the prior information. For example, regressing a matrix containing Raman spectra, as in the PCA example above, onto a vector of sample concentrations, thus enabling prediction of the concentration of an unknown sample based on the whole spectrum, rather than on particular peak (Geladi and Kowalski, 1986). Extensions to such methods have also been developed that enable non-linear regressions, for example kernel partial least-squares (KPLS) regression (Shawe-Taylor and Cristianini, 2004), which uses an additional function to project the original data into a higher dimensional feature space in which the regression can be analysed as if it were linear. Much like PCA, there are too many examples of the use of multivariate regression methods for Raman data analysis to include here and so some good recent examples have been chosen as an illustration (Gracie et al., 2014, Brewster et al., 2011, Nicolaou et al., 2011).

1.6. Aim of the Thesis

The overall aim of this research is to build upon the developments in enhanced Raman scattering techniques detailed in the previous section. In particular, sensitive and quantitative analyses will be targeted to enable the measurement of analytes at low concentration. Specific attention will be paid to the accuracy and reproducibility of the data
recorded, to ensure that any analytical methods developed will be robust and provide a strong platform for further work in the analytical field.

For example, bacterial spore detection has been previously demonstrated using SERS but the most sensitive existing techniques require bulky and expensive instrumentation, as well as carefully synthesised surface substrates. In addition, when samples are dried onto solid surfaces, as is often the case with confocal SERS analysis, the truly quantitative nature of the technique is lost. This work aims to tackle these issues by improving the sensitivity and demonstrating that truly quantitative and reproducible SERS measurements are possible from samples in the solution phase.

Novel SERS substrates will be investigated in order to explore the possibility of extending the range of SERS excitation wavelengths available into the deep UV. This would have the potential to improve the sensitivity of resonance Raman scattering at these wavelengths where interference from fluorescence is minimal, as well as exploiting charge transfer interactions at the surface to enhance the scattering intensity of analytes that would not normally be resonant at these wavelengths.

The ultra high resolution of TERS will be explored in attempt to characterise protein glycosylation in a greater level of detail than is possible using conventional Raman spectroscopy. Due to the complicated nature of TERS data, particular attention will be paid to the data analysis strategy in order to ensure the observed results are genuine and unbiased.

Finally, the sensitivity and reproducibility of solution-phase SERS and carefully considered data analysis mentioned previously will be brought together to show the adaptability of enhanced Raman scattering and its potential for use in high throughput
analysis of mixed analytes, through the design and demonstration of an on-line SERS detector for high performance liquid chromatography (HPLC). Despite promising beginnings, this field of research has recently drifted towards more time-consuming at-line methods. With the lessons learned from the previous sections of this work, it will be demonstrated that real-time SERS measurements can be sensitive and quantitative, providing a valid alternative approach to the conventional analyses of mixtures.
1.7. References


outliers and spectral deviations in near-field surface-enhanced Raman spectra. *Analytica Chimica Acta*, 446, 71-83.


58


2. Portable, Quantitative Detection of *Bacillus* Bacterial Spores Using Surface-Enhanced Raman Scattering
2.1. Declaration

This chapter consists of one published journal article:


This article has been reproduced in an unchanged format except for minor adjustments to incorporate it into this thesis.

As primary author on this publication, I carried out all of the experimental work, data analysis and write-up. Dr. Yun Xu authored scripts for data processing in Matlab, as well as providing general expertise and advice regarding data analysis strategy and methodology. Prof. Roy Goodacre, as Principal Investigator, provided the project brief and secured funding for the work to be carried out, in addition to providing advice and direction during the course of the research.
Portable, Quantitative Detection of Bacillus Bacterial Spores Using Surface-Enhanced Raman Scattering

Analytical Chemistry, 85, 3297-3302

Portable, quantitative detection of *Bacillus* bacterial spores using surface-enhanced Raman scattering.

David P. Cowcher, Yun Xu and Royston Goodacre*.

School of Chemistry and Manchester Institute of Biotechnology, University of Manchester,
131 Princess Street, Manchester, M1 7DN, UK

*Correspondence to roy.goodacre@manchester.ac.uk

Received: December 17, 2012. Accepted: February 14, 2013. Published: February 14, 2013

2.2. Abstract

Portable rapid detection of pathogenic bacteria such as *Bacillus* is highly desirable for safety in food manufacture and under the current heightened risk of biological terrorism. Surface-enhanced Raman scattering (SERS) is becoming the preferred analytical technique for bacterial detection, due to its speed of analysis and high sensitivity. However in seeking methods offering the lowest limits of detection, the current research has tended toward highly confocal, microscopy-based analysis, which requires somewhat bulky instrumentation and precisely synthesized SERS substrates. By contrast, in this study we have improved SERS for bacterial analyses using silver colloidal substrates, which are easily and cheaply synthesized in bulk, and which we shall demonstrate permit analysis using portable instrumentation. All analyses were conducted in triplicate to assess the reproducibility of this approach, which was excellent. We demonstrate that SERS is able to
detect and quantify rapidly the dipicolinate (DPA) biomarker for *Bacillus* spores at 5 ppb (29.9 nM) levels which are significantly lower than those previously reported for SERS and well below the infective dose of $10^4$ *B. anthracis* cells for inhalation anthrax. Finally, we show the potential of multivariate data analysis to improve detection levels in complex DPA extracts from viable spores.

2.3. Introduction

Spore-forming bacteria, such as *Bacillus*, are prevalent in soil and on plants and, as such, are commonly found in plant-based foods. They can easily be transferred to meat and dairy products during production, and although most forms are harmless, certain types, for example *B. cereus* and *B. subtilis*, release toxins that cause food poisoning (Granum, 1997). The most dangerous and notorious form, *B. anthracis*, causes anthrax, and its potential for use in bioterrorism was highlighted by the attacks on the U.S. postal service in 2001 (Dalton, 2001). A method of rapid and sensitive detection of spores, amenable for portable field use, is therefore highly desirable.

The vast number of *Bacillus* strains and their associated toxins means infective doses can range from $10^4$ - $10^{11}$ spores/ cells (Granum, 1997); therefore, the lowest end of this range is the necessary limit of detection (LOD). Biological analyses, such as PCR, are able to identify and discriminate between bacteria at the strain level, sometimes with as few as 10 spores (Gulledge et al., 2010), but even though analysis can be completed in the region of 1 h (Bell et al., 2002), it is still far too slow and interest has shifted toward small molecule biomarker detection. This results in a compromise between loss of discriminatory information and speed of analysis; the idea being that rapid detection allows precautionary measures to be taken while the specific danger level is more thoroughly assessed.
A particularly useful biomarker is dipicolinic acid (DPA), as it is found in bacterial spores (usually as a Ca$^{2+}$ chelate complex) but not in other spores, such as mold or pollen, or other commonly occurring environmental materials (Murrell, 1969). DPA can be detected and quantified by common laboratory-based techniques, such as liquid chromatography (Paulus, 1981), gas chromatography/ion mobility mass spectrometry (Dworzanski et al., 1997), and Fourier transform infrared spectroscopy (Helm and Naumann, 1995, Goodacre et al., 1999), although these techniques often lack the sensitivity required, particularly when sensitivity has to be compromised for portability (LODs in the tens of nanograms in the case of HPLC-UV (Paulus, 1981) and the prototype portable GC/IMS (Dworzanski et al., 1997)).

Appropriate sensitivity has been achieved in a time scale of a few minutes, by complexing Tb$^{3+}$ with DPA and measuring photoluminescence (Rosen et al., 1997, Pellegrino et al., 2002); however, this method is susceptible to interference from other molecules, either through DPA binding competition, leading to false positive results, or inhibition and luminescence quenching (Rosen et al., 1997), leading to false negatives. Despite significant improvement (Cable et al., 2009), interference in environmental samples cannot yet be completely eliminated.

Raman spectroscopy, in particular surface-enhanced Raman spectroscopy (SERS), has been shown to be a useful tool for DPA detection and quantification (Farquharson et al., 2004b, Farquharson et al., 2004a, Bell et al., 2005, Zhang et al., 2005, Zhang et al., 2006, Farquharson and Inscore, 2006, Jarvis et al., 2006, Cheng et al., 2009), as the vibrational spectrum allows definitive identification of DPA from interfering species, analysis time is less than 1 min, and portable instrumentation is readily available. Its use has even been
demonstrated with samples collected from mail sorting apparatus (Farquharson et al., 2004b), albeit at milligram level.

SERS exploits the enhancement in Raman signal intensity observed when an analyte is brought into close proximity with a metal surface that has nanoscale roughness features. These surfaces can be loosely divided into two groups: solid surfaces with immobilized nanoscale features and colloidal suspensions. DPA quantification has been reported using both.

Very low LODs have been reported on solid surfaces, down to 0.1 parts per billion (ppb) DPA (Cheng et al., 2009). While this is an impressive LOD, this technique requires the use of a Raman microscope, which greatly limits the portability of the method. In our opinion, the most practical approach to portable SERS detection of Bacillus is to use colloidal nanoparticles, as reported by Bell et al (Bell et al., 2005). This permits the use of smaller, simpler, and less expensive instrumentation and provides a true indication of analyte concentration, as the nanoparticles and adsorbed analyte move within the laser path under Brownian motion to give an averaging effect, rather than trying to manually focus on one fixed point. It is the improved detection of DPA using colloidal SERS that we report here along with experiments to detect DPA from B. cereus and B. subtilis due to their implications in food poisoning.
2.4. **Experimental**

Reagents. DPA (99%), manganese sulfate monohydrate (99%), trisodium citrate dihydrate (99.5%), glutaric acid (96%), and silver nitrate (99.9%) were used as supplied by Sigma Aldrich Ltd. (Dorset, U.K.). Nitric acid (70%) was used as supplied by Fisher Scientific (Loughborough, U.K.). Sodium sulfate decahydrate (99%) was used as supplied by BDH Ltd. (Poole, U.K.). Lab 28 blood agar base (without blood) was used as supplied by Lab M (Bury, U.K.).

2.4.1. **Colloid.**

Citrate-reduced silver colloid was prepared according to the Lee and Meisel method (Lee and Meisel, 1982). Briefly, 90.0 mg of AgNO$_3$ was dissolved in 500 mL of doubly distilled, deionized water and brought to boiling. Then 10.0 mL of 1% w/v aqueous sodium citrate solution was added with stirring, and boiling was resumed for 1 h. We have previously thoroughly characterized colloid prepared by this method using transition electron microscopy (TEM) (Shadi et al., 2009) and scanning electron microscopy (SEM) (Mabbott et al., 2013) and here compared the UV–vis absorbance spectrum to confirm that this colloid batch was comparable (data not shown).

2.4.2. **Instrumentation**

SERS spectra were recorded using a DeltaNu Advantage 200A portable spectrometer (DeltaNu, Laramie, WY, USA), equipped with a 633 nm HeNe laser giving a power of 3 mW at the sample, which was calibrated using a polystyrene internal standard as supplied by the instrument manufacturers. All analyses were conducted in triplicate.

2.4.3. **SERS Analysis of Pure DPA.**

DPA solutions of 1000 ppm (5.98 mM) were prepared in both deionized water and
20.0 mM aqueous nitric acid. These solutions were further diluted with their respective solvents to provide a range of different concentrations.

SERS analysis was performed as follows: 200 µL of silver colloid was added to a glass vial, followed by 200 µL of the DPA solution, and lastly 50 µL of 0.100 M sodium sulphate aggregating agent. The vial was capped and vortexed for 2 s, and then a 30 s SERS spectrum was recorded. In preliminary experiments (vide infra) aggregation times were assessed using sequential 30 s spectral collections for 30 min (60 measurements in total).

A color change of the colloid from light gray to darker gray indicates successful aggregation. In the case of the nitric acid solutions, it was apparent that self-aggregation was occurring before addition of the aggregating agent; therefore, for these samples the aggregating agent was omitted. In the experiments where direct comparisons were made between the two solvent systems, 50 µL of deionized water was added in place of the aggregating agent to maintain comparable volumes and concentrations. This was not required for all other measurements. For quantitative measurements, an internal standard of glutaric acid was added to the nitric acid solvent at a fixed concentration of 2.00 mM.

2.4.4. Culture of Bacterial Spores

*B. subtilis* and *B. cereus* spores were cultured according to the method reported by Goodacre *et al.* (Goodacre *et al.*, 1999). Briefly, Lab 28 nutrient agar was prepared containing 5.0 ppm MnSO₄ to promote sporulation. Cultures were incubated at 30 °C for 6 d, by which time the nutrient source had been depleted and the colonies were no longer growing. Spore colonies were scraped from the agar plates using an inoculating loop and suspended in deionized water. The spore suspensions were maintained at 50 °C in a water bath for 1 h to ensure maximum sporulation had occurred.
2.4.5. DPA Extraction and Analysis

Two 1.0 mL aliquots of the spore suspension were each washed once to remove any dissolved impurities, such as salts from the growth medium, by centrifuging at 17000g for 5 min and the supernatant was discarded. One wash and resuspension cycle was determined to be optimal (data not shown). One of the respective spore pellets was resuspended in deionized water, serial diluted, and used to perform a viable spore count. The other pellet was resuspended in 40.0 mM (40.0 mM enabled better extraction than the original 20.0 mM) nitric acid containing 2.00 mM glutaric acid internal standard and diluted with more of this solvent to provide a range of spore concentrations. These spore samples were ultrasonicated for 10 min and analyzed in the same manner as described above for the pure DPA samples.

2.4.6. Data Processing

All data analysis was performed using Matlab software version R2008a (The MathWorks, Natick, MA, USA). Univariate analysis comprised of normalizing the area under the curve of the most prominent feature: the pyridine ring-breathing mode at 1006 cm\(^{-1}\), to the C–H stretch of the glutaric acid internal standard at 2934 cm\(^{-1}\), after baseline correction using the asymmetric least-squares method (Eilers and Boelens, 2005) and scaling each spectrum on the intensity axis from 0 to 1.

Multivariate analysis was carried out using partial least-squares (PLS) and kernel partial least squared (KPLS) regression models, as reported previously by our research group (Nicolaou et al., 2011a, Nicolaou et al., 2011b), which analyze the whole SERS spectra. The raw spectra had been preprocessed using extended multiplicative scatter correction (EMSC) applied prior to PLS and KPLS analyses.
For all three analysis techniques, 40% of the data (the test set) were randomly selected to be left out and a calibration model was generated based on the remaining 60% (the training set). This model was used to predict the “unknown” concentrations of the data in the test set. Five different test sets were randomly chosen for each species.

2.5. Results and Discussion

For a starting point we repeated the method reported by Bell et al. (Bell et al., 2005) using a portable 633 nm Raman spectrometer, with the aim of applying similar methodology to the analysis of DPA extracted from viable spores. The spectra of DPA (Figure 2.1.) dissolved in deionized water were compared directly to those of DPA dissolved in 20.0 mM nitric acid, which has previously been shown to extract DPA from Bacillus spores (Zhang et al., 2005). It has also been reported that acidic conditions (pH 2−4) provide the optimum conditions for SERS of DPA (Cheng et al., 2009). The potassium thiocyanate internal standard described in the work by Bell et al. (Bell et al., 2005) was omitted at this stage to avoid undesirable reactions with the nitric acid.

This comparison lead to some interesting observations: first that the time taken for maximum SERS intensity to be achieved using the Na$_2$SO$_4$ aggregating agent significantly increased as the concentration of DPA was reduced (over 1 h approaching the 1 ppm LOD) and second that the nitric acid solution was capable of aggregating the colloid on its own, allowing any additional salt-based aggregating agent to be omitted from the nitric acid samples.
Figure 2.1. Baseline-corrected SERS spectra of glutaric acid internal standard, pure DPA, and DPA extracted from *B. cereus* and *B. subtilis* spores using nitric acid. Highlighted are the DPA ring breathing vibration at 1006 cm$^{-1}$ and the glutaric acid C–H stretch at 2934 cm$^{-1}$ that are used for quantification. Chemical structures of DPA and glutaric acid are shown.

The relative SERS intensities for comparable DPA concentrations after aggregation with Na$_2$SO$_4$ or HNO$_3$ are shown in Figure 2.2. Here we see that, for the same DPA concentration, nitric acid aggregation significantly lowers the time taken for maximum SERS intensity to be reached. It also increases the overall signal intensity and improves the reproducibility (as judged by the small standard deviation error bars) compared to aggregation using sodium sulfate.
**Figure 2.2.** Time-related plot showing the effect of two aggregating agents on 633 nm SERS signal from DPA. A constant concentration of DPA was used (5 ppm, 29.9 µM) and 30 s SERS signals were collected sequentially after the initiation of aggregation by 20 mM nitric acid and 0.100 M sodium sulphate. The points are the mean areas under the ring-breathing mode at 1006 cm\(^{-1}\) from baseline-corrected spectra, \(n = 3\). Error bars show one standard deviation.

While we cannot confirm the exact mechanism of aggregation that operates in either case, the presence of the nitric acid will significantly lower the pH of the system, which has been shown to rapidly initiate aggregation of silver nanoparticles even in the absence of an additional salt (Mabbott et al., 2013). In the same work, it is also observed that silver colloid aggregated with nitrate salt gives stronger SERS enhancement than with sulfate. We believe that a combination of these two factors allow for the dramatic improvement in SERS signal we observe over the sulfate-based aggregation. As the use of the thiocyanate internal standard would necessitate stricter safety controls in combination with the nitric
acid (due to the liberation of cyanide), we instead use glutaric acid, which will associate with the colloid in the same manner as DPA (via the carboxylate groups) but provide a C–H stretching mode not present in DPA for use as a reference.

### 2.5.1. Univariate Analysis of Pure DPA

The univariate calibration curve for DPA in 20.0 mM aqueous nitric acid is shown in Figure 2.3a. The SERS signal is now normalized to the internal standard to remove unavoidable effects such as colloid load differences or fluctuations in laser fluency. Normalization was achieved simply by dividing the area under the pyridine ring-breathing mode of DPA at 1006 cm$^{-1}$ by the area under the C–H stretch at 2934 cm$^{-1}$ from the glutaric acid (2.00 mM) internal standard. Note that in this plot the mean of three triplicate preparations is shown along with the standard deviation error bars, again illustrating the excellent reproducibility of this colloidal SERS approach.

In Figure 2.3a we see a curved relationship as we approach the LOD. This is likely to be due to binding competition at the nanoparticle surface between the citrate capping agent, DPA and glutaric acid internal standard: an effect that has been described in detail for mixed analyte detection by Stewart et al. (Stewart et al., 2012). The amount of glutaric acid that is bound to the nanoparticles appears to be unchanged in the curved region in Figure 2.3a, thereby still enabling it to be used as an internal standard. In fact, the glutaric acid is greatly in excess, particularly at the lowest detectable DPA concentrations, which is needed in order to detect the glutaric acid C–H stretching vibration. We would expect the DPA to associate preferentially with the nanoparticle surface due to the greater rigidity of its pyridine structure compared to the flexible backbone of the glutaric acid, making the second carboxylate group of DPA ideally placed for binding once the first has associated with the surface.
Figure 2.3. (A) Calibration curve showing the variation of the mean area under the
1006 cm\(^{-1}\) ring-breathing mode from DPA, normalized against the glutaric acid internal
standard C–H stretch at 2934 cm\(^{-1}\). Data are from baseline-corrected and scaled spectra and
the mean of \(n = 3\) measurements are shown. Error bars show one standard deviation and
the inset shows a close-up of the 0–100 ppb region. (B) Logarithm of the mean areas
shown in the inset in (A) against DPA concentration, providing evidence for a nonlinear
calibration.
Portable, Quantitative Detection of Bacillus Bacterial Spores Using Surface-Enhanced Raman Scattering

We only observe a decrease in the glutaric acid C–H stretch as the DPA concentration exceeds 200 ppb (data shown in the Supplementary Information, Figure 2.5): a tentative explanation for which is that due to its preferential binding affinity, the DPA will be as close as possible to the nanoparticle, with the glutaric acid having to occupy secondary and higher order spheres of association. As the DPA concentration increases and it begins to fill higher order spheres, the glutaric acid is forced out of the enhancement range, or to lose association with the nanoparticle altogether, hence its SERS signal decreases.

Plotting the logarithm (here we use base 10) of the peak area against concentration allows us to use a linear model for predicting DPA concentrations in the range 0–100 ppb, as shown in Figure 2.3b. Above this range we see a linear region from 100–200 ppb, and then as the nanoparticle surface becomes saturated, the signal levels out.

These data provide us with a LOD (based on 3 times the standard deviation in the blank) of 5 ppb (29.9 nM) and a LOQ (based on 10 times the standard deviation in the blank) of 45 ppb (269 nM). However, when 40% of the data are randomly removed from the calibration model to form the test set (see Table 2.1), the practical LOD rises to 10.2 ppb (60.1 nM). As far as we are aware, the lowest previously reported LOD of DPA using colloidal SERS is that by Bell et al. of 1 ppm (5.98 µM) using the sulphate-aggregated colloid; 2 orders of magnitude higher than the LOD we report here.

It is difficult to relate the concentration of pure DPA to a number of spores accurately, as different bacterial types and even spores of the same type are likely to have differing DPA content. However, a rough estimate can be calculated based on 10% of the spore mass being DPA (Paidhungat et al., 2000) and the approximate spore mass determined by Zhang et al. (Zhang et al., 2005) of $1.8 \times 10^{-11}$ g, giving $1.8 \times 10^{-12}$ g of DPA per spore. Our LOD
of 10.2 ppb therefore equates to approximately 1100 spores in our 200 µL sample volume; a level well below the infective dose of 10000 for inhalation anthrax (Granum, 1997).

The vials used for all the analyses detailed in this report are those commercially available to fit DeltaNu spectrometers, which are cylindrical and 8 mm in diameter, requiring at least 400 µL of sample to record a spectrum (200 µL colloid and 200 µL sample, including aggregating agent and internal standard). Given that the diameter of the laser beam is at most 1 mm, a custom-made vial would reduce the sample volume required further without reducing the sample interaction with the laser. We believe that this would reduce the number of spores detectable by a factor of at least 10.

### 2.5.2. Univariate Analysis of DPA Extracted from Bacillus Spores

The SERS spectra of the DPA extracted from spores were processed in exactly the same manner as those of the pure DPA; although instead of DPA concentration, our calibration is now based on spore count. Representative spectra of the extracts compared to pure DPA are shown in Figure 2.1. The LOD of DPA based on the DPA pyridine ring/glutaric acid C–H stretch in the *B. cereus* and *B. subtilis* extracts achieved was 1 954 000 spores and 2 230 000 spores respectively. This is obviously very high but is an issue with respect to DPA extraction rather than the SERS analysis per se.

### 2.5.3. Multivariate Data Analysis for Estimating DPA Concentrations and Spore Counts

In addition to the univariate analysis, we investigated multivariate calibration using the supervised methods PLS and KPLS regression. Example plots from these analyses are shown in Figure 2.4, with a comparison between them and the univariate analysis shown in Table 2.1.
Figure 2.4. Representative example plots of predicted DPA concentration or spore count against known DPA concentrations or spore counts. Three different analysis methods are shown based on univariate assessment (area under 1006 cm$^{-1}$), partial least-squares (PLS) and kernel PLS. Solid circles show the training data set (60%) and outline circles shown the test data set (40%).
Portable, Quantitative Detection of Bacillus Bacterial Spores Using Surface-Enhanced Raman Scattering

Table 2.1. Comparison between Predictions from Three Different Data Analysis Methods, Aimed at Predicting Pure DPA Concentration or Spore Counts from *B. cereus* or *B. subtilis*.\(^a\)

<table>
<thead>
<tr>
<th>Species</th>
<th>Metric</th>
<th>Univariate</th>
<th>PLS</th>
<th>KPLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure DPA</td>
<td>Factors</td>
<td>N/A</td>
<td>6 - 10</td>
<td>4 - 8</td>
</tr>
<tr>
<td></td>
<td>(Q^2)</td>
<td>0.842 (0.036)</td>
<td>0.786 (0.079)</td>
<td>0.839 (0.094)</td>
</tr>
<tr>
<td></td>
<td>(R^2)</td>
<td>0.930 (0.021)</td>
<td>0.979 (0.021)</td>
<td>0.971 (0.016)</td>
</tr>
<tr>
<td></td>
<td>RMSEP</td>
<td>10.2 (1.3)</td>
<td>11.8 (2.12)</td>
<td>9.973 (3.087)</td>
</tr>
<tr>
<td></td>
<td>RMSEC</td>
<td>8.35 (1.02)</td>
<td>4.18 (1.95)</td>
<td>5.179 (1.594)</td>
</tr>
<tr>
<td><em>B. cereus</em></td>
<td>Factors</td>
<td>N/A</td>
<td>3 - 6</td>
<td>2 - 7</td>
</tr>
<tr>
<td></td>
<td>(Q^2)</td>
<td>0.683 (0.225)</td>
<td>0.805 (0.159)</td>
<td>0.644 (0.280)</td>
</tr>
<tr>
<td></td>
<td>(R^2)</td>
<td>0.860 (0.081)</td>
<td>0.944 (0.070)</td>
<td>0.964 (0.047)</td>
</tr>
<tr>
<td></td>
<td>RMSEP</td>
<td>2760000 (332000)</td>
<td>1310000 (974000)</td>
<td>2880000 (917000)</td>
</tr>
<tr>
<td></td>
<td>RMSEC</td>
<td>2340000 (395000)</td>
<td>1330000 (612000)</td>
<td>1090000 (749000)</td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td>Factors</td>
<td>N/A</td>
<td>2 - 6</td>
<td>3 - 6</td>
</tr>
<tr>
<td></td>
<td>(Q^2)</td>
<td>0.116 (0.200)</td>
<td>0.419 (0.141)</td>
<td>0.692 (0.117)</td>
</tr>
<tr>
<td></td>
<td>(R^2)</td>
<td>0.475 (0.100)</td>
<td>0.906 (0.093)</td>
<td>0.954 (0.030)</td>
</tr>
<tr>
<td></td>
<td>RMSEP</td>
<td>123000 (34900)</td>
<td>83700 (9430)</td>
<td>59100 (4990)</td>
</tr>
<tr>
<td></td>
<td>RMSEC</td>
<td>109000 (9900)</td>
<td>43500 (18300)</td>
<td>30800 (12000)</td>
</tr>
</tbody>
</table>

\(^a\) Five different test data sets, each containing 40% of the whole data set were randomly chosen for each target and this table shows the mean values from the five data sets, with the standard deviation in brackets. Factors refers to the number of latent variables used in PLS or KPLS predictions, and these are chosen by training set validation. \(Q^2\) and \(R^2\) are the correlation coefficients from the test and training data sets respectively when the DPA concentration (or spore count) is predicted by the model. RMSEP and RMSEC are the root mean squared errors in the test and training data sets respectively for the same five plots.

