Screening of Human Serum/Plasma using Vibrational Spectroscopy for Early Disease Diagnostics and Therapeutic Drug Monitoring

Drishya Rajan Parachalil [Thesis]
Technological University Dublin, drishyarajan.parachalil@mytudublin.ie

Follow this and additional works at: https://arrow.tudublin.ie/sciendoc
Part of the Biochemistry Commons, and the Biophysics Commons

Recommended Citation

This Theses, Ph.D is brought to you for free and open access by the Science at ARROW@TU Dublin. It has been accepted for inclusion in Doctoral by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License
Screening of Human Serum/Plasma using Vibrational Spectroscopy for Early Disease Diagnostics and Therapeutic Drug Monitoring

Author: Drishya Rajan Parachalil

Supervisors: Prof. Hugh J. Byrne
Dr Jennifer McIntyre

A thesis submitted for the degree of Doctor of Philosophy

School of Physics & Clinical & Optometric Sciences,
and
FOCAS Research Institute
Abstract

Analysis of analytes present in the blood stream can potentially deliver crucial information on patient health and indicate the presence of numerous pathologies. Existing clinical techniques for this analysis can, however, be costly and time-consuming. The potential of Raman spectroscopic analysis of human plasma and/or serum for diagnostic purposes has been widely investigated and, increasingly, its feasibility for clinical translation has been explored. However, as the concentration of many analytes in plasma/serum is relatively low, to date such analysis has commonly been performed on air-dried drops deposited on substrates, leading to inhomogeneity and inconsistencies. This study explores the potential of Raman spectroscopy, coupled with fractionation and concentration techniques, as well as multivariate regression analysis, to quantitatively monitor diagnostically relevant changes in high and low molecular weight proteins as well as therapeutic drugs, in liquid plasma/serum. Having optimised the protocols for pure aqueous solutions and spiked serum samples, measurement protocols to detect the imbalances in plasma/serum analytes (fibrinogen, albumin, γ globulins, total protein content, glucose and urea), as an indicator of various diseases, and therapeutic monitoring of drugs (busulfan and methotrexate), using Raman spectroscopy were optimised in liquid serum, such that strategic clinical applications for early stage disease diagnostics and therapeutic drug monitoring can be evaluated. Furthermore, an adapted Extended Multiplicative Signal Correction algorithm was applied to raw spectra to remove background signal and spectral interferents. Using a validated partial least squares regression method, prediction models were built for the analytes, with accuracies which are
comparable with those reported for the conventional methods, without any additional sample preparation steps. This methodology was extended to determine the Limit of Detection (LOD) and Limit of Quantification (LOQ) for therapeutic drug monitoring in human serum, using the examples of Busulfan, a cell cycle non-specific alkylating antineoplastic agent, and, Methotrexate, a chemotherapeutic agent. This study demonstrates the options and alternatives that are available to make Raman spectroscopy suitable for the human bodily fluid analysis in the liquid form, leading to a better accuracy and repeatability and thus a better sensitivity.
Declaration

I certify that this thesis which I now submit for examination for the award Doctor of Philosophy (Ph.D.), is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. This thesis was prepared according to the regulations for postgraduate study by research of the Technological University Dublin and has not been submitted in whole or in part for another award in any other third level institution. The work reported on in this thesis conforms to the principles and requirements of the TU Dublin’s guidelines for ethics in research.

TU Dublin has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Signature ___________________ Date _____/_____/_____

Drishya Rajan Parachalil
Acknowledgements

First, I would like to thank my PhD mentor, Prof Hugh Byrne, for his support, incredible guidance, patience and encouragement during these past four years. His deep insights, all the useful discussions and brainstorming sessions, especially during the difficult conceptual development stage helped me at various stages of my research and kept me motivated throughout. I would also like to thank Dr. Jennifer McIntyre for being there for me always as a rock of support throughout my PhD. I will never forget her encouragement and help in the lab. Jen helped me adapt to FOCAS and overcome all the problems I faced during the early stages of my project. I could not have done it without her. Heartfelt thanks goes to Dr. Franck Bonnier for providing his scientific advice and knowledge, suggestions and insightful discussions about the research and for the The Irish-French PHC Ulysses 2018 Collaboration, which was a turning point in this study. Also, I would like to thank FOCAS Research institute for the wonderful work environment and facilities and DIT Fiosraigh scholarship for providing the funding to carry out this study.

PhD students often talk about loneliness during the course of their study but this is something that I never experienced at FOCAS. A special thanks to the very supportive, cheerful and fun loving FOCAS family, in particular, Neha, Damien, Isha, Fionn, Caroline, Dan, Uli and Naomi. Special thanks are also due to my best friend Jancy Jolly, for always being there for me. Words cannot express the feelings I have for my parents and sister for their unconditional love and support. I would not be here if it were not for you. Finally, I would like to acknowledge Alan, Christine and Gerry, for being a constant source of strength and inspiration. It is amazing to have family close by so far away from home.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATR</td>
<td>Attenuated total reflectance</td>
</tr>
<tr>
<td>ALAT</td>
<td>Alanine amino transferase</td>
</tr>
<tr>
<td>ASAT</td>
<td>Aspartate amino transferase</td>
</tr>
<tr>
<td>ALP</td>
<td>Alkaline phosphatase</td>
</tr>
<tr>
<td>Bu</td>
<td>Busulfan</td>
</tr>
<tr>
<td>CCD</td>
<td>Charged coupled detection</td>
</tr>
<tr>
<td>CRP</td>
<td>C reactive protein</td>
</tr>
<tr>
<td>CTD</td>
<td>Charge transfer devices</td>
</tr>
<tr>
<td>ECF</td>
<td>Extracellular fluid</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme linked immunosorbent assay</td>
</tr>
<tr>
<td>EMSC</td>
<td>Extended multiplicative signal correction</td>
</tr>
<tr>
<td>FSD</td>
<td>Fourier self deconvolution</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier-transform infrared spectroscopy</td>
</tr>
<tr>
<td>GC</td>
<td>Gas chromatography</td>
</tr>
<tr>
<td>GGT</td>
<td>Gamma glutanyl transferase</td>
</tr>
<tr>
<td>HCC</td>
<td>Hepatocellular carcinoma</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immuno deficiency virus</td>
</tr>
<tr>
<td>HDL</td>
<td>High-density lipoprotein</td>
</tr>
<tr>
<td>HMWF</td>
<td>High molecular weight fraction</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>HSA</td>
<td>Human serum albumin</td>
</tr>
<tr>
<td>ICF</td>
<td>Intercellular fluid</td>
</tr>
<tr>
<td>IF</td>
<td>Interstitial fluid</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IndGaAS</td>
<td>Indium-Gallium-Arsenic</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>IRF</td>
<td>Internal reflection element</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>LDL</td>
<td>Low-density lipoprotein</td>
</tr>
<tr>
<td>LMWF</td>
<td>Low molecular weight fraction</td>
</tr>
<tr>
<td>LOOCV</td>
<td>Leave one out cross validation</td>
</tr>
<tr>
<td>LOD</td>
<td>Limit of detection</td>
</tr>
<tr>
<td>LOPOCV</td>
<td>Leave one patient out cross validation</td>
</tr>
<tr>
<td>LOQ</td>
<td>Limit of quantitation</td>
</tr>
<tr>
<td>MI</td>
<td>Myocardial infarction</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MTX</td>
<td>Methotrexate</td>
</tr>
<tr>
<td>PCA</td>
<td>Principle component analysis</td>
</tr>
<tr>
<td>PLSR</td>
<td>Partial least squares regression analysis</td>
</tr>
<tr>
<td>RID</td>
<td>Radial immune diffusion</td>
</tr>
<tr>
<td>SERS</td>
<td>Surface enhanced Raman spectroscopy</td>
</tr>
<tr>
<td>SIRS</td>
<td>Systemic inflammatory response syndrome</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>SPEP</td>
<td>Serum protein electrophoresis</td>
</tr>
<tr>
<td>SVM</td>
<td>Supporting vector machine</td>
</tr>
<tr>
<td>TBG</td>
<td>Thyroxine binding globulin</td>
</tr>
<tr>
<td>TDM</td>
<td>Therapeutic drug monitoring</td>
</tr>
<tr>
<td>VDLP</td>
<td>Very low density lipoprotein</td>
</tr>
</tbody>
</table>
# Table of Contents

Abstract ........................................................................................................................................... i
Declaration ........................................................................................................................................ iii
Acknowledgements ......................................................................................................................... iv
List of abbreviations ....................................................................................................................... v
Table of Contents ........................................................................................................................... vii
List of Tables ..................................................................................................................................... xii
List of Figures .................................................................................................................................... xiv

**Chapter 1** .................................................................................................................................... 1

Introduction ........................................................................................................................................ 1
  1.1 Raman and Infrared Spectroscopy in the analysis of bodily fluids ........................................... 2
  1.2 Research question and hypothesis ............................................................................................. 6
  1.3 Thesis summary .......................................................................................................................... 6
  1.4 References ................................................................................................................................ 10

**Chapter 2** .................................................................................................................................... 15

Vibrational Spectroscopic Analysis and Quantification of Proteins in Human Blood Plasma and Serum .......................................................................................................................... 15
  2.1 Abstract ...................................................................................................................................... 15
  2.2 Introduction ................................................................................................................................. 15
  2.3 Analysis of Biofluids ................................................................................................................... 17
    2.3.1 Blood Sample: preparation of plasma versus serum ............................................................ 19
    2.3.2 Composition of Plasma and Serum ...................................................................................... 20
    2.3.3 Non-Protein constituents ...................................................................................................... 21
    2.3.4 Proteins ............................................................................................................................... 22
      2.3.4.1 Fibrinogen ..................................................................................................................... 24
      2.3.4.2 Albumin ......................................................................................................................... 25
      2.3.4.3 Globulins ....................................................................................................................... 26
      2.3.4.4 Immunoglobulins ......................................................................................................... 27
    2.4 Pathology of plasma proteins .................................................................................................. 29
      2.4.1 Abundant proteins ............................................................................................................. 29
      2.4.2 Low abundance proteins ................................................................................................. 32
2.4.2.1 Cytochrome C ................................................................. 34
2.4 Vibrational spectroscopic analysis of bodily fluids .......................... 35
  2.4.1 Vibrational Spectroscopy .................................................. 35
2.5 Experimental approaches ........................................................ 39
  2.5.1 Fourier-Transform Infrared Spectroscopy ......................... 39
  2.5.2 Instrumentation for Raman spectroscopy ..................... 44
2.6 Biospectroscopy ..................................................................... 46
2.7 Vibrational Spectroscopy of Protein ........................................... 50
  2.7.1 Spectroscopic signature of serum .................................. 58
  2.7.2 Quantitative analysis ....................................................... 65
2.8 Clinical Translation .................................................................. 69
2.9 References .............................................................................. 72

Chapter 3 .................................................................................. 89
Materials and Methods ................................................................. 89
  3.1 Introduction ........................................................................ 89
  3.2 Raman spectroscopy ............................................................. 89
  3.3 Sample substrates ............................................................... 90
  3.4 ATR-FTIR ......................................................................... 90
  3.5 Preparation of stock protein ............................................. 90
  3.6 Impact of centrifugal filtration .......................................... 91
  3.7 Observing the optimum volume to record spectra ............. 92
  3.8 Data Analysis of the recorded spectra ................................ 93
     3.8.1 Spectral Pre-processing ............................................ 93
     3.8.2 Partial Least Squares Regression .............................. 94
  3.9 Standardisation of measurement protocol .......................... 96
     3.9.1 Substrate selection ..................................................... 97
     3.9.2. Wavelength selection ........................................... 100
     3.9.3 Geometry selection ............................................... 102
     3.9.4. Protein measurement in the upright and inverted geometry 104
  3.10 ATR-FTIR analysis of varying concentrations of fibrinogen – A comparative study ..................................................... 109
  3.11 Summary ........................................................................... 111
  3.12 References ....................................................................... 112
Chapter 4 .......................................................... 114
Raman spectroscopic analysis of High Molecular Weight Proteins in solution—
considerations for sample analysis and data pre-processing .......................... 114
  4.1 Abstract ........................................................................................................ 114
  4.2 Introduction .................................................................................................. 115
  4.3 Materials and Methods ............................................................................... 118
    4.3.1 Preparation of stock protein and protein mixture ............................... 118
    4.3.2 Ultrasonication ................................................................................. 119
    4.3.3 Ion exchange chromatography ....................................................... 119
    4.3.4 Raman spectroscopy ....................................................................... 120
    4.3.5 Sample substrates ......................................................................... 120
    4.3.6 Spectral preprocessing .................................................................. 120
    4.3.7 Partial Least Squares Regression .................................................. 121
  4.4 Results ....................................................................................................... 122
    4.4.1 Standardisation of measurement protocol ....................................... 122
    4.4.2 Monitoring the concentration dependence of proteins in aqueous solution 124
  4.5 Discussion ................................................................................................ 137
  4.6 Conclusions .............................................................................................. 141
  4.7 References ................................................................................................. 143
  4.8 Electronic Supplementary information: ............................................... 149

Chapter 5 .......................................................... 155
Analysis of bodily fluids using Vibrational Spectroscopy: A direct comparison of Raman scattering and Infrared absorption techniques for the case of glucose in blood serum ...... 155
  5.1 Abstract ................................................................................................... 156
  5.2 Introduction .............................................................................................. 157
  5.3 Materials and Methods .......................................................................... 161
    5.3.1 Preparation of varying concentration of glucose in distilled water model ...... 161
    5.3.2 Preparation of glucose spiked in serum model ................................ 162
    5.3.3 Glucose levels in patient serum samples ...................................... 162
    5.3.4 Data collection using Raman spectrophotometer ....................... 165
    5.3.5 Data pre-processing and analysis .............................................. 165
    5.3.6 Partial Least Squares Regression ............................................ 167
  5.4 Results ................................................................................................... 168
    5.4.1 Monitoring the concentration dependence of glucose in distilled water ...... 169
    5.4.2 Monitoring the glucose concentration in spiked serum ............ 172
5.4.3 Monitoring the glucose concentration in patient samples ........................................ 174
5.5 Discussion .................................................................................................................. 178
5.6 Conclusion .................................................................................................................. 182
5.7 References .................................................................................................................. 183
5.8 Supplementary information ....................................................................................... 189
5.8.1 References ............................................................................................................... 189

Chapter 6 .................................................................................................................................. 190
Raman spectroscopic screening of High and Low molecular weight fractions of human serum .......................................................................................................................... 190
6.1 Abstract ......................................................................................................................... 190
6.2 Introduction ................................................................................................................... 191
6.3 Materials and Method .................................................................................................. 196
  6.3.1 Sample Preparation ................................................................................................. 196
  6.3.2 Data collection using Raman spectrometer ............................................................ 200
  6.3.3 Data pre-processing and analysis ........................................................................ 200
  6.3.4 Partial Least Squares Regression ......................................................................... 202
6.4. Results and discussion ............................................................................................... 204
  6.4.1 Quantification of total protein concentration and γ globulins in whole serum ... 206
  6.4.2 Quantification of albumin from the HMWF concentrate ...................................... 211
  6.4.3 Quantification of glucose and urea from the LMWF filtrate ................................. 213
6.5 Discussion ....................................................................................................................... 219
6.6 Conclusion ...................................................................................................................... 223
6.7 References ..................................................................................................................... 225
6.8 Electronic Supplementary information: ....................................................................... 236

Chapter 7 .................................................................................................................................. 243
Raman spectroscopy as a potential tool for label free therapeutic drug monitoring in human serum: the case of Busulfan and Methotrexate ................................................................................. 244
7.1 Abstract ......................................................................................................................... 245
7.2 Introduction ................................................................................................................... 245
7.3 Materials and Methods ............................................................................................... 248
  7.3.1 Materials ................................................................................................................ 248
  7.3.2 Raman spectroscopy ............................................................................................. 250
  7.3.3 Data pre-processing and analysis ........................................................................ 250
  7.3.4 Partial Least Squares Regression ......................................................................... 251
Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: Recent advances

8.1 Abstract
8.2 Introduction
8.3 Raman vs Infrared absorption spectroscopy
8.4 Measurement of Plasma vs Serum
8.5 Serum Fractionation
8.6 Data preprocessing
8.7 Data Analysis
8.8 Clinical Translation
8.9 Conclusion
8.10 References

Chapter 9
Conclusion

9.1 References

Appendix 1: Publications
Appendix 2: Conferences, Modules and Collaboration
List of Tables

Table 2.1: Relative content of abundant proteins in plasma and serum  23
Table 2.2: Globulin fractions with example of encompassed proteins  26
Table 2.3: Details of immunoglobulin subtypes  27
Table 2.4: Serum Protein Fractions and Conditions Associated with an Increased or Decreased Level  32
Table 2.5: Tentative peak assignments for FTIR spectral data, (i)-(vii) corresponds to Figure 9A (118-121)  49
Table 2.6: Tentative peak assignments for Raman spectral data, (i)-(viii) corresponds to Figure 9B (122-123).  50
Table 2.7: Assignment of Amide I band positions to secondary structure in H₂O for both IR and Raman (128,130,136,137,140)  55
Table 2.8: Summary of main amino acid side chain absorptions found in IR spectra between the 1400-1800cm⁻¹ region (128,136,144)  57
Table 3.1: Peak assignments for the Raman spectra of plasma proteins (7–10)  106
Table 5.1: List of measured glucose levels in patient samples. Glucose blood levels are quoted in terms of the SI unit of mmol L⁻¹ as well as mg dL⁻¹, commonly used in serology literature and in the study of Bonnier et al.(29)  164
Table 5.2: Comparison of the results of ATR-FTIR and Raman spectroscopic analysis of human serum spiked with varying concentrations of glucose  174
Table 5.3: Comparison of the results of ATR-FTIR (29) and Raman spectroscopic analysis of patient sample set for monitoring the glucose levels. FTIR results are normalised.

Table 6.1: Measured analyte levels in patient samples

Table 6.2: Summary of the results obtained from the patient samples

Table 8.1: LOD of glucose [44], busulfan [101], methotrexate [101], cholesterol, urea [69] and vitamin B12 calculated from the PLSR prediction plot of these analytes, compared to the maximum Raman intensity of maximum peak per unit acquisition time, per unit concentration
List of Figures

Figure 2.1: Obtaining (A) Plasma and (B) Serum from blood samples 20

Figure 2.2: Molecular structural figures as examples of (A) a sugar, (B) a lipid and an (C) amino acid 21

Figure 2.3: Structure of immunoglobulins 28

Figure 2.4: Disease patterns in serum protein electrophoresis (SPEP) 30

Figure 2.5: Common vibrational modes of chemical bonds 36

Figure 2.6: Depiction of light scattering by vibrating polarisation 38

Figure 2.7: The Michelson interferometer found in FTIR instruments. Red: incident beam, blue: reflected and purple: combined. Adapted from ref.(85) 40

Figure 2.8: Typical Instrumentation for Raman microspectroscopy (CCD; charge coupled detector) 46

Figure 2.9: Typical IR (A) and Raman (B) spectrum of human blood serum 48

Figure 2.10: Molecular vibrations of the amide group - Orange: Hydrogen, Red: Nitrogen, Purple: Carbon, Blue: Oxygen. Adapted from ref (132) 52

Figure 2.11: Raman spectra of (A) Albumin, (B) Fibrinogen and (C) Cytochrome C 53

Figure 2.12: Curve fitting of the Amide I band in serum, for IR (A) and Raman (B) Spectra 56

Figure 2.13: The inverted geometry used to analyse the serum focused by immersion objective 64

Figure 3.1: (A) Raman spectrum of water in the 96 polystyrene well plate and (B) an empty polystyrene well plate recorded using the 532nm laser line in the upright geometry with the x10 objective. This study indicates that 96 well polystyrene plates cannot be used as the substrate, as polystyrene peaks are observed superimposed on the water peaks 98

Figure 3.2: Raman spectrum of water recorded in the quartz well plate in the upright geometry using the 10x objective and the 532nm laser line. A strong water peak at 1650cm⁻¹ can be seen, with minimal background noise 99

Figure 3.3: (A) Raman spectra of water recorded in the Lab-Tek plate and (B) empty Lab-Tek plate recorded using the 532nm laser line in the upright geometry with the x 10 objective.
No interference from glass peaks at 850cm\(^{-1}\) can be seen in A. Therefore, the Lab-Tek plate can be used as an ideal substrate for Raman measurements.

**Figure 3.4:** Raman spectrum of water recorded using the 785nm laser line in the Lab-Tek plate in the upright geometry with the x 10 objective. Strong fluorescence background is observed due to absorption by glass impurities at 785nm.

**Figure 3.5:** (A) Raman spectrum of the distilled water recorded in the upright geometry with x10 objective (B) enhanced Raman spectrum of the distilled with considerably lower background and improved S/N, recorded in the inverted geometry focused with x60 water immersion objective.

**Figure 3.6:** The inverted geometry used to analyse the serum focused by immersion objective. (2)

**Figure 3.7:** Raman spectra of the stock solutions of albumin, fibrinogen, cytochrome c and vitamin B12 recorded in the finger print region in the inverted geometry focused by water immersion x60 objective. Well-defined Raman peaks with minimum background were obtained.

**Figure 3.8:** (A) Raman spectra of distilled water recorded from 1 µL to 1mL volume, (B) Raman spectra of fibrinogen solution recorded from 1 µL to 1mL volume, (C) Intensity versus volume graphs of distilled water and (D) Intensity versus volume graphs of fibrinogen stock. This shows a strong Raman signal can be recorded from a volume as small as 1µL in the inverted geometry using Lab-Tek Plate as substrate.

**Figure 3.9:** (A) ATR-FTIR spectra of varying concentration (1mg/mL to 50mg/mL) of fibrinogen in distilled water collected after depositing 2μL of the sample on the ATR crystal. (B) Intensity vs concentration plot of the Amide 2 peak shows a steady increase in the intensity until 30mg/mL and remains constant for 40 and 50mg/mL.

**Figure 4.1:** Raman spectra of the stock solutions of albumin, fibrinogen, cytochrome c and vitamin B12 recorded using the 532nm laser in the finger print region in the inverted geometry focused by water immersion x60 objective. Well defined Raman peaks with minimum background were obtained.

**Figure 4.2:** A: Raw Raman spectra of varying concentrations of albumin (5mg/mL – 50mg/mL) in distilled water, recorded using the 532nm laser B: Percent variance explained by the components, C: plot of PLSR coefficient with Albumin features, D: Linear predictive model built from the PLSR analysis.

**Figure 4.3:** A: Rubberband corrected Raman spectra of varying concentrations of Albumin (5mg/mL – 50mg/mL) in distilled water, B: % variance explained by the latent variables, C: plot of PLSR coefficient with Albumin features, D: Linear predictive model built from the PLSR analysis.
Figure 4.4: A: EMSC corrected of varying concentrations of albumin in simulated plasma, and B: Percent variance explained by the latent variables, C: PLSR coefficient showing albumin features, and D: Linear prediction model defined from the dataset

Figure 4.5: A: Raman spectra of varying concentration of sonicated fibrinogen background corrected using EMSC algorithm B: Percent variance explained by the latent variables C: PLSR coefficient plotted from the sonicated fibrinogen data set shows strong fibrinogen features, D: Linear predictive model built from the PLSR analysis showing correlation between concentration and peak intensity

Figure 4.6: Smoothed spectra of varying concentration of fibrinogen in simulated plasma (0.5mg/mL to 5mg/mL). The arrow indicates the order of increasing concentration

Figure 4.7: A: EMSC corrected data of varying concentrations of fibrinogen separated by ion exchange chromatography, and B: Percent variance explained by the latent variables, C: PLSR coefficient showing fibrinogen features and D: Linear prediction model defined from the dataset

Figure 4.S1: A: Raw Raman spectra of varying concentrations of albumin in simulated plasma (5mg/mL – 50mg/mL). The arrow indicates the order of increase in concentration, B: Percent variance explained by the latent variables, C: PLSR component showing the inverse albumin features and D: Linear predictive model built from the PLSR analysis

Figure 4.S2: A: Raman spectra of varying concentrations of fibrinogen in distilled water corrected using the “rubberband” method (0.5mg/mL – 5mg/mL), B: Percent variance explained by the latent variables, C: PLSR coefficient plotted from the data set shows negative peaks and no fibrinogen features, D: Linear predictive model built from the PLSR analysis

Figure 4.S3: A: Spectra corrected by EMSC algorithm of varying concentration of fibrinogen (0.5mg/mL – 5mg/mL), B: Percent variance explained by the latent variables, C: PLSR coefficient showing strong fibrinogen features, D: Linear predictive model built from the PLSR analysis

Figure 4.S4: Raman spectra of fibrinogen stock recorded before (Black) and after (Red) sonication. Sonication helped to increase the overall intensity of the Raman signal by increasing the solubility of the protein

Figure 4.S5: A: EMSC corrected data of varying concentrations of fibrinogen (0.5mg/mL to 5mg/mL), B: percent variance explained by the latent variables, C: PLSR coefficient showing the inverse peaks of albumin (1089cm⁻¹ and 1102cm⁻¹), and D: The predictive model built from the dataset

Figure 4.S6: Raman spectrum of the concentrate obtained after centrifugal filtration of simulated plasma using 100kDa filters showing albumin features (899cm⁻¹ and 1102cm⁻¹)
Figure 4.87: A: EMSC corrected spectra of varying concentrations of fibrinogen separated by ion exchange chromatography (0.5mg/mL to 5mg/mL), in the absence of sonication B: percent variance explained by the latent variables, C: PLSR coefficient showing inverse fibrinogen peak, and D: The predictive model built from the dataset

Figure 5.1: Raman spectrum of an aqueous glucose solution, concentration 450g/L. Example signature peaks at 450cm$^{-1}$, 911cm$^{-1}$, 1125cm$^{-1}$, 1340cm$^{-1}$ and 1460cm$^{-1}$ are labelled in the figure

Figure 5.2: (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of varying concentrations of glucose (5 x100mg/dL, 5 x 450mg/dL and 5 x 1000mg/dL, offset for clarity), in distilled water and signature peaks of glucose are highlighted with asterisks, (B): Evolution of the RMSECV on the validation model, (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R$^2$ values were calculated as 10.93mg/dL and 0.9705 respectively

Figure 5.3: (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of glucose spiked in serum (spiked concentrations 5 x 0mg/dL, 5 x 120mg/dL and 5 x 220mg/dL, offset for clarity) and the signature peaks of glucose are highlighted by asterisks, (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R$^2$ values were calculated as 1.66mg/dL and 0.9914 respectively

Figure 5.4: (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of patient samples (5 x 52.25mg/dL, 5 x 75.67mg/dL, 5 x 93.69mg/dL, 5 x 210.81mg/dL and 5 x 434.35mg, offset for clarity) and the signature peaks are marked by asterisks, (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R$^2$ values were calculated as 1.84mg/dL and 0.84 respectively

Figure 5.5: PLSR validation of patient samples on Clarke’s error grid. The RMSECV was found to be 1.84 mg/dL and R$^2$ value was calculated as 0.84

Figure 5.51: (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of serum samples, (B): Evolution of the RMSECV on the validation model, (C): PLSR coefficient shows a negative peak ~1000cm$^{-1}$, (D): Predictive model built from the PLSR analysis

Figure 6.1: Schematic overview of steps in fractionation of patient serum samples to separate γ globulin, albumin, and urea/glucose
Figure 6.2: Raman spectra of patient serum collected using Raman spectroscopy. (A) whole serum, (B) concentrate from 50 kDa filtration and (C) filtrate from 10 kDa filtration. Spectra have been off set for clarity.

Figure 6.3: Spectra of γ globulins (red- ~38% of serum) and albumin (blue - ~50% of serum) showing similar spectral features. Identifying signature peaks of γ globulins at 1240 cm\(^{-1}\) and 1553 cm\(^{-1}\) and of albumin at 940 cm\(^{-1}\) are highlighted with asterisks.

Figure 6.4: (A) EMSC corrected Raman spectra of total protein content from patient serum samples (4200 mg/dL, 5800 mg/dL, 6400 mg/dL and 7900 mg/dL). The spectra have been offset for clarity (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient for regression against total serum protein, (D): Linear predictive model for total serum protein built from the PLSR analysis. The RMSECV, R\(^2\) and standard deviation values were calculated as 114.7 mg/dL, 0.82 and 5.69 mg/dL.

Figure 6.5: (A) EMSC corrected Raman spectra of γ globulin of patient serum samples (329 mg/dL, 690 mg/dL, 836 mg/dL and 1404 mg/dL), the spectra has been offset for clarity. (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient for regression against γ globulin concentration shows the signature peaks of γ globulin (highlighted by asterisk), (D): Linear predictive model for γ globulin concentration built from the PLSR analysis. The RMSECV, R\(^2\) and standard deviation values were calculated as 126 mg/dL, 0.88 and 4.62 mg/dL.

Figure 6.6: (A): EMSC corrected Raman spectra of concentrate obtained after filtration with 50 kDa filters of patient sample (2710 mg/dL, 3580 mg/dL, 4140 mg/dL and 4780 mg/dL), the spectra have been offset for clarity (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient from regressing against albumin concentrations, (D): Linear predictive model for albumin built from the PLSR analysis. The RMSECV, R\(^2\) and standard deviation values were calculated as 90.097 mg/dL, 0.9072 and 1.1692 mg/dL.

Figure 6.7: (A) Reference spectrum of urea (1000 mg/dL) and (B) Reference spectrum of glucose (45000 mg/dL). The reduced spectral regions selected for PLSR analysis are indicated by dotted lines in each spectrum.

Figure 6.8: (A): EMSC corrected Raman spectra of filtrate obtained after filtration with 10 kDa filters of patient samples (2.5 mg/dL, 10.64 mg/dL, 19.04 mg/dL and 78.99 mg/dL). The spectra have been offset for clarity (B): Evolution of RMSECV of the data set regressed against urea concentrations (C): plot of PLSR coefficient with strong features of urea, (D): Linear predictive model built from the PLSR analysis. The RMSECV, R\(^2\), and standard deviation values were calculated as 1.736 mg/dL, 0.9232 and 2.89 mg/dL.

Figure 6.9: (A): EMSC corrected Raman spectra of filtrate obtained after filtration with 10kDa filters of patient samples (2.5 mg/dL, 10.64 mg/dL, 19.04 mg/dL and 78.99 mg/dL). Spectra have been offset for clarity, (B): Evolution of RMSECV of the data set regressed against urea concentration (C): plot of PLSR coefficient with strong features of urea, (D): Linear predictive model for urea concentration built from the PLSR analysis. The RMSECV,
R² and standard deviation values were calculated as 2.52 mg/dL, 0.9722 and 1.1418 mg/dL.

Figure 6.81: Plot of the concentrations of albumin and immunoglobulin for each patient

Figure 6.82: Plot of the concentrations of urea and glucose for each patient

Figure 6.83: Raman spectrum of β-carotene used for EMSC correction of human serum from patient samples

Figure 6.84: Raman spectrum of human serum used for EMSC correction of total protein and globulin from patient samples

Figure 6.85: (A) EMSC corrected Raman spectra of γ globulin of patient serum samples from 800 cm⁻¹ to 980 cm⁻¹ (329 mg/dL, 690 mg/dL, 836 mg/dL and 1404 mg/dL), the spectra has been offset for clarity. (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient for regression against γ globulin concentration shows the peaks of γ globulin, (D): Linear predictive model for γ globulin concentration built from the PLSR analysis

Figure 6.86: (A): Rubberband corrected Raman spectra of varying concentrations of Albumin from 5 mg/mL to 50 mg/mL (500 mg/dL to 5000 mg/dL) in distilled water, (B): Evolution of RMSECV on the validation model, (C): plot of PLSR coefficient with Albumin features, (D): Linear predictive model built from the PLSR analysis. The RMSECV is calculated as 1.58 mg/mL (158 mg/dL)

Figure 6.87: (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of varying concentrations of glucose (5 x 100 mg/dL, 5 x 450 mg/dL and 5 x 1000 mg/dL, spectra offset for clarity), in distilled water and signature peaks of glucose are highlighted with asterisks, (B): Evolution of the RMSECV on the validation model, (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R² values were calculated as 10.93 mg/dL and 0.9705 respectively

Figure 6.88: (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of glucose spiked in serum (spiked concentrations 5 x 0 mg/dL, 5 x 120 mg/dL and 5 x 220 mg/dL, offset for clarity) and the signature peaks of glucose are highlighted by asterisks, (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R² values were calculated as 1.66 mg/dL and 0.9914

Figure 6.89: (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of patient samples (5 x 52.25 mg/dL, 5 x 75.67 mg/dL, 5 x 93.69 mg/dL, 5 x 210.81 mg/dL and 5 x 434.35 mg, offset for clarity) and the signature peaks...
are marked by asterisks, (B): Evolution of RMSECV of the data set, (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation The RMSECV and $R^2$ values were calculated as 1.84mg/dL and 0.84 respectively.

**Figure 6.S10**: (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of urea spiked in water (1mg/dL to 1000mg/dL), (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient with urea features, (D): Linear predictive model built from the PLSR analysis. The RMSECV, $R^2$ and overall standard deviation values values were calculated as 70.4044mg/dL, 0.9048 and 1.0975mg/dL.

**Figure 7.1**: Schematic representation of the ultra-filtration, Raman analysis, data pre-processing and PLSR analysis of the Bu/MTX serum samples. LOD and LOQ is calculated from the prediction plot.

**Figure 7.2**: Reference spectrum of Bu used for EMSC correction. The chemical structure of Bu is shown in the inset. B: Reference spectrum of MTX used for EMSC correction. The chemical structure of MTX is shown in the inset. C: PLSR coefficient plot of Bu (400-1800cm$^{-1}$) from serum filtrate concentrations showing spectral features similar to the Bu reference at 1097cm$^{-1}$ and 1453cm$^{-1}$. D: PLSR coefficient plot of regression against MTX from 1200-1800cm$^{-1}$ showing spectral features similar to the MTX reference at 1351cm$^{-1}$, and 1593cm$^{-1}$. RMSECV were calculated to be 0.0003mg/mL for Bu and 4.02 µM for MTX, respectively.

**Figure 7.3**: Linear predictive model for (A) Bu and (B) MTX built from the PLSR analysis. The LOD and LOQ for Bu were calculated to be 0.0002±0.0001mg/mL and 0.00073±0.00010mg/mL, whereas the LOD and LOQ of MTX were calculated to be 7.8 ±5.0 µM and 26 ±5µM.

**Figure 7.S1**: EMSC corrected and smoothed dataset of A Busulfan from 0mg/mL to 0.05mg/mL in the fingerprint region. B: Methotrexate from 0µM to 500µM from 1200cm$^{-1}$ to 1800cm$^{-1}$. The spectra are offset for clarity.

**Figure 7.S2**: Evolution of RMSECV for (A) Busulfan and (B) Methotrexate. The RMSECV values are calculated to be 0.0003mg/mL for Bu and 4.02µM for MTX.

**Figure 8.1**: Early disease diagnosis, prognosis and treatment is possible with real-time analysis of patient serum using the inverted Raman spectral analysis.

**Figure 8.2**: Thin glass bottomed Lab-Tek plate combined with inverted Raman analysis can collect spectral data from very low amount of samples making it an ideal tool for clinical laboratory analysis.

**Figure 8.3**: PCA scatter plot of Raman data of serum without scaling the analyte spectra (A) and after scaling the analyte spectra to the water content (B). Figure 3B displays less
scatter when compared to Figure 3A, indicating less variability among the spectra.

**Figure 8.4:** Plot of LOD vs Intensity of glucose, busulfan, methotrexate, cholesterol, urea and vitamin B12.

Error! Bookmark not defined.
Chapter 1

Introduction

Vibrational spectroscopic techniques, both Raman and Infrared absorption, are certainly one of the most important analytical techniques available to scientists, as they provide the unique opportunity to investigate the molecular composition of both organic and inorganic compounds. They are routine, standard techniques used for fingerprinting and identifying chemicals, as they can give specific biochemical information without the use of extrinsic labels and without being extremely invasive or destructive to the system studied. Since both techniques are truly label-free, their potential for diagnostic applications has been well investigated and demonstrated, notably in various pathologies (1–7) and therapeutic drug monitoring (8–10).

Bodily fluids (e.g. plasma, serum, saliva or urine) are emerging as a potentially important source of samples for disease diagnosis and therapeutic monitoring, as their collection is largely non-invasive, cost effective and easy (11–17). Blood plasma/serum is not only the primary clinical specimen that contains a large number of proteins, but has also been studied even before genes were known to exist (18). Increasingly, many diagnostic tests are performed on bodily fluids, as sample collection is minimally invasive, compared to biopsy based techniques. However, in hospitals or medical centers, there is a time lapse between collection of bodily fluids, such as urine or blood, from patients and analysis of these samples for diagnose diseases, as these analysis are performed in specialised laboratories. Bodily fluids are usually collected from a large number of patients in a hospital, further delaying the
performance of the analysis and availability of results, which may in turn delay the therapy, and prolong patient anxiety. The accuracy of the test kits that enable point-of-care-testing is often poor and they are frequently avoided because of high cost (19,20). There is a need for objective and cost effective methods capable of accurate early disease diagnosis from bodily fluids in a point-of-care clinical setting. For this purpose, vibrational spectroscopic techniques are appropriate, as they are nondestructive, label-free, rapid, cost-effective, easy to operate, and require minimal sample preparation. Moreover, the use of plasma, serum or urine for diagnostics has the added advantage of being relatively non-invasive compared to conventional diagnostic methodologies such as biopsies.

This study aims to optimise protocols for Raman spectroscopic analysis of human blood plasma/serum in the liquid form, and to explore applications for the analysis and quantification of pathologically significant imbalances of constituent components, as well as to monitor therapeutic drug administration.

1.1 Raman and Infrared Spectroscopy in the analysis of bodily fluids

Vibrational spectroscopy usually refers specifically to the optical techniques of Raman and infrared (IR) absorption spectroscopy. Measurement using these techniques provides a great deal of information that is potentially useful in the medical environment and can be exploited for disease diagnosis and therapy (21). The quest for biomarker identification and analysis through bio-spectroscopy in general is an emerging field with huge potential and has recently been explored through vibrational spectroscopic approaches (6,17,21–23). The sensitivity to detect subtle changes in the biochemical composition and ability to detect the presence of specific biomarkers makes vibrational spectroscopy an ideal diagnostic tool. Considering the
advancement in spectroscopic technologies, and data analysis capabilities, coupled with filtration and fractionation techniques; bodily fluids can be analysed rapidly and non-invasively to detect disease related fluctuations in protein concentration (24–27).

Raman spectroscopy is a complementary tool to IR spectroscopy and is compatible with aqueous samples. This technique allows the analysis to be carried out in the native state of bodily fluids, and, therefore, the additional drying step can be eliminated. A comprehensive proof of concept has been designed and conducted to detect hepatocellular carcinoma (HCC) from patient serum using micro-Raman spectroscopy (22). The aim of the study was to differentiate serum samples of patients with HCC and patients without HCC. Two groups of patients were classified with an overall accuracy of 84.5% to 90.2% for dried serum drops and 86% to 91.5% for freeze-dried serum. A Surface-enhanced Raman scattering (SERS) based immunoassay has been developed to monitor levels of the mucin protein MUC4 in patient serum, which could help in the early detection and diagnosis of pancreatic cancer (28). Another study demonstrated the ability of Raman spectroscopy to detect trace amounts of glucose in urine that was ten-fold diluted, with an accuracy of 92% for abnormal (8mg/mL) and normal (0.15mg/mL) urine samples (29). In recent years, SERS has also been reported to be a good candidate for therapeutic drug monitoring (TDM) of various drugs in biological fluids (8,9,30),(31),(10), since quantitative analysis of drugs can be performed rapidly and higher sensitivity. Critical issues of using SERS for TDM include development of standardised substrates, intense surface enhanced resonance SERS responses from other biological molecules such as carotenoids and also the spectral interference from the fluorescence that could interfere with the drug detection (9,32).
IR absorption spectroscopic analysis of bodily fluids has been explored for disease screening and monitoring numerous conditions including cancer, arthritis, heart diseases, liver diseases and diabetes (21,25,33). A comprehensive study conducted using mid-infrared (mid-IR) spectroscopy of dried serum films demonstrated good accuracy for total cholesterol, triglycerides, total protein, urea and glucose, whereas this method was found to be less suitable for creatinine and uric acid (34). Another study demonstrated that both low density lipoprotein and high density lipoprotein cholesterol can be independently quantified using IR spectroscopy of dried serum films (35). Attenuated Total Reflectance-Fourier Transform Infrared (ATR-FTIR) was successfully used to detect and differentiate blood, saliva, semen and vaginal secretions (36). Freshly collected bodily fluids were deposited on to the analysis stage and measurements were carried out upon immediate deposition, using further scans at regular time intervals until the sample was completely dry. The identification was based on the unique spectral pattern and peak frequencies corresponding to sugars, proteins and phosphates in each bodily fluid. Blood components, such as serum and plasma, can be exploited for liquid biopsies as they contain biomarkers that can be used for disease diagnosis. ATR-FTIR has been proven to be a promising screening tool for diagnosing ovarian cancer from human blood serum (37). Plasma/serum based ATR-FTIR successfully discriminated ovarian cancer from controls with a success rate of ~97% from plasma and ~95% when compared to the histological diagnosis by a pathologist. Centrifugal filtration techniques coupled with ATR-FTIR of human serum have been employed to deplete the abundant high molecular weight proteins and subsequently enhanced the ability to monitor changes in the low molecular weight constituents. Glucose was used to spike the serum and it was demonstrated that fractionating the serum prior to spectroscopic analysis improves the
quantitative model based on the partial least squares regression (PLSR) algorithm (38). As IR spectroscopy is an absorption based technique, water cannot be used as a solvent due to its intense absorption in the IR region (39). IR analysis of the serum samples has therefore been predominantly performed on air dried samples which leads to chemical and physical inhomogeneity due to the adsorption of serum proteins on to the substrate surface based on their differing affinities, known as the Vroman effect (40–42). The Vroman effect is the main limiting factor of using air-dried samples for spectral analysis. Thus, the use of dried plasma samples will give rise to variations in the spectral features due to chemical and physical inhomogeneity, leading to unreliable results and, is not the ideal protocol for diagnostic applications.

An improved protocol of Raman spectroscopy set up, coupled with fractionation of serum using centrifugal filters to concentrate and separate low molecular weight proteins was demonstrated (43). Whereas FTIR spectra were recorded in aqueous solutions of gelatin at concentrations as low as 100mg/L, using Raman spectroscopy, high quality spectra of gelatin solutions as low as 10mg/L was achieved. Spectral features of human serum were found to be weak and partially obscured by water features. Dried deposits are shown to be physically and chemically inhomogeneous resulting in unreliable results. Concentration of the serum using commercially available centrifugal filter devices resulted in enhanced spectral intensity and quality. Improved analysis of serum using Raman spectroscopy was reported when the sample was analysed in the inverted geometry using the water immersion objective with a 785nm laser as source.

This project aims to evaluate the potential of Raman spectroscopy for the analysis of blood plasma/serum in the liquid state in order to detect subtle variations in the sample composition
and specific biomarkers linked to numerous pathologies and therapeutic drug monitoring. The initial stages involve optimisation of the analysis protocol, exploring the available range of laser source wavelengths, sample measurement substrates and measurement geometries. A simulated blood plasma is employed to establish the appropriateness and sensitivities of the combination of Raman spectroscopic, data pre-processing and multivariate analysis techniques. Centrifugal filtration and ion exchange chromatography are explored as a method for fractionation and concentration of the constituent proteins to improve the measurement sensitivity.

1.2 Research question and hypothesis

**Question:**

Can Raman spectroscopy detect imbalances in plasma/serum analytes and small molecule biomarkers or drugs to enable early disease diagnosis or monitor patient therapies?

**Hypothesis:**

A point-of-care, rapid, cost-effective disease/therapeutic monitoring system can be developed using Raman spectroscopy.

1.3 Thesis summary

**Chapter 2** is a reproduction of the submitted book chapter titled as ‘Vibrational Spectroscopic Analysis and Quantification of Proteins in Human Blood Plasma and Serum’, Vibrational Spectroscopy in Protein Research, Elsevier. It details the application of vibrational spectroscopic analysis for quantification of proteins in human plasma/serum. This
Chapter also gives an insight into the constituents of serum/plasma and the pathologies related to serum/plasma analytes. Moreover, this chapter describes the fundamental principle underlying the Raman and IR spectroscopic analysis and the basic instrumentation.

**Chapter 3** focuses on the methodology that has been used throughout the study in detail. The measurement protocol of Raman spectroscopy for liquid serum analysis, serum fractionation techniques and the use of multivariate analysis techniques to extract information from the Raman data sets are discussed. This chapter will also describe in detail the optimal Raman measurement protocol for aqueous samples in terms of geometry of the instrument, substrate, sample volume and laser line.

**Chapter 4** reproduces the published journal article entitled ‘Raman spectroscopic analysis of high molecular weight proteins in solution: considerations for sample analysis and data pre-processing’. Analyst. 2018;143(24):5987–98. It explores the potential of Raman spectroscopy, coupled with multivariate regression and protein separation techniques, to monitor diagnostically relevant changes in high molecular weight proteins in liquid plasma. Measurement protocols to detect the imbalances in plasma proteins as an indicator of various diseases using Raman spectroscopy are also optimised. Two types of data preprocessing methods (Rubberband and Extended Multiplicative Signal Correction (EMSC)) are introduced and compared in this chapter.

**Chapter 5** consists of the published journal article entitled ‘Analysis of bodily fluids using vibrational spectroscopy: a direct comparison of Raman’. Analyst. 2019;144:3334 – 3346. It demonstrates the use of Raman spectroscopy, similarly coupled with centrifugal filtration and multivariate analysis techniques, to monitor diagnostically relevant changes of glucose
in liquid serum samples, and compares the results with similar analysis protocols using infrared spectroscopy of dried samples. This study has been conducted on 25 patient serum samples.

**Chapter 6** reproduces the published journal article entitled ‘Raman spectroscopic screening of High and Low molecular weight fractions of human serum’ Analyst. 2019, DOI: 10.1039/C9AN00599D. It illustrates the suitability of Raman spectroscopy as a bioanalytical tool, when coupled with centrifugal filtration and multivariate analysis, to detect imbalances in both high molecular weight (total protein content, γ globulins and albumin) and low molecular weight (urea and glucose) fractions of the same samples of human patient serum, in the native liquid form. The strategy to digitally remove spectral interferents such as β-carotene is demonstrated here.

**Chapter 7** is adapted from a submitted journal article titled as ‘Raman spectroscopy as a potential tool for label free therapeutic drug monitoring in human serum: the case of Busulfan and Methotrexate’ Analyst, 2019, which describes the methodology, based on Raman spectroscopy coupled with multivariate analysis, to determine the Limit of Detection (LOD) and Limit of Quantification (LOQ) for TDM in human serum, using the examples of Busulfan, a cell cycle non-specific alkylating antineoplastic agent, and, Methotrexate, a chemotherapeutic agent and immune system suppressant.

**Chapter 8** summarises the main findings from this thesis and provides an overview of the recent advancements in the field of liquid Raman spectroscopic analysis of bodily fluids. The impact of scaling the data to the water content using EMSC algorithm and inverse correlation between LOD and spectral intensity is also highlighted in this chapter. The chapter has been
submitted in its entirety, as a review article titled as ‘Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: Recent advances’, Analytical and Bioanalytical Chemistry, 2019.

**Chapter 9** outlines the overall conclusion of the thesis, focusing on the shortcomings and challenges of the work presented, methods to improve the accuracy of the methodology and potential applications of liquid Raman spectroscopy in various fields.

In the chapters which are adapted from published works, the source referencing format of the original has been retained, but changes in figure and section numbering have been made, as required for the thesis format. Supplementary material related to the corresponding work is also added at the end of the chapters.
1.4 References


Chapter 2

Vibrational Spectroscopic Analysis and Quantification of Proteins in Human Blood Plasma and Serum

The following chapter reproduces the submitted book chapter entitled ‘Vibrational Spectroscopic Analysis and Quantification of Proteins in Human Blood Plasma and Serum’, Vibrational Spectroscopy in Protein Research, Elsevier

Author List- Clément Bruno, James M. Cameron, Drishya Rajan Parachalil, Matthew J. Baker, Franck Bonnier, Holly J. Butler, Hugh J. Byrne.

CB primarily contributed to sections 2.1, 2.2, 2.3, 2.4 and 2.5, JC contributed to sections 2.4, 2.5, 2.6 2.7 and 2.8, DRP contributed to sections 2.1, 2.2, 2.3, 2.4 and 2.5, while MJB, HJBu and HJBy contributed to chapter design, formatting and proofing.

2.1 Abstract

In this chapter, we outline current the clinical procedures for blood tests, examine the capability of biomedical vibrational spectroscopy for disease diagnostics and monitoring, and discuss the potentiality for the techniques to be successfully translated into the clinic. Alterations in biomolecular components in human blood are commonly used as an indication of disease states, namely differences in protein concentration. Unfortunately, conventional test kits currently employed in hospitals suffer from long time delays, meaning patients often
have to wait anxiously for their test results. Vibrational spectroscopic techniques, such as infrared and Raman, have the ability to replace current practices, as they are label-free, cost-effective, easy to operate, and require minimal sample preparation. The sensitivity to subtle changes in biochemical composition makes them ideal diagnostic tools, and recent advances in technology and data analytics means bodily fluids can be analysed rapidly and non-invasively to detect disease related fluctuations in protein concentration.

**Keywords:** Vibrational Spectroscopy, Infrared, Raman, Plasma, Serum, Diagnostics, Patient Monitoring, Protein, Quantification

### 2.2 Introduction

Bodily fluids (e.g. plasma, serum, saliva or urine) are emerging as an important source of samples for disease diagnostics and therapeutic monitoring, as their collection is largely non-invasive, cost effective and relatively simple. Analyses are performed in specialised laboratories in hospitals or medical centres, and there is a considerable time lapse between collection of bodily fluids from patients and delivery of the results, while the samples are analysed for disease diagnostic biomarkers. Bodily fluids are collected from a large number of patients in a hospital, further delaying the performance of the analysis and availability of results, which may in turn delay the therapeutic intervention, and prolong patient anxiety to the detriment of patient management. The accuracy of the test kits that enable point-of-care-testing is often poor and they are frequently avoided due to high cost [1], [2]. Furthermore, some diseases are still without biomarkers for diagnosis and disease stage monitoring. Although in recent years, global analysis techniques based on ‘omics’ have emerged, using multi-component concentrations as specific biomarkers, there is still a need for improved,
objective and cost effective methods capable of accurate early disease diagnosis from bodily fluids in a point-of-care clinical setting, whilst also developing new tools for the identification of novel biomarkers. For both these purposes, vibrational spectroscopic techniques are appropriate, as they are non-destructive, label-free, rapid, cost-effective, easy to operate, and require minimal sample preparation.

The vibrational spectroscopic techniques Raman and Infrared absorption, are valuable analytical techniques available to scientists, as they provide the unique opportunity to investigate the molecular composition of both organic and inorganic compounds. They are routine, standard techniques used for fingerprinting and identifying chemicals, as they can give molecularly specific chemical information without the use of extrinsic labels and without being extremely invasive or destructive to the system studied. Since both techniques are truly label-free, their potential for diagnostic applications has been well investigated and demonstrated, and increasingly, interest has turned towards their application in the analysis of bodily fluids, particularly human blood plasma and/or serum for disease diagnostics and patient monitoring [3], [4].

2.3 Analysis of Biofluids

For many years, the field of medicine has used bodily fluids, or biofluids, for patient monitoring and screening for disease. Within the body there are two major fluid categories; intracellular fluid (ICF) and extracellular fluids (ECF). ICF refers to fluids contained inside the cell plasma membrane, whilst EFC can be further differentiated into the interstitial fluid (IF) surrounding cells (~75%), the fluid component of the blood called plasma (~25%), and the transcellular fluids, a small specialised fraction of excreted biofluids (urine, saliva, gastric
fluids etc.). An adult human body is composed of 50-60% water, about two thirds of which is found in the ICF and one third in the ECF, enabling the crucial physiological process of osmosis to take place. Described as the passage of water across a semipermeable membrane, along a concentration gradient, osmosis is essential for cells to maintain homeostasis. Key to this transfer of biochemical constituents between the extra and inter cellular environment is the blood (plasma/serum), and therefore its composition can evolve over time, reflecting changes in patient health.

However, within this complex matrix, only a limited number of molecules, or biomarkers, have well-established symptomatic deviations from normal ranges that can be used for diagnosing specific pathologies. As a consequence, biomarker discovery has become an active field of research, aiming to provide tools for early diagnosis associated with better prognosis for patients. Cerebrospinal, synovial, pleural, pericardial, peritoneal, lymphatic fluids, the aqueous humour and mucosal secretions, such as urine, saliva, sputum, are thus considered as part of the ECF system. These fluids derive or interact with blood and can therefore similarly contain sentinel indicators of cellular and organ function or dysfunction. Thus, human biofluids are considered to be powerful sources of clinical biomarkers [5], [6]. In terms of disease diagnostics and prognosis, bodily fluids are an interesting alternative to cells and tissues [7]. It is expected that modification of the overall composition of the biochemical state of bodily fluids could deliver crucial information about patient health and disease states, enabling early disease diagnosis and administration of treatment [8]. Disease diagnosis from bodily fluids could potentially be developed into a dynamic diagnostic environment that will enable early disease diagnosis even before the disease becomes symptomatic.
These biofluids, which provide organ-specific information, are increasingly used for diagnosis; however, blood is considered the largest biomarker reservoir of the body. Since most of the clinical analytical instruments are accurate for both serum and plasma, these two terms are used interchangeably in most clinical tests [9]. Notably, many studies that have been reported to be carried out in serum were in fact carried out in plasma [10]-[14].

2.3.1 Blood Sample: preparation of plasma versus serum

A range of blood-based tests are routinely conducted in a clinical setting. Each analysis protocol has strict pre-analytical requirements, according to which molecular or cell specific analysis assays are to be used, even starting with the choice of sampling tubes. For example, measurement of clotting factors level (by mean of factor consumption and degradation) or complete blood cell count (through cell lysis and segregation in clot) cannot be analysed in dry tubes (i.e. without anti-coagulant reagents) [15]. Thus, only citrate-treated tubes are compatible with haemostasis evaluation. While plasma and serum are both cell-free fluids obtained from blood samples by centrifugation, they differ on the basis of whether the sample has been allowed to clot, or not (Figure 2.1).

For plasma preparation, tubes treated with anticoagulant (such as citrate- or heparin) are used for blood collection. Routinely, refrigerated centrifugation for 10 minutes at 2,000 x g concentrates unwanted cells and platelets. For serum preparation, whole blood is allowed to clot at room temperature for about 15–30 minutes. The clot is removed by refrigerated centrifugation at 1,000–2,000 x g for 10 minutes, often separated by a gel component to avoid contamination. It is important to immediately transfer the supernatant (plasma or serum) into a clean polypropylene tube and maintain samples at 2–8°C while handling. If the samples are
not analysed immediately, they should be stored at –20°C or preferably lower. It is also recommended to avoid freeze-thaw cycles, because this may have detrimental effects on many serum components.

**Figure 2.1.** Obtaining (a) Plasma and (b) Serum from blood samples

### 2.3.2 Composition of Plasma and Serum

Blood serum and plasma are predominantly composed of water (~90%), minerals, organic substances and gas (oxygen, carbon dioxide). Proteins are the predominant molecular components of blood plasma, the remaining constituents being carbohydrates, lipids and amino acids (Figure 2.2). Serum albumin, globulins, fibrinogen and a handful of other abundant proteins account for 99% of total serum proteins, while the remaining 1% is composed of low abundance circulatory proteins. Additionally, plasma or serum contain
more than 114,000 known metabolites at varying concentration level (<1 nmol/L to mmol/L) [16]-[18]. In general, to measure factors other than coagulants, serum is preferred as it is less complex and more sensitive.

Figure 2.2. Molecular structural figures as examples of a sugar (A), a lipid (B) and an amino acid (C)

2.3.3 Non-Protein constituents

Electrolytes are a key constituent of blood plasma and serum that play key roles in the human body, including maintaining pH balance and cellular communication. While sodium, chloride and bicarbonate (approximately 140 mmol/L, 100 mmol/L and 25 mmol/L, respectively) are considered highly concentrated, others such as potassium, phosphate, calcium and magnesium (4.5 mmol/L, 2.5 mmol/L, 2.5 mmol/L and 1 mmol/L, respectively) are found with lower contents [16]. Other minerals such as iron, copper, zinc, aluminium, lead can be found in even lower concentrations (≤μmol/L).

Energy is provided to cells in the form of glucose, and thus blood contains a large amount of glucose and its derivatives. Normal glucose levels are between 3.3-5.6 mmol/L, and is strictly regulated by endocrine system [18]. Diabetes mellitus is a disorder where this regulation has
failed, either due to insulin resistance or non-production, and can be identified by the glucose concentration in the blood. Hyperglycemia is defined as above the normal range, and hypoglycaemia as below.

Lipid concentrations also are indicators for pathological conditions and are therefore routinely monitored. A healthy range of total cholesterol is between 100-199 mg/dL in adults, whilst triglycerides are routinely below 150 mg/dL [16]; both are linked to risk of heart and blood vessel disease and can be used to screen patients to prevent severe conditions leading myocardial infarction.

Amino acids, the building blocks of peptides and proteins, are found at levels 2300-4000 µmol/L and are produced from protein catabolism. The most abundant amino acids, glutamine and alanine, are found at concentrations of 600 µmol/L and 300 µmol/L, respectively [16]. Amino acids metabolism is complex and depends on a thin balance between protein anabolism and catabolism. Any dysfunction in a metabolic pathway leads to hyperammoniemia (>400 µmol/L).

Nitrogen metabolites are also used routinely in clinics to witness body dysfunctions. Ammonia (6-35 µmol/L) is a highly toxic substance and thus is transported in blood mostly in the form of glutamine [16]. An increased level of ammonia can indicate hepatic dysfunction or hyper catabolism. Urea is the end product of protein (or amino acid) degradation and is formed in the liver. Increased blood levels above 2.5-8.3 mmol/L generally indicate elimination dysfunction, and thus can highlight renal failure.
2.3.4 Proteins

In total, more than 20,000 human proteins are known. Of these, more than 1,500 are extracellular [19]. Blood born proteins form a heterogeneous group comprising more than 300 types of proteins. The total concentration of serum proteins (serum total protein) ranges from 60-80 g/L and is dependent on many factors: rate of the synthesis and degradation, distribution in body fluids, hydration and elimination. Physiologically higher concentrations can be observed in the standing position and in association with increased muscle activity. On the other hand, lower concentrations are physiological for children, pregnant women or after prolonged fasting [20]. Some of the abundant proteins like fibrinogen and clotting factors are only found in plasma - about 8% of total proteins - while albumin and globulins can be measured in both plasma and serum. Albumin is the most abundant protein, accounting for about 55% of total proteins, which, combined with globulins, represents about 90% of the overall protein content of serum in healthy patients (Table 2.1). The dynamic range of concentrations in plasma and serum can be one of the greatest challenges in analysing the proteome, requiring advanced and intensive methods such as mass spectrometry-based protocols. However, for a large number of clinically relevant protein biomarkers, routine analysis is required to enable early disease diagnosis.
Table 2.1. Relative content of abundant proteins in plasma and serum

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Protein type</th>
<th>Function</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma</td>
<td>Fibrinogen</td>
<td>Coagulation</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Clotting Factors</td>
<td>Fibrin formation from fibrinogen</td>
<td>&lt;1%</td>
</tr>
<tr>
<td></td>
<td>Albumin</td>
<td>Blood vessel integrity</td>
<td>55%-65%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carrier for insoluble molecules</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue growth and healing</td>
<td></td>
</tr>
<tr>
<td>Serum and</td>
<td>α-1 globulin</td>
<td>High density lipoprotein (HDL)</td>
<td>1-5%</td>
</tr>
<tr>
<td>Plasma</td>
<td>α-2 globulin</td>
<td>Haptoglobin binds haemoglobin (iron)</td>
<td>5-11%</td>
</tr>
<tr>
<td></td>
<td>B globulin</td>
<td>Carrier and part of defence system against infection</td>
<td>7-13%</td>
</tr>
<tr>
<td></td>
<td>γ globulin</td>
<td>Antibodies – immune system</td>
<td>10-18%</td>
</tr>
</tbody>
</table>

2.3.4.1 Fibrinogen

The main difference between blood plasma and serum is the removal of clotting factors, predominantly fibrinogen, to produce the latter. Fibrinogen is a dimeric 340 kDa (0.4% in human plasma) plasma glycoprotein synthesised by the liver and plays a major role in blood coagulation. When blood clotting is activated, the circulating fibrinogen turns into fibrin and a stable clot is formed [21]. The normal concentration of fibrinogen in the human body is ~3 mg/mL, and any variation in this concentration can be an indicator of disease states. Many clinical studies have consistently shown elevated levels of fibrinogen in patients with cardiovascular disease and thrombosis [22]-[24]. A study conducted by Aleman et al. has shown that both elevated circulating fibrinogen (hyperfibrinogenemia) and abnormal fibrinogen levels are observable in plasmas from patients with venous thrombosis [25].
However, the study undertaken by Klovaite et al. indicated that elevated plasma fibrinogen levels are associated with increased risk of pulmonary embolism rather than deep venous thrombosis [26]. Another study conducted by Toss et al. showed that increased fibrinogen levels are associated with persistent Chlamydia pneumonia infection in unstable coronary artery disease [27]. A study conducted by the Emerging Risk Factors Collaboration, UK, also predicted a positive correlation between C-Reactive Protein (CRP), fibrinogen, and cardiovascular disease [28]. These studies and others suggest a strong correlation between increased levels of plasma fibrinogen and heart diseases that could be used as a diagnostic indicator [29], [30].

2.3.4.2 Albumin

In humans, albumin is the most abundant plasma protein, normally constituting about 50% of human plasma protein, and has a molecular weight of 66 kDa. Albumin is a protein made by the liver and its main role is to maintain the osmotic pressure of the blood compartment, provide nourishment of the tissues, and transport hormones, vitamins, drugs, and other substances such as calcium throughout the body [31]. The normal concentration of albumin in the human body is 30 g/L. Acute dehydration is the only clinical situation that is found to cause an increase in albumin concentration [8]. In the event of critical illness, the rates of synthesis and degradation of albumin are altered, leading to an abnormal distribution of albumin between the intravascular and extravascular compartments. The concentration of albumin decreases dramatically in critically ill patients and does not increase until the recovery phase of the illness [32]. Increased capillary leakage is the main reason for the altered distribution of albumin in critical illness. This is reported to occur in sepsis and after
major surgical stress [33], [34]. Cirrhotic patients are highly prone to suffer from septic shock. Several studies have demonstrated that the functions of albumin, such as ligand binding and transport of various molecules, can be applied in the treatment of cirrhotic patients and patients suffering from other end stage liver diseases [35]. It is clear that closely monitoring the variation in albumin concentration could act as an indicator of liver diseases and other related pathologies.

2.3.4.3 Globulins

Globulins represent fractions of proteins identified in the plasma/serum known as alpha-1 globulins, alpha-2 globulins, beta globulins and gamma globulins (Table 2.2). Each one of these fractions have various biological functions, including immune functions.

<table>
<thead>
<tr>
<th>Alpha-1 globulins Range: 1-4 g/L</th>
<th>Alpha-2 globulins Range: 3-9 g/L</th>
<th>Beta globulins Range: 7-15 g/L</th>
<th>Gamma globulins Range: 8-16 g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-1 antitrypsin</td>
<td>Very Low Density Lipoprotein (VLDL)</td>
<td>Transferrin</td>
<td>Immunoglobulins</td>
</tr>
<tr>
<td>α -1 acid glycoprotein (orosomucoid)</td>
<td>Haptoglobin</td>
<td>Low density lipoprotein (LDL)</td>
<td>C-reactive protein (CRP)</td>
</tr>
<tr>
<td>α -1 Antichymotrypsin</td>
<td>Alpha-2 macroglobulin</td>
<td>Complement protein C3</td>
<td>Lysozyme</td>
</tr>
<tr>
<td>High-density lipoprotein (HDL)</td>
<td>Ceruloplasmin</td>
<td>Fibrinogen</td>
<td>Alpha fetoprotein</td>
</tr>
<tr>
<td>Thyroxine-binding globulin (TBG)</td>
<td>Glycoprotein</td>
<td>Hemopexin</td>
<td></td>
</tr>
<tr>
<td>Prothrombin</td>
<td>Complement protein C4</td>
<td>Properdin</td>
<td></td>
</tr>
</tbody>
</table>
2.3.4.4 Immunoglobulins

Immunoglobulins (Ig) are gamma globulin proteins present in blood plasma and mucosal secretions, and are delivered to sites of inflammation within tissue. They are antibodies produced by B lymphocytes, white blood cells. They are among the most abundant proteins in the blood, comprising 20% (by weight) of total plasma/serum proteins (Table 2.3). Based on structure and protein composition, immunoglobulins are divided into five classes: IgG, IgA, IgM, IgD and IgE based on structure and protein composition (Table 2.3).

Table 2.3. Details of immunoglobulin subtypes

<table>
<thead>
<tr>
<th>Subtype</th>
<th>IgM</th>
<th>IgG</th>
<th>IgA</th>
<th>IgE</th>
<th>IgD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular weight</td>
<td>900 kDa</td>
<td>150 kDa</td>
<td>320 kDa</td>
<td>200 kDa</td>
<td>180 kDa</td>
</tr>
<tr>
<td>Normal concentrations</td>
<td>0.5-2 g/L</td>
<td>10-16 g/L</td>
<td>1-4 g/L</td>
<td>10-400 µg/L</td>
<td>0-0.4 g/L</td>
</tr>
<tr>
<td>Abnormal situations</td>
<td>Myeloma ↑ Immunodeficiency ↓</td>
<td></td>
<td></td>
<td>Allergic reactions ↑</td>
<td></td>
</tr>
</tbody>
</table>

Ig classes vary by the number of units; one unit (Ig G, IgD and IgE), two units (IgA) or five units (IgM). The simplest antibody molecule can be represented as a Y shape (Figure 2.3), composed of four polypeptide chains: two identical heavy chains (H) and two identical light chains (L) oriented parallel to each other and linked by di-sulphide bonds [36]. Both light
and heavy chains are characterised by variable and constant regions. L chains only consist of one variable (VL = variable light) and one constant (CL = constant light) domain, while heavy chains have one VH (variable heavy) and three to four CH (constant heavy) domains respectively, corresponding to IgG/IgA and IgM/IgE. The terminal end of both L and H chains present extremely variable (“hypervariable” regions), forming the antigen binding site. There are 2 types of L chains, λ (lambda) or κ (kappa) that can be found in all types of Ig, but never together. H chains are Ig specific and designated γ (IgG), α (IgA), μ (IgM), ε (IgE) and δ (IgD). Each heavy chain has about twice the number of amino acids and molecular weight (~50 kDa) compared to light chains (~25 kDa), resulting in a total immunoglobulin monomer molecular weight of approximately 150kDa. The constant region confers immunoglobulin its biologic activity while the variable regions form a complex, conformational molecular arrangement for the attachment of each specific antigen.

Figure 2.3. Structure of immunoglobulins
2.4 Pathology of plasma proteins

2.4.1 Abundant proteins

Generally, routine serological evaluation of a patient includes analysis of a panel of serum electrolytes, total protein and glucose concentrations. Estimating total protein, globulin and albumin content is important to assess the global and nutritional status of patients [8], [37]. The Biuret assay is the most common method used to quantify total protein levels in blood serum [38], which is the most compatible with routine application, in terms of sensitivity and linearity of range, as well as in terms of reaction time. Except in some known physiological states, a low concentration is a sign of undernutrition, malabsorption or hepatic disease and occurs in the case of severe protein loss (burn victims, renal failure). On the other hand, hyperproteinaemia is found in conditions such as dehydration, myeloma, and systemic lupus erythematosus.

Although total serum protein estimation has limited diagnostic potential when compared to albumin or globulin, due to lack of specificity, its relevance in the evaluation of patients with clinical conditions such as malnutrition, renal malfunction, liver diseases and immune disorders cannot be ignored [37], [39]. A normal level of total serum protein between 60-83 g/L indicates healthy nutritional status and normal liver function. Reduced serum total protein is predominantly found in patients with kidney disorders, HIV and the elderly [40], [41].

γ-globulins, produced by lymphocytes and plasma cells in lymphoid tissue, are large protein molecules that consists of the immunoglobulins: IgM, IgA, and IgG [42]. An elevation in γ-globulins is a characteristic abnormality in the serum proteins in liver diseases and carcinoma.
of the gastro-intestinal tract or breast. Testing globulin levels in serum routinely provides key information that helps diagnose various conditions and diseases that affect the immune status. Liver diseases, chronic inflammatory diseases, haematological disorders, infections and malignancies cause excess Ig levels, whereas humoral immunodeficiencies cause low Ig levels. Radial immunodiffusion (RID) is the gold standard method for measuring globulins. All conventional methods used for testing total protein content and globulin measurement make use of expensive disposables and are labour intensive. With escalating medical costs and budget constraints, a cost effective alternative technology is desirable.

Serum protein electrophoresis (SPEP) is a cost effective method for separation of proteins, based on their net charge (positive or negative), size and shape. It enables visualisation of the two major types of protein present in the serum: albumin and the globulin proteins.

![Figure 2.4. Disease patterns in serum protein electrophoresis (SPEP)](image)

The largest peak, closest to the positive electrode, reflects the high concentration of albumin, while globulins are represented by multiple smaller features (Figure 2.4). However, the
pattern of five globulin categories (alpha-1, alpha-2, beta-1, beta-2, and gamma) contains the most relevant information for diagnosis. A dense narrow band that is composed of a single class of immunoglobulins is categorised as monoclonal gammopathy. It is the result of a malignant clone producing only one type of antibody that leads to a thin peak in protein electrophoresis. A broad-based band in the gamma region suggests a polyclonal increase in immunoglobulins (Polyclonal gammopathy). When the β-globulins and γ-globulins do not separate, it can be specific to liver disease (cirrhosis), but is also common in autoimmune disease, chronic viral or bacterial infections. SPEP is a rapid technique to detect a number of conditions based on qualitative and quantitative patterns of the serum fractions. Table 2.4 provides examples of conditions with increased or decreased levels in protein fractions.
### Table 2.4. Serum Protein Fractions and Conditions Associated with an Increased or Decreased Level

<table>
<thead>
<tr>
<th>Serum Protein</th>
<th>Decreased</th>
<th>Increased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albumin</td>
<td>Malnutrition, Cachexia (wasting syndrome), liver disease, nephrotic syndrome, Impaired liver function, protein-losing enteropathies, Haemorrhage, severe burns</td>
<td>Indicated for dehydration</td>
</tr>
<tr>
<td>Alpha-1</td>
<td>Alpha-1 antitrypsin deficiency, Nephrotic syndrome, Liver dysfunction</td>
<td>Pregnancy, Inflammatory states</td>
</tr>
<tr>
<td>Alpha-2</td>
<td>Haemolysis, Liver disease</td>
<td>Nephritic syndrome, Adrenal insufficiency, Adenocorticosteroid therapy, Advanced diabetes mellitus, Hyperthyroidism</td>
</tr>
<tr>
<td>Globulins</td>
<td>Protein malnutrition</td>
<td>Biliary cirrhosis, Hypothyroidism, Nephrosis, Polyarteritis nodosa, Obstructive jaundice, Cushing’s disease, Third-trimester pregnancy, Iron-deficiency anemia</td>
</tr>
<tr>
<td>Beta 1-2</td>
<td>Agammaglobulinemia, hypogammaglobulinemia</td>
<td>Cirrhosis, Multiple myeloma, Hodgkin’s disease, Chronic lymphocytic Leukaemia, Amyloidosis, Rheumatoid</td>
</tr>
</tbody>
</table>

#### 2.4.2 Low abundance proteins

Currently, there are various proteins that are often used in the diagnosis and monitoring of different pathologies. For example hepatic function is evaluated by ASAT (aspartate amino transferase), ALAT (alanine amino transferase), ALP (alkaline phosphatase) and GGT
(gamma glutamyl transferase) activities. For cardiac injuries, such as acute coronary syndrome, diagnostic and patient care is dependent on the troponin test. Recently, a number of studies have proven that imbalances in plasma protein levels can be linked to the presence of numerous disease states [48], [49]. A method incorporating the ability of polyethylene glycol fractionation and immunoaffinity depletion to detect plasma biomarkers has been reported [50]. This method successfully identified 135 low abundance proteins with concentration levels less than 100ng/mL. A high accuracy mass spectrometry based proteomics method has been reported to characterise proteins in the plasma of patients with an acute bone fracture, leading to the discovery of several new proteins which were not previously reported in plasma [51]. Addonna et al. developed a pipeline by integrating the proteomic technologies used from the various stages of discovery of plasma biomarker identification, to identify early biomarkers of cardiac injury [52]. Patients were allowed to be their own controls by sampling blood directly from patient hearts before, during and after controlled myocardia injury. Liquid chromatography mass spectroscopy – LCMS – detected 121 highly differentially expressed proteins and >100 novel candidate biomarkers for myocardial infarction (MI) [53]. Ray et al. identified 18 signalling proteins in blood plasma that can be used to differentiate Alzheimer's disease samples and control subjects with close to 90% accuracy and to identify patients who had mild cognitive impairment that progressed to Alzheimer's disease 2–6 years later. This molecular test for Alzheimer's disease could lead to early diagnosis and better treatment [54].

2.4.2.1 Cytochrome c
Cytochrome c is a water soluble, ~12kDa heme protein found loosely attached to the inner membrane of the mitochondrion. Cytochrome c normally resides in the mitochondrion and is released into the blood in the event of cell death, triggering inflammation [55]. This protein is essential in mitochondrial electron transport and also acts as an intermediate in apoptosis [56], [57]. More recently, it was reported that cytochrome c can be used as an indicator of the apoptotic process in the cell [58]. This study demonstrated that cytochrome c can act as an in vivo apoptosis indicator and prognostic marker during cancer therapy using a cytochrome c enzyme linked immune sorbent assay (ELISA) kit that was modified to increase sensitivity. In another study, a sandwich ELISA method was used to measure serum cytochrome c levels to quantify the extent of apoptosis in systemic inflammatory response syndrome (SIRS) [59]. The prognostic significance of cytochrome c concentrations was investigated and the ability of this method to assess the severity of organ dysfunction and help to predict the prognosis of SIRS was demonstrated. Release of cytochrome c into circulation has been reported in patients with myocardial infarction [60] and several liver diseases [61]. The mean cytochrome c level recorded in patients with liver diseases was found to be 187.1 ng/mL, whereas that of healthy controls was 39.8 ng/mL. A number of studies have been undertaken to show that cytochrome c can be used as a potential clinical marker of molecular and cellular damage [55]. Cytochrome c was identified in mitochondrial damage-associated molecular patterns, along with interleukin-6 (IL-6), as a marker of inflammation in haemodialysis patients [62]. An elevated level of cytochrome c in human plasma/serum can be indicative of the presence of various pathologies. Therefore, this mitochondrial protein has a huge potential to be used as a clinical marker for these diseases at an early stage.
2.5 Vibrational spectroscopic analysis of bodily fluids

It is recognised, however, that conventional test kits commonly employed in a hospital environment for plasma/serum analysis suffer from long time delays due to the need for specialised laboratories, which may in turn delay the therapy, and prolong patient anxiety [63]. The development of optical methods for biomedical applications is an emerging field with huge potential [64], and has recently been explored through vibrational spectroscopic approaches [65], [66]. The sensitivity to subtle changes in biochemical composition makes vibrational spectroscopy an ideal diagnostic tool. Considering the advancement in spectroscopic technologies, and data analysis capabilities, coupled with filtration and fractionation techniques, bodily fluids can be analysed rapidly and non-invasively to detect disease related fluctuations in protein concentration [67]-[70].

2.5.1 Vibrational Spectroscopy

Vibrational spectroscopy usually refers specifically to the optical techniques of Infrared (IR) absorption and Raman scattering spectroscopy, as well as inelastic neutron scattering. It is a subset of spectroscopy which analyses vibrations within a molecule (or material). The vibrations are characteristic of the molecular structure and, in polyatomic molecules, give rise to a spectroscopic “fingerprint”. The spectrum of vibrational energies can thus be employed to characterise a molecular structure, or changes to it due to the local environment or external factors (e.g. radiation, chemical agents).

The number of vibrational modes for a given molecule will depend on its structure. A molecule with N number of atoms will have 3N degrees of freedom. Generally, non-linear
molecules will exhibit $3N-6$ vibrational modes, the 6 non vibrational degree of freedom corresponding to three translational and three rotational modes around the x, y and z axes. In contrast, as linear molecules are unable to rotate upon their axis, one of the rotational degrees of freedom is lost, and hence they can be described as having $3N-5$ vibrational modes [71]. Bond stretching and bending are the two fundamental types of molecular vibration; symmetric or asymmetric stretching alters the bond length, while bending vibrations consist of changing the bond angle, by twisting, rocking, wagging and scissoring (Figure 2.5).

![Figure 2.5. Common vibrational modes of chemical bonds.](image)

Vibrational energies fall within the mid - Infrared (IR) region of the electromagnetic spectrum and are commonly probed through IR absorption spectroscopy. Following the discovery of IR radiation by Herschel in 1800 [72], initial applications of IR absorption spectroscopy were limited to astronomy and astrophysics [73]. In material sciences, significant advances were made by 1900 when Abney and Festing recorded spectra for 52 compounds, correlating absorption bands with molecular structures [74]. Coblentz helped
establish IR spectroscopy as a routine analytical tool, cataloguing the spectra of hundreds of substances, both organic and inorganic. Technological developments post WWII aided considerably in establishing IR spectroscopy as a routine laboratory characterisation technique, but none more so than the development of commercial Fourier Transform IR (FTIR) spectrometers in the 1960s and 70s [75], [76] and FTIR microscopes in the late 1980s [77].

Raman spectroscopy is a complementary technique with origin in the discovery of the Raman effect in 1928 [78], for which C.V. Raman was awarded the Nobel prize in physics in 1930. In 1998 the Raman Effect was designated an ACS National Historic Chemical Landmark, in recognition of its importance in materials and process analysis. Raman spectroscopy remained largely a curiosity until the advent of the laser in the 1960s, and the revolution in Charged Coupled Detector (CCD) arrays in the 1980s and 1990s added to the benefits of high laser source intensities. In addition, the development of narrow band laser line rejection filters meant that the huge losses in signal from traditional triple monochromator systems could be overcome with the combination of a filter set and a single spectroscopic grating. Furthermore, the significant reductions in acquisition time with multichannel signal detection enabled significant improvements in signal to noise ratios [79]. The combination of technological developments led to a new range of Raman spectroscopic microscopes in the 1990s, establishing Raman spectroscopic microscopy as a relatively inexpensive benchtop laboratory tool to complement conventional infrared spectroscopy.
Both IR and Raman spectroscopy entail the coupling of incident radiation with molecular vibrations and the resultant spectrum is characteristic of the compound or material. However, whereas IR spectroscopy involves the absorption of radiation, inducing transitions between vibrational states, Raman spectroscopy is a scattering technique (Figure 2.6), whereby the incident radiation couples with the vibrating polarisation of the molecule and thus generates or annihilates a vibration. The differing underlying mechanisms gives rise to a complementarity of the two techniques. For a vibration to be active in IR spectroscopy, a change in dipole is required, whereas to be Raman active, a change in polarizability is required. As a rule of thumb, vibrations of asymmetric polar bonds tend to be strong in IR spectra, whereas Raman is particularly suitable as a probe of symmetric, nonpolar groups. Notably, O-H vibrations of water are very strong in IR spectra, whereas they are extremely weak in Raman spectra, rendering Raman a potentially more suitable technique for biomedical applications, particularly in vivo.
A further implication of the differing physical origins of the techniques is that, whereas IR directly monitors the absorption of IR radiation, Raman scattering can be employed in the UV, visible or near-IR regions of the spectrum. Raman scattering thus offers intrinsically higher spatial resolution for mapping or profiling, the diffraction limit being determined by the wavelength (<1µm for Raman, ~5-10µm for IR). For many applications, however, near IR is favoured as a source for Raman analysis, to minimise interference from scattering, fluorescence, or photodegradation of the sample [80].

2.6 Experimental approaches

2.6.1 Fourier-Transform Infrared Spectroscopy

FTIR spectrometers have replaced traditional dispersive instruments, due to their superior speed and sensitivity [81]. They utilise a Michelson interferometer, which is a device that splits a single beam of IR light into two paths, and then recombines them after a variable path difference has been introduced (Figure 2.7) [82]. The interferometers are composed of a fixed mirror, a movable mirror and a beam splitter. The purpose of the beam splitter is to reflect some of the radiation toward the fixed mirror, meanwhile partially transmitting the rest to the adjustable mirror. When the waves return to the beam splitter, they interact and are then further reflected and transmitted. The split beams travel different pathlengths as a result of the moving mirror, and hence produce waves of different of intensity when recombined [83]. As a function of time, the light field varies spectrally, and can be converted to the frequency domain through Fourier transformation. The absorption, reflection or scattering of light by a sample can thus be recorded in the time domain by a single detector. FTIR absorption
spectroscopy monitors the vibrational bending and stretching modes of molecules that are active within the infrared region. The wavelengths at which they absorb the IR radiation are measured, and as every compound has a characteristic set of absorption bands, it results in a unique spectroscopic fingerprint [84].

Figure 2.7. The Michelson interferometer found in FTIR instruments. Red: incident beam, blue: reflected and purple: combined. Adapted from ref.[85]

There are three main sampling modes involved in FTIR spectroscopy; transmission, reflection and attenuated total reflection. By default, most FTIR instruments use transmission mode, in which traditionally IR light is irradiated through a sample on an IR transparent window, such as calcium fluoride, and is collected by a detector on the other side [86]. The
coupling of transmission mode to microscopy has allowed FTIR imaging to emerge in biomedical research [87], [88]. There are, however, a number of flaws related to transmission mode. Sample preparation can be exhaustive, and short pathlengths (<10µm) are required to prevent full absorption of the IR radiation by the sample before reaching the detector. This limiting factor also affects aqueous samples, since water is highly IR active [89]. Furthermore, IR transparent substrates that are required for this technique are fragile and often rather expensive to replace [90].

In transmission/reflection, or so called transflection mode, the incident IR beam initially travels through a sample, and is then reflected back off an IR reflecting substrate, and again passes through the sample toward the detector. It can be advantageous in comparison to transmission, in that the substrates are generally inexpensive low emissivity (low e) slides, and the approximate sample thickness can usually be smaller than that required for transmission measurements (1-4µm c.f. 2-8µm) which can be beneficial when sample quantity is limited. On the other hand, as the pathlength is effectively doubled there is also a maximum thickness limitation. Transflection mode may also be prone to standing wave artefacts that cause spectral variance, although the implications of this effect for diagnostic applications are still being assessed [86], [91], [92].

In assessing the suitability of the measurement mode for analysis of biological samples which are highly physically and chemically inhomogeneous, it is important to understand the physical processes involved. When a sample is measured in reflection, or transflection, a proportion of the light registered derives from the top surface, the reflectance of which is governed by the real component of the refractive index of the material. The transmitted light, measured in transmission or transfection, can be reduced by the intrinsic absorptions of the
constituent molecules, giving rise to the desired fingerprint of the sample, but can also be reduced by “Mie-like” scattering from structures (cells and cell nuclei) which have dimensions similar to the wavelength of light employed (5-20µm). This scattering is resonantly enhanced in the neighbourhood of an absorption, and can give rise to spectral “artefacts” in reflection, transfection and transmission modes [93], [94]. These resonant Mie effects can be ameliorated by application of specific pre-processing methods [95].

The development of attenuated total reflection-Fourier transform infrared (ATR-FTIR) spectroscopy has attracted wide interest in the field in recent years. The technique is unique in that the incident IR beam does not actually travel through the sample, but is directed through a substrate with a high refractive index, such as diamond, germanium or silicon, known as an internal reflection element (IRE). The sample must be placed in direct contact with the IRE, as when the incident radiation reflects off the internal surface of the IRE, an evanescent wave projects orthogonally into the sample, which then attenuates the IR beam before exiting the IRE to the detector [89]. The refractive index of the chosen IRE and the sample govern the basic ATR phenomenon, as shown in Eq.1, where the critical angle $\theta_c$, can be calculated from $n_1$ and $n_2$, which are the refractive indices of the IRE and sample, respectively.

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right)$$

**Equation 1**: Calculation of the critical angle $\theta_c$, where $n_1$ and $n_2$ are the refractive indices of the IRE and sample, respectively.

The IR radiation undergoes total internal reflection when the angle of incidence at the sample-crystal interface is greater than the critical angle, hence materials with a high
refractive index are commonly chosen to minimise the critical angle [96]. An important factor is the depth of penetration $d_p$ of the evanescent wave into the sample (Equation 2), as it determines how much of the sample is actually analysed [85]. The penetration depth is dependent upon the angle of incidence, refractive indices and the wavelength; at longer wavelengths, the evanescent wave will penetrate deeper into the sample.

$$d_p = \frac{\lambda}{2\pi n_1 \sqrt{\sin^2 \theta - (n_2/n_1)^2}}$$

**Equation 2:** The depth of penetration $d_p$ where $\theta$ is the angle of incidence and $\lambda$ is wavelength.

One limitation with this approach, is that the IRE must be kept clean to ensure there is no cross-contamination between samples, inhibiting throughput. Traditional IREs can also be expensive, for example a fixed diamond crystal, preventing multi-IRE systems. However, multi-IRE systems have been developed that will enable a high throughput analysis [97]. Scratches on the surface of the IRE are known to affect the sample-IRE contact, and loss of sensitivity is common due to the shallow penetration depths [82]. That said, ATR has become extremely popular in FTIR spectroscopy, as it has numerous advantages over the other IR techniques [85]. In contrast to transmission mode, for which the sample usually has to be pressed into a pellet or thin film, the ATR-FTIR mode negates the need for time consuming preparation, as the sample can be examined directly on to the IRE, in liquid or solid state. The shorter pathlength makes it more applicable for aqueous samples, as there is less IR radiation is lost through water absorbance, compared to transmission measurements [89]. Likewise, minimal scattering effects and relatively high signal-to-noise ratio (SNR) are valuable attributes [86]. Biological samples, such as human blood serum, are well suited to
ATR analysis as only small volumes of biofluid drops are required to dry efficiently onto the IRE. The size of the crystal governs the volume of sample that is required, and to ensure intimate contact occurs the sample should cover the whole IRE, allowing effective penetration of the evanescent wave [98].

2.6.2 Instrumentation for Raman spectroscopy

A Raman spectrometer typically consists of three major components: an excitation source, a sampling apparatus, and a detector. While these three components have evolved in varying forms over the years, modern Raman instrumentation has developed around using a laser as an excitation source, a spectrometer for the detector, and either a microscope or a fibre optic probe for the sampling apparatus. Figure 2.8 shows a schematic of a typical Raman setup. The laser sources provide a stable and intense beam of radiation. A wide range of lasers can be used as the light source, although, for biological applications, longer wavelength, near infrared sources are commonly employed, to minimise photodamage, scattering and/or fluorescence [99]. Band pass “interference” filters are employed to clean the laser spectrum and remove plasma lines. Dispersive instruments make use of a notch filter coupled with a high quality grating monochromator. Double, or triple, grating monochromators, rejection filters, super notch filters, holographic notch or edge filters and holographic filters are employed to separate relatively weak Raman lines from intense Rayleigh scattered radiations [100]. Charge transfer devices (CTDs) such as charge-coupled devices and charge-injection devices act as detectors and are used in the form of arrays. The role of the CTD arrays are to convert the incoming optical signal into charge which are then integrated and transferred to readout devices [101]. CTDs are commonly made of silicon so laser wavelengths of less than
1 μm can be detected, while laser wavelengths of greater than 1 μm use single element detectors based on a low band-gap semiconductor, such as Germanium (Ge) or Indium–Gallium–Arsenic (InGaAs). The grating is used to disperse the light and the groove density determines the spectral resolution. Other factors that play a key role in determining spectral resolution include the wavelength, shorter wavelengths having a higher spectral resolution, and the spectrometer length, which is the distance between grating and the detector, longer distances providing higher resolution. The objective lens both delivers the incident light and collects the scattered light. The objective lens delivers the incident light to the sample, and the Raman scattered light can be observed at any angle. In the commonly employed backscattering geometry, the Raman is collected by the objective lens and delivered to the grating. The spectrally dispersed, detected Raman scattered light is displayed as a Raman shift from the source wavelength, which in converted to units of wavenumber (cm\(^{-1}\)), such that Raman spectra can easily be compared and contrasted with equivalent FTIR spectra.
Figure 2.8. Typical Instrumentation for Raman microspectroscopy (CCD; charge coupled detector)

2.7 Biospectroscopy

FTIR spectroscopy has become an accepted tool in biophysics for analysis of the structure and interactions of proteins, lipids, carbohydrates and nucleic acids [102]-[105]. Applications to tissue samples for diagnostic applications were first reported in the early 90s, and since this time a range of pathologies has been investigated [106]-[108]. The application of Raman spectroscopy to biomolecules and even tissues was first demonstrated as early as the 1960s, and by the mid 1970s biomedical applications were explored [109]-[111]. Whole cell and
tissue studies have been carried out on a range of pathologies [112]-[114] and *in vivo* studies [115], [116] have demonstrated the prospective for diagnostic applications. The potential of vibrational spectroscopy in conjunction with multivariate analysis techniques as a diagnostic tool has thus been well demonstrated and the concept of spectral cytopathology has been coined [117]. In this respect, Raman and infrared can be viewed as rival technologies, but to best advance the understanding of the potential of the techniques a combination of the two complementary techniques is recommended.