Although the univariate analysis is excellent (comparable to the multivariate analysis) for the pure DPA, we see that multivariate analysis is better for analysis of the spore extracts.
This is perhaps as one would expect, given that the acidic extract samples contain other components of the bacterial spores rather than just the DPA, which may adversely affect the accurate measurement of the area under the curve in the univariate analysis. The standard deviations for the univariate analysis are for the most part lower than those for the multivariate techniques, which shows that in order to develop a better quantitative model, the algorithms have needed to use subtler, and therefore variable, changes in the spectra from positions other than under the most prominent pyridine ring feature.

This is the first study to use colloidal SERS to quantify DPA extracted from spores and to attempt to relate these data to spore numbers. The limitations of the method as it stands lie in the extraction of the DPA from spore samples, and this is why the lowest number of spores detectable is considerably higher than desirable. The spectroscopic method is clearly able to detect pure DPA at the required concentrations (equivalent to 100 spores in 20 µL of sample). Once an improved extraction technique has been developed, we are confident that detection of spores will be possible at the required low levels.

2.6. Conclusion

We have shown that SERS can achieve appropriate detection levels of DPA biomarker for Bacillus bacterial spores in food safety and bioterrorism contexts using cheap, easily produced colloidal substrates and portable instrumentation. The current limiting factor of this method is the extraction of the DPA from the spores but we have shown that multivariate data analysis can significantly improve quantification in complex extracts. With improved DPA extraction, we believe this approach will be readily suitable for rapid field analysis.
2.7. References


Portable, Quantitative Detection of Bacillus Bacterial Spores Using Surface-Enhanced Raman Scattering


2.8. Supplementary Information

**Figure 2.5.** Area of the 2.0 mM glutaric acid C–H stretch peak that was used as an internal standard. The high standard deviations shown by the error bars (n=3) and lack of a general trend in the data from 0 - 100 ppb DPA shows that the peak area is not dependent on the concentration of DPA and is therefore a suitable internal standard. Above 200 ppb DPA however, a decreasing trend in the C–H stretch peak area is observed, along with much lower standard deviations, indicating that the amount of glutaric acid bound to the colloid is decreasing as it becomes saturated with DPA.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

3. An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.
3.1. Declaration

This chapter consists of one piece of work written in the format of a journal article but not currently submitted for publication:

As primary author, I carried out all of the experimental work, data analysis and write-up. Prof. Roy Goodacre, as Principal Investigator, provided the project brief and secured funding for the work to be carried out, in addition to providing advice and direction during the course of the research.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

David P. Cowcher and Royston Goodacre

School of Chemistry and Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

*Correspondence to roy.goodacre@manchester.ac.uk

3.2. Abstract

UV excitation wavelengths are useful in the analysis of biomolecules as any fluorescent background interference is avoided and the excitation energy is often resonant with electronic excitations within the analyte, creating a resonance enhancement effect. Whilst this resonance enhancement can increase the measured intensity of certain bands within the analyte spectrum, care has to be taken to avoid thermal damage to the sample as a result of the high energy of a UV laser, and therefore integration times need to be kept to a minimum.

Surface-enhanced Raman scattering (SERS) is capable of providing significant enhancements to Raman scattering intensities, but has thus far been limited to excitation at wavelengths longer than the near-UV region of the electromagnetic spectrum due to the necessary properties of the enhancing medium.

Here we show an investigation into a novel candidate for a UV-SERS substrate, boron nitride: a semiconductor material with a band gap of the appropriate magnitude to enable SERS via the chemical enhancement mechanism under deep-UV excitation. Despite
promising initial experimental results and some evidence of an interaction between the analyte rhodamine 6G and the BN substrate, no conclusive evidence of SERS enhancement was observed and the background spectrum of the substrate itself interfered significantly with the analyte spectra.

### 3.3. Introduction

First discovered in the 1970s (Fleischmann et al., 1974, Jeanmaire and Van Duyne, 1977, Albrecht and Creighton, 1977), surface-enhanced Raman scattering (SERS) is a method of increasing the signal response of conventional Raman scattering through the interaction of the incident light photons with a suitable surface in close proximity to the analyte.

The surfaces that are required to provide an enhancement are most commonly metals with nanoscale roughness features, which can be prepared in a number of ways, including chemically roughened electrodes, galvanic replacement (Mabbott et al., 2012), film-over-nanosphere (Zhang et al., 2005) and most commonly in the form of nanoparticles, of which there are many shapes, sizes and associated synthetic routes. The most popular are based on reduction of a metal salt with sodium borohydride, trisodium citrate (Frens, 1973, Lee and Meisel, 1982) or hydroxylamine hydrochloride (Leopold and Lendl, 2003). Generally, the greatest enhancements are observed on silver surfaces, followed by gold, then copper and the other noble metals. As such, the vast majority of SERS literature involves silver and gold-based methods.

Enhancement occurs in two independent but multiplicative ways. The electrochemical enhancement mechanism is capable of providing the greatest enhancements (up to a factor of $10^{14}$ in special cases (Nie and Emory, 1997, Kneipp et al., 1997)). This occurs by the
constructive interference between the electronic component of the incident electromagnetic radiation and regions of localised electron density at the nano-feature surface imparted by the same incident radiation, known as localised surface plasmons (Moskovits, 1978). The chemical enhancement method can be thought of as a variant of the resonance Raman scattering (RRS) mechanism (Hirakawa and Tsuboi, 1975): whereas with RRS the virtual excited energy level associated with the light scattering event is resonant with a real energy level in an electronic excited state of the analyte molecule, here the virtual excited energy level is resonant with a charge transfer between the surface and the analyte or vice versa. This is explained in more detail in the review by Campion and Kambhampati (Campion and Kambhampati, 1998).

Higher energy lasers give higher scattering intensities (Lewis and Edwards, 2001) but until the wavelength approaches the ultraviolet (UV) part of the spectrum, fluorescence interference becomes a limiting factor for some analytes. SERS is able to quench fluorescence significantly, but it does not eliminate it entirely. As SERS is essentially brought about by a resonance of one form or another, the magnitude of the SERS enhancement is highly dependent on the incident laser wavelength and for the vast majority of metal SERS surfaces, the appropriate laser wavelengths fall within the visible to near infrared (NIR) region of the spectrum. As an alternative SERS approach, interest has therefore developed in preparing SERS-active surfaces that provide enhancements into the UV region, away from any fluorescence excitation (Ren et al., 2003, Yang et al., 2011).

Non-metallic SERS-active surfaces are also becoming an increasingly popular area of investigation, including carbon materials such as graphene (Ling et al., 2009) as well as semiconducting materials including silicon and germanium-based structures (Wang et al., 2011). The latter are unable to support localised surface plasmons due to their lack of free
electrons that would be found in a conducting material and are therefore only able to promote SERS by the chemical enhancement mechanism.

In this study we used boron nitride, a semiconducting material with a band gap of 5.96 eV (Evans et al., 2008), which corresponds to a laser wavelength of 208 nm, to investigate whether it can be used to promote SERS enhancement by the chemical enhancement method when the incident radiation is from a UV laser at 244 nm.

Hexagonal boron nitride has a structure similar to, and isoelectronic with, graphite, existing in layers of alternating boron and nitrogen atoms that form a hexagonal lattice (Dean et al., 2010). Boronic compounds are known for behaving as Lewis acids (Housecroft and Sharpe, 2005) and at the BN surface, there is the potential for the electron-deficient boron to accept electron pairs from analytes with electron rich functional groups, such as carboxylates and alcohols. The lone pairs of electrons on the surface nitrogen atoms may also permit electron pair donation to electron poor analytes such as positively charged species. Bonding interactions between the BN surface and the analyte are important for chemically-enhanced SERS to occur.

3.4. Experimental

3.4.1. Reagents

Boron nitride refractory brushable paint (31% BN as supplied, 94% BN when dry) was obtained from Alfa Aesar, Heysham, UK. DL-phenylalanine (98.5%) was obtained from BDH chemicals, Poole, UK. Rhodamine 6G (95%), crystal violet (90%), DL-tyrosine
(98%) and L-tryptophan (98%) were all obtained from Sigma-Aldrich, UK. Graphite sticks and chalk sticks were purchased from a local art supplies store.

3.4.2. Preparation of surfaces and samples:
Boron nitride refractory brushable paint was diluted with deionised water to make it less viscous than as it was supplied. The diluted paint was dropped onto a standard glass microscope slide using a disposable pipette until it covered the whole surface of the slide. The slide was then left overnight to dry. Standard white drawing chalk was filed down from one side to give a flat surface (semicircular cross section). Graphite sticks were used as supplied.

Rhodamine 6G and crystal violet were dissolved in deionised water at a concentration of 1.00 mM and then an aliquot of each was diluted to 0.100 mM. The three aromatic amino acids were each dissolved in deionised water at a concentration of 0.05 M.

25 µL of each sample was spotted onto the relevant surface and allowed to dry before analysis.

3.4.3. Instrumentation and measurement:
Raman spectra were recorded using a Renishaw Raman microscope (Renishaw, Wotton-under-edge, UK), equipped with a Lexel Model 95 Argon Ion Laser, emitting at 244 nm and providing ~0.2 mW at the sample.

The samples were analysed on a rotating stage (~3 rpm) built in-house so that the sample spot was in constant motion under the microscope objective, thus minimising heating and the risk of damage to the sample by the laser.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

3.4.4. Data analysis:

All data analysis was performed in Matlab R2008a (The Mathworks, Natick, MA, USA) using scripts written in-house. Where spectra have been baseline-corrected, this has been done using the asymmetric least-squares (ALS) method (Eilers and Boelens, 2005).
3.5. Results and Discussion

Rhodamine 6G (R6G) has become a standard marker for SERS research due to the fact it is highly fluorescent in most of the visible region of the electromagnetic spectrum and only reveals its intense and characteristic spectrum when it undergoes the fluorescence quenching and signal enhancement effects of SERS. Its hydrochloride structure is shown in Figure 3.1a and has the potential to bond with the BN surface via the positively charged nitrogen species, but more likely through electron donation from the carbonyl oxygen to the Lewis acidic boron. Here we use R6G as our primary analyte to be in keeping with the norm, although our excitation wavelength of 244 nm is too low to promote fluorescence in this case. This means that whilst a typical R6G spectrum would no longer necessarily indicate a SERS event, we are in the unusual position of being able to compare the UV-Raman spectrum of R6G powder with that of our potential SERS system, as shown in Figure 3.2.

![Figure 3.1](image1.png)  
**Figure 3.1.** Structures of the hydrochloride salts of (a) rhodamine 6G and (b) crystal violet.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

**Figure 3.2.** Mean (n=3) UV Raman spectra of R6G powder (red) and 1.00 x 10^{-3} M aqueous solution of R6G spotted onto a BN surface (blue) showing a shift in the peak at ~1200 cm^{-1} as indicated by the dashed vertical red lines. In black is the spectrum of the BN surface. These spectra have been scaled between 0 and 1 on the intensity axis.

These initial results were encouraging: although there is a significant background peak from the BN itself at 1376 cm^{-1}, the peaks of the R6G spectrum seem better resolved and very intense given that there are vastly fewer molecules of R6G being interrogated on the BN slide than in the powder sample. There is also a measureable shift in the xanthene ring deformation peak (Jensen and Schatz, 2006) from 1197 cm^{-1} to 1213 cm^{-1}, which is an indication of an interaction between the R6G and the BN, such as we would expect to observe during SERS enhancement.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

In order to assess whether it was indeed SERS we were observing, two control surface substrates were used. Graphite is isoelectronic with BN but is an electrical conductor and therefore has no band gap. R6G samples were prepared on graphite surfaces in order to try and ascertain whether the semiconducting properties of the BN are useful in promoting SERS activity. Chalk (calcium carbonate) is an electrical insulator and would not be able to convey a SERS enhancement; however it has a similar porous structure to the BN surface and was used to eliminate the greatly increased surface area of the porous substrate as the factor responsible for the increased Raman signal intensity. Figure 3.3 shows a comparison of the UV-Raman spectra of 1.00 x 10^{-4} M R6G solutions spotted onto the different substrates.

Both the chalk and the graphite have their own characteristic background peaks at 1040 cm\(^{-1}\) and 1170 cm\(^{-1}\) (chalk) and 1610 cm\(^{-1}\) (graphite), which have the potential to mask peaks from the R6G, although the intensity of the R6G signal was much higher for the BN substrate than for either the graphite or the chalk, particularly the xanthene ring deformation (1200 cm\(^{-1}\)) and the various xanthene ring stretching vibrations (1500-1660 cm\(^{-1}\)) (Jensen and Schatz, 2006). This was again a positive indication that a SERS enhancement is being observed, although there were no apparent peak shifts between the R6G spectra on the BN and on the other substrates, meaning at this stage it was not able to say conclusively whether this is the result of SERS or another effect, such as increased surface area.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

Figure 3.3. Mean \((n=3)\) UV Raman spectra of (a) BN (blue), chalk (red) and graphite (black) surfaces, and (b) \(1.00 \times 10^{-4}\) M R6G solution spotted on the same surfaces.

Another commonly used analyte in the demonstration of SERS is crystal violet, whose structure is shown in Figure 3.1b. This dye molecule again has an intense SERS spectrum and a good affinity for metal surfaces, which makes it ideal for nanoparticle-based SERS measurements. Whilst this is not particularly relevant in this case, it would allow for comparison with existing literature.

Samples of crystal violet were prepared and analysed in the same manner as R6G, with example spectra shown in Figure 3.4. As was observed with R6G, peaks associated with the crystal violet are visible on the BN and chalk substrates, although the difference in intensity between the spectra is less pronounced. On the BN substrate, the background
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

signal masks the in-plane aromatic C-C vibrations of the crystal violet at 1600 cm\(^{-1}\) to some extent. On the chalk substrate, these same vibrations, as well as the C-H bending modes between 1000 cm\(^{-1}\) and 1200 cm\(^{-1}\), are also coincident with peaks from the surface material, meaning that neither substrate displays any useful enhancement of the crystal violet features (Lai et al., 2011).

![Figure 3.4.](image)

**Figure 3.4.** Mean Raman spectra (\(n=3\)) of crystal violet (a) 1.00 mM on BN, (b) 1.00 mM on chalk, (c) 1.00 mM on graphite and (d) powder. Dashed lines show the spectra of the bare surfaces. The tops of the peak from the BN background signal is cut off to make the crystal violet peaks easier to see.

The combination of this and the fact that no shifts in the crystal violet peak positions are observed mean there is no evidence that BN is creating a chemical-bond SERS effect here, which might be to be expected, given that the only means of interaction with the BN
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

The surface would be between the positively charged nitrogen, which is hindered by the two methyl groups (Figure 3.1b), and the lone pair of a surface nitrogen. Crystal violet peaks are not definitively observed at all on the graphite substrate but this is likely to be due to the fact the graphite is less porous than the other two substrates and there will be a smaller surface area of sample exposed to the incident light.

In order to investigate the BN substrate with some biochemically relevant samples, 50.0 mM solutions of the three aromatic amino acids, phenylalanine, tryptophan and tyrosine, all three of which have electron rich groups capable of interacting with the BN surface, were prepared, spotted onto the BN surfaces and analysed. The mean spectra are shown in Figure 3.5.

While there is some evidence of features in the spectra of the samples on the BN surfaces that correspond to peak positions in the crystalline reference samples, these are limited to very small, broad features and are not distinctive peaks that could be used for identification or quantitative analysis. It is unclear whether this is simply due to the strong background signal from the BN itself or whether the analyte concentration on the surface is too low for a spectrum to be observed. Either way, there is no evidence that the BN substrate is providing any significant SERS enhancement with these analytes.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

Figure 3.5. (a) Mean ($n=3$) Raman spectra of 50.0 mM solutions of the amino acids spotted onto BN surfaces: phenylalanine (solid blue line), tryptophan (dashed red) and tyrosine (dot-dashed black). The spectrum of the bare BN surface is shown in yellow. The top of the BN peak is cut off in each case to make the amino acid features clearer. (b) Mean ($n=3$) Raman spectra of the crystalline amino acids on a glass slide: phenylalanine (solid green line), tryptophan (dashed purple) and tyrosine (dot-dashed grey).

With no evidence of a SERS effect being observed on the BN substrate with any analyte except for the R6G, the R6G analysis was repeated to examine whether the apparent SERS effect was reproducible. The results of the repeat analysis are shown in Figure 3.6.

Similar to the results of the first experiment, the chalk and BN substrates gave significantly higher signal intensity than the graphite substrate but the differences between the relative intensities of the R6G peaks on the chalk and BN showed no significant difference. There was also no apparent shift in the R6G peak position between the two substrates.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

**Figure 3.6.** Repeat UV-Raman analysis of R6G (a) mean spectra \( (n=3) \) of 25 µL spots of 1.00 mM solution on BN (solid blue), chalk (dashed red) and graphite (dot-dashed black); (b) scaled mean spectra \( (n=3) \) of R6G powder (solid blue), and 25 µL spots of 1.00 mM R6G solution on BN: original data set (dot-dashed black) and repeat dataset (dashed red).
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

Peak shifts were observed however between the two repeat analyses of R6G on the BN substrates, as shown in Table 3.1. The shift of the B-N vibration from the substrates of the respective samples (19 cm⁻¹) is the maximum error that could reasonably be assigned to calibration error, although care was taken to ensure the instrument was accurately calibrated before each set of measurements was taken by using a diamond calibration standard. Several of the peak shifts are in excess of this and furthermore, the lack of consistency between the shifts relating to different peaks indicates that this is not merely a calibration error and that some interaction between the analyte and substrate has occurred.

Table 3.1. Peak shifts in the spectra of R6G spots on the BN substrate between the two repeat measurements and R6G powder.

<table>
<thead>
<tr>
<th>Tentative peak assignment</th>
<th>A/cm⁻¹</th>
<th>B/cm⁻¹</th>
<th>C/cm⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out-of-plane C-H bend</td>
<td>843</td>
<td>-19</td>
<td>-15</td>
</tr>
<tr>
<td>In-plane xanthene ring deformation</td>
<td>935</td>
<td>-27</td>
<td>-9</td>
</tr>
<tr>
<td>In-plane xanthene ring deformation, C-H bend, C-N bend</td>
<td>1213</td>
<td>-23</td>
<td>-16</td>
</tr>
<tr>
<td>B-N stretch (boron nitride)</td>
<td>1376</td>
<td>-19</td>
<td>N/A</td>
</tr>
<tr>
<td>Xanthene ring stretch, in-plane CH-bend</td>
<td>1449</td>
<td>-27</td>
<td>-12</td>
</tr>
</tbody>
</table>

b All peaks are from the R6G unless indicated otherwise. Assignments from Jensen and Schatz (Jensen and Schatz, 2006). A: peak position in 1.00 mM sample on original BN substrate. B: relative peak position in 1.00 mM sample on repeat BN substrate. C: relative peak position in R6G powder sample.

Whilst evidence of an interaction between the R6G and BN substrate is apparent, the enhancement factor is not significantly large when compared to other porous substrates.
An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.

such as chalk. Even if it could be proven conclusively that this interaction lead to a SERS enhancement, the fact that the BN peak itself is such a prominent feature in the spectrum and appears in an important region for biological sample characterisation, would severely limit the usefulness of BN as a suitable substrate for SERS analyses of biological samples.

3.6. Conclusion

These results have shown evidence of an interaction between BN substrates and R6G, manifested by shifts in the relative peak positions in the Raman spectra of R6G on the substrates compared with R6G powder. There is also limited evidence of a SERS effect via the chemical enhancement method, but this has not been possible to confirm.

While this would suggest BN substrates could be used to enhance the Raman signal intensity from biological samples, the presence of the intense, broad B-N stretching peak in the region around 1370 cm⁻¹, combined with a higher baseline intensity when using the BN substrates, interferes to the extent that the Raman peaks from relatively concentrated solutions of aromatic amino acids spotted onto the surface are barely observable. This negates the usefulness of the substrates for such analyses.
3.7. References


An Investigation of boron nitride as a substrate for surface-enhanced Raman scattering at 244 nm excitation.


4. The detection of protein glycosylation using TERS
4.1. Declaration

This chapter consists of one journal article, currently pending submission:


This article has been reproduced in an unchanged format except for minor adjustments to incorporate it into this thesis.

As primary author on this publication, I carried out all of the protein sample preparation, data analysis and write-up. Dr. Tanja Deckert-Gaudig was responsible for all of the experimental TERS measurements, gold nanoplate synthesis and sugar sample preparation, as well as providing advice on data interpretation. Dr. Victoria Brewster provided assistance and advice regarding sample handling, analysis and interpretation of protein Raman data based on her previous work. Dr. Lorna Ashton obtained the funding for the work to be carried out and provided assistance with data interpretation. Prof. Volker Deckert, as Principal Investigator of the research group at University of Jena where the experimental work was conducted, provided the instrumentation and other laboratory equipment that enabled the work to be completed. Prof. Roy Goodacre, as Principal Investigator of my research group in Manchester, provided the project brief and secured additional funding for the work to be carried out, in addition to providing advice and direction during the course of the research.
The detection of protein glycosylation using TERS

David P. Cowcher¹, Tanja Deckert-Gaudig², Victoria L. Brewster¹, Lorna Ashton¹, Volker Deckert²,³ and Royston Goodacre¹*.  

¹ School of Chemistry and Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

² Institute of Photonic Technology – IPHT, Albert-Einstein- Str. 9, 07745 Jena, Germany

³ Institut f. Physikalische Chemie, and Abbe Center of Photonics Friedrich-Schiller Universität, Helmholtzweg 4, 07743 Jena, Germany

*Correspondence to roy.goodacre@manchester.ac.uk

4.2. Abstract

The correct glycosylation of biopharmaceutical glycoproteins and their formulations is essential for them to have the desired therapeutic effect on the patient. It has recently been shown that Raman spectroscopy can be used to quantify the proportion of glycosylated protein from mixtures of native and glycosylated forms of bovine pancreatic ribonuclease (RNase). Here we show the first steps towards not only the detection of glycosylation status, but the characterisation of glycans themselves from just a few protein molecules at a time using tip-enhanced Raman scattering (TERS). Whilst this technique generates complex data that are very dependent on the protein orientation, with the careful development of combined data preprocessing, univariate and multivariate analysis
techniques, we have shown that we can distinguish between the native and glycosylated forms of RNase. Many glycoproteins contain populations of subtly different glycoforms, therefore with stricter orientation control, we believe this has the potential to lead to further glycan characterisation using TERS, which would have use in biopharmaceutical synthesis and formulation research.

4.3. Introduction

It is estimated that glycoproteins account for almost two thirds of all protein species (Apweiler et al., 1999) and with the level of research and investment into protein-based therapeutic products ever increasing (Hamrang et al., 2013), the accurate characterisation of post-translational modifications (PTMs) is vital for therapy. In particular, determining glycosylation status and glycan structure is becoming an important area of analytical science due to the potential adverse drug reactions for incorrect formulations and the need to have the correct protein glycoform for efficacious therapy.

In addition to the need for protein therapeutics to maintain the correct secondary and tertiary structure from the point of manufacture to their intended point of interaction with the patient, glycoproteins need to have the correct glycan attached in the correct place in order to function as intended. Incorrect glycosylation may result in misfolding, attenuation of efficacy as a result of compromised sorting/directing, ligand binding, biological activity, plasma half-life, stability and immunogenicity (Walsh and Jefferis, 2006).

Raman spectroscopy is an increasingly popular analytical tool in the field of biomedicine (Ellis et al., 2013) and has been frequently used for the characterisation of biopharmaceuticals (Wen, 2007). Raman has particular advantages in that it is non-
The detection of protein glycosylation using TERS

destructive and can be applied through a transparent window into a vessel, giving it the potential for online use in the analysis of a dynamic system (Marison et al., 2013). It has been used to characterise glycoproteins in the past, (Kopecky et al., 2003, Cui et al., 2005), as well as Raman optical activity (ROA) (Bell et al., 1994, Zhu et al., 2005, Johannessen et al., 2011), and including spectral differences as a result of chemical deglycosylation (Kikuchi et al., 1987). A combination of Raman spectroscopy and chemometric techniques has enabled the distinction between native and glycosylated forms of bovine pancreatic ribonuclease (RNase) protein, including quantification of the relative amounts of each form from mixtures (Brewster et al., 2011). The quantitative detection of glycated haemoglobin using Raman spectroscopy has also been demonstrated (Barman et al., 2012).

Even though Raman microscopy (with a typical interrogation diameter of 1 µm) is a confocal technique, the number of molecules within the laser focus is still vast and the spectra recorded are thus ensemble averages of multiple layers of protein molecules. While the majority of these molecules may be correctly glycosylated, if a small percentage (<5 %) are not, it is highly unlikely they would be detected (Brewster et al., 2011). From a formulation point of view, there is also no information about the specific error in the glycosylation, for example whether the protein is glycosylated in the wrong place, or whether the incorrect glycan is attached. Lazar et al. have constructed a useful review of the current progress in the analysis of glycoproteins and their glycans, particularly in relation to mass spectrometry (Lazar et al., 2013). However, a rapidly advancing variant of Raman spectroscopy, Tip-enhanced Raman scattering (TERS), may offer an alternative solution. TERS, like its more widely known relation surface-enhanced Raman scattering (SERS), exploits the phenomenon of metals with appropriately (nano)-sized features being able to localise regions of charge density at their surface when illuminated by appropriate
electromagnetic radiation. These so-called surface plasmons are able to interact constructively with the electric field component of the incident light, resulting in an enhancement of many orders of magnitude to the light intensity, thus like SERS significantly amplifying the intensity of Raman scattered photons (Moskovits, 1978). Whilst SERS uses nano-patterned surfaces or colloids to enhance microscopic or bulk-phase measurements (Smith, 2008), TERS utilises a single nanoparticle, mounted on the end of the tip of a scanning probe microscope, such as an atomic force microscope (AFM). The tip is scanned across a sample and provides the spatial resolution of AFM whilst collecting the vibrational structural information of an enhanced Raman measurement (Bailo and Deckert, 2008a).

TERS has already been used in a number of high-resolution biological applications; for example in the analysis of live bacterial cells (Neugebauer et al., 2006, Neugebauer et al., 2007) and the investigation of individual nucleic acid strands (Bailo and Deckert, 2008b); the reader is also directed to a nice review by Treffner et al. (Treffer et al., 2012) for further application areas.

In this study we set out to explore the potential of TERS to probe protein glycosylation at a near-molecular level by immobilising a monolayer of protein molecules on a surface and taking TERS measurements. The high spatial resolution of the technique has the potential to detect subtle structural variations that indicate incorrect glycosylation; variations that would very likely be masked by the ensemble averaging effect of normal Raman microscopy. Here we show the first steps towards realising this goal using the same simple RNase model system that was employed with conventional Raman spectroscopy (Brewster et al., 2011). We show for the first time that TERS can be used to distinguish successfully
between glycosylated and non-glycosylated proteins from the measurements of just a few molecules within a monolayer.