Lipids, proteins, nucleic acids and carbohydrates are the four biomolecular groups characteristically found in a biological spectrum, as measured using either FTIR or Raman spectroscopy. Figure 2.9 shows, for example, the (A) FTIR and (B) Raman microscopic spectra of human blood serum, where the protein-related bands are highlighted, while Tables 2.5 and 2.6 shows typical band assignments across the full spectra. In the "high wavenumber region", >2500cm\(^{-1}\), of the FTIR spectrum, the distinctive vibrations of N-H, C-H and O-H of lipids and proteins can be found, whereas in the "fingerprint region", <1800cm\(^{-1}\), the features are typically more complex combinations, including the Amide I (1650cm\(^{-1}\)) and Amide II (1520cm\(^{-1}\)) modes of proteins, nucleic acid phosphate stretching modes at 1070cm\(^{-1}\) and 1250cm\(^{-1}\) and lipidic derived features at 1310cm\(^{-1}\) and 1750cm\(^{-1}\). It should be noted that, although complementary techniques, the features in the respective spectra of FTIR and Raman have similar origin. Thus, the Raman spectrum of the nucleus exhibits similarly prominent signatures associated with proteins and lipids across the fingerprint region, as well as large peaks related to DNA and RNA at 785cm\(^{-1}\).
A notable difference between the two spectra is the spectral range is presented (Figure 2.9). FTIR covers the full spectral range in a single scan of the interferometer, which is recorded by a single detector. In contrast, Raman spectroscopy is currently predominantly performed by dispersive techniques, by which a spectral range is dispersed onto a CCD pixel array. The range covered depends on the dispersion element (grating) and use of higher resolution gratings means that multiple windows are required to cover the whole spectral range. Hence, Raman spectra are frequently presented as either the fingerprint or high wavenumber region. On the other hand, either Raman or FTIR can be limited in the lower wavenumber limit by instrument detectors, optical elements and or substrate used. Typically, Raman spectra are recorded as low as 400cm$^{-1}$, FTIR only as low as 1000cm$^{-2}$ or 600cm$^{-1}$. 

Figure 2.9. Typical IR (A) and Raman (B) spectrum of human blood serum
<table>
<thead>
<tr>
<th>Approximate Wavenumbers (cm⁻¹)</th>
<th>Vibration</th>
<th>Biochemical Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3300 (i)</td>
<td>ν(N-H)</td>
<td>Amide A of proteins/peptides</td>
</tr>
<tr>
<td>3100 (ii)</td>
<td>ν(N-H)</td>
<td>Amide B of proteins/peptides</td>
</tr>
<tr>
<td>2957</td>
<td>νas(CH₃)</td>
<td>Lipids</td>
</tr>
<tr>
<td>2920</td>
<td>νas(CH₂)</td>
<td></td>
</tr>
<tr>
<td>2872</td>
<td>νs(CH₂)</td>
<td></td>
</tr>
<tr>
<td>2850</td>
<td>νs(CH₂)</td>
<td></td>
</tr>
<tr>
<td>1740</td>
<td>ν(C=O)</td>
<td>Phospholipid esters</td>
</tr>
<tr>
<td>1715-1680</td>
<td>ν(C=O)</td>
<td>Nucleic acids</td>
</tr>
<tr>
<td>1650 (iii)</td>
<td>ν(C-N), δ(N-H)</td>
<td>Amide I of proteins</td>
</tr>
<tr>
<td>1645</td>
<td>γ(HOH)</td>
<td>Water</td>
</tr>
<tr>
<td>1550 (iv)</td>
<td>~60% δ(N-H), ν(C-N), δ(C-O), ν(C-C)</td>
<td>Amide II of proteins</td>
</tr>
<tr>
<td>1453</td>
<td>γ(CH₂)</td>
<td>CH₂ Scissoring</td>
</tr>
<tr>
<td>1450 (v)</td>
<td>δas(CH₃)</td>
<td>Lipid/Proteins</td>
</tr>
<tr>
<td>1395 (vi)</td>
<td>δ(CH₃)</td>
<td>Lipid/Proteins</td>
</tr>
<tr>
<td>1395</td>
<td>ν(C=O)</td>
<td>Carboxylate COO⁻</td>
</tr>
<tr>
<td>1380</td>
<td>γs(CH₃)</td>
<td>Phospholipid/triglyceride</td>
</tr>
<tr>
<td>1350-1250 (vii)</td>
<td>δ(N-H), ν(C-N), γ(C=O), ν(C-C)</td>
<td>Amide III – peptide/protein/collagen</td>
</tr>
<tr>
<td>1242</td>
<td>νas(PO²⁻)</td>
<td>DNA/RNA/phospholipid</td>
</tr>
<tr>
<td>1170</td>
<td>νas(C-O)</td>
<td>Ester</td>
</tr>
<tr>
<td>1150</td>
<td>ν(C-O), γ(COH)</td>
<td>Carbohydrates</td>
</tr>
<tr>
<td>1090</td>
<td>νs(PO²⁻)</td>
<td>DNA/RNA/phospholipid</td>
</tr>
<tr>
<td>1086</td>
<td>ν(C-O), ν(C-C), def(CHO)</td>
<td>Carbohydrates</td>
</tr>
<tr>
<td>1079</td>
<td>ν(C-C)</td>
<td>Glycogen</td>
</tr>
<tr>
<td>1065</td>
<td>ν(C-O)</td>
<td>DNA and RNA ribose</td>
</tr>
<tr>
<td>1050</td>
<td>ν(C-O)</td>
<td>Phosphate ester</td>
</tr>
<tr>
<td>1028</td>
<td>def(CHO)</td>
<td>Glycogen</td>
</tr>
<tr>
<td>965</td>
<td>ν(PO²⁻)</td>
<td>DNA and RNA Ribose</td>
</tr>
<tr>
<td>710-620</td>
<td>def(O=C-N)</td>
<td>Amide IV</td>
</tr>
</tbody>
</table>

ν = stretching; δ = bending; γ = wagging, twisting and rocking; def = deformation; as = asymmetric; s = symmetric
Table 2.6. Tentative peak assignments for Raman spectral data, (i)-(viii) corresponds to Figure 2.9B [122], [123].

<table>
<thead>
<tr>
<th>Approximate Wavenumbers (cm(^{-1}))</th>
<th>Vibration</th>
<th>Biochemical Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>785-788</td>
<td>(v_\text{s} (PO^2^-))</td>
<td>Nucleic Acid</td>
</tr>
<tr>
<td>1004(i)</td>
<td>(v\text{(ring breathing)}) Phenylalanine</td>
<td>Protein</td>
</tr>
<tr>
<td>1090(ii)</td>
<td>(v_\text{s} (PO^2^-))</td>
<td>Nucleic Acid</td>
</tr>
<tr>
<td>1127(iii)</td>
<td>(v(C-N))</td>
<td>Protein</td>
</tr>
<tr>
<td>1262</td>
<td>(v(C-C))</td>
<td>Lipid</td>
</tr>
<tr>
<td>1262</td>
<td>(v_\text{s} (PO^2^-))</td>
<td>Nucleic Acid</td>
</tr>
<tr>
<td>1319(iv)</td>
<td>(\gamma(CH_2))</td>
<td>Lipid/Nucleic Acid</td>
</tr>
<tr>
<td>1341(iv)</td>
<td>def(CHO)</td>
<td>Protein</td>
</tr>
<tr>
<td>1341(iv)</td>
<td>(v_\text{s} (PO^2^-))</td>
<td>Nucleic Acid</td>
</tr>
<tr>
<td>1451(v)</td>
<td>(\delta(CH_2))</td>
<td>Protein/Lipid</td>
</tr>
<tr>
<td>1554 (vi)</td>
<td>(\delta(N-H), v(C-N))</td>
<td>Amide II of Proteins</td>
</tr>
<tr>
<td>1619(vii)</td>
<td>(v(C=C))</td>
<td>Protein</td>
</tr>
<tr>
<td>1662(viii)</td>
<td>(v(C=O))</td>
<td>Amide I of Proteins</td>
</tr>
<tr>
<td></td>
<td>(v(C=C))</td>
<td>Lipid</td>
</tr>
</tbody>
</table>

\(v = \text{stretching}; \delta = \text{bending}; \gamma = \text{twisting}; \text{def} = \text{deformation}; s = \text{symmetric}\)

2.8 Vibrational Spectroscopy of Protein

For the investigation of proteins, vibrational spectroscopy is particularly useful, as protein related bands are dominant within biological spectra. Stretching vibrations are found in the higher-wavenumber region (3500-2500 cm\(^{-1}\)), such as C-H, N-H and O-H stretches, whereas bending and carbon skeleton fingerprint vibrations tend to occur in the lower-wavenumber regions. The most important spectral region in relation to biological materials is the information-rich fingerprint region (1800-400 cm\(^{-1}\)), wherein the Amide I and II peaks exist (1700-1500 cm\(^{-1}\)) [86].
The wealth of information that exists in a vibrational spectrum of a biological sample, detailed in Tables 2.5 and 2.6, renders the techniques as interesting tools for investigating molecular systems ranging from amino acids, peptides and protein complexes [124]-[127]. Vibrational spectroscopy can enhance the understanding of protein function, as it is sensitive to changes to the protonation state of amino acid side chains and the strength of hydrogen bonding between amide bonds [128], [129]. In both IR and Raman spectra, most characteristic bands are associated with the CONH group, referred to as Amide A (NH stretching, ~3300 cm\(^{-1}\)), Amide B (NH stretching, ~3100 cm\(^{-1}\)) and Amide I to VII (I: 1600–1700 cm\(^{-1}\), II: 1480–1580 cm\(^{-1}\), III: 1230–1300 cm\(^{-1}\), IV: 625–770 cm\(^{-1}\), V: 640–800 cm\(^{-1}\), VI: 540–600 cm\(^{-1}\), VII: 200 cm\(^{-1}\)) (Figure 2.10) [118], [130]. The Amide A band (~3300 cm\(^{-1}\)) originates from the NH stretching vibration, which are often present as a resonance doublet with the weakly absorbing Amide B (~3170 cm\(^{-1}\)), arising from a Fermi resonance between the first overtone of Amide II [118]. The Amide I, which absorbs near 1650 cm\(^{-1}\), is primarily caused by the C=O stretching vibrations, with smaller contributions from CN stretching, deformation of CCN and NH in plane bending vibrations. The out-of-phase combination of the NH bending and the CN stretching vibrations, as well as minor contributions from the CO in-plane bend and the CC and NC stretching vibrations, give rise to the Amide II band at ~1550 cm\(^{-1}\) [131].
Figure 2.10. Molecular vibrations of the amide group - Orange: Hydrogen, Red: Nitrogen, Purple: Carbon, Blue: Oxygen. Adapted from ref [132]

Similar analyses can be performed using IR and/or Raman spectroscopy, and even Raman Optical Activity (ROA), which is particularly sensitive to molecular chirality [133]. Although infrared absorption and Raman scattering spectroscopy probe the same physical phenomenon of molecular vibrations, the spectral profile is discernibly richer in substructure in the case of Raman, for similar instrumentational specifications in terms of spectral and spatial resolution [134]. Infrared absorption involves an electric dipole transition between two vibrational states, each of which has its own homogeneously and inhomogeneously
broadened line width. The resultant spectral bandwidth is a convolution of these two individual line widths. Although often represented as a transition between a real vibrational level of the manifold of an electronic state and a virtual electronic level, Raman is a scattering process. In the representation of the transition to a virtual state, the bandwidth of that state is infinitesimally small, and so the scattering line width is intrinsically less than that of an equivalent infrared absorption transition, giving rise to more distinct spectral features.

A characteristic feature of Raman spectra of many proteins, not observed in IR spectra, is the strong and often dominant feature at $\sim 1004 \text{ cm}^{-1}$, ascribed primarily to the ring breathing mode of the phenylalanine residue. Note, however, that its prominence in the spectrum does not reflect a similar relative prominence over other residues in the protein structure, but rather the large Raman scattering cross section of the highly polarisable, $\pi$-conjugated, ring structure. Similarly, the porphyrin moieties of cytochrome c contributes strongly to the Raman spectrum at $\sim 1585 \text{ cm}^{-1}$, and can be resonantly enhanced at source wavelengths of $<550\text{nm}$ (Figure 2.11) [135].

![Figure 2.11. Raman spectra of (A) Albumin, (B) Fibrinogen and (C) Cytochrome c.](image)
Many studies have looked at the potential of vibrational spectroscopy to predict protein secondary structure, as the Amide I band is highly sensitive to hydrogen bonding pattern, dipole-dipole interaction and the geometry of the polypeptide backbone [136]. A series of overlapping components that represent different structural elements, such as α-helices and β-sheets, are present in the broad Amide I band [89], [137].

Rygula et al. have reviewed the analysis of the secondary structure of proteins using Raman spectroscopy of 26 different proteins [130]. Raman spectroscopy for analysis of protein secondary has focused largely on the correlation of the position of the amide I and amide III vibrations with the crystallographically determined fraction of each secondary structural element present in the protein. Associated with the amide I and amide III modes, wavenumbers assignments for features associated with α-helix and β-sheet structures include: 1662–1655 and 1272–1264 cm\(^{-1}\) (a), 1674–1672 and 1242–1227 cm\(^{-1}\) (b), respectively, and the review also classified structures as “mixed structures (α/β, α + β, s)” and “others” [138]. The bands relating to protein secondary structure of the Amide I band are summarised in Table 2.7.

Particularly for the case of IR studies, protein secondary structures have been studied experimentally using both H\(_2\)O and D\(_2\)O. This is mainly due to the H\(_2\)O absorbance overlapping with the Amide I band, but it is also thought to be easier to obtain spectra in D\(_2\)O as the bands occur at lower wavenumbers than H\(_2\)O, meaning the region between 1400-1800cm\(^{-1}\) exhibits relatively low absorbance (or scattering), providing an ideal window to observe the weaker bands of solubilised protein [136]. That said, using H\(_2\)O as a solvent is still preferable to D\(_2\)O when looking at protein structure, as D\(_2\)O can slightly alter the
flexibility of proteins; for example, D_2O has been shown to increase the rigidity of most protein structures [139].

Table 2.7. Assignment of Amide I band positions to secondary structure in H_2O for both IR and Raman [128], [130], [136], [137] and [140]

<table>
<thead>
<tr>
<th>Secondary Structure</th>
<th>IR band Position (cm(^{-1}))</th>
<th>Raman band Position (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)-Helix</td>
<td>1648-1657</td>
<td>1650–1659</td>
</tr>
<tr>
<td>(\beta)-Sheets</td>
<td>1623-1641 &amp; 1674-1695</td>
<td>1669-1674</td>
</tr>
<tr>
<td><strong>Turns</strong></td>
<td>1662-1686</td>
<td>1680-1690 &amp; 1653-1656</td>
</tr>
<tr>
<td><strong>Disordered/Random Coil</strong></td>
<td>1642-1657</td>
<td>1640-1651</td>
</tr>
</tbody>
</table>

Fourier self-deconvolution (FSD) and derivative filtering are commonly employed methods for the investigation of protein secondary structure. FSD mathematically reduces bandwidths, so that the overlapping bands can be resolved [141]. This can also be achieved by differentiating the spectrum, commonly by calculating the second derivative, which exhibits a negative peak for every band or shoulder in the spectrum. For quantification, both FSD and second derivative spectra require curve-fitting (Figure 2.12), and the fractional areas of the fitted components correspond to the relative quantities of the different types of secondary structure [136]. Thus, the band areas are directly proportional to the relative amount of secondary structure that is represented in that spectral region. The quantity of each component is expressed as a percentage, which provides a clear picture of overall protein structure. For example, second derivative analysis by Dong et al. predicted that Immunoglobulin G contains mostly \(\beta\)-sheets (64%) and turns (28%), with few random coils.
(5%) and α-helix (3%), a finding which was found to be closely correlated to the values obtained through X-ray crystallography [142].

**Figure 2.12.** Curve fitting of the Amide I band in serum, for IR (A) and Raman (B) Spectra

Moreover, the estimation of amino acid side-chain absorption must be considered in the analysis of protein spectra. Amino acid residues - arginine, asparagine, glutamine etc. also absorb in the Amide I/II spectral region (Table 2.8). In the IR spectra of some globular proteins, the contribution of side-chains can be as high as 10-30% of the overall absorbance [137]. This can be a potential difficulty as the contribution from the amino acids will depend on their protonation state, which can be challenging to evaluate. The quantitative estimation of these groups can allow more refined predictions of the secondary structure of proteins and polypeptides by FTIR [136], [143].
Table 2.8. Summary of main amino acid side chain absorptions found in IR spectra between the 1400-1800 cm\(^{-1}\) region [128], [136], [144]

<table>
<thead>
<tr>
<th>Side chain assignments</th>
<th>Approximate band position in IR (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp, (\nu(C=O))</td>
<td>1716</td>
</tr>
<tr>
<td>Glu, (\nu(C=O))</td>
<td>1712</td>
</tr>
<tr>
<td>Asn, (\nu(C=O))</td>
<td>1678</td>
</tr>
<tr>
<td>Arg, (\nu_{as}(CN_3H_5^+)))</td>
<td>1673</td>
</tr>
<tr>
<td>Gln, (\nu(C=O))</td>
<td>1670</td>
</tr>
<tr>
<td>Arg, (\nu_s(CN_3H_5^+)))</td>
<td>1633</td>
</tr>
<tr>
<td>Lys, (\delta_{as}(NH_3^+)))</td>
<td>1629</td>
</tr>
<tr>
<td>Asn, (\delta(NH_2))</td>
<td>1622</td>
</tr>
<tr>
<td>Gln, (\delta(NH_2))</td>
<td>1610</td>
</tr>
<tr>
<td>Tyr, Ring-(O^-)</td>
<td>1602</td>
</tr>
<tr>
<td>Asp, (\nu_{as}(COO^-)))</td>
<td>1574</td>
</tr>
<tr>
<td>Glu, (\nu_{as}(COO^-)))</td>
<td>1560</td>
</tr>
<tr>
<td>Lys, (\delta(NH_3^+)))</td>
<td>1526</td>
</tr>
<tr>
<td>Tyr, Ring-(OH)</td>
<td>1518</td>
</tr>
<tr>
<td>Phe, Ring</td>
<td>1494</td>
</tr>
</tbody>
</table>

\(\nu = \text{stretching}; \delta = \text{bending}; \text{as} = \text{asymmetric}; \text{s} = \text{symmetric}\)

Rygula et al. also identify features of the Raman spectra that permit the description of the environment of numerous amino acid chains. These include the amino acid side-chain modes (e.g. tryptophan doublet 1360/1340 cm\(^{-1}\), tyrosine doublet 860/833 cm\(^{-1}\)) or the sulphur-containing residues in the different physical states (C–S stretching with H at the trans position of the S atom: 640–680 cm\(^{-1}\) and C–S stretching with C at the trans position of the S atom;
Over the past decade, there has been a rapid increase of proof of concept publications for spectroscopic disease diagnostics, highlighting its potential for progression into the clinical environment. The majority of the publications in the biomedical vibrational spectroscopy field have been based on the analysis of human tissue, with pilot studies showing it is possible to differentiate between healthy and cancerous tissue, as well as benign and malignant tumours [84]. Malignancies from various organs, such as breast, lung, colon and prostate tissues, have previously been studied which has provided a platform of promising results [145]-[147], [119], [106]. Despite the high volume of published research, the technique has yet to make a successful transition into the clinic [70].

More recently, there has been further interest in biofluid spectroscopy due to the ease of collection and handling, and minimal sample preparation is required. Blood components such as serum and plasma are commonly analysed for clinical reasons, carrying information regarding intra- and extra-cellular events. Biobanks exist as a valuable stock both serum and plasma, with the ability to repeat analysis or monitor treatment or disease progression [69]. More specifically, blood serum is the most complex biofluid, containing over 20,000 different proteins. As it perfuses all body organs, it gains proteomes from surrounding tissues and cells [148]. The low molecular weight fraction of serum - the peptidome - is information rich, therefore the spectroscopic biosignature of serum is ideal for detecting disease states [149].
ATR-FTIR spectroscopy has been proven to be a promising screening tool for detecting ovarian cancer from human blood, where both serum and plasma were used to discriminate ovarian cancer patients from healthy controls with a success rate of ~95% and ~97%, respectively [150]. Backhaus et al. used serum spectroscopy to differentiate between patients in good health and those with breast cancer, by applying unsupervised and supervised methods, reporting sensitivities and specificities of >92% for both [151]. Another pilot study found that the serum biosignature for cirrhotic patients, with and without hepatocellular carcinoma (HCC), could be successfully separated using support vector machine (SVM) classification and leave-one-out cross validation [152]. Furthermore, patients with extensive fibrosis in the liver have been separated from those without fibrosis, which is a common disorder in the early developmental stages of HCC, by using their FTIR serum spectra. Ollesch et al. introduced automated sampling for the first time, robotically spotting serum for high throughput FTIR measurements, in their quest to identify and validate spectroscopic biomarker candidates for urinary bladder cancer [153]. By using as little as 1μl of blood serum for ATR-FTIR analysis, Hands et al. were not only able to distinguish between serum of brain tumour patients and controls, but could effectively predict tumour grade by separating low grade lesions from patients with glioblastoma (high grade), highlighting the great potential of ATR-FTIR spectroscopy of blood serum for determining the severity of brain tumours [154]. Paraskevaidi et al. demonstrated that ATR combined with chemometrics was capable of differentiating patients with various neurodegenerative diseases. Alzheimer’s disease (AD) was identified with a sensitivity and specificity of 70%, and the AD patients were further segregated from those with dementia with 90% accuracy [155].
In order to maximise classification accuracy, feature extraction techniques can be utilised to pick out the most salient properties of the IR spectra. These methods isolates the features that are most highly correlated with a target set and ranks them based on similarity, which in turn allows discrimination between classes and maximizes the intergroup differences [156]. In a further study, Hands et al. used feature extraction to select the most discriminatory spectral regions in their dataset [157]. Blood serum from cohort of 433 patients, with and without a brain tumour, were collected for ATR-FTIR analysis. This substantial dataset consisted of control samples (non-cancer) and various brain tumour types; high- and low-grade glioma, meningioma and metastatic tumours. The variable ranking function highlighted the wavenumber variables that were most salient between each spectral class. This technique proved to be effective in exposing the changes in the spectral signature between classes and tumour grade, and was capable of differentiating between: cancer and non-cancer; glioma and meningioma; metastatic and brain cancer; and high- and low-grade glioma, based on the most discriminatory spectral regions; Amide I, Amide II, C-O stretch of lipids/proteins, CH$_2$ of lipids/proteins and PO$_2^-$ of DNA/RNA. Furthermore, vibrations of C-O, C=O and C-H of lipids/proteins, as well as the PO$_2^-$ stretch from nucleic material, enabled discrimination between the three organs of origin of the metastatic cancer samples (lung, melanoma and breast). Previous studies have also highlighted these spectral regions when analysing brain cancer and metastatic states via tissue spectroscopy [158], [159].

To further explore the dataset gathered by Hands et al., machine learning techniques were employed in a computational study by Smith et al. [160]. The cancer versus non-cancer spectra were classified by RF, which utilised a Gini impurity metric to elucidate the most important wavenumber regions for the classification. The carbohydrate region (997-1003 cm$^{-1}$) was
found to be of highest RF importance, followed by the phosphate (1290-1294cm\(^{-1}\)), lipid CH\(_2\) (1462-1464cm\(^{-1}\)), Amide II (1527-1533cm\(^{-1}\)), carbohydrate (1028-1034cm\(^{-1}\)) and protein COO\(^-\) (1387-1390cm\(^{-1}\)) regions. 2D correlation analysis was also performed alongside RF. The features highlighted by the 2D correlations were highly comparable to the RF results, further clarifying the main spectral differences. The combination of these machine learning methods permitted successful discrimination of cancer and non-cancer, with sensitivities and specificities of 92.8% and 91.5% respectively, verifying the plausible use of orthogonal techniques to examine salient information for more accurate and rapid diagnostics [160].

Although ATR is preferred for biofluids analysis, the strong absorbance of water is still evident in the fingerprint region, which can obscure protein absorbance in liquid samples. Hence, the analysis of biofluids has been predominantly performed on air dried samples, which can lead to chemical and physical inhomogeneity. The complex patterns that arise from dried biofluid drops have been of great interest over the past few decades, and various models have been published in an attempt to explain the complicated drying behaviour [161]. In analysing such dried droplets using single-point transmission FTIR, Hughes et al. found the absorbance of the Amide I/II region to be highly variable across a minute drop of blood serum. Spectra obtained from random locations across the dried serum spot, showed evidence of differences in sample thickness and heterogeneity. IR transmission imaging verified that there were biochemical differences across the drop [162]. Furthermore, the loss of light due to scattering, caused by the presence of cracks throughout the sample, along with varied drop thickness led to the conclusion that samples need to be smooth and evenly spread for transmission measurements [163]. Deegan et al. proposed the coffee ring effect, whereby capillary flow forces biomolecules to move out towards a drop’s edge leaving behind dense
ring at the periphery [164]. This is common when drying biofluid drops, which is a concern for spectroscopists, as the centre of the drop may not be representative of the whole sample. Specifically in blood serum, the process is known as the Vroman effect, whereby a series of molecular displacements arise through protein exchange [165]. When a biological liquid is applied to a solid surface, low molecular weight proteins attach to the surface first, before being displaced by larger protein molecules over time. Therefore, the adsorption of proteins on to the substrate surface will be based on their differing affinities [166]-[168]. Gelation and cracking patterns have also been observed in dried biofluid drops, which are thought to be dependent on protein concentration [88], [169]. These are the main limiting factors of using dried biofluids, as the surface inhomogeneity can cause peak shifts and alterations in band intensities [162]. Environmental (temperature and humidity) and experimental (volume and concentration) conditions have been shown to affect the drying patterns, hence it is vital the drying conditions are controlled and optimal protocols are developed in order to obtain a more homogenous deposition across the sample [161], [170]. When measured in the ATR mode, however, the sample deposit can be completely contained within the area of the crystal, such that the evanescent wave measures the average of the entire drop, averaging out any inhomogeneities.

Raman spectroscopy is a complementary tool to IR spectroscopy and is compatible with aqueous samples. This technique allows the analysis to be carried out in the native state of bodily fluids, and, therefore, the additional drying step can be eliminated. A comprehensive proof of concept has been designed and conducted to detect hepatocellular carcinoma (HCC) from patient serum using micro-Raman spectroscopy using dried and freeze-dried serum drops [171]. The aim of the study was to discriminate serum samples of patients with HCC.
and patients without HCC. Two groups of patients were classified with an overall accuracy of 84.5% to 90.2% for dried serum drops and 86% to 91.5% for freeze-dried serum. Although not specifically protein based, Mahmood et al. demonstrated the ability of Raman microspectroscopic analysis of dried blood serum to identify the presence of dengue infection, and to correlate the spectroscopic response with viral load [172]. A surface-enhanced Raman scattering (SERS) based immunoassay has been developed to monitor levels of the mucin protein MUC4 in patient serum, which will help in the early detection and diagnosis of pancreatic cancer [173].

Raman and surface-enhanced Raman spectroscopy (SERS) using silver nanoparticles, have been employed to identify signatures linked to ovarian cancer in blood plasma [174]. Both techniques provided satisfactory diagnostic accuracy for the detection of ovarian cancer, Raman achieving 94% sensitivity and 96% specificity, and SERS 87% sensitivity and 89% specificity. For early ovarian cancer, Raman achieved sensitivity and specificity of 93% and 97% respectively, while SERS had 80% sensitivity and 94% specificity.

As Raman microspectroscopy, by default, measures a very limited area of the sample, defined by the spot size of the objective focus, its suitability for analysis of chemically and physically inhomogeneously dried biofluid droplets is limited. However, an improved protocol for Raman spectroscopic analysis coupled with fractionation of serum using centrifugal filters to concentrate and separate low molecular weight proteins has been demonstrated [175]. FTIR spectra were recorded in aqueous solutions of gelatin at concentrations as low as 100 mg/L, using Raman spectroscopy, high quality spectra of gelatin solutions as low as 10 mg/L was achieved. Spectral features of human serum were found to be weak and partially obscured by water features. Dried deposits were shown to be physically and chemically inhomogeneous,
resulting in unreliable results. Concentration of the serum using commercially available centrifugal filter devices resulted in enhanced spectral intensity and quality and a hundred per cent recovery of the analytes. Improved analysis of serum using Raman spectroscopy was reported when the sample was analysed in the inverted geometry using the water immersion objective with a 785nm laser as source (Figure 2.13). A drop of water is used to minimise the differences in the refractive indexes between sample, objective and the substrate. However, the water drop does not contribute to the data collected, as it is outside the focus of the beam.

**Figure 2.13.** The inverted geometry used to analyse the serum focused by immersion objective

Using this analytical set up, Parachalil et al. recently reported on a systematic investigation of sample preparation considerations and data processing, for the analysis of blood plasma and serum [176]. In a solution of proteins, mixed in physiologically relevant concentrations, it was clearly seen that the poorly water soluble fibrinogen presented significant challenges to measurement in the liquid phase, as it caused extensive Mie scattering of the source laser as well as the Raman scattering from itself, and the other protein constituents, preventing any
quantitative measurement. This is a strong indication that the analysis of blood serum rather than plasma is favourable for optical based techniques. In cases where the determination of fibrinogen levels is desired, the study showed that fibrinogen aggregates were broken down through mild sonication of the mixture, which significantly reduced the scatter.

2.8.2 Quantitative analysis

The ability of vibrational spectroscopy to quantify levels of biomolecules in blood serum has enabled various clinical studies over the past few decades [177]. In recent years, the technique has been introduced for potential disease screening and monitoring for numerous health conditions, including arthritis, diabetes, heart disease and a variety of cancers [178]-[180]. The development of a robust spectroscopic protocol could enable the replacement of the quantification methods currently used in medical practice.

An early study reported the quantitative analysis of dried serum films provided high accuracy for total protein, triglycerides, cholesterol, urea and glucose, although the protocol was found to be less suitable for creatinine and uric acid [177]. Another demonstrated that both low density lipoprotein and high density lipoprotein cholesterol can be independently quantified using IR spectroscopy of dried serum films [181]. A later study showed patients with multiple myeloma are distinguishable from healthy patients through FTIR analysis of serum immunoglobulins, where the band intensity ratios in the spectral profile of the myeloma patients were higher than those with normal immunoglobulin levels [182].

Perez-Guaita et al. published an extensive quantitative study, using ATR-FTIR and partial least squares regression (PLSR) to determine the concentration of various proteins in human serum [183]. They established models for albumin and immunoglobulin, as well as total
globulin and the albumin/globulin ratio, which are routinely determined in clinical practice as they can be indicative of diseased states [184]. As the selected proteins exhibit different secondary structures, their analysis was focused on the behaviour of the Amide I/II bands. Their findings suggest that this technique could be useful in clinical practice as a routine assay for protein determination, as their prediction capability was extremely high for both albumin and total globulin, the root mean square error of prediction (RMSE) being lower than 5% for both. The RMSE for immunoglobulin and the albumin/globulin coefficient were slightly higher, between 7-14%, but the authors highlighted the ATR-FTIR method, coupled with PLSR, provides a promising quantification tool, especially at screening level [183].

Similar quantitative models using ATR-FTIR were constructed using PLSR in a recent study by Spalding et al., whereby human pooled serum was spiked with commercial human serum albumin (HSA) and immunoglobulin G (IgG). The spiked serum was analysed in both liquid and dried state to determine the optimal experimental protocol. In this particular study, in order to maintain reproducible spotting of the sample, the 10% air dried sample preparation was deemed to be the preferred approach. Using the same method, they analysed serum gathered from 20 patients, to test the predictive power of the PLSR model when looking at more complex samples. The model was tested by two validation method; leave one patient out cross validation (LOPOCV) and k-fold cross validation. For the prediction of total protein concentration, both models produced excellent results; the k-fold cross validation produced an RMSE of 1.986 ± 0.778mg/mL and an R^2 value of 0.934, whereas there was an RMSE of 1.534 ± 1.14mg/mL and an R^2 value of 0.926 for the LOPOCV model. Both blind testing methods produced similar trends, with the prediction of the individual HSA and IgG concentrations not as effective as the total protein. The prediction of IgG concentration was
inferior to that of HSA, which Spalding et al. suggested could be due to the inability of FTIR to distinguish between the variable contributions of five major types of immunoglobulins that are present in human serum (IgA, IgD, IgE, IgG and IgM) [185].

In comparison, Raman spectroscopy has been used for the quantitative analysis of blood to a lesser extent. Berger et al. used Raman microscopic analysis of liquid whole human blood and serum samples to quantify the content of six analytes, namely glucose, cholesterol, triglyceride, urea, total protein and albumin [186]. Rohleder et al. undertook a comparison of mid-infrared and Raman spectroscopy for the quantitative analysis of serum, based on mid-infrared and Raman spectra of the sera obtained from 247 blood donors [187]. PLSR analysis was used for the quantification of total protein, cholesterol, high- and low-density lipoproteins, triglycerides, glucose, urea and uric acid. IR measurements were performed on dried samples, whereas Raman was performed on liquids. For all analytes, comparable RMSEC (calibration) and RMSEV (validation) were achieved.

Centrifugal filtration devices have been utilised to improve the sensitivity of quantitative analysis by both Raman and IR spectroscopy, by separating the molecules according to their molecular weight. The proteins that are highly abundant in serum dominate the spectral profile, and through the removal of these proteins (albumin and globulins) the ability to monitor changes in the lower molecular weight fraction (LMWF) is enhanced. Note, however, the importance of following a strict rinsing protocol to remove residual glycerine has been highlighted [67]. The centrifugal fractionation of human serum using both ATR-FTIR and Raman spectroscopy was evaluated by Bonnier et al. [188]. In this study, whole human serum was spiked with a wide range of known concentrations of glycine, between 0.5 mg/mL and 50 mg/mL, in order to examine the capabilities of both ATR-FTIR and Raman
spectroscopy in human serum monitoring. As glycine has a molecular weight of 75Da, existing as the smallest amino acid, it can freely diffuse through the 10kDa centrifugal filter membrane, making it an ideal target for quantitative analysis of the filtrate. Small aliquots (0.5μl) of each concentration was measured, in both liquid and dry state, after air drying for 10 minutes. Bonnier et al. used principal component analysis (PCA) to explore and quantify the spectral variability caused by the adulteration of glycine to the human serum. Interestingly, the liquid samples produced a linear model (R² = 0.9993), whereas the results from the dried drops deviated from linearity above 10 mg/mL, and the relationship between the glycine concentration and spectral variations were expressed by a polynomial expansion (R² = 0.9978). In order to test the predictive power of both regression models, the lowest concentration, 0.5 mg/mL, was used as a blinded sample. The average predicted values for the liquid and dry models were 0.45 mg/mL ± 0.16 mg/mL and 0.383 mg/mL ± 0.007 mg/mL, respectively. Using the same protocol, more clinically relevant concentrations of glycine (0.01-2.5 mg/mL) were added to a centrifugal filtered LMWF stock solution for further analysis. Similar to the model for whole serum in the dried state, the LMWF model followed a polynomial fit achieving an R² value of 0.9981. In this case, the depletion of the abundant HMW proteins greatly enhanced the sensitivity of the regression model for the detection of glycine, delivering a predictive value of 0.011 mg/mL ± 0.006 mg/mL for the 0.01 mg/mL serum sample [188]. In the same study, it was demonstrated that the concentrate of the centrifugal process has been considerably concentrated (by factors up to x10 depending on the filter pore size), greatly enhancing the signal to (water) background levels for liquid phase Raman analysis. This indicates the potential for the prediction of other biomolecules that
exist within the LMWF with this method, and with further research, such techniques could be translated into the clinical environment as a rapid tool for screening and monitoring.

### 2.9 Clinical Translation

The high volume of research in the field of biomedical vibrational spectroscopy has indicated the potential utility of vibrational spectroscopy in a clinical environment. Numerous diagnostic and disease monitoring studies have reported extremely promising results, in some cases achieving sensitivity and specificity values greater than 90%. However, the techniques have not yet been successfully translated into the clinic [65]. The major hurdle to successful translation is arguably acceptance by health technology regulatory agencies, who determine which technologies are made available for public health. Criteria for successful acceptance require statistically verified clinical trials to prove clinical utility, but also clear understanding of the current clinical pathway in order to determine the economic and clinical impact of new technologies.

While promising proof-of-principal studies have supported clinical suitability, there have been few reported clinical trials employing either IR or Raman spectroscopy to confirm the utility in a prospective patient population. One of the examples closest to translation is the analysis of whole blood for the detection of malarial infection, currently being tested in a prospective cohort in Papua New Guinea [189]. This application of ATR-FTIR spectroscopy also looks to quantify levels of parasitaemia in blood, providing clinically relevant information more rapidly than current methods [190]. The analysis of blood serum using ATR-FTIR spectroscopy for the early detection of brain cancer is also approaching clinical use, currently analysing a prospective patient cohort. This application is also significant due
a clearly defined health economic study recently published, highlighting the clinical and economic benefits of such a test into the current diagnostic pathway for brain cancers [191].

In summary, this study shows that a serum blood test at the primary care level could prioritise patients for neuroimaging, improving patient survival and quality of life, whilst also providing cost savings by reducing unnecessary brain scans. Another benefit of this application is the potential of a seamless transition into the current clinical pathway. Blood tests at the primary care level are commonly ordered, and the addition of a triage test into the clinical pathway would not significantly disrupt current clinic practices.

Current advances in technology may also facilitate the uptake of vibrational spectroscopy into standard clinical practice in the near future. Automated or high-throughput instrumentation would be best suited to clinical settings so as to minimise pressure on personnel resource. High-throughput technologies are available in IR transmission (or transflection) systems, largely attributed to the development of multichannel detectors and IR sources for discrete frequency spectroscopy, as well as the use of sample substrates which can be batch processes. This could have specific impacts on the translation of tissue imaging applications which have the potential to complement histopathology [107]. On the other hand, ATR-FTIR spectroscopy is inherently limited to a single point of analysis, the IRE, which restricts the overall sample throughput, particularly when taking into account cleaning the crystal between measurements as well as background subtraction. However, the development of low cost IREs may provide a disposable substrate for ATR-FTIR spectroscopy, similarly enabling batch processing of samples and high-throughput spectral acquisition alongside the development of novel instrumentation [192]. High throughput systems have similarly been explored for Raman analysis and measurement in the liquid form.
avoids any further delay in the clinical work flow [193], [194]. Although the impact of reduced sampling times, and consequent signal to noise ratio, on classification and quantification accuracies have not yet been systematically explored.
2.9 References


the structure of proteins in solution by fourier-transform infrared spectroscopy’, Progress in

studied by infrared spectroscopy’, Prog. Biophys. Mol. Biol., vol. 72, no. 4, pp. 367–405,
1999.

[126] P. I. Haris and D. Chapman, ‘Does Fourier-transform infrared spectroscopy provide
useful information on protein structures?’, Trends Biochem. Sci., vol. 17, no. 9, pp. 328–

Vibrational Spectroscopy, J. M. Chalmers and P. R. Griffiths, Eds. Chichester, UK: John

[128] A. Barth, ‘Infrared spectroscopy of proteins’, Biochimica et Biophysica Acta (BBA)

2009.

[130] A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk, and M. Baranska,
‘Raman spectroscopy of proteins: a review: Raman spectroscopy of proteins’, Journal of


[132] K. Spalding, ‘Developing Spectroscopic Biofluid Diagnostics: Monitoring and
Therapeutic Profiling of Melanoma Patients’, University of Strathclyde, Glasgow, 2018.

from Raman or Raman optical activity spectra using random forests’, Protein Science, vol.

for imaging human skin tissue sections’, Analytical Methods, vol. 5, no. 9, p. 2281, 2013.

assignment of cytochrome c resonance Raman spectra via enzymic reconstitution with
isotopically labeled hemes’, Journal of the American Chemical Society, vol. 115, no. 26,


83


Chapter 3

Materials and Methods

3.1 Introduction

In the previous chapter, the background and aims of the study are discussed and the importance of liquid Raman spectroscopy for detection and monitoring plasma/serum analytes is highlighted. This chapter is designed to provide more information about the methodology which were employed throughout the study. Further details on sample preparation and general methodology are provided in the relevant chapters.

3.2 Raman spectroscopy

A Horiba Jobin-Yvon LabRam HR800 spectrometer with a 16-bit dynamic range Peltier cooled CCD detector was used to record the Raman spectra throughout this work. The spectrometer was coupled to either an Olympus BX41 upright or an Olympus 1X71 inverted microscope and a x10 (UMPlanFL N, Olympus) or a x60 water immersion objective (LUMPlanF1, Olympus) was employed. In the following experiments, two laser lines, 532nm and 785nm were used with the 600 lines/mm grating and 300 lines/mm grating respectively and the backscattered Raman signal was integrated for 3×80 seconds over the spectral range from 400-3500cm⁻¹.
3.3 Sample substrates

A range of substrates was tested, including a polystyrene 96 well cell culture plate, a quartz 96 well plate, and Lab-Tek plate. The polystyrene 96 well cell culture plate was purchased from True Line, USA. The quartz well plate was purchased from Hellma Analytics, Germany. The Lab-Tek plate (154534) has a 0.16-0.19mm thick, 1.0 borosilicate cover glass, and was purchased from Thermo Fischer Scientific, Ireland.

3.4 ATR-FTIR

A comparative study was conducted to investigate the potential of ATR-FTIR to detect similar variations in protein concentrations. ATR-FTIR spectra were recorded with the Perkin Elmer (MA, USA) Spotlight 400N Universal Attenuated Total Reflectance accessory of the spectrometer. A germanium crystal with a refractive index of 4.0 was employed for this analysis. In ATR mode, spectral data is the result of 16 scans from each sample with a spectral resolution of 2 cm\(^{-1}\), over the spectral range from 600cm\(^{-1}\) – 4000cm\(^{-1}\). 2 µl of the different liquid samples were deposited on the crystal and left to dry for 10-15 minutes before recording the spectra. Prior to recording, a background spectrum was also recorded from the crystal and automatically subtracted by the software.

3.5 Preparation of stock protein

Albumin (A9511) was purchased from Sigma Aldrich, Ireland. A 100mg/mL stock solution of albumin was prepared in distilled water. Fibrinogen (F3879) was purchased from Sigma
Aldrich, Ireland. A 100mg/mL stock solution of fibrinogen was prepared using warm phosphate buffered saline. For some experiments, a Sonics VCX-750 Vibra Cell Ultra Sonic Processor (Sonics & Materials Inc., USA), equipped with a model CV33 Sonic Tip was used to sonicate the fibrinogen stock solution for 5-10 seconds at 30% amplitude to improve the dispersion of the fibrinogen. Cytochrome c (C2506) was purchased from Sigma Aldrich, Ireland. A 1mg/mL stock solution of cytochrome c was prepared in distilled water. Vitamin B12 (V2876) was purchased from Sigma Aldrich, Ireland. A 10mg/mL stock solution of vitamin B12 was prepared in distilled water. The stock solutions of albumin, cytochrome c and vitamin B12 were stored at 4ºC. A fresh stock of fibrinogen was prepared before each experiment, as it was observed to be susceptible to precipitation when stored. Individual protein solutions of varying concentration were prepared for spectroscopic analysis, to explore the limit of detection of each protein and sensitivity of vibrational spectroscopic techniques to subtle changes in the protein concentrations in its native state.

3.6 Impact of centrifugal filtration

Amicon Ultra 0.5mL centrifugal filter devices (Merck, Germany) of various cut off points, 3 kDa, 10 kDa, 50 kDa and 100 kDa, were employed to concentrate and separate the proteins in the simulated plasma, based on their molecular weight. The centrifugal filtration procedure that was previously reported by Bonnier et al. was followed [1]. Pre-rinsing of the filter devices with 0.1M NaOH prior to plasma analysis is essential to avoid glycerine interference in the analysis [2]. The optimised washing and rinsing procedure includes spinning 0.5mL 0.1M NaOH at 14000×g for 30 minutes followed by three rinses with distilled water by spinning 0.5mL distilled water for 30 minutes at 14000×g. Every 30 minute wash and rinse
must be followed by spinning the device in the inverted position at 1000×g for 2 minutes, to remove the residual solution contained in the filter. After washing, 0.5mL sample is transferred to the 100kDa filter and centrifuged at 14000×g for 30 minutes. The solution that flows out from the 100kDa filter is the filtrate, which contains mostly water and molecules smaller than 100kDa. The remainder of the sample, known as the concentrate, is collected by placing the filter device upside down and spinning for 1000×g for 2 minutes. The resultant concentrate, ~50µL, contains molecules with molecular weight larger than 100kDa, and is concentrated by a factor of ~10. The filtrate is transferred to the 50kDa centrifugal filter device and is spun at 14000×g for 30 minutes. The same steps are repeated with the 10kDa and 3kDa filters. The final filtrate that flows out of the 3kDa filter contains mostly water and molecules that are smaller than 3kDa. Raman spectra of all the four concentrates and filtrates were recorded.

3.7 Observing the optimum volume to record spectra

Bio-fluid samples collected from patients are usually limited in volume, so it is crucial to be able to record the Raman spectra from such minute amounts of sample. In optimising the measurement protocol, the aim is to standardise the amount of sample volume required for obtaining Raman spectra. In order to perform such optimisation, water and stock solutions of fibrinogen were analysed in volumes from 1µL to 1mL.
3.8 Data Analysis of the recorded spectra

It is necessary to remove the background signal as much as possible from the raw Raman spectra in order to be able to distinctly visualise the Raman peaks, and in particularly to obtain accurate predictions for the purpose of disease detection. Spectral data processing and analysis are therefore critical considerations [3].

3.8.1 Spectral Pre-processing

The main contributing components of the spectra recorded using Raman spectroscopy are the Raman signal of the analyte, the background signal and noise. Pre-processing techniques are essential to remove the background signal and reduce the noise, before further analysis. Initially, protein spectra were analysed with no preprocessing. Smoothing of the raw data was done by Savitzky–Golay at a polynomial order of 5 and window 13. Then, pre-processing steps were applied to subtract the background from the data and the effects of these algorithms on subsequent regression analysis (Section 3.8.2) were investigated.

For the sake of comparison, two pre-processing techniques, Extended Multiplicative Signal Correction (EMSC) [4] and the “rubberband” method [5], were trialed on the raw dataset of the proteins. The rubber band method can be explained as a ‘rubberband’ of a defined length wrapped around the ends of the spectrum to be corrected by fitting against the curved profile of the spectrum. By assuming the convex polynomial configuration, the rubber band is wrapped around the curve profile of the spectrum from below and has the ability to exactly fit the real data with a line shape on both the sides of the spectrum [6], [7]. Rubberband correction was performed in Matlab. EMSC was developed in the 1980s for applications to
near IR analysis in food science [4]. EMSC is reported to be effective in removing the background signal of glass and water from Raman spectra of single cells [8], as well as having the additional benefit of baseline correcting the spectra, a step that is required prior to performing data analysis. In this study, EMSC was employed for the pre-processing of protein data to remove the underlying water spectrum. The EMSC algorithm uses a water spectrum generated upon illumination with 532nm, a spectrum of the relevant protein with minimum amount of water as reference, and a baseline of chosen polynomial order N as input; resulting in a modelled background consisting of water signal and a slowly varying baseline curve which can be subtracted to obtain spectra that are free from water signal. The reference was prepared by adding a few drops of distilled water to the known concentration of analyte powder and a thick paste is made. A Raman spectrum of the paste is recorded by focusing 532nm laser is used as the reference spectrum. The EMSC algorithm was implemented based on the algorithm of Kerr et al. [8] incorporating Matlab’s polyfit function.

3.8.2 Partial Least Squares Regression

Partial Least Squares Regression (PLSR) is a multivariate statistical method which aims to establish a model that relates the variations of the spectral data to a series of relevant targets. This method can be used to improve the limit of detection of Raman bio-sensing [9].

The PLSR model attempts to elucidate factors which account for the systematic majority of variation in predictors ‘X’ (spectral data) versus associated responses ‘Y’ (target values of analyte concentration) [10]. The spectral data (X matrix) is thus related to the targets (Y matrix) according to the linear equation $Y = XB + E$, where B is a matrix of regression coefficients and E is a matrix of residuals.
The PLSR algorithm allows for the construction of a regression model which can be used to predict the outcome in varying concentration of proteins, and the performance of the PLSR model in predicting varying protein concentration was evaluated in this study. In this case, the examples used are concentration and Raman signal, and therefore the algorithm can be used to predict the detection of Raman signal for a particular protein concentration. Leave – One - Out Cross Validation (LOOCV) or 20 fold cross validation was applied to assess the validity of the model. The number of latent variables was assessed, enabling the assessment of the performance of a model when applied to an unknown data set.

In the LOOCV method, a single element from the data set is used as a ‘test’ set and the remaining elements are used as the training set. This process is repeated until every single observation in the data set is used once as the ‘test’ set. The number of latent variables used for building the PLSR model is optimised by finding the value that is equivalent to the minimum of the Root Mean Square Error of Cross Validation (RMSECV) and percent variance explained by the latent variables. The spectral data obtained from the samples were split as 20% training and 30% test sets and the RMSECV was calculated. The training values are used to construct the model and its efficacy to predict the values from the test data set determines the quality of the fit. RMSECV is used to evaluate the predictive capacity of the constructed model [11]. The percent variance plot explains the number of components required for maximum variation in the input data.

The 20 fold cross-validation approach involves randomly dividing the set of observations into approximately equal size, ~50% of the spectral data were randomly selected as test set, while the remaining ~50% is used as the training set [12]. Spectra were divided into two groups of test and training spectra. The cross-validation process is then repeated 20 times.
(the folds), whereby all observations are used for both training and testing, and each observation is used for testing exactly once. The results from the folds can then be averaged to produce a single estimation. The RMSECV is calculated from the 20 iterations to measure the performance of the model for the unknown cases within the calibration set. The correlation between the concentration and spectral intensity is given by the $R^2$ value. The standard deviation was calculated to find the variation between each spectrum calculated from the same sample.

The appropriateness of various pre-processing methods can be determined through the performance of the PLSR model. In this study, PLSR analysis was performed on the pre-processed data as well as raw data. To improve the performance of PLSR model, outlier samples were removed from the original data set. The outlier samples are the ones which look completely different compared to other samples and cannot present the same (X, Y) linear relationship as other samples.

3.9 Standardisation of measurement protocol

Most of the screening methods for disease diagnosis are dependent on invasive procedures that might cause discomfort to the patients [13]. The major advantage of using body fluids as a screening target, specifically blood, urine, or saliva, is that their collection is significantly less invasive than the conventional screening methods, such as biopsy. The relatively weaker water signal in Raman spectroscopy allows the Raman spectral measurements to be taken of aqueous samples, for example those of bodily fluids. Liquid bio-spectroscopy is possible and the requirement of drying the samples can be eliminated with the use of Raman spectroscopy.
Moreover, the spectra of the biomolecules can be recorded in their native state, without any additional sample preparation steps.

A challenge of Raman spectroscopy measurements is choosing the optimum configuration for the experiment, in terms of laser line, substrate, sample volume, and instrument geometry. An ideal substrate should produce low background signals, be compatible with the test samples, and be cost effective. Many studies have reported the use of substrates such as aluminium, calcium fluoride and quartz [14]. However, a good substrate that produces minimal background signals is often expensive and can produce variable results in Raman measurements [8]. The main objective of the current work is to establish the optimal measurement settings in terms of cost and performance for liquid bio-spectroscopy experiments. To study the performance of each substrate systematically, Raman spectra of distilled water were recorded in the following substrates; polystyrene 96 well plate, quartz well plate and Lab-Tek plate, in the upright and inverted geometry, using the 532nm and 785nm laser lines of the Horiba Labram HR800 system described in Section 3.2.

3.9.1 Substrate selection

As they are commonly employed in a laboratory and clinical setting, 96 well polystyrene plates were initially tested as a suitable substrate for body fluid measurement, in the upright geometry. Distilled water was used as the sample and Raman spectra were recorded using the 532nm laser line focused by a x10 objective. When a 96 well polystyrene plate was used to record the spectrum of water using the 532nm laser line, two strong peaks were observed at ~1640cm\(^{-1}\) and ~1000cm\(^{-1}\) (Figure 3.1A). The broad peak at ~1640cm\(^{-1}\) is attributed to the OH bending of water, whereas the sharp band at ~1000cm\(^{-1}\) corresponds to a ring breathing
mode of the phenyl ring in the polystyrene. The same peak at ~1000 cm\(^{-1}\) can be seen when the Raman spectrum of an empty 96 well polystyrene well plate was recorded (Figure 3.1B). Therefore, the spectrum of Figure 3.1A indicates a significant contribution from the polystyrene substrate to the Raman spectrum of water. Such contributions may interfere with the analysis of subtle changes in the spectrum of the body fluids.

**Figure 3.1:** (A) Raman spectrum of water in the 96 polystyrene well plate and (B) an empty polystyrene well plate recorded using the 532nm laser line in the upright geometry with the x10 objective. This study indicates that 96 well polystyrene plates cannot be used as the substrate, as polystyrene peaks are observed superimposed on the water peaks.

The Raman spectrum of distilled water in the quartz 96 well plate substrate shows a strong water band at ~1640 cm\(^{-1}\) (Figure 3.2). No other additional bands are seen, which makes the
quartz substrate an ideal substrate to use for Raman measurements. However, despite its advantages as an ideal substrate, the quartz substrate is very expensive and is not suitable for day-to-day lab experiments, or ultimately for translation to the clinical environment.

Figure 3.2: Raman spectrum of water recorded in the quartz well plate in the upright geometry using the 10x objective and the 532nm laser line. A strong water peak at 1650cm$^{-1}$ can be seen, with minimal background noise.

As an alternative, a glass bottomed Lab-Tek plate was explored as substrate. The plates are also commonly employed in the laboratory setting, and the thin glass bottom, of 0.16 - 0.19 mm thickness, may not contribute significantly to the Raman signal. Figure 3.3A shows the Raman spectrum of water recorded using the 532nm laser, focused by the x10 objective, and Figure 3.3B is the Raman spectrum of an empty Lab-Tek plate recorded by focusing the laser on the glass bottom. The spectrum recorded from the empty Lab-Tek plate was that of only glass, which has a strong Raman band at 850cm$^{-1}$, attributed to the presence of orthosilicate [15]. Figure 3.3A demonstrates that, since the bottom of the Lab-Tek plate is made of thin glass; it does not influence the Raman spectrum of the sample. Therefore, the Lab-Tek plate was chosen as the appropriate substrate for the further Raman measurements.
Figure 3.3: (A) Raman spectra of water recorded in the Lab-Tek plate and (B) empty Lab-Tek plate recorded using the 532nm laser line in the upright geometry with the x 10 objective. No interference from glass peaks at 850cm\(^{-1}\) can be seen in A. Therefore, the Lab-Tek plate can be used as an ideal substrate for Raman measurements.

3.9.2. Wavelength selection

Another major challenge in Raman measurements is the selection of a suitable wavelength that is compatible with the substrate and gives maximum Raman signal of the samples used. The objective here is to select the best laser line which provides high quality Raman spectra with enhanced spectral details. From the results so far, it is clear that the 532nm laser line is compatible with Lab-Tek plate substrates and provides a strong Raman signal of water with
minimal background interference. It is always desirable to use a short wavelength with high excitation power source, as there is a linear relationship between the power of the scattered light and the intensity of the incident light and the power of the scattered light is inversely proportional to the fourth power of wavelength. In order to gauge the performance of a higher power longer wavelength laser, with a liquid sample and Lab-Tek plate as substrate, the Raman spectrum of distilled water was recorded using the 785nm laser line focused by the x10 objective.

**Figure 3.4**: Raman spectrum of water recorded using the 785nm laser line in the Lab-Tek plate in the upright geometry with the x 10 objective. Strong fluorescence background is observed due to absorption by glass impurities at 785nm

Figure 3.4 presents the Raman spectrum of water recorded with the 785nm laser line using a Lab-Tek plate as substrate. The Lab-Tek plate exhibits a strong fluorescence background signal which will swamp the Raman measurements of the sample. Since impurities in glass are absorbing and highly fluorescent in the NIR region, a laser line in the visible region of the spectrum is more desirable when working with glass substrates [16]. In terms of source wavelength, the 532nm laser line was chosen as the ideal laser line for further measurements,
as it gives no background fluorescence and a high quality Raman signal can be recorded from the sample.

### 3.9.3 Geometry selection

The next step in the experimental set-up is to choose the geometry of Raman measurements. The objective here is to identify the best instrumental set up that enables an increase in the overall spectral intensity accompanied by an improved signal to noise (S/N) ratio with small sample volume.

![Figure 3.5](image)

**Figure 3.5:** (A) Raman spectrum of the distilled water recorded in the upright geometry with x10 objective (B) enhanced Raman spectrum of the distilled with considerably lower background and improved S/N, recorded in the inverted geometry focused with x60 water immersion objective.
Bonnier et al. have already reported the advantages of the inverted Raman set up [1]. To verify, the Raman spectrum of water was recorded using both the upright geometry using the x10 objective and the inverted geometry using the x60 water immersion objective. From the results shown in Figure 3.5 A and B, it is notable that an enhanced water spectrum with considerably lower background was recorded in the inverted geometry. Higher collection efficiency of Raman signals could be obtained in the inverted geometry when focused with the water immersion objective.

![Figure 3.6: The inverted geometry used to analyse the serum focused by immersion objective. [1]](image)

An improved protocol of Raman spectroscopy set up, coupled with fractionation of serum using centrifugal filters to concentrate and separate the low molecular weight proteins, was demonstrated by Bonnier et al. [1]. Better analysis of serum using Raman spectroscopy was reported when the sample was analysed in the inverted geometry using a water immersion...
objective with 785nm laser line and CaF$_2$ substrate (Figure 3.6). In this study, a x60 water immersion objective is used with 532nm laser line and the substrate used is a Lab-Tek plate. A drop of water is used to minimise the differences in the refractive indexes between sample, objective and the substrate. However, the water drop does not contribute to the data collected, as it is outside the focus of the beam.

3.9.4. Protein measurement in the upright and inverted geometry

The stock solutions of albumin (100mg/mL), fibrinogen (100mg/mL), cytochrome c (1mg/mL) and vitamin B12 (10mg/mL) were prepared and Raman spectra were recorded in the optimised inverted geometry with the Lab-Tek plate as the substrate. The raw spectra of the proteins were baseline corrected suing the rubberband method and smoothed using the Savitzky–Golay algorithm (polynomial 5, window 13). Figure 3.7 shows the Raman spectra of the fingerprint region of the stock solution of proteins recorded in the inverted geometry using the water immersion objective. In comparison to the measurements taken in the upright geometry, the spectral features are well defined and significantly enhanced. Therefore, the inverted microscope set-up with the x 60 water immersion objective is ideal for liquid bio-spectroscopy measurements. Characteristic Raman peaks of the proteins are listed in Table 3.1. [11], [17]–[19]
Figure 3.7: Raman spectra of the stock solutions of albumin, fibrinogen, cytochrome c and vitamin B12 recorded in the finger print region in the inverted geometry focused by water immersion x60 objective. Well-defined Raman peaks with minimum background were obtained.
Table 3.1 Peak assignments for the Raman spectra of plasma proteins [11], [17]–[19]

<table>
<thead>
<tr>
<th>Plasma Proteins</th>
<th>Peak position (cm(^{-1}))</th>
<th>Assignments of Raman Vibrational Modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albumin</td>
<td>828</td>
<td>Tyrosine</td>
</tr>
<tr>
<td></td>
<td>899</td>
<td>v (CN)</td>
</tr>
<tr>
<td></td>
<td>940</td>
<td>v (CCN)(_\text{sym}), v(CC)</td>
</tr>
<tr>
<td></td>
<td>1002</td>
<td>Phenylalanine band</td>
</tr>
<tr>
<td></td>
<td>1089</td>
<td>v (CN)</td>
</tr>
<tr>
<td></td>
<td>1102</td>
<td>v (CN)</td>
</tr>
<tr>
<td></td>
<td>1336</td>
<td>C-H deformation</td>
</tr>
<tr>
<td></td>
<td>1450</td>
<td>C-H deformation</td>
</tr>
<tr>
<td></td>
<td>1655</td>
<td>Amide I (C=O stretching mode of proteins, α-helix conformation)</td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>758</td>
<td>Symmetric ring breathing of tryptophan</td>
</tr>
<tr>
<td></td>
<td>878</td>
<td>Arginine</td>
</tr>
<tr>
<td></td>
<td>1003</td>
<td>Phenylalanine band</td>
</tr>
<tr>
<td></td>
<td>1250</td>
<td>Amide III (C–N stretching mode, mainly α-helix conformation)</td>
</tr>
<tr>
<td></td>
<td>1336</td>
<td>C-H deformation</td>
</tr>
<tr>
<td></td>
<td>1450</td>
<td>C-H deformation</td>
</tr>
<tr>
<td></td>
<td>1552</td>
<td>Tryptophan</td>
</tr>
<tr>
<td></td>
<td>1659</td>
<td>Amide I (C=O stretching mode of proteins, α-helix conformation)</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>750</td>
<td>Heme breathing</td>
</tr>
<tr>
<td></td>
<td>1127</td>
<td>v(CH(_3))</td>
</tr>
<tr>
<td></td>
<td>1314</td>
<td>All bonds of heme</td>
</tr>
<tr>
<td></td>
<td>1585</td>
<td>C=C</td>
</tr>
<tr>
<td>Vitamin B12</td>
<td>722</td>
<td>5,6-Dimethylbenzimidazole</td>
</tr>
<tr>
<td></td>
<td>1160</td>
<td>Corrin ring</td>
</tr>
<tr>
<td></td>
<td>1197</td>
<td>5,6-Dimethylbenzimidazole</td>
</tr>
<tr>
<td></td>
<td>1498</td>
<td>Corrin ring</td>
</tr>
</tbody>
</table>
The spectra of albumin and fibrinogen shown in Figure 3.7 clearly reveal the common Raman peaks of these two proteins. These include the amide I band around ~1659cm⁻¹, a relatively sharp band at 1003cm⁻¹ associated with phenylalanine, intense bands at ~1336cm⁻¹ and ~1450cm⁻¹ due to C-H deformation, and a vibration band at ~940cm⁻¹ related to C-C stretching mode backbone of α-helix structure. The signature peaks of albumin that differentiate it from fibrinogen are bands at 899cm⁻¹ and 1102cm⁻¹, that can be related to ν(CC) and ν(CN). The signature peaks of fibrinogen are sharp bands observed at 758cm⁻¹ and 1552cm⁻¹ that can be assigned to tryptophan. Raman bands of cytochrome c and vitamin B12 are unique and can be easily distinguished; as evidenced in Figure 3.7.

3.9.5 Observing the optimum volume to record spectra

In order to determine the smallest volume that can be measured efficiently with the inverted microscope set up, Raman spectra of distilled water and a stock solution of fibrinogen were recorded from varying volumes of 1 μL to 1mL, in the Lab-Tek plate. The raw spectra ware baseline corrected using the rubberband method.

The intensity of the water spectrum varied slightly for different volumes, even though the variation is not large (Figure 3.8A). The intensity of the water peak at ~1640cm⁻¹ was measured and plotted against the volume (Figure 3.8C). From both the graphs, it is noteworthy that the variation in intensity is negligible as no signal is lost when the volume is reduced from 1mL to 1μL. The key point here is that high quality, consistent Raman spectra were obtained from a sample volume as low as 1μL. Freshly prepared fibrinogen stock solution was also measured as a function of volume. To maintain a level of consistency,
Raman spectra were always recorded as soon as the sample was transferred to the middle of the Lab-Tek plate to prevent drying of the sample and the acquisition time was 120 seconds for 3 accumulations. From the results (Figure 3.8A), it is clear that the intensity of the fibrinogen spectra did not vary significantly with volumes. In order to perform a quantitative evaluation of each of the recorded fibrinogen spectra, the intensity of the phenylalanine peak at 1003 cm\(^{-1}\) was measured for all the volumes and a graph was plotted (Figure 3.8D).

**Figure 3.8:** (A) Raman spectra of distilled water recorded from 1 μL to 1mL volume, (B) Raman spectra of fibrinogen solution recorded from 1 μL to 1mL volume, (C) Intensity versus volume graphs of distilled water and (D) Intensity versus volume graphs of fibrinogen stock. This shows a strong Raman signal can be recorded from a volume as small as 1μL in the inverted geometry using Lab-Tek Plate as substrate.

The results in this section demonstrate that enhanced Raman spectra with defined spectral features can be obtained from samples of volume as low as 1μL. This method is potentially advantageous in a clinical set up, where the sample volume is usually very low.
3.10 ATR-FTIR analysis of varying concentrations of fibrinogen – A comparative study

The application of ATR-FTIR to collecting high quality spectra from highly concentrated bodily fluids such as plasma has been previously investigated [20], [21]. This is a reliable, rapid and convenient approach to analyse the bodily fluid samples, but suffers from a major drawback of the Vroman effect. The objective of this study is to explore the dependence of the absorbance on the concentration of the sample. A physiologically relevant concentration range of fibrinogen from 1mg/mL to 50mg/mL was analysed in the ATR mode from 650-4000cm\(^{-1}\). 2μL samples were deposited directly onto the crystal and were air-dried prior to the measurements to remove the water content completely. Despite the rich composition of fibrinogen in higher concentrations, the intense water contribution from the liquid samples obscured the spectral features of the protein. Hence, air-drying is important to visualise all the spectral details clearly in the ATR-FTIR measurements. A drying time of 4-6 minutes was required for the highest concentration (50mg/mL) and about 8-10 minutes for the lowest concentration (1mg/mL). Following the air drying, the fibrinogen samples display numerous well defined peaks at 1536cm\(^{-1}\) (amide II band), 3280cm\(^{-1}\) (H-O-H stretching), 2957cm\(^{-1}\) (asymmetric CH\(_3\) stretching), 1080 cm\(^{-1}\) (C-O stretch) and 1453 cm\(^{-1}\) (CH\(_2\) scissoring) (Figure 3.9A). Figure 3.9A shows that higher concentrations yield high intensity peaks and the low concentrations give low intensity peaks. However, a plot of absorbance vs concentration (Figure 3.9 B) shows that the linearity of the plot is lost after 30mg/mL, due to saturation of the IR absorbance. According to the Lambert-Beer law, when the thickness of the deposit from the sample droplet exceeds the penetration depth of the ATR evanescent wave, an increase in solution concentration results in a linear correlation between the sample
concentration and signal intensity. However, linearity is lost above 30mg/mL, when the thickness of the dried fibrinogen sample exceeds the depth of penetration of evanescent wave. No further increase in the absorbance after 30mg/mL suggests loss of Lambert-Beer type dependence of the absorbance. Thus, a linear dependence of the absorbance and the intensity, and therefore quantitative potential, is observed only at lower concentration range [1].

![Figure 3.9: (A) ATR-FTIR spectra of varying concentration (1mg/mL to 50mg/mL) of fibrinogen in distilled water collected after depositing 2μL of the sample on the ATR crystal. (B) Intensity vs concentration plot of the Amide 2 peak shows a steady increase in the intensity until 30mg/mL and remains constant for 40 and 50mg/mL](image)
Therefore, sample drying is a major obstacle in the ATR-FTIR measurements as it leads to chemical and physical inhomogeneity and saturation of the absorbance at higher concentrations. Since Raman can perform spectroscopy in aqueous solutions, it is the best choice for liquid bodily fluid measurements.

3.11 Summary

Raman spectroscopy is a reliable, label free, rapid and cost effective technique, delivering a molecular fingerprint of the sample. In the present study, an optimised protocol for recording Raman spectra from small volumes of liquid protein samples has been demonstrated. An improved spectral quality with significant reduction in the background signal is achieved by using a water immersion objective in the inverted geometry, rather than the upright geometry. Using the water immersion objective has the additional advantage of minimising the refractive index differences between substrate, sample and objective without adding to the signal collected. A Lab-Tek plate was found to be the appropriate substrate as it provided strong Raman spectra using the 532nm laser line, focused with a water immersion objective, with negligible background in the inverted geometry. Additionally, in the inverted geometry, Raman spectra can be recorded from a sample volume which can be as small as 1 μL, which increases the potential of using this method in clinical applications. The limitations of performing ATR-FTIR of dried deposits have also been demonstrated. The demonstrated protocol for Raman analysis of protein samples in the liquid form can be applied to a wide range of body fluids and has huge potential in clinical applications.
3.12 References


Chapter 4

Raman spectroscopic analysis of High Molecular Weight Proteins in solution– considerations for sample analysis and data pre-processing

The following chapter reproduced the published journal article titled as ‘Raman spectroscopic analysis of High Molecular Weight Proteins in solution– considerations for sample analysis and data pre-processing’, Analyst, 2018, 143, 5987-5998.

Author List: Drishya Rajan Parachalil, Brenda Brankin, Jennifer McIntyre, and Hugh J Byrne

DRP performed all experimental analysis and authored the publication. BB assisted with the Ion Exchange Chromatography, JMcC and HJB input to conceptual design of the work, drafting and proofing of the manuscript.

4.1 Abstract

This study explores the potential of Raman spectroscopy, coupled with multivariate regression and protein separation techniques (ion exchange chromatography), to quantitatively monitor diagnostically relevant changes in high molecular weight proteins in liquid plasma. Measurement protocols to detect the imbalances in plasma proteins as an indicator of various diseases using Raman spectroscopy are optimised, such that strategic
clinical applications for early stage disease diagnostics can be evaluated. In a simulated plasma protein mixture, concentrations of two proteins of identified diagnostic potential (albumin and fibrinogen) were systematically varied within physiologically relevant ranges. Scattering from the poorly soluble fibrinogen fraction is identified as a significant impediment to the accuracy of measurement of mixed proteins in solution, although careful consideration of pre-processing methods allows construction of an accurate multivariate regression prediction model for detecting subtle changes in the protein concentration. Furthermore, ion exchange chromatography is utilised to separate fibrinogen from the rest of the proteins and mild sonication is used to improve the dispersion and therefore quality of the prediction. The proposed approach can be expeditiously employed for early detection of pathological disorders associated with high or low plasma/serum proteins.

4.2 Introduction

Raman spectroscopy has emerged over the past 20 years as an increasingly routine analytical technique for a wide range of applications, as it provides specific biochemical information without the use of extrinsic labels. This technique can provide intrinsic vibrational signatures of the material of interest in a non-destructive fashion, and its potential for diagnostic applications has been well demonstrated, notably in human serum and plasma [1]–[4]. Raman spectroscopy provides a vibrational signature of a complex biological mixture which is a result of the contributions from all the major components from that mixture, and changes in the concentrations of the components will give rise to notable changes in the Raman signal. However, although both Raman and Fourier-Transform Infrared (FTIR) spectroscopy have been widely explored to study bodily fluids over the last two decades, most of these studies
have been carried out on air dried samples, in order to avoid the water contribution in the case of FTIR, and to increase the concentration of the analytes in the case of Raman [5]–[9]. The major limiting factor in the use of dried samples is the so-called “coffee-ring” effect, or, specifically in terms of blood serum, the Vroman effect [10]–[12], whereby different analytes precipitate from solution at different rates, giving rise to variations in the spectral features due to chemical and physical inhomogeneity. This leads to spatially varying chemical compositions and sample thicknesses, and unreliable results[13]. Ultimately, it is desirable to undertake the analysis in the native state of bodily fluids, in which the chemical composition is averaged out by molecular motion over the measurement time, and additional drying steps can be eliminated. This aim naturally favours Raman analysis, as water is a relatively weak Raman scatterer.

In this paper, the sensitivity of Raman spectroscopy to detect subtle changes in a simulated plasma protein-mixture concentration is explored, specifically for the higher molecular weight proteins. Albumin is the most abundant plasma protein, normally constituting about 50% of the plasma protein and has a molecular weight of 66kDa [14]. The normal concentration of albumin in the human body is 30mg/mL, although it dramatically decreases in critically ill patients and does not increase again until the recovery phase of the illness [15]. Several studies have demonstrated that the functions of albumin, such as ligand binding and transport of various molecules, can be applied to the treatment of cirrhotic patients and patients suffering from other end stage liver diseases [16]–[18]. It is clear that closely monitoring the variation in albumin concentration could act as an indicator of liver diseases and other related pathologies. Fibrinogen is a 340kDa (0.4% in human plasma) dimeric plasma glycoprotein synthesised by the liver and plays a major role in blood coagulation [19].
The normal concentration of fibrinogen in human body is ~3mg/mL, and any variation in this concentration can be an indicator of disease states [20]–[22]. Many clinical studies have consistently shown elevated levels of fibrinogen in patients with cardiovascular disease and thrombosis [22]–[25].

The conventional test kits available in a hospital for plasma/serum analysis suffer from long time delays for the availability of results due to the need of specialised laboratories, which may in turn delay the therapy, and prolong patient anxiety. The potential of vibrational spectroscopy techniques coupled with multivariate analysis techniques have been previously investigated for a range of clinical applications [1]–[9], [26]–[29]. This paper evaluates the potential of Raman spectroscopy as a diagnostic tool to detect minute changes in the plasma protein concentrations in aqueous samples and systematically explores the challenges to such liquid based biopsy techniques, including sample scattering and abundance of individual constituent components, while presenting some potential solutions to improve the protocols of liquid biopsy monitoring using Raman spectroscopy.

A simulated plasma protein mixture of high and low molecular weight proteins, i.e. albumin, fibrinogen, cytochrome c and vitamin B12, at physiologically relevant concentrations, was prepared and variations were made to these concentrations over physiologically relevant ranges. Separation of proteins in the solution was performed by ion exchange chromatography to separate high molecular weight proteins from low molecular weight proteins, and high molecular weight fraction proteins from each other. The efficiency of data pre-processing methods (rubberband and Extended Multiplicative signal Correction (EMSC)) in removing the background, to build an accurate prediction model, was explored and mild sonication was used to improve the dispersion of fibrinogen. The standardisation of
measurement protocol and other experimental parameters is detailed and the results of concentration dependence study of proteins, in isolation and protein mixtures, and the chemometric methods used to build the prediction model are presented.

4.3 Materials and Methods

4.3.1 Preparation of stock protein and protein mixture

Albumin (A9511), fibrinogen (F3879), cytochrome c (C2506) and vitamin B12 (V2876) were purchased from Sigma Aldrich, Ireland. Solutions of individual proteins of varying concentration were prepared, to initially explore the accuracy of detection of each protein and sensitivity of vibrational spectroscopic techniques to subtle changes in the protein concentrations in its native state. Furthermore, in order to assess the ability of Raman spectroscopic techniques to detect subtle changes in the concentration of the protein in a more complex mixture, potentially usable as biomarkers of various disease states, varying concentrations of each protein in the protein-mixture were prepared. Concentrations of albumin and fibrinogen, were varied in the protein mixture in the physiologically relevant ranges, from 5mg/mL to 50mg/mL [15] and 0.5mg/mL to 5mg/mL [22] respectively while maintaining the concentrations of cytochrome c and vitamin B12 constant. The stock solutions and the protein-mixture solutions and analysed in the liquid form using Raman spectroscopy.
4.3.2 Ultrasonication

A Sonics VCX-750 Vibra Cell Ultra Sonic Processor (Sonics & Materials Inc., USA), equipped with a model CV33 Sonic Tip was used to sonicate the fibrinogen stock solution for 5-10 seconds at 30% amplitude at room temperature to explore the effect of improved dispersion of the fibrinogen on the measurement procedure. Previous studies have demonstrated negligible effects of ultrasonication on the integrity of the fibrinogen structure [30].

4.3.3 Ion exchange chromatography

Carboxymethyl-cellulose (C9481) was purchased from Sigma Aldrich, Ireland. It acts as a weak cationic exchanger and binds to the positively charged molecules [31]. Glycine (G8898) was purchased from Sigma Aldrich, Ireland and glycine buffer of pH 10 was prepared as the elution buffer [32]. 1mL of the protein-mixture was pipetted into a centrifuge tube and 0.08g of carboxymethyl-cellulose. The solution was mixed for 10 minutes on a Spira-mix roller and then centrifuged at 14000g for 5 minutes. The unbound material was present in the supernatant and was transferred to a fresh tube. The pellet was washed using 2mL glycine buffer by repeated inversion, followed by centrifugation at 14000g for 5 minutes. The supernatant that contains the fibrinogen was carefully transferred to a fresh centrifuge tube and Raman analysis was performed.
4.3.4 Raman spectroscopy

A Horiba Jobin-Yvon LabRam HR800 spectrometer with a 16-bit dynamic range Peltier cooled CCD detector was used to record the Raman spectra throughout this work. The spectrometer was coupled to Olympus 1X71 inverted microscope and a x60 water immersion objective (LUMPlanF1, Olympus) was employed. In all the following experiments, a 532nm laser of 12mW at the sample was used with the 600 lines/mm grating and the backscattered Raman signal was integrated for 3 accumulations and a total acquisition time of 80 seconds over the spectral range from 400-1800 cm\(^{-1}\).

4.3.5 Sample substrates

The Lab-Tek plate (154534) was chosen as the optimal substrate for this study. It has a 0.16-0.19mm thick glass bottom, 1.0 borosilicate cover glass, and was purchased from Thermo Fischer Scientific, Ireland.

4.3.6 Spectral preprocessing

Pre-processing techniques are essential to remove the background signal and reduce the noise, before further analysis. Smoothing of the raw data was done by Savitzky–Golay at a polynomial order of 5 and window 13. Two pre-processing techniques, Extended Multiplicative Signal Correction (EMSC) and the rubberband method, were trialed on the raw dataset of the proteins in Matlab, at different stages of the study. EMSC was employed for the pre-processing of protein data to remove the underlying water spectrum [33], which has an OH bending vibration at ~1640 cm\(^{-1}\) [34] which can obscure the protein signals at low
concentrations. The reference for EMSC was prepared by adding a few drops of distilled water to the known concentration of protein powder and a thick paste is made (~10mg/mL). A Raman spectrum of the paste was recorded using the 532nm laser as source and used as the reference spectrum. Rubberband correction was carried out in Matlab by wrapping a ‘rubberband’ of defined length around the ends of the spectrum to be corrected and fitting against the curved profile of the spectrum [35].

4.3.7 Partial Least Squares Regression

The Partial Least Squares Regression (PLSR) algorithm was applied to construct a regression model that can be used to predict the outcome in varying concentration of proteins, and the performance of the model in predicting varying protein concentration was evaluated [36]. The PLSR model attempts to elucidate factors that account for the systematic majority of variation in predictors ‘X’ (spectral data) versus associated responses ‘Y’ (target values of protein concentration). The spectral data (X matrix) is thus related to the targets (Y matrix) according to the linear equation $Y = XB + E$, where B is a matrix of regression coefficients and E is a matrix of residuals. Leave - One - Out cross validation was applied to assess the validity of the model. In this case, the number of latent variables was assessed, enabling the assessment of the performance of a model when applied to an unknown data set. The number of latent variables used for building the PLSR model is optimised by finding the value that is equivalent to the minimum of the Root Mean Square Error of Prediction (RMSEP) and percent variance explained by the latent variables. The spectral data obtained from the 30 samples were split as 20% training and 30% test sets and the RMSEP was calculated. RMSEP is used to evaluate the predictive capacity of the constructed model [37]. The percent variance
plot explains the number of components required for maximum variation in the input data. The appropriateness of various pre-processing methods can be determined through the performance of the PLSR model. In the following, for each aspect of the study, the optimum conditions for model development are detailed.

4.4 Results

4.4.1 Standardisation of measurement protocol

For the analysis of liquid protein samples, an optimised inverted set-up, previously demonstrated by Bonnier et al. [13] was used. Better analysis of serum using Raman spectroscopy was reported when the sample was analysed in the inverted geometry using a water immersion objective with a 785nm laser and CaF$_2$ substrate. In this study, a x60 water immersion objective is used with a 532nm laser and the substrate used was a Lab-Tek plate. The 532nm laser was chosen as it is compatible with (thin glass bottomed) Lab-Tek plate substrates and provides a strong Raman signal of water with minimal background interference. A drop of water is used to minimise the differences in the refractive indexes between sample, objective and the substrate. However, the water drop does not contribute to the data collected, as it is outside the focus of the beam. This set-up also has an added advantage of providing high quality, consistent Raman spectra from a sample volumes as low as 1μL.

Figure 4.1 presents the spectra of the fingerprint region of the stock solutions of proteins recorded in the inverted geometry. The raw spectra of the proteins were baseline corrected using the rubberband method and smoothed using the Savitzky–Golay algorithm (polynomial
Measurement in the inverted geometry, using a water immersion objective, is found to be the best instrumental set up that enables an increase in the overall spectral intensity accompanied by an improved signal to noise (S/N) ratio with small sample volume.

**Figure 4.1**: Raman spectra of the stock solutions of albumin, fibrinogen, cytochrome c and vitamin B12 recorded using the 532nm laser in the finger print region in the inverted geometry focused by water immersion x60 objective. Well-defined Raman peaks with minimum background were obtained.

The spectra of albumin and fibrinogen shown in Figure 1 clearly reveal the common Raman peaks of these two proteins. These include the amide I band around $\sim1659\text{cm}^{-1}$, a relatively sharp band at $1003\text{cm}^{-1}$ associated with phenylalanine, intense bands at $\sim1336\text{cm}^{-1}$ and $\sim1450\text{cm}^{-1}$ due to C-H deformation, and a vibration band at $\sim940\text{cm}^{-1}$ related to C-C
stretching mode backbone of α-helix structure. The signature peaks of albumin that differentiate it from fibrinogen are bands at 899cm\(^{-1}\) and 1102cm\(^{-1}\), that can be related to \(\nu(\text{CC})\) and \(\nu(\text{CN})\) [38]. The signature peaks of fibrinogen are sharp bands observed at 758cm\(^{-1}\) and 1552cm\(^{-1}\) that can be assigned to tryptophan [39]. Raman bands of cytochrome c and vitamin B12 are highly specific and can be easily distinguished, as evidenced in Figure 4.1 [40], [41].

4.4.2 Monitoring the concentration dependence of proteins in aqueous solution

**Albumin.** Protein solutions were prepared by varying the concentration of albumin in order to achieve the physiologically relevant range from 5mg/mL to 50mg/mL. Figure 4.2A show the raw unpre-processed spectra, which exhibit a steady increase in the spectral intensity when the concentration is increased from 5mg/mL to 50mg/mL. The spectrum of the highest concentration clearly shows albumin features, whereas those of the lower concentrations are dominated by water, which has a characteristic OH bending mode at \(\sim 1640\text{cm}^{-1}\). As the concentration of albumin increases, a notable increase in the background can also be observed, which can be attributed to scattering. Although many studies suggest that the broad background present in Raman spectra is due to fluorescence [42], albumin is optically transparent at 532 nm, so the background is rather due to scattering of the source laser as well as the Raman scattered light, which enters the spectrometer as stray light, and is dispersed across the CCD in a wavelength independent fashion [43]. In order to analyse the spectral variations and the albumin concentrations, the PLSR algorithm was applied. The percent variance plot in Figure 5.2B gives a rough indication of how the algorithm progressively fits
the spectral data, showing that nearly 68% of the variance is explained by the first component, while as many as four additional components make significant contributions.

**Figure 4.2** (A): Raw Raman spectra of varying concentrations of albumin (5mg/mL – 50mg/mL) in distilled water, recorded using the 532nm laser (B): Percent variance explained by the components, (C): plot of PLSR coefficient with Albumin features, (D): Linear predictive model built from the PLSR analysis

Based on the percent variance explained by the latent variables and the minimum value of RMSEP, the optimum number of latent variables to reach the best model is determined. The PLSR coefficient plot displayed in Figure 5.2C, confirms the correlation of the data in Figure
5.2D is based on albumin features, such as the peaks at ~1665cm$^{-1}$, ~1448cm$^{-1}$ and ~1337cm$^{-1}$. Finally, after selecting the optimum number of components for the data set analysed, a predictive model is built from the PLSR analysis (Figure 4.2D), to compare the observations to the known concentrations of albumin in the samples with the estimated concentrations from the spectral data sets. Figure 4.2D indicates that a good linear model could be obtained with the raw data set. However, the PLSR coefficient is not a clean albumin spectrum and has a large background due to scattering, indicating that scattering could have influenced the model. Furthermore, the minimum value of RMSECV was found to be 22.59mg/mL, indicating a poor accuracy of prediction over the range 5mg/mL to 50mg/mL. Analysis of the raw albumin concentration dependence serves as an initial illustration of some of the issues presented by measurement of high molecular weight macromolecules in solution. Appropriate pre-processing steps could help to minimise the background from scattering effects. Hence, rubberband pre-processing steps were performed on the data set before PLSR analysis and the model obtained is displayed in Figure 4.3.
Figure 4.3 (A): Rubberband corrected Raman spectra of varying concentrations of Albumin (5mg/mL – 50mg/mL) in distilled water, (B): % variance explained by the latent variables, (C): plot of PLSR coefficient with Albumin features, (D): Linear predictive model built from the PLSR analysis

Figure 4.3A shows the albumin data set after background correction using the rubberband method. Figure 4.3B shows the percent variance explained by the latent variables, indicating that three components accounted for the majority of the variance. Five latent variables were chosen for this model and the resultant PLSR coefficient exhibits strong albumin features, as shown in Figure 4.3C. A linear predictive model can be defined from the rubberband corrected data set of varying concentration of albumin in water Figure 4.3D. The RMSEP was found to be 1.58mg/mL after applying the rubberband pre-processing steps for the same
data set. The results suggest that there is a significant improvement in the predictive capacity of the constructed model when rubberband pre-processing steps are applied to the data set.

Simulated “pathological” plasma protein mixtures were prepared by varying the concentration of albumin in order to achieve the physiologically relevant range from 5mg/mL to 50mg/mL and by maintaining the concentrations of fibrinogen, cytochrome c and vitamin B12 constant at the concentrations of the “healthy” human plasma. The concentrations for hypoalbuminaemia (>30mg/mL) and hyperalbuminemia (<30mg/mL) have been deliberately included in the set of samples being prepared. Based on the results of Figure 4.3, rubberband correction was applied to the dataset in an attempt to improve the accuracy of the prediction by performing baseline correction. Notably, the Raman spectral features of the protein mixture were seen to decrease with increasing albumin concentration (Figure 4.S1A in supplemental material), and the PLSR coefficient obtained from this data shows inverse albumin features (Figure 4.S1C), indicating that the model built from this dataset is not reliable, as the high degree of scattering is effecting the dataset and the prediction model is not based on the albumin features. Hence, the EMSC based algorithm was applied to the data set in an attempt to eliminate the scattering associated with the albumin data in the simulated plasma and subsequently improve the prediction model. EMSC of polynomial order 4 was performed on the data set of varying concentration of albumin in simulated plasma protein mixture. It has previously been shown that the reference used for EMSC does not have to be a precise match to the sample of interest [44] and therefore a spectrum of albumin recorded with 532nm was used, as it is the most abundant protein in the mixture. To minimise any scattering, but also contribution of water, the powder has been diluted with a minimum amount of water (~1mL to 10mg).
Figure 5.4A displays the albumin spectra after performing background correction using the EMSC algorithm. The amide I band at 1665 cm\(^{-1}\) and CH\(_2\) deformation band at 1445 cm\(^{-1}\) can be clearly seen in the corrected spectra. Based on the percentage variance explained by the latent variables (Figure 4.4B) and the minimum value of RMSEP, seven latent variables were found to be optimal for this model. The PLSR coefficient shows albumin features (Figure 4.4C), indicating that the prediction is now based on the variation in the albumin peak intensity. A linear prediction model was achieved from this model (Figure 4.4D). The minimum value of RMSEP is 1.5844 mg/mL, indicating an improved prediction capacity. This value is the same as the minimum value of RMSEP recorded for the varying concentration of albumin in distilled water, indicating that the PLSR model of EMSC corrected simulated plasma spectra is as accurate as the PLSR model of rubberband corrected spectra of varying concentrations of pure albumin in water. The results demonstrated in this section suggest that this model can be effectively used to detect variations in the concentration of albumin in human plasma, as a result, for example, of liver disorders at an early stage. A strong reduction in the RMSEP indicates that the EMSC algorithm can efficiently subtract the background without altering the albumin features, which in turn improves the prediction of the model.
Figure 4.4 A: EMSC corrected of varying concentrations of albumin in simulated plasma, and B: Percent variance explained by the latent variables, C: PLSR coefficient showing albumin features, and D: Linear prediction model defined from the dataset.

**Fibrinogen.** Fibrinogen solutions were prepared by diluting the stock solution of 100mg/ml to the more physiologically relevant range of 0.5mg/mL to 5mg/mL. Raman spectra were recorded from the protein samples and smoothed using Savitzky–Golay (polynomial 5, window 13). When the rubberband method was applied on this dataset to perform baseline correction, the PLSR coefficient spectrum obtained was an inverse water spectrum, as shown in supplementary information (Figure 4.S2). Fibrinogen is poorly soluble in water, such that the fibrinogen solution is visually cloudier than the albumin solution. This significant problem of lack of solubility due to the protein aggregation leads to scattering of the more
pronounced Raman signal of the water, in a concentration dependent fashion. Hence, EMSC with a polynomial of order 4 was performed on the same data set to pre-process the data prior to PLSR analysis. The reference spectrum was obtained under similar conditions as the albumin reference, from a fibrinogen paste with minimal amount of water. A polynomial of order 3 resulted in the best correction. The output, however, is a very noisy spectral data set with some indication of fibrinogen features in the spectra, notably at ~758cm\(^{-1}\), ~1650cm\(^{-1}\), ~1450cm\(^{-1}\), ~1336cm\(^{-1}\) and ~1250cm\(^{-1}\) (Figure 4.S3 in supplemental).

In an attempt to overcome the lack of solubility of the protein, the stock solution was ultrasonicated to enhance the dispersion of fibrinogen and obtain a clear solution. Ultrasonication for approximately 10 seconds at 30% amplitude resulted in a clear solution of fibrinogen with a significantly improved Raman signal (Figure 4.S4 in supplemental). Varying concentrations of fibrinogen samples in the physiologically relevant range were prepared using the ultrasonicated fibrinogen stock.
Figure 4.5 A: Raman spectra of varying concentration of sonicated fibrinogen background corrected using EMSC algorithm B: Percent variance explained by the latent variables C: PLSR coefficient plotted from the sonicated fibrinogen data set shows strong fibrinogen features, D: Linear predictive model built from the PLSR analysis showing correlation between concentration and peak intensity.