4.4. Experimental

4.4.1. Reagents

Ribonuclease A and B from bovine pancreas (lyophilised powder), D-mannose and N-acetyl-D-glucosamine (GlcNAc) were used as supplied from Sigma Aldrich Ltd. (Dorset, UK).

4.4.2. Protein sample preparation

Ultraflat gold nanoplates were prepared according to the method described by Deckert-Gaudig and Deckert (Deckert-Gaudig and Deckert, 2009). The proteins were dissolved in deionised water at a concentration of 0.25 g L\(^{-1}\) (18.2 µM). Glass coverslips with gold nanoplate-coated surfaces were soaked in the protein solution for 19 h to enable the protein molecules to immobilise on the gold. The coverslips were then removed from the solution, rinsed three times with water and dried under vacuum.

4.4.3. Instrumentation and TERS measurement

NT-MDT non-contact AFM tips were evaporated with 25 nm silver and stored under argon till used. AFM images were acquired on a Nanowizard I (JPK Instruments AG, Germany), mounted on an inverted microscope (Olympus IX70, Japan) and equipped with an oil immersion objective, 60×, numerical aperture (NA) = 1.45 (Olympus, Japan). The objective was piezo-mounted (PIFOC, Physik Instrumente, Germany), thereby adjusting the laser focus to the z-movement of the TERS tip.
For excitation, a krypton ion laser was used (Innova 300c, wavelength 530.9 nm, U.S.A.). Spectra were collected by a confocal Raman spectrometer (LabRam HR, Horiba Jobin Yvon, France) with a liquid nitrogen-cooled CCD camera (ISA Spectrum One, Horiba Jobin Yvon, France). The power on the sample was 820-850 µW.

The lateral steps of the tips were controlled by an additional 100 µm × 100 µm sample scanning stage (P-734, Physik Instrumente, Germany). Spectra were recorded with acquisition times of 2 s, 5 s or 10 s (depending on the tip). After each grid measurement a reference spectrum next to the gold plate was acquired to exclude tip contamination. Grids of various sizes with point-to-point distances of 2-10 nm were set on the gold nanoplates. In total, 1977 spectra were recorded for the RNase A sample and 2345 spectra for RNase B, respectively.

Control samples of the individual sugars that make up the glycans on RNase B, mannose and GlcNAc, were also prepared and measured. 1.5 µL of 1.00 mM solutions of the sugars were dropped onto cleaned glass slides and dried under vacuum. Raman spectra were recorded from large micrometre sized crystals. On areas appearing ‘clean’ under the microscope, TERS spectra were also recorded. The nanometre sized crystals were approximately 100-300 nm in width and 10-20 nm in height. The point-to-point distance for the measurements was mostly 5 nm. It was more difficult to obtain good spectra from mannose than GlcNAc and a total of 489 and 2650 spectra were recorded from each sugar respectively.
4.4.4. Data analysis

Data analysis was carried out using Matlab version R2013a (The Mathworks, Natick, MA, USA). First, the spectra that showed no features except for the silicon background were removed from the data set. Cosmic rays were removed by inserting a linear fit between the start and end positions of the peak. The spectra were subsequently baseline-corrected using the asymmetric least-squares (ALS) method (Eilers and Boelens, 2005) with the smoothing parameter, $\lambda = 1000$ and the asymmetry parameter, $p = 0.001$.

Three different approaches were applied independently to negate accurately the influence of the Si peaks from the TERS tip:

(i) A Gaussian was fitted under the Si fundamental peak at 520 cm$^{-1}$ and the parameters that described this Gaussian (centre position, width and height) were used to plot a predicted first overtone peak. The fitted fundamental Gaussian and predicted overtone Gaussian were then subtracted from the relevant part of each spectrum.

(ii) Spectra of the gold nanoplates without protein were averaged (mean) and then each spectrum was normalised against this mean.

(iii) A secondary ALS baseline correction was applied to the region of the spectrum containing the Si overtone feature, with the parameters $\lambda = 100$ and $p = 0.01$. The region of the spectrum containing the Si fundamental peak was set to zero.

For each of the three datasets now with the Si spectrum removed, each spectrum was row normalised to 1 of its total signal.

At each stage of the data processing, principal components analysis (PCA) was performed on the data, with 10 principal components (PCs) projected in each case.
### 4.5. Results and Discussion

The RNase model proteins were chosen for this work as both the native form, RNase A, and the glycosylated form, RNase B, are available commercially in high purity. More importantly, as we have measured them using Raman microscopy, we also had a solid background model in place for comparison (Brewster et al., 2011). The inset in Figure 4.1 shows the structure of the protein and its associated glycan. Raman and TERS spectra of the individual sugar monomers are also provided in Figure 4.1.

Ultraflat gold nanoplates (Deckert-Gaudig and Deckert, 2009) were chosen to immobilise the RNase proteins for TERS analysis as they are suitably level for AFM measurement, transparent to allow TERS measurement through the sample, and will bind effectively to the disulfide bridges in the RNase proteins. An example of the typical gold nanoplate topography is shown in Figure 4.7. The diameter of the nanoparticle at the end of the TERS tip is approximately 20 nm, which means each measurement will interrogate just a few protein molecules at a time, assuming a RNase diameter of 3.8 nm (Ramm et al., 1985, Deckert-Gaudig et al., 2012).

TERS measurements were performed as previously described (Deckert-Gaudig and Deckert, 2009). At first the TERS tip was scanned through the laser spot and positioned at the location giving the highest reflection in the optical response image. After switching from tip-scanner to sample-scanner mode the topography was scanned and an appropriate nanoplate was selected. TERS spectra were recorded by way of grids on 1-10 nm distant points. The acquisition time depended on the enhancing ability of the respective tip and was set to 1s, 5 s or 10 s.
Figure 4.1. (a) Raw TERS spectra of RNase A from one grid with examples of cosmic rays highlighted by asterisks. (b) A typical background TERS spectrum with peaks from the Silicon tip highlighted with arrows. Also shown are examples of the (c) TERS and (d) Raman spectra of mannose (solid line) and N-acetylglucosamine (dashed line), the sugars that make up the glycan of RNase B. The TERS spectra (c) have had the silicon background removed. Inset shows a cartoon representation of the native form of bovine RNase drawn from atomic coordinates in the PDB (5RSA) using PyMOL. Also shown is the glycosylation point (Asn34 residue) and the RNase B glycan. Optional mannose refers to the variation in number and possible arrangements of mannose between the different glycoforms that occur in RNase B.
Data preprocessing is often necessary with Raman spectra so that one can extract the useful information from the background signal, especially when enhancement techniques such as SERS and TERS are used to probe analytes that are weak Raman scatterers, such as sugars (Walton and Blackwell, 1973), very low concentration, or as in this case, both. This was particularly true here, as the raw data from the TERS measurements were highly variable. Even between spectra that were from the same grid, for both the sugar monomers and the protein samples certain points gave no significant TERS response due to lost feedback and amongst those that did, there was seemingly little consistency from spectrum to spectrum in terms of peak position. Variations in band position and intensity are common observations in TERS and can be ascribed to the small number of molecules interacting with the tip and thereby inducing change of orientation (Ichimura et al., 2007, Treffer et al., 2011).

This was, however, not entirely surprising. The protein sample concentration was kept low to deposit molecules at almost monolayer coverage onto the gold nanoplates; the aim being to get as close as possible to looking at just one protein molecule at a time. We might expect to see more peaks in the TERS spectrum of a sample compared to the Raman due to the relaxation of selection rules (Berweger and Raschke, 2009), resulting in more allowed vibrational modes. For example, we believe this to be why strong peaks were seen at ~1575 cm\(^{-1}\) in the TERS spectra of crystalline mannose, but not in the Raman (Figure 4.1c and d). Also, the vastly greater spatial resolution of TERS compared to normal Raman microscopy meant that instead of getting an ensemble averaging effect from the varying peak positions due to the sugars’ and proteins’ random orientation, we observed sharper peaks that are more susceptible to shifts as a result of differing orientation to the nanoparticle surface on the end of the AFM tip. This variability can be clearly observed in
Figure 4.1a, which shows the raw spectra from just one grid of data, and was also observed when the crystalline glycan monomers were measured, hence mean spectra are shown in Figure 4.1c.

When comparing the protein samples, one of the main regions of interest was between 800 – 1100 cm\(^{-1}\), where peaks from the sugar molecules are expected but not strong peaks from the proteins. Despite this, the inconsistency of peak positions within such a large dataset meant it was not immediately possible to identify individual peaks that would be indicative of glycosylation status by visual inspection only. PCA was therefore used to analyse the spectra as a whole in an attempt to extract useful information.

In order to minimise the influence of random environmental factors on the PCA analysis, cosmic rays (such as those indicated by asterisks in Figure 4.1a) were removed and the spectra were baseline corrected to eliminate any baseline drift and minimise fluorescence interference. PCA was then performed on the data and the resultant scores plots are shown in Figure 4.2b and c. From these PCA scores plots of PC 1 against PC 2 (Figure 4.2b), there appeared to be some clustering of spectra according to the different proteins, particularly in PC 1; whilst PC 2 seemed to show some variation within certain sub-groups of these clusters. When the loadings were plotted however (Figure 4.2c), it was apparent that this separation, particularly in PC 1, was mainly due to the Si fundamental peak from the TERS tip at 520 cm\(^{-1}\) and not from any Raman features associated with the protein. In fact, PCA was able to group the samples from each respective measurement grid quite well based mainly on the features associated with the Si tip, revealing the subtle variation between measurement grids as a result of the background signal, which would need to be accounted for to gain reliable results. These data are shown in the Supplementary Information (Figure 4.8).
The detection of protein glycosylation using TERS

**Figure 4.2.** Following cosmic ray removal and baseline correction by the asymmetric least squares method, (a) a typical RNase A spectrum, (b) PCA scores plots of all of the RNase A (blue circles) and RNase B (red squares) with % explained variance in brackets and (c) PCA loadings of PC1 (solid purple) and PC2 (dashed green). Following the above, (d)-(f) spectra normalised against the mean blank spectrum and (g)-(i) after the Si overtone was removed by further asymmetric least squares baseline correction (see supplementary information Figure 4.9) and the region containing the Si fundamental set to zero. Parts (d) and (g) are example spectra from the same measurement. (e) and (h) are PCA scores plots of all of the RNase A (blue circles) and RNase B (red squares) with % explained variance in brackets. Insets show a zoomed-in view. Parts (f) and (i) are PCA loadings plots of PC1 (solid purple) and PC2 (dashed green).
An example of the background spectrum of the gold nanoplates is shown in Figure 4.1b and it is clearly dominated by fundamental Si-Si stretch from the silicon tip. The Si fundamental peak was so intense that it masked any protein or carbohydrate peaks in the same region. The Si first overtone peak at ~960 cm\(^{-1}\) appeared as a small but broad increase in the baseline and is unfortunately located in the information-rich part of the spectrum where we might expect to see differences between the protein samples, with sharper sample peaks often visible on top. This made the removal of the Si overtone somewhat more difficult without adversely affecting the useful peaks in the same spectral region. It was therefore important to accurately remove this overtone and the three different approaches detailed below were sequentially evolved to reduce the influence of the Si peaks on the PCA model.

The first method involved fitting Gaussians under both the Si fundamental and first overtone peaks of each spectrum. The centre position, width and height of each were recorded and the ratio of each parameter between the fundamental and overtone features were calculated. The median value of the ratios for each parameter was then used to plot a predicted first overtone peak based only on the properties of the measured fundamental peak, so as to remove the influence of any other peaks that may occupy the same region of the spectrum as the overtone. The median was chosen over the mean to negate the influence of a small number of extreme outliers. The fitted fundamental Gaussian and predicted overtone Gaussian were then subtracted from the relevant part of each spectrum.

The predicted overtone peaks fitted the general shape of the baseline very well (data not shown) but as there were overlapping features either side of this background peak, it was difficult to match up the predicted Gaussian region accurately with the continuing
spectrum, leading to the insertion of false features. As such, the Gaussian-fitting method was not considered reliable.

The second method used to remove the Si features was to normalise each spectrum against the mean of the blank spectra; that is to say, the spectra of the gold nanoplates without protein sample present. The idea was to be able to remove the background peaks in a way that did not destroy any analytically useful protein information in their immediate vicinity. A typical spectrum before and after this correction procedure is shown in Figure 4.2a and d, where it can be seen that the intensity of a number of features in the spectrum have been emphasised as a result of this processing. The PCA plot (Figure 4.2e) is now completely different to that generated previously (Figure 4.2b), although the loadings plot (Figure 4.2f) still shows some influence from the Si fundamental peak. This is likely caused by the position of the same peak in the blank spectrum not matching exactly that of the sample, as evident in Figure 4.2d; however, this Si fundamental vibration is no longer the dominant feature in PC 1 and other peaks at higher wavenumber are having greater influence.

The final method was intended to remove the influence of the Si fundamental from the PCA model entirely whilst maintaining features in the region of the Si overtone. As such, the region containing the Si fundamental peak was set to zero and whilst there was the potential to also remove other peaks in the same region, the complete removal of the Si peak was considered more beneficial. A second ALS baseline correction was applied to the region containing the overtone peak with the parameters set to fit the baseline very closely to the original data. Due to the broad nature of the overtone peak, the fitted baseline follows its shape reliably, but does not fit too closely to any sharper features that appear on top. An example of this baseline fitting is shown in Figure 4.9 (Supplementary Information).
The detection of protein glycosylation using TERS

The example spectrum used previously is shown following this data processing, along with the subsequent PCA scores and loadings plots in Figure 4.2g-i respectively. Features of the spectra are not emphasised as strongly as when they are normalised against the blank but the core cluster of the PCA scores plot shows a similar shape and the corresponding loadings show that very similar parts of the spectrum are influential; with the exception of the Si fundamental which has of course been removed.

Despite the apparent success in minimising the influence of the silicon background, PCA had still not conclusively identified regions of the spectra where differences were observable between the two forms of the protein; as evidenced by the lack of separation of RNase A (blue circles) and B (red squares) in Figure 4.2e and h. One other area that had not been considered thus far was the enhancement factor of the TERS measurements. Depending on the proximity and position of the TERS tip relative to the protein molecules whose spectra were being observed, the magnitude of the TERS enhancement will have been different for each measurement. This was very apparent from the individual spectra, which differed greatly in terms of maximum intensity and this may have inadvertently influenced the PCA scores plots in Figure 4.2. Thus after the removal of the Si features in the TERS spectra, data were normalised to their own total signal intensity (each spectrum was row normalised to 1 of its total signal).

The PCA scores and loadings following this row normalisation are shown in Figure 4.3a and b for the blank-normalised data and in Figure 4.3c and d for the ALS-corrected data. The fact that the PCA scores plots are now roughly circular in shape with no obvious clustering indicates that this normalising step has been successful. However, with this bias removed, PCA did not show any separation between the two different forms of RNase protein, even when higher PCs were investigated (data not shown). The fact that the
explained variance for PC1 is < 10% indicates that the spectra are too variable for PCA to be useful, with no similarities between spectra, even from features unrelated to the glycosylation status.

**Figure 4.3.** PCA plots of all data following normalisation against total signal. (a) and (c) scores plots with RNase A in blue circles and RNase B in red squares, with % explained variance in brackets. (b) and (d) loadings plots of PC1 (solid purple) and PC2 (dashed green). (a) and (b) following the normalisation against blank Si and (c) and (d) following the ALS Si-removal route.
With no obvious markers (Raman features/bands) identifiable by eye or through the use of multivariate PCA, despite strategic processing of the data, a more general approach was taken. In the previous work by Brewster et al. using conventional Raman spectroscopy (Brewster et al., 2011), after multivariate data processing using multivariate supervised learning (partial least squares), six regions of the Raman spectra were identified that were significant in separating the two RNase types. Therefore the area under the curve (AUC) of these same regions was measured in the pre-processed TERS spectra (in this example using the normalisation against the blank method for Si peak removal), and the results represented in a box and whisker plot in Figure 4.4. In general, the glycosylated form of the protein has a higher AUC for the regions 850-900 cm\(^{-1}\), 1220 – 1300 cm\(^{-1}\), 1420 – 1490 cm\(^{-1}\) and 1700 – 1800 cm\(^{-1}\) and a lower AUC for 780 – 820 cm\(^{-1}\) and 950 – 1000 cm\(^{-1}\). Assignments for these regions are shown in the Table 4.1, where it is clear that the TERS was in agreement with protein conformational changes rather than the detection of the sugar per se.

Figure 4.5 shows box and whisker plots of the AUC for other regions of the TERS spectra that are not expected to contain any glycan specific information (Brewster et al., 2011). With the exception of the 300 – 780 cm\(^{-1}\) region that includes any artefacts from the removal of the Si fundamental peak, the AUCs show little difference between the two forms of the protein (the two medians are very close as are the IQRs), indicating that we may indeed be able to classify the proteins based on features within the ‘sugar regions’ of the spectra.
Figure 4.4. Box and whisker plot of the area under the preprocessed TERS spectra in the wavenumber regions indicated that showed differences between the two protein glycoforms using normal Raman spectroscopy. For each region, the shaded box and whisker represents RNase B. In each case the blue box represents the interquartile range (IQR), the red line the median, and the black dashed whiskers the remaining data, with the exception of any data that lie outside the IQR by more than 1.5 x IQR (±2.7σ), which are shown by red crosses. Vibration assignments for the regions are given in Table 4.1. In this example, the data were preprocessed using the normalisation against blank Si.
Table 4.1. Tentative Raman band assignments for the regions of the TERS spectra that indicate differences between RNase A and RNase B (Brewster et al., 2011).

<table>
<thead>
<tr>
<th>Region</th>
<th>Protein assignment</th>
<th>Sugar assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>780 – 820 cm(^{-1})</td>
<td>Tyrosine ring</td>
<td>Glycosidic ring</td>
</tr>
<tr>
<td>850 – 900 cm(^{-1})</td>
<td>C-C backbone</td>
<td>C-O-O stretch</td>
</tr>
<tr>
<td>950 – 1000 cm(^{-1})</td>
<td>β-sheet(^1), disordered secondary structure(^2), Phenylalanine ring</td>
<td>Glycosidic ring</td>
</tr>
<tr>
<td>1220 – 1300 cm(^{-1})</td>
<td>Amide III</td>
<td>NH(_2) twist</td>
</tr>
<tr>
<td>1420 – 1490 cm(^{-1})</td>
<td>CH, CH(_2), CH(_3)</td>
<td>Glycosidic ring</td>
</tr>
<tr>
<td>1700 – 1800 cm(^{-1})</td>
<td>Side-chain C=O</td>
<td>n/a</td>
</tr>
</tbody>
</table>

\(^1\)(Howell et al., 1996)  \(^2\)(Ashton et al., 2007)

These six sugar regions (Figure 4.4, Table 4.1) were isolated and PCA performed on each region individually. As illustrated for the 950 – 1000 cm\(^{-1}\) and 1700 – 1800 cm\(^{-1}\) regions, when all the spectra are used (Figure 4.6a and d respectively), there appears to be some separation between RNase A (blue circles) and B (red squares). Furthermore, when the spectra that fell outside the interquartile range of the box and whisker plot in Figure 4.4 were removed (the less typical spectra), the clustering is improved dramatically, with a complete separation between the two protein types observed for the 950 – 1000 cm\(^{-1}\) region (Figure 4.6b and e). Typical representative spectra, selected from the centres of the clusters are shown in Figure 4.6c and f and the features responsible for the separation, tentatively assigned to β-sheet/disordered secondary structure/phenylalanine and side-chain carbonyl vibrations respectively (\textit{vide infra}) are also highlighted in yellow.
The detection of protein glycosylation using TERS

Figure 4.5. Box and whisker plot of the area under the preprocessed TERS spectra in the wavenumber regions did not show differences between the two protein glycoforms using normal Raman spectroscopy. For each region, the shaded box and whisker represents RNase B. In each case the blue box represents the interquartile range (IQR), the red line the median, the black dashed whiskers the remaining data, and any outliers are indicated by the red crosses. In this example, the data were preprocessed using the normalisation against the blank Si method.
Figure 4.6. Following all of the preprocessing, (a) shows PCA scores of the area under the curve of $950 – 1000 \text{ cm}^{-1}$ and (b) shows PCA scores of the interquartile range of the same data, with RNase A shown by blue circles, RNase B shown by red squares and % explained variance in brackets. (c) Shows representative spectra from each group in (b) with RNase A in solid blue, RNase B in dashed red and the selected region highlighted in yellow. (d)-(f) show the same information as (a)-(c) but following the ALS Si-removal preprocessing route and the area under the curve of $1700 – 1800 \text{ cm}^{-1}$. (a) and (d) comprise 643 RNase A spectra and 1116 RNase B spectra, while after selection of the spectra within the IQR (b, e) 321 and 558 spectra are retained for the RNase A and B respectively.
Extra peaks attributed to the glycan in the spectra of RNase B were not definitively observed, a possible explanation for which might be the weak scattering nature of the mannose and GlcNAc residues. From the TERS spectra of neat mannose and GlcNAc it is evident that even in crystalline form these molecules do not always show highly enhanced signals in the characteristic carbohydrate region (900-1100 cm$^{-1}$). Instead, it is thought that the distinction we observe results from differences in the way the two forms of the protein adsorb onto the gold substrate. More specifically, that the native form of the protein (RNase A) can adsorb in many different orientations, but the presence of the glycan in RNase B somewhat restricts the number of possibilities. We have not been able to characterise the structure of the protein monolayer on the gold, but there are four disulphide bridges within the molecule that are able to adsorb onto the surface (Di Felice et al., 2002), as well as numerous side-chain functional groups. It is known that sugar molecules do not readily interact with gold surfaces (Stuart et al., 2005) and therefore both sterically and chemically, the glycan is highly likely to direct adsorption on the gold substrate so that the glycan itself is preferentially on top, thus giving it the potential to mask at least some of the protein vibrations.

The major distinguishing feature is an extra peak in the spectrum of RNase A in the 950–1000 cm$^{-1}$ region. In this region one might expect to observe peaks associated with protein β-sheet (Howell and Li-Chan, 1996), disordered secondary structure (Ashton et al., 2007) and possibly shifted phenylalanine ring-breathing (Tuma, 2005). The lack of this feature in the glycosylated protein spectra could be due to masking by the weakly-scattering glycan if it is attributed to a β-sheet or phenylalanine vibration, or alternatively it is possible that due to its relatively unhindered side-chain functionality, the native form of the protein undergoes partial unfolding upon adsorption on the gold substrate, leading to the
appearance of an additional peak attributed to disordered secondary structure. Unfortunately we have no data to confirm either hypothesis at this stage.

Also as a consequence of fewer carbonyl groups from the amino acid side-chains of RNase B being able to interact with the gold substrate (as they are facing the wrong way) compared to RNase A, they are more able to be ‘seen’ by the TERS tip. This provides a tentative explanation for the increased signal intensity in the 1700 – 1800 cm\(^{-1}\) region for RNase B compared to RNase A (Figure 4.6d-f). Whilst the TERS intensity in this region is generally higher for RNase B than RNase A, peaks are still observed in some RNase A spectra in this region, as would be expected if multiple orientations were possible: in some cases the RNase A will have adsorbed in the preferred orientation for RNase B purely by chance. This would also explain why there is some overlap in the PCA plot (Figure 4.6e). The crystal structure of RNase B has previously been characterised (Williams et al., 1987) and interactions between the glycan and the amino acid side-chains were not detected, thus we do not cite this as a possible explanation for the spectral differences observed.

It was clear that like conventional Raman spectroscopy, TERS does contain enough information to allow the differentiation between two proteins that differ in whether they are glycosylated or not. The data processing steps required to obtain this useful information may be quite involved; however, care has been taken to eliminate bias with the use of exclusively unsupervised multivariate methods, meaning at no point were the analyses developed based on prior knowledge of the sample groups to which the TERS data belonged to.

The aim of the experiment had been to detect directly the glycans on the glycosylated protein molecules in order to look for differences in the glycans themselves, but this does not appear to have been achieved here. While it is possible that this information is present,
the high level of variability in peak positions makes it extremely difficult to make a definitive assignment of spectral features to the sugars in the glycans. The TERS spectra of the crystalline mannose and GlcNAc show that sugar bands in the 800 – 1100 cm\(^{-1}\) region are only detected if the bands in the 1400 – 1600 cm\(^{-1}\) region are very strongly enhanced and also that the spectra of the individual sugar monomers are highly dependent on orientation with respect to the tip.

We do believe this research is a valuable first step towards demonstrating the potential of TERS to characterise glycosylated proteins, given that this is the first time that anyone has been able using TERS to distinguish between glycosylated and native forms of a protein without any control of protein-surface interaction. The next step would be to apply orientation control with the design of functionalised substrates that would force the proteins to orientate themselves on the surface in a reproducible way, thereby minimising the variation in peak position from similar vibrations. With this level of control in place, it would become more likely that structural changes as a result of incorrect glycosylation would be detectable.

4.6. Conclusion

We have demonstrated for the first time that TERS can be used to distinguish between glycosylated and native forms of proteins. This is the first step towards the goal of being able to characterise correct glycosylation of just a few protein molecules at a time using this powerful high spatial resolution technique. We have shown that although TERS data of protein samples can be very complex, carefully considered data processing steps can reveal useful information. Whilst TERS does not compete with the high-throughput
potential of conventional Raman for manufacturing process monitoring and quality control, after careful design of surface substrates to control sample orientation, TERS has the potential for use in glycoprotein synthesis and formulation research.

4.7. Acknowledgements

DC and RG would like to thank the Manchester Institute of Biotechnology for funding this research and the BBSRC and Avacta Group plc. for funding DC’s Ph. D. studentship. He would also like to thank the rest of the Deckert group at Universität Jena for their help and support, as well as Dr. Catherine Winder for her help with sample transport.
4.8. References


4.9. Supplementary information

Figure 4.7. AFM topography of a typical gold nanoplate. The area used to collect a grid of 62 TERS spectra from this example is shown by the white square.
**Figure 4.8.** PCA scores plot of all data following cosmic ray removal and baseline correction. This is the same plot that is shown in Figure 4.2b but this time coloured to show how the PCA is grouping the data by measurement grid, based mainly on the spectral features of the silicon background, as shown by the loadings plot (Figure 4.2c).
The detection of protein glycosylation using TERS

Figure 4.9. The fitting of asymmetric least-squares baselines under the Si overtone peak.

The dot-dashed black line shows the fitted baseline as used to correct the whole spectrum ($\sigma = 1000$, $p = 0.001$) and the dashed red line shows the fitted baseline used to remove the Si overtone from this region ($\sigma = 100$, $p = 0.01$). The original TERS spectrum is shown in solid blue.
5. Quantitative on-line LC-SERS of purine bases
Quantitative on-line LC-SERS of purine bases

5.1. Declaration

This chapter consists of one journal article, currently pending submission:


This article has been reproduced in an unchanged format except for minor adjustments to incorporate it into this thesis.

As primary author on this publication, I carried out all of the sample preparation, detector design and construction, experimental measurements, data analysis and write-up. Dr. Roger Jarvis carried out preliminary investigations relating to the research project and provided preliminary data. Prof. Roy Goodacre, as Principal Investigator, provided the project brief and secured additional funding for the work to be carried out, in addition to providing advice and direction during the course of the research.
Quantitative on-line LC-SERS of purine bases

David P. Cowcher, Roger Jarvis and Royston Goodacre*

School of Chemistry and Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

*Correspondence to roy.goodacre@manchester.ac.uk

5.3. Abstract

Raman spectroscopy has been of interest as a detection method for liquid chromatographic separations for a significant period of time, due to the structural information it can provide, allowing the identification and distinction of co-eluting analytes. Combined with the rapidly advancing field of enhanced Raman techniques, such as surface-enhanced Raman scattering (SERS), the previous low sensitivity of Raman measurements has also been alleviated. At-line LC-SERS analyses, where SERS measurements are taken of fractions collected during or after HPLC separation have been shown to be sensitive and applicable to a wide variety of analytes; however, quantitative, real-time, on-line LC-SERS analysis, applicable to high-throughput experiments, has not been previously demonstrated with better limits of detection than more conventional techniques such as UV absorbance. Here we show that by introducing silver colloid, followed by an aggregating agent into the post-column flow of an HPLC system, we can quantitatively and reproducibly analyse mixtures of purine bases, with limits of detection the region of 100 – 500 pmol. The analysis is performed without the use of a flow cell, thereby eliminating previously detrimental memory effects.
5.4. Introduction

The purine bases adenine and guanine are constituents of both DNA and RNA. They are found in tissues, cells and physiological fluids as a result of nucleic acid catabolism, tissue degradation, diet and nucleotide salvage (Sheng et al., 1991). They have been shown to be susceptible to deamination by reactive nitrogen species in the body to form the breakdown products hypoxanthine and xanthine (Lim et al., 2006). The accurate quantification of these purine bases, together with their metabolites, is valuable, as differences in the amounts present in excreted fluids such as urine have been detected between healthy and non-healthy patients with various cancers (Di Pietro et al., 2001, Liu et al., 2008), as well as providing valuable information for food storage applications (Lou, 1998).

Analysis of purine bases and their derivatives by reversed-phase high-performance liquid chromatography (RP-HPLC) is not new and there are many previously published methods, with a few examples given here (Edmonds et al., 1985, Yao et al., 1990, Porcelli et al., 1994, Kamel et al., 1999, Fryčák et al., 2002, Kaneko et al., 2009). However, the most commonly used detection method is UV absorbance (Brown et al., 2002), and in order to obtain suitable resolution of different purines, sophisticated chromatography is required; for example by the inclusion of ion-pairing species (del Moral et al., 2005) and even then, definitive identification from a complex mixture based on retention time alone is likely to prove problematic. In order to overcome these chemical species’ resolution difficulties, HPLC has been coupled with detection techniques that can provide more detailed analyte identification; most popular are mass spectrometry (Edmonds et al., 1985, Kamel et al., 1999, Chen et al., 1999, Oliveira and Watson, 2000, Locatelli et al., 2012, Seger et al., 2013) and nuclear magnetic resonance (NMR) spectroscopy (Exarchou et al., 2005, Seger et al., 2013).
Vibrational spectroscopic techniques such as infrared (IR) (Kuligowski et al., 2010) and Raman (Sheng et al., 1991, Marquardt et al., 1999, Dijkstra et al., 2005, Carrillo-Carrión et al., 2011, Carrillo-Carrion et al., 2012) are also emerging as useful, novel detectors for HPLC analyses, as they can provide detailed structural information, allowing definitive analyte identification. Furthermore, their analyses are rapid and the instrumentation relatively small and inexpensive compared to techniques such as MS and NMR, making on-line analysis a realistic possibility. The development of vibrational spectroscopic detection for HPLC has not progressed at the speed of other techniques due to the fact that many of the solvents required for the mobile phase, particularly in RP-HPLC, give a strong background signal often masking the analytes of interest. However, with recent advances in data processing techniques, solvent backgrounds are able to be more effectively removed (Kuligowski et al., 2010).

A particularly exciting area of vibrational spectroscopy for many applications where low analyte levels are found is surface-enhanced Raman scattering (SERS). This approach provides vastly superior sensitivity compared to conventional Raman scattering, it retains structural and chemical information, and provides the benefit over IR of minimal interference from water vibrations. SERS uses predominantly metal surfaces with nanoscale features to enhance the incident and scattered light in their vicinity, thereby increasing the detected signal by many orders of magnitude (Campion and Kambhampati, 1998).

The combination of HPLC and SERS detection is not new and various examples have been reported previously (Dijkstra et al., 2005). These can be divided into two main groups: on-line (Freeman et al., 1988, Sheng et al., 1991, Cabalin et al., 1993) and at-line (Somsen et al., 1997, Sägmüller et al., 2003, Trachta et al., 2004, Carrillo-Carrión et al., 2011,
Carrillo-Carrion et al., 2012). The former interfaces SERS detection directly with the HPLC system; meaning analysis is conducted in real time. Whilst this approach is ideal in that it allows high-throughput analysis and data collection that takes no longer than the chromatographic run itself, there are inherent difficulties. For example, with colloid-based SERS approaches, the nanoparticles need to be introduced into the mobile phase flow and the colloid allowed to mix and aggregate sufficiently for a maximum SERS response to be achieved, while at the same time minimising the distance between column and detector to prevent sample diffusion and loss of analyte resolution. There are also issues with memory effects: the colloid sticking to the inside of flow cells used to collect on-line SERS spectra, thus requiring cleaning between samples. A nice idea incorporating a windowless flow cell was presented by Cabalin et al. (Cabalin et al., 1993) but their reported limits of detection (LOD) were significantly higher than UV absorbance detection. Likewise with fixed substrate-based SERS detection, problems arise in the cleaning of the substrate to ensure a reproducible SERS response is obtained from sample to sample. These memory effects were the main reason why on-line HPLC-SERS using roughened electrodes has ceased (Dijkstra et al., 2005).

At-line SERS detection deals with the separation and detection parts of the analysis separately, with the separation being completed first and effluent fractions collected throughout the run. These fractions are then analysed off-line using a separate Raman instrument. As there is no longer a time-critical element to the analysis, care can be taken to optimise the SERS conditions for each fraction independently and there is no need to compromise SERS response for separation efficiency, as is often the case with on-line methods. This does mean of course that at-line methods are so far nowhere near as high-throughput as on-line ones.
Both on-line (Sheng et al., 1991) and at-line (Carrillo-Carrión et al., 2011) SERS detection has been demonstrated for purine base separations. The former used heated colloid as the SERS substrate, which was introduced directly into the mobile phase flow post-column. Thermal aggregation was used to increase the SERS intensity but quantitative detection was not possible. The slower latter method was able to quantify a mixture of purine and pyrimidine bases using micro-LC separation, followed by fraction collection in microtitre plates. Silver quantum dots were then added to these fractions and the pH adjusted to provide optimum SERS conditions before analysis with a relatively expensive and bulky confocal Raman microscope. While this at-line work (Carrillo-Carrión et al., 2011) provides a very robust analysis, incorporating a high level of control, the extra time required for analysis of each sample will significantly affect its applicability to high throughput experiments and the large sample sets often encountered with metabolomics experiments (Goodacre et al., 2004).

In this study we have shown for the first time, to the best of our knowledge, that introducing a separate aggregating agent into the mobile phase flow after the silver colloid allows for a fully quantitative, on-line analysis of mixed analytes. By removing the traditional flow cell altogether, we can also eliminate the memory effects that have plagued previous on-line SERS attempts. Using purine base samples as an example, even when the HPLC peaks are co-eluting, it is possible to use a simple portable Raman spectrometer to generate data that rival traditional UV detection, and the developed LC-SERS method has much lower limits of detection.
5.5. Experimental

5.5.1. Reagents and materials

Adenine (99%), guanine (99%), hypoxanthine (≥99%), xanthine (≥99%), acetic acid (LCMS-grade) and trisodium citrate (analytical grade) were all used as obtained from Sigma Aldrich Ltd. (Dorset, UK). Potassium dihydrogen phosphate and silver nitrate (both analytical grade) were used as obtained from Fisher Scientific (Loughborough, UK). Phosphoric acid (85%) and potassium nitrate (99%) were used as obtained from Alfa Aesar (Heysham, UK). Perchloric acid (60%) was used as obtained from BDH Ltd, (Poole, UK). All water used for this work was of HPLC purity and supplied by Sigma Aldrich (Dorset, UK) unless otherwise stated.

5.5.2. Nanoparticle synthesis

Citrate-reduced silver colloid was synthesised according to the method described by Lee and Meisel (Lee and Meisel, 1982) in 1.5 L batches. All glassware was pre-cleaned using aqua regia and washed thoroughly with deionised water. 270 mg of silver nitrate was dissolved in 1500 mL of water and heated to boiling. 30.0 mL of 1 % weight/volume (w/v) trisodium citrate in water was added dropwise with vigorous stirring and the mixture left to boil for ~1 h. The gradual appearance of a green-grey suspension indicated successful nanoparticle formation. After ~1 h the nanoparticle suspension was allowed to cool and stored protected from light. The nanoparticle size distribution was characterised by UV-vis spectrophotometry to allow comparison of different batches and was very similar to colloid prepared by us previously (Mabbott et al., 2013, Cowcher et al., 2013). The colloid was agitated before use to ensure any settled particles were resuspended.
5.5.3. Sample preparation

1.00 mM stock solutions of the purine bases were prepared according to the method described by Carrillo-Carrión et al. (Carrillo-Carrión et al., 2011). Adenine was dissolved in water and guanine, hypoxanthine and xanthine in 2.00 M perchloric acid. Samples for individual analysis were then prepared by diluting the stock solutions with the appropriate amount of 0.010 M pH 3.0 potassium phosphate buffer. Stock solutions for mixture analysis were prepared in the same way at four times the concentration required in the final mixture for each analyte (4.00 x 10^{-4} M, 2.00 x 10^{-4} M, 1.00 x 10^{-4} M, 4.00 x 10^{-5} M and 2.00 x 10^{-5} M). Equal aliquots of each of the relevant stocks were then mixed together (four per mixture), giving the desired final concentrations according to Table 5.1.

Table 5.1. Final concentration of individual analytes in the mixed solutions, obtained by mixing the appropriate stock solutions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0 x 10^{-4} M</td>
<td>5.0 x 10^{-6} M</td>
<td>1.0 x 10^{-5} M</td>
<td>2.5 x 10^{-5} M</td>
</tr>
<tr>
<td>2</td>
<td>5.0 x 10^{-5} M</td>
<td>1.0 x 10^{-5} M</td>
<td>5.0 x 10^{-6} M</td>
<td>1.0 x 10^{-5} M</td>
</tr>
<tr>
<td>3</td>
<td>2.5 x 10^{-5} M</td>
<td>5.0 x 10^{-5} M</td>
<td>1.0 x 10^{-4} M</td>
<td>5.0 x 10^{-6} M</td>
</tr>
<tr>
<td>4</td>
<td>1.0 x 10^{-5} M</td>
<td>2.5 x 10^{-5} M</td>
<td>5.0 x 10^{-5} M</td>
<td>1.0 x 10^{-4} M</td>
</tr>
<tr>
<td>5</td>
<td>5.0 x 10^{-6} M</td>
<td>1.0 x 10^{-5} M</td>
<td>2.5 x 10^{-5} M</td>
<td>5.0 x 10^{-5} M</td>
</tr>
</tbody>
</table>

5.5.4. Instrumentation

HPLC separation was conducted using a Shimadzu LC-10AVP HPLC system (Kyoto, Japan), consisting of an on-line degasser, two mobile phase pumps, autosampler, column oven and variable wavelength UV absorbance detector.
SERS analysis was performed using a small DeltaNu Advantage 200A portable Raman spectrometer (Laramie, WY, USA), with 633 nm laser excitation. An optical extension tube was fitted to the front of the spectrometer in place of the usual sample holder to allow external Raman excitation with a power of 3 mW at the sample. Scattered light was collected by the same aperture at 180° to the excitation direction.

Colloid for SERS analysis was introduced from a syringe using a syringe driver (KD Scientific, Holliston, MA, USA).

The HPLC system itself was unaltered and operated in the standard UV absorbance detection mode. All modifications to the solvent flow occurred downstream of the UV detector. Silver nanoparticles were introduced to the mobile phase from a 60 mL syringe on a syringe driver, by means of a T-mixer. Immediately downstream of this mixer, 0.500 M potassium nitrate aggregating agent was introduced from a second HPLC mobile phase pump by means of a second T-mixer. Following the introduction of the colloid and aggregating agent, the HPLC tubing ended in front of the aperture of the Raman spectrometer, where the mobile phase/colloid/aggregating agent mixture dripped from the tube as Raman spectra were recorded of the forming droplets. The position of the tube end relative to the spectrometer aperture was optimised by collecting 1 s spectra in continuous mode until the signal response was maximised. The instrument setup is shown in Figure 5.1 and pictures of the apparatus are provided in Figure 5.5.
5.5.5. Sample analysis

HPLC separation was conducted using a 150 x 4.6 mm, 3 μm particle size, Supelcosil LC-18-T C₁₈ column (Supelco, Bellefonte, PA, USA), fitted with a column guard and maintained at 25 °C in a column oven. For each injection the run time was 8.0 min. The mobile phase was 0.010 M aqueous potassium phosphate buffer at pH 3.0, pumped at a flow rate of 1.0 mL min⁻¹. 50 μL of each sample was introduced using an auto-injector. UV absorbance detection was measured at 260 nm, as determined from UV-vis spectra of the individual analytes (data not shown).

Silver colloid for SERS measurements was introduced at 1.0 mL min⁻¹ and the potassium nitrate aggregating agent at 0.2 mL min⁻¹. There was sufficient time to change over the colloid syringe between the end of one sample’s analysis and the injection of the next. This did not affect the SERS data, meaning this was not a limiting factor in the length of a
sample batch. SERS spectra were recorded using the multi-acquire mode of the DeltaNu spectrometer, with an individual integration time of 1 s.

5.5.6. Data analysis

All data were exported from the respective instruments’ operating software and analysed using Matlab R2013a (The Mathworks, Natick, MA, USA). No preprocessing was conducted on the UV absorbance data. In order to remove the background citrate peaks, SERS spectra were row normalised against their total signal and then normalised against the mean SERS spectrum from a sample injection of just the potassium phosphate buffer. The code for these preprocessing steps was written in-house, and is available on request.
5.6. Results and Discussion

The analyte solvent and mobile phase were prepared at pH 3.0 in order to ensure the purine bases remained preferentially in their lactam form (Sheng et al., 1991), which reduces the effect of vertical base stacking (Brown et al., 2002). This tautomerisation is shown in Scheme 5.1 (Supplementary Information). As such, the four components eluted from the HPLC column within 8 min.

In previous nanoparticle-based SERS experiments, for example bacterial spore detection (Cowcher et al., 2013), we have found that mixing equal volumes of sample and colloid gives good SERS spectra. Therefore Ag colloid was pumped at the same flow rate as the HPLC mobile phase (1 mL min$^{-1}$), and this was found to provide good SERS data via preliminary experimentation (data not shown). A syringe driver was chosen to introduce the silver colloid to the system, as the use of more sophisticated liquid delivery instruments would risk lasting contamination and damage to their internal workings through deposition of nanoparticles.

A diagram of the full instrumentation setup is shown in Figure 5.1. In order to test the extent of the colloid memory effect with on-line aggregation (Dijkstra et al., 2005) the SERS detector was initially set up with a quartz capillary tube passing though the conventional sample holder of the DeltaNu instrument. This allowed successful collection of SERS data along with straightforward alignment; however the internal walls of the capillary tube quickly developed a coating of silver as reported previously; a photo of which is shown in Figure 5.6. Therefore the quartz capillary was removed and the sample holder replaced with an optical extension tube that allows analysis of samples external to the Raman instrument. The end of the HPLC tubing was positioned just above the Raman aperture so that as a droplet of the sample formed, it was optimally within the focus of the
Quantitative on-line LC-SERS of purine bases

spectrometer. The precise position of this tubing was optimised by pumping the HPLC mobile phase, colloid and aggregating agent and observing the SERS response of the citrate background, with the instrument in continuous mode (Photographs of this setup are show in Figure 5.5). While the short length of tubing between the aggregating agent T-mixer and spectrometer is technically able to experience colloid memory effects, this was not observed during any of the sample analyses. Replicate samples were not analysed sequentially in order to make any differences from memory effects show up in the data.

The potassium nitrate aggregating agent was introduced at 0.2 mL min$^{-1}$, which means that at a concentration of 0.500 M there is more aggregating agent present than would normally be necessary for a good SERS measurement; however the excess ensures that aggregation occurs rapidly, producing optimum SERS conditions by the time the mixture drips in front of the Raman aperture. The typical time from aggregate addition to measurement is <1 s.

During sample analysis, sequential Raman spectra were recorded with a 1.0 s integration time, using the instrument's ‘multiacquire’ setting. With the combined flow rate of 2.2 mL min$^{-1}$, multiple individual droplets pass in front of the Raman aperture during the collection of each spectrum, thereby allowing a small level of eluent averaging. Ideally the instrument would begin recording the next spectrum as soon as it had finished the previous one but while the individual spectra are only recorded for 1.0 s, the current instrument takes time to save the data file and move onto the next spectrum. As a result of this, spectra were only able to be recorded approximately every 7.5 s. The retention time axis was constructed accurately by dividing the total number of measurements recorded during each injection by the total run time.

The chromatographic conditions were deliberately kept straightforward, with a single-component, isocratic mobile phase. This avoided the requirement for additional ion-pairing
reagents or organic modifiers that might interfere with the SERS detection. Unfortunately this increased peak tailing, particularly with adenine (*vide infra*), but this compromise was made in order to allow us to demonstrate that the SERS spectra can be used to quantify co-eluting peaks of different analytes.

SERS spectra throughout the run are dominated by peaks from the citrate capping agent of the colloid. This background was very intense and highly reproducible, with intensity fluctuations only coming from the lensing affect of the liquid droplets forming and falling in front of the spectrometer. However, when analyte(s) was present, the citrate was readily replaced at the surface of the colloid as evident by the citrate signal intensity dropping off significantly. This provided us with the advantage that we were able to isolate the analyte signal very effectively, first by normalising each spectrum against its total signal (row-normalisation) and then by dividing each spectrum by the mean row-normalised spectrum of a blank sample from the same chromatographic run, as shown in Figure 5.7, along with the citrate background spectrum itself. A similar method has previously been described by Quintás *et al.* ([Quintás et al., 2008](#)) for HPLC-IR spectroscopic data.

Following these data processing steps, example spectra of the four individual purine bases are shown in Figure 5.2. Using vibrations specific to each of the analytes their theoretical limit of detection (LOD) and limit of quantification (LOQ), as well as the reproducibility using UV absorbance and SERS, were determined simultaneously by repeatedly injecting three replicates of individual purine base samples at varying concentrations. In addition a leave-one-out (LOO) analysis was performed, in which a single data point was removed from the dataset and a calibration model (again from individual analyte-specific SERS peaks) developed based on the remaining data. This model was then used to predict the concentration of the data point that was left out. The whole process was repeated for each
individual data point in turn. These results are summarised in Table 5.2, with the associated calibration plots and plots of the leave-one-out predictions shown in Figures 5.8 – 5.11. LODs are calculated from the mean blank value plus three times its standard deviation and the LOQs from the mean blank value plus ten times its standard deviation.

Figure 5.2. SERS spectra of the four nucleobases at their peak retention times in the range 400-2000 cm\(^{-1}\). (a) xanthine, (b) hypoxanthine, (c) guanine and (d) adenine. The spectra here are preprocessed as they would be for analysis: normalised against total signal and against the mean blank spectrum. To provide a better visual comparison, these examples have also been scaled on the y-axis between 0 and 1.
Table 5.2. Metrics for HPLC analysis of individual nucleobase samples using simultaneous UV and SERS detection.

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Adenine</th>
<th>Guanine</th>
<th>Hypoxanthine</th>
<th>Xanthine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention time / s</td>
<td>268</td>
<td>226</td>
<td>254</td>
<td>311</td>
</tr>
<tr>
<td>Linearity</td>
<td>0.9841</td>
<td>0.9892</td>
<td>0.9935</td>
<td>0.9839</td>
</tr>
<tr>
<td>LOD (conc.) / M</td>
<td>4.25 x 10^{-6}</td>
<td>2.97 x 10^{-6}</td>
<td>1.95 x 10^{-6}</td>
<td>2.79 x 10^{-6}</td>
</tr>
<tr>
<td>LOD (on column) / mol</td>
<td>2.13 x 10^{-10}</td>
<td>1.49 x 10^{-10}</td>
<td>9.76 x 10^{-11}</td>
<td>1.40 x 10^{-10}</td>
</tr>
<tr>
<td>LOQ (conc.) / M</td>
<td>7.37 x 10^{-6}</td>
<td>7.62 x 10^{-6}</td>
<td>3.45 x 10^{-6}</td>
<td>6.13 x 10^{-6}</td>
</tr>
<tr>
<td>LOQ (on column) / mol</td>
<td>3.69 x 10^{-10}</td>
<td>3.81 x 10^{-10}</td>
<td>1.72 x 10^{-10}</td>
<td>3.06 x 10^{-10}</td>
</tr>
<tr>
<td>UV Mean linearity of leave-one-out calibration</td>
<td>0.9839</td>
<td>0.9891</td>
<td>0.9934</td>
<td>0.9838</td>
</tr>
<tr>
<td>Linearity of leave-one-out predicted vs actual conc.</td>
<td>0.9812</td>
<td>0.987</td>
<td>0.9916</td>
<td>0.9779</td>
</tr>
<tr>
<td>RMS error of leave-one-out predicted vs actual conc. / M</td>
<td>2.27 x 10^{-6}</td>
<td>1.87 x 10^{-6}</td>
<td>1.50 x 10^{-6}</td>
<td>2.48 x 10^{-6}</td>
</tr>
<tr>
<td>Peak position / cm^{-1}</td>
<td>744</td>
<td>666</td>
<td>734</td>
<td>672</td>
</tr>
<tr>
<td>Peak assignment</td>
<td>Ring breathing (Carrillo-Carrióń et al., 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linearity</td>
<td>0.9816</td>
<td>0.9976</td>
<td>0.992</td>
<td>0.9863</td>
</tr>
<tr>
<td>LOD (conc.) / M</td>
<td>1.58 x 10^{-6}</td>
<td>9.38 x 10^{-7}</td>
<td>1.16 x 10^{-6}</td>
<td>1.98 x 10^{-6}</td>
</tr>
<tr>
<td>LOD (on column) / mol</td>
<td>7.88 x 10^{-11}</td>
<td>4.68 x 10^{-11}</td>
<td>5.79 x 10^{-11}</td>
<td>9.86 x 10^{-11}</td>
</tr>
<tr>
<td>LOQ (conc.) / M</td>
<td>2.26 x 10^{-6}</td>
<td>1.79 x 10^{-6}</td>
<td>4.40 x 10^{-6}</td>
<td>3.39 x 10^{-6}</td>
</tr>
<tr>
<td>LOQ (on column) / mol</td>
<td>1.13 x 10^{-10}</td>
<td>8.9 x 10^{-11}</td>
<td>2.20 x 10^{-10}</td>
<td>1.70 x 10^{-10}</td>
</tr>
<tr>
<td>SERS Mean linearity of leave-one-out calibration</td>
<td>0.9821</td>
<td>0.9976</td>
<td>0.995</td>
<td>0.9862</td>
</tr>
<tr>
<td>Linearity of leave-one-out predicted vs actual conc.</td>
<td>0.9624</td>
<td>0.9962</td>
<td>0.9933</td>
<td>0.9822</td>
</tr>
<tr>
<td>RMS error of leave-one-out predicted vs actual conc. / M</td>
<td>3.26 x 10^{-6}</td>
<td>1.01 x 10^{-6}</td>
<td>1.34 x 10^{-6}</td>
<td>2.20 x 10^{-6}</td>
</tr>
</tbody>
</table>
The SERS data showed excellent quantitative behaviour, with better linearity and lower LOD/LOQ than the UV absorbance detection for each analyte. The LOO analysis was also better for the SERS data than the UV in most cases, with just the adenine results marginally worse in the SERS. However, this was affected by a single anomalous result at the high extreme of the adenine calibration, which is likely to be due to the non-ideal peak tailing that we see for adenine due to the chromatographic conditions chosen, which suggest overloading of this analyte. The inset of the adenine plot in Figure 5.11 shows that at lower concentration the predicted results are very accurate and reproducible.

Mixed-analyte samples were then analysed under the same conditions and following the same data processing steps, although different SERS peaks had to be chosen for adenine and hypoxanthine compared to the individual analyses, due to the spectral overlap in conjunction with the chromatographic co-elution, which prevented the two analytes being accurately quantified using the most intense (ring breathing) vibrations. An example of the data produced from a mixed sample is shown in Figure 5.3 and it can be clearly seen how the different SERS band positions can readily be used to quantify co-eluting analytes, especially when compared to the simultaneously collected UV absorbance data. Once again, individual calibration plots were made for both the UV and SERS data (Figure 5.12) to allow comparison; although due to co-elution, conclusive identification of each analyte by UV was only possible in the very few cases where the concentration of one analyte was particularly high relative to the other(s). Therefore accurate quantification from the mixtures by UV was not possible and the LODs quoted represent the lowest sample concentration from which a peak could be measured, rather than the statistical LOD. Leave-one-out analysis was also employed for the SERS calibration, with the resulting predictions shown in Figure 5.4. The full calibration data are reported in Tables 5.2 to 5.4.
Quantitative on-line LC-SERS of purine bases

Figure 5.3. Example LC-SERS data from a single injection of a sample containing $2.5 \times 10^{-5}$ M guanine, $5 \times 10^{-5}$ M hypoxanthine, $1 \times 10^{-5}$ M adenine and $1 \times 10^{-4}$ M xanthine, eluted in that order. Each SERS spectrum throughout the data collection has been normalised against its total signal and against the mean blank spectrum. (a) shows the 3D LC-SERS data plot for the spectral region between 400 – 2000 cm$^{-1}$. (b) shows the corresponding UV chromatogram (top) along with scaled UV chromatograms from
injections of the individual analytes below. (c) - (f) show SERS intensity with respect to retention time (line), with scaled SERS chromatograms from injections of the individual analytes (solid colour) and the area measured for quantification (shaded) for (c) adenine at 1322 cm$^{-1}$, (d) guanine at 666 cm$^{-1}$, (e) hypoxanthine at 1713 cm$^{-1}$ and (f) xanthine at 672 cm$^{-1}$.