The spectrum of sonicated fibrinogen after background correction using the EMSC algorithm with polynomial of order 3 displays strong fibrinogen features with higher intensity over the same concentration range, compared to the non-sonicated fibrinogen samples (Figure 4.5A). Applying PLSR, it is clear from Figure 4.5B that a total of six components made significant contributions to explain the variance in the sonicated fibrinogen spectra. Based on the percent variance explained, six latent variables were used to build the prediction model. The PLSR
coefficient plot shows signature peaks of fibrinogen, indicating that the prediction was based on variation in the fibrinogen spectral intensities (Figure 4.5C). A linear prediction model was defined from the data set, showing correlation between the Raman peak intensity and concentration (Figure 4.5D). The minimum value of RMSEP is found to be 0.0615mg/mL. The reduction in the RMSEP value recorded for fibrinogen data after sonication indicates that the accuracy of the model increases as a result of the improved solubility following sonication. Hence, it can be concluded that sonication improves the solubility of the fibrinogen and increases the spectral intensity, in turn leading to a considerable improvement in the predictive capacity of the model.

Simulated “pathological” plasma protein-mixture was prepared by varying the concentration of fibrinogen stock in order to achieve the physiologically relevant range from 0.5mg/mL to 5mg/mL and by maintaining the concentrations of albumin, cytochrome c and vitamin B12 constant at the normal concentrations in healthy human plasma. The concentrations for heart disorders (<3mg/mL) and liver disorders (<3mg/mL) have been deliberately included in the concentration range. The raw spectra of varying concentrations of fibrinogen in simulated plasma were smoothed by Savitzky–Golay, polynomial of 5, window 13 (Figure 5.6).
Figure 4.6: Smoothed spectra of varying concentration of fibrinogen in simulated plasma (0.5mg/mL to 5mg/mL). The arrow indicates the order of increasing concentration.

The arrow indicates that both the background and spectral features themselves decrease with increasing concentration of fibrinogen. However, noting that albumin is the dominant contributor to the Raman signal, and that fibrinogen is the dominant scatterer, this can be understood as a (fibrinogen) concentration dependent loss of (albumin) Raman scattering.

The PLSR coefficient obtained after pre-processing the data using the EMSC based algorithm shows an inverse spectrum of albumin rather than fibrinogen, as shown in figure 4.S5 in supplemental. As in the case of the water dispersions, the dominant effect of increasing concentrations of the poorly soluble fibrinogen is the scattering of the dominant Raman spectrum. Hence, although the predictive model built from this dataset shows a good correlation with fibrinogen concentration, it is not based on the characteristic spectroscopic signature of fibrinogen, and the variation of the albumin signal could equally be due to any other scatterer.
Ultra-filtration using 100kDa centrifugal filters failed to separate fibrinogen from the rest of the protein in the protein mixture. Figure 4.S6 shows that the Raman spectrum of the concentrate obtained has pronounced characteristic albumin features at 899 cm\(^{-1}\) and 1102 cm\(^{-1}\). Ion exchange chromatography was therefore explored as an alternative method for fibrinogen separation from the protein mixture, based on its charge. Carboxymethyl-cellulose acts as a weak cationic exchanger and fibrinogen is eluted out by altering the net charge of the bound protein, and thus its matrix binding capacity. Fibrinogen was detected in the unbound fraction. Albumin was not detected in the unbound fraction by Raman spectroscopy and it is concluded adsorption of the albumin fraction to the carboxymethyl cellulose resin occurred at the pH values employed. Other studies have shown carboxymethyl cellulose may form insoluble complexes with serum albumin [45].

Fibrinogen was extracted from the protein mixtures over the full concentration range, and Raman spectra were recorded from the separated fibrinogen and EMSC was performed on the data set before doing PLSR analysis. In the absence of sonication the prediction model performed poorly, due to the high degree of scattering, as seen in figure 4.S7. Mild sonication can be employed to improve the solubility of and reduce the scattering from fibrinogen, and thus the performance of the prediction model.
The spectrum of sonicated fibrinogen separated by ion exchange chromatography after background correction using the EMSC algorithm displays strong fibrinogen features. In Figure 4.7B, it is clear that nine components made significant contributions to the variance in the sonicated fibrinogen spectra. The minimum value of RMSEP is found to be 0.0568mg/mL. The PLSR coefficient plot shows the signature peaks of fibrinogen (Figure 4.7C), indicating that the linear prediction model obtained was based on the correlation between the Raman spectral intensities of fibrinogen and concentration (Figure 4.7D). Hence,
it can be concluded that ion exchange chromatography can successfully separate fibrinogen for Raman analysis from the protein mixture within 30 minutes and an accurate prediction model can be built from the Raman data to detect subtle changes in the fibrinogen concentration. Early detection of fibrinogen concentration could help to prevent disorders that are associated with increased fibrinogen level in plasma such as thromboembolism [46], various cardiovascular events and post-surgical arterial re-occlusion [47].

4.5 Discussion

In monitoring biological molecules in their native aqueous state in biofluids, Raman spectroscopy offers the potential advantage over other spectroscopic techniques such as infrared absorption, that water was a relatively low scattering cross section. However, applications of the technique face several challenges related to detection of relatively low concentrations and variations of concentrations of analytes, and low quality signals from poorly dispersed components, and there remains a considerable number of issues relating to the fundamental process of recording and extracting the spectral details using chemometric techniques.

Raman analysis in the inverted geometry using a water immersion objective is found to be the optimal method to record well defined spectra with minimal background, and notably samples of volumes as low as 1µL can be measured. In a sample set of varying concentrations over physiologically relevant ranges, the albumin contributions to the spectrum dominate over those of the water, and, after minimal preprocessing, PLSR can be employed to establish a regression model whose predictive performance shows a close correlation between the concentrations of the proteins and the Raman spectral profile. However, in a the more
complex simulated plasma mixture of proteins, improved data preprocessing techniques are required to account for the increased spectral background.

Although the broad background to Raman spectra is often attributed to fluorescence, this cannot be the case for materials with are nonresonant at the Raman source wavelength. Proteins such as albumin and fibrinogen can, however, contribute to stray Mie scattered light by causing diffusely scattered radiation that is not well collimated by the collection objective of the Raman microscope, enters the spectrometer effectively as stray light, and is dispersed across the detector [21]. The rubberband pre-processing method appeared to efficiently remove the background from the data set of varying concentration of albumin in water, but failed to satisfactorily deal with the background of varying concentrations of albumin in the simulated plasma protein mixture. The more sophisticated EMSC based algorithm helped eliminate the scattering associated with the albumin data in the simulated plasma, improving the prediction model, and also helped to extract the spectral features of fibrinogen from water.

In both cases, before subtraction, the primary effect of varying the protein concentrations was to decrease the contribution of the dominant Raman scatterer, which can be understood in terms of the presence of the poorly soluble, highly Mie scattering fibrinogen component. This proposed method can be efficiently used to detect albumin as a standard biomarker for detecting diseases associated with hypoalbuminemia (<30mg/mL), such as liver diseases, gastrointestinal protein loss, edema and hyperalbuminemia (>30mg/mL), such as severe dehydration and abnormal increase in body fat [48], [49]. The accuracy of the proposed method is comparable to that of the most commonly used method for detecting albumin from biological fluids, the enzyme linked immunosorbent assay (ELISA) [50][51], which is
sensitive and selective but is very time consuming and requires extensive sample preparation steps. 

In varying concentrations of fibrinogen in aqueous solution, the Raman signal of the water itself is diffusely scattered, increasingly so with increasing fibrinogen concentration, and thus the PLSR identifies a decreasing Raman contribution of water as the dominant concentration dependent effect. In the case of albumin in the simulated protein mixture, a concentration dependent Mie scattering of the Raman signal of albumin itself is the dominant effect of increasing albumin concentration. While one would expect a linear concentration dependent increase in the Raman signal of albumin, the inability of the ultra-filtration technique to separate the two high molecular weight proteins may suggest an interaction between the albumin and fibrinogen, such that increased albumin Raman scattering is overwhelmed by increased Mie scattering. 

Mild sonication is seen to improve the dispersion of fibrinogen in aqueous solutions, and significantly improve the Raman signal. Removing the water contribution using EMSC is seen to significantly improve the predictive model (Figure 4.5). 

Separation of the fibrinogen by ion exchange chromatography from the plasma protein mixture and application of the ultrasonication technique to reduce aggregation helped to detect fibrinogen features from the plasma solution even at a concentration as low as 0.5mg/mL. The RMSEP of 0.0568mg/mL compares favourably with similar observations, for example for attenuated total reflection – Fourier transform infrared absorption monitoring of glucose in blood serum [29]. The accuracy of this study is closer to that of the most commonly used gold-standard method i.e, the Clauss assay, which has a detection limit of ~0.4 mg/mL [52]. The Clauss assay is relatively time consuming and suffers from
inconsistencies in the results due to calibration standards, methodologies and variation in the reagents from various manufacturers [46]. These steps are relevant only in the case of human plasma and can be avoided while working with human serum as fibrinogen is absent in the serum. Notably, although this study has targeted the HMWF proteins to highlight the associated challenges, the optimised protocol can be adapted to detect low molecular weight proteins or other biomarkers in bodily fluids, which may have additional diagnostic potential, after depletion of the abundant proteins [13].

Ion exchange chromatography is a quick method to separate the proteins from each other by altering their net surface charge, making it an ideal tool for separating all the protein constituents and a better alternative to ultra-filtration. In this case, ultra-filtration failed to separate HMWF proteins from one another, as they tend to form hydrophobic bonds and nonspecific binding interactions with the membrane material (Fig 4.S6). However, the ion exchange chromatographic method has to be tailored to the specific protein, depending on its charge, and cannot be applied as a ‘one-for-all’ separation kit for all the proteins.

It is worth noting that PLSR seeks to correlate systematic variations in the spectroscopic profiles with the external target variable, in this case protein concentration. As such the method is not inherently specific to any molecular variations, in the case where multiple species vary simultaneously over the same range. It has been demonstrated, in Raman microspectroscopic studies of the action of drugs in cells in vitro, that independent regression of the same dataset against the applied dose and resultant viability of the cell population yielded could be employed to yield information about independent processes in the cells [53], [54]. However, where the study of two simultaneously varying molecular species may require
further analysis of the PLSR co-efficient, and/or regression over different spectral regions, which contain specific markers of the respective species of interest.

4.6 Conclusions

The potential advantages of using vibrational spectroscopy for disease diagnosis based on bodily fluids have been extensively explored over the last two decades. However, little consideration has been given to date to the optimisation of a Raman analysis protocol involving proteins in their native aqueous state, leading to irreproducible results due to high complexity of the plasma proteins. This study is a proof of concept that Raman spectroscopy can be successfully used to detect subtle changes in individual plasma protein concentration from simulated plasma samples to disease diagnostics purposes.

It has been shown that measurement in the inverted geometry using a water immersion objective yields high quality spectra and the sample volume can be as small as 1μL. This experimental set up is advantageous for clinical purposes where the volumes of patient samples are minimal. In the simulated plasma protein mixture, the poorly soluble fibrinogen component was seen to obscure the systematic variations of the protein concentrations, due to the high degree of scattering. Extraction of the fibrinogen by ion exchange chromatography is seen to be more specific than by ultra-filtration, such that the variations of fibrinogen levels themselves can be quantified. In general, the scattering problems caused by fibrinogen favour the use of blood serum for the analysis of the remaining lower molecular weight fractions.

However, to further ensure relevancy and consistency of these results, experiments need to be carried out in pooled plasma/serum. The use of Raman spectroscopy coupled with
chemometric techniques not only gives a mere estimate of whether the protein levels are high or low but also gives higher accuracy of quantification. Once appropriate experimental methods are established, a hypothesised point-of-care device that can be used in real clinical applications for spectroscopic analysis of body fluids can be realised. The proposed approach can be expeditiously employed for early detection of pathological disorders associated with high or low plasma proteins.
4.7 References


[22] I. O. Tekin, B. Pocan, A. Borazan, E. Ucar, G. Kuvandik, S. Ilikhan, N. Demircan,


Figure 4.S1. A: Raw Raman spectra of varying concentrations of albumin in simulated plasma (5mg/mL – 50mg/mL). The arrow indicates the order of increase in concentration. B: Percent variance explained by the latent variables, C: PLSR component showing the inverse albumin features and D: Linear predictive model built from the PLSR analysis.
Figure 4.S2A: Raman spectra of varying concentrations of fibrinogen in distilled water corrected using the “rubberband” method (0.5mg/mL – 5mg/mL), B: Percent variance explained by the latent variables, C: PLSR coefficient plotted from the data set shows negative peaks and no fibrinogen features, D: Linear predictive model built from the PLSR analysis
Figure 4.S3A: Spectra corrected by EMSC algorithm of varying concentration of fibrinogen (0.5mg/mL – 5mg/mL), B: Percent variance explained by the latent variables, C: PLSR coefficient showing strong fibrinogen features, D: Linear predictive model built from the PLSR analysis.
Figure 4.S4. Raman spectra of fibrinogen stock recorded before (Black) and after (Red) sonication. Sonication helped to increase the overall intensity of the Raman signal by increasing the solubility of the protein.

Figure 4.S5 A: EMSC corrected data of varying concentrations of fibrinogen (0.5mg/mL to 5mg/mL), B: percent variance explained by the latent variables, C: PLSR coefficient showing
the inverse peaks of albumin (1089\textsuperscript{cm\(^{-1}\)} and 1102\textsuperscript{cm\(^{-1}\)}), and D: The predictive model built from the dataset

![Raman spectrum](image)

**Figure 4.56.** Raman spectrum of the concentrate obtained after ultra-filtration of simulated plasma using 100kDa filters showing albumin features (899\textsuperscript{cm\(^{-1}\)} and 1102\textsuperscript{cm\(^{-1}\)})
Figure 4.87A: EMSC corrected spectra of varying concentrations of fibrinogen separated by ion exchange chromatography (0.5mg/mL to 5mg/mL), in the absence of sonication B: percent variance explained by the latent variables, C: PLSR coefficient showing inverse fibrinogen peak, and D: The predictive model built from the dataset.
Chapter 5

Analysis of bodily fluids using Vibrational Spectroscopy: A direct comparison of Raman scattering and Infrared absorption techniques for the case of glucose in blood serum

The following chapter has been reproduced from the published journal article entitled ‘Analysis of bodily fluids using Vibrational Spectroscopy: A direct comparison of Raman scattering and Infrared absorption techniques for the case of glucose in blood serum’, Analyst, 2019, 144, 3334-3346

Author list: Drishya Rajan Parachalil, Franck Bonnier, Clément Bruno, Hélène Blasco, Igor Chourpa, Matthew J. Baker, Jennifer McIntyre, and Hugh J. Byrne

DRP performed all experimental analysis and authored the publication. CB, HB and IC assisted clinical aspects of the study, FB provided the samples, assistance in harmonising the Raman and IR analysis protocols. FB, MJB, JMcC and HJB input to conceptual design of the work and drafting and proofing of the manuscript
5.1 Abstract

Analysis of biomarkers present in the blood stream can potentially deliver crucial information on patient health and indicate the presence of numerous pathologies. The potential of vibrational spectroscopic analysis of human serum for diagnostic purposes has been widely investigated and, in recent times, infrared absorption spectroscopy, coupled with ultra-filtration and multivariate analysis techniques, has attracted increasing attention, both clinical and commercial. However, such methods commonly employ a drying step, which may hinder the clinical work flow and thus hamper their clinical deployment. As an alternative, this study explores the use of Raman spectroscopy, similarly coupled with ultra-filtration and multivariate analysis techniques, to quantitatively monitor diagnostically relevant changes of glucose in liquid serum samples, and compares the results with similar analysis protocols using infrared spectroscopy of dried samples. The analysis protocols to detect the imbalances in glucose using Raman spectroscopy are first demonstrated for aqueous solutions and spiked serum samples. As in the case of infrared absorption studies, centrifugal filtration is utilised to deplete abundant analytes and to reveal the spectral features of Low Molecular Weight Fraction analytes in order to improve spectral sensitivity and detection limits. Improved Root Mean Square Error of Cross Validation (RMSECV) was observed for Raman prediction models, whereas slightly higher $R^2$ values were reported for infrared absorption prediction models. Summarising, it is demonstrated that the Raman analysis protocol can yield accuracies which are comparable with those reported using infrared absorption based measurements of dried serum, without the need for additional drying steps.
Human bodily fluids (e.g. blood serum/plasma, urine, saliva, tears and cerebrospinal fluid) are considered to be a rich reservoir of clinical biomarkers and are an interesting alternative to cells and tissues in terms of disease diagnosis and prognosis, owing to advantages such as minimal invasiveness, low cost, and rapid sample collection and processing (1–4). The biochemical composition of human serum can provide crucial information on patient health and indicate the presence of numerous pathologies, as it encompasses a vast range of proteins and biochemical products accumulated while perfusing various organs (5). Moreover, alterations in the biochemical composition of the serum/plasma could reflect changes of physiological states due to disease, enabling early disease diagnosis and treatment (6–8). For example, it is reported that carcinoembryonic antigens, CA 15-3 and CA 27.29, can be considered as serum biomarkers for breast cancer (9), KL-40 and MMP-9 have been found to be potential serum biomarkers for high grade gliomas (10), prostate specific antigen was found in higher levels in the serum of patients with prostate cancer (11,12), and carcinoembryonic antigen has been reported to be serum biomarker for colorectal cancer (13). In recent decades, extensive studies have explored the detection of serum biomarkers for various diseases, and quite naturally the reference tools used are the conventional, complex analytical techniques such as chromatography, electrophoresis or mass spectroscopy (14–18). More recently, the field of serum proteomics has exploded in the literature, as this emerging field has gained world-wide attention due to its potential to reveal important information regarding the pathogenesis or progression of diseases (19–22). Analysis of serum biomarkers based on protein content is inherently challenging, because of the low concentrations and the vast variety and dynamic range of protein abundance. The
complex nature of the serum poses as a huge problem for the detection of small molecule biomarkers (2,23). Human serum contains more than 10,000 different proteins, with an overall concentration ranging from 60-80mg/mL. Furthermore, circulating species present in the serum, such as metabolites, peptides, sugars, and lipids, add to its complexity. Conventional proteomic methods struggle to handle large dynamic range of abundances of its constituent components (17). The characteristics of serum are usually dominated by the high molecular weight fraction (HMWF) of proteins, which includes albumin (57-71%) and globulin (8-26%) and masks the features of low molecular weight analytes which are present in trace amounts (<5%) (24). Chromatographic fractionation can significantly enhance the sensitivity and specificity of the data recorded, and coupling with spectroscopic analysis techniques such as IR absorption in hybrid techniques such as Liquid Chromatography (LC-IR) or Gas Chromatography (GC-IR) (25) can improve the performance further. Such chromatographic techniques, although extensively used in pharmaceutical, chemical or food science applications are time consuming and costly, however (26–28), and commercially available centrifugal filters have been employed to deplete the HMWF before analysis to facilitate analysis of the low molecular weight fraction (LMWF), including proteins and molecular biomarkers (29).

Vibrational spectroscopic techniques, both Raman and infrared absorption, have emerged over the past 20 years as increasingly routine analytical techniques for a wide range of applications, as they reveal specific biochemical information without the use of extrinsic labels. Although they are often considered complementary techniques (30) there are also specific considerations for each, for specific applications such as measurement of bodily fluids, as described by Bonnier et al. (31). They provide intrinsic vibrational signatures of
the material of interest in a non-destructive fashion, and the potential for diagnostic applications has been well demonstrated, notably in human serum and plasma (7,32–34). However, although both Raman, Fourier-Transform Infrared (FTIR) and Attenuated Total Reflectance-FTIR (ATR-FTIR) spectroscopy have been widely explored to study bodily fluids over the last two decades, most of these studies have been carried out on air dried samples, in order to avoid the water contribution in the case of FTIR, and to increase the concentration of the analytes in the case of Raman (35–40). Two major limiting factors in the use of dried samples are the drying time (41) and also the so-called “coffee-ring” effect, or, specifically in terms of blood serum, the Vroman effect (42–44), whereby different analytes precipitate from solution at different rates, giving rise to variations in the spectral features due to chemical and physical inhomogeneity. Previous studies have clearly shown that the IR spectrum of dried (aggregates of) molecular species is not the same as that in solution (45). It has further been shown that, for dried deposits from solutions of varying concentrations of analyte, the linearity of the Beer-Lambert law, and therefore the quantitative nature of the measurement is compromised as the concentration of the analyte is increased (45). In the case of the “coffee ring effect”, the thickness, and therefore the quantitative accuracy of the measurement is spatially inhomogeneous, and the technique is not ideally suited for quantitative measurements, such as those considered here. Notably, Spalding et al. have used a serum dilution technique to improve the reliability of the technique for quantitative measurement (46). The issues with inhomogeneity can potentially be overcome by micropipetting and sampling the whole drop, and there have been studies that show excellent specificity/sensitivity of classification of diseased state using dried samples (40,47,48). A study conducted by Hands et al. have shown that a 1 µL spot of serum
on an ATR-FTIR crystal takes 8 minutes to dry to a state where excellent spectra can be collected (49). Nevertheless, in terms of clinical translation, the requirement of a drying step adds considerably to the sampling time and complexity of the workflow (41).

Protocols for monitoring changes in HMWF serum proteins in their native liquid form using Raman spectroscopy have recently been demonstrated, however, without the need for drying (50). Analysing in the native liquid state, the chemical composition is averaged out by molecular motion over the measurement time, greatly reducing the variability of the measurement. Bonnier et al. compared FTIR and Raman for measuring gelatin in solution and described protocols for isolation of LMWF from serum (31). This study indicated that Raman in the liquid form could deliver similar sensitivities to ATR-FTIR for measurement of LMWF species.

The aim of this study is to investigate the sensitivity and accuracy of Raman spectroscopy in the liquid state, coupled with centrifugal filtration, as a biochemical tool to detect clinically relevant changes in the biochemical composition of serum and compare to the ATR-FTIR technique, using the specific example of glucose. The study is specifically designed according to the ATR-FTIR study of glucose in serum by Bonnier et al., using identical parameters and protocols of ultra-filtration and multivariate regression analysis, such that a direct comparison of the two techniques can be made (29). The protocol is first demonstrated using aqueous solutions and human serum spiked with systematically varied concentrations of glucose, before exploring the sensitivity and accuracy of the technique in patient samples.
5.3 Materials and Methods

5.3.1 Preparation of varying concentration of glucose in distilled water model

D-glucose (G8769) was purchased from Sigma Aldrich, Ireland and 6 glucose solutions were prepared over the concentration range 100mg/dL to 1000mg/dL. Amicon Ultra 0.5mL centrifugal filter devices (Millipore- Merck, Germany), with 10kDa cut off point, were employed to concentrate and fractionate the serum samples. The centrifugation procedure previously reported by Bonnier et al. was followed (31). A further study published by Bonnier et al. reported 100% recovery of LMWF using 10kDa, hence only 10kDa cut off filtration has been used in the present study (45). Pre-rinsing of the filter devices with 0.1M NaOH prior to plasma analysis is essential to avoid glycerine interference in the analysis (32). The optimised washing and rinsing procedure includes spinning 0.5mL 0.1M NaOH at 14000×g for 30 minutes followed by three rinses with distilled water by spinning 0.5mL distilled water for 30 minutes at 14000×g. Every 30 minute wash and rinse must be followed by spinning the device in the inverted position at 1000×g for 2 minutes, to remove the residual solution contained in the filter. After washing, 0.5mL of glucose solution is transferred to the 10kDa filter and centrifuged at 14000×g for 30 minutes. The solution that passes through the 10kDa filter is the filtrate, which contains mostly water and molecules smaller than 10kDa. The remainder of the serum, known as the concentrate, is collected by placing the filter device upside down and spinning for 1000×g for 2 minutes. The resultant concentrate, ~50µL, contains molecules with molecular weight larger than 10kDa, concentrated by a factor of ~10, and can be employed for study of the HMWF (50). All the filtrate solutions were analysed using Raman spectroscopy and five replicate measurements from different positions
have been performed for each sample of ~50µL. In subsequent analysis, each patient is represented by all the spectra recorded from that patient, rather than the mean.

5.3.2 Preparation of glucose spiked in serum model

Sterile, human serum (H6914) and D-glucose (G8769) were purchased from Sigma Aldrich, Ireland, for the preparation of in vitro spiked models. The commercial human serum was spiked with glucose in the concentration range that matches the study of Bonnier et al. (29). Since glucose is already present in normal human serum at concentrations of 70-110mg/dl (51), the final concentration of the spiked samples covers the physiologically relevant ranges of normal (80-120mg/dL) (52), and hyperglycaemia (>120mg/dL). The centrifugal processing step described in Section 2.1 was performed on the spiked serum samples to obtain the filtrate. Note, in the analysis of Figure 3, the regression is performed over spiked, rather than total glucose concentration, consistent with the approach of Bonnier et al. (29).

5.3.3 Glucose levels in patient serum samples

Patient serum samples were donated by the University Hospital CHU Bretonneau de Tours (France) and the ethical procedures were followed. The blood samples were collected from the individuals as routine blood check-ups and 1 mL per patient was provided for spectroscopic analysis. A total of 25 patient samples were included in the present study. Samples were collected by personnel of the University Hospital CHU Bretonneau de Tours, under standard clinical protocols and ethical procedures approved by the hospital. The samples were serologically profiled, for other purposes, and the anonymised, residual discard
samples, along with their serological profiles were donated to the Université François-
Rabelais de Tours, for further study. No further specific ethical approval or patient consent
is required. Glucose concentrations were obtained by routine biochemical analysis using a
COBAS analyser, following the in house guidelines for routine biochemical analysis. The
principle of the test is based on the enzymatic reference method with hexokinase, which
catalyses the phosphorylation of glucose to glucose-6-phosphate by ATP methods (53,54).
Subsequently, glucose-6-phosphate is oxidised by glucose-6-phosphate dehydrogenase, in
the presence of NADP, to gluconate-6-phosphate. This reaction is specific, with no other
carbohydrate being oxidised. The rate of NADPH formation during the reaction is directly
proportional to the glucose concentration and is measured photometrically in the UV.
Measured glucose levels in the patient samples are listed in Table 6.1, and can be seen to
cover the range 55-435 mg dL\textsuperscript{-1}. Note, that the distribution of glucose levels covers a broader
range than that of the study of similar (n=15) patients by Bonnier \textit{et al.} (61-208 mg dL\textsuperscript{-1})
(29).
Table 5.1. List of measured glucose levels in patient samples. Glucose blood levels are quoted in terms of the SI unit of mmol L\(^{-1}\) as well as mg dL\(^{-1}\), commonly used in serology literature and in the study of Bonnier et al. (29)

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Glucose blood levels</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mmol L(^{-1})</td>
<td>mg dL(^{-1})</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.9</td>
<td>52.25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>55.85</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.9</td>
<td>70.27</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.9</td>
<td>70.27</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.2</td>
<td>75.67</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4.2</td>
<td>75.67</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.3</td>
<td>77.47</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4.5</td>
<td>81.08</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5.1</td>
<td>91.89</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5.1</td>
<td>91.89</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>5.2</td>
<td>93.69</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>5.7</td>
<td>102.70</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>109.10</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>6.1</td>
<td>109.90</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>6.4</td>
<td>115.31</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>6.4</td>
<td>115.31</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>7.2</td>
<td>129.72</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>8.8</td>
<td>158.55</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>9.9</td>
<td>178.37</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>11.5</td>
<td>207.20</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>11.7</td>
<td>210.81</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>13.4</td>
<td>241.44</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>15.7</td>
<td>282.88</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>15.8</td>
<td>284.68</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>24.1</td>
<td>434.23</td>
<td></td>
</tr>
</tbody>
</table>
5.3.4 Data collection using Raman spectrophotometer

The measurement conditions used for screening HMWF proteins in solution have recently been detailed (50). Raman spectra of all the liquid samples were recorded at stabilised room temperature (18°C) using a Horiba Jobin-Yvon LabRam HR800 spectrometer with a 16-bit dynamic range Peltier cooled CCD detector. The spectrometer was coupled to an Olympus 1X71 inverted microscope and a x60 water immersion objective (LUMPlanF1, Olympus) was employed. The substrate used was a Lab-Tek plate (154534) with a 0.16-0.19mm thick glass bottom, 1.0 borosilicate cover glass, and was purchased from Thermo Fischer Scientific, Ireland. Since the scattering efficiency is inversely proportional to the fourth power of the wavelength, it is always desirable to use a short wavelength Raman source. In the filtered serum, there are no molecular species which are resonant with 532nm, and therefore neither fluorescence nor photodamage are limiting factors.

5.3.5 Data pre-processing and analysis

The raw spectra were subjected to pre-processing techniques in Matlab before further analysis, to remove the background signal and reduce the noise. Smoothing of the raw data was done using the Savitzky–Golay method, with a polynomial order of 5 and window 13 and the Extended Multiplicative Signal Correction (EMSC) algorithm was applied to the smoothed spectra to remove the underlying water spectrum, whose OH bending feature at 1640 cm\(^{-1}\) can interfere with the protein spectra, particularly at low concentrations (55).
The principle of EMSC for subtraction of a specific measureable background spectrum and the associated Matlab codes have previously been published by Kerr and Hennelly, 2016 (55), and their description has been adapted in the following. In the case of measurement in aqueous solution, the raw sample spectrum, $S$, consists of Raman spectrum of the analyte of interest, $R$, a baseline signal, $B$, and the water signal, $W$.

$$S = R + B + W$$ (55)  

The Raman spectrum of interest can be represented by a reference spectrum of the analyte of interest, $r$, and it can be assumed that $R$ is the product of this reference spectrum and a certain scalar weight, $C_r$, which describes the concentration dependence (56,57).

$$R \sim C_r \times r$$ (55)  

Similarly, a spectrum, $w$, is recorded from water directly in order to represent the spectral contribution of water in $W$, as the product of pure water spectrum and a certain scalar weight.

$$W = C_w \times w$$ (55)  

The baseline, $B$, is now represented by a polynomial:

$$B_N = C_0 + C_1X + C_2X + \ldots + C_NX^N$$ (55)  

where $N$ is the order of polynomial and $C_m$ for $m = 0 \rightarrow N$ represents various coefficients of polynomial (58). The EMSC algorithm is used to obtain estimates of the scalar values $C_r$, $C_m$ and $C_w$. These estimates are obtained from an optimal fit of the various vectors in Equation 5.

$$S \sim [C_r \times r] + [C_w \times w] + [\sum_{m=0}^{N} C_mX^m]$$ (55)  

The background corrected, concentration dependent analyte spectra, $T$, can be represented as:

$$T = \frac{S - [C_w \times w] - [\sum_{m=0}^{N} C_mX^m]}{C_w}$$ (55)
Note, that division by $C_w$ has the effect of scaling the analyte spectra, assuming a constant water contribution to all sample spectra. The Raman spectrum of the stock glucose solution (450g/L) is used as the reference for EMSC, and polynomial of order 5 was used in all cases. The glucose reference spectrum and water spectrum used in EMSC were recorded using the 532nm laser line and Lab-Tek plate as substrate. The x60 objective brings the laser to a focus within the liquid, beyond the thin (0.16mm-0.19mm) glass substrate. No significant contribution from the glass to the recorded spectrum was apparent, and therefore no correction was required. The Rubberband method (59) was used to baseline correct the glucose reference spectrum after smoothing it using the Savitzky–Golay method.

5.3.6 Partial Least Squares Regression

Partial Least Squares Regression (PLSR) was employed to establish a model that relates the variations of the spectral data to a series of concentrations. This regression model can be used to improve the limit of detection of Raman bio-sensing (60). Constructed based on the spectra of samples of known glucose content, either solutions of varying concentrations of glucose (in water or in commercial serum), or those of the patient serum, the model is then validated using a rigorous cross validation procedure which evaluates its performance in accurately predicting glucose concentrations. A 20 fold cross validation approach has been employed to validate the robustness of the method. This approach involves randomly dividing the set of observations into approximately equal size, 50% of the spectral data were randomly selected as test set, while the remaining 50% is used as the training set (61). In the current case, (5 x 25 spectra) were divided into two groups of 65 (test) and 60 (training) spectra. The cross-validation process is then repeated 20 times (the folds), with all
observations are used for both training and testing, and each observation is used for testing exactly once. The results from the folds can then be averaged to produce a single estimation. The Root Mean Square Error of Cross Validation (RMSECV) is calculated from the 20 iterations to measure the performance of the model for the unknown cases within the calibration set. The correlation between the concentration and spectral intensity is given by the $R^2$ value. The standard deviation was calculated to find the variation between each spectrum calculated from the same sample. The number of latent variables used for building the PLSR model is optimised by finding the value that is equivalent to the minimum of the RMSECV.

5.4 Results

In this study, Raman spectra of the samples were recorded in the inverted geometry using a x60 water immersion objective with a 532nm laser and the Lab-Tek plate was used as the substrate. The 532nm laser was chosen as it is compatible with the thin glass bottomed Lab-Tek plate and provides a strong Raman signal of the sample with minimal background interference. This set-up was previously reported by Bonnier et al. (31) to yield better analysis of serum using Raman spectroscopy when the sample was analysed in the inverted geometry using a 785nm laser and CaF$_2$ substrate. The added advantage of this setup is that it provides high quality, consistent Raman spectra from sample volumes as low as 1μL. The protocol using 532nm was more recently further explored for analysis of HMWF serum proteins (50).

Figure 5.1 presents the spectra of the fingerprint region of the pure glucose solution recorded in the inverted geometry. The raw spectra of the glucose were baseline corrected using the
rubberband method and smoothed using the Savitzky–Golay algorithm (polynomial 5, window 13). Example signature peaks of glucose (indicated by asterisks) appear at ~450 cm\(^{-1}\), associated with an endocyclic \(\delta(C-C-O)\) ring mode, \(\delta(C1-H1)\) vibration at 911 cm\(^{-1}\), a peak at 1060 cm\(^{-1}\) due to \(\nu(C1-OH)\) stretching, a relatively sharp peak at 1125 cm\(^{-1}\) which can be assigned to the \(\delta(C-O-C)\) angle-bending mode and the sharp peaks at 1340 cm\(^{-1}\) and 1460 cm\(^{-1}\), related to the \(\delta(C-C-H)\) vibration and pure CH\(_2\) group vibration, respectively(62).

![Figure 5.1](image)

**Figure 5.1.** Raman spectrum of an aqueous glucose solution, concentration 450 g/L. Example signature peaks at 450 cm\(^{-1}\), 911 cm\(^{-1}\), 1125 cm\(^{-1}\), 1340 cm\(^{-1}\) and 1460 cm\(^{-1}\) are labelled in the figure.

### 5.4.1 Monitoring the concentration dependence of glucose in distilled water

In order to establish the analysis protocol for the patient samples, a PLSR prediction model was first built and applied to the set of varying concentrations of glucose in distilled water, as well as spiked into serum. The first step in this study was to optimise the measurement
protocol and also to evaluate the efficacy of the centrifugal filtration technique in separating and concentrating glucose from the HMWF proteins. For this, different amounts of pure glucose were spiked into the distilled water. The normal glucose concentration and the concentration and hyper-glycaemia were deliberately included to simulate physiologically relevant concentrations.

Figure 5.2 (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of varying concentrations of glucose (5 x100mg/dL, 5 x 450mg/dL and 5 x 1000mg/dL, offset for clarity), in distilled water and signature peaks of glucose are highlighted with asterisks, (B): Evolution of the RMSECV on the validation model, (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted
with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and $R^2$ values were calculated as 10.93mg/dL and 0.9705 respectively.

Figure 5.2 A displays examples of the EMSC (polynomial order 5) corrected low (5 x 100mg/dL), medium (5 x 450mg/dL) and high (5 x 1000mg/dL) datasets, showing glucose features (indicated by asterisks) of greater intensity at higher concentration, whereas the lower concentrations shows weaker features of glucose. The groups of spectra are off set for clarity. The concentration 1000mg/dL was deliberately included in the dataset to evaluate the consistency of the glucose spectral features as the concentration increases from 100mg/dL to 1000mg/dL. In order to analyse the spectral variations and the glucose concentrations, the PLSR algorithm was applied.

Based on the percent variance explained by the latent variables and the minimum value of RMSECV (Figure 5.2B), the optimum number of latent variables to reach the best performance is determined to be 4. The PLSR coefficient plot displayed in Figure 5.2C confirms the correlation of the data in Figure 5.2D is based on glucose features, such as the peaks at $\sim$1060cm$^{-1}$, $\sim$1125cm$^{-1}$, 1450cm$^{-1}$ and $\sim$1340cm$^{-1}$. Finally, after selecting the optimum number of components for the data set analysed, a predictive model is built from the PLSR analysis (Figure 5.2D), to compare the observations to the known concentrations of glucose in the samples with the estimated concentrations from the spectral data sets. Figure 5.2D indicates that a satisfactory linear model could be obtained with the raw data set and that the concentration dependence of the sample set is conserved by centrifugal filtration. From the results shown in figure 5.2, it is evident that the predicted values are in good agreement with the reference concentrations and the corresponding correlation coefficient
(R^2) is calculated as 0.9705. Note, each concentration point has five independent measurements, and the mean standard deviation of each measurement is 4.8mg/dL. The RMSECV calculated from the 20 iteration of cross validation is 10.93mg/dL, thereby indicating that PLSR provides accurate predictions for the glucose Raman data over the entire concentration range of 100mg/dL to 1000mg/dL.

5.4.2 Monitoring the glucose concentration in spiked serum

In an attempt to extend the optimised protocol to a more complex environment, PLSR analysis was performed on the EMSC corrected data set recorded from the filtrate of the serum samples with spiked glucose concentrations varying from 0mg/dL to 220mg/dL. When the entire fingerprint region was selected for PLSR analysis, the resultant PLSR coefficient displayed a negative peak at ~1000cm\(^{-1}\) which could potentially derive from other LMWF species such as urea (63) (Figure 5.5.1C in Supplemental). Therefore, the spectral region from 1030cm\(^{-1}\) to 1400cm\(^{-1}\), which contains strong glucose features at ~1050cm\(^{-1}\) and at~1340cm\(^{-1}\) but minimal interference from urea, has been chosen for PLSR analysis (62). Figure 5.3A shows the glucose data set after background correction using the EMSC algorithm (offset for clarity by 5.0 units). The optimum number of latent variables were chosen by calculating the lowest value of RMSECV (Figure 5.3B). Six latent variables were chosen for this model and the resultant PLSR coefficient exhibits strong glucose features, as shown in Figure 5.3C. A linear predictive model can be defined from the EMSC corrected data set of varying concentration of glucose in serum Figure 5.3D. The RMSECV was found to be 1.66mg/dL and R^2 value was calculated as 0.9914. The mean standard deviation of each measurement
from 20 iterations of cross validation was calculated to be 3.2mg/dL. The results suggest that this optimised protocol can be applied to the patient samples to build a quantitative model.

![Graphs](image)

**Figure 5.3.** (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of glucose spiked in serum (spiked concentrations 5 x 0mg/dL, 5 x 120mg/dL and 5 x 220mg/dL, offset for clarity) and the signature peaks of glucose are highlighted by asterisks, (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R² values were calculated as 1.66mg/dL and 0.9914.
Bonnier et al. have previously illustrated the strategy of centrifugal filtration to fractionate the samples to eliminate the influence of HMW in order to screen potential LMWF biomarkers using glucose as a model analyte, measured by ATR-FTIR [33]. The results of the analysis of similarly centrifugally filtered, glucose spiked samples of commercial serum are reproduced in Table 5.2, and directly compared to the results of the current study using Raman spectroscopic analysis in the native liquid state. Although each method yields similar $R^2$ values, the RMSECV of Raman spectroscopic analysis is significantly lower, which indicates an increased accuracy of the measurement protocol in the liquid state.

**Table 5.2.** Comparison of the results of ATR-FTIR and Raman spectroscopic analysis of human serum spiked with varying concentrations of glucose.

<table>
<thead>
<tr>
<th>Measurement type</th>
<th>Concentration range (mg/dL)</th>
<th>RMSECV(mg/dL)</th>
<th>Standard deviation</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTIR (Min-Max normalised)</td>
<td>0-220</td>
<td>2.199</td>
<td>0.250</td>
<td>0.995</td>
</tr>
<tr>
<td>Raman spectroscopy</td>
<td>0-220</td>
<td>1.665</td>
<td>3.2</td>
<td>0.991</td>
</tr>
</tbody>
</table>

**5.4.3 Monitoring the glucose concentration in patient samples**

Figure 5.4A displays the Raman spectra from patient samples after performing background correction using the EMSC algorithm, in the reduced spectral range of 1030-1400 cm$^{-1}$. The spectra are offset for clarity by 5.0 units. The glucose bands at 1120cm$^{-1}$ and 1340cm$^{-1}$, related to the $\delta$(C-C-H) vibration and pure CH$_2$ group vibration, and the peak at 1060cm$^{-1}$ due to $\nu$(C1-OH) stretching can be clearly seen. Based on the minimum RMSECV value
(Figure 5.4B), 12 latent variables were found to be optimal for constructing a PLSR based model. The PLSR coefficient clearly shows glucose features (Figure 5.4C), indicating that the prediction is based on the variation in the glucose peak intensities.

**Figure 5.4.** (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of patient samples (5 x 52.25mg/dL, 5 x 75.67mg/dL, 5 x 93.69mg/dL, 5 x 210.81mg/dL and 5 x 434.35mg, offset for clarity) and the signature peaks are marked by asterisks, (B): Evolution of RMSECV of the data set, (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and $R^2$ values were calculated as 1.84mg/dL and 0.84 respectively.
In the prediction model of Figure 6.4D, the minimum value of RMSECV is 1.84mg/dL and the $R^2$ value is calculated as 0.82, indicating a high prediction capacity. The mean standard deviation was determined to be 3.48mg/dL, indicating acceptable repeatability between the cross validation iterations.

**Table 5.3.** Comparison of the results of ATR-FTIR (29) and Raman spectroscopic analysis of patient sample set for monitoring the glucose levels. FTIR results are normalised.

<table>
<thead>
<tr>
<th>Measurement type</th>
<th>Concentration range (mg/dL)</th>
<th>RMSECV (mg/dL)</th>
<th>Standard deviation (mg/dL)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATR-FTIR (Min-Max normalised)</td>
<td>61.25-210</td>
<td>3.1</td>
<td>1.90</td>
<td>0.9957</td>
</tr>
<tr>
<td>Raman Spectroscopy</td>
<td>52.25-210</td>
<td>1.6</td>
<td>2.31</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>52.25-440</td>
<td>1.84</td>
<td>3.48</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Table 5.3 directly compares the results of analysis of a similar patient sample-set, with glucose levels which varied over the range 0-210mg/dL, similarly centrifugally filtered and analysed using ATR-FTIR (29) and the results of the current study of patient samples over the same (patients 1-21, Table 5.1) and extended (patients 1-25, Table 5.1) range using Raman spectroscopic analysis in the native liquid state. It is noteworthy that Raman spectroscopy yields significantly lower values of RMSECV for all the Raman prediction models, suggesting higher sensitivity and accuracy. However, the standard deviation was found to be higher and the $R^2$ value was found to be considerably lower for the shorter as
well as larger concentration range. The reduced $R^2$ value and higher standard of deviation could be attributed to the variability in the spectral response of the patient samples. Nevertheless, the precision of the model expressed by the RMSECV values indicates the suitability of this technique to discriminate patients with very close concentrations of blood glucose. The results demonstrate that Raman spectroscopy is able to detect subtle variations in the glucose concentrations with similar accuracy in the native liquid state, to the ATR-FTIR method in dried samples.

In order for a glucose detection method to be viable, it should be able to detect glucose in the clinically relevant range (10-450mg/dL). Post PLSR analysis, the dataset is presented in the Clarke’s error grid (Figure 5.5), the most common standard for evaluating the performance of a glucose detection method used since 1987 (64).

![Figure 5.5. PLSR validation of patient samples on Clarke’s error grid. The RMSECV was found to be 1.84 mg/dL and $R^2$ value was calculated as 0.84](image)

Figure 5.5. PLSR validation of patient samples on Clarke’s error grid. The RMSECV was found to be 1.84 mg/dL and $R^2$ value was calculated as 0.84
Data points that fall in zone A and B are acceptable values. Values that fall outside A and B result in erroneous diagnosis. On the Clarke’s error grid of patient samples (Figure 5.5), 98% of the PLSR validation dataset falls within zone A and B, which is the zone of clinical accurate measurement with no effect on clinical actions. Error can be attributed to the intrinsic variability of patient samples which reflects their physiological state on the day. However, the results from this study are promising, indicating that Raman spectroscopy coupled with multivariate analysis and centrifugal filtration techniques can be used as a biochemical tool for detecting potential small biomarkers from human serum/plasma.

5.5 Discussion

The potential advantages of using vibrational spectroscopy for biomarker assessment in bodily fluids have been extensively explored in the last two decades. However, little consideration has been given so far to protocols involving Raman analysis in the native liquid state of proteins. Liakat et al. reported the in vitro prediction of physiologically relevant concentration of glucose using mid-IR transmission light with respect to a Clarke error grid (52). Near IR and Raman spectroscopy was also used for measurement of glucose from artificial plasma with high precision and accuracy (65). Although these are initial steps towards developing Raman spectroscopy into a biochemical tool for serum/plasma analysis, measurements should be performed on patient samples in order to ensure the relevancy of these methods. Using Raman spectroscopy as a biochemical tool, it has been possible to detect differences in peak intensities of altered serum compared to normal ones for glucose and lipid compounds (66), multicomponent blood analysis (67) and also to determine blood
glucose concentration of blood samples with above-physiological levels of glucose within 5 min (68).