**Figure 5.4.** Plots of mean ($n = 3$) predicted concentration against actual concentration for each of the four nucleobases using leave-one-out calibration. All data are from injections of nucleobase mixtures using SERS detection. Error bars show one standard deviation. Perfect predictions would lie on the red lines.
Table 5.3. UV metrics for HPLC analysis of mixtures of nucleobase samples using simultaneous UV and SERS detection. Overlap of peaks in the UV chromatograms meant that a zero measurement and confident integration of the peaks from certain low-concentration samples was not possible. Therefore, the LOD quoted in brackets is the lowest concentration sample from which a peak could be measured, rather than the statistical LOD. None of the other metrics reported for the SERS analysis (Table 5.4) were able to be determined.

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Adenine</th>
<th>Guanine</th>
<th>Hypoxanthine</th>
<th>Xanthine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention time / s</td>
<td>268</td>
<td>226</td>
<td>254</td>
<td>311</td>
</tr>
<tr>
<td>Linearity</td>
<td>0.9323</td>
<td>0.9595</td>
<td>0.9175</td>
<td>0.9957</td>
</tr>
<tr>
<td>LOD (conc.) / M</td>
<td>(5 x 10^{-5})</td>
<td>(5 x 10^{-6})</td>
<td>(1 x 10^{-5})</td>
<td>(2.5 x 10^{-5})</td>
</tr>
<tr>
<td>LOD (on column) / mol</td>
<td>(2.5 x 10^{-9})</td>
<td>(2.5 x 10^{-10})</td>
<td>(5 x 10^{-10})</td>
<td>(1.25 x 10^{-10})</td>
</tr>
</tbody>
</table>
Quantitative on-line LC-SERS of purine bases

Table 5.4. SERS metrics for HPLC analysis of mixtures of nucleobase samples using simultaneous UV and SERS detection.

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Adenine</th>
<th>Guanine</th>
<th>Hypoxanthine</th>
<th>Xanthine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention time / s</td>
<td>268</td>
<td>226</td>
<td>254</td>
<td>311</td>
</tr>
<tr>
<td>Peak position / cm$^{-1}$</td>
<td>1322</td>
<td>666</td>
<td>1713</td>
<td>672</td>
</tr>
<tr>
<td>Peak assignment</td>
<td>-N7C5+C8N7 stretch$^1$</td>
<td>Ring breathing$^{1,2}$</td>
<td>C6=O stretch$^3$</td>
<td>Ring breathing$^2$</td>
</tr>
<tr>
<td>Linearity</td>
<td>0.9817</td>
<td>0.9874</td>
<td>0.9795</td>
<td>0.9672</td>
</tr>
<tr>
<td>LOD (conc.) / M</td>
<td>$2.75 \times 10^{-6}$</td>
<td>$3.45 \times 10^{-6}$</td>
<td>$9.43 \times 10^{-6}$</td>
<td>$4.07 \times 10^{-6}$</td>
</tr>
<tr>
<td>LOD (on column) / mol</td>
<td>$1.38 \times 10^{-10}$</td>
<td>$1.73 \times 10^{-10}$</td>
<td>$4.72 \times 10^{-10}$</td>
<td>$2.04 \times 10^{-10}$</td>
</tr>
<tr>
<td>LOQ (conc.) / M</td>
<td>$5.91 \times 10^{-6}$</td>
<td>$5.06 \times 10^{-6}$</td>
<td>$1.89 \times 10^{-5}$</td>
<td>$1.15 \times 10^{-5}$</td>
</tr>
<tr>
<td>LOQ (on column) / mol</td>
<td>$2.95 \times 10^{-10}$</td>
<td>$2.53 \times 10^{-10}$</td>
<td>$9.47 \times 10^{-10}$</td>
<td>$5.75 \times 10^{-10}$</td>
</tr>
<tr>
<td>Mean linearity of leave-one-out calibration</td>
<td>0.9818</td>
<td>0.9874</td>
<td>0.9794</td>
<td>0.9856</td>
</tr>
<tr>
<td>Linearity of leave-one-out predicted vs actual conc.</td>
<td>0.9753</td>
<td>0.9838</td>
<td>0.9745</td>
<td>0.9806</td>
</tr>
<tr>
<td>RMS error of leave-one-out predicted vs actual conc.</td>
<td>$5.51 \times 10^{-6}$</td>
<td>$4.47 \times 10^{-6}$</td>
<td>$5.72 \times 10^{-6}$</td>
<td>$4.91 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

$^1$(Otto et al., 1986) $^2$(Carrillo-Carrión et al., 2011) $^3$(Ulicny et al., 1994)

Unsurprisingly, the SERS linearity and LOD/LOQ, whilst very respectable, are not quite as good as for the individual sample injections. Whilst we note this, the difference is by no means extreme, and certainly demonstrates that the SERS detection is quantitative in a situation where UV absorbance detection is clearly not possible, and a more powerful detector that yields structural information is needed.

In comparison to the at-line fraction collection method (Carrillo-Carrión et al., 2011), the LODs and LOQs of the on-line SERS data from the individual analyte injections are significantly improved for guanine and hypoxanthine, and comparable for adenine and
xanthine. From the mixed samples, the on-line SERS LODs and LOQs are only very slightly higher in all cases. The major improvement here comes as the data collection is completed at the same time as the chromatography, rather than requiring chromatography first, followed by lengthy offline SERS analysis using a large Raman microscope. We believe we have demonstrated that our method is far more applicable to high throughput measurements. Of course Carrillo-Carrión’s at-line method allows for far greater control of SERS conditions, particularly in terms of pH, which thus far limits our on-line method to analyte mixtures that are amenable to SERS at similar pH; however, we believe there is the potential for on-line pH modification post-column where in the future we shall attempt to introduce the buffer in the same manner as the aggregating agent. While this would require analysis-specific validation and for the buffer to be programmed for each analysis, it would allow the on-line analysis of more complex mixtures using SERS detection.

Finally in comparison to at-line techniques, our method requires large quantities of colloid. However, as citrate-reduced silver colloid is cheap and straightforward to produce in large quantities, this should not be a significant drawback. In regular use, the syringe driver would easily be replaced with a pump that could handle larger volumes of colloid than the 60 mL syringes used here, thereby allowing large sample runs to continue unattended for many hours. The on-line aggregation of the colloid means the waste liquid simply needs to be filtered to remove the silver metal for safe disposal.
5.7. Conclusion

We have shown for the first time that on-line SERS detection for HPLC can be fully quantitative and provide data in real time, simultaneous to other non-destructive on-line detection techniques, such as UV absorbance. This approach brings with it a significant time saving over at-line methods that require SERS analysis following chromatographic separation and cumbersome fractionation, making this method more applicable to high throughput applications. The introduction of silver colloid to the post-column solvent flow, followed by the introduction of potassium nitrate aggregating agent, gave reproducible SERS spectra from mixtures of purine bases that were able to be quantified even when the analytes were co-eluting, based on their unique Raman spectra. The LODs were comparable and sometimes superior to those of current at-line methods, in the region of 100 – 500 pmol on-column. Our technique does lack the post-column control of SERS conditions, such as pH, that previously published at-line methods do, but we believe this problem could be addressed in the future and allow SERS quantification of mixtures of a wide range of analytes.

5.8. Acknowledgements

DC would like to thank the BBSRC and Avacta Group plc. for funding this work and his Ph. D. studentship. RG thanks U.K. EPSRC and BBSRC and the industrial members of the Bioprocessing Research Industry Club (BRIC) for funding. We would also like to thank Dr. Catherine Winder for the use of the syringe driver.
5.1. References


Quantitative on-line LC-SERS of purine bases


5.2. Supplementary Information

Figure 5.5. Photographs of the SERS detector during analysis with the enclosure removed. Side view of the sample droplet forming at the end of the LC tubing in front of the laser aperture (top), front-on view (lower left) and a new droplet forming in the laser path as soon as the previous droplet has fallen (lower right).
Figure 5.6. Quartz capillary used for LC-SERS analysis, showing the inside of the tube coated with silver.
Quantitative on-line LC-SERS of purine bases

**Figure 5.7.** Extracting useful information from the LC-SERS data. (a) shows a 3D plot of the raw data obtained from a LC-SERS injection. In this case the sample is $5 \times 10^{-5}$ M guanine. The spectra are clearly dominated by the vibrations of the Ag nanoparticles’ citrate capping species, shown in (b). This becomes more apparent when the spectra are normalised against the total signal (c). A loss of signal intensity is observed when the analyte displaces the citrate on the surface. Normalisation against the mean blank spectrum (entirely citrate peaks) reveals the relatively low-intensity guanine spectrum (d).
Figure 5.8. Calibration plots for injections of the individual nucleobases using UV detection. The mean UV peak area \( n = 3 \) is shown with error bars equal to one standard deviation. Insets show a close-up of the main plot in the low concentration region.
Figure 5.9. Plots of mean ($n = 3$) predicted concentration against actual concentration for each of the four nucleobases using leave-one-out calibration. All data are from injections of individual nucleobase samples using UV detection. Error bars show one standard deviation. Insets show a close-up of the main plot in the low concentration region.
**Figure 5.10.** Calibration plots for injections of the *individual* nucleobases using SERS detection. The mean SERS peak area ($n = 3$) at the relevant wavelength (see Table 5.2) is shown with error bars equal to one standard deviation. Insets show a close-up of the main plot in the low concentration region.
Figure 5.11. Plots of mean ($n = 3$) predicted concentration against actual concentration for each of the four nucleobases using leave-one-out calibration. All data are from injections of individual nucleobase samples using SERS detection. Error bars show one standard deviation. Insets show a close-up of the main plot in the low concentration region.
Figure 5.12. Calibration plots for injections of mixtures of nucleobases using SERS detection. The mean SERS peak area \( n = 3 \) at the relevant wavelength (see Table 5.4) is shown with error bars equal to one standard deviation. Insets show a close-up of the main plot in the low concentration region.
Scheme 5.1. Tautomerisation of adenine between the lactam and lactim forms.
6. Discussion
6.1. Discussion and future perspectives

The research presented in this thesis has demonstrated several new developments in the discriminatory and quantitative application of enhanced Raman scattering techniques, and even though the specific application of each chapter is different, there are some important overall philosophies that have been shown from the thesis as a whole.

First, and perhaps most importantly, all of the chapters highlight the fact that the data produced from enhanced Raman experiments are complicated, particularly when low concentration samples are used. Therefore, suitable data handling and analysis must be carefully considered, not just to extract the useful information from the background, but also to be confident that the apparent findings are genuinely as a result of the hypothesis under examination and not due to other external factors.

Second, this thesis has shown that whilst state-of-the-art multivariate data analysis techniques, such as the PLS and KPLS regression analyses used in Chapter 2, are becoming very much the initial go-to approach for many researchers, these techniques are not always the most appropriate for the dataset being analysed. It is important not to forget the value of the more straightforward, yet established univariate techniques, even for the analysis of complex datasets, as exemplified in the TERS analysis of proteins in Chapter 4.

Finally, all of the chapters show that understanding and control of the interactions between analytes and competing species with the surface of the enhancing medium is vital if fully quantitative analyses are to be achieved using enhanced Raman scattering.

As set out in Chapter 2, bacterial spore detection via the DPA biomarker has been an active area of research within the SERS community (Farquharson et al., 2004, Bell et al., 2005, Zhang et al., 2005, Farquharson and Inscore, 2006, Cheng et al., 2009) due to the
Discussion

sensitivity and selectivity, but above all the speed and potential portability of SERS measurements. This being said, the latest research in the area had drifted away from colloid-based SERS substrates in favour of fixed substrates, usually on a slide that needs to be placed under a Raman microscope for analysis (Zhang et al., 2005, Cheng et al., 2009). Although this had brought benefits in sensitivity and reproducibility over the current colloidal-based analyses, due to the confocality of a microscope instrument and the ability to precisely engineer the substrate, the monetary cost is vastly increased and the portability is lost, two of the fundamental benefits to using SERS analysis in the first place. The work presented in this thesis has been able to show the analytical community that colloid-based SERS is still applicable to the analysis of bacterial spores, giving almost three orders of magnitude increase in sensitivity over previously reported colloid-based methods. Furthermore, this was demonstrated using a small, portable instrument, making the analysis appropriate for field use.

The use of an internal standard for SERS analysis is rarely reported and adds further strength to the analytical method as a whole. It also showed that competition effects at the nanoparticle surface (Stewart et al., 2012) are significant when the analyte concentrations are so low and this had to be accounted for in the quantitative analysis.

Whilst the sensitivity to the DPA biomarker was dramatically increased in its pure form, it was disappointing not to be able to show the same sensitivity with DPA extracted from Bacillus spores. The SERS spectra of the concentrated spore extract showed that DPA was definitely present, but at low concentrations the signal was lost in the background. It is likely that this is caused by inefficient extraction of the DPA from the bacterial spores but could also be a result of further competition at the nanoparticle surface due to the vast array of other species that are present in a biological sample. The improved extraction and
analysis of DPA from live spores is therefore the next direction required for this area of research, which will come from a better understanding of the competing species within the spore extract, as well as exploring alternative extraction methods. Techniques such as liquid chromatography-mass spectrometry (LCMS) (Donato et al., 2012) would be able to provide detailed information about the composition of the extract and allow identification of any likely surface competition. With better understanding of the precise surface kinetics of the DPA analysis, it is hoped that colloidal SERS methods will be applicable to other biological systems in a quantitative manner in the future.

SERS enhancement by non-metallic substrates has only recently started to be explored and for the most part is limited to the chemical enhancement mechanism, due to the insulating properties of most non-metallic species (Wang et al., 2011). Chemically enhanced SERS in the UV excitation region, as explored in Chapter 3, had the potential to allow Raman enhancement away from the interference of fluorescence effects. Whilst the data initially appeared to show interactions between the Rhodamine 6G analyte and the BN substrate, repeat measurements were not conclusive and the observations were not transferrable to other analytes. Perhaps the biggest limitation of the substrate itself, even if it were to be subsequently proven to genuinely enable a Raman enhancement, is the B-N stretching vibration at 1370 cm\(^{-1}\) that appears in an important region of the spectrum for biomolecule analysis. It is therefore likely that this would preclude the use of the BN substrate for future SERS research. Despite this negative outcome, the results of this work are still important in the understanding of the dependence of SERS enhancements on the specific interactions between substrate and analyte, and in a more general sense emphasise the importance of rigorous data analysis to ensure the interpretations made are correct.
Careful consideration and exploration of different data analysis techniques was responsible for the highly positive outcome of the TERS work presented in Chapter 4. The need for accurate glycan characterisation was clearly identified (Walsh and Jefferis, 2006) and TERS was quite rightly identified as a potential method of choice. It was perhaps somewhat optimistic to set out to achieve full glycan characterisation by TERS in the first set of experiments, given the lack of previous literature in this area, and the complexity of the data produced soon gave an indication as to perhaps why this had not previously been attempted. The variability in the data from measurement to measurement showed just how sensitive a technique TERS is, and how much the peak positions depend on the orientation of the analyte with respect to the TERS tip.

Throughout the analysis of the TERS data, which took considerable time, care was taken to maintain a lack of bias. This approach adds strength to the results achieved, especially since at no stage was any prior information about the class each sample belonged to used in developing the analytical model. Special care was taken to ensure any data processing techniques used did not add additional variability to the dataset, not only to avoid adding bias, but to avoid compounding the problem of data variability by increasing it unnecessarily. This is why the silicon background removal method by Gaussian fitting, which was probably the most sound theoretically of the methods proposed, was rejected based on its propensity to include artefacts.

Despite the inability of the TERS analysis in its current state to provide specific glycan characterisation, the ability to determine glycosylation status is nevertheless a valuable first step towards achieving this goal and shows that TERS has the ability to discriminate between different proteins when measuring only a few molecules at a time. Far greater knowledge of the way the protein interacts with the TERS substrate is required if
Discussion

orientation is to be adequately controlled in future experiments. If this is achievable, perhaps through the design of self-assembled monolayers or controllable derivatisation of the protein species themselves, then TERS has a promising future in the characterisation of protein glycosylation and other post-translational modifications on single protein molecules.

In the final results chapter of this thesis, colloid-based SERS enhancement was again shown to be sensitive and quantitative. There was a clear need for a new detection system for HPLC separations that could provide the detailed structural information necessary to discriminate and quantify co-eluting components of complex samples, where due to the co-elution of two or more analytes, crowded chromatograms are unavoidable (Sheng et al., 1991).

In a similar manner to the bacterial spore biomarker detection, emphasis in the analytical community had migrated from straightforward on-line techniques, to highly confocal and controllable, but expensive and time-consuming at-line methods (Carrillo-Carrión et al., 2011). Whilst such methods have been shown to be sensitive and reproducible, they are unsuitable for high-throughput experiments with large sample sets, as the SERS analysis itself has to be conducted after the chromatography, as well as each fraction being carefully prepared beforehand.

One of the main reasons for on-line methods falling out of favour was the memory effects caused by colloid sticking to the inside of the various designs of flow cell (Carrillo-Carrión et al., 2011). This problem was effectively alleviated by removing the flow cell altogether and measuring spectra of forming droplets of the sample/colloid/aggregating agent mixture. Indeed it is the novel introduction of aggregating agent following the introduction
of the nanoparticles that allows more intense and reproducible SERS spectra than have previously been reported.

The purine base example used herein to demonstrate the on-line SERS detection for HPLC is perfectly valid in its own right, but to extend this work, it would be beneficial to explore separations of other analyte mixtures. In its current state, the instrumentation only allows analysis of analytes that are amenable to SERS detection at the same pH; however with modification to the post-column colloid and aggregating agent introduction, it is anticipated that pH control could also be imparted, further diversifying the application potential of the technique.

An unexpected outcome of the HPLC-SERS work was that the displacement of the citrate capping agent of the colloid by the purine analyte was very obvious and reproducible. With further investigation, real-time liquid mixing in a flowing system, coupled with on-line detection, could provide a useful insight into the kinetics of analyte competition at the nanoparticle surface.

The outcome of each individual section of this thesis has been positive, including Chapter 3, where although no publishable work resulted, the importance of rigorous repeat analysis and investigation of multiple analytes was demonstrated. This thesis as a whole has provided valuable support for the ever-advancing field of enhanced Raman scattering and in particular has contributed to the revitalisation of colloid-based SERS analyses in areas where it was beginning to be dismissed as unreliable. It has been shown that with appropriate experimental design and control, sensitive and reproducible SERS analysis need not be at the expense of speed, cost-effectiveness and portability.
Discussion

In summary, the work presented in this thesis has shown that colloidal SERS can be used to detect and quantify the DPA biomarker for *Bacillus* bacteria at far lower levels than those reported previously and below the infections dose of inhalation anthrax, as well as how the use of an internal standard can improve the reproducibility of SERS data. It has shown that although boron nitride surfaces may have the potential to enhance Raman signals via a chemical enhancement mechanism, the background spectrum of the surface itself interferes significantly. TERS was shown to be able to distinguish between glycosylated and non-glycosylated forms of RNase protein in a monolayer with the use of a carefully designed data analysis strategy and finally, the on-line interfacing of HPLC and colloidal SERS was demonstrated to be sensitive and quantitative for the first time.

In future it is hoped that other SERS researchers explore the use of internal standards as a means of improving the reproducibility of their analyses. There is certainly more to be learned in terms of the competition at the nanoparticle surface and how this can be accurately modelled and accounted for. Hopefully readers of this work will take away the idea that planning the way their data is analysed is just as important as collecting the data itself, particularly for quantitative analyses of trace amounts. Of all the work presented here, on-line HPLC-SERS has the highest potential for future development and use, having already been shown to work well, even with minimal attention to the design of the chromatographic system. It provides a real alternative to existing techniques and the fundamental methods presented here could be adapted relatively easily in order to be applied to a wealth of different analytical systems. The flowing SERS system itself also has potential outside of the HPLC application for the study of interactions at the colloid surface.


6.2. References


Appendix I: Published work in original format
Publication 1.


*Analytical Chemistry*, 85, 3297-3302.
Portable, Quantitative Detection of *Bacillus* Bacterial Spores Using Surface-Enhanced Raman Scattering

David P. Cowcher, Yun Xu, and Royston Goodacre*

School of Chemistry and Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN United Kingdom

**Supporting Information**

**ABSTRACT:** Portable rapid detection of pathogenic bacteria such as *Bacillus* is highly desirable for safety in food manufacture and under the current heightened risk of biological terrorism. Surface-enhanced Raman scattering (SERS) is becoming the preferred analytical technique for bacterial detection, due to its speed of analysis and high sensitivity. However in seeking methods offering the lowest limits of detection, the current research has tended toward highly confocal, microscopy-based analysis, which requires somewhat bulky instrumentation and precisely synthesized SERS substrates. By contrast, in this study we have improved SERS for bacterial analyses using silver colloidal substrates, which are easily and cheaply synthesized in bulk, and which we shall demonstrate permit analysis using portable instrumentation. All analyses were conducted in triplicate to assess the reproducibility of this approach, which was excellent. We demonstrate that SERS is able to detect and quantify rapidly the dipicolinate (DPA) biomarker for *Bacillus* spores at 5 ppb (29.9 nM) levels which are significantly lower than those previously reported for SERS and well below the infective dose of $10^4$ *B. anthracis* cells for inhalation anthrax. Finally we show the potential of multivariate data analysis to improve detection levels in complex DPA extracts from viable spores.

Spore-forming bacteria, such as *Bacillus*, are prevalent in soil and on plants and, as such, are commonly found in plant-based foods. They can easily be transferred to meat and dairy products during production, and although most forms are harmless, certain types, for example *B. cereus* and *B. subtilis*, release toxins that cause food poisoning. The most dangerous and notorious form, *B. anthracis*, causes anthrax, and its potential for use in bioterrorism was highlighted by the attacks on the U.S. postal service in 2001. A method of rapid and sensitive detection of spores, amenable for portable field use, is therefore highly desirable.

The vast number of *Bacillus* strains and their associated toxins means infective doses can range from $10^6$ to $10^{11}$ spores/cells; therefore, the lowest end of this range is the necessary limit of detection (LOD). Biological analyses, such as PCR, are able to identify and discriminate between bacteria at the strain level, sometimes with as few as 10 spores, but even though analysis can be completed in the region of 1 h, it is still far too slow and interest has shifted toward small molecule biomarker detection. This results in a compromise between loss of discriminatory information and speed of analysis; the idea being that rapid detection allows precautionary measures to be taken while the specific danger level is more thoroughly assessed. A particularly useful biomarker is dipicolinic acid (DPA), as it is found in bacterial spores (usually as a Ca$^{2+}$ chelate complex) but not in other spores, such as mold or pollen, or other commonly occurring environmental materials. DPA can be detected and quantified by common laboratory-based techniques, such as liquid chromatography, gas chromatography/ion mobility mass spectrometry, and Fourier transform infrared spectroscopy, although these techniques often lack the sensitivity required, particularly when sensitivity has to be compromised for portability (LODs in the tens of nanograms in the case of HPLC-UV and the prototype portable GC/IMS$^2$).

Appropriate sensitivity has been achieved in a time scale of a few minutes, by complexing Tb$^{3+}$ with DPA and measuring photoluminescence,$^{10,11}$ however, this method is susceptible to interference from other molecules, either through DPA binding competition, leading to false positive results, or inhibition and luminescence quenching,$^{10}$ leading to false negatives. Despite significant improvement,$^{12}$ interference in environmental samples cannot yet be completely eliminated.

Raman spectroscopy, in particular surface-enhanced Raman spectroscopy (SERS), has been shown to be a useful tool for DPA detection and quantification,$^{13-20}$ as the vibrational spectrum allows definitive identification of DPA from interfering species, analysis time is less than 1 min, and portable instrumentation is readily available. Its use has even been demonstrated with samples collected from mail sorting apparatus,$^{14}$ albeit at milligram level.
Analytical Chemistry

SERS exploits the enhancement in Raman signal intensity observed when an analyte is brought into close proximity with a metal surface that has nanoscale roughness features. These surfaces can be loosely divided into two groups: solid surfaces with immobilized nanoscale features and colloidal suspensions. DPA quantification has been reported using TEM.

Very low LODs have been reported on solid surfaces, down to 0.1 parts per billion (ppb) DPA.19 While this is an impressive LOD, this technique requires the use of a Raman microscope, which greatly limits the portability of the method. In our opinion, the most practical approach to portable SERS detection of Bacillus is to use colloidal nanoparticles, as reported by Bell et al.21 This permits the use of smaller, simpler, and less expensive instrumentation and provides a true indication of analyte concentration, as the nanoparticles and adsorbed analyte move within the laser path under Brownian motion to give an averaging effect, rather than trying to manually focus on one fixed point. It is the improved detection of DPA using colloidal SERS that we report here along with their implications in food poisoning.

■ EXPERIMENTAL SECTION

Reagents. DPA (99%), manganese sulfate monohydrate (99%), trisodium citrate dihydrate (99.5%), glutaric acid (96%), and silver nitrate (99.9%) were used as supplied by Sigma Aldrich Ltd. (Dorset, U.K.). Nitric acid (70%) was used as supplied by Fisher Scientific (Loughborough, U.K.). Sodium sulfate decahydrate (99%) was used as supplied by BDH Ltd. (Poole, U.K.). Lab 28 blood agar base (without blood) was used as supplied by Lab M (Bury, U.K.).

Colloid. Citrate-reduced silver colloid was prepared according to the Lee and Meisel method.21 Briefly, 90.0 mg of AgNO₃ was dissolved in 500 mL of doubly distilled, deionized water and brought to boiling. Then 10.0 mL of w/v aqueous sodium citrate solution was added with stirring, and boiling was resumed for 1 h. We have previously thoroughly characterized colloid prepared by this method using transition electron microscopy (TEM)22 and scanning electron microscopy (SEM)23 and here compared the UV–vis absorbance spectrum to confirm that this colloid batch was comparable (data not shown).

Instrumentation. SERS spectra were recorded using a DeltaNu Advantage 200A portable spectrometer (DeltaNu, Laramie, WY, USA), equipped with a 633 nm HeNe laser giving a power of 3 mW at the sample, which was calibrated using a polystyrene internal standard as supplied by the instrument manufacturers. All analyses were conducted in triplicate.

SERS Analysis of Pure DPA. DPA solutions of 1000 ppm (5.98 mM) were prepared in both deionized water and 20.0 mM aqueous nitric acid. These solutions were further diluted with their respective solvents to provide a range of different concentrations.

SERS analysis was performed as follows: 200 μL of silver colloid was added to a glass vial, followed by 200 μL of the DPA solution, and lastly 50 μL of 0.100 M sodium sulfate aggregating agent. The vial was capped and vortexed for 2 s, and then a 30 s SERS spectrum was recorded. In preliminary experiments (vide infra) aggregation times were assessed using sequential 30 s spectral collections for 30 min (60 measurements in total).

A color change of the colloid from light gray to darker gray indicates successful aggregation. In the case of the nitric acid solutions, it was apparent that self-aggregation was occurring before addition of the aggregating agent; therefore, for these samples the aggregating agent was omitted. In the experiments where direct comparisons were made between these two solvent systems, 50 μL of deionized water was added in place of the aggregating agent to maintain comparable volumes and concentrations. This was not required for all other measurements. For quantitative measurements, an internal standard of glutaric acid was added to the nitric acid solvent at a fixed concentration of 2.00 mM.

Culture of Bacterial Spores. B. subtilis and B. cereus spores were cultured according to the method reported by Goodacre et al.22 Briefly, Lab 28 nutrient agar was prepared containing 5.0 mM MnSO₄ to promote sporulation. Cultures were incubated at 30 °C for 6 d, by which time the nutrient source had been depleted and the colonies were no longer growing. Spore colonies were scraped from the agar plates using an inoculating loop and suspended in deionized water. The spore suspensions were maintained at 50 °C in a water bath to ensure maximum sporulation had occurred.

DPA Extraction and Analysis. Two 1.0 mL aliquots of the spore suspension were each washed once to remove any dissolved impurities, such as salts from the growth medium, by centrifuging at 17000g for 5 min and the supernatant was discarded. One wash and resuspension cycle was determined to be optimal (data not shown). One of the respective spore pellets was resuspended in deionized water, serial diluted, and used to perform a viable spore count. The other pellet was resuspended in 40.0 mL of 40.0 mM MnSO₄ enabled better extraction than the original 20.0 mM nitric acid containing 2.00 mM glutaric acid internal standard and diluted with more of this solvent to provide a range of spore concentrations. These spore samples were ultrasonicated for 10 min and analyzed in the same manner as described above for the pure DPA samples.

Data Processing. All data analysis was performed using Matlab software version R2008a (The MathWorks, Natick, MA, USA). Univariate analysis comprised of normalizing the area under the curve of the most prominent feature: the pyridine N–H stretch of the amine group, 3298−3302 cm⁻¹, to the C–H stretch of the glutaric acid internal standard at 2934 cm⁻¹, after baseline correction using the asymmetric least-squares method24 and scaling each spectrum on the intensity axis from 0 to 1. Multivariate analysis was carried out using partial least-squares (PLS) and kernel partial least squares (KPLS) regression models, as reported previously by our research group,25,26 which analyze the whole SERS spectra. The raw spectra had been preprocessed using extended multiplicative scatter correction (EMSC) applied prior to PLS and KPLS analyses.

For all three analysis techniques, 40% of the data (the test set) were randomly selected to be left out and a calibration model was generated based on the remaining 60% (the training set). This model was used to predict the "unknown" concentrations of the data in the test set. Five different test sets were randomly chosen for each species.

■ RESULTS AND DISCUSSION

For a starting point we repeated the method reported by Bell et al.21 using our portable 633 nm Raman spectrometer, with the aim of applying similar methodology to the analysis of DPA extracted from viable spores. The spectra of DPA (Figure 1)
Appendix I: Published work in original format

Analytical Chemistry

Figure 1. Baseline-corrected SERS spectra of glutaric acid internal standard, pure DPA, and DPA extracted from *B. cereus* and *B. subtilis* spores. Highlighted are the DPA ring breathing vibration at 1006 cm\(^{-1}\) and the glutaric acid C–H stretch at 2934 cm\(^{-1}\) that are used for quantification. Chemical structures of DPA and glutaric acid are shown.

Dissolved in deionized water were compared directly to those of DPA dissolved in 20.0 mM nitric acid, which has previously been shown to extract DPA from Bacillus spores. It has also been reported that acidic conditions (pH 2–4) provide the optimum conditions for SERS of DPA. The potassium thiocyanate internal standard described in the work by Bell et al. was omitted at this stage to avoid undesirable reactions with the nitric acid.

This comparison lead to some interesting observations: first that the time taken for maximum SERS intensity to be achieved using the Na\(_2\)SO\(_4\) aggregating agent significantly increased as the concentration of DPA was reduced (over 1 h approaching the 1 ppm LOD) and second that the nitric acid solution was capable of aggregating the colloid on its own, allowing any additional salt-based aggregating agent to be omitted from the nitric acid samples.

The relative SERS intensities for comparable DPA concentrations after aggregation with Na\(_2\)SO\(_4\) or HNO\(_3\) are shown in Figure 2. Here we see that, for the same DPA concentration, nitric acid aggregation significantly lowers the time taken for maximum SERS intensity to be reached. It also increases the overall signal intensity and improves the reproducibility (as judged by the small standard deviation error bars) compared to aggregation using sodium sulfate.

While we cannot confirm the exact mechanism of aggregation that operates in either case, the presence of the nitric acid will significantly lower the pH of the system, which has been shown to rapidly initiate aggregation of silver nanoparticles even in the absence of an additional salt. In the same work, it is also observed that silver colloid aggregated with nitrate salt gives stronger SERS enhancement than with sulfate. We believe that a combination of these two factors allow for the dramatic improvement in SERS signal we observe over the sulfate-based aggregation. As the use of the thiocyanate internal standard would necessitate stricter safety controls in combination with the nitric acid (due to the liberation of cyanide), we instead use glutaric acid, which will associate with the colloid in the same manner as DPA (via the carboxylate groups) but provide a C–H stretching mode not present in DPA for use as a reference.

**Univariate Analysis of Pure DPA.** The univariate calibration curve for DPA in 20.0 mM aqueous nitric acid is shown in Figure 3a. The SERS signal is now normalized to the internal standard to remove unavoidable effects such as colloid load differences or fluctuations in laser fluency. Normalization was achieved simply by dividing the area under the pyridine ring-breathing mode of DPA at 1006 cm\(^{-1}\) by the area under the C–H stretch at 2934 cm\(^{-1}\) from the glutaric acid (2.00 mM) internal standard. Note that in this plot the mean of three triplicate preparations is shown along with the standard deviation (error bars). The points are the mean areas under the ring-breathing mode from DPA, normalized against the glutaric acid internal standard C–H stretch at 2934 cm\(^{-1}\). Data are from baseline-corrected and scaled spectra and the mean of *n* = 3 measurements are shown. Error bars show one standard deviation and the inset shows a close-up of the 0–100 ppb region. (B) Logarithm of the mean areas shown in the inset in (A) against DPA concentration, providing evidence for a nonlinear calibration.

![Figure 2](image1.png) Time-related plot showing the effect of two aggregating agents on SERS signal from DPA. A constant concentration of DPA was used (5 ppm, 29.9 μM) and SERS signals were collected sequentially after the initiation of aggregation by nitric acid and sodium sulfate. The points are the mean areas under the ring-breathing mode at 1006 cm\(^{-1}\) from baseline-corrected spectra, *n* = 3. Error bars show one standard deviation.

![Figure 3](image2.png) (A) Calibration curve showing the variation of the mean area under the 1006 cm\(^{-1}\) ring-breathing mode from DPA, normalized against the glutaric acid internal standard C–H stretch at 2934 cm\(^{-1}\). Data are from baseline-corrected and scaled spectra and the mean of *n* = 3 measurements are shown. Error bars show one standard deviation and the inset shows a close-up of the 0–100 ppb region. (B) Logarithm of the mean areas shown in the inset in (A) against DPA concentration, providing evidence for a nonlinear calibration.
Appendix I: Published work in original format

Analytical Chemistry

Table 1. Comparison between Predictions from Three Different Data Analysis Methods, Aimed at Predicting Pure DPA Concentration or Spore Counts from B. cereus or B. subtilis

<table>
<thead>
<tr>
<th>target</th>
<th>metric</th>
<th>univariate</th>
<th>PLS</th>
<th>KPLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>pure DPA conc./ppb</td>
<td>factors</td>
<td>N/A</td>
<td>6−10</td>
<td>4−8</td>
</tr>
<tr>
<td></td>
<td>$Q^2$</td>
<td>0.842 (0.036)</td>
<td>0.766 (0.079)</td>
<td>0.839 (0.094)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.930 (0.021)</td>
<td>0.979 (0.021)</td>
<td>0.971 (0.016)</td>
</tr>
<tr>
<td></td>
<td>RMSEP</td>
<td>10.2 (1.3)</td>
<td>11.8 (2.12)</td>
<td>9.973 (3.087)</td>
</tr>
<tr>
<td></td>
<td>RMSEC</td>
<td>8.35 (1.02)</td>
<td>4.18 (1.95)</td>
<td>5.179 (1.594)</td>
</tr>
<tr>
<td>B. cereus spore count</td>
<td>factors</td>
<td>N/A</td>
<td>3−6</td>
<td>2−7</td>
</tr>
<tr>
<td></td>
<td>$Q^2$</td>
<td>0.683 (0.225)</td>
<td>0.805 (0.159)</td>
<td>0.644 (0.280)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.860 (0.081)</td>
<td>0.944 (0.070)</td>
<td>0.964 (0.047)</td>
</tr>
<tr>
<td></td>
<td>RMSEP</td>
<td>2760000 (332000)</td>
<td>1310000 (974000)</td>
<td>2880000 (917000)</td>
</tr>
<tr>
<td></td>
<td>RMSEC</td>
<td>2340000 (395000)</td>
<td>1330000 (612000)</td>
<td>1090000 (749000)</td>
</tr>
<tr>
<td>B. subtilis spore count</td>
<td>factors</td>
<td>N/A</td>
<td>2−6</td>
<td>3−6</td>
</tr>
<tr>
<td></td>
<td>$Q^2$</td>
<td>0.116 (0.200)</td>
<td>0.419 (0.141)</td>
<td>0.692 (0.117)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.475 (0.100)</td>
<td>0.906 (0.093)</td>
<td>0.954 (0.030)</td>
</tr>
<tr>
<td></td>
<td>RMSEP</td>
<td>123000 (34900)</td>
<td>83700 (9430)</td>
<td>59100 (4990)</td>
</tr>
<tr>
<td></td>
<td>RMSEC</td>
<td>109000 (9900)</td>
<td>43500 (18300)</td>
<td>30800 (12000)</td>
</tr>
</tbody>
</table>

“Five different test data sets, each containing 40% of the whole data set were randomly chosen for each target and this table shows the mean values from the five data sets, with the standard deviation in brackets. Factors refers to the number of latent variables used in PLS or KPLS predictions, and these are chosen by training set validation. $Q^2$ and $R^2$ are the correlation coefficients from the test and training data sets respectively when the DPA concentration (or spore count) is predicted by the model. RMSEP and RMSEC are the root mean squared errors in the test and training data sets respectively for the same five plots.”

deviation error bars, again illustrating the excellent reproducibility of this colloidal SERS approach.

In Figure 3a we see a curved relationship as we approach the LOD of DPA using colloidal SERS is that by Bell et al. of 1 ppb. As far as we are aware, the lowest previously reported LOD of DPA using colloidal SERS is that by Bell et al. of 1 ppb (5.98 μM) using the sulfate-aggregated colloid; 2 orders of magnitude higher than the LOD we report here.

The SERS spectra of the DPA extracted from spores B. cereus and B. subtilis extracts achieved was 200 ppb (data shown in the Supporting Information, Figure S1): a tentative explanation for which is that due to its H stretch as the DPA concentration exceeds 100 ppb, as shown in Figure 3b. Above this range we see a linear region from 100−200 ppb, and then as the nanoparticle surface becomes saturated, the signal levels out.

These data provide us with a LOD (based on 3 times the standard deviation in the blank) of 5 ppb (29.9 nM) and a LOQ (based on 10 times the standard deviation in the blank) of 45 ppb (269 nM). However, when 40% of the data are randomly removed from the calibration model to form the test set (see Table 1), the practical LOD rises to 10.2 ppb (60.1 nM). As far as we are aware, the lowest previously reported LOD of DPA using colloidal SERS is that by Bell et al. of 1 ppb (5.98 μM) using the sulfate-aggregated colloid; 2 orders of magnitude higher than the LOD we report here.

It is difficult to relate the concentration of pure DPA to a number of spores accurately, as different bacterial types and even spores of the same type are likely to have differing DPA content. However, a rough estimate can be calculated based on 10% of the spore mass being DPA26 and the approximate spore mass determined by Zhang et al. of 1.8 × 10−11 g, giving 1.8 × 10−12 g of DPA per spore. Our LOD of 10.2 ppb therefore equates to approximately 1100 spores in our 200 μL sample volume; a level well below the infective dose of 10000 for inhalation anthrax.1

The vials used for all the analyses detailed in this report are those commercially available to fit DeltaNu spectrometers, which are cylindrical and 8 mm in diameter, requiring at least 400 μL of sample to record a spectrum (200 μL colloid and 200 μL sample, including aggregating agent and internal standard).

Given that the diameter of the laser beam is at most 1 mm, a custom-made vial would reduce the sample volume required further without reducing the sample interaction with the laser. We believe that this would reduce the number of spores detectable, by a factor of at least 10.

Univariate Analysis of DPA Extracted from Bacillus Spores. The SERS spectra of the DPA extracted from spores were processed in exactly the same manner as those of the pure DPA; although instead of DPA concentration, our calibration is now based on spore count. Representative spectra of the extracts compared to pure DPA are shown in Figure 1. The LOD of DPA based on the DPA pyridine ring/glutaric acid C−H stretch in the B. cereus and B. subtilis extracts achieved was 1 954 000 spores and 2 230 000 spores respectively. This is obviously very high but is an issue with respect to DPA extraction rather than the SERS analysis per se.

dx.doi.org/10.1021/ac303657k | Anal. Chem. 2013, 85, 3297–3302
Appendix I: Published work in original format

Multivariate Data Analysis for Estimating DPA Concentrations and Spore Counts. In addition to the univariate analysis, we investigated multivariate calibration using the supervised methods PLS and KPLS regression. Example plots from these analyses are shown in Figure 4, with a comparison between them and the univariate analysis shown in Table 1.

Although the univariate analysis is excellent (comparable to the multivariate analysis) for the pure DPA, we see that multivariate analysis is better for analysis of the spore extracts. This is perhaps as one would expect, given that the acidic extract samples contain other components of the bacterial spores rather than just the DPA, which may adversely affect the accurate measurement of the area under the curve in the univariate analysis. The standard deviations for the univariate analysis are for the most part lower than those for the multivariate techniques, which shows that in order to develop a better quantitative model, the algorithms have needed to use more subtle, and therefore variable changes in the spectra from positions other than under the most prominent pyridine ring feature.

This is the first study to use colloidal SERS to quantify DPA extracted from spores and to attempt to relate these data to spore numbers. The limitations of the method as it stands lie in the extraction of the DPA from spore samples, and this is why the lowest number of spores detectable is considerably higher than desirable. The spectroscopic method is clearly able to detect pure DPA at the required concentrations (equivalent to 100 spores in 20 μL of sample). Once an improved extraction technique has been developed, we are confident that detection of spores will be possible at the required low levels.

CONCLUSION

We have shown that SERS can achieve appropriate detection levels of DPA biomarker for Bacillus bacterial spores in food safety and bioterrorism contexts using cheap, easily produced colloidal substrates and portable instrumentation. The current limiting factor of this method is the extraction of the DPA from the spores but we have shown that multivariate data analysis can significantly improve quantification in complex extracts. With improved DPA extraction, we believe this approach will be readily suitable for rapid field analysis.

ASSOCIATED CONTENT

Supporting Information

Figure showing the area of the 2.0 mM glutaric acid C–H stretch peak that was used as an internal standard. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: roy.goodacre@manchester.ac.uk.

dx.doi.org/10.1021/ac303657k | Anal. Chem. 2013, 85, 3297−3302
Appendix I: Published work in original format

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank BBSRC for funding. D.C. also thanks Avacta Group Plc. for his CASE award studentship, as well as Dr. Catherine Winder for her expertise in the microbiological aspects of the experimentation.

■ REFERENCES

Appendix II: Additional publications
Publication 2

Optimization of Parameters for the Quantitative Surface-Enhanced Raman Scattering Detection of Mephedrone Using a Fractional Factorial Design and a Portable Raman Spectrometer

Samuel Mabbott, Elon Correa, David P. Cowcher, J. William Allwood, and Royston Goodacre®

School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

Supporting Information

ABSTRACT: A new optimization strategy for the SERS detection of mephedrone using a portable Raman system has been developed. A fractional factorial design was employed, and the number of statistically significant experiments (288) was greatly reduced from the actual total number of experiments (1722), which minimized the workload while maintaining the statistical integrity of the results. A number of conditions were explored in relation to mephedrone SERS signal optimization including the type of nanoparticle, pH, and aggregating agents (salts). Through exercising this design, it was possible to derive the significance of each of the individual variables, and we discovered four optimized SERS protocols for which the reproducibility of the SERS signal and the limit of detection (LOD) of mephedrone were established. Using traditional nanoparticles with a combination of salts and pH it was shown that the relative standard deviations of mephedrone-specific Raman peaks were as low as 0.51%, and the LOD was estimated to be around 1.6 μg/mL (9.06 × 10⁻⁶ M), a detection limit well beyond the scope of conventional Raman and extremely low for an analytical method optimized for quick and uncomplicated in-field use.

Recently, the recreational drugs market has seen an influx of designer drugs that exhibit many structural similarities to illegal substances but contain residual groups that help to bypass conventional drug laws. The failure of synthetic drug laws and the lack of effective classification has led to changes in the drug market. In general, mephedrone is considered to be a legal high. The drug is classified as a synthetic cathinone, a phenyl isopropylphenone, and amphetamine, which can be isolated from the flowering plant Catha edulis native to East Africa and the Arab Peninsula. Mephedrone belongs to a family of synthetic cathinone derivatives that also include butylone and methylene. In 1929, Saem de Burnaga Sanchez described the first chemical synthesis of mephedrone. One of the simplest synthetic routes for mephedrone is displayed in Scheme S1 in Supporting Information, whereby 4-methylpropophenone is used as the starting material. Structurally, mephedrone is very similar to amphetamine, as a phenyl ring and two carbon atoms separate amine groups in both. Similar short-term effects have also been reported, including increased alertness, confidence, and talkativeness. The effect of mephedrone on the body surrounds the stimulated release and inhibited reuptake of monoamine neurotransmitters; it is also believed that the drug promotes serotonin-mediated ADH activity similar to the mechanism of "euphoria." Both actions however are responsible for the short-term effects and long-term problems associated with the "highs" usage. The increased number of hospital admissions and deaths relating to the drug has served to highlight the dangers of mephedrone. In addition to the many side effects including skin rashes, minor amnesia, numbness, short-term memory effects, and headaches, the drug has also been implicated in more serious effects such as acute myocarditis. Mephedrone was made illegal in the U.K. on April 16, 2010, following similar reclassifications in many countries that had previously banned the drug. Mephedrone was graded as a class B drug, alongside previously illegal amphetamines. Even though mephedrone usage has been banned, surveys suggest that supplies of the drug are no longer available from "head shops" or the Internet, mostly come from street dealers who charge a mean price of £16 per

Received: September 3, 2012
Accepted: December 3, 2012
Published: December 3, 2012
A number of strategies for the qualitative and quantitative analysis of mephedrone and cathinone derivatives have been developed using liquid chromatography mass spectrometry (LC-MS) and gas chromatography mass spectrometry (GC-MS). Although these methods conduct a high level of qualitative and quantitative analysis, their inability to be operated by a nonspecialist, high sample costs, and lack of portability severely limit their application to in-field analysis of mephedrone. One previously unexplored method for the detection of mephedrone, which offers a high level of sensitivity while being accessible and portable, is surface-enhanced Raman scattering (SERS). Previous work has highlighted the use of SERS in the quantification of amphetamine and methamphetamine. Sülle et al. described how the modification of amphetamine using 2-mercaptonicotinic acid and a roughened silver foil made the detection of amphetamine and methamphetamine possible at concentrations of 19 ppm and 17 ppm, respectively. Sigmüller and co-workers used matrix-stabilized silver halides to detect MDMA contained in methanolic solutions at a concentration of $10^{-3}$ M, and Faulds and colleagues demonstrated that amphetamine sulfate could be detected at concentrations of $10^{-6}$ M when using colloidal suspensions. The coupling of mephedrone to gold or silver nanoparticles in theory should be a straightforward procedure; however, one SERS caveat is reproducibility of the Raman signal generated. It is therefore important that a robust and extensive optimization strategy is developed to maintain the spectral reproducibility to acceptable analytical standards without sacrificing low detection limits. The control of dynamic solution-based SERS systems can often be difficult, requiring careful selection of metallic nanoparticles, pH, aggregating agents, and aggregation time. To explore all the possible variables for the SERS optimization of just one analyte would involve a substantial amount of sample preparation and scrutiny, making the technique impractical. It is therefore desirable that mathematical strategies are applied to the experimental design to develop a much shorter but statistically significant and robust analysis strategy. One way of satisfying such a criterion is to use a fractional factorial design. Here we show that SERS can be combined successfully with a fractional factorial design that allows for the exploration of multiple variables while resulting in the optimization of a portable SERS system for the detection of mephedrone in solution.

**MATERIALS AND METHODS**

**Materials.** Trisodium citrate, sodium hydroxide, potassium hydroxide, sodium chloride, sodium sulfate, potassium nitrate, and potassium sulfate were supplied by Fisher chemicals (Fisher Scientific UK Ltd., Loughborough, U.K.). Gold(III) chloride trihydrate (99.9%) and silver nitrate (99.9999%) were purchased from Sigma Aldrich (Sigma Aldrich, Dorset, U.K.). Mephedrone was bought from an online supplier (www.buychem.co.uk), a site that has been closed down since the drug was made illegal. The purity of the mephedrone was verified using $^1$H NMR, $^{13}$C NMR, direct infusion accurate mass spectrometry, and elemental microanalysis. All solvents used in this analysis were HPLC grade, and all chemicals that were purchased were used as supplied.

**Preparation of Glassware.** All glassware used in the synthesis of gold and silver nanoparticles were washed thoroughly with aqua regia (3:1 HCl/HNO$_3$) and then scrubbed with soap solution and rinsed with copious amounts of water. The glassware was left to dry at 50°C before use.

**Synthesis of Gold Nanoparticles.** Gold nanoparticles were synthesized using the Turkevich method. A 100 mL volume of 50 mg gold(III) chloride hydrate solution was added to 850 mL of boiling water. While the solution was being vigorously stirred 50 mL of 1% trisodium citrate was added. Boiling was maintained for 1 h, but after around 30 min it was observed that the solution had turned to a deep red color, signifying nanoparticle formation. The sol was left to cool and stored at 4°C in a blacked out container. UV-–vis spectrophotometry identified the $\lambda_{\text{max}}$ of the nanoparticles as 529 nm.

**Synthesis of Silver Nanoparticles.** Silver nanoparticles were synthesized using a method outlined by Lee and Meisel. A 90 ng amount of silver nitrate was added to 500 mL of water and heated along with vigorous stirring. To a boiling silver nitrate solution was added 10 mL of 1% trisodium citrate dropwise. The solution was stirred for 1 h. Formation of a milky green sol indicated that the synthesis had been successful. The sol was left to cool and stored at 4°C in a blacked out container. UV–vis spectrophotometry identified the $\lambda_{\text{max}}$ of the nanoparticles as 421 nm.

**pH Adjustment of Nanoparticles, Salts, and Mephedrone Solutions.** To measure SERS response in a variety of environments, the gold and silver nanoparticles were separated into three 200 mL portions and pH corrected. The pHs selected for this purpose were 3, 7, and 10. Modification of the pH was carried out using either citric acid (5 M) or sodium hydroxide (5 M). The acid and base were specifically chosen because both citrate and sodium ions are present in the nanoparticle sol synthesis. One microliter portions of the acid or base were added to the nanoparticles until the desired pH had been reached. Three × 1 L flasks of water were also adjusted to pH 3, 7, and 10. These were then used to make up salt solutions at a concentration of 0.5 M and to dissolve the mephedrone to a concentration of 5.65 $\times$ 10$^{-3}$ M. Altogether, there were twelve salt solutions (three pHs × four different salts) and three mephedrone solutions (pH 3, 7, and 10).

**SERS Analysis.** SERS spectra were collected using a DeltaNu Advantage benchtop Raman spectrometer (Intevac Inc., Santa Clara, CA). The instrument is equipped with a 633 nm HeNe laser with a power output of 3 mW at the sample. Spectra were acquired for 20 s over a range of 200–3400 cm$^{-1}$, and spectral resolution was 10 cm$^{-1}$. Solution samples were placed in an 8 mm glass vial and subjected to laser irradiation once loaded into the sample cell attachment. The instrument was calibrated using toluene to find the ideal distance from laser to sample.

**Fractional Factorial Design.** The SERS analysis described in this work follows the model of a quantitative experiment. In this type of experiment, data were obtained for dependent variables (the specific SERS peaks under study in our case) as a function of a set of independent variables (the mixtures of metal, salt, and pH). Considering all the parameters under study, the full factorial design or the total number of unique quantitative experiments that could be generated is $[(2 \text{ metals}) \times (7 \text{ concentrations}) \times (4 \text{ salts}) \times (10 \text{ concentrations}) \times (3 \text{ pH}) + (2 \text{ metals}) \times (7 \text{ concentrations}) \times (3 \text{ pH})] = 1722$ mixtures. This is a large number of experiments to perform in the lab, and...
Appendix II: Additional publications

Table 1. Tentative SERS Band Assignments for Mephedrone \(^{55,60}\)

<table>
<thead>
<tr>
<th>Raman shift (cm(^{-1}))</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>797</td>
<td>para-disubstituted benzene ring vibration</td>
</tr>
<tr>
<td>669</td>
<td>CH(_2) rocking vibration combined with CH stretch</td>
</tr>
<tr>
<td>1034</td>
<td>CH in plane deformation from para-disubstituted benzene</td>
</tr>
<tr>
<td>1134</td>
<td>aliphatic amine symmetrical C–N–C stretch</td>
</tr>
<tr>
<td>1323</td>
<td>para-disubstituted benzene ring vibration</td>
</tr>
<tr>
<td>1356</td>
<td>CH(_2) symmetrical deformation</td>
</tr>
<tr>
<td>1666</td>
<td>benzene derivative C=C=C stretch</td>
</tr>
<tr>
<td>1078</td>
<td>C–C stretch</td>
</tr>
<tr>
<td>1282</td>
<td>CH(_2) symmetrical stretch</td>
</tr>
<tr>
<td>1039</td>
<td>CH stretch of aromatic</td>
</tr>
</tbody>
</table>

Figure 3 show the results of PCA carried out on the extracted intensities from all 288 experiments. Figure 3A–C shows separation based on the type of metal, concentration, type of salt, and pH, respectively. These plots show that silver nanoparticles are responsible for a much greater enhancement factor for mephedrone than gold, a fact which is well recognized in the field of SERS. Although the concentration of salt can be variable at pH 7, the presence of salts at this pH is influential in propagating an appreciable amount of SERS enhancement; however, at pH 3, no salt is needed to generate increased Raman signal. This strongly suggests that the signal enhancement is directly related to the degree of aggregation imposed on the systems. At pH 7, the nanoparticles are stable and do not self-aggregate, and the salt is therefore needed to bring the nanoparticles into close contact, resulting in the formation of areas of high enhancement. It is well documented that at pH 3, the nanoparticles can self-aggregate, thus resulting in an increased signal with the salt absent. While the percentage of salt is variable at pH 7, the results do show that 70% silver nanoparticle content is a vital requirement for the signal to be maximized. Samples at pH 3 show no such dependence on nanoparticle content, with all percentages of salt being represented among some of the most optimal systems (Figure 3D).