Using glucose as a model analyte, this study successfully demonstrated the feasibility of employing Raman spectroscopy for detecting small biomarkers in serum after depletion of HMWF proteins. It has been shown that optimal experimental set up for Raman analysis for this experiment is Lab-Tek plates as substrate and measurement in the inverted geometry using water immersion objective and the sample volume can be as small as 1μL. This experimental set up is advantageous for clinical purposes where the volumes of patient samples are minimal. After the depletion of the abundant proteins, the dominant water peak from the filtrate collected after centrifugal filtration using 10kDa can be removed by using the EMSC algorithm, and PLSR analysis applied to obtain a prediction model relating the glucose concentrations and the intensity of glucose features. Even though the EMSC algorithm removed the underlying water spectra effectively, there could be interference from other LMWF analytes, namely, urea (7-20mg/dL). Thus, as presented in the present study, the spectral range from 1030cm⁻¹ to 1400cm⁻¹ was chosen for data analysis, as this region does not contain signature peaks of urea.

The depletion of HMWF proteins using centrifugal filtration to detect glucose in serum using ATR-FTIR was previously reported by Bonnier et al. (29). While the work carried out by Bonnier et al. showed excellent results, the requirement of a drying step is potentially a major drawback. Indeed, the drying process required for ATR-FTIR has been identified by Cameron et al. as a potentially significant impediment to translation of the technique to clinical applications (41). Since Raman is compatible with aqueous samples, sample drying can be avoided and data can be recorded from the native environment, which makes the
proposed method an ideal alternative to IR. The results summarised in Table 5.2 and Figure 5.3 suggest that Raman spectroscopy maintained high level of accuracy and predictive power and the relationship between spectral variation and protein concentrations is linear, with minimal standard deviation. The PLSR model built on varying concentration of glucose spiked in serum provided an accurate prediction model ($R^2=0.9914$, RMSECV= 1.66mg/dL) after applying pre-processing steps using the EMSC based algorithm. Having established the optimal sample preparation and analysis protocol using the spiked serum model, the same protocol was applied to patient samples. In the case of patient samples, RMSECV and $R^2$ values were calculated to be 1.84mg/dL, and 0.84 respectively. Although ATR-FTIR provides better standard deviation and higher $R^2$ value, the Raman prediction model gives lower a RMSECV value, indicating higher accuracy. The PLSR coefficient plot shows clear features of glucose, indicating the prediction model is built on variations in glucose concentrations. The added advantage of Raman spectroscopy is that the analysis can be performed on liquid samples and no additional time delays associated with sample drying are introduced. Although the quantitative capability of Raman can be easily demonstrated using glucose spiked with a pooled serum model, the analysis of patient samples can be more complicated, the reason being the intrinsic variability of individual samples depending upon the physiological state of the individual on that day.

For the purposes of a direct comparison of the two techniques, glucose concentration in human serum was chosen, and the results detailed in Table 5.3 indicate that Raman in the liquid state provides higher accuracy than FTIR-ATR. Although vibrational spectroscopic techniques are unlikely to replace current techniques for routine glucose monitoring, in a more general sense, an argument for serological applications of vibrational spectroscopic
techniques has been made (28, 37). The advantages of using Raman spectroscopy over common biochemical assays to quantify urea and creatinine has previously been reported (63). Moreover, Raman spectroscopy as a biochemical tool for serum analysis is cost effective, rapid and a non-destructive method as compared to currently employed gold standard clinical methods such as spectrophotometric analysis (e.g. COBAS analyser) (69). The COBAS analyser has a standard deviation of 0.04 mmol.L\(^{-1}\) (which is equivalent to 0.721 mg.dL\(^{-1}\)), as mentioned in the Material and Methods section, and the lower detection limit is 2mg/dL (70) with a correlation coefficient of 0.975 reported for immunoassays (71) and \(R^2\) of 0.990 for urea, creatinine, sugar, total protein and calcium (72). Similar \(R^2\) values were calculated for pure glucose solutions and glucose spiked serum solutions using Raman spectroscopy, suggesting that Raman spectroscopy is well-suited for routine use as a biochemical tool for glucose analysis. However, the analysis using COBAS analyser is complex and uses various enzymatic reagents, increasing the cost and chance of inaccuracy in the results obtained. Employing well trained personnel to operate the equipment enhances the reliability but increases the cost. Hence, Raman spectroscopy offers several advantages as it is a onetime investment, easy to operate and provides rapid results with wider information without destroying the medium. This could be translated as an alternative method for glucose monitoring, especially in the case of hyperglycaemia. Further studies need to be conducted to investigate other LMWF metabolites from human serum using Raman spectroscopy. However, to further ensure relevancy of the results, the study should ultimately be conducted on large number of patient samples.
5.6 Conclusion

Summarising, the work presented showcases the development of an optimal methodology for the detection of LMWF analytes from human serum using Raman spectroscopy with minimal sample preparation steps and without the use of any extrinsic labels. The proposed approach can be expeditiously employed for early detection of pathological disorders associated with high or low serum/plasma proteins/biomarkers. Disease diagnosis from bodily fluids can be developed into a dynamic diagnostic environment that will enable early disease diagnosis even before the disease becomes symptomatic. Thus, analysis of bodily fluids has emerged as one of the promising approaches to deliver crucial information about patient health and monitor disease progression and/or therapy. Given the remarkable advances in the field over the last two decades, including sample preparation, protein fractionation, quantitation and chemometrics, it is conceivable that vibrational spectroscopic techniques can be developed as a point-of-care disease monitoring system. Ultimately, the proof of concept presented in this study can be easily transferable to any other low molecular weight biomarkers or therapeutic drugs.
5.7 References


5.8 Supplementary information

1. When the entire fingerprint region was selected for PLSR analysis, the resultant PLSR coefficient displayed a negative peak at ~1000cm\(^{-1}\) which could potentially derive from other LMWF species such as urea (1)

![Figure 5.S1](image)

**Figure 5.S1** (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of serum samples, (B): Evolution of the RMSECV on the validation model, (C): PLSR coefficient shows a negative peak ~1000cm\(^{-1}\), (D): Predictive model built from the PLSR analysis.

5.8.1 Reference

Chapter 6

Raman spectroscopic screening of High and Low molecular weight fractions of human serum

The following chapter has been adapted from the published journal article titled as ‘Raman spectroscopic screening of High and Low molecular weight fractions of human serum’, Analyst, 2019, DOI: 10.1039/C9AN00599D.

**Author List:** Drishya Rajan Parachalil, Clément Bruno, Franck Bonnier, Hélène Blasco, Igor Chourpa, Jennifer McIntyre, and Hugh J. Byrne

DRP performed all experimental analysis and authored the publication. CB, HB, IC assisted clinical aspects of the study; FB provided the samples, assistance in harmonising the Raman and IR analysis protocols. FB, JMcC and HJB input to conceptual design of the work, drafting and proofing of the manuscript

6.1 Abstract

This study explores the suitability of Raman spectroscopy as a bioanalytical tool, when coupled with ultra-filtration and multivariate analysis, to detect imbalances in both high molecular weight (total protein content, γ-globulins and albumin) and low molecular weight (urea and glucose) fractions of the same samples of human patient serum, in the native liquid
form. Ultra-filtration was employed to separate and concentrate the high and low molecular weight fractions of the serum. Initially, aqueous solutions of the respective molecular species, covering physiologically relevant concentration ranges, were analysed to optimise the measurement protocols. An adapted Extended Multiplicative Signal Correction (EMSC) algorithm was applied to raw spectra to remove water background signal and spectral interferents (β-carotene). Using a validated partial least squares regression modelling method, $R^2$ values, Root Mean Square Error of Cross Validation (RMSECV) and standard deviations were established for the quantification of γ-globulin, total protein, albumin, urea and glucose content of the patient serum samples. The study demonstrates that Raman spectroscopy in the liquid form is a viable alternative and/or adjunct to current clinical practice for the parallel analysis of high and low molecular weight fractions, and simultaneous analysis of multiple analytes in the low molecular weight fraction, of human serum for diagnostic applications.

6.2 Introduction

Increasingly, medical diagnostic tests are performed on bodily fluids, as sample collection is minimally invasive, compared to histology or cytology based techniques. However, conventional methods for biofluid analysis can be inconsistent and are associated with high cost and long-time delays (1,2). Therefore, there is a need for more sensitive and cost effective methods with higher accuracy for early disease diagnosis from bodily fluids in a point-of-care clinical setting. Vibrational spectroscopic techniques, both Infrared absorption and Raman scattering, are among the most important analytical techniques available to scientists, as they provide detailed molecularly specific fingerprints of both organic and
inorganic compounds, without the use of extrinsic labels and without being extremely invasive or destructive to the system studied. Since both techniques are truly label-free, their potential for medical diagnostic applications has been well investigated and demonstrated (3–7). Such techniques are particularly attractive for routine analysis of biofluids, as they are easy to apply, require minimal sample preparation and are readily adaptable to analysis of various bodily fluids (1-5). Notably, it has been demonstrated that, whereas IR analysis of biofluids usually requires a drying step (8–10), Raman spectroscopic analysis can be applied to bodily fluids in their native liquid state (11,12). As water is a relatively weak scatterer, Raman spectroscopy is much more amenable than IR for analysis of biofluids in their native aqueous state (10, 11), and its potential has been well demonstrated, notably in human serum and plasma (13–16).

Estimating serum total protein, γ-globulin and albumin content is important to assess the nutritional status of patients (17,18). Although total serum protein estimation has limited diagnostic potential when compared to albumin or globulin, its relevance in the evaluation of patients with clinical conditions such as malnutrition, renal malfunction, liver diseases and immune disorders cannot be ignored (18,19). A normal level of total serum protein (6000-8300 mg/dL) indicates healthy nutritional status and normal liver function. Reduced serum total protein is predominantly found in patients with kidney disorders, HIV and aging in the elderly (20,21). The biuret assay is the most common method used to quantify total protein levels in blood serum (1).

γ-globulins (<150kDa (22), ~38% of serum proteins), produced by lymphocytes and plasma cells in lymphoid tissue, are large protein molecules that include the immunoglobulins: IgM, IgA, and IgG (23). The most characteristic abnormality in serum proteins in liver diseases
and carcinoma of the gastro-intestinal tract or breast is an elevation in \( \gamma \)-globulins (24,25). Testing globulin levels in serum routinely provides key information that helps diagnose various conditions and diseases that affect the immune status. Liver diseases, chronic inflammatory diseases, haematological disorders, infections and malignancies cause excess globulin levels (>1600mg/dL) (26), whereas humoral immunodeficiencies cause low globulin levels (<700mg/dL) (27). Radial immunodiffusion (RID) (28) is the gold standard method for measuring globulins. All the conventional methods used for testing total protein content and globulin measurement make use of expensive disposables and are labour intensive. With escalating medical costs and budget constraints, a cost effective alternative technology is desirable.

Albumin is the most abundant high molecular weight fraction (HMWF) serum protein, normally constituting about 50% of the total serum content, and has a molecular weight of 66 kDa (29). The normal concentration of albumin in the human body is 3000 mg/dL, although it dramatically decreases in critically ill patients and does not increase again until the recovery phase of the illness (17). Several studies have demonstrated that the functions of albumin, such as ligand binding and transport of various molecules, can be applied to the treatment of cirrhotic patients and patients suffering from other end stage liver diseases (30–32). Strong correlation between cardiovascular disease and the level of albumin concentration have also been reported (33–38), which means a normal concentration of albumin in bodily fluids is considered a sign of good health. It is clear therefore that closely monitoring the variation in albumin concentration could act as an indicator of liver diseases and other related pathologies. Conventional methods used to determine the level of albumin include absorption spectroscopy or electrochemical based assays, immunoassays and high
performance liquid chromatography (HPLC) and can be time consuming and very expensive (20, 26–32). Therefore, a sensitive, rapid, cost-effective method such as Raman spectroscopy is highly desirable to quantitatively analyse albumin.

Urea is a colourless crystalline compound of molecular weight 60.056 g/mol (60.056 Da) and is the main nitrogenous by-product produced by the liver when the body metabolises proteins (45). The normal concentration of urea in human blood serum is 5–20 mg/dL. The kidneys normally filter out this waste from the body, and therefore it is important to monitor blood urea as higher (<20 mg/dL) or lower levels (>5 mg/dL) could indicate various pathologies such as kidney or liver malfunction (46–49). The standard technique for assessment of urea from blood or urine is based on colorimetric measurement, by which a specific reagent reacts with the sample and absorption at a given wavelength is used to identify urea concentration (48,50).

Glucose is a low molecular weight carbohydrate that must be monitored on a regular basis in the case of diabetes patients (51). The normal concentration of glucose in blood is 70 to 130 mg/dL and changes in this level would lead to hyper (>130 mg/dL) or hypoglycaemia (<70 mg/dL) (51). While currently available methods of routinely self-monitoring blood glucose have become low cost, such glucose meters often suffer from significant errors that can lead to unreliable results (52,53). Hence, there is an unmet need for an accurate glucose monitoring tool, notably in a clinical setting.

Over the past decades, there have been numerous studies of analytes in biofluids using vibrational spectroscopy, and, in recent years, attenuated total reflection (Fourier Transform) IR (ATR-FTIR) has become popular for rapid screening of biofluids, particularly blood
serum (54). Multianalyte serum analysis has previously been reported using mid-infrared spectroscopy, for the simultaneous quantitation of eight serum analytes: total protein content, albumin, triglycerides, cholesterol, glucose, urea, creatinine and uric acid (55) and for simultaneous quantification of glucose and urea analytes along with malaria parasitemia quantification using ATR-FTIR (56). Notably, however, both the analyses were conducted on dried serum samples. Berger et al., as long ago as 1999, used Raman microscopic analysis of liquid whole human blood and serum samples to quantify the content of six analytes, namely glucose, cholesterol, triglyceride, urea, total protein and albumin (57). Rohleder et al., performed a direct comparison of Raman and FTIR spectroscopies of multiple analytes in human serum, concluding that the techniques produced similar accuracies(58). More recently, Parachalil et al., demonstrated that, using the same serum fractionation and regression analysis protocols, Raman microspectroscopic analysis in the liquid state performed at least as well as ATR-FTIR of dried samples, in the quantification of glucose levels in human serum (11).

This study aims to further evaluate the potential of Raman spectroscopy for the analysis of blood serum in the liquid state to simultaneously detect and quantify subtle variations in the whole serum (total protein content and γ-globulins), HMWF (albumin) and LMWF (urea and glucose) as specific biomarkers linked to numerous pathologies, adding improved sampling techniques, sample fractionation and selected spectral ranges to improve over the previous work using ATR-FTIR and Raman spectroscopy.

Human serum is highly complex and diverse, however, due to its heterogeneous nature, consisting of dynamic range of biomolecules such as proteins, lipids, carbohydrates etc. (59). Raman spectra of serum are dominated by features of the abundant proteins (notably
globulins and albumin) and water, making it difficult to visualise the spectral features of less abundant analytes and small molecules which can act as biomarkers. Both urea and glucose are classified as Low Molecular Weight Fraction (LMWF) analytes, and their low abundance makes it harder to detect these analytes from the full serum samples. Therefore, measures are taken to make data interpretation easier, such as fractionation of serum using centrifugal filters prior to Raman analysis, followed spectral pre-processing.

6.3 Materials and Method

6.3.1 Sample Preparation

γ-globulins (G4386), albumin (A9511), urea (F3879) and β-carotene (C9750-5G) were purchased from Sigma Aldrich, Ireland. Individual solutions of varying concentrations of urea (1-1000 mg/dL) were prepared in distilled water by varying the concentrations over a physiologically relevant range.

Patient serum samples were donated by the University Hospital (CHRU) Bretonneau de Tours (France) and the ethical procedures were followed. The blood samples were collected from the individuals as routine blood check-ups and 1 mL per patient was provided for spectroscopic analysis. A total of 25 patient samples were included in the present study. Samples were collected by personnel of CHRU, under standard clinical protocols and approved ethical procedures (Comité de Protection des Personnes, Tours Region Central Oest 1- PP/ANSM- PHAO15-HB-METABOMU, registered internationally: ClinicalTrials.gov ID: NCT02670226). The samples were serologically profiled, for other purposes, and the anonymised, residual discard samples, along with their serological profiles were donated to
the Université de Tours, for further study. No further specific ethical approval or patient consent is required. Albumin, γ-globulins (IgG, IgM and IgA), total protein, urea and glucose concentrations were obtained by routine biochemical analysis using a COBAS analyser, following the CHRU guidelines for routine biochemical analysis. The principle for the test of albumin utilises the formation a blue-green complex to test for albumin, and a purple-coloured biuret complex to test for total protein. While albumin displays a sufficiently cationic character to be able to bind with an anionic dye, Bromosresol green (BCG), at a pH of 4.1 (60) and divalent copper reacts in alkaline solution with protein peptide bonds in the case of total protein (61). The urea test is based on a coupled enzyme reaction (urease, followed by glutamate dehydrogenase), whereby measurement of NADH (converting to NAD+) is made at 340 nm (62). γ-globulin concentrations were obtained using immunoturbidometric assays, and were summed to provide a value for total γ-globulin content (63). Measured analyte levels in the patient samples are listed in Table 6.1. Concentrations are expressed in mg/dL for consistency with other studies (9). Notably, no correlation was found between the concentrations of HMWF and LMWF analytes in the patient serum samples (Figure 6.S1 and 6.S2 in supplemental).
### Table 6.1. Measured analyte levels in patient samples

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Total Protein (mg/dL)</th>
<th>γ-globulin (mg/dL)</th>
<th>Albumin (mg/dL)</th>
<th>Urea (mg/dL)</th>
<th>Glucose (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6300</td>
<td>1327</td>
<td>3130</td>
<td>78.99</td>
<td>108.0</td>
</tr>
<tr>
<td>2</td>
<td>6900</td>
<td>1010</td>
<td>4510</td>
<td>15.97</td>
<td>81.0</td>
</tr>
<tr>
<td>3</td>
<td>6400</td>
<td>2254</td>
<td>2710</td>
<td>21.85</td>
<td>91.8</td>
</tr>
<tr>
<td>4</td>
<td>6700</td>
<td>1259</td>
<td>3580</td>
<td>7.00</td>
<td>115.2</td>
</tr>
<tr>
<td>5</td>
<td>5800</td>
<td>404</td>
<td>3050</td>
<td>8.40</td>
<td>241.2</td>
</tr>
<tr>
<td>6</td>
<td>5600</td>
<td>670</td>
<td>3120</td>
<td>5.88</td>
<td>207.0</td>
</tr>
<tr>
<td>7</td>
<td>7600</td>
<td>1395</td>
<td>4220</td>
<td>50.14</td>
<td>115.2</td>
</tr>
<tr>
<td>8</td>
<td>6700</td>
<td>712</td>
<td>4450</td>
<td>75.63</td>
<td>210.6</td>
</tr>
<tr>
<td>9</td>
<td>6000</td>
<td>875</td>
<td>3690</td>
<td>59.38</td>
<td>75.6</td>
</tr>
<tr>
<td>10</td>
<td>6200</td>
<td>1164</td>
<td>2970</td>
<td>5.88</td>
<td>282.6</td>
</tr>
<tr>
<td>11</td>
<td>6900</td>
<td>1409</td>
<td>3690</td>
<td>2.52</td>
<td>75.6</td>
</tr>
<tr>
<td>12</td>
<td>7500</td>
<td>1843</td>
<td>3910</td>
<td>29.41</td>
<td>102.6</td>
</tr>
<tr>
<td>13</td>
<td>5800</td>
<td>726</td>
<td>3880</td>
<td>21.57</td>
<td>129.6</td>
</tr>
<tr>
<td>14</td>
<td>6100</td>
<td>840</td>
<td>3340</td>
<td>10.64</td>
<td>93.6</td>
</tr>
<tr>
<td>15</td>
<td>4200</td>
<td>541</td>
<td>2600</td>
<td>11.20</td>
<td>178.2</td>
</tr>
<tr>
<td>16</td>
<td>7900</td>
<td>1478</td>
<td>3890</td>
<td>17.65</td>
<td>77.4</td>
</tr>
<tr>
<td>17</td>
<td>6600</td>
<td>1187</td>
<td>3950</td>
<td>10.92</td>
<td>91.8</td>
</tr>
<tr>
<td>18</td>
<td>7600</td>
<td>1201</td>
<td>4780</td>
<td>17.28</td>
<td>70.2</td>
</tr>
<tr>
<td>19</td>
<td>7700</td>
<td>1484</td>
<td>4420</td>
<td>26.33</td>
<td>158.4</td>
</tr>
<tr>
<td>20</td>
<td>6500</td>
<td>1396</td>
<td>3380</td>
<td>52.94</td>
<td>52.2</td>
</tr>
<tr>
<td>21</td>
<td>7400</td>
<td>1663</td>
<td>4360</td>
<td>8.12</td>
<td>109.8</td>
</tr>
<tr>
<td>22</td>
<td>7600</td>
<td>924</td>
<td>4250</td>
<td>39.50</td>
<td>70.2</td>
</tr>
<tr>
<td>23</td>
<td>6600</td>
<td>1556</td>
<td>3150</td>
<td>19.05</td>
<td>284.4</td>
</tr>
<tr>
<td>24</td>
<td>6500</td>
<td>1171</td>
<td>3750</td>
<td>8.12</td>
<td>55.8</td>
</tr>
<tr>
<td>25</td>
<td>7100</td>
<td>1236</td>
<td>4140</td>
<td>25.49</td>
<td>433.8</td>
</tr>
</tbody>
</table>

Amicon Ultra 0.5 mL centrifugal filter devices (Merck, Germany), of cut-off point 100 kDa, 50 kDa and 10 kDa, were employed to concentrate and separate the analytes of interest from the individual solutions and patient samples. The centrifugation procedure that has been previously reported by Bonnier *et al.* was followed (13). The optimised washing and rinsing procedure includes spinning 0.5 mL 0.1 M NaOH at 14000×g for 30 minutes followed by three rinses with distilled water by spinning 0.5 mL distilled water for 30 minutes at 14000×g.
Every 30 minute wash and rinse must be followed by spinning the device in the inverted position at 1000×g for 2 minutes, to remove the residual solution contained in the filter.

**Figure 6.1.** Schematic overview of steps in fractionation of patient serum samples to separate γ-globulin, albumin, and urea/glucose

Figure 6.1 is a schematic representation of the patient serum fractionation steps employed in this study. Non-fractionated serum was directly analysed to collect data for total protein content and γ-globulin. In order to fractionate the patient serum samples, 0.5 mL sample was first spun for 30 minutes using 100 kDa filters to remove the analytes larger than 100 kDa (γ-globulin) in the concentrate and the transmitted filtrate, which contains molecules smaller than 100 kDa, was used for further fractionation. For albumin isolation, the filtrate obtained was spun using 50 kDa filters for 30 minutes and the resultant concentrate was collected for Raman analysis. The filtrate transmitted by the 50 kDa filter was collected and was spun using 10 kDa filter for 30 minutes and the resultant filtrate was collected for Raman analysis.
to detect the presence of urea/glucose. Using a model sample based on glycine spiked serum, Bonnier et al. demonstrated the reproducibility of the HMW proteins depletion to be quite efficient and acceptable (8).

6.3.2 Data collection using Raman spectrometer

The measurement conditions used were the same as those recently reported for screening HMWF proteins (11) and glucose (64) in solution. Raman spectra of the liquid samples were recorded at stabilised room temperature (18°C) using a Horiba Jobin-Yvon LabRam HR800 spectrometer with a 16-bit dynamic range Peltier cooled Charged Coupled Device. The spectrometer was coupled to an Olympus 1X71 inverted microscope and a x60 water immersion objective (LUMPlanF1, Olympus) was employed. A 532nm laser was used, which had a power of ~30 mW at the sample, with a 600 lines/mm grating and the backscattered Raman signal was integrated for 3×80 seconds over the spectral range from 400-1800 cm⁻¹. The substrate used was a Lab-Tek plate (catalog number 154534) with a 0.16-0.19 mm thick glass bottom, 1.0 borosilicate cover glass, and was purchased from Thermo Fischer Scientific, Ireland.

6.3.3 Data pre-processing and analysis

The raw spectra were subjected to pre-processing techniques in Matlab before further analysis, to remove the background signal and reduce the noise. Smoothing of the raw data was done using the Savitzky–Golay method (polynomial order of 5 and window 13) and the rubberband method (11) was found to be appropriate to baseline correct the smoothed reference spectra of all the analytes and the smoothed spectra of varying concentrations of
albumin spiked in distilled water (11). The ‘rubberband’ correction was carried out by wrapping a ‘rubberband’ of defined length around the ends of the spectrum to be corrected and fitting against the curved profile of the spectrum. An adapted Extended Multiplicative Signal Correction (EMSC) algorithm (65) with a 5th order polynomial was applied to the raw dataset to remove the spectral interferents from the data. EMSC is applied to remove the underlying water spectrum from all the dataset, whose OH bending feature at 1640 cm⁻¹ can interfere with the protein spectra and also scales the analyte spectra, assuming a constant water contribution to all sample spectra (65). The principle of EMSC for subtraction of a specific measureable background spectrum and the associated Matlab codes have previously been published by Kerr and Hennelly, 2016 (65), and their description is adapted in the following. The raw spectrum, S, consists of Raman spectrum of interest, R, a baseline signal, B, and the water signal, W.

\[ S = R + B + W \] (65) \hspace{1cm} (1)

The Raman spectrum of interest can be represented by a reference spectrum of the material of interest, r, and it can be assumed that R is the product of this reference spectrum and a certain scalar weight, \( C_r \), which describes the concentration dependence (66,67)

\[ R \sim C_r \times r \] (65) \hspace{1cm} (2)

Similarly, a spectrum, w, is recorded from water directly in order to represent the spectral contribution of water in W, as the product of pure water spectrum and a certain scalar weight.

\[ W = C_w \times w \] (65) \hspace{1cm} (3)

The baseline, B, is now represented by an appropriate order of polynomial (N) as:
\[ B_N = C_0 + C_1X + C_2X + \ldots + C_NX^N \quad (65) \]

where \( N \) is the order of polynomial and \( C_m \) for \( m = 0 \rightarrow N \) represents various coefficients of polynomial. The EMSC algorithm is used to obtain estimates of the scalar values \( C_r, C_m \) and \( C_w \). These estimates are obtained from an optimal fit of the various vectors in Equation 5.

\[ S \sim [C_r \times r] + [C_w \times w] + \left[ \sum_{m=0}^{N} C_mX^m \right] \quad (65) \]

The background corrected, concentration dependent analyte spectra, \( T \), can be represented as:

\[ T = \frac{S-[C_w \times w]-[\sum_{m=0}^{N} C_mX^m]}{C_w} \quad (65) \]

Note, that division by \( C_w \) has the effect of scaling the analyte spectra, assuming a constant water contribution to all sample spectra.

In the case of total protein, \( \gamma \)-globulins and albumin, EMSC is also applied to remove the \( \beta \)-carotene. Since the Lab-Tek plate is made up of thin glass bottom, no glass correction was required. Raman spectra of the pure serum, \( \gamma \)-globulin (\(~100\ \text{mg/mL}\)), albumin (\(~100\ \text{mg/mL}\)), glucose (\(45\text{mg/mL}\)) and urea (\(~100\ \text{mg/mL}\)) prepared with minimal amount of water are used as the reference for EMSC.

**6.3.4 Partial Least Squares Regression**

Partial Least Squares Regression (PLSR) is a multivariate statistical method which aims to establish a model that relates the variations of the spectral data to a series of relevant targets.
The PLSR model attempts to elucidate factors which account for the systematic majority of variation in predictors ‘X’ (spectral data) versus associated responses ‘Y’ (target values of protein concentration) (68). The spectral data (X matrix) is thus related to the targets (Y matrix) according to the linear equation $Y = XB + E$, where B is a matrix of regression coefficients and E is a matrix of residuals. The PLSR algorithm allows for the construction of a regression model which can be used to predict the outcome in varying concentration of analytes, and the performance of the PLSR model in predicting varying analytes concentration was evaluated in this study. In this case, the examples used are concentration and Raman signal, and therefore the algorithm can be used to predict the detection of Raman signal for a particular analytes concentration. The performance of the regression model in predicting varying concentrations of the analyte over a particular range was evaluated and this method can be employed to improve the limit of detection of Raman bio-sensing (68).

Constructed based on the spectra of samples of known analyte content, either solutions of varying concentrations in distilled water or those of the patient serum, the model is then validated using a rigorous cross validation procedure which evaluates its performance in accurately predicting analyte concentrations. For consistency with previous studies (64), a 20 fold cross validation approach has been employed to validate the robustness of the method. This approach involves randomly dividing the set of observations into approximately equal size, ~50% of the spectral data were randomly selected as test set, while the remaining ~50% is used as the training set (69). In the current case, (5 x 25 spectra) were divided into two groups of 65 (test) and 60 (training) spectra. The cross-validation process is then repeated 20 times (the folds), whereby all observations are used for both training and testing, and each observation is used for testing exactly once. The results from the folds can then be averaged.
to produce a single estimation. The Root Mean Square Error of Cross Validation (RMSECV) is calculated from the 20 iterations to measure the performance of the model for the unknown cases within the calibration set. The correlation between the concentration and spectral intensity is given by the $R^2$ value. The standard deviation was calculated to find the variation between each spectrum calculated from the same sample. The number of latent variables used for building the PLSR model is optimised by finding the value that is equivalent to the minimum of the RMSECV.

6.4. Results and discussion

Figure 6.2A shows the mean spectra of the whole serum samples derived from patients. For comparison, Figure 6.3 shows the spectra of globulin and albumin. Unexpectedly, additional strong Raman bands at 1167 cm$^{-1}$ and 1535 cm$^{-1}$ (12,70) are observed, which dominate the protein features of serum, which can be ascribed to the presence of β-carotene (Figure 6.S3). β-carotene is a red-orange-coloured carotenoid widely distributed in fruits and vegetables and plays a major role in the maintenance of normal healthy skin, vision, immune system and mucous membranes (71,72). Regular intake of β-carotene rich foods results in higher levels in blood serum/plasma (73,74) and is also associated with decreased risk of cardiovascular disease, heart disease, cancer, and other causes of mortality (75,76). As shown by a community-based study of French adults, serum β-carotene and vitamin C concentrations are useful biomarkers of vegetable and fruit consumption in the French diet (77). β-carotene features became less apparent in the spectra recorded from the concentrate of 50kDa filtration, as shown in Figure 6.2B and no β-carotene features were found in the spectra collected from the filtrate of 10kDa filtration, as seen in Figure 6.2C. The molecular size of
β-carotene is less than 10 kDa (~537Da) (78), and therefore the presence of β-carotene in the concentrate of 50 kDa rather than filtrate of 10kDa must be attributed to the binding of β-carotene to albumin (79). Some studies have shown high binding affinity of β-carotene to albumin due to its identical left-right symmetric structure that has great potential for association of albumin molecules and protects β-carotene against oxidative degradation (79,80). In terms of quantitative analysis of the high molecular weight constituents, the strong contribution of the β-carotene is undesirable. Therefore, EMSC was applied to subtract the water signal, β-carotene signal as well as background noise from the initial serum data, and from that of the 50 kDa filtered data.

![Figure 6.2](image)

**Figure 6.2.** Raman spectra of patient serum collected using Raman spectroscopy. (A) whole serum, (B) concentrate from 50 kDa filtration and (C) filtrate from 10 kDa filtration. Spectra have been off set for clarity.
Figure 6.3. Spectra of γ-globulins (red - ~38% of serum) and albumin (blue - ~50% of serum) showing similar spectral features. Identifying signature peaks of γ-globulins at 1240 cm\(^{-1}\) and 1553 cm\(^{-1}\) and of albumin at 940 cm\(^{-1}\) are highlighted with asterisks.

6.4.1 Quantification of total protein concentration and γ-globulins in whole serum

A total serum protein test analysis measures the total amount of protein in the serum, as well as the amount of two major proteins in the serum; albumin and globulin. The normal concentration of total protein content in human serum is between 6000-8300 mg/dL and a reduced serum protein level is an indication of kidney disorders, HIV and aging in the elderly (20,21). A pure serum spectrum was used as the reference for the EMSC correction (Figure 6.S4), along with spectra of pure β-carotene and water, measured under identical conditions. Figure 7.4A presents the EMSC corrected Raman spectra of all the 25 patients from unfiltered serum samples. As expected, the characteristic bands of both globulin and albumin, highlighted in Figure 6.3, can be observed, as indicated by the asterisks (81,82). Notably, the
strong bands of β-carotene are no longer as prominent. The background-subtracted, smoothed spectra of the whole patient serum from the range 400 cm$^{-1}$ to 1800 cm$^{-1}$ were provided to the PLSR algorithm, regressed against total serum protein concentrations of Table 6.1, and a prediction model for total protein concentrations was built. The RMSECV plot of the PLSR model shows a steady decrease within the first 10 components and stabilises after 12$^{th}$ component, indicating the data is well modelled (Figure 6.4B). Note that the data range is not well balanced, and particularly that there is only one patient with total serum content below 5500 mg/dL. Nevertheless, the PLSR co-efficient of Figure 6.4C shows spectral features at 940 cm$^{-1}$, 1553 cm$^{-1}$ and 1176 cm$^{-1}$, indicating the prediction model (Figure 6.4D) was built on the spectral features of albumin (50% of total protein) and γ-globulin (38% of total protein). The RMSECV, $R^2$ and standard deviation values were calculated as 114.7 mg/dL, 0.82 and 5.69 mg/dL. For comparison, the prediction accuracy achieved by Berger et al., for measuring total protein using near infrared Raman spectroscopy with 830nm was reported to be 190mg/dL(57), whereas Shaw et al., (55) and Rohleder et al., (83) reported prediction accuracies of 310mg/dL, 176mg/dL using Mid-IR spectroscopy of dried films indicating that the present method is more feasible for total protein analysis from serum.
Figure 6.4. (A) EMSC corrected Raman spectra of total protein content from patient serum samples (4200 mg/dL, 5800 mg/dL, 6400 mg/dL and 7900 mg/dL). The spectra have been offset for clarity (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient for regression against total serum protein, (D): Linear predictive model for total serum protein built from the PLSR analysis. The RMSECV, $R^2$ and standard deviation values were calculated as 114.7 mg/dL, 0.82 and 5.69 mg/dL.
Figure 6.5. (A) EMSC corrected Raman spectra of γ-globulin of patient serum samples (329 mg/dL, 690 mg/dL, 836 mg/dL and 1404 mg/dL), the spectra has been offset for clarity. (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient for regression against γ-globulin concentration shows the signature peaks of γ-globulin (highlighted by asterisk), (D): Linear predictive model for γ-globulin concentration built from the PLSR analysis. The RMSECV, R² and standard deviation values were calculated as 126 mg/dL, 0.88 and 4.62 mg/dL.

In the case of determination of γ-globulin content in patient serum, the background-subtracted and smoothed fingerprint region of the unfiltered serum spectra was also analysed, but regressed this time against the γ-globulin concentration. As expected, the RMSECV prediction plot (Figure 6.5B) shows a steady decrease until the 12th component, indicating that 12 components should be utilised in this model. The lowest value of RMSECV was recorded as 126 mg/dL and the linearity of the model was calculated as R²=0.88. However,
it should be noted that the spectral profile of the regression co-efficient of Figure 6.5C is not strikingly similar to the spectrum of γ-globulin in Figure 6.3, and the characteristic features of at 1240 cm$^{-1}$ and 1553 cm$^{-1}$ are not prominent. γ-globulin is composed of IgG, IgM and IgA, the relative contributions of which are seen to vary considerably from patient to patient. Immunoglobulin G (IgG), itself, for example, is well known for its heterogeneity, which in itself makes it an excellent biomarker of a person's general state of health (84,85). Regression against each of the individual components did not produce a good correlation. The spectrum of each of them, and the composite g-globulin can also be influenced by protein/protein and any other interactions in the complex serum mixture, which can give rise to conformational changes in the immunoglobulin structure. Thus, it may not be surprising that the individual patient globulin spectral contributions are variable from patient to patient, and that the regression co-efficient of Figure 6.5C, effectively an average of them all, does not exactly match the spectral of the commercial sample in Figure 6.3. Nevertheless, the R$^2$ of this model is comparable to that of the method reported by Guaita et al. (86) to detect total globulin using ATR-FTIR from blood serum, and it should be noted that no correlation between globulin content and that of the other major protein, albumin was observed (Figure 6.S1). The background-subtracted and smoothed fingerprint region of the unfiltered serum spectra was also analysed over a reduced spectral region (1150 cm$^{-1}$ to 1600 cm$^{-1}$) to avoid spectral interference from albumin, but regressed this time against the γ-globulin concentration. Analysis over this reduced range failed to produce an improved performance, however (Figure 6.S5). Excess globulin levels (>1600mg/dL) are usually an indicator of liver diseases, chronic inflammatory diseases, haematological disorders, infections and malignancies (26), whereas humoral immunodeficiencies cause low globulin levels (<700mg/dL) (27).
As shown in Figure 6.3, the Raman spectra of albumin and γ-globulin are extremely similar and, since the concentrations of albumin and γ-globulin are not correlated (Figure 6.S1 in supplemental), PLSR cannot predict both from the same dataset; invalid correlation for the entire population from which the training set is drawn would lead to spurious prediction models (57). Hence, patient serum was spun using 100kDa centrifugal filter tubes to separate and remove the γ-globulins in the concentrate and the remainder filtrate was transferred to 50kDa filter tubes for further concentration of albumin. The spectra recorded from the concentrate of 50 kDa filtrate was used for building a PLSR prediction model for albumin from patient samples.

6.4.2 Quantification of albumin from the HMWF concentrate

For albumin determination, the spectra recorded from albumin paste (Figure 6.3 blue) and β-carotene (Figure 6.S3) were used as the reference spectrum for the EMSC algorithm to perform background correction and the background-corrected and smoothed spectra were fed into the PLSR algorithm to build the prediction model by regressing against albumin concentration (Table 6.1). The signature albumin bands that can be seen are the amide I band around ~1659 cm⁻¹, a relatively sharp band at 1003 cm⁻¹ associated with phenylalanine, intense bands at ~1336 cm⁻¹ and ~1450 cm⁻¹ due to C-H deformation, bands at 899 cm⁻¹ and 1102 cm⁻¹, which can be related to ν(CC) and ν(CN) and a vibration band at ~940 cm⁻¹, related to C-C stretching mode backbone of α-helix structure (87). The absence of the Raman peaks at 1240 cm⁻¹ and 1553 cm⁻¹ confirms that there is no interference from γ-globulin. PLSR analysis of aqueous solutions of albumin over the range (500-5000 mg/dL) has previously been reported, and the results of the prediction model are summarised by Figure
6.56 in the supplemental material. Leave-One-Out cross validation was used to test the robustness of this model and the RMSECV value calculated was 158 mg/dL (11).

EMSC successfully subtracted the background without altering the albumin features (Figure 6.6A), such as bands at ~1336 cm\(^{-1}\) due to C-H deformation and a vibration band at ~940 cm\(^{-1}\), related to C-C stretching mode backbone of α-helix structure. The strength of the albumin features is, however, seen to vary considerably from patient to patient, as expected, due to the variation in albumin content from 2710 mg/dL to 4780 mg/dL (Table 6.1). The normal level of albumin in human blood is 3000mg/dL (17), any decrease in this level could be an indicator of liver diseases (30-32). Applying PLSR to the dataset, a strong decrease in the RMSECV is observed within the first 8 latent variables, as shown in Figure 6.6B, which is followed by a stabilisation of the values after 12 latent variables, indicating that 12 latent variables should be used to build the prediction model. The PLSR coefficient plot displayed in Figure 6.6C shows visible albumin features, indicating that the prediction model was built on their variation. Finally, a linear predictive model is built from the PLSR analysis (Figure 6.6D) to compare the observations to the known concentrations of albumin in the samples with the estimated concentrations from the spectral data sets, yielding an \(R^2\) value of 0.9072 and RMSECV value of 90.097 mg/dL. The prediction accuracy of the present method for measuring albumin from serum is higher than the prior studies on Raman (120mg/dL) (57) and Mid IR spectroscopy (220mg/dL)(55). The overall standard deviation is calculated to be 1.1692 mg/dL. These values are comparable to the previously reported values obtained from varying concentration of pure albumin in water (11).
6.4.3 Quantification of glucose and urea from the LMWF filtrate

Glucose and urea are both present in the filtrate obtained after depletion of the HMWF of human serum using filtration. Notably, there is no obvious correlation between the levels of the two analytes per patient, as shown in Figure 6.S2 in supplemental.
Figure 6.7 shows the reference spectra of urea and glucose used for EMSC correction of the raw spectral data. The signature peak of urea is a relatively sharp band seen at 1006 cm$^{-1}$, as seen in Figure 6.7A, which can be attributed to the symmetric stretching of C-N (88). Figure 6.7B shows Raman peaks of glucose at 1060 cm$^{-1}$, due to ν(C1-OH) stretching, and a sharp peak at 1125 cm$^{-1}$, which can be assigned to the δ(C-O-C) angle-bending mode (89).

In the case of glucose, Figure 6.S7, 6.S8 and 6.S9 in the supplemental display the previously reported prediction models built from varying concentrations of glucose in distilled water, spiked in commercial serum, and patient serum, respectively (64). The RMSECV value calculated for the model built from varying concentrations in distilled water (100-1000 mg/dL) was 10.93 mg/dL and $R^2$ was calculated as 0.9705. Notably, in the patient samples, the overlap of the strong spectral features of urea and glucose can interfere with the PLSR prediction of glucose using the full spectral range, as shown in Figure 6.S9. In order to facilitate efficient prediction of glucose from the filtrate, the spectral region from 1030 cm$^{-1}$ to 1400 cm$^{-1}$ was chosen to build a PLSR prediction model of glucose from patient samples (Figure 6.S8), (52.25-440 mg/dL) resulting in an RMSECV value of 1.84 mg/dL and $R^2$ value of 0.84, comparable with the values obtained for the pure aqueous solutions (64). The results suggest that this method could detect the concentration of glucose in the ranges of hyper (>130 mg/dL) or hypoglycaemia (<70 mg/dL) (51).
Figure 6.7. (A) Reference spectrum of urea (1000 mg/dL) and (B) Reference spectrum of glucose (45000 mg/dL). The reduced spectral regions selected for PLSR analysis are indicated by dotted lines in each spectrum.

Prior to the analysis of urea in patient samples, aqueous solutions of varying concentrations of urea were made up and Raman analysis with the optimised protocol was performed. The recorded spectra were pre-processed and PLSR analysis was performed. EMSC efficiently subtracts the water spectrum, which has an OH bending vibration at ~1640 cm⁻¹ (64) and obscures the protein signals at low concentrations. Figure 6.S11 shows the spectra corrected with the EMSC algorithm, showing strong features of urea increasing as the concentration is increased from 1 mg/dL to 1000 mg/dL. The PLSR algorithm, using 6 latent variables, was applied to the smoothed and background-subtracted spectra and a linear predictive model was built, yielding a lowest value of RMSECV of 70.4 mg/dL and an $R^2$ value of 0.9048. The overall standard deviation was calculated to be 1.0975 mg/dL.

Having verified the protocol for analysis of urea in water, it was applied to the filtrate obtained after filtration from patient serum. Higher (<20 mg/dL) or lower levels (>5 mg/dL) of urea in the patient serum could indicate various pathologies such as kidney or liver.
malfuction (46–49). Figure 6.8A shows the EMSC-corrected data of the fingerprint region and Figure 6.8B indicates 12 latent variables should be used to build the model. The PLSR coefficient shows the features of urea, as displayed in 6.8C and Figure 6.8D shows the linear model built from this dataset. The linearity is indicated by \( R^2 = 0.9232 \) and the RMSECV was calculated as 1.736 mg/dL. The overall standard deviation was calculated to be 2.89 mg/dL.

Figure 6.8. (A): EMSC corrected Raman spectra of filtrate obtained after filtration with 10 kDa filters of patient samples (2.5 mg/dL, 10.64 mg/dL, 19.04 mg/dL and 78.99 mg/dL). The spectra have been offset for clarity (B): Evolution of RMSECV of the data set regressed against urea concentrations (C): plot of PLSR coefficient with strong features of urea, (D): Linear predictive model built from the PLSR analysis. The RMSECV, \( R^2 \), and standard deviation values were calculated as 1.736 mg/dL, 0.9232 and 2.89 mg/dL.
Notably, the typical serum concentrations of urea are considerably higher than those of glucose, as shown in Table 6.1, and therefore the characteristic features of urea dominate over those of glucose. Nevertheless, having observed the improvement of the sensitivity of the model for the case of glucose, a reduced spectral region, from 800 cm\(^{-1}\) to 1030 cm\(^{-1}\), was also tested for the urea regression.

Figure 6.9. (A): EMSC corrected Raman spectra of filtrate obtained after filtration with 10kDa filters of patient samples (2.5 mg/dL, 10.64 mg/dL, 19.04 mg/dL and 78.99 mg/dL). Spectra have been offset for clarity, (B): Evolution of RMSECV of the data set regressed against urea concentration (C): plot of PLSR coefficient with strong features of urea, (D): Linear predictive model for urea concentration built from the PLSR analysis. The RMSECV, \(R^2\) and standard deviation values were calculated as 2.52 mg/dL, 0.9722 and 1.1418 mg/dL.
Figure 6.9A shows the EMSC corrected spectra of urea over the reduced spectral range, and Figure 6.9C displays the PLSR coefficient, which is dominated by the 1006 cm⁻¹ peak of urea. 12 latent variables were used to build the model after calculating the lowest value of RMSECV, as seen in Figure 6.9B. Figure 6.9D confirms that the linear predictive model is based on systematic variations of the features of urea. The value of RMSECV, $R^2$ and overall standard deviation calculated are 1.1418 mg/dL, 0.9722 and 2.52 mg/dL. Better $R^2$ and RMSECV values were obtained when the PLSR analysis was performed over the reduced spectral range of urea as compared to the full fingerprint region, suggesting that the sensitivity and the accuracy of the model could be considerably improved. The proposed method also has a better prediction accuracy than the previously reported methods on Raman by Berger et al. (3.8mg/dL) (57) and mid IR spectroscopy by Shaw et al. (3.08mg/dL) (90), Petrich et al. (16mg/dL) (58) and Rohleder et al. (2.1mg/dL) (83).

Table 6.2. Summary of the results obtained from the patient samples

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Analysis range</th>
<th>Concentration range</th>
<th>$R^2$</th>
<th>RMSECV</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total protein</td>
<td>400-1800 cm⁻¹</td>
<td>4200-7900 mg/dL</td>
<td>0.82</td>
<td>115 mg/dL</td>
<td>5.7 mg/dL</td>
</tr>
<tr>
<td>(whole serum)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ-globulin</td>
<td>1150-1600 cm⁻¹</td>
<td>329-1404 mg/dL</td>
<td>0.88</td>
<td>126 mg/dL</td>
<td>4.6 mg/dL</td>
</tr>
<tr>
<td>(whole serum)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albumin (HMWF)</td>
<td>400-1800 cm⁻¹</td>
<td>2600-4780 mg/dL</td>
<td>0.90</td>
<td>90 mg/dL</td>
<td>1.2 mg/dL</td>
</tr>
<tr>
<td>Glucose (LMWF)</td>
<td>1030-1400 cm⁻¹</td>
<td>52.5-434.2 mg/dL</td>
<td>0.84</td>
<td>1.8 mg/dL</td>
<td>3.5 mg/dL</td>
</tr>
<tr>
<td>Urea (LMWF)</td>
<td>400-1800 cm⁻¹</td>
<td>2.52-78.99 mg/dL</td>
<td>0.92</td>
<td>1.7 mg/dL</td>
<td>2.9 mg/dL</td>
</tr>
<tr>
<td>Urea (LMWF)</td>
<td>800-1030 cm⁻¹</td>
<td>2.52-78.99 mg/dL</td>
<td>0.97</td>
<td>1.1 mg/dL</td>
<td>2.5 mg/dL</td>
</tr>
</tbody>
</table>

218
6.5 Discussion

The study demonstrates that Raman spectroscopy, coupled with ultra-filtration and PLSR analysis, can be employed to detect variations in multiple analytes from the same serum samples with a high degree of higher accuracy. The routine serum analysis techniques used for the measurement of total protein (biuret method) and γ-globulin (RID and Turbidometric Immuno assay (TIA)) are expensive, require technically competent laboratory personnel and cannot measure multiple analytes from the same serum samples (91). The strategy demonstrated in this study enables the simultaneous estimation of total protein level and detection of imbalance in γ-globulin concentration accurately from whole serum, without the use of any reagents and without destroying the sample being studied. The proposed method has many advantageous over the biuret method, as the required sample volume can be as low as 10μL, it is rapid and non-destructive to the medium being studied, whereas the biuret method is considered impractical due to the requirement of large sample volume and laborious sample processing steps (1,92). Current methods used for determination of γ-globulins are time and temperature dependent and are time consuming, which is a major impediment for early diagnosis (93–96). Moreover, the linearity reported for a comparative study of RID and TIA is 0.59 (1), lower than the R² calculated in the present study. The gold standard methods used for albumin measurements, such as BCG and Bromocresol purple assays have been reported to be susceptible to overestimation of albumin content, especially at lower albumin concentrations (97,98). In this study, it has been demonstrated that PLSR can be used to extract concentration predictions in order to build a prediction model from blood serum. With the use of filtration, the albumin was isolated from patient serum and the
prediction model had a prediction accuracy significantly superior to that of the BCG method (2.2 g/dL) used to determine albumin concentrations from cirrhotic patients (99).

Serum fractionation considerably enhanced the capability to detect and quantify LMWF analytes (9) and it has been shown that reducing the spectral region helped to avoid the spectral interference from other analytes and yielded improved accuracy (64). It has previously been reported that selecting the spectral region from 1030 cm\(^{-1}\) to 1400 cm\(^{-1}\) improved the sensitivity and specificity for the prediction model of glucose from patient samples over the concentration range 52.5-434.2 mg/dL, and the technique was demonstrated to be at least as accurate as ATR-FTIR of similar patient samples (9), measured in the dried state and closer to the accuracy of colorimetric methods, 1.4 mg/dL urea (100) and 2 mg/dL glucose (101). As anticipated, similarly higher prediction accuracy (RMSECV=1.14 mg/dL) was attained when PLSR analysis was performed on a reduced range for urea from patient samples over the concentration range 2.52-78.99 mg/dL, compared to the full range (RMSECV=1.73 mg/dL).

The strategy illustrated in this and previous studies enables simultaneous detection of various analytes from human serum using Raman spectroscopy with minimal sample preparation, no labelling and no additional sample drying steps. Depletion of the HMWF is a non-optional step in the investigation of protein imbalances or disease related biomarkers in the LMWF of serum. Mass spectroscopy, the most commonly used technique for LMWF analysis also makes use of the HMWF depletion in order to potentially target small circulating biomarkers (102). While most of these methods found in the literature are based on chemical extraction/precipitation, the impacts of these solvents are unknown and cross contamination from these solvents is highly probable (102–104). Therefore, ultra-filtration coupled with
chemometrics is a viable option to investigate the less studied LMWF and to interpret the
results by overcoming the intricacies of the multidimensional dataset.

Several methods have been previously described in literature for the quantification of
multiple analytes in serum samples. Perez-Guaita et al., established models for the
determination, in serum samples, of albumin, γ-globulin, total globulin, and
albumin/globulin coefficients, using ATR-FTIR. Values of RMSECV determined for
albumin (126 mg/dL) and γ-globulin (138 mg/dL) are comparable, or larger than the
RMSECV determined in the current study (albumin 90 mg/dL, γ-globulin 126 mg/dL).
Although the spectral co-efficient of prediction of Figure 7.5C is somewhat unsatisfactory,
in that the specific distinctive features of γ-globulin are not prominent, the γ-globulin levels
are not correlated with albumin levels, and the predictive model shows a relatively high
degree of linearity and low standard deviation, giving some assurance that it is indeed
predicting γ-globulin levels. In the study reported by Roy et al., ATR-FTIR spectroscopy
enabled the simultaneous quantification of glucose and urea analytes along with malaria
parasitemia quantification from a spectrum obtained from a dried drop of blood on a glass
slide. The specificity for the PLS-DA was found to be 98% for parasitemia levels but low
sensitivity of 70% was achieved because of the negative samples in the model. The RMSECV
for parasite concentration (0-5%), glucose (0-400 mg/dL) and urea (0-250 mg/dL) spiked
samples were 0.58%, 16% and 17%, respectively (56), whereas the present study showed a
considerably lower RMSECV value of 1.84 mg/dL (52-440 mg/dL) and 1.69 mg/dL (2-79
mg/dL) for glucose and urea from patient samples. Another study conducted by Shaw et al.,
based upon the infrared spectra for 8 serum analytes, reported standard errors of 2.8 g/L (total
protein), 22 mg/dL (albumin), 0.23 mmol/L (triglycerides), 0.28 mmol/L (cholesterol), 7.4
mg/dL (glucose) and 6.6 mg/dL for urea, with correlation coefficients of 0.95 (55). In an investigation conducted on 247 serum donors by Rohleder et al. using Raman spectroscopy of $\lambda_{ex}$-785nm, the standard errors for 7 analytes were reported as 176mg/dL (total protein), 20.7mg/dL (triglycerides), 11.0mg/dL (high density lipoprotein), 15.7mg/dL (low density lipoprotein), 0.81mg/dL (uric acid), 2.1mg/dL (urea) and 6.8mg/dL (glucose) and the efficiency of ultra-filtration technique in improving the prediction accuracy of glucose and urea was demonstrated (83). Although the samples were dried in all these, superior prediction accuracy was afforded by Raman analysis of liquid samples. Berger et al. demonstrated the use of Raman spectroscopy to measure concentrations of serum and whole blood components to simultaneously predict the content of 6 analytes in serum from a 66 patient data set. The quoted prediction errors for albumin, urea and glucose were 120 mg/dL, 3.8 mg/dL and 26 mg/dL, respectively (57), and therefore the combination of improved measurement protocols, serum fractionation, and sectioning the spectral region for regression results in considerable improvement in the prediction accuracy.

It should be noted that direct comparisons of the two techniques of ATR-FTIR and Raman, and even different studies using the same technique, can only be tenuous at best, given the number of potential variables in samples, measurement protocols and data analysis techniques. Notably, a systematic study of the PLSR model construction, validation and testing protocols has not as yet been carried out. The previous study of Parachalil et al., (11) tried to minimise these variabilities by utilising the same sample preparation/processing as well as data analysis techniques in the direct comparison of ATR-FTIR of dried serum with Raman of the liquid state, for the quantification of glucose in serum. The present study has further demonstrated the capacity for prediction of serum content for total protein, $\gamma$-globulin,
albumin, urea and glucose from the same patient samples, with a high degree of accuracy, consolidating the prospect of establishing Raman spectroscopy as a biomedical tool to rival and/or augment conventional approaches such as Mass spectroscopy or chromatography, currently used to deliver crucial information relevant for diagnosis. Further improvements in the sensitivities and variabilities of the techniques ultimately rely on the reproducibility of the measurement, and the signal to noise ratio. The former could potentially be improved by an automated focussing and sampling methodology, while the latter is largely an instrumental consideration. As longer accumulation times to reduce the noise are not recommended, because of sample evaporation and also speed of measurement throughput, optimisation of instrumentation for higher signal throughput could be explored, for example by sacrificing spectral resolution.

6.6 Conclusion

In summary, the potential of Raman spectroscopy combined with filtration and chemometrics to detect variations in total serum protein, γ-globulin, HMWF and LWMF analytes from the same patient serum is successfully demonstrated. Firstly, the prediction model was built in the spiked samples of the analytes in water and was then translated to the patient serum. Although Raman spectroscopy can build quantitative models with higher accuracy in spiked samples, the analysis of patient samples can be affected by numerous factors such as multi-parametrical variability, normally observed in clinical application. Nevertheless, the proposed approach successfully built the prediction model from the whole serum (total protein, γ-globulin), concentrate (albumin) by removing β-carotene and filtrate (urea and glucose) of patient samples at higher accuracy and sensitivity. Given its low cost, easy
implementation, rapid results and higher sensitivity and specificity, the technique may become an alternative for the existing laboratory screening methods. Furthermore, the methodology presented in this work can be applied to a wider range of bodily fluids and can be implemented as a next generation point of care biochemical diagnostic tool in a clinical setting.
6.7 References


42. Choi S, Choi EY, Kim DJ, Kim JH, Kim TS, Oh SW. A rapid, simple measurement


52. Ginsberg BH. Factors affecting blood glucose monitoring: Sources of errors in


Biointerfaces. 2018;171(June):123–33.


6.8 Electronic Supplementary information:

1. No correlation was found between the concentrations of globulin and albumin in patient samples

![Figure 6.S1](image1.png)

**Figure 6.S1.** Plot of the concentrations of albumin and immunoglobulin for each patient

2. No correlation was found between the concentrations of urea and glucose in patient samples

![Figure 6.S2](image2.png)

**Figure 6.S2.** Plot of the concentrations of urea and glucose for each patient
3. Reference spectrum of β-carotene

**Figure 6.3.** Raman spectrum of β-carotene used for EMSC correction of human serum from patient samples

4. Reference spectrum of human serum

**Figure 6.4.** Raman spectrum of human serum used for EMSC correction of total protein and globulin from patient samples
5. PLSR was performed on shorter spectral region of γ globulin from patient serum samples.

**Figure 5.55.** (A) EMSC corrected Raman spectra of γ globulin of patient serum samples from 800 cm\(^{-1}\) to 980 cm\(^{-1}\) (329 mg/dL, 690 mg/dL, 836 mg/dL and 1404 mg/dL), the spectra has been offset for clarity. (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient for regression against γ globulin concentration shows the peaks of γ globulin, (D): Linear predictive model for γ globulin concentration built from the PLSR analysis.
6. PLSR was performed on varying concentration of albumin in water (5-50mg/dL)

Figure 6.S6. (A): Rubberband corrected Raman spectra of varying concentrations of Albumin from 5mg/mL to 50mg/mL (500mg/dL to 5000mg/dL) in distilled water, (B): Evolution of RMSECV on the validation model, (C): plot of PLSR coefficient with Albumin features, (D): Linear predictive model built from the PLSR analysis. The RMSECV is calculated as 1.58mg/mL (158mg/dL)
7. PLSR result of varying concentrations of glucose in distilled water (100-1000mg/dL) (2)

![Graph A](image1.png)  ![Graph B](image2.png)  ![Graph C](image3.png)  ![Graph D](image4.png)

**Figure 6.S7.** (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of varying concentrations of glucose (5 x100mg/dL, 5 x 450mg/dL and 5 x 1000mg/dL, spectra offset for clarity), in distilled water and signature peaks of glucose are highlighted with asterisks, (B): Evolution of the RMSECV on the validation model, (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and $R^2$ values were calculated as 10.93mg/dL and 0.9705 respectively.
8. PLSR analysis performed on the filtrate collected after centrifugal filtration of glucose spiked in serum samples using 10kDa filters (2)

**Figure 6.58.** (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of glucose spiked in serum (spiked concentrations 5 x 0mg/dL, 5 x 120mg/dL and 5 x 220mg/dL, offset for clarity) and the signature peaks of glucose are highlighted by asterisks, (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and R² values were calculated as 1.66mg/dL and 0.9914
9. PLSR analysis performed on the filtrate collected after centrifugal filtration of patient samples using 10kDa filters (2)

Figure 6.S9. (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of patient samples (5 x 52.25mg/dL, 5 x 75.67mg/dL, 5 x 93.69mg/dL, 5 x 210.81mg/dL and 5 x 434.35mg, offset for clarity) and the signature peaks are marked by asterisks, (B): Evolution of RMSECV of the data set, (C): plot of PLSR coefficient with glucose features, (D): Predictive model built from the PLSR analysis. The value displayed in the PLSR model is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. The RMSECV and $R^2$ values were calculated as 1.84mg/dL and 0.84 respectively.
10. PLSR performed on varying concentrations of urea in water

Figure 6.S10 (A): EMSC corrected Raman spectra of filtrate obtained after centrifugal filtration with 10kDa filters of urea spiked in water (1mg/dL to 1000mg/dL), (B): Evolution of RMSECV of the data set (C): plot of PLSR coefficient with urea features, (D): Linear predictive model built from the PLSR analysis. The RMSECV, $R^2$ and overall standard deviation values were calculated as 70.4044mg/dL, 0.9048 and 1.0975mg/dL.

6.8.1 References


Chapter 7

Raman spectroscopy as a potential tool for label free therapeutic drug monitoring in human serum: the case of Busulfan and Methotrexate

The following chapter has been reproduced from the submitted journal article entitled ‘Raman spectroscopy as a potential tool for label free therapeutic drug monitoring in human serum: the case of Busulfan and Methotrexate’, Analyst, 2019.

**Author List:** Drishya Rajan Parachalil, Deirdre Commerford, Franck Bonnier, Igor Chourpa, Jennifer McIntyre, and Hugh J. Byrne

DRP performed all experimental analysis, with the assistance of DC, and authored the publication. FB, IC, JMcC and HJB input to conceptual design of the work and drafting and proofing of the manuscript
7.1 Abstract

A methodology is proposed, based on Raman spectroscopy coupled with multivariate analysis, to determine the Limit of Detection (LOD) and Limit of Quantification (LOQ) for therapeutic drug monitoring in human serum, using the examples of Busulfan, a cell cycle non-specific alkylating antineoplastic agent, and, Methotrexate, a chemotherapeutic agent and immune system suppressant. In this study, ultra-filtration is employed to fractionate spiked human pooled serum to efficiently recover the drug in the filtrate prior to performing Raman analysis. The drug concentration ranges were chosen to encompass the recommended therapeutic ranges and toxic levels in patients. Raman spectra were collected from the filtrates in the liquid form, using an inverted backscattering microscopic geometry, using 532nm as source. Finally, prediction models were built by using Partial Least Squares Regression (PLSR) and LOD and LOQ were calculated directly from the linear prediction models. The LOD calculated for Busulfan is 0.0002 ± 0.0001 mg/mL, 30-40 times lower than the level of toxicity, enabling the application of this method in target dose adjustment of Busulfan for patients undergoing, for example, bone marrow transplantation. The LOD and LOQ calculated for Methotrexate are 7.8 ± 5 µM and 26 ±5 µM, respectively, potentially enabling high dose monitoring. The promising results obtained from this study suggest the potential of Raman spectroscopy for therapeutic drug monitoring of drugs in bodily fluids.

7.2 Introduction

Therapeutic drug monitoring (TDM) refers to the clinical practice of management of a patient's drug dosage within a targeted therapeutic window, based on measurement of
concentration of the drug in the bloodstream at timed intervals. For drugs with a narrow therapeutic range, such monitoring is essential to provide individualised patient treatment, while maintaining the efficacy of drugs and minimising drug toxicity and related adverse effects (1,2).

TDM has also been increasingly advocated to improve the standard of chemotherapy, in which side effects can be substantial and life threatening (3–6). The currently available technique of chemotherapeutic-dosage calculation based on dose intensity and body surface area has been reported to be inaccurate for patients undergoing sustained chemotherapeutic treatment (7,8). In the era of rising cost of healthcare, it is necessary to develop a rapid, sensitive, and cost-effective, point-of-care technique for TDM, which can quantitatively measure the serum concentration of drugs, such that the dosing strategy can be tailored to the metabolism of an individual patient for a personalised therapeutic regime.

Busulfan (Bu) is a bi-functional alkylating agent (see chemical structure in inset of Figure 2A) used in the chemotherapy-based conditioning regimen for hematopoietic stem cell transplantation (HSCT) (9–17). Bu has a very narrow therapeutic index, and higher systemic exposure to Bu is related to hepatic sinusoidal obstruction syndrome, neurotoxicity or interstitial pneumonia, while low levels have been shown to be associated with increased incidence of graft rejection (18–21). Measurement of individual serum Bu levels during oral or intravenous dosing is likely to provide the necessary elements to monitor the drug disposition, ensuring efficacy, reduced incidences of toxicity and graft rejection (18–22). Several analytical methods, including chromatographic techniques coupled with a number of detection methods, have been described for analysing Bu in biological fluids; Gas chromatography (GC) with electron capture detection, high performance liquid
chromatography (HPLC) with UV detection, GC-Mass spectroscopy (MS) with selected ion monitoring, and Enzyme linked immunosorbent assays (ELISA) have been reported to have high sensitivity for monitoring Bu in biological fluids (10,11,23–26). However, the translation of these techniques to a routine analytical tool in a clinical setting for TDM is impractical, owing to their complexity and cost.

Methotrexate (MTX) is a folate antagonist (see chemical structure in the inset of Figure 2B) widely used as an anti-cancer agent to treat various malignancies, such as leukemia, breast cancer, lymphomas and autoimmune diseases (27). MTX is administered in both low and high dosage (LDMTX and HDMTX), and monitoring serum MTX concentrations is essential to avoid high dosage related side effects (28). Serum MTX concentrations can vary from 10 nM to 1 mM for different patients, due to pharmacokinetic variability (28). The serum MTX concentration should reach between 10µM (0.001mM) and 100µM (0.01mM) after 12-36 hours of HDMTX infusion and should reduce to 0.2µM after 72 hours. From the clinical point of view, it is essential to be able to detect the serum concentrations of MTX between 0.1µM and 10µM, as high toxicity related adverse effects are associated with concentrations >10µM (28,29). Various sophisticated analytical tools such as Enzyme multiplied immunoassay technique (30), radioimmunoassay (31), enzyme exhibition assays (32), capillary zone electrophoresis (33) and liquid chromatography coupled with tandem mass chromatography (HPLC-MS/MS)(6,34–39) have been reported for TDM of MTX from biological fluids. Although immunoassays (40) and separation techniques (38) are routinely employed due to, they suffer from major limitations such as interferences from other compounds and lack of availability for all the drugs currently monitored (4). HPLC-MS/MS
is considered the gold standard method for MTX (35-39) analysis due to its high sensitivity and robustness; however, it is time consuming, expensive and requires skilled personnel.

In recent years, Surface enhanced Raman spectroscopy (SERS) has been reported to be a good candidate for TDM of MTX (28,29,41), doxorubicin (42), paclitaxel and cyclophosphamide (3) in biological fluids, since quantitative analysis of drugs can be performed rapidly and higher sensitivity. By comparison, quantification of Bu in biological fluids using spectroscopic techniques has not been explored. Critical issues of using SERS for TDM include development of standardised substrates, intense surface enhanced resonance SERS responses from other biological molecules such as carotenoids and also the spectral interference from the fluorescence that could interfere with the drug detection (28,43). Therefore, new techniques that are inexpensive, less complex and faster are essential to quantitatively determine the concentration of drugs in a clinical setting. Herein, a rapid drug screening strategy using Raman spectroscopy coupled with ultra-filtration and multivariate analysis technique for Bu and MTX from liquid serum that yields a significant improvement in detection capabilities and minimises error is explored.

7.3 Materials and Methods

7.3.1 Materials

Methotrexate (A6770), Busulfan (B058) and human pooled serum (H6914) were purchased from Sigma Aldrich, Ireland. Stock solutions of 0.1mg/mL Bu in methanol and 1mM MTX in 0.1M NaOH were prepared. The spiked concentrations of Bu in serum are expressed in mg/mL and MTX in µM to be consistent with previous studies (4,29). The commercial
human serum was spiked with Bu and MTX over the therapeutically relevant concentration ranges, to achieve the final concentrations of (0 - 0.05 mg/mL) for Bu and (0 – 100 µM) for MTX. The normal therapeutic range for Bu is 0.0005mg/mL to 0.005mg/mL and any concentration below 0.0005mg/mL can cause transplant failure, or higher than 0.005mg/mL, transplant related mortality (11), whereas for MTX, 1µM to 10µM and <10µM is considered toxic (29). Raman spectra of highly concentrated Bu and MTX drug solutions prepared with a minimal amount of water (~1mg/mL) are used as the reference for the Extended Multiplicative Signal Correction algorithm (see Data pre-processing and analysis).

Amicon Ultra 0.5mL centrifugal filter devices (Millipore- Merck, Germany), with a cut off point of 10kDa, were employed to fractionate the serum samples. The centrifugation procedure previously reported by Bonnier et al. was followed (44). The optimised washing and rinsing procedure includes spinning 0.5mL 0.1M NaOH at 14000×g for 30 minutes, followed by three rinses with distilled water by spinning 0.5mL distilled water for 30 minutes at 14000×g. Every 30 minute wash and rinse must be followed by spinning the device in the inverted position at 1000×g for 2 minutes, to remove the residual solution contained in the filter. After washing, 0.5mL of spiked serum solution is transferred to the 10kDa filter and centrifuged at 14000×g for 30 minutes. The filtrate that passes through the 10kDa filter contains mostly water and molecules smaller than 10kDa. All the filtrate solutions were analysed using Raman spectroscopy and five replicate measurements from different positions have been recorded. In subsequent analysis, each dosed serum sample is represented by all the spectra recorded from that sample, rather than the mean.
7.3.2 Raman spectroscopy

The measurement conditions used for screening analytes in human serum in the liquid form have recently been detailed (45,46). Raman spectra of all the liquid serum filtrate samples and references were recorded at stabilised room temperature (18ºC) using a Horiba Jobin-Yvon LabRam HR800 spectrometer with a 16-bit dynamic range Peltier cooled CCD detector. A 532nm laser was used, which had a power of ~30 mW at the sample, with a 600 lines/mm grating and the backscattered Raman signal was integrated for 3×80 seconds over the spectral range from 400-1800 cm\(^{-1}\). The spectrometer was coupled to an Olympus 1X71 inverted microscope and a x60 water immersion objective (LUMPlanF1, Olympus) was employed. The substrate used was a Lab-Tek plate (154534) with a 0.16-0.19mm thick, 1.0 borosilicate glass bottom, and was purchased from Thermo Fischer Scientific, Ireland.

7.3.3 Data pre-processing and analysis

The raw spectra were subjected to pre-processing techniques in Matlab before further analysis, to remove the background signal and reduce the noise. Smoothing of the raw data was done using the Savitzky–Golay method (polynomial order of 5 and window 13) and the rubberband method (45) was found to be appropriate to baseline correct the smoothed reference spectra of both the drugs. The ‘rubberband’ correction was carried out by wrapping a ‘rubberband’ of defined length around the ends of the spectrum to be corrected and fitting against the curved profile of the spectrum. An adapted Extended Multiplicative Signal Correction (EMSC) algorithm (47), with a 3rd order polynomial, was applied to remove the
underlying water spectrum from all the dataset, whose OH bending feature at 1640 cm\(^{-1}\) can interfere with the analyte spectra, and also scales the analyte spectra, assuming a constant water contribution to all sample spectra (47).

### 7.3.4 Partial Least Squares Regression

Partial Least Squares Regression (PLSR) was employed to establish a model that relates the variations of the spectral data to a series of concentrations. This regression model can be used to establish the limit of detection and quantitation of Raman bio-sensing of drugs (48,49). Constructed based on the spectra of samples of known drug content, over a range of varying concentrations of drug (in commercial serum), the model is then validated using a rigorous cross validation procedure which evaluates its performance in accurately predicting drug concentrations. For consistency with previous studies (45,46), a 20 fold cross validation approach has been employed to validate the robustness of the method. This approach involves randomly dividing the set of observations into approximately equal size, 50% of the spectral data were randomly selected as test set, while the remaining 50% is used as the training set (50). The cross-validation process is then repeated 20 times (the folds), whereby all observations are used for both training and testing, and each observation is used for testing exactly once. The results from the folds can then be averaged to produce a single estimation. The Root Mean Square Error of Cross Validation (RMSECV) is calculated from the 20 iterations to measure the performance of the model for the unknown cases within the calibration set. The correlation between the true and predicted concentrations is given by the \( R^2 \) value. The standard deviation was calculated to quantify the amount of variation in the
dataset. The number of latent variables used for building the PLSR model is optimised by finding the value that is equivalent to the minimum of the RMSECV. The Limit of Detection (LOD) and Limit of Quantification (LOQ) of these two drugs for this method were calculated from the PLSR prediction plot, using a IUPAC-consistent approach previously reported for multivariate regression analysis by Ostra et al. (48).

\[
\text{LOD} = 3 \times S_{\text{blank}} \times b 
\]

\[
\text{LOQ} = 10 \times S_{\text{blank}} \times b 
\]

where, \( S_{\text{blank}} \) is the standard deviation of a blank (zero concentration sample) and \( b \) is the slope of the regression (inverse calibration) model, in the region of linearity. The slope was calculated for the linear region of the prediction plot, including the standard deviation of each point, by initially regressing over the higher concentrations, and progressively adding smaller concentrations to the regression range, until the calculated slopes were seen to begin to reduce.

### 7.4 Results and Discussion

Figure 7.1 shows the schematic diagram of the strategy used to collect the Bu and MTX data from the serum samples to build the prediction models.
Figure 7.1. Schematic representation of the ultra-filtration, Raman analysis, data pre-processing and PLSR analysis of the Bu/MTX serum samples. LOD and LOQ is calculated from the prediction plot.

The advantages of employing an inverted geometry to record Raman spectra have been detailed by Bonnier et al (51). The feasibility of using a Lab-Tek plate as substrate (45) and impact of ultra-filtration coupled with multivariate analysis techniques in detecting low molecular weight fraction analytes have also previously been reported (46,52,53). The Raman spectra recorded from the 10kDa filtrate of Bu and MTX spiked serum samples were subjected to pre-processing steps followed by PLSR analysis. The whole finger print region (400-1800cm\(^{-1}\)) was chosen to build PLSR prediction for Bu, whereas a shorter region, in which there are strong bands of MTX (1200-1800cm\(^{-1}\)), was chosen to facilitate efficient prediction of MTX by increasing the sensitivity. The improvement of the sensitivity of the prediction model for the case of glucose and urea when regressed over a reduced spectral region was previously reported (46,54). The normal and toxic ranges of the Bu and MTX were encompassed by the range of spiked serum samples.

Figure 7.2A shows the reference spectrum of Bu, and the signature peaks of Bu at 1097cm\(^{-1}\) and 1453cm\(^{-1}\), which can be ascribed to a CH\(_2\) scissoring mode and C-C stretching, respectively (55). In the case of MTX, the signature peaks are a strong band at 1593cm\(^{-1}\),
which can be ascribed to the scissoring of the NH$_2$ group, while the sharp band at 1351cm$^{-1}$ can be ascribed to CH$_2$ scissoring vibrations (Figure 7.2B) (4,29,41). The pre-processed data set of systematically varied concentration of spiked, filtered serum (Figure 7.S1A and B) is fed into the PLSR algorithm to build a prediction model that correlates the known concentration and the predicted concentration, based on the variation in spectral intensity, for each drug. On the basis of the percent variance explained by the latent variables and the lowest value of RMSECV, the total number of latent variables used to reach the best performance is calculated to be 3 for both Bu and MTX (Figure 7.S2A and B). The RMSECV values are calculated to be 0.0003mg/mL for Bu and 4.02µM for MTX. The PLSR coefficient plots (Figure 7.2C and D) of Bu and MTX display Raman bands in good accordance with the reference ones, namely, the presence of peaks at 1097cm$^{-1}$ and 1453cm$^{-1}$ for Bu, and 1593cm$^{-1}$ and 1351cm$^{-1}$ for MTX, respectively.
Figure 7.2. Reference spectrum of Bu used for EMSC correction. The chemical structure of Bu is shown in the inset. B: Reference spectrum of MTX used for EMSC correction. The chemical structure of MTX is shown in the inset. C: PLSR coefficient plot of Bu (400-1800 cm\(^{-1}\)) from serum filtrate concentrations showing spectral features similar to the Bu reference at 1097 cm\(^{-1}\) and 1453 cm\(^{-1}\). D: PLSR coefficient plot of regression against MTX from 1200-1800 cm\(^{-1}\) showing spectral features similar to the MTX reference at 1351 cm\(^{-1}\), and 1593 cm\(^{-1}\). RMSECV were calculated to be 0.0003 mg/mL for Bu and 4.02 µM for MTX, respectively.
Figure 7.3. Linear predictive model for (A) Bu and (B) MTX built from the PLSR analysis. The LOD and LOQ for Bu were calculated to be 0.0002±0.0001mg/mL and 0.00073±0.00010mg/mL, whereas the LOD and LOQ of MTX were calculated to be 7.8 ±5.0 µM and 26 ±5µM.

Figure 7.3A and B indicate that the concentration dependence of the sample set is conserved by centrifugal filtration and a satisfactory linear model could be obtained for Bu and MTX from the filtrate of the serum samples. A linear prediction plot with a correlation accuracy ($R^2$) of 0.97 was obtained for Bu, with an LOD of 0.0002 ±0.0001mg/mL and LOQ of 0.00073 ±0.0001mg/mL ($b=0.96$ and $s_{blank}=0.00008mg/mL$), both in the acceptable range of clinical use. Samples with Bu concentrations higher than 0.002mg/mL are frequently observed in many hospitals, but, Bu concentrations less than 0.0005mg/mL are rarely seen (11). This Raman spectral response was validated to be linear over the entire range of 0.0003mg/mL to 0.0125mg/mL, which has not been fully validated in previous studies. In comparison to the earlier studies based on LC-MS, the present method shows similar performances in precision and recovery (13,16,24,56). However, considering the time needed for sample preparation for the chromatography based methods, the present method has an added advantage that it does not require complex sample preparation steps. Besides, the
amount of sample required for this method (1-50 μL) is slightly less than that for other commonly employed methods (50 to 200 μL)(11). Thus, this proposed approach can be expeditiously implemented in laboratories in clinical settings for introduction of TDM of Bu to achieve safe and proper dosing.

Similarly, the correlation accuracy (R^2) is as high as 0.96 for MTX. The LOD was calculated to be 7.8 ±5.0μM and the LOQ to be 26 ±5 μM (b=0.94 and S_{blank} = 2.8 μM). High risk of toxicity related adverse effects are associated with serum MTX concentrations of >10μM (4,29). The concentrations outside of safety values of MTX are >10 μM at 24 hours or >1 μM at 48 hours, and the serum MTX concentration should drop down to 0.2μM after 72 hours to reach the safety value (4,29,57,58). For most drugs, the process of drug elimination is a first-order rate process, and so, in a given patient, can be characterised by a rate constant (59). Therefore, from the clinical point of view, regular monitoring of MTX levels in patient serum can be used to determine a rate of drug elimination and help establish a personalised dosing regime for each patient. In previous studies, many researchers have reported the use of SERS substrates to detect MTX in plasma/serum with a LOD as low as 0.17μM (29). Preparation of these substrates are time consuming and expensive and are prone to experimental errors, however (43). In contrast, the proposed method with inverted Raman spectroscopy is cost-effective and easy to use, that can be translated as a point-of-care diagnostic tool for high-dosage MTX in bodily fluids. Notably, in determining the LOD and LOQ, while the slope of the concentration dependent response is dependent on the Raman scattering cross section of the analyte, the standard deviation of the blank is a measurement parameter, and instrument specific, and could potentially be improved by reduced noise and/or signal variability.
7.5 Conclusion

In summary, a rapid, sensitive, cost effective and reproducible method to determine the Bu and MTX levels in human serum has been demonstrated. In clinical practice, identification of individual therapeutic concentration of drugs is crucial for specific drugs with narrow therapeutic window by measuring the levels of these drugs at designated intervals in the serum/plasma, as the drug concentration in serum/plasma largely varies with time for different individuals based on their age, body weight, pregnancies, temporary illnesses, infections, emotional and physical stresses, accidents, and surgeries (12,58,60–62). TDM takes these factors into consideration and accommodates them while establishing an individual therapeutic concentration to fit the specific needs of a patient. This simple approach of Raman spectroscopy coupled with ultra-filtration and multivariate analysis technique allows to effectively preserve the information in the filtrate while enabling easy detection of the drug concentration with higher accuracy. This strategy could be widely adopted for monitoring a variety of other drugs and small molecules. The present method accurately determines MTX concentrations at 7.8±5.0µM, suggesting that this method can be applied for high dose monitoring of MTX. On the other hand, this method determined the concentration of Bu as low as 0.0002±0.0001mg/mL, which is 30-40 fold below the lowest Bu level that may present a risk for toxicity (19), thus ensuring effective and safe therapy for patients undergoing bone marrow transplant. Therefore, this can be a useful protocol for TDM of Bu to achieve safe and appropriate dosing. Further studies are needed to investigate the determination of these drugs in patient serum to ensure successful implementation of this method as a diagnostic tool. Thus far, this study is a proof of concept that simple Raman
spectroscopy combined with multivariate analysis technique and ultra-filtration has the potential to be used as a diagnostic tool for therapeutic drug monitoring from human serum.
7.6 References

10. Desire S, Mohanan EP, George B, Mathews V, Chandy M. A rapid & sensitive liquid chromatography- tandem mass spectrometry method for the quantitation of busulfan levels in plasma & application for routine therapeutic monitoring in


17. Ringde O, Ljungman P, Hassan M. High busulfan concentrations are associated with
increased transplant-related mortality in allogeneic bone marrow transplant patients. Bone Marrow Transplantation, 1997;20:909–913


7.7 Electronic Supplemental information

1. EMSC corrected dataset of Busulfan and Methotrexate

Figure 7.S1. EMSC corrected and smoothed dataset of A Busulfan from 0mg/mL to 0.05mg/mL in the fingerprint region. B: Methotrexate from 0µM to 500µM from 1200cm⁻¹ to 1800cm⁻¹. The spectra are offset for clarity.
2. Root Mean Square Error Cross Validation (RMSECV) plots for Busulfan and Methotrexate

Figure 7.82. Evolution of RMSECV for (A) Busulfan and (B) Methotrexate. The RMSECV values are calculated to be 0.0003mg/mL for Bu and 4.02µM for MTX
Chapter 8

Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: Recent advances

This chapter has been reproduced from the submitted article entitled ‘Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: Recent advances’, Analytical and Bioanalytical Chemistry, 2019.

Author List: Drishya Rajan Parachalil, Jennifer McIntyre, and Hugh J. Byrne

DRP performed all experimental analysis and authored the publication. JMcC and HJB input to conceptual design of the work and drafting and proofing of the manuscript.

8.1 Abstract

There is compelling evidence in the literature to support the application of Raman spectroscopy for analysis of bodily fluids in their native liquid state. Naturally, the strategies described in the literature for Raman spectroscopic analysis of liquid samples have advantages and disadvantages. Herein, recent advances in the analysis of plasma/serum in the liquid state are reviewed. The potential advantages of Raman analysis in the liquid form over the commonly employed infrared absorption analysis in the dried droplet form are initially highlighted. Improvements in measurement protocols based on inverted microscopic
geometries, clinically adaptable substrates, data preprocessing and analysis, and applications for routine monitoring of patient health as well as therapeutic administration are reviewed. These advances suggest that clinical translation of Raman spectroscopy for rapid biochemical analysis can be a reality. In the future, this method will prove to be highly beneficial to clinicians for rapid screening and monitoring of analytes and drugs in the biological fluids, and to the patients themselves, enabling early treatment, before the disease becomes symptomatic, allowing early recovery.

8.2 Introduction
Vibrational spectroscopic techniques, both Raman and Infrared (IR) absorption, have been extensively explored over the last two decades for obtaining the biochemical composition of bodily fluids in the field of biomedical analysis [1–10]. The sensitivity to detect subtle changes in the biochemical composition and ability to detect the presence of specific biomarkers or drugs makes vibrational spectroscopy an ideal tool for the early diagnosis of various pathologies [11–14] and therapeutic drug monitoring [15–17]. Both techniques are truly label free, rapid, cost-effective, easy to operate, non-destructive and provide the unique molecular fingerprint of the sample with minimal sample preparation steps. Such techniques are particularly attractive for routine analysis of biofluids, as they are easy to apply, require minimal sample preparation and are readily adaptable to analysis of various bodily fluids [1-5], potentially reducing clinical analysis time, and alleviating patient angst (figure 1).

Over the past decades, there have been numerous studies of analytes in biofluids using vibrational spectroscopy, and, in recent years, attenuated total reflection (Fourier Transform) IR (ATR-FTIR) has become popular for rapid screening of biofluids, particularly blood plasma and serum [13]. Notably, however, ATR-FTIR is predominantly conducted on dried
droplets of bodily fluids, adding to the complexity of the measurement and the clinical workflow [18,19]. In comparison, the prospect of using Raman spectroscopy for the label-free extraction of biochemical information from biological fluids is attractive from various perspectives; liquid sample analysis, no requirement for additional reagents, ease of use, speed, cost-effectiveness and low sample volume requirement. The application of Raman spectroscopy to biomolecules and even tissues was first demonstrated as early as the 1960s, and by the mid 1970s biomedical applications were explored [20–22]. Whole cell and tissue studies have been carried out on a range of pathologies [23–25] and in vivo studies [26,27] have demonstrated the prospective for diagnostic applications. Raman microspectroscopy potentially lends itself naturally to the analysis of liquid biofluids and such applications have attracted considerable attention in recent years. There have been, however, multiple strategies that are used to perform such real-time analysis and these options must be carefully considered to achieve optimum Raman spectra from liquid samples [28,29]. Several studies have reported the proof-of-concept of liquid sample analysis using Raman spectroscopy [30–37]. However, no systematic validation and testing of protocols has yet been carried out to consider this technique for real-time clinical applications. This review will summarise the recent advances in the standardisation of measurement protocol for biological fluid analysis in the liquid state using Raman spectroscopy, in terms of the optimal wavelength and substrate, serum fractionation methods, data collection, pre-processing and post processing steps to obtain results with higher accuracy and sensitivity. Applications for both analysis of imbalances of high and low molecular weight serum constituent components of pathological and clinical significance, as well as therapeutic drug monitoring will be considered. Note, that the studies considered do not require signal enhancement techniques such as surface
enhanced Raman spectroscopy, and therefore such techniques are not considered within the scope of the review. However, as they has been extensively explored in recent times, a brief comparison of the techniques to infrared absorption based techniques is provided.

**Figure 8.1.** Early disease diagnosis, prognosis and treatment is possible with real-time analysis of patient serum using the inverted Raman spectral analysis.

**8.3 Raman vs Infrared absorption spectroscopy**

Although Raman and IR spectroscopy are considered complementary techniques, the fundamental physical phenomena governing them are very different, and thus result in distinct technical challenges to their clinical implementation [6,38]. As IR spectroscopy is based on absorption due to electric dipole transitions associated with molecular vibrations,
water cannot be used as a solvent, due to its intense absorption in the IR region due to the highly polar OH groups [39,40]. IR analysis of bodily fluids has therefore been predominantly performed on air-dried samples, which leads to chemical and physical inhomogeneity due to the so–called “coffee ring” effect and thereby inconsistencies in the results obtained [18,19,39]. Raman spectroscopy is an inelastic scattering technique based on the Raman effect, i.e., the coupling of the oscillating electronic polarisabilities of the molecular bond with the source electromagnetic field [41,42]. The distinct advantage of this technique is that it is compatible with water-rich samples such as serum/plasma, as water molecules have a relatively low scattering cross section and do not mask the scattering from the solutes in the aqueous solutions [43–45]. Zhao et al., developed a fast and reliable approach to identify and quantify liquid injectables in the liquid state for spurious/falsely-labelled/falsified/counterfeit medical products (SFFCs) using Raman spectrophotometer of 785nm excitation source [46]. Principle Component Analysis (PCA) combined with Classical Least Squares (CLS) chemometric methods were used to overcome the problems of the interference signals of glass containers and solutions, and weak signals from active pharmaceutical ingredients were finally extracted. Water was used as an internal standard for normalisation and CLS quantitation models were established. When Raman predicted values were compared with HPLC reference results, the relative error of eight doxofylline liquid injectable samples were within 5% and three low-concentration Levofloxacin in Levofloxacin Lactate and Sodium Chloride Injection samples were within 10%, demonstrating this approach to be a reliable and rapid screening method to detect SFFCs in liquid dosage forms. In order to extend this Raman spectroscopic set up for the study of
biological fluids, further studies should be conducted on biological fluids to extract clinically relevant information.

Multianalyte, dried serum analysis has previously been reported using mid-infrared spectroscopy, for the simultaneous quantitation of eight serum analytes: total protein content, albumin, triglycerides, cholesterol, glucose, urea, creatinine and uric acid [47] and for simultaneous quantification of glucose and urea analytes along with malaria parasitemia quantification using ATR-FTIR [11]. As a direct comparison of the techniques of mid-IR and Raman spectroscopy, together with multivariate data analysis, for the quantitative analysis of serum, Rohleder et al. analysed the serum of 247 blood donors [36]. The IR analysis was undertaken on dried droplets, whereas the Raman analysis was of the liquid serum and/or serum filtrates. Under their investigation for the quantification of glucose, urea, uric acid, LDL cholesterol, HDL cholesterol, total protein, cholesterol and triglycerides, Raman and mid-infrared spectroscopy delivered similar accuracies for the prediction of physiologically relevant analyte levels [36]. More recently, Parachalil et al., undertook a similar comparison of Raman compared to ATR-FTIR, using identical sample preparation and analysis protocols, to quantitatively monitor diagnostically relevant changes of glucose [44], indicating that Raman spectroscopy in the liquid state can perform at least as well as ATR-FTIR, without the need for the drying step.

Notably, whereas many of the previous studies of liquid serum samples employed an upright microscopic geometry, that of Parachalil et al., employed an inverted geometry, as previously demonstrated by Bonnier et al. [62]. Improved analysis of serum using Raman spectroscopy was reported when the sample was analysed in the inverted geometry using a water immersion objective with a 785nm laser and CaF$_2$ substrate. A drop of water is used to
minimise the differences in the refractive indices between sample, objective and the substrate, and thus improve the optical coupling. However, the water drop does not contribute to the data collected, as it is outside the focus of the beam. As a much better cost-effective, clinically adaptable option, Parachalil et al. introduced a commercial, cover slip (of 0.16-0.19 mm thickness) bottomed vesicle (Lab-tek plate) as the substrate. The use of glass precludes the use of a 785nm source [48,49], and thus a 532nm laser was chosen as the source, which provides a strong Raman signal of water with minimal background interference (Figure 2). This set-up also has the added advantage of providing high quality, consistent Raman spectra from a sample volumes as low as 1μL. Medipally et al. also reported the benefits of using inverted Raman geometry to analyse plasma samples of small volume (20μL) from prostate cancer patients [50]. Note, enhanced methods such as Surface Enhanced Raman Spectroscopy (SERS) are not considered in this review.