Figure 3E shows the model coefficients of a partial least-squares regression \(^{54}\) applied to the 10 SERS peaks under study and using metal, salt, and pH as the independent variables. The absolute values of these coefficients indicate the significance of each variable in the determination of signal intensity; the two most crucial elements of the systems are the type of metal and pH. Although the type of salt is of lesser significance, it does have a major influence on the spectral intensity as demonstrated by the PCA plots. From the factorial design and Figure 3D (which for each of the 288 experiments performed shows a circle proportional to the level of the SERS enhancement), it was possible to identify four SERS protocols which gave the greatest enhancement of the mephedrone signal. These were then subjected to reproducibility and LOD tests. Two experimental solutions were selected that had a pH of 3 and no salt present, while another two protocols were chosen which had a pH of 7 with salt present. The systems selected were as follows:

- pH 3 (no salt):
  - pH Opt 1: 10% silver sol (50 \(\mu\)L), 10% mephedrone solution (50 \(\mu\)L), and 80% water (400 \(\mu\)L)
  - pH Opt 2: 40% silver sol (200 \(\mu\)L), 10% mephedrone solution (50 \(\mu\)L), and 50% water (250 \(\mu\)L)

- pH 7 (including salt):
  - pH Opt 1: 10% silver sol (50 \(\mu\)L), 10% mephedrone solution (50 \(\mu\)L), and 80% water (400 \(\mu\)L)
  - pH Opt 2: 40% silver sol (200 \(\mu\)L), 10% mephedrone solution (50 \(\mu\)L), and 50% water (250 \(\mu\)L)
Appendix II: Additional publications

Figure 1. Example spectra generated using gold colloid and the factorial design. (A) 40% Colloid (300 µL), no salt, 10% analyte (50 µL), and 50% water (250 µL) at pH 3. (B) 70% Colloid (350 µL), no salt, 10% analyte (50 µL), and 20% water (100 µL) at pH 3. (C) 40% Colloid (200 µL), 10% K2SO4 (50 µL), 10% analyte (50 µL) at pH 3 and 40% water (200 µL). (D) 70% Colloid (350 µL), 2% KNO3 (10 µL), 10% analyte (50 µL), and 18% water (50 µL) at pH 7.

Figure 2. Example spectra generated using silver colloid and the factorial design. (A) 10% Colloid (50 µL), 2% K2SO4 (10 µL), 10% analyte (50 µL), and 78% water (390 µL) at pH 3. (B) 10% Colloid (50 µL), 6% KNO3 (30 µL), 10% analyte (50 µL), and 74% water (370 µL) at pH 10. (C) 70% Colloid (350 µL), 0% salt (50 µL), 10% analyte (50 µL), and 20% water (100 µL) at pH 7. (D) 70% Citrate (350 µL), 10% K2SO4 (50 µL), 10% analyte (50 µL), and 10% water (50 µL) at pH 7.

Salt Opt 1: 70% silver sol (350 µL), 10% mephedrone solution (50 µL), 18% water (90 µL), and 2% KNO3 solution (10 µL)
Salt Opt 2: 70% silver sol (350 µL), 10% mephedrone solution (50 µL), 14% water (70 µL), and 6% NaCl solution (30 µL)

From now on, the optimized systems will be referred to as pH Opt 1, pH Opt 2, Salt Opt 1, and Salt Opt 2 as outlined above.

Reproducibility Assessment of Optimized Systems.
To carry out the reproducibility analysis, six replicate samples were prepared for each of the four optimized systems identified, using the fractional factorial design. The ten peaks used for
Appendix II: Additional publications

![Figure 3. PCA scores plot computed on the SERS intensities of the 10 mephedrone peaks under study. Plots A-D are identical and only the labeling is different: experiments labeled (A) by metal type and concentration, (B) by salt type and concentration, and (C) by pH value. Plot D identifies each experiment using a closed circle, and the size of the circle is proportional to the enhancement obtained over the 10 mephedrone peaks under study, i.e., sum of the 10 peaks for that particular experiment. (E) PLS model coefficient values considering mephedrone peak intensities as dependent variables.](image)

Finding the optimum SERS systems (Table 1) were also used to assess the systems reproducibility. A $5 \times 10^{-4}$ M mephedrone solution was used in all the assessments. Initially, the maxima of each of the 10 peaks were defined and the minima (start and end of the peaks) were selected using data points three places (bins or columns) either side of the maxima. This method of peak picking allowed for fast data processing, bypassing any unnecessary manual assignments. Each of the individual peaks were then extracted and baseline corrected using an asymmetric least-squares algorithm in Matlab. It was essential to make sure that the minima of each of the peaks had a Y value equal to 0, to remove any unavoidable background shifts. Two individual methods were used to extract peak area information: these were a trapezoidal methodology which calculates the definite integral of the peaks and a sum method which estimates the area of a curve based on the intensity of data points contained within the defined peak parameters. The peaks and their corresponding areas can be seen in Tables 2 and 3 (pH and salts). Reproducibility analysis of pH Opt 1 and pH Opt 2 show very low, acceptable relative standard deviation (RSD) values for each of the individual peaks analyzed from the two conditions. The RSDs for pH Opt 1 range from 1.09 for the peak at 2913 cm$^{-1}$ to 6.87 for the peak at 1203 cm$^{-1}$. pH Opt 2 also has a minimum RSD for the peak at 2913 cm$^{-1}$ while its highest RSD is demonstrated by the peak at 699 cm$^{-1}$. The reproducibility analysis performed on the pH 7 systems including salt also displayed very low RSDs with a range of 0.51–8.89 recorded for Salt Opt 1 and RSDs ranging from 1.80 to 11.88 for Salt Opt 2. All four optimal solutions identified by the factorial design demonstrated low levels of signal variability, which is highly encouraging for a technique that is often associated with a lack of reproducibility. Both methods used in the analysis of peak area generated values that were in agreement, as only four peaks analyzed displayed an RSD of greater than 1%. Therefore, both methods are suitable for the assessment of peak area. Examples of the raw spectra retrieved using each of the four conditions are shown in Figure 4. The spectra show little change with respect to whether the pH or salt optima were used; this highlights the fact that the interaction of the nanoparticles with the mephedrone is similar under both conditions. Subtle differences in the spectra are difficult to explain but are most likely due to small differences in the system dynamics created by the continuous exchange of citrate and mephedrone molecules at the surface of the nanoparticles. It is assumed that the interaction occurs between the silver and the amino moiety of the drug. It can however be said with a reasonable degree of accuracy that the interaction is noncovalent due to the absence of an Ag–N vibration around 210–245 cm$^{-1}$. Establishing the LOD of MeMephedrone. As citrate is used in the sol synthesis to stabilize the nanoparticles, it is essential that spectral bands arising from the mephedrone were discriminated from the background citrate bands. Although the citrate peaks cannot be seen when the mephedrone concentration is sufficiently high, they do appear in the spectra when the mephedrone is at low concentrations and certainly become visible in the colloidal blanks (Figure S1, Supporting Information). If this selective step was omitted, it could potentially lead to false assessment of the LOD due to spectral contributions from citrate scattering. Figure S1 shows the SERS
Appendix II: Additional publications

Table 2. Summary of the Results for the Reproducibility Analysis of the Two pH Optima (A: pH Opt 1, B: pH Opt 2)

<table>
<thead>
<tr>
<th>Raman shift (cm(^{-1}))</th>
<th>peak area (trapezoidal method)</th>
<th>peak area (sum method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>start</td>
<td>end</td>
<td>maxima</td>
</tr>
<tr>
<td>pH Opt 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>768</td>
<td>806</td>
<td>797</td>
</tr>
<tr>
<td>959</td>
<td>973</td>
<td>949</td>
</tr>
<tr>
<td>1025</td>
<td>1044</td>
<td>1034</td>
</tr>
<tr>
<td>pH Opt 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>788</td>
<td>806</td>
<td>797</td>
</tr>
<tr>
<td>999</td>
<td>978</td>
<td>969</td>
</tr>
<tr>
<td>1025</td>
<td>1044</td>
<td>1034</td>
</tr>
<tr>
<td>pH Opt 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>768</td>
<td>806</td>
<td>797</td>
</tr>
<tr>
<td>959</td>
<td>973</td>
<td>949</td>
</tr>
<tr>
<td>1025</td>
<td>1044</td>
<td>1034</td>
</tr>
<tr>
<td>pH Opt 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>788</td>
<td>806</td>
<td>797</td>
</tr>
<tr>
<td>999</td>
<td>978</td>
<td>969</td>
</tr>
<tr>
<td>1025</td>
<td>1044</td>
<td>1034</td>
</tr>
<tr>
<td>Salt Opt 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>start</td>
<td>end</td>
<td>maxima</td>
</tr>
<tr>
<td>Salt Opt 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>start</td>
<td>end</td>
<td>maxima</td>
</tr>
</tbody>
</table>

Table 3. Summary of the Results for the Reproducibility Analysis of the Two pH Optima (A: Salt Opt 1, B: Salt Opt 2)

Spectra of 5 × 10^{-4} M mephedrone with the colloidal blanks overlaid; the highlighted areas show the positions of the five peaks used to calculate the LOD of the amphetamine derivative, and these were established at 1188 cm\(^{-1}\), 1212 cm\(^{-1}\), 1606 cm\(^{-1}\), 2922 cm\(^{-1}\), and 3650 cm\(^{-1}\).

While earlier in this study the peak area was used to explore reproducibility, here peak intensity was used as an alternative. The reason behind using intensity-based evaluation of the peaks is that at decreasing concentrations of mephedrone, peaks often shift and become distorted as the background signal arising from the citrate-stabilized nanoparticles becomes more prominent. One adverse effect of this is that while the minima of the peak from which the area is being derived stays constant, background contributions still arise from the mephedrone/citrate-covered surface especially when the drug is at low concentrations, hence resulting in values being falsely attributed.

...
Appendix II: Additional publications

Figure 4. Example raw SERS spectra of mephedrone (5 × 10^{-4} M) acquired using all the optimized conditions identified by the factorial design. (A) pH Opt 1: 10% silver sol (50 μL), 10% mephedrone solution (50 μL), and 80% water (400 μL) at pH 4.3. (B) pH Opt 2: 40% silver sol (50 μL), 10% mephedrone solution (50 μL), and 18% water (90 μL) and 2% KNO₃ solution (10 μL). (C) 70% Silver sol (350 μL), 10% mephedrone solution (50 μL), 14% water (70 μL), and 6% NaCl solution (30 μL).

Although the aggregation is influential in the propagation of SERS signal from the mephedrone, controlling the extent of the aggregation can often be difficult, resulting in certain concentrations having a large variance in peak intensity. This lack of control is most likely to have resulted in the large standard deviations seen in the LOD plots for pH Opt 2. The Salt Opt 1 LOD plots both show a greater degree of linearity, with R² values ranging from 0.95 to 0.99. The addition of a salt allows for much greater control over the aggregation state of the system, and this is reflected in the low standard deviations exhibited across all mephedrone concentrations and analyte peaks.

The LOD is defined as the concentration of the analyte that is required to produce an instrument response that is three times as large as the standard deviation of the noise level, and this equation is supplied in the Supporting Information. The LODs established for each of the peaks under the four conditions are displayed in Table 4. The lowest LOD of 9.07 × 10^{-3} M is

Table 4. Limits of Detection Established Using the Intensity of Signal Arising from the Five Assignable Mephedrone Peaks

<table>
<thead>
<tr>
<th>peak maxima (cm⁻¹)</th>
<th>limit of detection, M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salt Opt 1</td>
<td>Salt Opt 2</td>
</tr>
<tr>
<td>1194</td>
<td>1.05 × 10⁻⁴</td>
</tr>
<tr>
<td>1212</td>
<td>1.07 × 10⁻⁴</td>
</tr>
<tr>
<td>1606</td>
<td>9.70 × 10⁻⁴</td>
</tr>
<tr>
<td>2922</td>
<td>8.95 × 10⁻⁴</td>
</tr>
<tr>
<td>3050</td>
<td>1.32 × 10⁻³</td>
</tr>
</tbody>
</table>

Salt Opt 1, Salt Opt 2, pH Opt 1, and pH Opt 2 refer to the four optimal conditions identified from the factorial design.
recorded for mephedrone when Salt Opt 1 and peak 1666 cm\(^{-1}\) were analyzed. However, Salt Opt 2 appears to offer the most consistent LODs which are all within 2.03 × 10\(^{-5}\) M to 9.60 × 10\(^{-5}\) M. The lowest LOD recorded for the pH systems arises from pH Opt 1 for the peak at 1184 cm\(^{-1}\) which has an LOD of 6.09 × 10\(^{-5}\) M.

The lowest detection limit of 9.07 × 10\(^{-6}\) M translates to 1.6 μg/mL mephedrone. This is a fully acceptable LOD for mephedrone present in solution and has shown improvement over previous detection limits for amphetamines using SERS as highlighted earlier in this paper. It is approximated that the single user dosage ranges between 5 and 90 mg, with around 1–4 g being consumed in a session.\(^{630}\) This demonstrates that SERS coupled with a portable spectrometer is capable of in-field analysis and is sufficiently sensitive to detect mephedrone at concentrations well below the detection limit of conventional Raman\(^{60}\) estimated to be ~0.1 M.

## CONCLUDING REMARKS

It has been demonstrated that when a fractional factorial design is employed, it is possible to explore multiple SERS parameters and optimise the systems for the detection of specific analytes of interest, in this case the illicit drug mephedrone. The calicids used in this experiment were unspecific, meaning they were not specifically adapted at high cost to invoke favorable binding to the mephedrone; this shows that the technique is accessible to non-SERS specialists and also extends the use of conventional silver nanoparticles and SERS to a variety of analytes. The focus on developing a methodology, which can be applied to real world problems, has been achieved, as the LODs and reproducibility of the SERS phenomenon is demonstrated at a highly sensitive and accurate level.

## ASSOCIATED CONTENT

\* Supporting Information Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org.\*

## AUTHOR INFORMATION

**Corresponding Author**

*Email: roy.goodacre@manchester.ac.uk*

**Notes**

The author declares no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank the RSC and EPSRC for funding S.M.'s Analytical Ph.D. Studentship. E.C. and R.G. also acknowledge the EU Commission (www.epi7project.com) under the 7th Framework Programme for Research and Technological Development.

## REFERENCES

Appendix II: Additional publications

Analytical Chemistry

(57) Eds, P. H. C.; Boles, H. F. M. Baseline correction with asymmetric least squares smoothing; Leiden University Medical Center Report, 2005.
Appendix II: Additional publications

Publication 3

**Introduction**

The properties of light have been of interest to experimentalists for millennia. From the publication of Ibn al-Haytham’s seven volume treatise the ‘Book of Optics’ (Kitab al-Manazir) in the early 11th century (leading him being regarded as the father of modern optics), through to one of the very first publications in a scientific journal, Isaac Newton’s paper on the theory of the properties of light itself followed some years later by his famous book, Opticks, in 1704. Whilst separated by seven centuries, those two polymaths and their respective bodies of work shared similarities, perhaps the most important being that both were the result of systematic and methodical experimentation. In the century following the publication of Newton’s...
Opticks, as early as the mid-1800s, the interaction of light within tissue was being used by physicians to assist in disease recognition. It was during a sea voyage from India to England in 1921 that the brilliant Indian physicist C.V. Raman, another renowned experimentalist, undertook some on-board experiments which were later submitted to *Nature* in a letter called the “The Colour of the Sea.” Unable to accept Lord Rayleigh's explanation that the colour of the sea was just a reflection of the colour of the sky, Raman’s experiments showed that the colour of the sea was in fact a direct result of the molecular scattering of light and independent of absorption or the reflection of light from the sky. This was very closely followed by another letter to *Nature* concerning the molecular scattering of light in liquids and solids. These experiments opened up a deep interest in C.V. Raman and a new field of research on his return to Calcutta (Kolkata) on the scattering of light, as well as the publication of another article on the molecular diffraction and quantum structure of light in the following year.86 Raman and collaborators such as K.S. Krishnan began a series of seminal experiments concerning the scattering of light in a large number of liquids, as well as theories about the potential applications of their experiments, which culminated in their discovery of the inelastic scattering effect named after Raman in 1928 on 28 February,87 and his award of the Nobel Prize for Physics in 1930. This discovery was also independently observed by Landsberg and Mandelstam later in 1928 (see Fig. 1 for a timeline of events in Raman spectroscopy). There was a great deal of interest in the Raman effect and this not only added the fundamental understanding about the quantum nature of light, and its interaction with matter at the molecular level, but also opened up a completely new area of optics and spectroscopic research that is, particularly in terms of biological and biomedical applications,88–90 accelerating at a greater rate during the last decade than at any time since its inception. Whereas infrared (IR) spectroscopies measure the absorption of energy, Raman spectroscopy measures the exchange of energy with electromagnetic (EM) radiation of a particular wavelength, usually provided by a monochromatic light source such as a laser in the visible or near-IR portion of the EM; although it is also possible to conduct experiments in the near- and deep-UV. From the exchange in EM energy a measurable Raman shift in the wavelength of incident laser light is observed, this is also referred to as the inelastic light scattering effect.82,84 It is usually the Stokes shift which is measured, as

**Lorna Ashton** studied for her BSc with the Open University before joining the University of Manchester where she obtained her PhD in Biomedical Sciences and Raman Optical Activity. She has since worked as a post-doctoral research associate specialising in Raman Spectroscopy and two-dimensional correlation analysis for the characterisation of conformational transitions in biomolecules. Lorna is currently working as part of a BBSRC-funded project for rapid evolution of enzymes and synthetic microorganisms developing Raman spectroscopy as a high-throughput analytical technique for industrial bioprocesses.

**Steve O’Hagan** gained BSc and MSc degrees in Chemistry at the University of Manchester. The MSc research was to characterise ‘Molecular Beam Sources’ using mass spectrometry and TOF kinetic energy measurements; computer simulations were also used. Steve gained a PhD in Chemistry at the University of Warwick; several mass spectrometry and chemometric techniques were employed to analyse engine oils and additives. After several years working for commercial laboratories, he joined the University of Manchester as a Computer Officer, working with Doug Kelk’s and Roy Goodacre’s research groups. Interests include: genetic programming, laboratory automation; analytical laboratory data analysis; chemometrics; as well as scientific visualization.

**Roy Goodacre** is a PhD graduate from the University of Bristol (UK) where he studied mass spectrometry of microbial systems. After a postdoc, Welcome Trust fellowship and lecturership in the University of Wales, Aberystwyth, he is now Professor of Biological Chemistry at the University of Manchester (UK). His group’s main areas of research [http://www.biospec.net/](http://www.biospec.net/) are broadly within analytical biotechnology, metabolomics and systems biology. His expertise involves many forms of Raman spectroscopy (including deep UV resonance Raman and SEIRS), FT-IR spectroscopy, and mass spectrometry, as well as advanced chemometrics, machine learning and evolutionary computational methods. He is Editor-in-Chief of the journal Metabolomics, on the Editorial Advisory Boards of *Analyst* and *Journal of Analytical and Applied Pyrolysis*, a founding director of the Metabolomics Society and a director of the Metabolic Profiling Forum.
Appendix II: Additional publications

Critical Review

Event/discovery

2011 First Raman microscope taking Raman spectra in situ
2010 Visible in vivo ATR Raman imaging by Stott et al.
2010 GID-DOM proposed by Graham, Facchini, Douce and Narasimhan
2009 Doyle and Smith awarded Nobel Prize in Physics for CEMS
2005 GID (renamed) by Paultre
2000 LIFS developed by Volker Scholz
1997 Single molecule NIS by Karin Munk
1996 Significant increase in commercial instrumentation
1993 Enhancement of 100 times by Richard E. Davey
1992 Use of Ag and Au reagents for GID reported by Lee and Mered
1990s CCD detectors used for Raman
1980s Three-spiral cross-sections with Ramanscope
1979 1H nuclear and colloid surface enhanced Raman scattersing (SERS)
1976 Raman coupled with fluorescence
1970s Detection of charge-coupled devices (CCD) by K.D. Burke and Steven Smith
1964 Raman microscopes and Raman microspectroscopy
1961 Use of Ge, Si and NaCl crystals for Raman scattering
1960s Raman scattered light source for Raman spectrometers
1952 Application of Raman spectrometers
1920 Observation of resonance-Raman (RR) spectra
1916 Excitation of Raman scattering
1898 E.O. Hutchinson’s monograph on ‘The Molecular Diffusion of Light’
1897 Albert Einstein’s paper on the photoelectric effect

Fig. 1 Timeline of events and discoveries in the history of Raman spectroscopy. Nobel Prize awarded to C. V. Raman in the ICS of Patna Science, India, by Prof Paul C. J. R. Smir, School of Chemistry, University of Manchester.

This has a higher probability of occurring than an anti-Stokes. It is important to note that this shift is complementary to IR absorption and a spectroscopic fingerprint of the same sample can be analyzed and constructed by both vibrational spectroses. Whilst mid-IR spectroscopy is known to be intensely sensitive to and a high absorber of water, this is generally not the case with Raman spectroscopy as water is a weak scatterer. For biomedical and routine clinical applications and low laser output power at the point of contact, this allows for the direct collection of Raman spectra. The utility of Raman spectroscopy has been demonstrated for a diverse and wide range of potential biological and biomedical applications, such as bacterial identification, chemical hazards and illicit substance detection, as well as food and product authentication, with great deal of interest and research into its potential for disease diagnosis and use in biomedical applications seen during the last decade (Fig. 2).

Some of the previously documented limitations of Raman spectroscopy for biomedical applications have been discussed, including issues such as weak scattering signals, subsequently.

Fig. 2 Results of bibliometric analysis of the number of publications (A) and citations (B) per year listed on ISI Thomson Web of Science® (http://wut.imnas.missouri.edu) for the period 1992–2012, using the search terms (A) Raman AND disease, (B) spatially offset Raman AND disease, (C) SERS AND disease, (D) Raman endoscopy, (E) Raman imaging AND disease.

This journal is © The Royal Society of Chemistry 2013
Appendix II: Additional publications

long spectral acquisition times, fluorescence from biological samples, and interference from silica within fibres. However, within this exciting and highly research active era of biomedical Raman spectroscopy, and with the hard work and application of experimentalists (in the long tradition of those already mentioned above), these so-called limitations are being overcome by groups of scientists and engineers around the globe who are not content to remain within the bounds of current knowledge, or the confines of commercially available optics for that matter, and who are constantly pushing the field forward. This introductory overview focuses on some of the recent developments within this exciting field, highlighting a small number of high-impact studies in imaging, endoscopy, stem cell research, and other recent developments such as spatial offset Raman scattering (SORS), coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS), amongst others.

Imaging

Biomedical imaging is an extremely useful and important field which allows for the collection and processing of highly complex biochemical and physiological data, and the creation, manipulation, and in-depth analysis of three-dimensional colourised images to aid molecular sensing, drug transport, characterization of cells and tissues, and, perhaps the ultimate goal, the rapid diagnosis of disease (vide infra). Label-free optical imaging, particularly in vivo, would be highly advantageous as dyes or fluorescent labels, which are needed as contrast agents, can be toxic or perturbative to the cell/tissue. Optical techniques have also been said to have the potential to be complementary to existing techniques such as magnetic resonance imaging (MRI), and they offer superior sensitivity and high spatial resolution (compared to MRI, see Fig. 3). Whilst mass spectrometry (MS) has mass appeal for metabolomics for example and can be used for chemical imaging of tissues and cells — including in 3D — as well as being combined with vibrational spectroscopy, imaging MS is a destructive analysis and the practical spatial resolution for MALDI-MS is 50-200 μm and for SIMS is 1 μm, thus Raman spectroscopy has considerable opportunities to shed light on disease. In any review of biomedical imaging, and particularly one focussed on Raman spectroscopy, it would be a heinous omissions not to include any one of a number of studies by the Sunney Xie group. A major contributor to biomedical imaging, based at Harvard University (http://bernstein.harvard.edu), this group’s body of work includes both coherent Raman scattering techniques, comprising of coherent anti-Stokes Raman scattering (CARS) microscopy and stimulated Raman scattering (SRS) microscopy. Descriptions of these techniques can be found in Table 1 and it should also be noted that the signal from CARS is non-linearly proportional to species concentration whilst the signal from SRS is linearly proportional.
Appendix II: Additional publications

Table 1 Definition of some of the main approaches of Raman spectroscopy discussed in this review

<table>
<thead>
<tr>
<th>Technique</th>
<th>Acronym</th>
<th>Short definition</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface enhanced Raman spectroscopy</td>
<td>SERS</td>
<td>Requires close proximity/adsorption onto a roughened metal surface, a colloidal solution or a roughened electrode (usually Ag or Au). Enhancement explained by two processes: an electromagnetic enhancement effect (thought to dominate) and a charge transfer mechanism, known as chemical enhancement. Has a fluorescence quenching effect. Can tune to a specific chromophore for additional resonance enhancement (known as SERS). Enhancements over normal Raman scattering of typically $10^7$ to $10^9$</td>
<td>19, 122, 136 and 137</td>
</tr>
<tr>
<td>Coherent anti-Stokes Raman spectroscopy</td>
<td>CARS</td>
<td>A multiphoton form of Raman spectroscopy based on a non-linear conversion of two lasers into a coherent high intensity beam in the anti-Stokes region. The emission is usually many orders of magnitude greater than spontaneous Raman scattering. Useful for obtaining spectra of fluorescent samples. Nonresonant background can complicate spectral assignment, limit sensitivity, and affect quantitative interpretation.</td>
<td>37, 38, 41, 45 and 118</td>
</tr>
<tr>
<td>Stimulated Raman scattering</td>
<td>SRS</td>
<td>A multiphoton technique analogous to stimulated emission, where two lasers collide on a sample. The sample is excited by collinear and tightly focused pump and Stokes beams. When the differences in known frequences match a molecular vibration in the sample, the Stokes beam intensity increases and pump beam intensity decreases as a result of the coherent excitation of molecular vibration. Unlike CARS, it does not exhibit a nonresonant background and is a good technique for sensitive, high spatial resolution 3D imaging.</td>
<td>25, 26, 19, 63, 53 and 119</td>
</tr>
<tr>
<td>Spatially offset Raman spectroscopy</td>
<td>SOUS</td>
<td>Raman spectra are collected from locations spatially separated from the point of laser illumination on the sample surface. SOUS allows for the isolation of chemically rich spectral information from distinct substrates or layers and through other barriers, not accessible via spontaneous Raman. Typical wavelengths used for biological tissue are 785 or 830 nm. Ideal for detecting disease in cells and tissue underlying other tissue types, such as bone through skin, cancer cells through muscle and lipid tissue for example.</td>
<td>25, 26, 19, 63, 53 and 119</td>
</tr>
</tbody>
</table>

As briefly mentioned above, the signal from spontaneous Raman scattering is known to be weak, with potentially long integration times, which impacts by significantly reducing imaging speed (an important factor as in vivo samples are in constant motion at the micro scale). Although these nonlinear methods were first invented in the 1960s, coherent Raman scattering techniques can be said to have very recently surmounted any technical difficulties and imaging speed hurdles by enhancing the Raman signal level by up to five orders of magnitude, and image collection speeds by three orders of magnitude. Indeed, on the subject of speed, one of the more memorable articles propelling CARS microscopy as a rapid method over a decade ago (for the identification of bacterial spores), had the unforgettable acronym FAST-CARS (femtosecond adaptive spectroscopic techniques for coherent anti-Stokes Raman spectroscopy). More recent reports have highlighted some of the potential limitations of CARS, which has been said to make quantitative interpretation and applications other than lipid imaging unnecessarily challenging. These limitations can be particularly noticeable in the fingerprint region where spectral signals are said to be congested and separating out weak from strong signals in this region is thus more problematic for CARS microscopic imaging. That being said, CARS microscopy has been successfully demonstrated as a biomedical imaging method for a range of ex vivo biological samples such as brain structures, colon tissue, and arterial tissue for example, as well as in vivo studies including spinal nerve tissue and atherosclerotic plaque deposits in animals, with a recent report also demonstrating the use of CARS microscopy in skin imaging of humans in vivo. Recent articles within the last half decade have explored SERS as an alternative imaging technology, which, unlike CARS, does not exhibit a nonresonant background and so is inherently more quantitative, though like other nonlinear multiphoton techniques (including CARS), does allow for sensitive, high spatial resolution 3D imaging. Whilst initial reports demonstrated stimulated Raman scattering's advantages over CARS (e.g. a lack of nonresonant background complications), it was said to not be suitable for biomedical due to sample photodamage from excessive laser power and had a limited spatial and spectral resolution as well as a slower image acquisition rate than CARS. These challenges were rapidly overcome by the Xic group the following year using a multiaperture approach which lowered peak rates by three orders of magnitude (resulting in no photodamage), optimised spectral resolution, increased sensitivity by four orders of magnitude than the previous year's report and surpassed the detection limit previously stated for SERS microscopy.
Not content with these and other leaps forward, three bio-imaging applications were then elegantly presented by Xie and co-workers. The first application monitored and imaged the uptake of polysaturated omega-3 fatty acids by living human cancer cells, specifically eicosapentaenoic acid (EPA). Using the Raman band at 3015 cm\(^{-1}\) (attributable to unsaturated fatty acids), and in contrast, the Raman band at 2920 cm\(^{-1}\) whose peak intensity is similar for saturated and unsaturated fatty acids, they were able to use SERS to follow the uptake of EPA by living cells, concluding that EPA is taken up by the cells and more strongly enriched in lipid droplets compared to other cellular organelles. The second application presented the potential for tissue imaging without the requirement for staining, highlighting the 3D sectioning capability and subcellular resolution of SERS, using the CH\(_2\) stretching vibration at 2845 cm\(^{-1}\) (see Table 2 for Raman band frequencies of biological interest).

This was demonstrated in a variety of mouse tissues, from neuron bundles in corpus callosum (highlighting myelin sheaths), thick brain tissue, and from depth profiles of intact tumors in additional comparison SERS and CARS images of stratum corneum (visually illustrating how the nonresonant background from CARS can complicate image interpretation). Finally, this seminal series of experiments showed the use of SERS to monitor drug delivery, namely deuterated dimethyl sulfoxide (DMSO), a skin penetration enhancer and retinoic acid (RA) used to treat a range of skin conditions and acute promyelocytic leukemia. Drug delivery into fresh mouse skin was monitored by tuning into the vibrations for DMSO at 670 cm\(^{-1}\) and RA at 1570 cm\(^{-1}\) (as well as lipids of subcutaneous fats at 2845 cm\(^{-1}\)). This series of experiments elegantly demonstrated the potential for SERS as a new approach for studying pharmacokinetics in vivo, sensitive label-free imaging, and molecular sensing in 3D in living cells and tissue.

To highlight the range of new possibilities unleashed by the use of SERS, we now consider the potential for Raman imaging on a new platform, in vivo imaging in biomedicine, and perhaps acutely mindful of the challenges required to achieve this goal, as well as the fact that SERS imaging had not yet been accomplished in vivo (in animals or humans), an article was published in 2010 demonstrating that in vivo SERS was not only possible (both in mice and humans), but that it could also be performed at video-rate speeds. This high-speed imaging was achieved by modulating the intensity of the Stokes beam at 20 MHz and the group building their own custom-built system raster scanned across a sample with a line rate of 8 kHz (100 ns per pixel at 512 × 512 pixels with up to 25 frames per s), increasing imaging speed by three orders of magnitude. In addition, to significantly increase the collection of backscattered light from in vivo samples further in-house modifications were undertaken, such as changing the geometry of the photodetector and microscope objective. This involved exciting light from the objective through an aperture in the detector placed between the microscope objective and the sample, and specially designing a filter to block the modulated Stokes beam whilst transmitting the pump beam.
Appendix II: Additional publications

Critical Review

The effectiveness of this new video-rate SRS imaging technique was then demonstrated through a series of experiments such as skin imaging in living mice. The CH\(_2\) stretching vibration at 2950 cm\(^{-1}\) was shown to mainly highlight proteins, as well as residual lipid structures and individual red blood cells in the viable epidermis, whilst the water signal at 3250 cm\(^{-1}\) was coincident with sebaceous glands in both positive and negative contrast. In vivo imaging of drug penetration of trans-retinal was demonstrated in living mouse skin, visualising with SRS imaging and 3D depth projection that penetration of the topical applied drug occurs along the hair shaft. This pathway was not observed in previous experiments with excised fresh tissue, which was said to further highlight the importance of in vivo imaging, as it can lead to insights into the transport mechanisms of small molecules in living organisms.

Finally, a series of in vivo SRS imaging experiments were performed on human skin of a volunteer’s forearm showing cell layers of the viable epidermis to a depth of 50 μm, and the boundary (via observation of varying nuclear sizes) between the viable epidermis and stratum corneum when tuned into the CH\(_2\) stretching vibration at 2950 cm\(^{-1}\). The penetration-enhancing small molecule DMSO (which was deuterated) was applied to the skin, and its accumulation in the hair shaft and lack of complete penetration into the hair itself was imaged using the C-D stretching vibration at 2125 cm\(^{-1}\). With image acquisition times of 150 ms and 37 ms respectively, these blue-free high-speed in vivo vibrational imaging experiments demonstrated the diagnostic potential of this technology in humans.

Stem cells

Stem cells, stem cell therapy, and stem cell engineering are extremely important and highly topical areas of scientific research with huge potential benefits for the treatment of a wide range of diseases and biomedical applications. The ability to identify the phenotypic purity of live cells is absolutely crucial (and a noninvasive optical method would be highly beneficial), as excessive proliferation of unwanted phenotypes (i.e., uncontrollable differentiation) following transplantation could result in undesirable consequences, such as tissue overgrowth and tumour formation for example. One of the recent studies by the Notinker group applied Raman spectroscopy to determine if it could be used as a label-free, noninvasive method to detect and image intrinsic chemical differences that could be used as molecular markers among highly heterogeneous stem cells populations; specifically, cardiomyocytes (CMs) derived from human embryonic stem cells (hESCs).

Using a custom-built instrument, Raman detection and imaging of molecular markers specific to hESC-derived CMs was carried out, along with retrospective phenotypic identification of all cells via immunofluorescence imaging integrated with the Raman microscope. Multivariate statistical analysis of Raman spectra and cross-validation methods were used to develop a model (from 50 CMs and 40 non-CMs within the same heterogeneous populations) to determine the true accuracy of phenotypic identification of CMs, the sensitivity and specificity parameters, and select discriminatory Raman bands. Raman spectral images corresponding to the Raman bands identified by both the multivariate model and immunostaining of the same cells allowed for the accurate assignment of Raman molecular markers. The conclusions drawn from the results were that spectral differences were more attributable to glycerol and myosin, with glycerol being responsible for discrimination of CMs (with a band assignment at 860 cm\(^{-1}\)), and myosin providing a lesser contribution (with a band assignment at 838 cm\(^{-1}\)). This study demonstrated the potential of Raman spectroscopy for noninvasive phenotypic identification of stem cells, though the authors themselves stated that it was not yet practical for medical applications due to their long spectral acquisition times.

However, a later study by the same group on the same hESC-derived cardiomyocytes reduced spectral acquisition times by a hundredfold, from minutes per cell to 5 s per cell, without the need for ratiometric scanning. When incorporating high-powered commercial lasers, this could be further reduced to cell sorting speeds of approximately 10 cells per s. In addition, a recent study by the same group has successfully applied Raman technology, image and quantify the differentiation status of live neural stem cells in vitro, where this time the spectral differences were said to be related to cytoplasmic RNA. A number of other interesting studies involving Raman analysis of stem cell populations have also appeared in the literature very recently and this area of interest, not surprisingly, continues to grow and flourish.

As well as cellular differentiation, the location of specific drugs inside cells is also of interest as this may allow the elucidation of the drug’s site of action. A recent study addressing this enabled mapping the site of action of the HIV protease inhibitor indinavir and lopinavir in cervical carcinoma cells expressing the E6 oncoprotein from human papilloma virus (HPV). This study demonstrated that indinavir undergoes enhanced nuclear accumulation in E6 expressing cells, indicating this as the site of action for this compound against the HPV. Further interesting studies showing the plethora of Raman imaging applications include a report on stimulated Raman photoacoustic imaging, surface enhanced Raman scattering (SERS) imaging using nanotags in live mice, as a potential multiplexed imaging detection method for multiple biomarkers in living subjects associated with a specific disease, quantitative multiplexed SERS imaging, noninvasive time-course imaging of apoptotic cells, and multivariate image reconstruction methods for Raman hyperspectral datasets. For those specifically interested in coherent nonlinear optical imaging, which includes stimulated Raman scattering, the reader is directed to a recent review by Min et al., others include a review of CARS microscopy, as well as gold nanoparticles and imaging in medicine.

Endoscopy

The clinical potential for in vivo Raman endoscopy has been the subject of research for over a decade, since the first published report of in vivo Raman spectra of human gastrointestinal tissue measured during routine clinical endoscopy in 2000. In the same year work by the Stone group incorporated Raman spectroscopy and clinical endoscopy to discriminate between

This journal is © the Royal Society of Chemistry 2013
Appendix II: Additional publications

normal, dysplastic and cancerous laryngeal tissue — Nick Stone’s group, based in the UK (http://www.exeter.ac.uk) have played a major role in pioneering the field of clinical optical diagnostics with (in addition to other vibrational spectromicroscopies) well received studies using Raman spectroscopy/endoscopy to analyse a range of diseases/disorders such as Barrett’s oesophagus\(^\text{16}\) and bladder and prostate cancer.\(^\text{17}\) One of the recent studies by the Stone group involved the evaluation of the suitability of a custom-built fibre-optic Raman probe for the potential \emph{in vivo} diagnosis of early onset oesophageal neoplasia. Whilst this involved \emph{ex vivo} sampling, the results clearly demonstrated the potential for the rapid and accurate differentiation between benign tissue, Barrett’s oesophagus and both dysplastic and malignant tissue.

The culmination of this study was a custom-built confocal Raman probe constructed by the group. This novel probe had been reported by the group previously\(^\text{16}\), although by the time of the latest study, the probe had undergone some modifications which had improved its performance for spectral acquisition during oesophageal endoscopy. The 90 cm long, 2.7 mm diameter fibre-optic probe was designed to fit into the instrument channel of a standard clinical endoscope and to have direct contact with oesophageal epithelial tissue. The optics had been modified by incorporating a grating index lens at the tip (and the output power regulated to 60 mW), so that a sampling depth of 160 to 200 μm could be ensured. Collecting spectra from this depth meant that signals from deeper tissue structures would not obscure those collected from mucosal abnormalities such as early neoplastic changes,\(^\text{16}\) and that these abnormalities could be quickly classified in timescales suitable to a clinical setting. The performance of the probe \emph{ex vivo} had been evaluated for translational use for \emph{in vivo} sample collection, with lower laser power at the probe tip and short spectral acquisition times said to enable its routine use for oesophageal endoscopy and paving the way for \emph{in vivo} clinical trials.\(^\text{17}\) This group have recently published a review on \emph{in vitro} and \emph{in vivo} Raman spectroscopy as a potential routine tool for the rapid, noninvasive, early diagnosis of lesions and preventing development of cancer in the oesophagus.\(^\text{17}\)

Another group who have made a significant contribution to the field of Raman endoscopy in recent years is headed by Zhiwei Huang and based at the Optical Imaging Laboratory in Singapore (http://www.bioeng.nus.edu.sg/ophthoimaging/huang/). Several of these studies utilised image-guided Raman endoscopy, which the group reported for the first time in 2009.\(^\text{17}\) Whilst image-guided endoscopy is by no means novel,\(^\text{23}\) the integration of image-guided techniques with Raman endoscopy is relatively recent. The first report on this technique involved integrating Raman spectroscopy with trimodal imaging techniques (white-light reflectance, autofluorescence and narrowband) and the development of a novel 1.8 mm Raman probe which filtered out interference from fluorescence as well as interference from silica from within optical fibres. This was demonstrated via the rapid collection of Raman spectra (<1 s) and the corresponding endoscopic images of different locations of the upper gastrointestinal tract of a healthy volunteer in real-time and \emph{in vivo}.\(^\text{23}\)

Since then this group have published several articles using image-guided Raman endoscopy, demonstrating its potential as an \emph{in vivo} real-time detection method for a variety of diagnostic applications. These studies have included, perhaps not surprisingly, the \emph{in vivo} diagnosis of oesophageal cancer using this technique in conjunction with biomolecular modelling.\(^\text{24}\) This involved collecting spectra from 75 oesophageal tissue sites from 27 patients of normal tissue (squamous mucosa) and malignant tumours. The cancerous tissue was said to show distinct Raman signals mainly associated with cell proliferation, lipid reduction, abnormal nuclear activity and neovascularisation. To estimate the biochemical composition of oesophageal tissue, biomolecular modelling was employed using six basis reference spectra from actin, collagen type I, DNA, histones, titin and glycogen. This allowed for the construction of a linear discriminant analysis (LDA) model with a sensitivity of 97% and specificity of 95.2% for \emph{the in vivo} diagnosis of oesophageal cancer.\(^\text{25}\) These results have since been said to be extremely promising, but that image-guided Raman endoscopy is yet to be used to detect dysplasia or the early onset of cancer.\(^\text{25}\)

Nevertheless, the Huang group have previously demonstrated the potential of Raman spectroscopy for the detection of dysplasia, with a ball-Hess fibre-optic probe, in the high wavenumber region (HW) (2800–1700 cm\(^{-1}\)) for the \emph{in vivo} detection of cervical dysplasia (a HW Raman probe was first presented by Gewin Puppels and co-workers for \emph{in vitro} measurements of brain tissue in 2005\(^\text{26}\)). The perceived main advantages of HW Raman were said to be a significant reduction in fluorescence and background signal from optical fibres, more intense Raman signals (compared to the fingerprint region), and the possibility of an unfiltered single fibre Raman probe design for \emph{in vivo} clinical use.\(^\text{28}\) In addition to this study, several more recent reports have continued to investigate this area.\(^\text{27–42}\)

The \emph{in vivo} detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy has also been demonstrated, and significant differences between normal and cancerous gastric tissue were reported.\(^\text{42}\) More recently, a study has been published showing the development of an online automated spectral diagnostics system integrated with image-guided Raman endoscopy for \emph{real time} \emph{in vivo} diagnosis during endoscopic examination.\(^\text{43}\) This system was built on a database of 2655 normal and 283 cancerous gastric tissue spectra acquired from 305 patients, with the system employing a variety of diagnostic algorithms. Other so-called \emph{in vivo} real-time applications recently reported have involved transnasal image-guided Raman endoscopy of the larynx and nasopharynx, which the authors hoped would pave the way for realising early diagnosis and detection of cancers and precancers of the head and neck.\(^\text{26}\) All of these studies show promise and many have similarities in terms of rapid spectral acquisition times (absolutely crucial in \emph{in vivo} studies), and novel developments in optical engineering, but there remains more to accomplish, such as the ability not only to discriminate cancer, but to classify and grade both cancerous and precancerous cells and tissue.\(^\text{26}\)
Appendix II: Additional publications

Critical Review

Recent developments

In terms of noninvasive biomedical Raman other very exciting developments are spatially offset Raman spectroscopy (SORS) as well as transmission Raman, both techniques directly resulting from research involving depth profiling using Raman Kerr-gating methods. SORS was invented and developed in the Central Laser Facility of the Rutherford Appleton Laboratory (http://www.clf.dii.dcu.ie) in 2005 by Pavel Matousek and readers are directed to an excellent overview of this fast moving field by Matousek and Stone. The central difference in SORS is that Raman spectra are collected from different locations, spatially separated (offset) from the point of laser excitation on the sample (Fig. 4). As a consequence of photon diffusion processes within tissue the resultant Raman spectra contain different relative contributions from different depths within the sample and thus allow for the highly accurate chemical analysis of subsurface objects of interest. These spectra acquired from different spatial offsets are processed to reveal pure Raman spectra of subcompartments from separate depth locations within tissue. SORS can be applied in a number of Raman collection and beam delivery geometries including single point collection, ring, and other pattern illumination. In respect to biomedical imaging this approach is said to be particularly useful when configured as inverse SORS, where the sample is illuminated by an adjustable ring-shaped laser beam (generated by a conical (axicon) lens) and the Raman light collected via fibres in the centre of the ring. The radius of this ring is said to define the spatial offset and as it is adjustable, this can be optimised to suit both the scattering properties and dimensions of each sample and covers a wider illumination zone on a sample surface than conventional SORS. SORS is said to be effective at tissue depths in excess of 500 μm, which is well beyond the accessible range of conventional confocal Raman spectroscopy. Applications can be as diverse and wide-ranging as noninvasive detection of pharmaceuticals through packaging, detection of hidden explosives and drug precursors behind opaque plastics and garments, as well as agricultural and food product analysis.

From the outset, Matousek and collaborators immediately realised the diagnostic potential of SORS with its ability for noninvasive subsurface probing of for example bone through skin and dermatology studies. This foresight was quickly realised with the first SORS spectra of excised bone collected in 2006-2007 from animal and human cadavers, and in vivo SORS demonstrated by Matousek during the same year. There have been several studies since including accurate in vivo assessment of bone composition using the carbonate (1070 cm⁻¹) to phosphate (958 cm⁻¹) ratio through the skin of live mice, in vivo evaluation of bone grafts, transcutaneous in vivo monitoring of glucocorticoid induced osteoarthritis and in vivo measurement and evaluation of subcutaneous fat. A significant body of pioneering work forwarding Raman spectroscopy as a tool for breast cancer diagnosis was undertaken in the George R. Harrison Spectroscopy Lab at MIT (http://web.mit.edu/spectroscopy/) under the directorship of Michael S. Feld. This included identifying chemical differences in microcalcifications from benign and malignant breast lesions, demonstrating the real time capabilities of an in vivo Raman system during femoral bypass and breast lumpectomy surgeries, and also involved testing of spectral diagnostic algorithms for breast cancer diagnosis. The results from the Raman research on the chemical composition and identification of the different types of microcalcifications has led directly to others investigating the potential of SORS as a possible clinical adjunct to mammography, for the noninvasive diagnosis of breast cancer. As the changing concentration of carbonate substitution for phosphate ions in the calcium hydroxide lattice in microcalcifications may relate to the process of tumour cell metastasis, and the ability to measure the magnitude of this (as well as soft tissue signals) by Raman, could indicate the potential progression of this cancer.

Research by Stone, Matousek, and collaborators, demonstrated the proof-of-principle of SORS for potential in vivo breast cancer diagnosis in a model system using three calcification standards overlaid with various preparations of animal tissue (i.e. chicken breast tissue, with and without skin). For this work they utilised a continuous wave 827 nm laser with a spatial offset for collection of 3 mm, which enabled the probing and collection of Raman spectra from calcifications through up to 10 mm of tissue. Previous results by this group using Raman-gated Raman techniques had achieved penetration depths of 1 mm in comparison. This SORS study demonstrated the collection of high quality Raman spectra and biochemical information measured through 8.7 mm of tissue, identifying the difference between three calcification standards.

Subsequent experiments by the same group applied transmission Raman spectroscopy in combination with chemometrics in similar model systems collecting Raman calcification signals from depths of 20 mm, this time through porcine tissue, as a breast tissue ‘phantom’. This was said to reach the lower range for clinically relevant breast tissue thicknesses from...
Appendix II: Additional publications

mammographic screening (1.9 cm), with further work said to be required to reach the top of this range (5 cm)."** Whilst these studies utilized the carbonate-to-phosphate band ratio (similar to the SORS bone studies above)**, a more recent study showed that the intense Raman phosphate band at 960 cm⁻¹ broadens and shifts as carbonate concentration (from calcium hydroxyapatite) increases in the calcifications."** This study was said to pave the way toward a new generation of noninvasive breast cancer screening methods based around SORS and transmission Raman spectroscopy."** Some very recent studies by the Maldeby-van Jansen group have investigated SORS for the real-time, intraoperative assessment of breast cancer tumour margins."** This same group then developed and tested a SORS probe with multiple source detector offset limits specific to this type of analysis, and published results acquired from 35 freeze-thaw breast cancer samples in vitro."**

Variants of SORS have also recently been used to augment analysis and include in vivo transcutaneous glucose sensing in rats"** and further demonstrated the accuracy and functionality of this in a later study."** These studies employed what they term surface enhanced SORS (SESORS), first published in 2010,"** and presented at the Federation of Analytical Chemistry and Spectroscopy Societies (FACSS) meeting in the USA during the same year. SESORS aims to use SERS techniques using nanoparticles and nanospheres with the subsurface probing of SORS, with a further study exploring this technique's potential for Raman imaging."**

On the subject of SERS, which has not been mentioned in any great detail here, many studies have appeared in the literature during the last few years related to disease detection and we would like to highlight just a modest selection of these. They of course include the multiplexed imaging of SERS nanoparticles in vivo already mentioned above,"** as well as other multiplexing studies, such as multiplex single nucleotide polymorphism (SNPs) genotyping coupled to SERS,"** multiplexed in vivo cancer detection using SERS NIR nanotags, demonstrating the excellent sensitivity, stability and tumour specificity of three bioconjugated nanotags,"** recent reviews on the area of SERS multiplexed detection for disease diagnostics,"** as well as SERS cancer detection and imaging and the potential of SERS agents for targeted drug delivery and photothermal therapy."**

Several SERS studies involving immunoasays, including cancer detection,"** detection of a potential pancreatic cancer marker in serum,"** and on-chip immunoasays using hollow gold nanospheres."** Fluorescent SERS SERS gold-coated nanorod probes for in vivo imaging of lymph node mapping and tumor targeting in mice,"** high sensitivity in vivo detection of inflammation using gold nanoclusters conjugated to monoclonal antibodies"** and the use of functionalised nanoparticles and SERS for the detection of DNA relating to disease."** With the recent interest concerning the potential resurgence in microbial disease, as well as bioterrorism, an article demonstrating a portable quantitative SERS system for detection of Bacillus spores at levels significantly lower than those previously reported for SERS has been published."** Fig. 5 shows SERS spectra of a biomarker from the spores of Bacillus, a species of which Bacillus anthracis is the cause of the acute, and mostly lethal, disease anthrax. The detection limit of this dioplic acid biomarker is estimated at 3 ppb (30 nM).**

In addition to those primary articles and reviews already mentioned above, other recently published reviews and books on an array of Raman-based topics which may be of interest include those concerning in vivo and in vitro analysis."** Raman scattering in pathology,"** optical tumour margin identification in the laboratory,"** data-classification algorithms for spectral analysis,"** recent advances in gold nanoparticle based assays for detecting and identifying microbes,"** and emerging Raman applications and techniques in biomedical and pharmaceutical fields."**

Concluding remarks

We hope that whilst only an introductory overview to some of the more recent work in this field, the range and scope of the studies shown here elegantly demonstrate the development and exploitation of Raman spectroscopy as a medical diagnostic tool. We believe this truly is an exciting field, where those at the forefront are propelling it forwards and are simply not content to remain within the confines of current knowledge or commercially available optics. These studies highlight the necessary interdisciplinary nature of this field, with inspirational contributions from a range of analytical scientists and technologists including clinicians, biologists, chemists, optical engineers, as well as statisticians and chemometrics. These multidisciplinary inputs and fresh approaches are invigorating this field, unlocking the doors to new insights,"** adding to knowledge, and opening up new dimensions and avenues of study, as well as potential clinical applications, for biomedical Raman spectroscopy.
Appendix II: Additional publications

Acknowledgements

DIE would like to thank Analyst for their kind invitation to write this article and the patience of all in the editorial team during its construction, as well as Prof. Paul O’Brien FRSE, Dr Steve O’Hagan and Hwebee Muhammad Ali for assistance with graphics and software. RG, LA and DPC are very grateful to the UK BBSRC, EPSRC, Avacta group plc, and industrial members of the Boprocessing Research Industry Club (BRIC) for funding this work.

References

2 I. M. Newton, Phil. Trans., 1671, 6, 3075-3087.
6 J. Strutt, Philos. Mag., 1889, 47, 375-394.
Appendix II: Additional publications
Appendix II: Additional publications


134. Emerging Raman Applications and Techniques in Biomedical and Pharmaceutical Fields (Biological and Medical Physics, *Analyst*, 2013, 138, 3871-3884 | 3883)
Appendix II: Additional publications