![Image of Lab-Tek plate and inverted Raman analysis](image)

**Figure 8.2.** Thin glass bottomed Lab-Tek plate combined with inverted Raman analysis can collect spectral data from very low amount of samples (1μL) making it an ideal tool for clinical laboratory analysis. Spectra can be recorded in less than 1 minute.

### 8.4 Measurement of Plasma vs Serum

Human blood plasma or serum are the most commonly studied bodily fluids for disease diagnosis, biomarker discovery and therapeutic drug monitoring [51–57]. While plasma and
serum are both cell-free fluids obtained from blood samples by centrifugation, they differ on the basis of whether clotting has been allowed or not. One question which inevitably arises in performing biochemical estimation of blood, therefore, is what is the right choice of medium to be used, plasma or serum? Plasma is commonly obtained as the supernatant layer after refrigerated centrifugation of blood collected with anticoagulants (such as potassium-Ethylenediaminetetraacetic acid, sodium-citrate and lithium-heparin) for 10 minutes at 2,000 x g to concentrate unwanted cells and platelets [58]. For serum preparation, whole blood is allowed to clot at room temperature for about 15–30 minutes, whereupon the clot is removed by refrigerated centrifugation at 1,000–2,000 x g for 10 minutes, often separated by a gel component to avoid contamination [58]. It is important to immediately transfer the supernatant (plasma or serum) into a clean polypropylene tube and maintain samples at 2–8°C while handling. If the samples are not analysed immediately, they should be stored at –20°C or preferably lower [59]. The time delay between centrifugation and separation of the globular fraction and the improper storage conditions could negatively impact on the protein profile obtained [59]. Therefore, standardisation of sample collection, processing and storage protocols is crucial to ensure reproducibility and consistency in the results obtained. It is also recommended to avoid freeze-thaw cycles, because this may have detrimental effects on many serum components [60].

Blood serum and plasma are predominantly composed of water (~90%), minerals, organic substances and gas (oxygen, carbon dioxide) [60]. Proteins are the predominant molecular components of blood plasma, the remaining constituents being carbohydrates, lipids and amino acids. Serum albumin, globulins, fibrinogen and a handful of other abundant proteins account for 99% of total plasma/serum proteins, while the remaining 1% is composed of low
abundance circulatory proteins [54]. Additionally, plasma or serum contain more than 114,000 known metabolites at varying concentration level (<1 nmol/L to mmol/L) (60).

Since most of the clinical analytical instruments are accurate for both serum and plasma, these two terms are used erroneously interchangeably in most clinical tests [7]. Notably, many studies that have been reported to be carried out in serum were in fact carried out in plasma [61–66].

Medipally *et al.* investigated the effect of different instrumental and sample preparation parameters to identify a combination that would reduce the overall acquisition time for recording spectra from blood plasma with minimal of sample preparation steps [50]. Out of the four different laser lines (785 nm, 660 nm, 532 nm and 473 nm) tested, only the 785 nm laser line gave a reliable biochemical signature of liquid plasma samples. Fluorescence was observed when the 660 nm laser line was used and a resonance Raman effect due to the presence of β carotene was observed when 532 and 473 nm laser lines were used. Plasma samples from 10 prostate cancer patients and 10 healthy volunteers were used in this study. A 96 well plate (cover glass bottomed) was used to hold samples of 20 µL and the Raman spectra were recorded in the inverted geometry. Spectral preprocessing steps and principal component analysis – linear discriminant analysis (PCA-LDA) was performed in the R environment. The classification resulted in a sensitivity and specificity of 96.5% and 95% respectively. Although a cost effective approach to perform rapid analysis of liquid plasma is demonstrated in this study, no attempts have been made to reduce the effect of spectral interferents while using the 532 nm laser line.

In validating the protocols for analysis of bodily fluids using inverted Raman microscopic analysis using 532 nm, Parachalil *et al.* utilised a simulated blood plasma mixture of albumin,
fibrinogen, cytochrome C, and vitamin B12 [44]. The findings from this study with simulated plasma protein mixture show that the poorly soluble fibrinogen component obscured the systematic variations of the protein concentrations due to high degree of scattering. Mild sonication of the aqueous solution helped to improve the solubility of fibrinogen and significantly improved the Raman spectral intensity by minimising scattering effects. Since centrifugal filtration failed to separate fibrinogen from rest of the proteins, ion exchange chromatography had to be applied to separate the fibrinogen by altering its net surface charge. Although ion exchange chromatography is a quick method to separate proteins, this method has to be tailored for a specific protein depending on its charge and cannot be used as a ‘one-for-all’ separation kit for all the proteins. In terms of the applications of Raman Spectroscopy for analysis of blood content, it is therefore recommended that fibrinogen content should be extracted from plasma, and mildly sonicated for quantitation, whereas blood serum can be readily further analysed for High Molecular Weight Fraction (HMWF) and Low Molecular Weight Fraction (LMWF) quantification.

8.5 Serum Fractionation

As long ago as 1999, Berger et al. used near IR light at 830nm to perform Raman microscopic analysis of liquid whole human blood and serum samples to quantify the content of six analytes, namely glucose, cholesterol, triglyceride, urea, total protein and albumin [28]. The total acquisition time per sample was 5 min and the samples were continuously stirred in a quartz cuvette to minimise heating artefacts from high intensity laser. However, no attempts were made to fractionate the serum to deplete the HMWF analytes.
Centrifugal filtration devices have been utilised to improve the sensitivity of quantitative analysis by both Raman and IR spectroscopy, by separating the molecules according to their molecular weight [68]. The proteins that are highly abundant in serum dominate the spectral profile, and by the removal of these proteins (albumin and globulins,) the ability to monitor changes in the lower molecular weight fraction (LMWF) is enhanced. Pre-rinsing of the filter devices with 0.1M NaOH prior to plasma analysis is essential to avoid glycerine interference in the analysis [67]. The optimised washing and rinsing procedure includes spinning 0.5mL 0.1M NaOH at 14000×g for 30 minutes, followed by three rinses with distilled water by spinning 0.5mL distilled water for 30 minutes at 14000×g. Every 30 minute wash and rinse must be followed by spinning the device in the inverted position at 1000×g for 2 minutes, to remove the residual solution contained in the filter. After washing, 0.5mL sample is transferred to the filter and centrifuged at 14000×g for 30 minutes. The solution that flows out from the filter is the filtrate, which contains mostly water and molecules smaller than the pore size of the chosen filter. The remainder of the sample, known as the concentrate, is collected by placing the filter device upside down and spinning for 1000×g for 2 minutes. The resultant concentrate, ~50µL, contains molecules with molecular weight larger than chosen pore size, and is concentrated by a factor of ~10. This indicates the potential for the prediction of other biomolecules that exist within the LMWF with this method, and with further research, such techniques could be translated into the clinical environment as a rapid tool for screening and monitoring.

In the study conducted using 247 blood donors by Rohleder et al. [37], a key was that the prediction accuracy of glucose and urea were improved by up to a factor of 2 by depletion of the HMWF using centrifugal filtration techniques. The processing protocol has been
extensively explored for both ATR-FTIR and Raman analysis [39,68] whereby the HMWF can be isolated in the unfiltered concentrate, while the LMWF filtrate can be further fractionated and/or concentrated. Notably, Bonnier et al. have highlighted the importance of appropriate rinsing of the filters to avoid contamination of the filtrate [67]. In the study of Rohleder et al. [37], Raman spectroscopy was used for quantitative analysis of serum and serum ultrafiltrate with an accuracy, within the range of clinical interest [37], by using 785nm laser as the excitation wavelength and a quartz cuvette to hold the sample. In this measurement setup, a minimum of 200 µL sample was required. 10kDa centrifugal filters were used to deplete the HMWF from the serum and the spectra of glucose, urea and uric acid were recorded from the serum as well as the ultrafiltrate.

Parachalil et al. further demonstrated the suitability of Raman spectroscopy as a bioanalytical tool, when coupled with ultra-filtration and multivariate analysis, to detect imbalances in both HMWF (total protein content, γ globulins and albumin) and LMWF (urea and glucose) of the same samples of human patient serum, in the native liquid form [69]. Using a validated Partial Least Squares Regression (PLSR) method, the γ globulin and total protein analysis models, based on unfiltered patient serum, produced $R^2$ values of 0.88 and 0.82, and Root Mean Square Error of Crossvalidation (RMSECV) of 126 mg/dL and 115 mg/dL, respectively. Post fractionation of the patient serum samples by ultra-filtration using 100 kDa and 50 kDa filters, a similar analysis produced an $R^2$ value of 0.91 and RMSECV of 90 mg/dL for albumin, which is comparable to the values previously reported for a model of aqueous solutions of albumin over a similar concentration range. In the case of urea, $R^2$ and RMSECV values of 0.90 and 70.40 mg/dL for the range of aqueous solutions of varying concentrations were achieved, and 0.92 and 1.73 mg/dL, for the low molecular weight (<10
kDa) filtrate of patient samples, when the full spectral range of 400-1800 cm$^{-1}$ was employed. Reducing the spectral range of the analysis to 800 cm$^{-1}$ to 1030 cm$^{-1}$ considerably improved the prediction accuracy and sensitivity, resulting in an R$^2$ value of 0.97 and RMSECV of 1.14 mg/dL. In the case of glucose, a reduced spectral range from 1030 cm$^{-1}$ to 1400 cm$^{-1}$ was chosen to avoid interference from urea, resulting in an R$^2$ value of 0.84 and RMSECV value of 1.84 mg/dL in the filtrate from the same patient samples. Although both the proof-of-concept studies have been carried out on rather small populations (25 patient samples), they demonstrate that the method has potential for clinical implementation for early disease diagnostics from bodily fluids. In this work, ultrafiltration in conjunction with chemometric methods were used to overcome three problems: first, eliminating the interference from water and β-carotene; second, extracting the Raman signals from LMWF analytes; and third, solving the problems associated with multiple analyte variations in the serum. The sample preprocessing enables a fractionation and concentration of the different molecular weight fractions of the serum, enabling their analysis and quantification without the need for additional enhancement techniques. It has been demonstrated that, using Ag or Au colloids, intense and repeatable spectra are only obtained if the high molecular weight protein fraction is filtered out from the serum [70]. Enhanced signals of the low molecular weight fraction can then be obtained, although the process of addition of colloids would add to the workflow in terms of time and cost.
8.6 Data preprocessing

Data preprocessing, such as spikes removal, baseline correction, smoothing etc., aims to remove any perturbations to the spectra due to any distortions to the spectra which might arise during the measurement process (scattering, electronic noise). This pretreatment of data is critical to accessing the desired information without losing crucial information. The fundamental concepts and basic theory of most used chemometric tools in the pharmaceutical industry for pre-processing, processing and post-processing of the generated data have been detailed in the review article by Sacre et. al [71]. The majority of the commercial software, e.g. Labspec, includes real-time spikes correction, enabling visual inspection of the spectra, and/or manual spike removal when very few spikes are present. Some algorithms have been developed to perform automated spike removal in the case of a large dataset, for which manual correction is not possible [72–76]. The most commonly used smoothing algorithm for de-noising the spectra without losing much information is that of Savitzky-Golay [77]. Drifts in the baseline occurring due to scattering or fluorescence may be corrected using different approaches such as asymmetric least squares [78], mixture models [79], polynomial filter [80,81] and the rubber band method [82]. In the study conducted by Medippally et. al., rubber band baseline correction, Savitzky–Golay smoothing algorithm and vector-normalisation was performed on the raw Raman spectra of liquid plasma prior to post processing, in R based statistical software [50]. In studies aimed at early diagnosis of oral cancer using Raman spectroscopy by Sahu et. al.[83] and differentiation Meningioma by Mehta et. al.,[84], raw liquid serum spectra were baseline corrected using a fifth order polynomial function, smoothed using Savitzky-Golay algorithm and vector normalised in Matlab based statistical software. Fifth order polynomial background subtraction from the
raw serum spectra and normalisation was also implemented in Matlab by Rohdeler et. al. [36] in their study to compare mid-IR and Raman spectroscopy in quantitative analysis of serum. Two different algorithms were used by the same group to perform pre-processing of spectra of serum spectra of ultrafiltrate [37]. Spectra originating from ultrafiltrate were scaled to the area under the water Raman band at 1640 cm\(^{-2}\) after determining the area by a Gaussian fit with linearly decreasing background from 1550 to 1770 cm\(^{-2}\). In the case of serum, spectra were scaled to the maximum intensity and a fifth order polynomial background was subtracted. In their study to detect alterations in glucose and lipid components in the serum using near-IR Raman spectroscopy, Borges et. al.[85] fitted a seventh-order polynomial function over the 400-1800 cm\(^{-1}\) region and subtracted from the raw serum spectrum to remove the unwanted background, providing an effective baseline correction, and spectra were normalised by the area under the curve prior to data analysis. Least squares fitting was used by Berger et al. to remove the background mathematically by subtraction of a sloped straight line from each serum spectrum without affecting the shape of the Raman peaks [28]. Jenkins et. al. subjected the raw Raman spectra of serum to wavenumber standardisation, background subtraction using a rolling circle filter algorithm and normalisation to the peak at 1004 cm\(^{-1}\) [29]. The phenyl alanine peak was chosen for normalisation as it is the sharpest and most intense peak within the serum spectra and it is observed that normalisation to this peak produced better diagnostic discriminatory results when compared to other normalisation methods such as vector normalisation.

Although the accepted dogma is that the contribution of water to Raman spectra is significantly less than to mid-IR spectra, water still has a significant contribution to the Raman signatures in the fingerprint region of bodily fluids, due to the OH stretching vibration.
at 1640 cm\(^{-1}\). Analysis of the ultracentrifugation concentrate reduced the relative contribution of the water signal [68] but further efforts to remove it by data preprocessing have been explored, to increase the relative contribution of the analyte. Kerr et al., [86] have compared a number of commonly employed data pre-processing techniques, and, especially in the case where the background contains a known interferent, such as substrate or background, demonstrate the benefits of the adapted EMSC model, which also contains a polynomial with linear and higher order components. Parachalil et al. adopted an adapted method of Extended Multiplicative Signal Correction (EMSC) [86] to subtract the known spectrum of water from their spectra, initially for the analysis of the HMWF in simulated plasma mixtures [44], and subsequently the measurement protocol was used to quantitatively monitor diagnostically relevant changes of glucose in liquid serum samples (spiked serum samples and patient samples), and the results were compared with similar analysis protocols using infrared spectroscopy of dried samples [43]. The analysis protocols to detect the imbalances in glucose using Raman spectroscopy were first demonstrated for aqueous solutions and spiked serum samples. As in the case of infrared absorption studies [87], centrifugal filtration was utilised to deplete abundant analytes and to reveal the spectral features of LMWF analytes, in order to improve spectral sensitivity and detection limits. After the depletion of the abundant proteins, the dominant water peak from the filtrate collected after centrifugal filtration using 10kDa can be removed by using the EMSC algorithm, and PLSR analysis applied to obtain a prediction model relating the glucose concentrations and the intensity of glucose features. Note, that the study introduced a water normalisation factor into the EMSC protocol, using the fitted co-efficient of the water scaling factor, which has the effect of scaling the analyte spectra, assuming a constant water contribution to all sample spectra. This
step helps to considerably reduce the spectral variability of the Raman spectra of glucose recorded from 25 patient samples (Figure 3) [44].

The principle of EMSC for subtraction of a specific measurable background spectrum and the associated Matlab codes have previously been published by Kerr and Hennelly, 2016 [86], and their description is adapted in the following. The raw spectrum, S, consists of Raman spectrum of interest, R, a baseline signal, B, and the water signal, W.

\[
S = R + B + W \quad [86] 
\]

(1)

The Raman spectrum of interest can be represented by a reference spectrum of the material of interest, r, and it can be assumed that R is the product of this reference spectrum and a certain scalar weight, \( C_r \), which describes the concentration dependence [88,89]

\[
R \sim C_r \times r \quad [86] 
\]

(2)

Similarly, a spectrum, w, is recorded from water directly in order to represent the spectral contribution of water in W, as the product of pure water spectrum and a certain scalar weight.

\[
W = C_w \times w \quad [86] 
\]

(3)

The baseline, B, is now represented by an appropriate order of polynomial (N) as:

\[
B_N = C_0 + C_1X + C_2X^2 + \ldots + C_NX^N \quad [86] 
\]

(4)

where \( N \) is the order of polynomial and \( C_m \) for \( m = 0 \rightarrow N \) represents various coefficients of polynomial. The EMSC algorithm is used to obtain estimates of the scalar values \( C_r, C_m \) and \( C_w \). These estimates are obtained from an optimal fit of the various vectors in Equation 5.

\[
S \sim [C_r \times r] + [C_w \times w] + [\sum_{m=0}^{N} C_mX^m] \quad [86] 
\]

(5)
The background corrected, concentration dependent analyte spectra, \( T \), can be represented as:

\[
T = \frac{S - \left[ C_w \times w \right] - \left[ \sum_{m=0}^{N} c_m x^m \right]}{C_w} \tag{6}
\]

Note, that division by \( C_w \) has the effect of scaling the analyte spectra, assuming a constant water contribution to all sample spectra.

**Figure 8.3.** (A) PCA scatter plot of Raman data of the filtrate obtained after ultrafiltration of patient serum without scaling the analyte (glucose) spectra from the previously published study by Parachalil *et. al* [44] and (B) after scaling the analyte spectra to the water content. Figure 3B displays less scatter when compared to Figure 3A, indicating less variability among the spectra.

However, although the EMSC algorithm removed the underlying water spectra effectively, interference from other LMWF analytes, namely, urea was noticed. Hence, a shorter spectral range from 1030cm\(^{-1}\) to 1400cm\(^{-1}\) was chosen for data analysis, as this region does not contain signature peaks of urea. Improved Root RMSECV was observed for Raman prediction.
models, whereas slightly higher $R^2$ values were reported for infrared absorption prediction models.

The adapted EMSC algorithm can also be employed to remove known spectral interferents. In several studies of human serum, β-carotene has been observed as a string contributor to the spectrum, particularly when using 532nm or lower as source, as the scattering from the conjugated antioxidant species are resonantly enhanced [29,50,83]. The LMWF species are not easily removed by centrifugal separation [69], but their contributions can be effectively “digitally” removed using the adapted EMSC protocol.

8.7 Data Analysis

Data analysis protocols can be differentiated into classification protocols, largely employed for diagnostic purposes, and regression protocols, predominantly used for quantitation of identified analytes. The latter have been explored for both imbalances in intrinsic blood constituents, as well as for therapeutic drug monitoring.

Depciuch et al. collected Raman and FTIR spectra from dried and liquid serum from depressed patients and their analysis indicated that both methods provided equally valid results to discriminate the patients with similar accuracies [33]. The light source wavelength was 780 nm and the sample volume used was 1.5ml, which could be an impediment for clinical translation. Other similar studies conducted by the same group using the same set up on phospholipid-protein balance in human serum and qualitative, quantitative changes in phospholipids and proteins in animal depression models. The role of zinc deficiency induced phospholipid-protein imbalance in serum of the animal models suggest that both IR and Raman spectroscopic techniques could be used as effective tools to identify the changes in
the blood serum [30–32]. To investigate similarities and differences between the serum samples of different types of depression in humans, Principal Components Analysis-Linear Discriminant Classification (PCA-LDC) was employed in their analysis of Raman and FTIR spectra from dried and liquid serum from depressed patients [33]. The results from both FTIR and Raman spectra unambiguously demonstrated that the levels of proteins and phospholipids are higher in healthy controls than in depressed subjects and that phospholipids affect the structure of proteins. Two measurement strategies were compared in order to determine the influence of water on the measured spectra; the first method entailed recording the water spectrum as back-ground, and subsequently automatically subtracting it from each serum spectrum, while the second method entailed recording a spectrum of the air prior to the spectrum of blood serum and subsequently the water spectrum was subtracted from the blood serum without air background. Both the methods provided identical serum spectra, suggesting that appropriate measurement of the background and the subtraction of water signal had the greatest impact on the reliability of the results.

Jenkins et al. developed a high-throughput (HT) serum Raman spectroscopy platform and compared dry and liquid data acquisition of serum samples for liquid biopsy of 30 colorectal cancer patients and 30 matched control patients [29]. Using a stainless steel high throughput substrate that allows up to 40 samples to be loaded at once, and 785 nm laser as the excitation wavelength, the maximum sensitivity and specificity obtained for discrimination of colorectal cancer patients were 77% and 81% respectively. In this study, Raman spectra were subjected to routine preprocessing such as wavenumber standardisation, background subtraction using a rolling circle filter algorithm and normalisation to the peak at 1004 cm$^{-1}$. Partial least squares discriminant analysis (PLS-DA) was used to investigate causes of
differences and variances within datasets. PLS-DA models were cross validated using k-fold cross validation with 5 folds. The sample volume required was 200 µL and the total time for the data collection was 12.5 min. In a clinical setting, it would be advantageous to reduce the data collection time in order to avoid delay in analysis. When excitation wavelength of 532 nm was used on the liquid platform, the specificity was found to be low, due to the prominence of carotenoids in the serum spectra. No measures were taken to remove the interference of carotenoids from the spectral data. The prominence of carotenoids due to resonance Raman was also reported by Sahu et al., in an exploratory study for detection of oral cancers [83] and by Mehta et al., in their study conducted to differentiate meningioma. Sahu et al. used 532 nm as excitation source and Raman spectra were collected by placing 30 µL serum samples on a glass slide and the laser was focused through a 50X Nikon objective. The spectra were vector normalised, baseline corrected, smoothed and subjected to PCA-LDA followed by cross-validation using leave-one-out cross-validation (LOOCV).

Although this study reported spectral differences between DNA, changes in the plasma amino acid profiles and β-carotene levels across the analysed groups, the strong bands of β-carotene could interfere with the detection of other analytes increasing the ambiguity in the results obtained. In the study conducted by Mehta et al., 25 patient samples from healthy and meningioma groups were subjected to PC and PC-LDA followed by LOOCV cross validation, yielding classification efficiency of 92% and 80% for healthy and meningioma respectively. Passively thawed 30 µL serum samples, on a calcium fluoride (CaF₂) slides were subjected to Raman analysis using 785 nm laser excitation source. Borges et al. [85], collected blood serum from 44 volunteers to discriminate between altered and normal concentrations of glucose, total cholesterol, triglycerides, low density (HDL) and high density lipoproteins.
(LDL). Raman analysis was performed using 830nm excitation source for a sample volume of 100µL. The data collected was subjected to PCA yielding a classification efficiency of 77% for total cholesterol, 81% for triglycerides, 59% for HDL and 60% for LDL.

In terms of quantification of systematically variable analyte concentrations, multivariate PLSR analysis is often the method of choice. PLSR is most commonly employed to construct a model that can relate variations of the measured spectral responses to a systematic variation of concentrations of the target analyte [90–92]. The constructed model can then be employed to identify spectral factors which account for the maximum variation in predictors ‘X’ (spectral data) versus associated responses ‘Y’ (target values of protein concentration) [91]. The spectral data (X matrix) can thus be related to the target concentrations (Y matrix) according to the linear relationship $Y = XB + E$, in which B and E are matrices of regression coefficients and residuals, respectively. The PLSR model can be used to predict the outcome of varying concentration of analytes based on the spectral data. The model is validated using a rigorous cross-validation procedure which evaluates its performance in accurately predicting analyte concentrations [44]. In the study of Berger et al., PLSR prediction models were built, using leave one out cross validation (LOOCV), for each of the six analytes after spectral background removal, providing an root mean square error of 26mg/dL, 12mg/dL, 29mg/dL, 3.8mg/dL, 0.19mg/dL and 0.12mg/dL for glucose, cholesterol, triglyceride, urea, total protein and albumin, respectively [28]. A similar approach was employed by Parachalil et al. [43] to compare the predictive capacity of Raman microscopy for quantitation of glucose in human serum, compared to the similar ATR-FTIR analysis of Bonnier et al. [87] indicating that the RMSECV for Raman of 1.84mg/dL is comparable or better than that of 3.1mg/dL for ATR-FTIR.
Having removed the spectral interferents using EMSC, PLSR can independently identify spectral correlation associated with target analyte range. Cross-validation of the PLSR model is required to evaluate the accuracy, often cited in terms of the Root Mean Square Error of Cross Validation (RMSECV) [93]. The number of latent variables used for construction of the PLSR model is optimised by establishing the value that is equivalent to the minimum of the RMSECV. The $R^2$ value provides an indication of the correlation between the analyte concentration and spectral intensity, while the standard deviation (STD) provides an indication of the variation between each spectrum calculated from the same sample. Commonly, a multiple fold cross validation approach is employed to validate the robustness of the method. Typically, the set of observations is randomly divided into approximately equal size, e.g. 50% of the spectral data randomly selected as test set, while the remaining 50% is used as the training set [101]. The cross-validation process is then repeated multiple times (the folds), such that all observations are used for both training and testing, and each observation is used for testing exactly once. The results from the folds can then be averaged to produce a single estimation. The Root Mean Square Error of Cross Validation (RMSECV) is calculated from the multiple iterations to measure the performance of the model for the unknown cases within the calibration set.

In order to advance the studies of Berger et al., [28] and Rohleder et al.,[36,37] using the improved inverted microscopy modality, Parachalil et al.[69] also used the EMSC correction, ultrafiltration protocols and PLSR analysis to construct models for the quantitation of the whole serum (total protein, γ globulin), serum concentrate (albumin) and serum filtrate (urea and glucose) of patient samples, resulting in higher accuracy and sensitivity of analysis. The
strategy demonstrated in this study enables the simultaneous estimation of total protein level and detection of imbalance in γ globulin concentration accurately from whole serum, without the use of any reagents and without destroying the sample being studied. The proposed method has many advantageous over the routinely used biuret method, Turbidimetric Immunoassay (TIA) and Radial Immunodiffusion (RID), as the required sample volume can be as low as 10µL, it is rapid and non-destructive to the medium being studied, whereas the conventional methods are considered impractical due to the requirement of large sample volume and laborious sample processing steps [94–96]. Moreover, the linearity reported for a comparative study of RID and TIA is 0.59 [97], substantially lower than the $R^2$ calculated by Parachalil et al. [44]. Using filtration, the albumin was isolated from patient serum and the prediction model had a prediction accuracy significantly superior to that reported for the Bromoscerol Green method (2.2 g/dL), used to determine albumin concentrations from cirrhotic patients [98]. It has previously been reported that selecting the spectral region from 1030 cm$^{-1}$ to 1400 cm$^{-1}$ improved the sensitivity and specificity for the prediction model of glucose from patient samples over the concentration range 52.5-434.2 mg/dL, and the technique was demonstrated to be at least as accurate as ATR-FTIR of similar patient samples [69], measured in the dried state and closer to the accuracy of colorimetric methods, 1.4 mg/dL urea [99] and 2 mg/dL glucose [100]. Similarly, higher prediction accuracy (RMSECV=1.14 mg/dL) was attained when PLSR analysis was performed on a reduced range for urea from patient samples over the concentration range 2.52-78.99 mg/dL, compared to the full range (RMSECV=1.73 mg/dL).

This standardised, optimised methodology was also applied to determine the Limit of Detection (LOD) and Limit of Quantification (LOQ) for therapeutic drug monitoring (TDM)
in human serum, using the examples of Busulfan, a cell cycle non-specific alkylating antineoplastic agent, and, Methotrexate, a chemotherapeutic agent and immune system suppressant [101]. Ultrafiltration of the spiked human pooled serum with 10kDa centrifugal filter efficiently recovered the drug in the filtrate prior to performing Raman analysis. The drug concentration ranges were chosen to encompass the recommended therapeutic ranges and toxic levels in patients. Finally, prediction models were built by using PLSR and LOD and LOQ were calculated directly from the linear prediction models. The LOD calculated for Busulfan is 0.0002 ± 0.0001 mg/mL, 30-40 times lower than the level of toxicity, enabling the application of this method in target dose adjustment of Busulfan for patients undergoing, for example, bone marrow transplantation. The LOD and LOQ calculated for Methotrexate are 7.8 ± 5 µM and 26 ±5 µM, respectively, potentially enabling high dose monitoring. Although SERS gives promising results in detecting drugs at low concentrations in biological matrices [16,102], qualitative variations within the SERS substrate, and the interference of other biomolecules with the SERS spectra, makes quantification in real samples a challenging task [103]. The simpler approach, adopted here, of Raman spectroscopy coupled with concentration of the high and low molecular weight serum fractions using commercially available centrifugal filter devices and multivariate analysis technique ensures that the information in the fractions is effectively preserved, while enabling easy detection of the analyte concentration with higher accuracy that in the unprocessed sample. Better analysis of serum using Raman spectroscopy was observed when the sample was analysed in the inverted geometry using the water immersion objective with a 532nm laser as source with Lab-Tek plate as substrate [44]. A drop of water is used to minimise the differences in the refractive indexes between sample, objective and the substrate. However, the water drop does
not contribute to the data collected, as it is outside the focus of the beam. The promising results from the systematic studies conducted by Parachalil et al.[69] using this analytical set up combined with chemometric techniques such as EMSC to measure concentrations of HMWF, LMWF and drugs in human serum strongly indicates a highly significant correlation between predicted and reference concentrations. These results suggest that Raman can on its own can detect and quantify concentrations close to 1% without the aid of any enhancement method [37].

Considering the more general applications of the technique to a broader range of drugs or analytes, it is notable that both the LOD and LOQ for a given set up are determined by the STD of the measurement of the control (serum or filtered serum), and the Raman scattering efficiency of the analyte. To further illustrate this point, the LOD of vitamin B12, cholesterol, urea and glucose from liquid samples for this measurement protocol were also determined using the method previously reported by Parachalil et al. [101] (Table 1). Figure 4 displays a plot of LOD versus the maximum intensity per unit acquisition time, per unit concentration. An approximately linear, inverse relationship can be seen, which simply emphasises the maxim that the stronger the Raman signal of the analyte, the easier it is to detect. The correlation also implies, however, that a simple measurement of the analyte in aqueous solution could be employed to predict the LOD, and therefore the suitability of the technique for the therapeutic monitoring application.
Table 8.1: LOD of glucose [44], busulfan [101], methotrexate [101], cholesterol, urea [69] and vitamin B12 calculated from the PLSR prediction plot of these analytes, compared to the maximum Raman intensity of maximum peak per unit acquisition time, per unit concentration.

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Intensity (arb. units)</th>
<th>LOD (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>1000</td>
<td>0.0006±0.0005</td>
</tr>
<tr>
<td>Busulfan</td>
<td>19704</td>
<td>0.0008±0.0001</td>
</tr>
<tr>
<td>MTX</td>
<td>1600</td>
<td>0.00076±0.0005</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>216718</td>
<td>0.0006±0.0001</td>
</tr>
<tr>
<td>Urea</td>
<td>540000</td>
<td>0.00033±0.00001</td>
</tr>
<tr>
<td>VB12</td>
<td>1266666</td>
<td>0.00014±0.00002</td>
</tr>
</tbody>
</table>

Figure 8.4. LOD of glucose (44), urea (69), busulfan (101), methotrexate (101), cholesterol, and vitamin B12 calculated from the PLSR prediction plot of these analytes, compared to the
‘maximum Raman intensity of maximum peak per unit acquisition time, per unit concentration’. The methodology used for calculating LOD was previously published (101)

8.8 Clinical Translation

Vibrational spectroscopic techniques have attracted a lot of attention as analytical methods of choice for analysing biological samples. The promise of the techniques is based on the abilities to objectively fingerprint the biochemical profile underlying early onset of disease in cells, tissues or bodily fluids [3,6,9]. The techniques have no requirement of specific reagent, and the minimal sample preparation means there should be fewer procedural errors and no major time delay in providing the results to the patients. In the case of FTIR, considerable progress towards using this technique as a routine diagnostic tool for disease diagnosis has been made. Recently, ClinSpec Diagnostic, based at the University of Strathclyde’s Technology & Innovation Centre, won a £1.2 million investment to further develop a “revolutionary” blood test using FTIR which could potentially improve brain cancer survival rates through early detection of the disease [104]. Researchers from Monash University have similarly developed FTIR based techniques to analyse disease-causing pathogens in human blood [105]. Advances in instrumentation have made FTIR an ideal tool of choice for an increasing number of clinical application even though the sample drying step is potentially a huge drawback [19,68,106].

Notably, it is demonstrated that the Raman analysis protocol can yield accuracies which are comparable with those reported using infrared absorption based measurements of dried serum, without the need for additional drying steps. However, there remain challenges to fulfilling the requirements for clinical translation. Distortions in the spectra due to the water contributions, low detectability of the low molecular weight fraction analytes, lack of
standardisation steps and financial factors could be the hindrance for clinical translation in the medical and clinical environment. Nevertheless, new strategies have been recently developed to address these potential limitations, namely the inverted Raman geometry, low cost (Lab-Tek plate) substrates, serum fractionation techniques, selective spectral region analysis and sophisticated data preprocessing (EMSC) and analysis (PLSR) techniques. These advances have maximised the diagnostic accuracy and are cost-effective solutions that are likely to be adopted in a clinical setting.

Application of the adapted EMSC based algorithm helped to eliminate the contribution from water and scattering associated with the HMWF, and linear predictive models were built from the PLSR analysis. It is worth noting that PLSR seeks to correlate systematic variations in the spectroscopic profiles with the external target variable, in this case protein concentration. As such, the method is not inherently specific to any molecular variations, in the case where multiple species vary simultaneously over the same range. The accuracy of the proposed method is comparable to that of the most commonly used method for detecting albumin from biological fluids, the Enzyme Linked Immunosorbent Assay (ELISA) [107,108], and the most commonly used gold-standard method i.e., the Clauss assay for fibrinogen [109]. The proposed approach can be expeditiously employed for early detection of pathological disorders associated with high or low plasma/serum analytes.

Bodily fluids are usually collected from a large number of patients in a hospital, potentially delaying the performance of the analysis and availability of results, which may in turn delay the therapy, and prolong patient anxiety. The accuracy of the conventional test kits that enable point-of-care testing can be poor, and they are can be avoided due to high cost [110,111]. The enormous advantage of employing this inverted Raman spectral analysis for
real-time serum/plasma analysis in a clinical setting is that, in the face of emerging diseases, one could get an early diagnosis in a cost effective manner before the disease becomes symptomatic (Figure 1). Integrated with smaller spectroscopic instrumentation, implementation of such a screening/diagnostic tool in developing countries where conventional diagnostics methods are scarce and avoided due to high cost, would bring about a dramatic impact on population screening.

8.9 Conclusion

The potential advantages of employing Raman spectroscopy of bodily fluids for disease diagnosis apart from its high sensitivity and specificity are: low cost, no reagents are required, samples can be probed in the native state and rapid results can be obtained. Although there is a wealth of information to support the application of Raman analysis for bodily fluids, this approach is still considered young and upcoming in the eyes of medical community, thereby hindering its clinical translation. The examples summarised in this article attest to the cost effective, easy to use, reproducible method of Raman measurement protocol for detecting imbalances in serum/plasma proteins and low molecular weight analytes, as well as therapeutic drug monitoring. Significantly improved sensitivity of Raman spectroscopic measurements of blood samples in liquid form is achievable by means of serum fractionation and the implementation of chemometric approaches. In the case of clinical translation, there remain challenges to overcome before it can be a reality. Nevertheless, it has been demonstrated that options and alternatives are available to overcome these challenges in using Raman spectroscopy for liquid sample analysis, leading to a better accuracy and repeatability and thus a better sensitivity. Naturally, the next phase includes not only refining
this method and improving the technical capabilities to match those of the current clinical needs, but also technical advances to translate them from research laboratory to clinical practice. In addition, to ensure further relevancy of this method, comparison studies between the gold standard of current diagnostic methods and the current method in a large multi-centred randomised clinical trials are required. Once these factors are taken into account, it is possible to envisage a routine platform providing clinical biochemical analysis at minimal cost.

8.10 References


Vib Spectrosc. 2018;99:50–8.

diagnosis of Alzheimer’s disease using spectrochemical analysis of blood. PNAS.
2017; 114 (38)

15. Yang J, Tan X, Shih W, Cheng MM. A sandwich substrate for ultrasensitive and
label-free SERS spectroscopic detection of folic acid / methotrexate. Biomed
Microdevices. 2014;16(5):673-9

16. Fornasaro S, Marta D, Rabusin M. Toward SERS-based point-of-care approaches for

al. Ultrasensitive SERS Substrate for Label-Free Therapeutic-Drug Monitoring of
Paclitaxel and Cyclophosphamide in Blood Serum. Anal Chem. 2019;91(3):2100-
2111.

rapid diagnosis of gliomas from serum samples using infrared spectroscopy and

19. Cameron JM, Butler HJ, Palmer DS, Baker MJ. Biofluid spectroscopic disease
diagnostics: A review on the processes and spectral impact of drying. J

20. Lord RC, Yu NT. Laser-excited Raman spectroscopy of biomolecules. I. Native

21. Tobin MC. Raman spectra of crystalline lysozyme, pepsin, and alpha chymotrypsin.


45. Leal LB. Photodiagnosis and Photodynamic Therapy Vibration spectroscopy and body bio fluids: Literature review for clinical applications. Photodiagnosis
Photodyn Ther. 2018;24:237–44.


58. Vaught JB, Henderson MK. Biological sample collection, processing, storage.


69. Parachalil DR, Bruno C, Bonnier F, Chourpa I, McIntyre J and Byrne HJ. Raman


88. Kohler A, Kirschner C, Oust A, Martens H. Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical


Controversies in the determination of serum albumin concentration in chronic liver

99. Smolcic S. Validation of methods performance for routine biochemistry analytes at

100. Luque-Garcia JL, Neubert TA. Sample preparation for serum/plasma profiling and
biomarker identification by mass spectrometry. J Chromatogr A. 2007;1153(1–
2):259–76.

101. Parachalil DR, Commerford D, Bonnier F, Chourpa I, McIntyre J, Byrne HJ. Raman
spectroscopy as a potential tool for label free therapeutic drug monitoring in human

102. Hidi IJ, Mühlig A, Jahn M, Liebold F, Cialla D, Weber K and Popp J. LOC-
2014;6:3943-3947

103. Fornasaro S, Marta D, Rabusin M. Toward SERS-based point-of-care approaches for

104. Burley H. Strathclyde spin-out gets £1.2m to advance brain cancer detection test.
The Scotsman [Newspaper on the internet]. 2019 May 01. Available from :
https://www.strath.ac.uk/whystrathclyde/news/nationalawardforcancerdiagnosisspin
outcompany/

105. Dropulich S. New device that can rapidly diagnose disease could save lives. Monash
University [Internet]. 2019 January 14th. Available from:
https://www.prnewswire.com/news-releases/new-device-that-can-rapidly-diagnose-


Chapter 9

Conclusion

Through a systematic assessment of the analytical conditions, comparing Raman spectroscopy to FTIR, use of the 785 nm laser line to the 532 nm laser line as Raman source, use of Lab-Tek plate to other substrates, upright to inverted geometry of Raman, and data preprocessing techniques, the settings and techniques used for liquid serum analysis were tailored to optimise the quality of the results generated in the study. It is observed in this study that inverted immersion Raman spectroscopy employing a 532 nm laser and Lab-Tek plate substrate generated spectra of better overall quality, with higher signal-to-noise ratios. Hence, these methods and settings were used for further investigation into the diagnostic capabilities of Raman spectroscopy.

The results from the study conducted to detect imbalances in high molecular weight plasma proteins from simulated plasma show that extraction of the fibrinogen from the protein mixture solution by ion exchange chromatography is more specific than ultra-filtration, allowing the quantification of variations of fibrinogen levels [1]. It was observed that mild sonication could be applied to increase the solubility of the fibrinogen. EMSC successfully removed the background associated with the scattering as well as the water signal, and a PLSR prediction model was built, correlating the systemic variations in the spectral profile with that of the high molecular weight protein concentrations. In general, the scattering problems caused by fibrinogen favour the use of blood serum for the biochemical analysis of
analytes other than coagulants. However, to further ensure the relevancy and consistency of these results, experiments need to be carried out in human plasma from healthy and patient samples. An alternative study should be conducted using Raman optical activity to ensure ultrasonication does not disintegrate the protein at lower frequencies.

The comparative study conducted between Raman and FTIR using glucose as a model analyte demonstrated that the Raman analysis protocol can yield accuracies which are comparable with those reported using infrared absorption based measurements of dried serum, without the need for additional drying steps [2]. It is noteworthy that Raman spectroscopy yields significantly lower values of RMSECV for all the Raman prediction models, suggesting higher sensitivity and accuracy and the suitability of this technique to discriminate patients with very similar concentrations of blood glucose. However, lower R² values and higher standard deviations were observed for Raman spectroscopy, which could be due to the intrinsic variability of individual samples, depending upon the physiological state of the individual on that day. Hence, the study should ultimately be conducted on a larger number of patient samples to further confirm the suitability of Raman as a biochemical tool for screening LMWF analytes in liquid serum samples.

The potential of Raman spectroscopy as a bioanalytical tool to simultaneously detect imbalances in the high and low molecular weight fraction of human serum has been demonstrated [3]. In this study, EMSC was employed to remove water and spectral interferents such as β-carotene that can interfere with the spectra of target proteins. It has also been demonstrated that selecting spectral regions where the bands of the target protein are stronger leads to higher specificity of the model and allows detection of multiple analytes from the same dataset. The strategy illustrated in this study enables simultaneous detection
of various analytes (total protein content, γ-globulin, albumin, urea and glucose) from human serum using Raman spectroscopy with minimal sample preparation, no labelling and no additional sample drying steps, with accuracy comparable to that of the conventional gold standard methods. Depletion of the HMWF is a non-optional step in the investigation of protein imbalances or disease related biomarkers in the LMWF of serum. Although this study provided promising results for the detection of total protein content, albumin, urea and glucose, the spectral co-efficient of prediction of γ-globulin is somewhat unsatisfactory with no prominent features of γ-globulin. Raman optical activity (ROA) could be a viable option to probe structural differences of globulins due to its stereochemical potential and could yield structural information not otherwise available [4], [5]. ROA and Tip enhanced Raman spectroscopy (TERS) have been proven to be effective techniques for quantification of glycosylated status of target proteins [6]-[8].

Further improvements in the sensitivities and reduced variabilities of the techniques ultimately rely on the reproducibility of the measurement, instrument calibration, using water as an internal standard and improved signal to noise ratio. The proposed Raman methodology could potentially be improved by integrating an automated focussing and sampling methodology, and regular maintenance of the instrument. As longer accumulation times to reduce the noise are not recommended, because of sample evaporation and also speed of measurement throughput, optimisation of instrumentation for higher signal throughput could be explored, for example by sacrificing spectral resolution.

This simple approach of Raman spectroscopy coupled with ultra-filtration, multivariate analysis has been extended to therapeutic drug monitoring of Busulfan and Methotrexate from humans serum and a method to calculate LOD, and LOQ with higher accuracy was
demonstrated [9]. Further studies are needed to investigate the determination of these drugs in patient serum to ensure successful implementation of this method as a diagnostic tool. The proof of concept shown in this study gives rise to the hope of utilising Raman spectroscopy as a biochemical tool for monitoring a variety of drugs and small molecules from bodily fluids.

Ultimately, the studies presented in this thesis are transferable to any other low and high molecular weight analytes or therapeutic drugs. The studies shown in this thesis attest to the cost-effective, easy to use, reproducible method of Raman measurement protocol for detection of imbalances in serum/plasma proteins, as well as therapeutic drug monitoring. Since Raman spectroscopy has established itself as a reliable and non-destructive technique, this proposed methodology can also find application as a rapid drug-screening tool for athletes or in forensic science, where the sample size may be extremely small and needs to be preserved. Several factors such as improving accuracy, consistency and standardised sample collection and measurement procedure still must be addressed before clinical translation as biofluids are biochemically heterogeneous. Once these factors are taken into account, it is possible to envisage a point-of-care routine biochemical-fingerprinting platform, a truly revolutionary step in biomedical science.
9.1 References


Appendix 1: Publications


2. Parachalil DR, McIntyre J, Byrne HJ. Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: Recent advances, Analytical and Bioanalytical Chemistry, 2019 (Submitted)


Appendix 2: Conferences, Modules and Collaboration

Conferences:

- Raman4Clinics, February 2016 – Poster presentation
- Faraday’s discussion, March 2016 – Poster presentation
- SPEC, June 2016 – Poster presentation
- CLIRSPEC Summerschool, July 2016
- CLIRCON, April 2017 – Poster presentation
- Raman4Clinics, July 2017 – Poster presentation
- SPEC, April 2018 – Poster presentation
- Symposium on Vibrational spectroscopy at Université de Tours, France, March 2019 – Oral presentation

Modules Taken:

- Advance course in R programming
- Programming in Matlab
- Learning, Teaching and Assessment
- Pedagogical Practice

Collaboration:

Secured the The Irish-French PHC Ulysses 2018 funding by Irish Research Council and Campus France with participation of the French embassy in Ireland to collaborate with Nanomédicaments et Nanosondes group in Université de Tours, France and Pure and Applied Chemistry group in University of Strathclyde